Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia



First published 1982
Reprinted 1985

Printed in Great Britain at the University Press, Cambridge

Library of Congress catalogue card number: 81–17070

*British Library cataloguing in publication data*
Cairns-Smith, A. G.
Genetic takeover and the mineral origins of life.
1. Evolution 2. Genetics
I. Title
575.1 QH430
ISBN 0 521 23312 7

# Contents

*To Dorothy Anne*



# Preface

The main idea in this book is that the first organisms on Earth had an altogether different biochemistry from ours – that they had a solid-state biochemistry. This idea was already in the opening sentence of the abstract of my first paper on the origin of life (1966): 'It is proposed that life on Earth evolved through natural selection from inorganic crystals'. It sounds an odd idea, but I meant it literally. I still do.

Other ideas have changed somewhat, especially on the question of where the first organic molecules came from. I still held to the usual view in my earlier book (1971) that the original source of organic molecules in organisms was a 'primordial soup' built up through non-biological processes. In writings since then (see References) I have moved increasingly away from this towards a more old fashioned notion that photosynthesis from atmospheric carbon dioxide has always been the source of organic molecules in organisms. In this book I suggest furthermore that the relevant photosynthesis was always biological; that organic molecules *in* organisms were never made other than *by* organisms – that in this respect anyway, life on Earth has always been as now.

A log-jam of objections has to be cleared before such a point of view can begin to make sense. I have to persuade you that the very first organisms need not have contained any organic molecules. I have to reinterpret results of experiments purporting to demonstrate that some of our present-day biochemicals would have been made by non-biological processes on the primitive Earth. I have to provide an account, at least in principle and in as much detail as possible, as to how early evolutionary processes could have transformed organisms with one kind of central control machinery into organisms based on control structures of an altogether different kind. (This is genetic takeover.) And I should explain too *why* evolution should

have had such devious beginnings, and what kinds of experiments and observations can best help us to understand the likely nature of that deviousness.

I do not pretend to have written at a particular level. I doubt if this would be feasible for a book of moderate length on a subject that touches on so many fields and which often requires detailed discussion of particular topics. Any reader is bound to find bits of it difficult (as I did). But if you have a broad interest in science; if you enjoy reading, say, *Scientific American*, then I think that the main ideas and arguments should be accessible. To help in this I have included a fair amount of review material and textbook material.

I have many people to thank. First of all my wife and family for putting up with my frequent abrupt absences ('he's writing his book') and also for help in reading drafts, etc. Many colleagues, especially in the Chemistry Department here, have contributed to ideas developed in this book – with their enthusiasm or scepticism: I could hardly have asked for a wider range of chemical expertise immediately around me. I should mention particularly research students who have worked with me on topics related in some way to the problem of the origin of life – Donald Mackenzie, George Walker, Douglas Snell and Chris Davis. They helped me to keep in view the practicalities of chemistry – an important recurring theme in this book. For example, Chris Davis' single-minded attempts to make a replicating organic polymer revealed, more clearly than any thought-experiment, the mountains of difficulty that lie in that approach to the origin of life.

I have been very fortunate in being able to find experts in different areas who were willing to read and comment on chapters of the manuscript – in several cases they read the whole book. I would like to thank particularly Paul Braterman, Colin Brown, John Carnduff, Alan Cooper, Charles Fewson, Jim Lawless, Karl Overton, Neil Spurway and Jeff Wilson. I usually took their advice, but not quite always. In any case only I am to blame for the mistakes and misconceptions that you may find.

The whole task of writing was greatly helped by two other experts, Janet McIntyre who word-processed the words (again and again), and Dougal McIntyre who provided the necessary hardware and software.

The electron micrographs are an important part of this book and I am particularly grateful for the prints that were provided by Mr Alan Craig and by Drs Baird, Dixon, Fryer, Gildawie, Güven, Horiuchi, Keller, Kirkman, Mackenzie, Morimoto, Mumpton, Posner, Rautureau, Weir, Wilson and Yoshinaga.



Permission by authors to use material for line drawings (as indicated in figure legends) is gratefully acknowledged as are the permissions granted by the following publishers to redraw, reproduce or otherwise use material for the figures and tables indicated:

Academic Press Inc. (figures 1.8, 6.1, 8.13).
Akademische Verlagsgesellschaft (figures 6.17, 6.20, 6.25, 6.35, 6.45*a*).
The American Chemical Society (figure 1.11).
The American Institute of Physics (figures 6.5, 6.7, 6.9).
Butterworth and Co. Ltd (figure 6.14).
*Canadian Journal of Chemistry* (figure 6.12).
Chapman & Hall Ltd (figure 9.4).
The Clay Minerals Society (figures 6.26, 6.29, 6.32, 6.34, 6.48, 6.52, 6.55, 8.5, 8.6).
Cornell University Press (figures 6.21, 6.23).
Elsevier Scientific Publishing Co. (figures 6.46, 8.4, 8.9).
Elsevier Sequoia S.A. (figure 8.14).
The International Union of Crystallography (figure 7.2*b*).
Japan Scientific Societies Press (Table 1.5).
Johann Ambrosius Barth (figure 6.13).
Longman Group Ltd (figures 6.5, 6.7, 6.22, 6.44).
Macmillan Journals Ltd (figures 6.24*a*, 6.50).
The Mineralogical Society of America (figures 6.31, 6.45*b*, 6.51, 7.2*d*, 7.8).
The Mineralogical Society of London (figures 6.28, 6.30, 6.33, 6.38, 6.57, 7.2*c*, 8.7).
North-Holland Publishing Co. (figure 1.2).
Oxford University Press (figure 6.16).
Pergamon Press Ltd (figures 5.3, 5.8, 6.26).
D. Reidel Publishing Co. (figure 1.5, Table 1.1).
The Royal Society of Chemistry (figure 6.24).
The Royal Swedish Academy of Sciences (figure 7.2*a*).
The Society of Economic Paleontologists and Mineralogists (figure 6.54).
Springer-Verlag (figure 6.47, Table 1.2, 1.4).
Wiley Interscience (Table 6.1).
John Wiley and Sons Inc. (figure 6.15).

Graham Cairns-Smith
Glasgow, Summer 1981

# Preview

*Evolution, organism, life* Let us start from Dobzhansky's (1973) assertion that 'nothing in biology makes sense except in the light of evolution'. **Evolution** is then the central idea and other ideas in biology are to be related to it.

Here evolution is not simply to be taken as any long-term trend or change (as we might talk of the evolution of the Universe or of the Earth's atmosphere). Biological evolution is a much more specific idea. It takes place through successions of physicochemical systems of a particular sort – systems that can reproduce and pass on specific characteristics to offspring. Such systems are **organisms**. What makes them special, what explains them, is precisely that they participate in evolution. A living organism is the latest link in a chain stretching far back into the past. Successions of organisms, and only successions of organisms, can be subject to natural selection and hence become adjusted to their environments in the kind of way that Darwin described.

Other ideas follow. What has to be passed on between generations of organisms, to maintain the line, cannot in the long term be a material: it can only be information if it is to be transmissible indefinitely. And what evolves is not the organisms exactly, but rather the information which they transmit and which determines their specific characteristics – **genetic information** it is called. Evolution depends absolutely on the existence of the means to transmit genetic information so that characteristics of parents can reappear in offspring: evolution depends on **heredity**.

All the same there must be material vehicles for genetic information. As we will discuss in Chapter 2, it looks as if the only chemically feasible way of transmitting more than trivial amounts of information would be through structures that, like DNA, replicate by some kind of templating mechan-

ism. We can call such things **genes** without implying that they necessarily have to be made of DNA or anything at all chemically similar to this, our present **genetic material**.

**Mutations** are also needed for a Darwinian evolution although they represent little difficulty in principle: they fall into the category of *mistakes*. Genetic information must sometimes change to produce altered characteristics, and these alterations must be inheritable. Such changes can be effectively random. Usually they will be deleterious. But not quite always: there is the possibility that an arbitrary design modification may improve the adaptation of an organism to its environment. Such improvements will tend to catch on because individuals incorporating them are more likely on average to survive and have offspring. Eventually, and more easily in sexually reproducing populations, the improved modification is likely to become universal. Mutations should not be too common or their destructive effects will outweigh their advantages: evolution over the long term depends on the replication of genetic information being usually very accurate.

If organisms can be said to be prerequisites for evolution, **life** can be taken as the rather intangible *product* of that process. Evolved organisms of the sort that we are familiar with conform to but are not wholly explained by the laws of physics and chemistry. In this they are like man-made machines. Indeed it is a useful informal definition of life that it is naturally occurring machinery.

Certainly if you look closely enough at a supposedly humble bacterium like *Escherichia coli*, you will find it as packed as an aeroengine with systems and subsystems that co-operate – that *must* co-operate for the bacterium to survive and be able to pass on its complicated message of how to survive. Surely *E. coli* is 'life' by our informal definition – and presumably the cleverness of its design is a product of evolution.

*The origin of life* Because 'life' is a rather vague idea it would not be possible to locate the origin of life precisely in time. Life would have emerged gradually during the evolution of organisms that would not actually have been alive to begin with. One might draw an analogy between the term 'life' and the term 'high ground'. If you go for a walk in the hills you will reach high ground sooner or later. You might give various reasons for saying that you had reached high ground at some stage – because of a change in the vegetation, perhaps, or an improvement in the view, or the need for another pullover. But although your companions might agree that a day's walk had reached high ground there would be much less agreement



as to exactly when. There would be arguments as to which were the definitive symptoms. In any case there would be no sharp line to be drawn.

Similarly with 'life' there are symptomatic characteristics that you can list if you like (metabolism, irritability, homeostasis, and such things), but the list, however erudite, will never amount to an unambiguous definition. 'Life', like 'high ground', is not a sufficiently sharp idea to be sharply defined. If there is no difficulty in practice in distinguishing now the living from the non-living, that is because evolution has by now climbed far beyond the regions of ambiguity. It is only because of this accident of viewpoint that life may seem to us to be a sharp idea.

'The origin of life', then, should be seen as describing a field of study rather than an event. Even the more particular question of the origin of life on the Earth has several parts to it. For example:

1 *Prevital conditions.* What was the Earth like when the first organisms arose?
2 *The origin of organisms.* What were those first systems like that were able to evolve but had not yet evolved? What were they made of? How did they arise?
3 *The emergence of life.* By what stage had early evolution given rise to seemingly designed, co-operative systems? What were organisms made of by then?
4 *The origin of our biochemical system.* How and when did evolution come to depend, as it now does, on nucleic acids and proteins?

Behind such questions there is the assumption that life arose spontaneously on the Earth – it was not brought here by spacemen or specially contrived by God. Here I am accepting a common view among scientists of the late twentieth century.

Another common view is that before there was life on Earth there had to be a build-up of the 'right' component molecules – amino acids, sugars, and such things. This seems to me to be a quite mistaken idea. The set of molecules that is now fixed at the centre of our biochemistry is bound in with a whole approach to molecular control that is highly sophisticated. As we shall see, it is doubtful whether our present 'biochemicals' would have accumulated in primordial waters; but even if they had, the first evolving systems could not in any sense have foreseen which of the molecules in their surroundings would have been particularly appropriate for the kind of machinery that was to emerge in the distant future. Natural selection is indeed an amazing engineer, but it works strictly without a

drawing board. It does not think ahead. It only knows a good thing when it sees it in operation. It is not at all clear, for example, that amino acids would have been particularly significant to begin with, long before there were accurately made proteins. It is not clear indeed that *any* of the now universal 'biochemicals' would have been appropriate at the very start of evolution, even if they had been there on the Earth. For evolvable systems it is not amino acids, sugars, or lipids that are critical, but somewhat higher order structures such as catalysts and membranes – and above all hereditary control machinery. I will try to show that there are easier ways of making such things than with the ultimate units on which today's biochemical high technology is based. My specific suggestion will be that crystalline inorganic minerals of colloidal dimensions – clays – would have provided the main materials out of which organisms of a first kind were made, and that organic molecules had little if any part to play right at the beginning. Such minerals would have been forming on the primitive Earth as they do on the Earth now. And the units from which they are made are simple, symmetrical inorganic species such as silicic acid and hydrated cations. These are effectively indestructible and their endless supply would be maintained through rock weathering processes. Often the clay minerals that are produced from weathering solutions seem to organise themselves fortuitously, in a rough and ready way, into the kinds of things that might be needed in a primitive organism.

For membranes, for example, there is no need to presuppose processes that might make amphiphobic organic molecules such as lipids: clay membranes are much more easily made – they are forming in the environment in vast quantities all the time. Again, inorganic materials are the readiest form of control devices such as specific adsorbants or catalysts. Even primitive genes can most easily be envisaged as arising through the continuous crystallisation of colloidal inorganic materials. Indeed the idea of a wholly inorganic (non-carbon) genetic material is at the heart and at the start of my thesis.

To avoid confusion I must stress the distinction between what is the generally held view of the role of minerals in the origin of life and the view being proposed here. It is usual to follow Bernal (1951) in seeing minerals as assisting in that putative prevital build-up of organic molecules suitable for life. Clays, for example, are seen as having concentrated organic molecules and catalysed reactions between them until these organic molecules formed themselves into systems able to reproduce and evolve under natural selection. I would agree that clays were important for the origin of life; but not in the way that Bernal said. On the view being proposed here clays were the materials, perhaps the sole materials, out of which the earliest organisms were *made*.

It will be a general theme, when thinking about primitive evolving systems, that what can start most easily is unlikely to be at all similar to what will be selected in the long run. Consider an analogy with human technological development. Primitive ways of doing things – with a clay tablet or a stone axe – depend on very little fabrication. These devices are made from materials to hand that just happen to have (more or less) appropriate properties. (The materials are literally rough and ready.) Later, as techniques of fabrication improve, much more sophisticated means to similar ends become possible. Often these are not simply elaborations of the original means. Perhaps the axe head became a spear, and the spear a bow and arrow, but the intercontinental missile has a quite different history, one that started more likely in the jewellery trade with the discovery of gold and metal working.

That sort of deviousness is quite like evolution too – at least it is like later episodes of evolution that are still visible to us. Our lungs are not improved gills, nor is our way of walking related to the locomotion of an amoeba. During evolution quite new ways of doing things become possible from time to time and these may displace older ways. Only in the short term can a succession of organisms be seen as a line: over the longer term it should be seen rather as a rope that is made up of fibres constituting the various subsystems in the organisms. The rope may be continuous that joins us to our ultimate ancestors, those makeshift systems that could first evolve under natural selection: but that is not to say that each or any fibre in that rope is continuous from the beginning to the end. Subsystems may come and go while an overlapping continuity is maintained over the long term. It is in this kind of way, as we shall discuss in Chapter 3, that evolution has been able to escape from original design approaches – through takeovers: first ways may not only be transformed, they can be *replaced* by later ways that are based on unrelated structures.

Our central biochemistry, so fixed now, was evolving at one time. I will suggest that during that evolution its subsystems were updated, perhaps some of them several times; that protein, in particular, is a comparatively new idea replacing earlier, more cumbersome approaches to biomolecular control – and that nucleic acid is only a little older. I will try to indicate that switches of function, so common in that part of evolution still visible to us, would have been expected too in that now invisible early biochemical evolution and would have made possible the takeover of the most central controller of all – the genetic material. From the considerations touched

on above – that it is unlikely that what is possible to begin with will resemble what is optimal in the end – there would be a tendency for radical change in the genetic material (as in other things) during very early biochemical evolution. Given also a mechanism – genetic takeover – such a change would seem almost inevitable.

Two of the main preoccupations in this book, then, will be with the questions: 'is genetic takeover possible?' and 'can we imagine genetic materials that could more easily have been generated on the primitive Earth than nucleic acid?' Genetic takeover is the key that allows us to contemplate altogether different kinds of materials for the most primitive genes.

Part I of this book will lead towards the idea of genetic takeover. Part II will use this key to allow us to shift our view of the material nature of the very first organisms. Part III is then a new story of how life originated on the Earth.

Here are the kinds of answers that will emerge to those four sets of questions put earlier.

1 The most favourable prevital conditions would have been on an Earth that had land and sea and weathering cycles and an atmosphere dominated by nitrogen and carbon dioxide.
2 The first organisms were a subclass of colloidal mineral crystallites forming continuously in open systems.
3 These mineral organisms evolved modes of survival and propagation that would have seemed highly engineered or contrived. That is to say they became a form of life.
4 Some evolved primary organisms started to make organic molecules through photosynthesis. This led to organisms that had both inorganic and organic genes. Eventually the control of their own synthesis passed entirely to the organic genes (nucleic acid) which by now operated through the synthesis of protein.

# PART I

# The problem

# 1

# Current doctrine

*A question of definition* The established view about how life first arose is embodied in the doctrine of chemical evolution. This is not easy to define as it is not a single idea, but it can be understood from accounts such as those by Bernal (1967), Calvin (1969), Kenyon & Steinman (1969), Lemmon (1970), Miller & Orgel (1974), Dickerson (1978) and others.

At a very general level the doctrine of chemical evolution is simply:

1 that life arose from systems that obeyed the normal laws of physics and chemistry (Calvin, 1969). I would not cavil with that; but there is a more restricted meaning for the term 'chemical evolution' that is more or less explicit in most contemporary accounts. Almost invariably there are two other notions present which seem to me to be much more dubious, namely:
2 that there was a prevital progression, a natural long-term trend analogous in a limited way to biological evolution, that proceeded from atoms to small molecules to larger molecules – and finally to systems able to reproduce and evolve under natural selection; and also:
3 that the relevant molecules in prevital processes were, broadly speaking, the kinds of molecules relevant to life now (see figure 1.1).

In this chapter we will be concerned with some of the main successes of the doctrine of chemical evolution and also with some of its more immediate difficulties. These will be mainly difficulties of a practical sort – although some are very serious. I will leave until Chapter 2 my more fundamental scepticism.

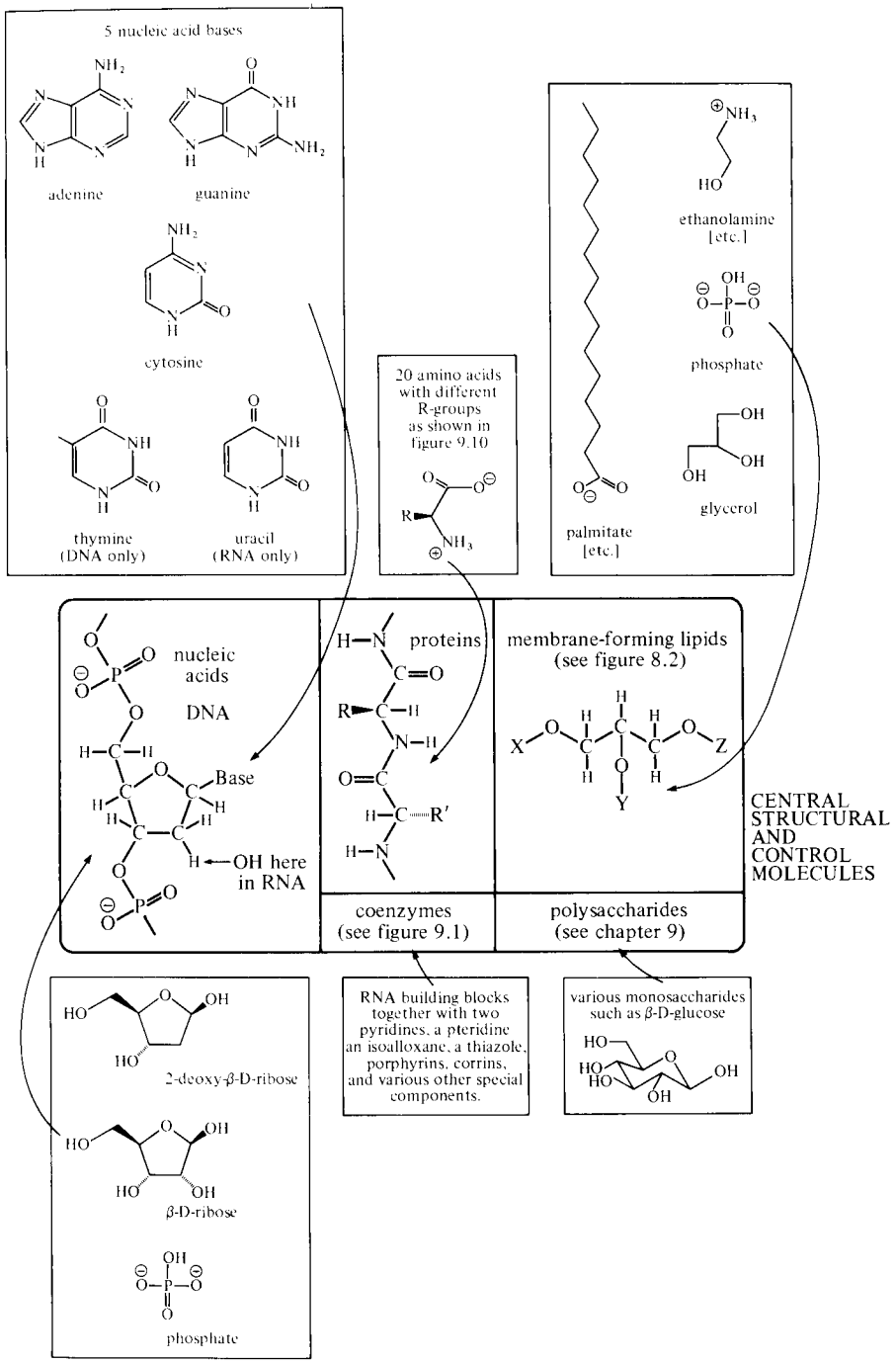



Figure 1.1.



### The Oparin–Haldane hypothesis

A. I. Oparin in 1924, and J. B. S. Haldane in 1929, introduced the idea of a 'hot dilute soup' (in Haldane's words) from which life emerged and on which early life fed. Haldane saw carbon dioxide ($CO_2$) as the atmospheric carbon source and ultraviolet light as the means of generating reduced organic species from this source. In this he followed Allen (1899), Moore (1913), Moore & Webster (1913) and Becquerel (1924), and he was influenced by the studies of Baly (e.g. Baly, Heilbron & Barker, 1921; Baly, Heilbron & Hudson, 1922) on *in vitro* photosynthesis. Haldane took a genetic view of first life: 'The first living or half-living things were probably large molecules synthesised under the influence of the sun's radiation, and only capable of reproduction in the particularly favourable medium in which they originated.' This idea of some kind of genetic molecule at the origin of life has been a recurring one. Troland (1914) and C. B. Lipman (1924) suggested something of the sort as did Morgan (1926), Muller (1929), Dauvillier (1938), Beadle (1949), and Blum (1951). In more recent formulations of this idea life is generally seen to have started from molecules of nucleic acid (Horowitz, 1959; Maynard Smith, 1975).

Oparin (1938, 1957) was later to discuss these and numerous other ideas. For Oparin the main atmospheric carbon source was not $CO_2$ but methane ($CH_4$). According to Oparin the early atmosphere was somewhat like that of Jupiter, composed mainly of non-metal hydrides – all of which are volatile. This vision of a highly reduced early atmosphere was in line with Urey's ideas as to how the planets formed (Urey, 1952). The vast excess of hydrogen in the cloud of dust and gas from which the solar system formed would only gradually have dispersed, he said, leaving a reservoir of reduced materials that would dominate the atmosphere of the Earth for a considerable time.

Figure 1.1. All life now on Earth depends on a central set of large and rather complex molecules that can be made by joining together 100 or more 'building blocks' (shown in the outer boxes). **Nucleic acids** provide the genetic material for all present-day organisms. They also provide much of the machinery for the synthesis of proteins (see figures 9.6–9.9). **Proteins**, together with **coenzymes** provide catalysts and reagents for chemical reactions; while, with **lipids** and **polysaccharides**, proteins also provide the 'glassware' and other apparatus needed to perform organic chemical operations. Between them the central molecules create the conditions for their own resynthesis by controlling the synthesis (or acquisition) of their building blocks and the assembly of these components in an appropriate way.

Oparin also strongly rejected the idea of life originating from a fortuitously forming genetic molecule. He regarded such an idea as unscientific on the grounds that it would be too improbable. 'How can one study a phenomenon which, at best, can only have occurred once in the whole lifetime of the Earth?' he asked. Oparin took a more 'metabolic' view of the nature and origin of life. Although reproduction was an essential prerequisite for Darwinian evolution, for the appearance of life in the full sense, this property of reproduction was only gradually arrived at through systems that had a kind of metabolism and which could even evolve to some extent through a kind of selection. He saw these systems as coacervate droplets – that is, aggregations of hydrophilic polymers such as those that had been studied in the laboratory by Bungenberg de Jong (1932). The small molecules made originally in the atmosphere and washed into the seas would have polymerised, he said, and formed discrete coacervate droplets which would have served to concentrate organic molecules generally. Already these might have some of the characteristics of life: these systems are individuals, with reactions presumably taking place inside them. In some droplets the reactions cause more droplet material to be formed. Such droplets will grow at the expense of others. Even reproduction of a sort is not very hard to imagine in a real ocean with breaking waves: big droplets will simply get broken into pieces which can then continue to grow. Thus slowly over a long period these coacervates become more life-like. Eventually they come to be able consistently to pass on characteristics to offspring; they can reproduce, that is, in the full biological sense, and evolution through natural selection is under way.

### *The primitive atmosphere*

On the two main points on which Oparin and Haldane differed – on primitive gases and primitive genes – it was Oparin's views that were to predominate. Not that the idea of a $CO_2$ carbon source was to disappear: that line continued particularly among those who saw sugars at the beginning of our biochemistry, for example Dauvillier (1938, 1965) and Lipmann (1965). And one of the earliest experiments simulating conditions on the primitive Earth was carried out by Calvin and his associates (Garrison *et al.*, 1951) – on the reduction of $CO_2$ by ionising radiation in the presence of hydrogen ($H_2$) and solutions containing iron (as $Fe^{2+}$). Subsequently, Getoff (1962, 1963a, b) was to show that ultraviolet radiation was effective too in making reduced species from $CO_2$ in water – in making, for example, carbon monoxide (CO), aldehydes and oxalic acid.



More recently, Hubbard, Hardy & Horowitz (1971) and Hubbard, Hardy, Voecks & Golub (1973), in studying simulated Martian conditions, have found that reduced organic species are formed when silicates are exposed to ultraviolet radiation in the presence of water and CO. (It is reasonable to suppose that there might have been some CO in a $CO_2$-containing atmosphere – through ultraviolet photolysis of $CO_2$ and from volcanoes: on the other hand, there is little CO in the Martian atmosphere.) More recently still, Bar-Nun & Hartman (1978) have found an efficient synthesis of alcohols, aldehydes and acids from ultraviolet photolysis of CO and $H_2O$.

So it is certainly plausible to suggest that reduced non-nitrogenous organic species could have been formed on the primitive Earth from $CO_2$ and sunlight, as both Allen and Moore had suggested so long ago. There had been doubts about this idea due to difficulties in repeating some of Baly's work (e.g. Mackinney, 1932), and this may have helped to suppress interest in the primordial fixation of $CO_2$.

Another consideration is that much less energy is needed to make organic molecules in a strongly reducing atmosphere (figure 1.2). And by the early nineteen fifties non-nitrogenous molecules, such as sugars, were no longer being seen at the origin of our biochemistry. Attention had shifted from metabolism itself to something still more central – to the control of metabolism, to proteins. Amino acids seemed more important than simple sugars. In 1953 Watson & Crick were to re-emphasise that life, or at least life now, is based on nitrogen-containing compounds. It became possible to imagine, perhaps, a minimal organism of nucleic acid or of nucleic acid and protein.

But it was Miller's experiment, also from 1953, that was to have the strongest effect in establishing that vision of Oparin's and Urey's of a strongly reducing early atmosphere for the Earth. Miller passed sparks through a mixture of $CH_4$, $NH_3$ (ammonia), $H_2O$ and $H_2$ for a week at somewhat under 100 °C in an apparatus that allowed water-soluble products to be removed. The gas was the atmosphere, the water trap the oceans, and the sparking was a thunderstorm on the primitive Earth. Among the products identified were glycine and alanine. Later (Miller, 1955; Miller & Urey, 1959) it was found that 15 % of the original carbon added as $CH_4$ appeared in a fairly limited range of identifiable organic molecules. These are shown in figure 1.3. The rest of the carbon was mainly in the form of an unanalysed, probably largely polymeric, tar. Among the identified species there were four of the protein amino acids.

This was a sensational result: you might say a textbook example of

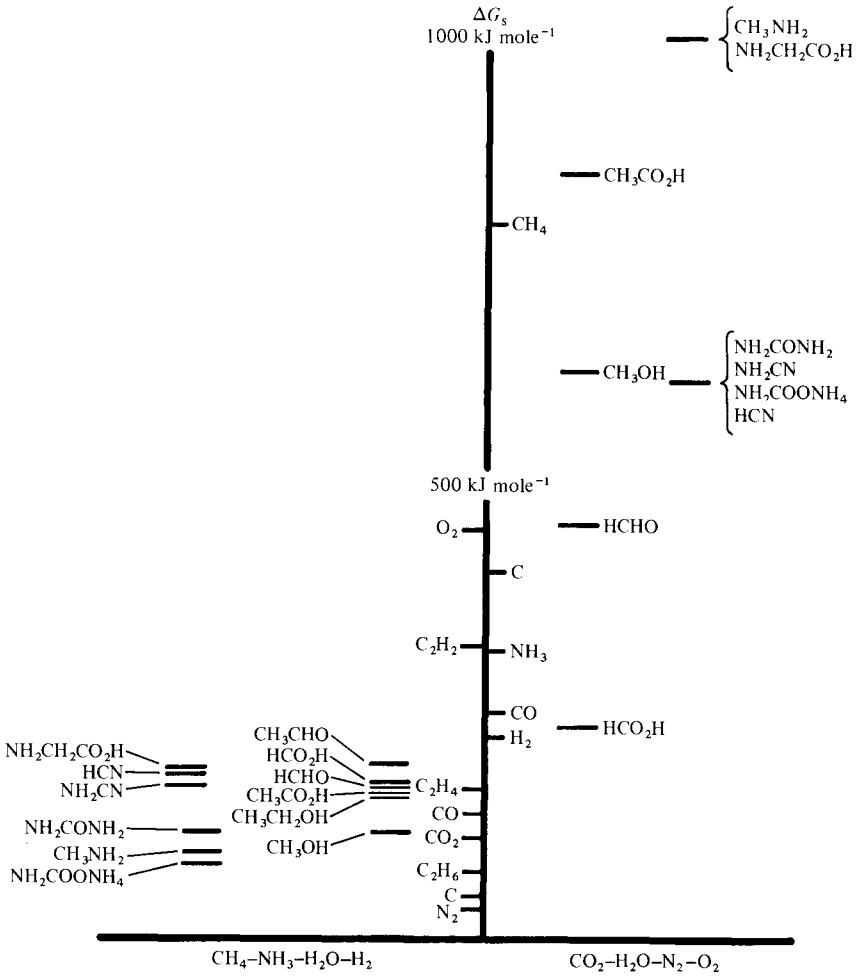



Figure 1.2. Standard free energies of synthesis of some molecules from a strongly reducing gaseous mixture (left) and from a strongly oxidising one (right). (After Toupance, Raulin & Buvet, 1971.)

scientific method at work. An hypothesis – in this case an hypothesis about important and remote events – had been tested by experiment, and not found wanting. Not only could organic molecules have been formed under a reducing primitive atmosphere, but there seemed to be a prejudice in favour of molecules at the centre of our biochemistry.

So the strongly reducing primitive atmosphere entered the general scientific imagination. Oparin's and Urey's vision was now generally



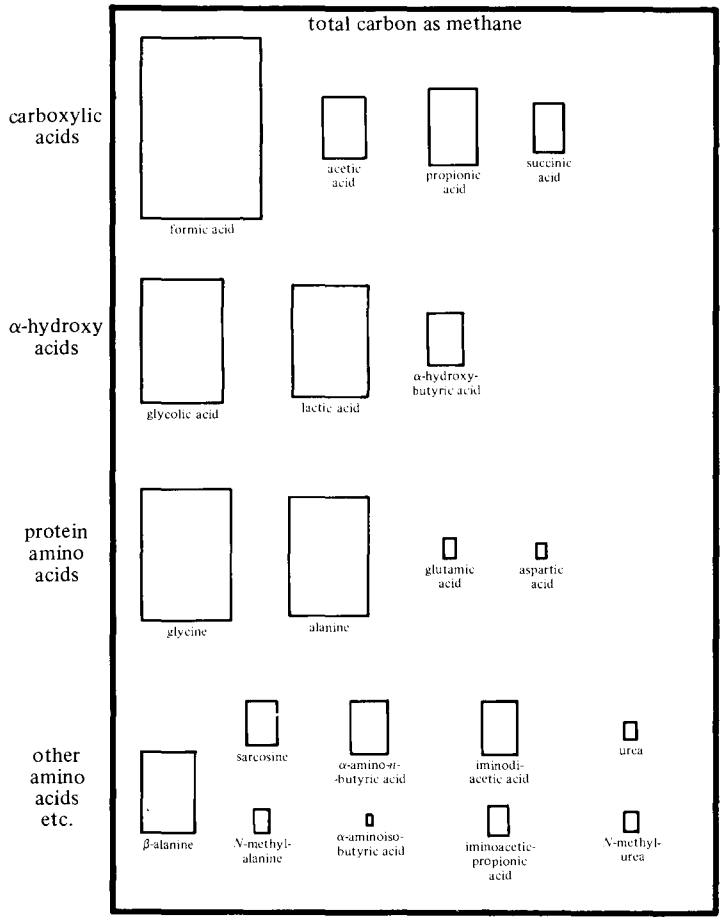



Figure 1.3. Yields from sparking a mixture of $CH_4$, $NH_3$, $H_2O$ and $H_2$. The large outer box represents the total carbon available initially as methane. The other boxes represent the proportions of that carbon that became fixed in identified compounds. (Calculated from data in Miller & Orgel, 1974.)

believed to be true: even perhaps to be necessary for a rational explanation of the origin of life. At any rate the idea became almost obligatory for first pages of elementary textbooks on biology and geology.

Yet opinions based on the latest geological and astronomical evidence are increasingly against an early strongly reducing atmosphere for the Earth (e.g. Kerr, 1980). The trend in recent years has been towards the view that by the time the Earth had settled down sufficiently from its birth

traumas to have sustained life, or to have started to accumulate a primordial soup, the general conditions would soon have been rather as now – but without the effects of life. The atmosphere would have been nitrogen ($N_2$) and $CO_2$, mainly, plus a small ration of noble gases with possibly some $H_2$ and CO as well.

Geologists have long favoured the idea that volcanoes have been the source of our atmosphere and hydrosphere (Rubey, 1951; Holland, 1962; Abelson, 1966). Even the earliest atmosphere, it would seem, must have been produced this way – it could not simply have been the residuum from the hydrogen-laden clouds from which the solar system formed. Had the first atmosphere been picked up in that way then the noble gases would have been caught at the same time – and the noble gases are not abundant enough on our Earth for that.

The question is: what was coming out of the volcanoes by the time the Earth had settled down from its birth traumas? The answer to that question depends on the oxidation state of the rocks in the crust and upper mantle of the early Earth. Holland (1962) had suggested that there might have been an early period when volcanic gases were quite strongly reducing – with plenty of hydrogen and carbon monoxide although not very much methane. This would have been possible if metallic iron had stayed near the surface for a time, only later sinking to the core.

Walker (1976a, b) has pointed to difficulties here. The $Fe^{3+}:Fe^{2+}$ ratio in present-day mantle-derived rocks is too high to suggest that these rocks were ever in equilibrium with metallic iron. Also there is too much nickel in surface rocks – if metallic iron had been present near the surface at one time, and then moved away, it should have taken the nickel with it (Clark, Turekian & Grossman, 1972). In any case recent ideas on planet formation suggest that the later accreting material, that would have formed the crust and upper mantle, would have been quite highly oxidised – anyway not metallic iron. Finally there is a fundamental problem in that any major change in the oxidation state of the mantle would have had cataclysmic effects at the surface – enormous amounts of gravitational energy would have been released as iron sank to the core. As Walker puts it: 'a highly reducing atmosphere, if it ever existed, would have had a short life and a violent end'. On the other hand, even present volcanoes produce a little hydrogen so that in Walker's view the primitive atmosphere might have contained a few per cent of hydrogen.

Admittedly such evidence is indirect. There is, however, direct evidence against there having been a strongly reducing atmosphere as long ago as 3.8 billion years. In Greenland there are metamorphosed sedimentary rocks



of that age (Moorbath, O'Nions & Pankhurst, 1973). These contain abundant carbonate (Allaart, 1976) showing that at least by that time carbon dioxide was in the atmosphere (Schidlowski, 1978). Indeed Moorbath (1977) suggests that since the Greenland rocks are not so very abnormal, conditions when they were laid down could not have been so very different from now.

As I said, there is only the indirect evidence against some still earlier strongly reducing phase. But the time available for that hypothetical phase is getting rather tight. There is evidence, from moon rocks and elsewhere, for heavy impact cratering throughout the inner solar system that finished at about 4.0 billion years ago (Goodwin, 1976; Hartmann, 1975; Kirsten, 1978). That, seemingly, was the final stage of the accretion of the inner planets, and it gives us an earliest likely starting date for the origin of life (or for the consistent build-up of a primordial soup if there ever was such a thing). If a strongly reducing atmosphere was needed for the origin of life there was only about 0.2 billion years when it might have been there.

In addition to these considerations there are arguments against very much $NH_3$ having been present in the primitive atmosphere – even in a strongly reducing one – it is too soluble in water and too quickly photolysed (Bada & Miller, 1968; Ferris & Nicodem, 1974; Schwartz, 1981).

As doubts arose about the constitution of the primitive atmosphere, experiments were performed under a wider range of conditions, with different mixtures of gases and with different energy sources. The overall result as reviewed by Kenyon & Steinman (1969) was remarkable: neither the composition of the starting gas mixture (provided it is free of oxygen) nor the energy source is particularly critical. Electrical discharges could cause amino acids to be formed in any of the following mixtures for example: $CH_4+NH_3+H_2O$; $CO+N_2+H_2O$; $CO+NH_3+H_2$; $CO+N_2+H_2$ (Abelson, 1956, 1957); or $CH_4+N_2+H_2O$+traces of $NH_3$ (Ring, Wolman, Friedman & Miller, 1972; Wolman, Haverland & Miller, 1972). Also, as was pointed out by Miller (1955), it had been known for a long time that glycine can be synthesised by the action of a silent discharge on $CO+NH_3+H_2O$ (Loeb, 1913).

Other energy sources that can generate amino acids from strongly reduced mixtures are ultraviolet radiation when $H_2S$ is present as sensitiser (Sagan & Khare, 1971); ultraviolet radiation or sunlight in the presence of platinised $TiO_2$ (Reiche & Bard, 1979); ionising radiation (Palm & Calvin, 1962); heating at 1000 °C with silica catalysts (Harada & Fox, 1964, 1965); heating $CO+H_2+NH_3$ with Ni/Al or clay catalysts (Hayatsu, Studier & Anders, 1971); by means of shock waves (Bar-Nun, Bar-Nun, Bauer &

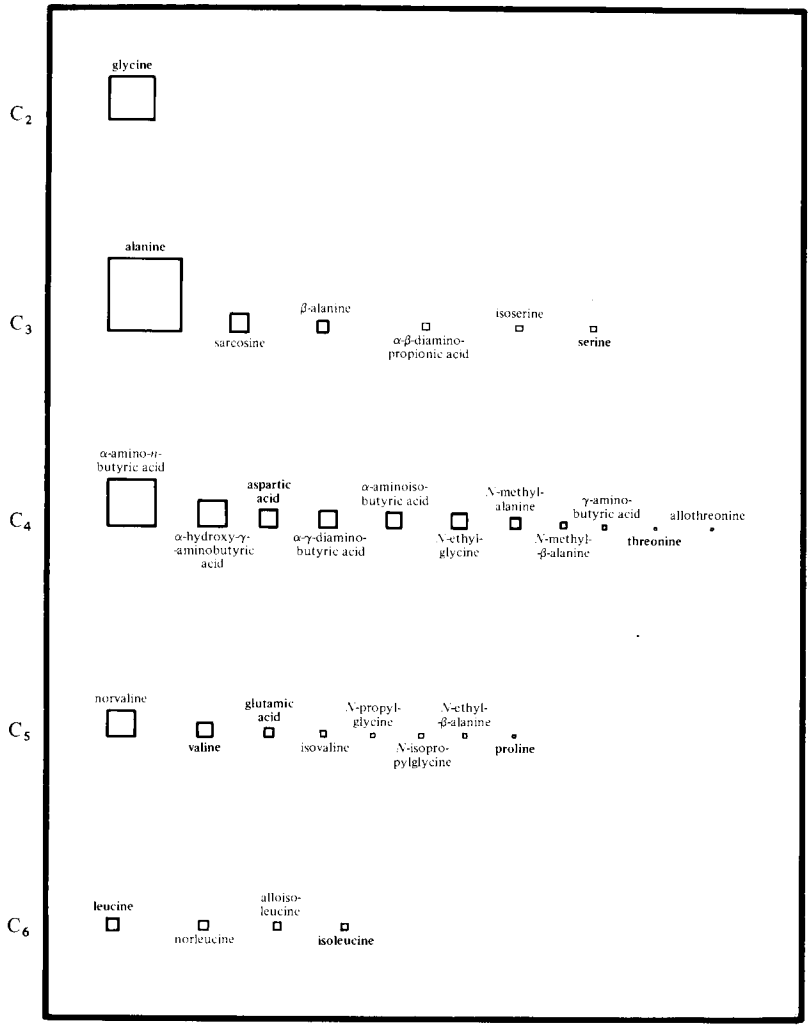



Figure 1.4. Yields of amino acids from sparking a mixture of $CH_4$, $N_2$, $H_2O$ and a trace of $NH_3$. The large outer box represents the total carbon available initially as methane. The other boxes represent proportions of that carbon that became fixed in identified amino acids. (Calculated from data in Miller & Orgel, 1974, from Ring, Wolman, Friedman & Miller, 1972, and Wolman, Haverland & Miller, 1972.)



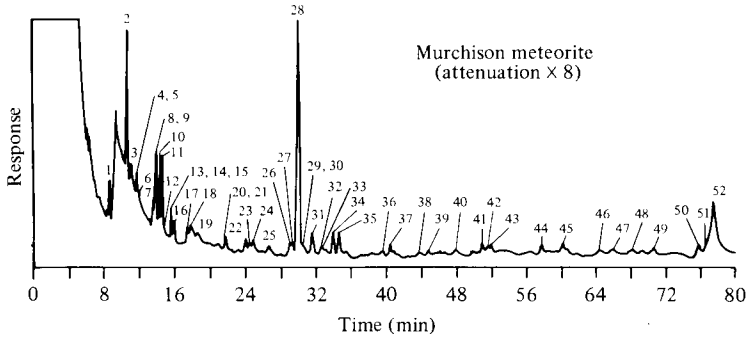



Figure 1.5. Gas chromatogram of the *N*-trifluoroacetyl-D-2-butyl esters of amino acids in an acid-hydrolysed extract of the Murchison meteorite. The numbers refer to Table 1.1. (From Lawless & Peterson, 1975.)

Sagan, 1970, 1971); by heating aqueous $NH_3+CH_2O$ (Fox & Windsor, 1970) or aqueous hexamethylene tetramine with hydrochloric acid (Wolman, Miller, Ibanez & Oró, 1971).

Even the kinds and relative amounts of amino acids and other products are not so strikingly affected by these widely different conditions: glycine and alanine are generally still at the head of the list of the protein amino acids formed (although alanine sometimes wins); there may be small amounts of other small hydrophobic amino acids such as valine; very often aspartic and glutamic acids are there with serine not uncommon.

Unfortunately absolute yields are not quoted in many cases and quite often it seems that only protein amino acids were looked for. (There is also the possibility that if protein amino acids were not found then work was not published.) These objections do not apply to the results summarised in figure 1.4 which is probably a fair example of the kinds of yields and product distributions obtained from more successful syntheses. This set of amino acids was formed in a sparking experiment similar to Miller's original one. The main differences were in the absence of $H_2$, the presence of $N_2$ and the near absence of $NH_3$. These changes did not greatly affect the outcome although the yields were smaller (cf. figure 1.3). This set of amino acid products is strongly similar also to the products of another presumably abiotic synthesis: the amino acids in the Murchison meteorite. (See figure 1.5 and Table 1.1.) The comparison with the amino acid distribution in *E. coli* is much less striking (Lawless & Peterson, 1975).

I will leave until Chapter 2 the question of why any modern biochemicals should turn up in meteorites, etc. In the meantime we may ask why it should

Table 1.1. *Amino acids in the Murchison meteorite (from Lawless & Peterson, 1975)*

| | |
|---|---|
| 1 Isovaline | 27 L-Pipecolic acid |
| 2 α-Aminoisobutyric acid | 28 Glycine |
| 3 D-Valine | 29 Neutral cyclic |
| 4 Linear neutral ($C_6$) | 30 Neutral cyclic |
| 5 L-Valine | 31 β-Alanine |
| 6 *N*-Methylalanine | 32 Neutral cyclic |
| 7 D-α-Amino-*n*-butyric acid | 33 Polyfunctional linear aliphatic |
| 8 D-Alanine | 34 D-Proline |
| 9 Linear neutral ($C_5$) | 35 L-Proline |
| 10 L-α-Amino-*n*-butyric acid | 36 Linear neutral ($C_5$) |
| 11 L-Alanine | 37 Unknown[b] |
| 12 Linear neutral ($C_5$) | 38 Unknown[b] |
| 13 $NH_3$[a] | 39 Linear neutral ($C_5$) |
| 14 Linear neutral ($C_5$) | 40 Unknown[b] |
| 15 Linear neutral ($C_6$) | 41 γ-Aminobutyric acid |
| 16 *N*-Methylglycine | 42 D-Aspartic acid |
| 17 *N*-Ethylglycine | 43 L-Aspartic acid |
| 18 D-Norvaline | 44 Polyfunctional linear aliphatic |
| 19 L-Norvaline | 45 Polyfunctional linear aliphatic |
| 20 Linear neutral ($C_5$) | 46 Polyfunctional linear aliphatic |
| 21 Linear neutral ($C_6$) | 47 Polyfunctional linear aliphatic |
| 22 D-β-Aminoisobutyric acid | 48 Polyfunctional linear aliphatic |
| 23 L-β-Aminoisobutyric acid | 49 Unknown[b] |
| 24 β-Amino-*n*-butyric acid | 50 D-Glutamic acid |
| 25 Unknown[b] | 51 L-Glutamic acid |
| 26 D-Pipecolic acid | 52 Unknown[b] |

[a] Present in blank.
[b] Peaks labelled 'unknown' do not appear to be amino acids. Their identification awaits the use of high-resolution mass spectrometry.

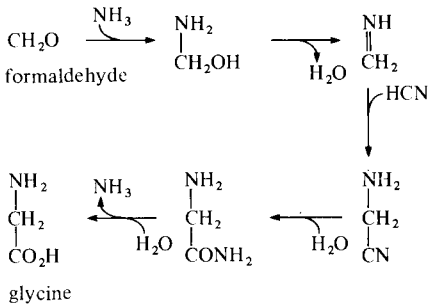
be that a fairly limited set of small molecules should keep on turning up under rather different conditions of synthesis. Alanine and glycine, for example, are particularly ubiquitous, with amino acids generally among favoured types. The answer is probably rather complicated. First we must not exaggerate: there are differences in the mixtures and most of the material is usually unanalysed polymeric tar. Second, the experiments are usually contrived so as to favour water-soluble species – if they can get into the water they escape the energy source and avoid being further transformed. And of course stable molecules will tend to be favoured in any case. But at least a large part of the explanation is probably kinetic:



because although the energy sources may be different these often generate the same small, high-energy, intermediate species. It is perhaps the limited number of these that leads to the relatively limited spread of products (Oró, 1965; Buvet & Le Port, 1973).

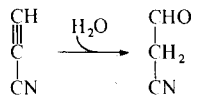
*Abiotic routes to 'biochemicals'*

*Cyanide* Cyanide (HCN) is one of those small, easily made high-energy species. Miller (1957) showed that in the early stages of his sparking experiments HCN was generated, along with aldehydes. He suggested that the mechanism for the formation of amino acids was a Strecker synthesis: for example glycine might be formed as follows:

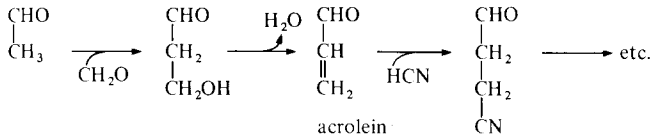



Other aldehydes should produce the corresponding amino acids – for example acetaldehyde would give alanine through similar reactions. The corresponding hydroxy acids (see figure 1.3) could be produced through the addition of HCN to aldehydes rather than, as in the above scheme, to imines. The relative amounts of amino and hydroxy acids that are produced depend on the concentration of free ammonia – the amounts are about equal when the concentration of free ammonia is about $10^{-2}$ M at 25 °C (Miller & Orgel, 1974).

Cyanoacetylene is another of the high-energy intermediates found in the sparking experiments, particularly with $CH_4 + N_2$ (Sanchez, Ferris & Orgel, 1966b). This hydrates to cyanoacetaldehyde:



which might be expected to yield asparagine and aspartic acid through a Strecker synthesis (the cyanide group in the initial aldehyde also hydrolysing, via primary amide, to acid). Glutamic acid, it has been suggested, could arise from acrolein obtained from the reaction of formaldehyde with acetaldehyde:



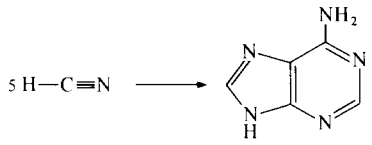

Serine might have arisen from glycolaldehyde – itself formed by condensing two formaldehyde molecules (the first step in the formose reaction discussed on page 28). These and other mechanisms are discussed by Miller & Orgel (1974), who emphasise that in the simulation experiments the yields of non-protein amino acids may be comparable to those of the common amino acids. Also the yields are always low by normal synthetic standards, and sometimes very low indeed.

Miller & Orgel suggest some (rather long) synthetic routes to the aromatic amino acids phenylalanine, tyrosine and tryptophan, although it is doubtful whether more than traces of these amino acids have ever been formed from very simple starting conditions. The basic amino acids lysine, arginine and histidine are still more difficult and lack any plausible prevital synthetic pathway. But these are by no means fatal difficulties for the doctrine of chemical evolution: it would be naive to suppose that all the present amino acids were needed for the very first organisms. Some 'primordial' set might have been enough to begin with – say glycine, alanine, valine, leucine, aspartic acid, glutamic acid and serine – the others being added during early biochemical evolution.

An additional source of aldehydes might have been available through the action of ultraviolet light on water vapour in an atmosphere that contained $CH_4$ or CO (Ferris & Chen, 1975a). If enough $NH_3$ had been present, HCN too could have been generated photochemically, as it probably still is in the atmosphere of Jupiter (Ferris & Chen, 1975b).

Already in 1955 Miller had noted that 'polymers' of hydrogen cyanide were formed during his experiments on the synthesis of amino acids by electric discharges. Stimulated partly by this and by a report from Abelson that such polymers gave amino acids on acid hydrolysis, Oró in the early sixties started investigating cyanide as a precursor for amino acids (Oró & Kamat, 1961). Interestingly cyanide also turned out to be a precursor for adenine (Oró, 1960; Oro & Kimball, 1961), which can be regarded as a pentamer of HCN:





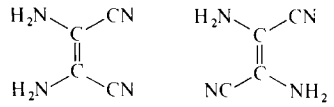

adenine

Ammonium cyanide was used in these early experiments with temperatures of 27–70 °C and concentrations in the range 1–15 M. Yields were of the order of half a per cent or so. More efficient syntheses of adenine, under admittedly very artificial conditions, were to follow: a 15–20 % yield from cyanide in liquid ammonia (Wakamatsu *et al.*, 1966), and a 40–50 % yield by heating formamide in a sealed vessel at 120 °C with $POCl_3$ (Morita, Ochiai & Marumoto, 1968). In this last case, although formamide can be regarded as the hydrate of HCN, HCN was *not* an intermediate in the reaction.

In Oró's experiments there was arguably too much cyanide to be realistic, but subsequent experiments under conditions nearer to those plausible for the primitive Earth have confirmed that amino acids, adenine and several other 'biochemicals' are formed from HCN (Lowe, Rees & Markham, 1963; Sanchez, Ferris & Orgel, 1967; Mathews & Moser, 1967; Ferris, Donner & Lobo, 1973; Ferris, Wos, Nooner & Oró, 1974; Ferris, Joshi, Edelson & Lawless, 1978; Ferris & Edelson, 1978).

A typical procedure is given by Ferris, Joshi, Edelson & Lawless (1978). A 0.1 M solution of HCN at pH 9.2 was left in a dark stoppered bottle at ordinary temperatures. After six months about half the cyanide had been converted to a complex mixture of HCN oligomers. The oligomeric material was then hydrolysed at 110 °C for 24 h in either 6 M HCl or at pH 8.5 to give a mix containing the 'biochemicals' shown in Table 1.2.

Of the oligomers of HCN the tetramers diaminomaleonitrile and diaminofumaronitrile:




are seen as important intermediates on the way to higher oligomers and to the amino acids, purines and pyrimidines, etc. shown in Table 1.2.

Table 1.2. *Biomolecules from hydrolysis of HCN oligomers (from Ferris, Joshi, Edelson & Lawless, 1978)*[a]

| | 6 N HCl | pH 8.5[b] |
|---|---|---|
| 4,5-Dihydroxypyrimidine | 0.7–0.9 % | – |
| Orotic acid | – | 0.009 % |
| 5-Hydroxyuracil | 0.003 % | nd[d] |
| Pyrimidine (structure unknown)[b] | – | + |
| Adenine | 0.03–0.04 % | + |
| 4-Aminoimidazole-5-carboxamide | + | +[e] |
| Glycine | 0.6 % | 0.1 % |
| Diaminosuccinic acid | 0.1 %[c] | + |
| Aspartic acid | + | + |
| Alanine | + | + |
| β-Alanine | + | + |
| α-Aminoisobutyric acid | + | + |
| Guanidinoacetic acid | 0.03 % | + |
| Guanidine | 0.2 % | 0.1 % |

[a] HCN oligomers were prepared and hydrolysed as described in Ferris *et al.* (1978). The yields shown are based on starting HCN. Conversions are approximately 2× as great since only about 50 % of the cyanide is consumed. It is assumed that 5 moles of HCN are required to form each mole of heterocyclic compound and 4 moles are required to form each mole of glycine. The identity of each compound was established by comparison of the mass spectrum of a volatile derivative with that of an authentic sample except for the following which were purified by chromatography and identified by their ultraviolet spectra and/or specific colour tests: adenine, 4-aminoimidazole-5-carboxamide, guanidinoacetic acid and guanidine. The presence of the compound is indicated by (+) and absence by (–). When quantitative analyses were not performed no analysis is indicated by (nd). The presence of guanine, xanthine and hypoxanthine is suggested by ultraviolet data but these were not reported because their identity has not been confirmed by mass spectral data.

[b] This compound gives a blue colour with diazotised sulphanilic acid and it is converted to 4,5-dihydroxypyrimidine by hydrolysis with 6 N HCl.

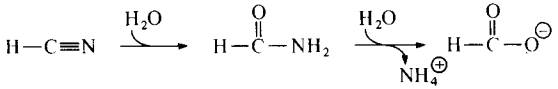
[c] Sum of the yields of *meso*- and racemic diaminosuccinic acid. They are formed in approximately a 2:1 ratio respectively.

[d] Control experiments demonstrated that this compound is destroyed by the hydrolysis conditions used in this study.

[e] 4-Aminoimidazole-5-carboxamide decomposes under acid hydrolysis conditions and when hydrolysed at pH 8.5 for 24 h. It can be detected in the hydrolysate of the HCN oligomers formed in the absence of oxygen when a 1.5 h hydrolysis time is used.



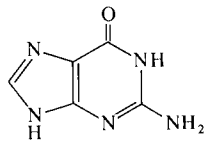
Whatever the mechanism, it is impressive that so many of the crucial types of nitrogen-containing molecules in our biochemistry can be made from a single very simple starting material. Yet there are difficulties. The yields are very small. And again there is a problem about initial concentrations: if the HCN is below 0.01 M there is no oligomerisation and the hydrolysis of cyanide to formate gets the upper hand:




The most efficient synthetic sources of cyanide would probably have been electric discharges or shock waves in a strongly reducing atmosphere. As we have seen, it is questionable whether there ever was a methane-laden atmosphere. Even then it is hard to imagine thunderstorms making enough cyanide. Even if the raindrops in the thunderstorm had been 0.1 M cyanide, by the time this had reached the oceans it would have been very dilute – the more so from the tendency of cyanide to hydrolyse to formic acid. Some efficient and rapid concentration mechanism would seem to be an essential part of this story.

Sanchez, Ferris & Orgel (1966a) suggested that freezing might have provided the necessary effect. On being cooled below 0 °C a dilute cyanide solution would become more concentrated as the ice separated from it. (Under ideal circumstances, a 75 % solution of cyanide could be created in this way – by cooling to −21 °C which is the eutectic temperature of $HCN/H_2O$ mixtures.) The oligomerisation reactions thus become possible (even if they would be rather slow) from initially very dilute solutions. At the same time the hydrolysis reaction would be strongly inhibited.

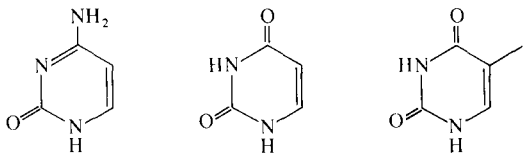
Guanine, xanthine and hypoxanthine were tentatively identified by Ferris *et al.* (1978). As the other purine of the nucleic acids, guanine is particularly interesting although it does not seem to be so easily come by as adenine.




guanine

Syntheses of guanine have been reported by Ponnamperuma (1965) on irradiation of HCN solutions, and from 4-aminoimidazole-5-carbonitrile by reaction with cyanate, urea or cyanogen (Miller & Orgel, 1974).

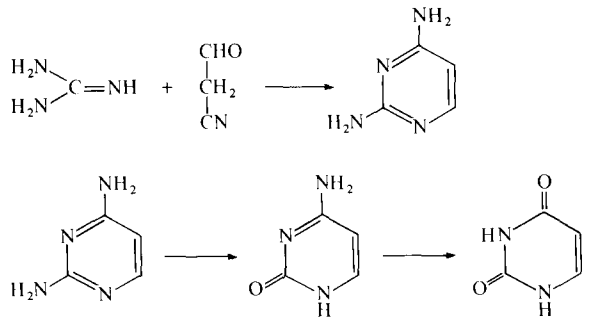
Of the three nucleic acid pyrimidines, cytosine, uracil and thymine:



cytosine uracil thymine

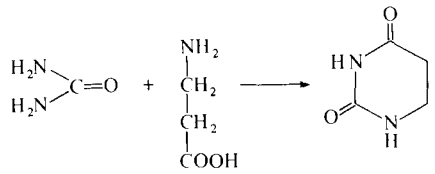
cytosine has been successfully made in 20 % yield from cyanate and cyanoacetylene (Sanchez *et al.* 1966b; Ferris, Sanchez & Orgel, 1968). Both cyanoacetylene and cyanate might be regarded as belonging to that class of molecules that are small, reactive and relatively easy to make from available energy sources: the first is a substantial product of sparking mixtures of methane and nitrogen, the second of ultraviolet irradiation of CO and $NH_3$ (Hubbard *et al.*, 1975). So in view of the high yield of the reaction between cyanoacetylene and cyanate it is perhaps legitimate to regard cytosine as a possible product of primitive Earth chemistry. But we should be cautious with reactions that presuppose starting materials that have to be themselves the products of other reactions. Could cyanoacetylene and cyanate have been made under the same general conditions, and if not, could they nevertheless have come together?

Given a synthesis for cytosine, then uracil looks easy since it forms in very good yield from cytosine by hydrolysis. Indeed the reaction is rather too easy: with a half life of 200 years at 30 °C there is the question of how enough of the cystosine would survive (Miller & Orgel, 1974). Guanidine (an HCN product; see Table 1.2) and cyanoacetaldehyde provide alternative possible routes to the synthesis of both of these pyrimidines (Ferris, Zamek, Altbuch & Freiman, 1974):





Yet another route to uracil that has been suggested starts from (the HCN products) urea and $\beta$-alanine giving, first, dihydrouracil:



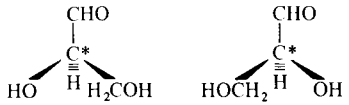
Such dihydropyrimidines can be dehydrogenated by irradiation with ultraviolet light in the presence of water vapour: this reaction is catalysed by clays – particularly montmorillonite (Chittenden & Schwartz, 1976).

The main interest in this last mechanism is perhaps that it can be modified to produce thymine – by adding acetate salts (Schwartz & Chittenden, 1977). Another synthesis of thymine – from uracil with formaldehyde and formic acid – has been reported recently (Choughuley, Subbaraman, Kazi & Chada, 1977).

We can see, then, that the entire set of nucleic acid bases *might* have been made on the primitive Earth. Routes from plausible intermediates can be formulated. If the 'wrong' pyrimidines predominate in the simpler kinds of synthesis this is not necessarily so important: as Ferris and his associates (1978) point out, the first kinds of nucleic acids might not have been quite the same as ours. These authors suggest that hydroxy and dihydroxy uracils might have been used and might have been more easily built into the necessary nucleosides. The main problems are in the yields: products are usually very minor constituents of tars.

*Formaldehyde* Cyanide and molecules related to it have so far dominated our discussions; but for the synthesis of sugars under loosely controlled conditions we must turn to formaldehyde. As long ago as 1861 Butlerov had discovered that formaldehyde generates a mixture of sugars in solutions of calcium hydroxide. Since then the formose reaction, as it has come to be called, has been demonstrated with a variety of catalysts, including common minerals, over a pH range from just acid to strongly alkaline (Reid & Orgel, 1967; Gabel & Ponnamperuma, 1967; Cairns-Smith, Ingram & Walker, 1972).

In principle the formose reaction is an oligomerisation of formaldehyde. You can regard glucose as $(CH_2O)_6$ rather as you can regard adenine as $(HCN)_5$. Like the oligomerisation of HCN the formose reaction is very

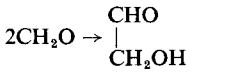



D-glyceraldehyde L-glyceraldehyde

Figure 1.6. Glyceraldehyde has one **asymmetric carbon atom** (starred), so called because this atom has four different things attached to it. As a result glyceraldehyde has two mirror-related forms designated D and L as shown. These are called **enantiomers**. (In general, objects that are not superposable on their mirror images are said to be **chiral**.) With most sugars there are two or more asymmetric carbon atoms so that, in addition to enantiomers, other kinds of **stereoisomers** become possible that are not mirror-related. (As a left-hand-plus-right-hand is not mirror-related to a left-hand-plus-left-hand.) Such isomers are called **diastereomers**. Usually if a molecule has $n$ asymmetric carbon atoms it will have $2^n$ stereoisomers consisting of $2^{n-1}$ pairs of enantiomers.

complex, and in practice it generates products other than sugars – particularly polyalcohols. It can be discussed in terms of seven main classes of reaction.

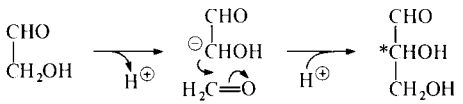
*1. Formaldehyde self-reaction:*




glycolaldehyde

This is a slow reaction the mechanism of which is uncertain.

*2. Aldol condensation of formaldehyde with a sugar:* for example the simplest sugar, glycolaldehyde, condenses to give glyceraldehyde:

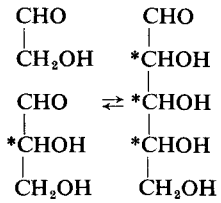



This reaction depends on the sugar having at least one 'active hydrogen', that is, at least one hydrogen atom attached to a carbon atom adjacent to a carbonyl group. (It is because there is no such active hydrogen in formaldehyde that its self-condensation is difficult.)

In making glyceraldehyde, as in the making of most amino acids from smaller molecules, an asymmetric carbon atom is generated so that there is the additional complication that two kinds of glyceraldehyde are produced (see figure 1.6 for a further explanation and definition of terms).

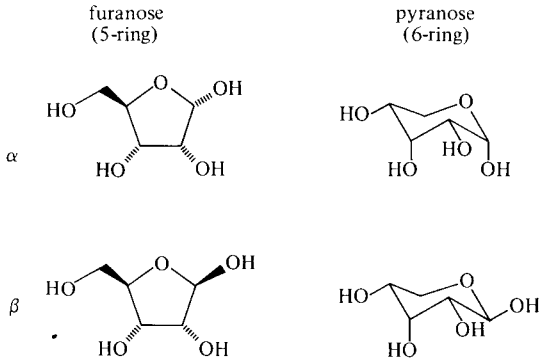


*3. Aldol and retro-aldol reactions of sugars:* for example between glycolaldehyde and glyceraldehyde:




There are three asymmetric centres now, in the aldopentose, so that really this corresponds to eight different structures: namely the 'right-handed' (D-) and 'left-handed' (L-) enantiomers of ribose, arabinose, xylose and lyxose. There is good reason to suppose that each of these sugars is formed to some extent and, in the case of right-handed and left-handed versions, in about equal amounts (Mizuno & Weiss, 1974). (A mixture containing an equal number of 'right-handed' and 'left-handed' enantiomers is described as a **racemic** mixture.)

*4. Ring formation:* ring structures become possible when there are more than three carbon atoms in a sugar. For D-ribose, for example, there are four possibilities:




As we will discuss later, this adds complications when it comes to using sugars as building blocks for larger molecules such as nucleotides or polysaccharides. Ring structures such as these are the predominent forms for most natural sugars, but the free sugars interconvert fairly easily via the open-chain structures – so we will keep to the simpler open-chain representations for the present.

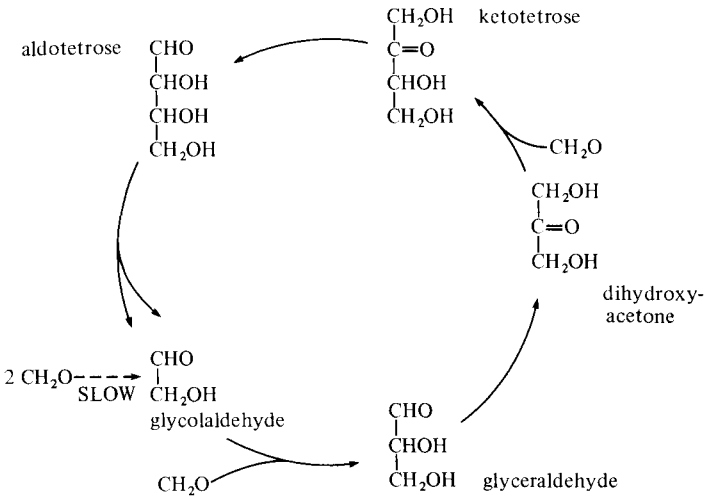
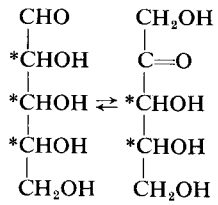
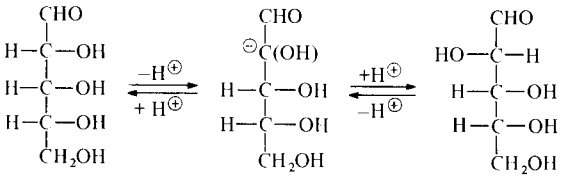



Figure 1.7. Breslow's autocatalytic cycle for the incorporation of formaldehyde into carbohydrate. The induction period for this reaction is explained by the need first to make some glycolaldehyde through a (slow) reaction between two formaldehyde molecules.

5. *Carbonyl shifts:* for example between aldopentoses and ketopentoses:



6. *α-Carbon epimerisation:* for example between D-ribose and D-arabinose:

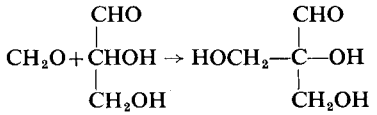


An induction period is typical of the formose reaction: a solution of formaldehyde and lime may sit for an hour with apparently nothing happening and then quite suddenly start to turn brown, which indicates that sugars (more strictly products of their further reaction to 'caramel') are being formed.



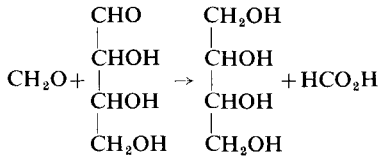
This induction period is explained by an autocatalytic mechanism such as that proposed by Breslow (1959) and shown in figure 1.7. This uses the reaction types 1, 2, 3 and 5 given above, but is very much a minimum mechanism: the true situation is certainly much more complex (Pfeil & Ruckert, 1961; Walker, 1971; Cairns-Smith & Walker, 1974). The various reaction types can, and probably do, produce all the possible sugars including branched sugars.

Indeed from more recent work (Mizuno & Weiss, 1974; Shigemasa, Sakazawa, Nakashima & Matsuura, 1978) it is clear that branched sugars are very often the most important products and it is easier to be selective in favour of branched-chain structures. Mizuno & Weiss describe the formose reaction as 'a unique method for producing branched-chain carbohydrates'. They report that while small proportions of many sugars are produced by a formose reaction catalysed by calcium hydroxide, the major products in the $C_4$–$C_6$ group are branched-chain aldoses and ketoses. Such structures readily arise either from type-2 or type-3 reactions, for example:



In addition to all this Cannizzaro reactions take place such as:

7. *Disproportionation between formaldehyde and a sugar:*



this adds complex polyalcohols – very often branched polyalcohols – to the formose mixture and these are often the predominant products (Shigemasa *et al.*, 1978). 'Caramelisation' reactions are a further complication; but even if you count only sugars the products are very complex indeed. For example Ruckert, Pfeil & Scharf (1965) identified 27 species by paper chromatography. A gas chromatogram of a trimethylsilylated formose mixture is illustrated in figure 1.8.

The oligomerisation of $CH_2O$, like that of HCN, produces, tantalisingly, some of the key molecules needed for our kind of biochemistry; but the products are very much more complicated mixtures than had previously been thought, containing only small amounts of natural sugars. Again with $CH_2O$, as with HCN, there are problems of concentration: the

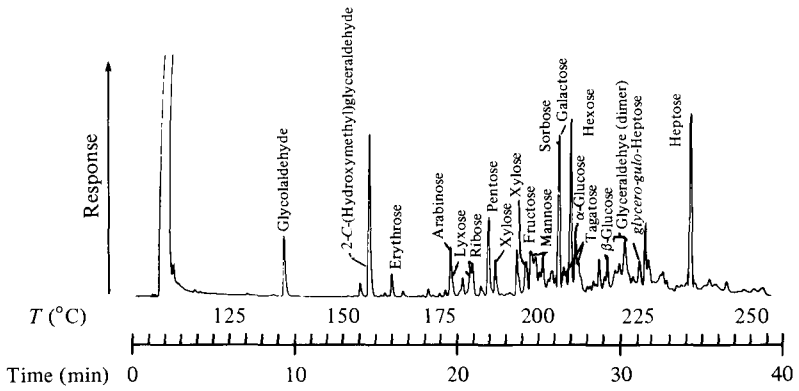



Figure 1.8. Gas chromatogram of a trimethylsilylated formose. (From Mizuno & Weiss, 1974.)

formose reaction has not been demonstrated with solutions more dilute than 0.01 M (Miller & Orgel, 1974). Because formaldehyde is so generally reactive it is hard to see how large volumes of sufficiently concentrated solutions could have accumulated anywhere on the primitive Earth.

*Carboxylic acids* Formic acid is perhaps the easiest of all organic molecules to synthesise under conditions simulating those on the primitive Earth. It was the major product in Miller's original experiment (figure 1.3): it was produced in 22 % yield by reduction of $CO_2$ by ionising radiation (Garrison *et al.*, 1951). Formic acid was produced also on ultraviolet irradiation of CO and $H_2O$ in the presence of silicates (Hubbard *et al.*, 1971, 1973).

Getoff (1962, 1963a, b) suggested that in the reduction of $CO_2$ in aqueous $Fe^{2+}$ solutions by ultraviolet radiation the effective reducing agent was the solvated electron and he proposed the following scheme for the production of formic and oxalic acids:

$$H^+ + e^-_{aq} \rightarrow H\cdot + H_2O$$
$$CO_2 + e^-_{aq} \rightarrow \cdot CO_2^- + H_2O \text{ (with } Fe^{2+} \text{ present)}$$
$$\cdot CO_2^- + H\cdot \rightarrow CO_2H^- \text{ (or } CO + OH^-)$$
$$\cdot CO_2^- + H^+ \rightarrow \cdot CO_2H$$
$$CO_2H^- + H^+ \rightarrow HCO_2H$$
$$2\cdot CO_2H \rightarrow \begin{matrix} CO_2H \\ | \\ CO_2H \end{matrix}$$

The intermediate carboxyl radical is generated also by ultraviolet irradiation of CO, $CO_2$, or both, adsorbed on silica gel (Tseng & Chang, 1975).



Formic acid is also produced by hydrolysis of cyanide (as discussed already) or CO:

$$CO + H_2O \rightarrow HCO_2H$$

In any complex system where redox processes are taking place – for example in the formose–Cannizzaro reactions – formic acid will often tend to be made and then persist since formate is not easily reduced.

There are fewer routes to acetic acid but it, and somewhat higher acids, are reasonably accessible. Propionic acid, for example, is ten times as abundant as glycine in the Murchison meteorite. On the other hand, the amount of any carboxylic acid in this meteorite falls away fairly rapidly as the molecular weight increases (Lawless & Yuen, 1979).

Long straight-chain fatty acids have proved to be very difficult to make under plausibly prevital conditions. Allen & Ponnamperuma (1967) obtained some monocarboxylic acids in the $C_2$–$C_{12}$ range by exposing a mixture of methane and water to a semicorona discharge: but branched chains seem to be produced this way. Leach, Nooner & Oró (1978) obtained straight-chain fatty acids under the rather extreme conditions of a 'Fischer–Tropsch-type' synthesis, that is, catalytic reaction of CO and $H_2$ at high temperatures (here in the range 375–485 °C). Yields in terms of initial CO were in the range of 0.002–0.05 %.

The small hydroxy acids, glycolic and lactic acid, were major products from Miller's experiments (figure 1.3) and occur in the Murchison meteorite in similar amounts to the corresponding amino acids (Lawless, 1980).

Radiolysis and ultraviolet photolysis of aqueous solutions of acetic acid have been shown to give polycarboxylic acids including succinic, tricarballylic and malonic acid (Negrón-Mendoza & Ponnamperuma, 1978). Yields, although not quoted, appear to be small. Fourteen dicarboxylic acids have been identified in the Murchison meteorite (Lawless *et al.*, 1974).

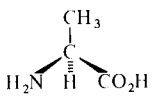
Production of glycerol by uncontrolled reaction does not appear to have attracted much attention: it is a minor product of the formose–Cannizzaro reaction, but that is not in itself very promising. Perhaps early membranes could have been made with less elaborate molecules than our lipids: but anything at all like our membranes would presumably depend on the consistent synthesis of long alkyl chains, or at least of large regular hydrocarbon moieties of some sort. In the event, on the basis of current evidence, even the simplest kinds of soapy micelles would be difficult to envisage on the prevital Earth.

### Chirality

If the production of relatively small molecules on the primitive Earth is to be seen in the light of chemical evolution as part of a progression through polymers to higher order systems, it is high time to pause to consider more seriously the question of chirality ('handedness', see figure 1.6). Unless the chirality of monomers, such as amino acids and sugars, is defined no well organised polymers can be made from them.

Through the looking glass there is that other world of objects that are obviously similar to the objects of our world – but very few are exactly the same. Most objects, that is, are chiral: you can not superimpose them on their mirror images. Small molecules are often sufficiently symmetrical to have superposable mirror images, and human artefacts are often approximately mirror symmetrical; but most things in Nature – pebbles, mountains, even moderately sized molecules – are very unlikely to be identical to their mirror images.

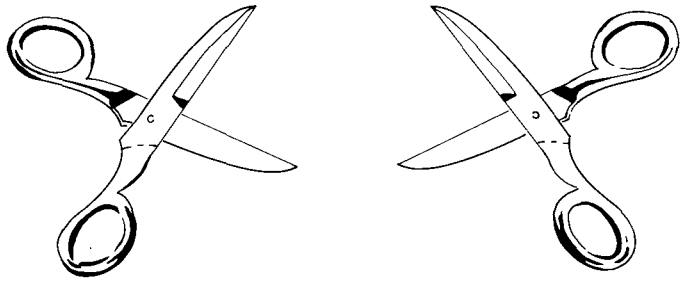
There is nothing particular, then, in an organism not being superposable on its mirror image: nor is it in the least odd to find chiral molecules in organisms – since organisms contain many very large molecules. What is odd is that the chirality of a given molecule in an organism is, on the whole, always the same. You can imagine the mirror image of a haemoglobin molecule, but you will never find a real molecule corresponding to it. Even among the smallest chiral molecules you will find at least a very strong predominance in organisms of one of the enantiomers – that is, for the 'left-handed' or 'right-handed' version of a particular chiral molecule. Sugars are like this – they are with few exceptions D-, that is, related to D-glyceraldehyde (figure 1.6). So are the amino acids (except glycine, which is achiral). Almost all amino acids, and all protein amino acids, are L- with their groups attached to the central carbon atom as in L-alanine:




We have to dig a little to get to the essential oddness of this situation. The teleological question 'why are organisms like this?' is quite easy to answer. Organisms are machines with molecules as their smallest working parts. In so far as these parts make contact interactions it is important that the chirality of the parts is specified, just as it is important that other features – materials and dimensions for example – are specified.

Even to make such a simple machine as a pair of scissors you have to be careful – because the normal design of scissors is chiral. The scissors in your hand and the scissors in the mirror would both work: but a pair made up of one blade from each would not (see figure 1.9).





A B

Figure 1.9. Each of the blades of a pair of scissors like these has to be either 'right-handed' or 'left-handed'. Either enantiomer, A or B, will do, but the diastereomer consisting of one 'right-handed' and one 'left-handed' blade would not work (cf. caption to figure 1.6).

You could design scissors with achiral blades if you wanted to; you might even be able to design a typewriter whose components (apart from the typeface) were achiral. But as the machinery becomes more complex the restrictions of having to use achiral components would be increasingly troublesome. Apart from anything else you would not be able to use nuts and bolts or screws.

It would be even more difficult to design a complex machine – say a typewriter – which contained chiral components but for which it did not matter whether these components were 'left-handed' or 'right-handed'. I do not know how you would design such a 'racemic' machine when even D, L-scissors don't work.

The situation, then, seems to be this: (i) If a complex machine is to be assembled from premade components, then it is likely that, at least for an optimal design, some of these components will be chiral. We only have to think about the complexity of modern biomolecular machinery to be unsurprised that Nature uses chiral components. (ii) We should not be surprised that, using chiral components, Nature has tight conventions about which enantiomer of each molecule to use.

But you might ask why, given that some set of chiral conventions is a good idea, we have life as it is rather than its mirror image (or both). This can be seen as part of a bigger puzzle: not only does all life on Earth agree about the chirality of the most central small molecules, it shares many

other quite intricate details of structure as well. The metabolic pathways are similar, and the choice of protein amino acids is identical throughout life on Earth – and the genetic code is very nearly universal. This points to an explanation in terms of descent of all life now extant from a common ancestor that was already highly evolved biochemically. As part of the complex package that we have all inherited from this last common ancestor there happened to be a particular set of chiral choices. It may well be that what gave our common ancestor the edge over organisms using other biochemical systems (that happened perhaps to have other chiral choices) had nothing to do with chirality: perhaps our ancestor discovered a neater way of fixing carbon dioxide, or happened first on the ribosome.

In Chapter 3 we will come to consider the interesting question of why all life should share such a late biochemical ancestor; but for the present the fact is enough: being a highly evolved intricate piece of machinery this ancestor would have had to have made some consistent set of chiral choices. And it is really no surprise that mirror life is not with us. Even if you were to imagine two groups of organisms that were identical except that one was biochemically the mirror image of the other, they would not stay identical for long. Each would be evolving separately, experiencing a different mutational history. One or the other would win: only one or the other would produce that particular design that was (much later still) to become the common design for all life.

### *The origin of chiral discrimination*

So far we have been mainly concerned with questions of chiral choice. Given that an efficient biochemistry must use chiral molecules, we have been thinking about how our biochemistry might have arrived at some single set of chiral choices. We may have little idea as to why it settled on the particular set that it did. But in principle, at least, all such questions of chiral choice can be understood in terms of evolution, as arising from some mix of historical accident and biochemical efficiency. These questions are difficult and interesting, but they are not the real nub of the matter. The most central of the nest of puzzles about biochemical chirality is not one of choices – of why early evolving systems chose D- this or L- that: it is how evolving systems came to see any difference between D's and L's in the first place. Before you can choose you must discriminate. The nub question is the origin of chiral discrimination.

Ordinarily, chemical reactions are indifferent to chirality. A reaction, such as the formose reaction, that produces chiral molecules is expected to



make racemic mixtures with 'left-handed' and 'right-handed' forms of the products equally represented. If biochemistry is different in this respect, if cells can treat enantiomers differently, then this is because the molecules in cells, especially the enzymes, are already chiral.

Bonner (1972) has reviewed the long history of this fundamental question of origin. In terms of the doctrine of chemical evolution there are two sorts of explanation that have been given: an external 'abiotic' explanation and an internal 'biotic' explanation, that is:

1 The environment created an asymmetric bias: in particular, at least somewhere, it provided especially L-amino acids and D-sugars.

and 2 The environment always produced racemic mixtures of chiral molecules, but at some stage in their evolution organisms came to select molecules of uniform chirality from these mixtures.

Bernal (1951) suggested that quartz crystals in the primitive environment might have biased a mixture of molecules locally. Quartz crystals come in 'left-handed' and 'right-handed' versions which are self-seeding; so quite extensive regions might well have contained only one of the kinds of quartz, and hence tended to accumulate or destroy enantiomers preferentially. The units out of which quartz crystals are built – silicic acid – are achiral, so in making a quartz crystal chirality is literally being generated. Presumably in any particular case this is from an initial chance that might have gone either way: an initial seed happened to start, say, left-handed quartz and it simply went on that way. Examples of such spontaneous generation of chirality have been found in the laboratory and are discussed by Bonner (1972) and Wald (1957). Wald was very sceptical of this idea, though, doubting whether more than a small bias could thus have been transferred from crystalline minerals to molecules in the environment. Modest effects have been reported since (Bonner, Kavasmaneck, Martin & Flores, 1974, 1975; Bonner & Kavasmaneck, 1976, 1977; Morimoto, Kawashiro & Yoshida, 1978) with 30 % enrichment in the most favourable case (about 1 % is more usual). Wald was sceptical too of the idea that D- and L-isomers might separate through crystallisation in the way that, say, Pasteur (1848) discovered for sodium ammonium tartrate. Wald lists the following problems: there would have to be supersaturated solutions; moderately sized areas would be bound to contain both kinds of crystals (because, unlike the quartz case, D would become more concentrated in solution as L crystallised out, and Pasteur would not have been there to separate the crystals); finally the two enantiomers would have to form

separate crystals, which enantiomers only rarely do. Perhaps Wald overestimated this rarity, but we might add that even if all three of these conditions were satisfied, the products of simulation experiments are so grossly contaminated that it looks as if prevital organic molecules would very seldom have crystallised at all.

Other asymmetric influences have been considered. In laboratory experiments circularly polarised light can show small discriminations in the photochemical decomposition of chiral organic molecules. But it is doubtful if there was much circular polarisation of light in primordial skies: the effect in any case could only have been very feeble. Other circularly polarised photons that have been considered are Bremsstrahlung $\gamma$-rays which are generated when longitudinally polarised $\beta$-decay electrons interact with matter. Thus, it has been suggested, a fundamental asymmetry of matter – expressed in the non-conservation of parity – might have left its mark at the chemical level: Bremsstrahlung photons are always polarised the same way. It has been shown recently (Bonner, Van Dort & Yearian, 1975) that longitudinally polarised electrons can have a different effect on enantiomers. But again the effect found was small – a 1 % enantiomeric bias in the destruction of D, L-leucine.

A recent suggestion that a soup might have been significantly chirally biased comes from the report by Bondy & Harrington (1979) that L-leucine, L-aspartic acid and D-glucose are preferentially adsorbed on bentonite. The total amount of binding was small, but it seemed that the 'correct' enantiomers were bound about ten times as much as the 'incorrect' ones.

Friebele, Shimoyama, Hare & Ponnamperuma (1981) found a very much smaller effect in the adsorption of amino acids by sodium montmorillonite – a 0.5–2 % preference for L-amino acids, while Youatt & Brown (1981), reporting negative results, suggest that the original effects may have arisen from an artefact of the technique being used.

Clays such as these would have adsorbed organic molecules on the primitive Earth as Bernal (1951) was the first to point out. It is not clear how such clays could have shown net asymmetric effects. One would have expected that there would have been a balanced mix of D-binding and L-binding features in clays on the primitive Earth – unless there had been a chiral bias in some particular environment within which crystallites were forming. Even if it turns out that some modern clays do show chiral preferences in certain cases, it will not necessarily follow that clays formed on the primitive Earth would have been chirally selective. Modern (chiral) biochemicals and their degradation products are very commonly present in the environment within which modern clays form. And we know that chiral organic molecules can impose a chiral bias on crystals forming in their presence (Pincock, Bradshaw & Perkins, 1974).



The most radical cause that has been suggested for a possible asymmetry in the early environment is that enantiomers are not, after all, chemically equivalent – only very nearly so (Yamagata, 1966). The idea here is that the fundamental asymmetry between matter and antimatter might show up directly, creating differences in the energies of enantiomers (the true enantiomer of, say, L-alanine being D-alanine made of antimatter). For example reaction rates, melting points and solubilities might be slightly different. (They cannot be very different or the phenomenon would be well known.) Evidence for such effects has been presented recently and amplification devices – for example through polymerisation or precipitation – have been discussed (see papers in the *Journal of Molecular Evolution*, 1974, and *Origins of Life*, 1981, in particular, papers by Thiemann and by Wagener; also Thiemann & Darge, 1974).

So where does this leave us: supposing that there was a primordial soup, was it chirally biased? 'I don't know' is clearly the best answer; but next to that we might venture that it might have been very slightly asymmetric in a global sense – because of possible minor effects arising from either Bremsstrahlung or non-equivalence of enantiomers or both; and since these effects would act consistently in one direction. Somewhat stronger biases might have existed locally, but these would probably not have persisted for very long by geological standards. In any case, anything more than a very moderate bias can not reasonably be hoped for within any moderately sized persisting environment.

The trouble is that a moderate bias in favour of particular enantiomers in the primitive environment would not have been nearly good enough. It might indeed be used, in 'biotic' theories, to explain why life was to end up with L-amino acids and D-sugars: life evolving under conditions in which one set of enantiomers are easier to come by than another might well eventually be pushed one way. But, as we saw, that is not the real problem: even without an imposed bias evolving organisms would flip one way or the other. The real problem is right at the start: it lies in the seemingly hopeless incompetence of mixed D/L machines on the one hand and in the need, seemingly, for working machinery before natural selection can get started. This problem afflicts too those 'biotic' theories where chiral discrimination is acquired gradually during evolution.

What would we mean by saying that evolving systems had acquired chiral discrimination? We would mean (i) that they were able to distinguish, if only partially, the enantiomers of some molecules. Operationally

this might show up in a particular appetite that the systems were showing for, say, L-alanine rather than D-alanine. Then (ii) we would mean that this peculiarity was indefinitely transmissable to offspring. There have to be, in some sense, receptors in the systems that are stereospecific to some degree to some molecules; and then these receptors must be reproducible, or the means by which they are produced must be reproducible. Condition (i) says what chiral discrimination is, while condition (ii) defines 'acquire'. It is because of condition (ii) that simple fluctuations within the systems are no good: the adventitious acquisition of a quartz crystallite, say, or a polymer tangle that momentarily happened to have a crevice in it that bound L-alanine. To reappear in subsequent generations, something much more distinct is required: some sort of hereditary machinery. The problem is how to make such machinery in the first place from parts that have only a slightly better than evens chance of having a given chirality (on the 'abiotic' model) or an evens chance (on the 'biotic' model).

This is a serious problem: for molecular machinery to work at all precisely – and indefinite heritability of characteristics would seem to need precisely working machinery – it looks as if there must be multiple contact interactions between components. But in that case, if the components are chiral, then their chiralities would have to be specified. I cannot be dogmatic, but I can no more imagine an effective racemic molecular biology than I can imagine an effective racemic typewriter (and the same goes for biased D, L molecular biologies or biased D, L typewriters). There is another possibility though. Let us stay with the typewriters.

People who assemble typewriters in factories may not have realised it, but there is an ingenious device of the management in such factories for cutting down errors in assembly: where components like screws or little bent levers are chiral only one of the enantiomers is provided. The situation is analogous to that presumed in the purely 'abiotic' theory of the origin of chiral discrimination. Mistakes that might have arisen through the lack of chiral discrimination initially were simply prevented by there being no choice. 'The Management' in the form of prior abiotic processes had it rigged.

Suppose that you had a typewriter factory in which the unfortunate operatives were provided with racemic boxes of nuts and bolts, racemic heaps of little bent levers and so on with which to make the machines. Even if the operatives could not tell right from left, provided they got the relationships between the parts correct, the result could be a production of competent typewriters – with, incidentally, half of them right-handed and half of them left-handed. But it would be a slow business. Half the times



when trying to put in a chiral component the operative would find that its relationship to the preassembled parts was wrong and he would have to try again. And, of course, this would only work if the operatives knew when a relationship was correct: we are assuming knowledge on their part that could not be ascribed to that operative, kinetic motion, that had to put together the first evolving systems. But let us go a little further and suppose that an over-ingenious management, instead of providing chirally pure components, had designed the typewriter in such a way that each piece as it was added would only fit if it was of the correct chirality in relation to the already assembled parts.

If we are to avoid the need either for a well resolved soup to begin with, or alternatively, for chirally mixed if not racemic initial hereditary systems, then I think that something like this sort of foolproof jigsaw puzzle of a hereditary machine is what we must be thinking about. Once an initial system has started with one chirality it must have to go on with the same chirality. And I mean *have* to, as near as matters. Anything less than that would not create a long-term persisting effect: there would just be trends that would fluctuate to and fro.

If this is correct, then it places very severe restrictions on plausible initial hereditary mechanisms. Under what circumstance can molecules self-assemble into a structure that is determined (a) by the form of the components and (b) by the way its assembly happened to start?

Do protein or nucleic acid molecules fall into this restricted class? Are there conditions under which, say, D, L mixtures of amino acids will polymerise into a mixture of D- and L-polypeptides rather than D, L-polypeptides? Or, if you seed such a reaction mixture with, say, L-polypeptide, does this tend to induce specifically the polymerisation of the L-amino acids? Wald (1957) discusses experiments that were designed to answer such questions. These experiments used 4 % solutions in anhydrous dioxan of the activated glutamic acid derivative $\gamma$-benzyl glutamic acid anhydride (Blout & Idelson, 1956; Doty & Lundberg, 1956). The results of these experiments could be taken to support the idea that life would be much better off making a decision one way or the other about D or L: the pure enantiomers of the monomer polymerised much more quickly than D, L mixtures. The reason was, apparently, that the pure enantiomers gave a more stable $\alpha$-helix and that the formation of such an $\alpha$-helix had an accelerating effect on polymer growth. There was little selectivity, however, between D and L: racemic monomers gave D, L-polymers rather than a mixture of D-polymers and L-polymers.

There are two reasons why selective seeding would be expected to be

much less effective for such polymer growth processes than for crystal growth processes. First there are more ways of going wrong. The configuration of the previous unit may influence the relative rates of incorporation of new D and L units, but polymers generally are too flexible, there is too much room around them, for this influence to be decisive: even the relatively rigid $\alpha$-helix was not able to insist that a new amino acid adding was to be of the same chirality as the units already there.

Then secondly, as Wald points out, the irreversible character of the reaction mitigates against selectivity. Mistakes once made cannot be put right. With crystal growth processes on the other hand, the conditions of packing are very restrictive: it may simply be impossible to fit the wrong kind of molecule or if it is possible the energy difference due to the distortions created is likely to be substantial. Since the forces holding molecules in crystals are generally reversible, the molecules can come and go repeatedly as the crystal is built up: they can feel their way to an ideal lowest energy arrangement. There is a fundamental difficulty here because the main biopolymers are all energetically uphill from their monomers in solution: thus 'error correction', so easy for crystal growth, is difficult and energetically expensive for biopolymers. We will be returning to this point in Chapter 5.

If we imagine replication of a nucleic acid molecule through the attachment of monomers to a separated strand as in Watson & Crick's original vivid picture, then it would be reasonable to suppose, as Wald did, that an initial chirally uniform strand would exert a selectivity on nucleotides in the surroundings: in particular that the template strand would select for the correct chirality (that is, it would select D-ribonucleotides if the template strand is a poly-D-ribonucleotide). As it has turned out, DNA does not replicate like this in modern cells. Instead the process, as it now operates, depends on very elaborate assistance of enzymes. (We will discuss this in detail in Chapter 5.) Nor is there any indication from experiment that a nucleic acid molecule is, on its own, sufficiently discriminating to replicate without some rather elaborate assistance. There has to be an 'operative', it seems, who knows something about how things are to be arranged: the selectivity is partly in the enzymes. That at least is a reasonable view on the basis of current knowledge, although it might be overturned by an experiment tomorrow that might show that there are fairly simple conditions, for example in the presence of some mineral surface, where activated nucleotides would polymerise spontaneously on a nucleic acid template. But no such experiment has yet been successful (Orgel & Lohrmann, 1974).



Part of the problem of the origin of chiral discrimination is to see that there is a problem at all. Always at the back of one's mind is the feeling that there must be some simple physical effect that will work because crystals are often chirally discriminating with such high efficiency. The crystal is the ideal simple example of the sort of foolproof jigsaw puzzle object that we had been thinking about: a crystal structure is implicit in the structure of its units under given conditions plus, in the case of chiral crystals, one extra bit of information that can be provided by a seed – whether to choose right or left. But having started one way it goes on that way.

The difficulties arise when you start to think out in detail how in practice that effect could be used. Its only obvious application is to the 'abiotic' mechanism: but there are severe difficulties with this mechanism as we have seen. To be used in 'biotic' mechanisms there would have to be crystals in the organisms that were of a particular chirality and which grew and broke up to seed subsequent generations and so maintain an original handedness. That sounds all right in principle, and indeed it is an old idea, but in practice it is very difficult to see generations of coacervates keeping with, say, L-glutamic acid, and leaving D-glutamic acid in the soup, through such a trick. (The coacervates would have to create supersaturated conditions; the solutions would have to be reasonably pure to crystallise at all; as time passed D-glutamic acid would accumulate in the surroundings and make adventitious seeding of this other enantiomer more likely. And so on.)

And yet efficient resolution of enantiomers through crystal seeding is at least possible under idealised laboratory conditions. That is more than can be said about resolution through polymerisation processes, where even in the laboratory under the most contrived conditions there are no good examples of polymers that, made from one enantiomer of a chiral monomer, can seed specifically further growth through selective addition of that enantiomer. Yet for Wald's mechanism to work, nothing much less will do.

We might recall that the modern biochemical technique of keeping control of chirality is largely through the specification of 'sockets' in proteins. And it is worth emphasising how unprimordial that technique seems – it looks like a way of maintaining chiral discrimination rather than of originating it. Consider: the control of specific sockets in proteins depends in turn, and in an intricate way, on the control of protein folding. Suppose that somehow a polypeptide sequence had been specified with 100 chiral amino acids in it. This sequence would only specify the folding if the chirality of the amino acids could be relied upon – otherwise there would be $2^{100}$ (i.e. approximately $10^{30}$) different structures possible, with side

chains pointing different ways. This would be bound to affect the folding which depends so critically on side-chain packing. A protein molecule that hardly ever folds the same way twice would be a poor basis for specifying a groove that could feel the difference between L-alanine and D-alanine.

Problems of molecular discrimination in early evolving systems go far beyond those of chirality. (As they do in making machines generally: before such a nice question arises as to whether one has the correct enantiomer of a component there is the cruder question of whether one has the correct component.) To imagine a DNA molecule replicating in the kind of way that Wald supposed, the parent strand must pick out nucleotides in the first place from a presumably complex mixture of molecules that, if it contained nucleotides in any quantity, would contain many nucleotide-like molecules. In a primitive unevolved environment chiral discrimination might seem the least of the problems; yet the discrimination would seemingly have to be that good: only a chirally uniform nucleic acid looks as if it could work (Wald, 1957; Miller & Orgel, 1974). (And these are not the only problems associated with the idea that there were nucleic acid molecules in the very first organisms, as we shall see later.)

Attitudes to the origin of chirality in organisms have varied from surprise that anyone should still think that there is a problem (Frank, 1953) through a refusal to be baffled (Miller & Orgel, 1974) all the way to talk of enigma (Kenyon & Steinman, 1969).

I would agree with Frank to the extent that I do not see the problem of the origin of chirality as raising fundamental philosophical issues. The difficulties are of a practical sort. But they are very great difficulties all the same. Perhaps we might settle for the comment by Briggs with which Bonner concludes his review of the origin of chirality that 'it presents problems to the hypothesis of chemical evolution that are at present insoluble'.

In Part III of this book I will give a possible solution to the problem of the origin of chirality that depends on the idea that the first systems able to evolve under natural selection were hereditary machines made from achiral components. Chiral discrimination could thus be a comparatively late evolutionary achievement since it was not needed to begin with. Without being drawn into details at the moment I can perhaps best indicate what I mean by adapting an analogy of Thiemann & Darge (1974). They see the origin of chirality in organisms as analogous to the spontaneous crystallisation of one enantiomer from a supersaturated racemic solution. On the view that I will be presenting the appropriate analogy becomes rather the spontaneous appearance of enantiomeric crystals (like quartz) from an



achiral mother solution. Chirality does not have to be built into units, it can appear at a higher level of organisation. Thus organisms that start with achiral units can come eventually to be able to handle chiral molecules (which later still take over). Of course such an idea only becomes possible if you are prepared to contemplate altogether different initial hereditary machinery. But if you are prepared to follow through this somewhat more complicated story you will find another advantage: crystallisation, with its error correction mechanisms intact, can become the means through which the initial hereditary machinery worked.

### Prevital polymers

Let us nevertheless pursue the story of the origin of life according to the doctrine of chemical evolution and consider in more detail the question of how biopolymers might have arisen. Which of these would have been the most important to begin with is a matter of some dispute; but it is commonly believed that proteins of a sort, or nucleic acids of a sort (or both) would have been necessary for the making of those first systems that could evolve under natural selection and so take off from the launching platform provided by prevital chemical processes.

We have already come to a major difficulty here: much of the point of protein and the whole point of nucleic acid would seem to be lost unless these molecules have appropriate secondary/tertiary structures; and that is only possible with chirally defined units. As we saw, the 'abiotic' way of circumventing this problem (by prevital resolution of enantiomers) seems hopelessly inadequate, and 'biotic' mechanisms depend on efficient machinery already in action. At all points these difficulties can be seen as technical – some effect, such as a chirally discriminating polymerisation, might be discovered to sweep these difficulties away. If you are a believer in chemical evolution then that is the attitude to take. But there are further severe difficulties ahead. A clutch of them appear when we come to consider how biopolymers might have been made consistently without biology.

#### *Purification*

First of all there is a problem which is seldom discussed. The starting monomers would have been grossly impure. On the basis of simulation experiments they would have been present in complex mixtures that contained a great variety of variously reactive molecules.

No sensible organic chemist would hope to get much out of a reaction

from starting materials that were tars containing the reactants as minor constituents. Perhaps because they are sensible organic chemists most experimenters, in trying to establish some prevital path to biopolymers, do not start with such complex mixtures. Instead they say something like this: 'monomer A has been shown to be formed under prebiotic conditions and so has reagent B: so we treated A and B (obtained from Maxipure Chemical Corporation) under prebiotic conditions such and such and made the biochemically significant molecule C'. Suggestions as to how A and B might have been purified under prevital conditions are seldom made.

In organic chemistry it is often the work-up rather than the reaction that causes most of the trouble. Think about the techniques that are used: pH adjustments, solvent extractions, chromatography, evaporations to dryness, recrystallisations, filtrations and so on. Now you can say that such things might have taken place fortuitously under primitive geological conditions. Each individual operation can be imagined – a transfer of a solution, a washing of a precipitate, an evaporation, and so on. But very many such operations would have had to take place consistently and in the right order. In a typical work-up procedure there are subtle things that can make the difference between success and mess – how long to wait, say, after the pH adjustment before filtering. Practical organic chemistry is not easy. Very much has to be engineered. It is not sensible to suppose that an uninformed geochemistry would fortuitously be expert in such things.

### *Concentration*

Next there is the problem of the concentrations of the monomers in primordial waters. It has been emphasised repeatedly that the idea of an oceanic primordial soup is difficult to sustain on thermodynamic and kinetic grounds. For example Hull (1960) says: 'First, thermodynamic calculations predict vanishingly small concentrations of even the smallest organic compounds. Second, the reactions invoked to synthesise such compounds are seen to be much more effective in decomposition.' Hull was discussing particularly the effects of ultraviolet radiation which he calculated would have destroyed 97 % of amino acids produced in the atmosphere before they reached the oceans. Glycine, he reckoned, would have formed at best a $10^{-12}$ M solution. Sillen (1965) was similarly pessimistic and talked of the 'myth of the probiotic soup', and Abelson (1966) pointed to the lack of any geological evidence for a thick oceanic soup in the past. The matter was pursued again by Hulett (1969) in an influential paper in which he considered various energy sources but stressed particularly photochemical effects. (The main point is that short-wavelength photons are needed for the synthesis of molecules such as aldehydes and amino acids, while the much more abundant longer wavelength photons are effective in decomposing them.) The importance of ultraviolet radiation in destroying organic molecules was again emphasised by Rein, Nir & Stamadiadou (1971). Dose (1974, 1975) has calculated that primitive oceanic concentrations of amino acids would have been around $10^{-7}$ M if photochemical effects are taken into account. Thus Dose points out that the concentration of amino acids in the 'soup' would have been about the same as their concentrations in the oceans now. A similar conclusion has been reached by Nissenbaum (1976) on the basis of other (non-biological) geochemical processes that scavenge organic molecules from the oceans (for example adsorption on sinking minerals). As a result the mean residence time for organic molecules is 1000–3000 years. A further effect that can be seen in action now on Mars would have tended to keep surface regions clean of organic molecules. Ultraviolet radiation that can penetrate an anoxic atmosphere can have an oxidising effect on inorganic surface materials creating species that are ready to destroy any organic molecules that are formed (Ponnamperuma *et al.*, 1977; Klein, 1978).



It can be reasonably argued that an oceanic soup was not needed, that what was required was that stocks of organic molecules of the right kind should have been able to build up somewhere to a sufficient concentration, and consistently enough, to have provided the materials to make the first evolving systems – and to have provided them with food. Perhaps at the margins of oceans there would have been evaporating rock pools that would have served to concentrate materials. It would take a lot of evaporating, though, from $10^{-7}$ M to the kinds of concentrations that might be needed. And the ultraviolet sunlight would have been a nuisance for any concentration process that uses evaporation. Here Bernal's idea seems better. He pointed out that organic molecules would have been adsorbed on solid particles, particularly clays, so that potential monomers would have been brought closer together. Hence, perhaps, amino acids might have been able to form polymers on clays even if the general solutions surrounding them were quite dilute.

Lahav & Chang (1976) have recently taken up this idea in an analysis of published data on the adsorption and condensation on clays of amino acids, purine and pyrimidine bases, sugars, nucleosides and nucleotides. None of these types of molecule binds very strongly to ordinary clays. At equilibrium, between ocean and clay sediments, they would be very thinly spread over the available clay surface. Such estimates are necessarily rough,

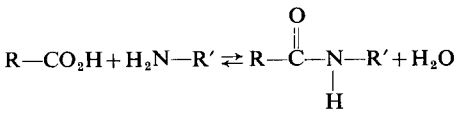
but even supposing that Dose (1975) was pessimistic by two orders of magnitude in his estimate of oceanic amino acid concentrations, the molecules would have been on average at least 10 nm apart. Additional factors would still be needed to concentrate the molecules sufficiently, for example evaporation or freezing.

*Condensation*

There is a third difficulty in prevital synthesis of biopolymers, and this is the most generally recognised: all the major biopolymers are metastable in aqueous solution in relation to their (deactivated) monomers. Left to itself in water, a polypeptide will hydrolyse to its constituent amino acids: Miller & Orgel (1974) estimate that the half life of alanylalanine is about $8 \times 10^7$ years at 0 °C and about $6 \times 10^5$ years at 25 °C.

Most experimental work on prevital simulation of polymer synthesis has been concerned with this third problem. There have been two main approaches.

*1. Devices for shifting the equilibrium.* The idea here is to separate the products, water and polymer, as they are formed so that the reaction is driven to the right, for example:



As this equilibrium is only reached very slowly at ordinary temperatures and near neutral pH, heat, catalyst or both are also needed.

The simplest technique is to heat the amino acids together so that the water is driven off as steam. This tends to decompose the amino acids (Katchalski, 1951), but Fox found that if a relatively high proportion of aspartic acid, glutamic acid or lysine were present in a dry mixture of amino acids, then a clean protein-like material, 'proteinoid', could be formed. The temperature required was usually around 150–180 °C (Fox, 1956, 1969; Fox & Harada, 1960; Fox & Dose, 1972). More recently amino acids have been polymerised at 65–85 °C (Rohlfing, 1976). Protenoid shows catalytic properties (see also Dose, 1971) and on contact with water forms very striking cell-like objects ('microspheres') which Fox sees as models for systems that are on the way to becoming cells.

Lahav, White & Chang (1978) have made oligomers (dimers and trimers mainly) of amino acids. In a typical experiment 1 ml of 0.023 M glycine



Table 1.3. *Oligomerisation of glycine on clays (after Lahav, White & Chang, 1978)*

| Expt no. | Glycine (nmoles/ mg clay) | No. of cycles | Net heating time (days) | Approx. yields of oligomers (nmoles/mg clay) Di | Tri | Tetra | Penta | Total yield of glycine in oligomers nmoles/ mg clay | approx. % |
|---|---|---|---|---|---|---|---|---|---|
| Kaolinite | | | | | | | | | |
| 1 | 374 | 11 | 33.7 | 2.3 | 0.5 | — | — | 5.9±0.9 | 1.6 |
| 3 | 374 | 27 | 67.4 | 2.3 | 1.0 | 0.3 | Trace | 8.9±0.6 | 2.4 |
| 5 | 123 | 27 | 67.4 | 1.0 | 0.4 | 0.1 | Trace | 3.5±0.3 | 2.8 |
| 7 | 791 | 27 | 67.4 | 3.5 | 1.6 | 0.6 | Trace | 14.14 | 1.8 |
| Bentonite | | | | | | | | | |
| 15 | 1070 | 11 | 32.8 | 6.4 | 0.2 | — | — | 13.3±1.7 | 1.2 |
| 17 | 1070 | 27 | 67.4 | 4.9 | 0.6 | Trace | — | 11.7±1.7 | 1.1 |
| 20 | 32000 | 11 | 57.0 | 37 | 8.2 | — | 2.5 | 111±28 | 0.3 |
| 23 | 93000 | 11 | 57.0 | 40 | 7.9 | 1.2 | 0.8 | 113±10 | 0.1 |

was added to 60 mg of kaolinite or 20 mg bentonite ($Na^+$ form) and put through cycles of wetting and drying, and temperature changes, as follows:

(i) dehydrate at 60 °C for 1–2 days
(ii) heat at 94 °C for 2–3 days
(iii) rehydrate with 1 ml of water
(iv) repeat (i)–(iii) *n* times

Some results are given in Table 1.3. One might imagine inland or coastal regions on the primitive Earth that were being alternately flooded and baked dry in the sun.

In similar experiments with bentonite Lawless & Levi (1979) have shown that yields may be improved considerably by transition metal ions in the exchange positions of the clays – $Cu^{2+}$ in particular, but also $Zn^{2+}$ and $Ni^{2+}$ (see Table 1.4).

As with the formation of proteinoid, the tyranny of thermodynamics is avoided in the above experiments by using open systems. In each case water is being removed and energy used to do that. The clays are also, presumably, exerting catalytic effects – particularly those clays which have exchangeable transition metal cations.

*2. Raising the energy of the monomer.* One problem with trying to polymerise amino acids at around neutral pH in water is that under these

Table 1.4. *Influence of exchangeable cations on the oligomerisation of glycine on bentonite*

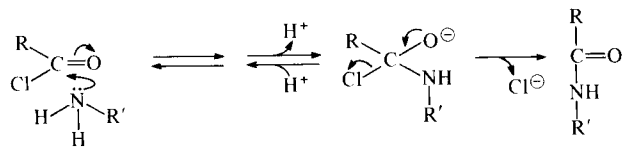
| Exchangeable cation | Yields of oligomers (nmoles glycine incorporated/mg clay) Di | Tri | Tetra | Penta | Glycine incorporated into oligomer (%) |
|---|---|---|---|---|---|
| $Na^+$ | 17.2 | — | — | — | 0.9 |
| $Ni^{2+}$ | 29.5 | 6.3 | — | — | 1.8 |
| $Cu^{2+}$ | 69.9 | 41.4 | 10.4 | 2.8 | 6.2 |
| $Zn^{2+}$ | 25.6 | 3.3 | — | — | 1.4 |

2 ml of $2 \times 10^{-2}$ M glycine at neutral pH were added to 20 mg of bentonite in each case and put through eight wet-dry cycles (after Lawless & Levi, 1979).

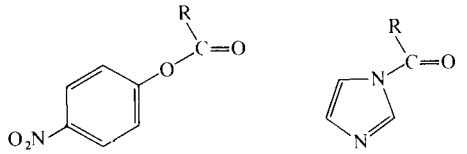
conditions neither the amino nor the acid group is very reactive – because the amino acid molecules spend most of their time as zwitterions:

When the molecule is in the zwitterion form the negative charge on the carboxyl group destroys the electrophilic character of its carbon atom. At the same time the nucleophilic character of the nitrogen is destroyed by protonation of its lone pair of electrons.

The reactivity of the carboxyl can be increased by changing its hydroxyl for a group that does not have a proton to lose – a methyl ester, for example, is a mildly activated form of the carboxyl group. An acid chloride is much more strongly activated since in this case the hydroxyl has been replaced with an electron-withdrawing group that can easily leave as an anion to push the reaction in the required direction:

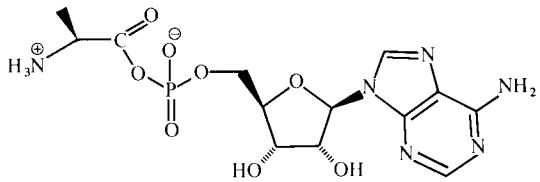


Indeed acid chlorides are rather too reactive to be useful in peptide synthesis – they react too easily with alcohols and with water for example. Somewhat milder forms of carboxyl activation are more selective for reaction with amines, and numerous alternative leaving groups have been used in peptide syntheses (Bodansky, Klausner & Ondetti, 1976). Two examples are *p*-nitrophenyl esters and imidazole amides:



There are several problems, though, in attempting to apply this idea to primitive chemistry. Firstly, energy must come from somewhere to activate an amino acid. Secondly, even mildly activated species are more or less susceptible to hydrolysis. Thirdly, the amino group of an amino acid will usually have to be protected during activation if only to prevent premature polymerisation. This means that separate deprotecting steps will also be required. In addition side groups of the more reactive amino acids will have to be protected and deprotected.

Amino acyl adenylates are activated forms of amino acids involved in protein biosynthesis (figure 9.9). For example alanyl adenylate is:



As we will discuss in the next section, such molecules would be very implausible prevital products but, interestingly, they can be made to polymerise in aqueous solutions in the presence of the clay montmorillonite (Paecht-Horowitz, Berger & Katchalsky, 1970; Paecht-Horowitz, 1974; although, see also Brack, 1976).

Instead of using isolable activated intermediates, amide links can be formed by means of coupling or condensing agents such as carbodiimides. These coupling agents first join to the carboxyl to create a good leaving group which is then expelled following attack by the amino nitrogen (see figure 1.10). Carbodiimides can also be used in making internucleotide phosphodiester bonds. Thus, with a coupling agent, activation and condensation can be performed in one operation.

Cyanamide is a tautomer of carbodiimide itself:

$$H_2N{-}C{\equiv}N \rightleftarrows HN{=}C{=}NH$$

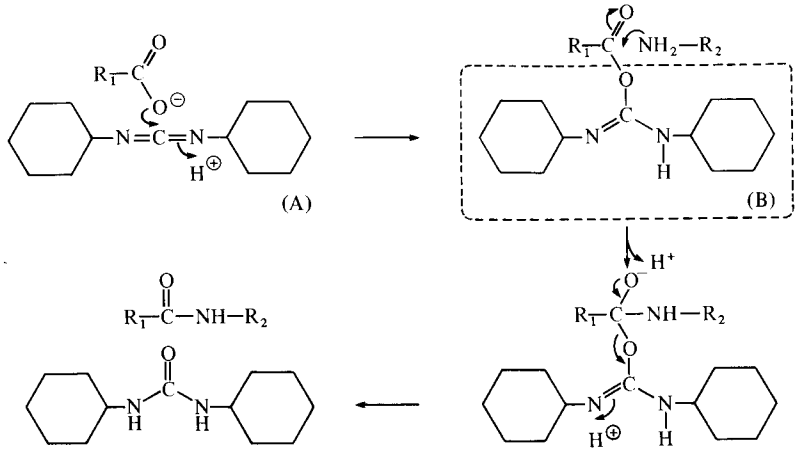



Figure 1.10. Dicyclohexylcarbodiimide (A) is a typical coupling agent. In joining an acid to an amine it reacts first with the acid to make a leaving group (B).

and has been suggested as a primordial coupling agent. Small yields of triglycine, leucylglycine and glycylleucine have been obtained with this reagent (Ponnamperuma & Peterson, 1965), but under conditions more acid than would have been expected in primordial oceans. Ponnamperuma (1978) has reviewed the possible prevital condensing agents shown in Table 1.5. Of the first five cyanide-derived agents he considers that only hydrogen cyanide tetramer operates at a sufficiently high pH to be realistic. This agent on the other hand is rather unstable. Ponnamperuma concludes that linear or cyclic polyphosphates are the best idea. On the other hand, Sherwood *et al.* (1978) suggest that the regions between clay layers might sometimes be sufficiently acidic to allow condensation reactions mediated by cyanamide.

It seems to me that the idea of coupling agents putting together polypeptides on a lifeless Earth adds another dimension of unreality to an already unreal line of thought. Remember that primordial simulations generally give only low yields of amino acids. Remember that the products are tars and that suggestions for prevital work-up procedures are usually absent. Remember the difficulties anyway in building up concentrations of solutions of amino acids or of the cyanide or phosphate to make a coupling agent. Remember that even from laboratory bottles the agents in question do not work very well. Remembering all that, now add the thought that coupling agents are rather unspecific. If a well chosen coupling agent under



Table 1.5. *Some possible prevital condensing agents (after Ponnamperuma, 1978)*

| | |
|---|---|
| Cyanamide | $H_2N-C\equiv N$ |
| Carbodiimides | $R-N=C=N-R$ |
| Dicyanamide | $N\equiv C-\underset{\displaystyle H}{\underset{\vert}{N}}-C\equiv N$ |
| Dicyandiamide | $H_2N-\overset{\displaystyle NH}{\overset{\Vert}{C}}-\underset{\displaystyle H}{\underset{\vert}{N}}-C\equiv N$ |
| Diaminomaleonitrile | $N\equiv C-\overset{\displaystyle NH_2}{\overset{\vert}{C}}=\overset{\displaystyle H_2N}{\overset{\vert}{C}}-C\equiv N$ |
| Linear polyphosphates | $O-\overset{\displaystyle O}{\overset{\vert}{P}}(=O)-O-\left[\overset{\displaystyle O}{\overset{\vert}{P}}(=O)-O\right]_n-\overset{\displaystyle O}{\overset{\vert}{P}}(=O)-O^{(n+4)\ominus}$ |
| Cyclic polyphosphates | $\left[\text{cyclo-}(P_3O_9)\right]^{(n+2)\ominus}$ |

well chosen laboratory conditions can effectively join the acyl group A to the nucleophile B that is because among the choices exercised by the experimenter was the crucial one of only putting A and B into a flask for the coupling agent to couple. Compared with such carefully arranged marriages the affairs of a primordial soup would have been grossly promiscuous.

One can get an impression of what is needed in practice for the synthesis of peptides by considering the machinery that is used in automated procedures. One such piece of equipment is shown in figure 1.11. Merrifield, Stewart & Jernberg (1966) describe its construction and operation in nine close pages of diagrams and descriptions. I quote (more or less at random) from the middle of their paper: '. . . the rear disk contains a center port

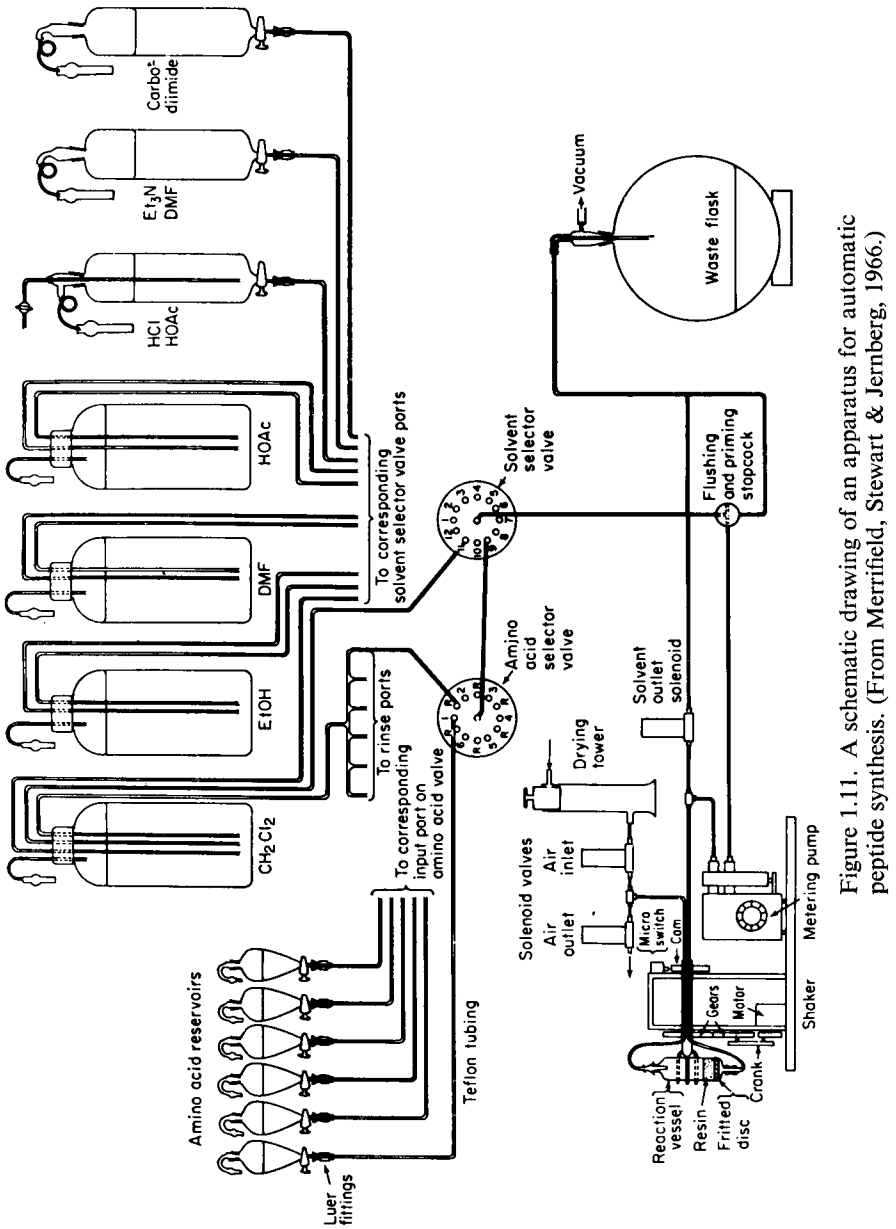



Figure 1.11. A schematic drawing of an apparatus for automatic peptide synthesis. (From Merrifield, Stewart & Jernberg, 1966.)



and one circumferential port which are joined by a 1.5 mm hole within the disk. As this disk is turned it connects one at a time the 12 inlet ports to the central outlet port. A leak-free seal between the two teflon disks of the valve . . .' And that is one of the less terse passages. Not shown in figure 1.11 is a programmer, like a musical box drum, that puts appropriate operations (mixings, rinsings, shakings, etc.) in sequence. There have to be many pegs on the drum because one cycle of the automatic synthetic procedure that extends the peptide chain by one unit requires nearly 90 steps.

Now I am not saying that for peptide synthesis without human intervention there has to be something physically like Merrifield's machine. There does not have to be that particular piece of engineering. But I think there has to be engineering.

Another example of automatic peptide synthesis is the synthesis by the ribosome in the modern cell. (We will be considering this in more detail in Chapter 9.) There are no tubes or valves or metering pumps here: but in the design of the ribosome, the adaptor RNA molecules and their activating enzymes; in the whole system, with its message tapes and its code, there is surely at least as much engineering as in Merrifield's machine. (See figures 9.6–9.10.)

Perhaps there is some other way of making peptides with more or less specified amino acid sequences; and perhaps this way does not need detailed control. Perhaps it could then have operated before there was life on Earth, before that engineer, natural selection, appeared on the scene. But it is difficult to see how this could have been so. I think we would know by now if there was some much easier way.

It is similarly difficult to imagine anything like polysaccharide being accumulated in primordial waters. As we saw, the monosaccharides could only have been made easily from formaldehyde, as far as anyone knows, and there is doubt if there could have been sufficient concentrations of that. In any case, as we saw, the product of the formose reaction is a very complex mixture that easily leads to higher polymers and to caramel.

More realistic is the suggestion by Kenyon & Nissenbaum (1976) that more random kinds of polymeric material 'melanoidin' and 'aldocyanoin' were implicated in the earliest precursors of life. Melanoidin was made by heating solutions of a sugar (e.g. 0.1 M glucose) and an amino acid (also 0.1 M) at around 100 °C for 2–4 days at around neutral pH. Aldocyanoin was made by keeping a solution of NaCN, $NH_4Cl$, $CH_2O$ and $NH_4SCN$ (each 0.2 M) at pH 9.3 in a stoppered flask at room temperature for 2–4 weeks. Each of these kinds of polymer gave microspheres – the melanoidin made two distinct sizes (about 3 $\mu$m and 8 $\mu$m) while in the turbid dark-

brown aldocyanoin reaction mixtures the microspheres were more uniform (at about 2 $\mu$m).

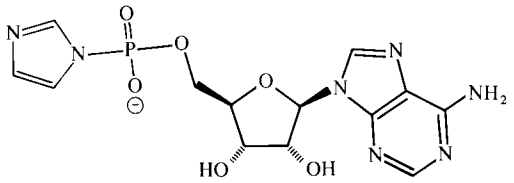
For such polymers to form, concentration mechanisms would still be needed; but presumably impure starting materials would suffice for these much less ambitious products. Then again the energy problems are lighter – at least if we can assume a supply of cyanide and formaldehyde since these products are energetically downhill from there.

### The implausibility of prevital nucleic acid

If it is hard to imagine polypeptides or polysaccharides in primordial waters it is harder still to imagine polynucleotides. But so powerful has been the effect of Miller's experiment on the scientific imagination that to read some of the literature on the origin of life (including many elementary texts) you might think that it had been well demonstrated that nucleotides were probable constituents of a primordial soup and hence that prevital nucleic acid replication was a plausible speculation based on the results of experiments. There have indeed been many interesting and detailed experiments in this area. But the importance of this work lies, to my mind, not in demonstrating how nucleotides could have formed on the primitive Earth, but in precisely the opposite: these experiments allow us to see, in much greater detail than would otherwise have been possible, just why prevital nucleic acids are highly implausible.

Let us consider some of the difficulties. *First*, as we have seen, it is not even clear that the primitive Earth would have generated and maintained organic molecules. All that we can say is that there might have been prevital organic chemistry going on, at least in special locations. *Second*, high-energy precursors of purines and pyrimidines had to be produced in a sufficiently concentrated form (for example at least 0.01 M HCN). *Third*, the conditions must now have been right for reactions to give perceptible yields of at least two bases that could pair with each other. *Fourth*, these bases must then have been separated from the confusing jumble of similar molecules that would also have been made, and the solutions must have been sufficiently concentrated. *Fifth*, in some other location a formaldehyde concentration of above 0.01 M must have built up. *Sixth*, this accumulated formaldehyde had to oligomerise to sugars. *Seventh*, somehow the sugars must have been separated and resolved, so as to give a moderately good concentration of, for example, D-ribose. *Eighth*, bases and sugars must now have come together. *Ninth*, they must have been induced to react to make nucleosides. (There are no known ways of bringing about this thermodynamically uphill reaction in aqueous solution: purine nucleosides have been made by dry-phase synthesis, but not even this method has been successful for condensing pyrimidine bases and ribose to give nucleosides (Orgel & Lohrmann, 1974).) *Tenth*, whatever the mode of joining base and sugar it had to be between the correct nitrogen atom of the base and the correct carbon atom of the sugar. This junction will fix the pentose sugar as either the $\alpha$- or $\beta$-anomer of either the furanose or pyranose forms (see page 29). For nucleic acids it has to be the $\beta$-furanose. (In the dry-phase purine nucleoside syntheses referred to above, all four of these isomers were present with never more than 8 % of the correct structure.) *Eleventh*, phosphate must have been, or must now come to have been, present at reasonable concentrations. (The concentrations in the oceans would have been very low, so we must think about special situations – evaporating lagoons and such things (Ponnamperuma, 1978).) *Twelfth*, the phosphate must be activated in some way – for example as a linear or cyclic polyphosphate – so that (energetically uphill) phosphorylation of the nucleoside is possible. *Thirteenth*, to make standard nucleotides only the 5′-hydroxyl of the ribose should be phosphorylated. (In solid-state reactions with urea and inorganic phosphates as a phosphorylating agent, this was the dominant species to begin with (Lohrmann & Orgel, 1971). Longer heating gave the nucleoside cyclic 2′,3′-phosphate as the major product although various dinucleotide derivatives and nucleoside polyphosphates are also formed (Österberg, Orgel & Lohrmann, 1973).) *Fourteenth*, if not already activated – for example as the cyclic 2′,3′-phosphate – the nucleotides must now be activated (for example with polyphosphate; Lohrmann, 1976) and a reasonably pure solution of these species created of reasonable concentration. Alternatively, a suitable coupling agent must now have been fed into the system. *Fifteenth*, the activated nucleotides (or the nucleotides with coupling agent) must now have polymerised. Initially this must have happened without a pre-existing polynucleotide template (this has proved very difficult to simulate (Orgel & Lohrmann, 1974)); but more important, it must have come to take place on pre-existing polynucleotides if the key function of transmitting information to daughter molecules was to be achieved by abiotic means. This has proved difficult too. Orgel & Lohrmann give three main classes of problem. (i) While it has been shown that adenosine derivatives form stable helical structures with poly(U) – they are in fact triple helixes – and while this enhances the condensation of adenylic acid with either adenosine or another adenylic acid – mainly to di(A) – stable helical structures were not formed when either poly(A) or poly(G) were used as templates. (ii) It was difficult to find a suitable means of

making the internucleotide bonds. Specially designed water-soluble carbodiimides were used in the experiments described above, but the obvious pre-activated nucleotides – ATP or cyclic 2′,3′-phosphates – were unsatisfactory. Nucleoside 5′-phosphorimidazolides, for example:



were more successful, but these now involve further steps and a supply of imidazole, for their synthesis (Lohrmann & Orgel, 1978). (iii) Internucleotide bonds formed on a template are usually a mixture of 2′–5′ and the normal 3′–5′ types. Often the 2′–5′ bonds predominate although it has been found that $Zn^{2+}$, as well as acting as an efficient catalyst for the template-directed oligomerisation of guanosine 5′-phosphorimidazolide also leads to a preference for the 3′–5′ bonds (Lohrmann, Bridson & Orgel, 1980). *Sixteenth*, the physical and chemical environment must at all times have been suitable – for example the pH, the temperature, the $M^{2+}$ concentrations. *Seventeenth*, all reactions must have taken place well out of the ultraviolet sunlight; that is, not only away from its direct, highly destructive effects on nucleic acid-like molecules, but away too from the radicals produced by the sunlight, and from the various longer lived reactive species produced by these radicals. *Eighteenth*, unlike polypeptides, where you can easily imagine functions for imprecisely made products (for capsules, ion-exchange materials, etc.), a genetic material must work rather well to be any use at all – otherwise it will quickly let slip any information that it has managed to accumulate. *Nineteenth*, what is required here is not some wild one-off freak of an event: it is not true to say 'it only had to happen once'. A whole set-up had to be maintained for perhaps millions of years: a reliable means of production of activated nucleotides at the least.

Now you may say that there are alternative ways of building up nucleotides, and perhaps there was some geochemical way on the early Earth. But what we know of the experimental difficulties in nucleotide synthesis speaks strongly against any such supposition. However it is to be put together, a nucleotide is too complex and metastable a molecule for there to be any reason to expect an easy synthesis. You might want to argue about the nineteen problems that I chose: and I agree that there is a certain arbitrariness in the sequence of operations chosen. But if in the compounding of improbabilities nineteen is wrong as a number that would be mainly because it is much too small a number. If you were to consider in more detail a process such as the purification of an intermediate you would find many subsidiary operations – washings, pH changes and so on. (Remember Merrifield's machine: for one overall reaction, making one peptide bond, there were about 90 distinct operations required.)



A ⟶ B ⟶ C ⟶ D

Figure 1.12. According to Horowitz (1945), a metabolic pathway would have evolved backwards. D was at first a vital molecule available in the environment. D gradually ran out, giving organisms time to evolve an internal source – by converting C, some simpler precursor, that was still in the environment. As C ran out there would then be selection pressures to find some other environmental molecule, B, and the means to convert it to C. Hence complex molecules that were originally provided by a primordial soup came to be made instead from simple commonly available molecules such as $CO_2$ and $N_2$.

### Problems for primitive heterotrophs

Let us suppose that all the difficulties that we have been discussing were somehow overcome, and let us now consider how the very first organisms might have fared. According to the doctrine of chemical evolution these organisms were heterotrophs, that is to say they depended on organic foods. The diet of primordial soup was so adequate, it is said, that these organisms had no need for metabolic pathways to begin with. Such pathways could evolve gradually as the foods ran out (by the mechanism proposed by Horowitz in 1945; see figure 1.12).

To have one's food provided sounds like an easy sort of life, but in reality there would be great difficulties with such an idea. There are problems of assimilation. To be a heterotroph implies an ability to recognise molecules, or at the very least to distinguish between classes of them. For the eventual evolution of metabolic pathways, specific recognition devices would be required. Thinking along the lines of current means of biomolecular control, some kind of structure would seem to be needed that could form specific sockets corresponding to the molecules in the environment. But until you have the ability to recognise at least some molecular units, how do you reach the point of being able to manufacture such specific devices? Organisms now can presuppose protein-synthesising machinery. And a great variety of transport proteins located in the cell

membranes can actively and selectively pull in particular molecules from the environment (Lin, 1970; Wilson, 1978; Rosen, 1978).

The trouble is that a socket (such as that in an enzyme or a transport protein) that can recognise another molecule is much more difficult to engineer than the molecule itself. Organic chemists are only just coming round to this means of molecular control, and the structures being made are very complex (Cram & Cram, 1978). So what were the control techniques? How was tarry chaos avoided?

If the enzymes in today's cells can cope so well this is partly because the molecules that they come across belong to a quite limited set. An enzyme may distinguish between D-glucose and D-fructose, because these are among the relatively few kinds of molecules that it encounters: but it can easily be confused by molecules from a larger range. *E. coli*'s enzymes work because, among other things, *E. coli*'s protein-loaded cell membrane is highly selective about the molecules it lets in. A primitive organism, lacking such customs control and living in a tarry 'broth' that contained for every 'correct' molecule a myriad of similar 'incorrect' ones would have to have far more accurate enzymes to bring about any particular sequence of reactions.

So that is the problem: how to evolve accurate recognising structures from a molecular technology that probably could not tell glycine from alanine, let alone D from L. Until you know one molecule from another how do you start to do the kind of sophisticated chemistry needed to make the membranes, the active centres and so on, on which molecular discrimination depends?

Another problem is heredity. In modern organisms, at least, the holding and printing of genetic information depends on an ability to do some very sophisticated chemistry – beyond the abilities of organic chemists, never mind primordial thunderstorms. Perhaps there are easier ways than ours of performing this most crucial of all functions; but attempts to replicate nucleic acids without enzymes have been unsuccessful and, as we saw, nucleotides are not among the expected components of a prevital soup. As mentioned earlier, one of the ways of attempting to overcome this last difficulty is by postulating a preliminary kind of semi-Darwinian evolution, where microsystems are not able to reproduce but they are already subject to selection. Hence organic chemical expertise could have evolved, it is said, to the point at which the synthesis of nucleic acids becomes feasible. In the next chapter we will take up again this very important question of whether a useful kind of evolution is possible without some means of storing and replicating information. I do not think it is.

# 2

# Three doubts

In this chapter we will move from the practical problems about the origin of life that mainly concerned us in Chapter 1 to more philosophical ones. We will be concerned here with three fundamental doubts about the doctrine of chemical evolution: doubts about the significance of Miller's experiment; doubts about the relevance of a probiotic evolution; and doubts about the whole idea that there have been any 'molecules of life' invariant through evolution from the start.

### *Puzzles or anomalies?*

Few would deny that there are difficulties in the doctrine of chemical evolution. The question at issue is whether these are to be taken as puzzles or anomalies in the sense defined by Thomas Kuhn (1970) in his well known account of how changes of view can take place in science. If the difficulties are puzzles then they are soluble within the framework of chemical evolution; the cracks can safely be ignored in the meantime (although they should not be papered over); something will turn up to account for the difficulties. But if these difficulties are anomalies they are to be seen rather as signs that the doctrine of chemical evolution is an inadequate frame within which to discuss the origin and early evolution of life; the cracks are structural and further building work will make them worse. So are the difficulties that we have been coming across puzzles or anomalies?

So far we have been discussing mainly acknowledged difficulties and treating them, as they usually are treated, as puzzles. Some seem to be very hard puzzles indeed – the origin of chiral discrimination, for example, or the problems of prevital work-up procedures: but never is it inconceivable

that a solution could be found tomorrow that will fit the established doctrine. Indeed, if you believe in the established doctrine, if you hold to the idea that the difficulties are puzzles, then they are positively to be welcomed as fining down the possibilities. For example, the difficulties about having to have sufficiently concentrated solutions of HCN and of preventing HCN from hydrolysing to formate can both be seen as indicators that the Earth was a cold place when life originated. What is seen as yet another *ad hoc* assumption by the sceptic is seen rather as clarification of details by the believer.

If you hold to the doctrine of chemical evolution, how do you explain the gigantic implausibility of prevital nucleic acids? There are several ways. In line with, say, Oparin or Bernal you may say that the question is irrelevant; that chemical systems can evolve, in the sense of becoming more competent at doing organic chemistry, before evolving in the full Darwinian sense – that to begin with hereditary machinery was not needed. Or you may say that hereditary processes can be efficient enough to begin with without replicating molecules. I will be trying to show later in this chapter, and in Chapter 3, that neither of these statements will do: that a genetic view of the origin of life is the only tenable one. But even so, you might accept a genetic view but deny the implausibility. Prevital nucleic acid only seems to be implausible you may say, but the very fact that it is so puzzling on present knowledge will make the explanation, when it turns up, the more convincing. If we could find one or a few unexpected effects that can account for all nineteen of those problem areas that we considered towards the end of Chapter 1, then we would not only have confirmed the doctrine of chemical evolution in a convincing way, we would know in some detail how life on Earth must have originated.

So there is nothing irrational in not giving in to difficulties. But as the difficulties accumulate the stakes get higher: success would be all the more resounding, but it becomes less likely. Sooner or later it becomes wiser to put your money elsewhere. The puzzles may be, after all, anomalies.

A pattern of success and failure has been emerging through the experimental work on the origin of life that started in the early nineteen fifties. The small molecules of our biochemistry have proved to be chemically much more accessible than one might have hoped, with amino acids and purines as perhaps the best examples. That fits the doctrine of chemical evolution: we can say that the doctrine is confirmed (in the weak scientific sense) by these findings. But, equally consistently, other things that are needed for the doctrine to work out have not been so readily forthcoming. Problems of concentrations are now well recognised: they are a recurring



theme. Then there are the much less well recognised problems of purification. And as one goes on to consider the larger molecules, such as lipids and nucleotides, chemical accessibility dissolves. These molecules are not at all easy to make. Part of the trouble is how to get up energy gradients; but that is not the only trouble. Because these molecules are bigger, and hence have richer possibilities for isomerism, the problems of discrimination – of which chiral discrimination is only one – become so difficult that there almost seems to be a conspiracy not to discuss them, but to design experiments as if they had been solved, by using purified starting materials. And then there are the biopolymers . . . In all there is an impression of a field that started well and then fell away.

*Some analogies* One of the troubles with starting well is that you may lock onto a point of view and then stay with it long after it has failed to maintain its initial promise. (This kind of thing is well described by catastrophe theory: cf. Zeeman, 1976.) There are many examples from the history of science. The phlogiston theory was one. This was the idea that an inflammable material contained a 'principle of fire' – phlogiston – which was given off when the material was burnt. This explains the obvious facts of combustion so well that it was quite natural to accept it and then interpret new data in terms of it. Suppose that as a respectable eighteenth century phlogistonist, you had been asked to explain why a candle goes out when a jar is put over it. Would you have taken this as a puzzle or an anomaly? Neither, you would have taken it as a confirmation of the phlogiston idea because it is so easily explained: the air in the jar has become saturated with phlogiston and cannot take up any more. What about the gain in weight when a metal is converted to a calx (i.e. oxide)? Here you would have put on a more patient expression. This is not so much a puzzle, you would explain, as a misunderstanding: phlogiston is the principle of fire, you see, not an isolable substance; metals all contain this principle (which is why incidentally they have so many properties in common). When the corpuscles of metals are imbued with phlogiston the effect of gravity is lessened . . . You would, I daresay, have been able to ward off all comers with such explanations. Nothing would force you to change sides. (Nothing ever forced Priestly.)

Then again the alchemists found plenty to confirm the idea that the synthesis of gold was a feasible project. The problem was seen as that of combining properties – heaviness, shininess, yellowness and so on – in one material. It was found that certain solutions (of polysulphides) could tinge metals yellow (for the recipe see Liecester, 1956). The more perceptive of

the alchemists knew, I daresay, that they had not yet made gold. But on one point all were surely agreed: this was a step in the right direction.

You can see the dangers of supposing that an entire doctrine is confirmed when experiments that test only a part of it come up with expected results. Much depends on whether there are other explanations for the positive results. What might be the signs of some other, yet unthought-of explanation? One sign might be where, in spite of tests of many predictions in different areas of the doctrine, success is restricted to only one or a few areas. As we have seen, this is the case for chemical evolution.

Another symptom might be where success (in some limited area) is too easy. Here is a story that caricatures what I mean:

Lady G has lost the diamond from her ring. She summons the Brownies to look for it. Late in the afternoon, sitting at the French windows, Lady G sees in the distance a small brown figure running towards her. Through her opera glasses she notes a look of delight on the Brownie's face and the glint of a small object held by excited fingers. Being of a scientific turn of mind, Lady G frames an hypothesis: 'the diamond has been found'. Then she notices another Brownie appearing from a similar direction also clutching a shining object. Then several more, some with several shining objects in their hands. Is Lady G's conjecture further confirmed? Why not? Surely the more glinting objects there are in view the better is the chance that one will be the diamond? But long before the first Brownie reaches Lady G she has abandoned her hypothesis in favour of 'There Must Be Some Other Explanation'. And of course there was: a car with a shattered windscreen had spread pseudo-diamonds on the road outside . . .

I think that the history of the prevital simulation experiments has something of that story in them. In the limited zone of the synthesis of small biochemical molecules the experiments have been embarrassingly successful – with 'blanks' often coming out positive, that is, biochemicals being favoured by conditions that would *not* have been present on the primitive Earth. Even among the plausibly prevital syntheses we are left with too many possibilities for their significance to be very clear. Was it from the sparks of lightning or the shocks of thunder or the glare of ultraviolet light or the infall of meteorites or the baking of volcanoes; or was it through a sweeping up of cosmic dust that we acquired the first biochemicals? Or was it a bit of each? With so many possibilities to choose from one might be as well not to settle on any. There is the thought that other possibilities will turn up; that the true explanation is perhaps still to be found.

### First doubt; the significance of Miller's experiment

It is really too naive simply to assert that the prevital simulation experiments confirm the doctrine of chemical evolution. Well, yes they do



(in the weak scientific sense) – but with crippling provisos. The whole subject of confirmation in science is notoriously tricky, and here we have an example, I think, of a hypothesis that has become *less* plausible the more it has been 'confirmed' (cf. Gardner, 1976).

It is those 'Other Explanations' that cause the trouble. Proposition A may be confirmed by observation L in the scientific sense, that is L is consistent with A. But suppose that observation L is also consistent with proposition B? Then L does not confirm A as against B, it only confirms A *or* B as against (some) other possibilities. Any observation inevitably confirms a whole set of propositions A, B, C . . . and only some of these you will have thought of. The normal way round this difficulty is to test the proposition in as many diverse ways as possible. Observation M, let us say, also confirms proposition A (including other unthought-of possibilities) but not proposition B. Then you have confirmed A as against B. But you have not eliminated the possibility of some unthought-of proposition that is common to the sets that are confirmed by both observations L and M. So you carry on with observations N, O, P, and so on – all as diverse as possible. If proposition A continues to hold out against these diverse tests, then it is reasonable to hope that there is not some unthought-of, quite different explanation that would also fit all these observations.

But, as we have seen, the really successful tests of chemical evolution have been restricted to one aspect – chemical accessibility of a number of the smallest molecular units. These molecules can be said to be chemically accessible because they are so often present in complex mixtures formed by uncontrolled reactions. The other aspects are mainly a catalogue of difficulties that have been getting worse: the early environment now looks less clement than was once thought; hopeful possibilities for resolution of enantiomers have not worked out; purification, concentration, activation, when any kind of solution has been proposed, all call for rather special situations, and they are anyway inefficient. And the significance of those seemingly easy routes to biochemicals from cyanide and formaldehyde becomes less clear when examined in more detail. As we saw, neither cyanide nor formaldehyde would easily have built up to sufficient concentrations. Yields in these uncontrolled syntheses are very low, in general, and the products always very complex. Many of these products – especially sugars – would not have survived long enough to have accumulated in a 'primordial soup'.

Is there, then, some Other Explanation? Is there an hypothesis that will account for the (rather general) accessibility of (some) of our biochemicals, and yet accommodate the difficulties?

*Biochemical economy*

**Biochemical economy** is a label that I will attach to the idea that on the whole you do not expect organisms to use molecules that are much more elaborate or unstable than they have to be. Whatever else, it is a likely *outcome* of evolution that a biochemistry be based on chemically accessible molecules.

Other things being equal, it would be an advantage to an organism to be able to use molecules that were relatively easy to make – because then there would be fewer catalysts needed, less genetic information, fewer metabolic products to interfere with each other, and so on. During the early evolution of our biochemistry there would be a continual pressure to be more efficient in this sense. Provided early evolutionary processes were flexible enough to allow experimentation with different subsystems, then the final outcome would predictably contain numerous fairly easily made molecules using metabolic pathways of least resistance.

One can make the expectation a little more precise. You might expect three rough classes of small molecules to have emerged as components of an evolved biochemistry – let us say classes A, B and C. The A-class are accessible molecules such as formate, glycine, alanine or adenine. They are accessible because thermodynamically favoured: they are simple and rather unreactive molecules. For that reason they naturally tend to emerge as products of abiotic processes of various sorts in which atoms are being more or less violently shuffled about. (Once you have made glycine, for example, it tends to persist because it is a zwitterion – page 50.) But one can easily see advantages for a biochemical system in having some molecules that are not only easily made but fairly stable. Such units would be less liable to misreact. A-Class molecules are simple basic nuts and bolts.

The B-class would again be accessible types but this time only because kinetically favoured. They would be easily made from available starting materials but would not be particularly stable. There are plenty of such comparatively short-lived molecules in our biochemistry. Sugars are evidently in this category. That sugars appear temporarily in the condensation of formaldehyde may have some relevance to how they were first biosynthesised: but you do not find sugars in mature products of non-biological semi-chaotic reactions – they are not in meteorites for example. B-class molecules are simple active metabolites.

The C-class biochemicals are not easy to make. They are common types in our biochemistry nevertheless because they perform critical functions that make their manufacture worthwhile. Nucleotides, coenzymes, chlorophyll and lipids are in this category. So are many of the amino acids – especially the aromatic and the basic amino acids. Higher animals can get the best of both worlds, avoiding the trouble of making some of their C-class molecules by depending instead on dietary intake – vitamins and essential amino acids are very often C-class biochemicals. (Dietary requirements vary of course, but for the white rat the essential amino acids are phenylalanine, tryptophan, lysine, histidine, arginine, threonine, methionine, leucine, isoleucine and valine; Mahler & Cordes, 1971.) If the A-class molecules are like nuts and bolts, those in the C-class are like more specialised, expensive pieces of a machine. C-class molecules are specially made for particular purposes.



Now only the first of these rough categories, only the A-class, might accumulate through probiotic processes to allow the possibility of a probiotic explanation for present biochemicals. If our biochemical system was really put together from components of some oceanic or localised soup then you would expect some minimum version of our system to be makeable entirely from A-class molecules. Yet that seems very far from the case – since sugars, lipids and especially nucleotides are absent from this supposedly primordial class.

*Why are nucleotides difficult to make?* That nucleotides are very difficult to make whereas several of the amino acids are quite easy can be understood in terms of biochemical economy, along the following lines.

Our whole biochemical system, including the choice of the molecular units out of which it is now made, is an outcome of fully Darwinian evolutionary processes. Always in making these choices there were questions of cost effectiveness; of taking the most easily made molecule that could do the job. Several easily made amino acids could be chosen for the protein set because it did not matter very much which were chosen – so long as there was some variety of amino acid types. If you doubt this just think of the variety of functions that protein can perform and how many very effective functions of protein, for example making feathers or antibodies, were only discovered long after the twenty amino acid set was decided on. The cleverness of protein lies only somewhat in the choice of the twenty amino acids. Most of the cleverness is in the stitching of these units into specific sequences. Functions in proteins are predominantly sequence-dependent rather than unit-dependent.

By contrast, the process of the replication of a nucleic acid molecule is independent of the sequence of monomers in that nucleic acid molecule. The whole point is that the sequence is replicated whatever it is. Hence the

cleverness has to be far more in the units themselves. Units for replicating organic molecules have far more design restraints on them: it would be lucky if something that happened to be easily made would do. (I will be going into this in greater detail in Chapter 5.)

*'Probiotic pathways'* In any case biochemical economy seems the most plausible general explanation for the tendency for reaction pathways to be similar within and without organisms. For example, Degani & Halmann (1967) found that the spontaneous decomposition of glucose takes place along routes that are similar to the glycolytic pathway. We might say, then, that if there had been a primordial soup and if sugars had formed in it, something like the glycolytic pathway might have come before organisms. This would be a misleading way of talking however. It might seem to imply that the glycolytic pathway had to be *rehearsed* in a probiotic soup before it could be taken up by organisms. To talk about a 'probiotic pathway' at all is misleading; it biases the imagination in favour of an historical connection which may not have been there.

Is it not more likely that early evolving systems simply discovered a bit of chemistry: not something out there in an actual soup, but something inherent in the way molecules behave? If the organic chemistry of meteorites; or of Fischer–Tropsch reactions; or of spark plasmas or shock waves in gases; or of reactions of formaldehyde in base; or of cyanide solution (and so on) bear some resemblances sometimes to present-day biochemistry (and to each other) that is not to say that there was ever any causal connection in the sense that any of these processes caused or copied any of the others. The connection, I think, is universal and parallel, not historical and serial. All these processes are confined by rules of chemistry.

*Biochemical economy is sufficient* According to the view to be elaborated in this book, biochemical economy provides virtually the whole explanation for any overlap there may be between the set of easily made molecules and the set of our biochemicals. We will have to take a more complex view then of very early evolution: we will have to find other approaches to the nature of the very first organisms; and we will have to explain the situation within which an evolving biochemistry would have been able to make choices about certain molecules when now there can be no more choice. We will see that these are not insuperable paradoxes, and that furthermore biochemical economy allows us to dispense with a primordial soup. But one advantage we can already see: biochemical economy



explains most easily why, although some biochemicals are stable and easily made, other very central ones are either not very stable or not at all easily made.

### Second doubt; the relevance of a probiotic evolution

*The two kinds of evolution*

The Universe evolves and with it the stars, the solar system, the Earth, and environments on the Earth. That is one kind of evolution. The other kind is biological evolution (what I mean in this book by the unqualified term 'evolution').

Environmental evolution was under way, presumably, before life started to evolve on Earth. So there is a serial connection in that the evolution of the Earth created the conditions for the evolution of life. But this is not to say that the two processes are the same or even similar, except that each is a kind of change. And it is hard to see *why* one kind of evolution should have led to the other.

Was 'chemical evolution' the connection? I do not think so. The building up of a primordial soup, if such a thing ever happened, would have been part of environmental evolution. The oceans would have accumulated organic molecules in much the same way as any other geochemical process would have taken place. Unless you take a religious or mystical view there was no guiding hand to contrive an outcome suitable for the origin of life. Mountains were made and worn down, the wind blew, the sun shone – and a soup did or did not form: all such processes were on an equal footing; it would only have been with an eye to the future that some of these processes might have been given a special label and called 'chemical evolution'.

Biological evolution, on the other hand, *is* special, as discussed in the opening pages of this book. Above all what makes it special is heredity. This is the great divide: either there is a long-term hereditary mechanism working or there is not. If there is not then there is no accumulation of 'know-how' as Kuhn (1976) put it: the survival or non-survival of some putative half-organism will not be 'remembered' in the distant future to have any effect. Things would change, systems such as coacervates would come and go, but you could not expect them to become more *efficient*: you would not expect them to become more efficient at organic chemical operations, for example. Only evolving organisms can progress in that sort of way.

Suppose that by chance some particular coacervate droplet in a pri-

mordial ocean happened to have a set of catalysts, etc. that could convert carbon dioxide into D-glucose. Would this have been a major step forward towards life? Probably not. Sooner or later the droplet would have sunk to the bottom of the ocean and never have been heard of again. It would not have mattered how ingenious or life-like some early system was; if it lacked the ability to pass on to offspring the secret of its success then it might as well never have existed.

So I do not see life as emerging as a matter of course from the general evolution of the cosmos, via chemical evolution, in one grand gradual process of complexification. Instead, following Muller (1929) and others, I would take a genetic view and see the origin of life as hinging on a rather precise technical puzzle. What would have been the easiest way that hereditary machinery could have formed on the primitive Earth?

### *Model for a biological evolution*

*Are genes necessary?* What is the minimum sort of hereditary system? Can we say what kinds of structures would be necessary? The reproduction of organisms now on the Earth depends on the replicability of molecules – DNA molecules: is it a necessary minimum requirement for a system to be subject to natural selection that it should contain replicable molecules of some sort?

Consider an ocean that is coacervating spontaneously, giving rise to droplets of a certain general composition. We are interested in one rather odd droplet that happens to contain $n$ particles of a catalyst X that increases the rate of growth of the droplet. In all other respects our droplet is typical. The trouble is that even if $n$ is $10^{18}$ – an unreasonably large number – the particles would have been diluted out after about 60 generations. Clearly to continue to have an effect indefinitely the number of catalyst particles in a droplet must, on average, double between droplet divisions. Anything less than that and the catalyst will sooner or later be 'forgotten'.

There are two ways in which the catalyst particles could increase in numbers between droplet divisions:

(*1*) *by assimilation* – they could be picked up from the environment – or

(*2*) *by synthesis* – they could be put together from other units (d, e, etc.) themselves ultimately acquired from the environment. In either case there is an additional requirement that X must exert some more or less specific effect in bringing about the increase in its numbers. Otherwise it might simply be that a droplet would contain X, not because its parent did, but because the environment was such that droplets coacervating from it contained X.



In the case of assimilation there must be some kind of X-receptor in the droplets: not necessarily a specific socket, but some effect of X must predispose the droplet to acquire more X. In the case of synthesis, X must either predispose the droplet to acquire suitable units (say d and e) or catalyse the synthesis of X from units that are there anyway, or both. This gives us seven cases to consider.

A1 (direct X-assimilation)
X is itself an X-receptor, or part of an X-receptor.

A2 (indirect X-assimilation)
X helps to make some other X-receptor: that is, X is a control structure (for example a catalyst) in the making of the X-receptor.

S1 (direct d, e-assimilation)
X is a receptor for components d and e and hence contributes to the making of more X, because elsewhere in the droplets d and e are converted into X (without the need for further control).

S2 (indirect d, e-assimilation)
X helps to make receptors for d and e by acting as a control structure in their synthesis. Hence indirectly it contrives to pull d and e from the environment: again these form spontaneously into more X without the need for further control.

S3 (X-synthesis)
The droplet is a receptor for d and e in any case (that is, to be a d, e-receptor is a general property of coacervates forming under the given conditions). X converts d and e into more X.

S4 (direct d, e-assimilation + X-synthesis)
As S3 but X is also a d, e-receptor.

S5 (indirect d, e-assimilation + X-synthesis)
X helps to make d, e-receptors and also converts d and e into more X.

At least formally, then, hereditary processes can be imagined that do not explicitly involve replicating molecules: only in the schemes S3, S4 and S5 do the growth- and/or fission-promoting structures X also control their own synthesis – and even here this control might not involve template-directed synthesis. The acid hydrolysis of an ester is a reaction in which protons can be said to be 'self-reproducing' species. This reaction is both catalysed by hydrogen ions and it produces them. Such autocatalyses are indeed well known in chemistry and no templating processes need be in-

volved; no process, that is, in which the form of a parent structure is imposed on a daughter structure. In the schemes S3, S4 and S5, X is simply an autocatalyst. In other cases X makes its reappearance more likely in some other way, by picking up X from the surroundings, by helping to collect the bits to make more X, and so on. In general and in principle what X has to do is to look after its own reappearance in the next generations of coacervate droplets. It must reproduce in the ordinary meaning of that word – it must be produced again – but there is seemingly nothing to insist that that reproduction should involve template copying (what I will from now on call **replication**).

Evolution through natural selection, though, requires more than just the inheritance of characteristics. It requires an indefinite future potential for those characteristics to be modified. It must be possible for X to change to X′, a modification of X, and then to X″, and so on: and then these modifications must be inheritable.

Horowitz (1959) stressed the insufficiency of simple autocatalytic processes as a basis for heredity. The key point is that autocatalytic processes do not in general have the property of reproducing modifications. If you feed protons to an aqueous solution of an ester you will get back more protons, but if you feed deuterons you will not get deuterons reproduced instead. Simple autocatalytic processes are not only limited in themselves, but they have no future, they are isolated. The species that is being reproduced does not belong to a set of similar species that could alternatively be reproduced. In simple autocatalysis only the information corresponding to the presence or absence of a particular species is reproduced – nothing about its detailed form.

Now one might imagine the evolution of a system taking place through the setting up of several autocatalytic processes. The catalyst set X, let us say, consists initially of an autocatalyst m: X then evolves, through X′, to X″ which consists of three autocatalysts l, m and n. The trouble here is that each reproducing molecular species only contributes a limited amount of genetic information (about one bit – the species is either there or it is not). Also it is very hard to imagine in practice several autocatalytic processes going on together without interfering with each other. (Here is a challenge to a chemist's ingenuity: a vessel contains a mixture such that if A is added, more A forms in the vessel; if B is added, more B, and if C, more C. It might just be possible to construct such a three-bit genome using non-template autocatalytic processes, but I think that if you were to try to design a multi-bit system you would find that the difficulties rapidly got out of hand.)



There would be a similarly low ceiling for the evolution of systems using hereditary mechanisms A1 or A2. Here what is inherited is the ability to pick up X from the environment rather than the ability to make X: X is 'autoacquisitive' rather than autocatalytic. Even if you could imagine a mechanism for this, such an evolution would be limited by the number of (non-interfering) 'autoacquisitive' catalysts in the environment. The whole idea seems absurd.

There is nothing wrong with the idea that the early environment might have provided catalysts for early organisms. The environment is still a source of metal ions, such as $Zn^{2+}$, that are important parts of enzymes. The trouble arises when one tries to imagine the environment as the source of the *specificity* of catalysts, or indeed of any other kind of specificity. If it is the environment that entirely controls the sequences of reactions in some microsystem – because, say, the environment contains catalysts that the microsystem passively acquires – then the microsystem is simply part of the environment. To evolve within the environment the microsystems must acquire specific characters of their own so that a million years later the microsystems could be different even if the environment had remained similar. That is not to say that some aspects of the microsystems' behaviour cannot be controlled by the environment – by catalysts and other things – what is important if a microsystem could be said to have evolved is that at least some aspects of its behaviour are not so controlled.

### *Expect replicating structures*

Let us summarise our conclusions about what is necessary for an extended evolution of microsystems, and then take the argument a little further.

Simple reproduction is no good: for example, coacervates growing and dividing would not be evolving if that is all they were doing; in that case their characteristics would be determined by the environment. (If you came back to look in a million years you would find no difference if the environment was no different.) Particular structures that happened to be possessed by some droplets – for example catalytic particles – might be passed on when the droplets divided and might provide some individuality for different lines of droplets, but these effects would soon be diluted out. Not the goods but their means of production must be inheritable. (Their means of acquisition was another idea but this turned out to be too limited.) And the means of production has to be in the form of specificity or information. Even this, though, is not enough. For evolution, the information must be

modifiable and the modifications must also be inheritable. This eliminates the kind of simple autocatalysis of which ester hydrolysis is a model. Indeed it implies copying of some sort.

Taking a more complex example, suppose that X catalyses the formation of Y and that likewise Y catalyses the formation of X. That would be a reproducing system with the structures of X and Y not necessarily related. But to be evolvable X′ must catalyse the formation of Y′ *and Y′ must catalyse the formation of X′*. It is this last consideration that demands some structural correspondence between the Xs and the Ys: X and Y could be + and − strands of a DNA molecule for example. It is not necessary that X, X′, X″, etc. are structurally identical to Y, Y′, Y″, etc. or even at all similar in composition, but there must be a detailed correspondence between them – and for an extended evolution there would have to be some vast set of such corresponding member states. Templating is still not forced on us as the means of ensuring such a correspondence. For example, a text can be reproduced by being translated first into a temporal sequence of blips, sent along a telephone wire and reconstructed a thousand miles away. With sufficiently sophisticated equipment, translation and back-translation could be effected by other means than the direct testing out of units for size and shape as happens, say, in DNA replication. All we can say is that, for a molecular system, templating looks the easiest way.

Our rejection of a set of independent autocatalytic processes was similarly based on chemical plausibility: in principle genetic information could be carried in the form of a check list of presences and absences, but this would be impracticable, and in any case inefficient from the point of view of information capacity. Suppose that there were four possible items on the check list. This would give, in all, sixteen possible lists (none; A; B; C; D; AB; AC; . . . ABCD) which is to say that the system could have an information capacity of four bits. By contrast, if it is not merely the existence or absence of items that constitutes the information but their arrangement – for example the arrangement of the four bases in DNA – the information capacity is limited only by the number (not by the number of kinds) of the units. The chemistry can thus remain simple while the information capacity tends to infinity.

Although other hereditary systems are possible in principle, it would seem that in practice any extended evolution of microsystems would depend on the copying of stable arrangements in some sort of genetic material. Such a genetic view of life is now widely held; but if you add to it the conclusion from Chapter 1 that prevital nucleic acid replication is



wholly implausible, then you arrive at a much less popular view: *there must have been at least one other kind of genetic material before nucleic acid.*

### Third doubt; original biochemical similarity

Perhaps the main barrier to accepting the above conclusion is the idea that the central biochemicals have always been more or less a fixed set. It is as if the origin of life was being seen as a kind of chess problem to be solved, if at all, by using only the pieces now visible on the board. I will conclude this chapter with comments on some of the rather superficial arguments that are sometimes put up in favour of original biochemical similarity.

*Perhaps life can only work in one way* Pirie (1957, 1959) has pointed out the emptiness of the assertion that life could only work in one way. This assertion becomes less plausible the more we know about the details of how our kind of life works. There are surely other catalysts than proteins, other membrane materials than lipids, other genetic materials than nucleic acid. Life is primarily about systems of a certain kind, not substances as such.

*Present biochemical similarity* The close biochemical similarities between all organisms now on Earth is not a good argument for the idea that organisms have always been biochemically similar to present-day organisms. When one looks more closely at the common features of all life now on Earth one sees machinery that could not have been picked out of a soup; that is evidently the product of prolonged evolution. All life now is evidently descended from a common ancestor that was high up the evolutionary tree – a considerable distance from the start of the evolutionary process that gave rise to that common ancestor. The crucial question is what happened over that distance.

*Miller's experiment* Then again, as discussed at length in this chapter, the simulation experiments, although promising to begin with, have turned out to be ambiguous: if anything they are better understood in terms of the idea that our present set of central biochemicals was a product of evolution rather than a precondition.

*Surely to make a machine you first need the components*? Then again it is a naive view of evolution, even if it sounds like logic, that to make a machine you have to start with the components. That is the way we make machines, but evolution does not work like this. I will devote much of the next chapter to the question of how natural selection brings machinery into existence; how subsystems come and go in the trying out of designs and combinations of them; how, eventually, mechanisms become frozen in; how one subsystem may become universal – not because it was the first but because it became the best. And we will think in detail about just why the first is unlikely to be at all similar in design to what eventually turns out to be the best.

*Should we not choose the simplest hypothesis?* Others would appeal to Occam's Razor in defence of the principle of original biochemical similarity. They would say that this is the simplest hypothesis and should be adopted in the absence of clear evidence one way or the other. Simplicity has its place in science: where there is nothing else to be said between two or more views – then choose the simplest by all means. But there is very much to be said about early evolution, and there is no alternative but to think through the arguments in detail. In the circumstances to hold to the simplest view may show less a leaning to elegance than to laziness. There is no indication that the very early evolution that made our biochemistry was a simple process. Rather the reverse, if later evolution is anything to go by.

In any case the place for Occam's Razor is in natural philosophy, not in natural history. Of historical questions we may ask simply 'is it true?' (is it true, say, that Harold got an arrow in his eye at the battle of Hastings?). Maybe we do not know, but the question is not one of formal elegance, it is one of fact: either it happened or it did not. Similarly, questions of early evolution are questions of fact, however remote these facts may seem. This is not at all like, say, physics where the choice between two ways of describing known facts may rest on which lets you write the simpler equations.

Where simplicity is most to be sought in biology is in the construction of the very first systems capable of indefinite evolution. They must have been simple, in the sense of uncontrived, if chance and chemistry put them together. It is surely at least as plausible to suggest that the first organisms were simple but dissimilar to us and that our biochemistry was the outcome of a subsequent (complicated) evolution, as to suppose that the first organisms were based on our complicated way of doing things, followed by a subsequent (simple) evolution.



*Would not the genetic material be (virtually) invariant?* Then again it is sometimes stated more or less explicitly that (even if the rest of evolution was devious) there could be no place for radical change at the very centre – that the genetic material itself must be invariant or nearly so. I will try to show in the following two chapters that, on the contrary, you would expect there to have been at least one radical switch in genetic materials during early evolution.

# 3

# Questions of evolution

> I define life . . . as a whole that is pre-supposed by all its parts.
> *S. T. Coleridge* (ca *1820*)
>
> . . . organisms – that is to say, systems whose parts co-operate.
> *J. B. S. Haldane* (*1929*)
>
> We shall regard as alive any population of entities which has the properties of multiplication, heredity and variation.
> *J. Maynard Smith* (*1975*)
>
> I suggest that these three properties – mutability, self-duplication and heterocatalysis – comprise a necessary and sufficient definition of living matter.
> *N. H. Horowitz* (*1959*)
>
> . . . it might be claimed that the most important fact about them [living things] is that they take part in the long term processes of evolution.
> *C. H. Waddington* (*1968*)

Definitions of life commonly fall into either of two classes. First, and most immediately understandable, there are definitions like those given by Coleridge and Haldane. Here organisms are to be distinguished from other physicochemical systems in being seemingly purpose-built. Organisms are to be seen as machines analogous in some detail to man-made machines which are made up of components and subcomponents acting together. According to modern science the machinery of life was not put together with forethought by an intelligent being. The only engineer was natural selection. Definitions of this sort are **teleonomic**, that is, they use the idea of apparent purpose.

The other kind are the **genetic** definitions of life which concentrate not on products but on prerequisites for evolution. The definitions according to Maynard Smith and Horowitz are of this kind. We discussed in the last chapter why replicating structures of some sort – genes in the most general



sense – would seem to be necessary for heredity and hence for evolution through natural selection. (The entities referred to by both Maynard Smith and Horowitz, for example, could be DNA molecules.)

Waddington's description neatly connects both kinds of definition of life and avoids some of the quibbles that can be put up against each. Even organisms that do not reproduce are part of the long-term processes of evolution – they are an outcome of these processes – although they no longer directly further them. Then again the very first organisms on Earth must fail to conform to the teleonomic definition if you bar miracles: these were not overtly machines in the sense of being highly contrived and seemingly improbable in their mode of survival and propagation. Yet they too can be claimed to have been part of life – in retrospect at least – as having been at the other end of those same long-term processes of evolution. Life in the teleonomic sense could only have emerged gradually during the very early evolution of systems that from the start conformed to the genetic definition.

As implied in the Preview, I am using the teleonomic definition of 'life' but the genetic definition of 'organism'. Hence I will never talk about unevolved life – which would be a contradiction – but I will have much to say about putative unevolved organisms. Indeed to search for the origin of life is to search, first, for such systems.

Much of the difficulty in the problem of the origin of life and much of the confusion in the definition of life has arisen, I think, from a confusion between prerequisites for evolution and products of evolution. Part of the trouble is that products have become prerequisites – that is, products of early stages of evolution have become prerequisites for evolution now.

In this chapter we will be concerned very largely with how evolution ties such knots and in doing so conceals the means by which it creates that most characteristic property of life – co-operation between components. But to begin at the beginning: what sort of system is it that we are looking for, that is not an improbable collaboration of parts and yet can evolve to become one?

## *What is a phenotype for?*

An organism is said to consist of a **genotype** and a **phenotype** in an environment. We discussed in the Preview and in the last chapter how any system if it is to evolve within an environment must be able to hold and replicate information of some sort – instructions about how to make its characteristic features. It is only such instructions that can be transmitted indefinitely between generations to provide the basis for a long-term

evolution. The replicable instructions constitute the genotype of an organism, while the characteristics are the phenotype. The environment has to be included in the description because in the end it is the environment that the genetic instructions instruct. The phenotype of a modern organism can be seen as a kind of go-between: itself a bit of the environment locally modified by the genetic information.

Following Sherrington (1940) you can think of the environment for life as a stream, with organisms as eddies in the stream that are locally moving against the entropy trend. On the surface of the Earth the main motive stream is the stream of photons from the Sun. Sooner or later the energy that arrives on the Earth will be reradiated into space in a more degraded form. But between this arrival and this departure the gigantic eddies of the weather and the water cycle are driven continuously. Much more intricate are the convolutions inserted by life on Earth between the absorption of sunlight in green leaves and the reradiation of that energy, eventually, into space.

But really it is only the phenotypes of modern organisms that should be viewed in this highly dynamic way, as being a dynamic part of the environment with 'the motion of the eddy . . . drawn from the stream', as Sherrington put it. By contrast, for an individual organism, the genotype is static. We might extend Sherrington's analogy by thinking of a stone of a certain shape placed in a stream and causing a specific kind of turbulence. The (static) shape of the stone is a genotype which when placed in a particular environmental stream informs that stream to generate a localised, persistent, but dynamic turbulence – a phenotype.

There are two counter-intuitive aspects here. Using higher animals as models we would be much more inclined to see the organism as dynamic and the environment as static. But the only bit of an organism that is unambiguously not part of the environment is the bit that is static – the genotype.

The other counter-intuitive idea is that, in computer jargon, it is software in organisms that lasts, while hardware is being perpetually replaced. Consider, for example, the instructions about how to make cytochrome *c* molecules: that software has remained little altered in essentials while mountain ranges have risen and been worn away many times. Yet the hardware, the actual individual protein molecules, individual DNA molecules, and so on, have been quite evanescent, flickering in and out of existence on a geological time scale. And this is very close to the heart of the problem: following our discussions in the last chapter we might say that life can begin to appear when mechanisms exist for retaining and propagating a kind of software – genetic information – indefinitely.



Perhaps the simplest kinds of organisms would be hardly more than pieces of unencumbered information-printing machinery – 'naked genes' as they have been called (Muller, 1929). To have the potential for indefinite evolution into the future, the potential information capacity of these naked genes would have to be very high.

As discussed in Chapter 1, the idea of a 'naked gene', as the simplest and first kind of organism, has a long history. It is somewhat out of favour now mainly on account of two kinds of argument that are put up against it.

First, there is a practical argument. Even if it could evolve in principle, it is said, such a structure would be too improbable in practice: it would be exceedingly unlikely to form, and the Earth would be exceedingly unlikely to continue to provide the highly specialised components needed to keep it replicating. If we think about a naked nucleic acid molecule such an attitude seems justified.

Second, there is a formal argument. To evolve, a system must have both a genotype and a phenotype. Pure information is no use: it is the phenotype on which selection operates to give genetic information a meaning. Formally this argument is impeccable, but it is largely irrelevant. A 'naked gene' would not be – could not be – pure genotype. Clearly what is meant by a gene, in this context at least, is some sort of structure that is holding information – something analogous to a DNA molecule or a punched card. Such a thing is not pure software as it includes the structure that is holding the information, and that is hardware. And at least some aspects of hardware could very well be phenotype.

There are two ways in which this could be so. First, we may note in passing that in a DNA molecule only the base sequence (specified by a copying process) is genotype: everything else about it is phenotype because the processes that give rise to all those other aspects of the DNA molecule – processes such as making deoxyribose, collecting the phosphate, and so on – are under the control of information disseminated throughout the genome. They arise by following rather than by copying instructions. Of more immediate interest, nucleic acid molecules can often be both genotype and phenotype in the sense that a sequence of bases directly specifies some functional object. The transfer RNAs and the ribosomal RNAs are like this: here a (potentially replicable) message tape can become a machine by twisting up on itself in a way that is determined by the message.

That this duality in RNA can allow RNA molecules to evolve has been beautifully illustrated by Spiegelman and his school (Spiegelman, 1970; Spiegelman, Mills & Kramer, 1975; Mills, Kramer & Spiegelman, 1973) in their studies of RNA molecules replicating with the aid of an enzyme system *in vitro*: natural selection was found to operate so as to favour

particular RNA molecules according to their size and to how they twist. For more recent discussions see Orgel (1979) and Eigen, Gardiner, Schuster & Winkler-Oswatitsch (1981).

So the formal argument against evolving 'naked genes' is too formal. All that is necessary is that, in addition to holding and printing some sort of pattern (analogous to nucleic acid base sequences), a primitive gene could have at least some of its properties affected by the pattern that it happened to be holding. Simple properties such as shape and size might do. Suitably shaped particles might tend to adhere to each other slightly better. In a particular environment this might have survival value – for example in a stream, by reducing the chances of a mass of such particles being washed away. To my mind it is only the practical arguments that hold up, but then these arguments are against naked nucleic acid genes, not against 'naked genes' as such.

The term 'naked gene' means 'consisting only of genetic material' and not 'lacking any phenotype'. The key point is that in so far as a gene has physical and chemical properties that are affected by the information that it contains; and in so far as those properties affect the survival chances of the gene, then that gene itself has phenotypic characteristics. Phenotype and genotype are separate ideas, but they are not necessarily separate structures.

In any case we must try to disengage from the idea that a phenotype has to be something elaborate. True, all the phenotypes that we have ever come across in real organisms are highly elaborate. Yet if we ask 'what is a phenotype for?' the answer does not seem to demand that it be particularly ingenious. A phenotype should provide conditions that encourage the preservation and/or the replication and/or the propagation of genetic information. In the most general case these conditions are strictly unspecified. It does not matter in what way the phenotype helps, or even if its assistance is only marginal. So long as it is some sort of consequence of genetic information and so long as it tends to promote that information then a phenotypic characteristic can catch on.

It is because nucleic acid is so difficult to make that the phenotypic machinery serving it has to be so complicated. We might say that the necessary complexity of the go-between machinery – machinery between a genetic material and its environment – depends on how mismatched a genetic material is with its environment. The modern cell is like a spacecraft providing life support systems for the fragile DNA molecules that contain its essential information. A naked DNA gene would be mismatched on our Earth as a spaceman without his space suit would be



mismatched on Mars. As Dawkins (1976) puts it, organisms are survival machines for their genes.

One possible way of explaining the present relationship between DNA and the external environment is to say that there was a time in the remote past on the Earth when the external environment was matched to DNA and that the go-between apparatus evolved gradually as the environment became less clement. That is logical, but from our discussions in Chapter 1 it does not seem to work. It really does not look as if any purely physico-chemical environment would have had the expertise to carry through the necessary cycles of operations for a nucleic acid genetic material.

Another way of imagining a first match between a genetic material and a primitive environment is to suppose that it was the genetic material, rather than the environment, that was different to begin with. We have already concluded something of the sort (page 75). But would such a change be evolutionarily possible? Would it fit with more general ideas on the nature of evolutionary change?

### What does evolution do?

*Four aspects of organisation*

Most people would agree that evolution has created highly organised systems. Indeed we tend to equate, in a rough sort of way, the level of evolution with the level of organisation in an organism. Unfortunately the question 'how organised is this system?' is ambiguous. A system might be said to be highly organised because it is very orderly, like a crystal; or because it is complicated like a television set; or because it is arranged into distinct organs like a bureaucracy; or because the parts out of which it is made are strongly dependent on each other, as in machines of many sorts. Or you might say that efficiency is the only measure of organisation. Organisation is difficult to measure just because it is a compound idea – only its elements are measurable in suitable circumstances. But we can say, I think, that an organised system is some arrangement of units that are at least to some extent orderly, efficient, complex and made of parts that co-operate. The question is which of these aspects is the most typical product of evolution.

*Orderliness* I will try to show now that orderliness is not a particularly significant attribute of organisms.

First let us define orderliness in a way that is in line with its meaning in statistical thermodynamics. We will say in a general way that to be well

ordered is to belong to a small subset of possibilities. Suppose that you have put some coffee cups, that had been lying about in the kitchen, into the kitchen cupboard. You would have been ordering the cups because their positions in the cupboard, although not completely defined, are more closely defined than they were when the cups were just somewhere in the kitchen. There are fewer possibilities for the state 'stowed away in the cupboard' than there are for the state 'somewhere in the kitchen'. If you could count the respective numbers of possibilities then you would have a measure of the amount of tidying up that you had done.

Molecular ordering processes can often be measured in this fashion, because there is a way of estimating molecular possibilities – the number of particular ways that a set of molecules could be arranged, including the distribution of their energy between them. Such a number, $W$, is vast, even for moderately sized physicochemical systems. But $W$ for a system can be estimated because it is related to the entropy ($S$) through the Boltzmann equation, $S = k \ln W$ ($k$ is a constant, *ca* $1.4 \times 10^{-23}$ JK$^{-1}$).

Now let us try to make a rough but very conservative estimate of the physicochemical orderliness of a bacterium such as *E. coli*. Let us think only about the water in *E. coli* – about $10^{-13}$ moles. The molar entropy of water under ordinary conditions of temperature and pressure is about 70 JK$^{-1}$: so the entropy of an *E. coli*-sized blob of water is about $7 \times 10^{-12}$ JK$^{-1}$. That gives us a value for $W$ of about $10^{200\,000\,000\,000}$. For the same mass of water vapour the entropy would be about $19 \times 10^{-12}$ JK$^{-1}$: $W$ in that case would be about $10^{600\,000\,000\,000}$. The ratio $W$ (vapour)/$W$ (liquid) represents a measure of the ordering that takes place when this tiny amount of water condenses from the vapour. It represents the 'improbability' of the condensation – or rather how improbable that condensation would seem if one did not take into account intermolecular forces. It is a big 'improbability' – about $10^{400\,000\,000\,000}$.

Biological systems have another kind of 'improbability'. Even if you take into account intermolecular forces, an organism still seems to be an improbable system if you do not also take into account its evolutionary history. To measure what we might call the biological orderliness of *E. coli* we should attempt to get some estimate of the constraints placed by evolution on molecular arrangements within *E. coli*. How much of the order in *E. coli* has been generated by evolution?

It is not difficult to make a maximum estimate for this. The product of evolution, the only thing that is passed on between organisms over the long term, is genetic information. We can estimate the maximum amount of that information in *E. coli* by counting the number of base pairs in its



DNA. This number is about $4.5 \times 10^6$. With four possibilities for each base pair the total number of ways in which *E. coli*'s DNA might have been arranged is $4^{4500\,000}$, i.e. $10^{2700\,000}$. Since no doubt there are many slightly different DNA sequences that would do to make an *E. coli*, this is very much a maximum measure of the 'improbability' of *E. coli*. It is a measure of the maximum amount of ordering that evolution could have been said to have achieved in making *E. coli*. Looked at one way $10^{2700\,000}$ is a staggering number, far exceeding, say, the number of electrons in the known Universe (around $10^{80}$). But compared with the coralling of possibilities achieved when a speck of dew condenses, it seems rather a poor performance. The generation of order as such cannot be what evolution is especially good at.

*Efficiency* Components of organisms are not just tidily made. They have functions. They are machines. As such it is appropriate to ask how efficient they are.

Natural selection can be seen as a kind of optimising procedure that leads to increasingly efficient machines; and we can discuss evolution, as we can discuss optimising procedures generally, in terms of movements across a diagram – what is sometimes called a 'functional landscape'.

A very simple example is shown in figure 3.1. This refers to some imagined machine with two variable features, say a carburettor with two adjusting screws on it. The contour lines represent screw settings of equal efficiency. To find the optimum setting for the carburettor you might first adjust screw $X$ in the direction that improved performance, until there was no further improvement, then similarly adjust $Y$, and then $X$ again, and so on – following a path like (*a*) in figure 3.1. Or you might make alternate small changes to $X$ and $Y$ sensing the gradient, as it were, and following a more direct path – like (*b*) in figure 3.1. If you imagine the contours defining an altitude so that more efficient positions are lower, then the path taken through an optimising procedure could be compared with that of a ball rolling in a basin. It too can find the lowest point *without testing every point*. That is the key idea: if there are smooth relationships between the efficiency of an object and some aspects of its structure, then the structure can be optimised without having to try out every possibility.

You can think of a protein molecule, say an enzyme, as a machine containing as many adjusting screws as there are amino acids in the molecule. Each screw has twenty possible settings corresponding to the twenty amino acids. Some of the screws – near the active centre particularly – are coarse adjusters. Others, more remote from the active centre, have much less

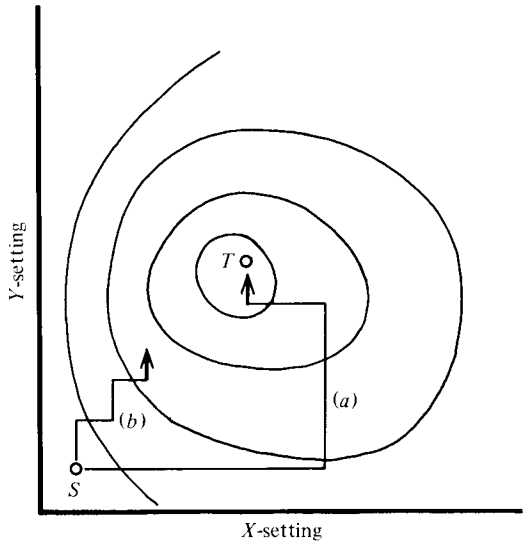



Figure 3.1. A very simple functional landscape corresponding to a machine with two adjusting screws that have some single optimum setting (*T*). The contours represent settings of equal efficiency. (*a*) and (*b*) correspond to alternative tuning procedures starting from some inefficient setting (*S*). If we regard more efficient settings as downhill this diagram becomes a contour map for a basin, and the optimisation of screw settings is analogous to a ball starting at *S* and finding the lowest point *T*.

effect – they are fine adjusters. Occasionally a mutation takes place that, in effect, flicks one of the screws to a new position. If as a result the enzyme works better, then the gene specifying it, and within which the mutation took place, will tend to become more prevalent and eventually displace the earlier 'setting' for that amino acid. (Because on average organisms that contain the improved modification are more likely to have offspring and pass on the gene specifying it. Even if the benefit is only marginal the long-term effect will be inexorable because always tending in the same direction. Eventually the improved modification will displace the other.)

For the evolution of a protein the relevant functional landscape contains not two but perhaps 200 structural dimensions, with the number of possible combinations of settings so vast that only a minute fraction of them could ever have been tried out in the whole history of the Earth. But in spite of the vast multidimensional volume of this protein hyperspace, distances in it are quite short. There are indeed $20^{200}$ sets of settings for the screws on a protein made up of 200 amino acids, but any two of these can be inter-



converted in no more than 200 moves. It is a question of knowing which direction to take.

It is sometimes said that since mutations are chance events the exploration of the space of protein sequences must be a random walk and hence the chances of any particularly effective sequence being stumbled on is too remote to be worth thinking about. But this fails to understand the nature of evolutionary exploration and the role of chance in it. The exploration of the protein space is not like a random walk, it is like a succession of random steps checked at each step to see if it seems to be in the right direction. A single step that improves function will be amplified and become the most likely basis for the next step. The most important effect of mutations is thus far from throwing up new random amino acid sequences on the off-chance that they might be useful. It is to make small random modifications to sequences that already are useful. Having hit on an idea that works, the immediate surroundings are explored. When a better one is found this in turn becomes established by the automatic process of selection and then the new surroundings are again explored. Then the best modification of the original modification is selected . . . Eventually some optimal sequence is discovered. Such a procedure is a progression from a member of some large set of possibilities – for example the set of all amino acid sequences that have at least some sort of activity of some very general kind – to members of smaller and smaller subsets. Eventually very particular structures can be arrived at that have efficient and specific action.

Now chance has a part to play in all this, as it has a part to play in chemical reactions or in the diffusion of a gas through a hole. But that is not to say that the course of evolution is 'pure chance' any more than chemical reactions are or the behaviour of gases is. Natural selection is an optimising procedure analogous to the procedure for tuning a carburettor, or to the way in which a ball in a basin discovers the lowest point. All points do not have to be visited if there is a reasonably smooth relationship between structures and functions. Optimisation procedures work on this basis: when you have hit on a good idea a better idea is most likely to be next door. Natural selection is no exception.

Natural selection could arrive at seemingly wildly improbable structures quite quickly if the circumstances were right. Once natural selection in moving from set to subset to sub-subset, etc. had made the equivalent of a series of 200 choices between equally probable alternatives, that is to say once 200 bits of genetic information had been accumulated, the *a priori* probability of any one system so selected would be one in $2^{200}$ (i.e. $10^{-60}$). Two hundred selections may not seem very many, but if they can

be accumulated in series, if the outcome of one selection can consistently become the basis for the next (and this is possible in Nature only for replicating structures) then these two hundred selections each between two alternatives becomes equivalent to a single selection from *ca* $10^{60}$ alternatives. When you consider that the number of atoms in the Earth is about $10^{50}$ it is clear that natural selection, because it operates on replicating structures, is capable of choosing quite quickly from a notional bag of possibilities that is far too big ever to have been actual.

One might discuss in similar terms the optimisation of other parts of organisms, say the shape, size, flexibility, etc. of a fly's wing, where the lowest point on some multidimensional diagram would be determined by a compromise between such things as efficiency in flight, ease of manufacture and invisibility to predators. The evolution of whole organisms can be described in terms of a huge imaginary diagram – a vast landscape with the net tendency to survive and leave offspring (biological fitness) as the only measure of efficiency.

Landscapes of this sort were introduced by Wright (1932) and are frequently used in the discussion of evolutionary processes (for example by Dobzhansky, Ayala, Stebbins & Valentine, 1977; Lewontin, 1978). The usual convention is to regard optimisation through natural selection as a climbing of mountains and to talk of 'adaptive peaks' in the evolutionary landscape. I prefer to think in terms of basins – downhill as fitter – because that is more like a potential energy diagram describing simple physical systems. Also one can imagine adaptation as being like a ball rolling in a basin.

This improving effect of natural selection (adaptation) is not the whole explanation for evolution by any means. Most of the time, indeed, natural selection operates to keep things as they are, with all subsystems near-optimised. It is one thing to be able to adjust a pre-existing machine, it is another to invent the machine in the first place. In terms of the functional landscape: once you are in a basin natural selection will find its lowest point, but how is the basin found? Then there is the problem of multiple minima. A difficulty with optimisation procedures is that they only seek out a best local solution, the local lowest point. But there may be some much lower point separated by intervening high ground.

The form of a functional landscape depends on the environment in which the organism finds itself (think of those flies: the optimised wing design could depend on, say, average wind speeds or on the predators that were around). So if the environment changes, the functional landscape may undergo 'earth movements': what was a basin may become a slope allowing movement to a nearby lower point. Indeed the functional landscape should be regarded as variable and mobile (Waddington, 1957). One can see the possible effects of this by imagining a collection of ball bearings placed in one hollow of a dynamically undulating and complex surface. They might well become spread out in time into separate collections occupying different hollows.



I will return to the question of how natural selection can become an inventor. In the meantime let us note that evolution is not completely to be described in terms of optimisation procedures. A man is not an optimised fish, or a fish an optimised bacterium.

What seems more serious still is that a man is not overtly more efficient than a fish, or a fish than a bacterium. The overall function of an organism – the function to which in the end all its most elaborate machinery is subservient – is to be fruitful and multiply. What is optimised overall is fitness – the tendency to leave offspring in the long term. Lower organisms are often very good at doing that.

So where does this leave us? Orderliness, we decided, was not sufficiently characteristic of biology to be the crucial product of evolution. Natural selection indeed generates order to the extent that it seeks out efficient subsets of possibilities, but the amount of ordering thus achieved is, as we saw, tiny compared to the ordering that goes on when, say, a teaspoonful of water freezes. Efficiency is more to the point in that it introduces the idea of function. The overall function of genetic information turns out to be something rather simple – to persist or at least to be continuous (although possibly changing) in time.

Rocks are good at persisting too. And the 'function' of a crystal structure might be said to be persistence, a stable polymorph being the most persistent arrangement of component molecules in the circumstances. To avoid these quibbles we must add that in biology what persists in the long term is not an object, such as a rock; or a thermodynamic fact, such as the stability of a crystal structure; but a 'plan'. In biology it is software that persists in the long term. But there is still a difficulty in seeing biological efficiency – fitness – as the key feature that increases with evolution. On that score man would be less highly evolved than many bacteria.

*Complexity* One has to be careful too with complexity as the measure of evolution. After all, any heap of rubbish can be complex (that is, needing a lot of information to specify it). Yet, organisms that are in our eyes highly evolved are more complex than primitive organisms. Why does evolution go with increasing complexity?

I think there are three main reasons. First, there is the point made by Maynard Smith (1969): the first organisms must have been simple, to become more complex was then the only way to go. This is not sufficient to explain why organisms became as complex as they did, however. Even the space of protein molecules of up to 200 amino acids long should have been roomy enough for a virtually endless exploration. Shortage of possibilities cannot have been the only consideration.

Second, we might very well argue, in answer to the last point of the previous section, that what evolution really does is not to increase efficiency, to create organisms of increasing fitness, but to come to produce organisms that can fill niches of increasing difficulty. To be a land animal is, in many ways, more difficult than to be a sea animal, and, on the whole, land animals are perhaps more interesting, more highly evolved we would say. To come to live up in the trees, our ancestors had to develop some pretty advanced technology in the way of eyesight, hand grip, and so on. As a result of all this they were no more biologically efficient than the soil bacteria beneath them; but they were a good bit more interesting.

Now any sort of advanced technology usually calls for complicated answers. Engineers may, for aesthetic or economic reasons, seek simple designs for machines, and natural selection will tend to favour simplicity in organisms. But in the event, in both cases, the most efficient machine is likely to be complex: to get something just right for some difficult task usually calls for nests of subsystems. The latest kind of aeroplane may be simple in conception, but infernally complicated when it comes down to the nuts and bolts. The much greater number of possibilities inherent in more complex structures improves the chance that among these possibilities something suitable will be found. Thus organisms, if they are to occupy difficult niches, may be pushed on to complex answers. And it is the organisms that have been so pushed that we choose to call 'highly evolved'.

Huxley (1974) defined evolutionary progress in terms of an increasing ability to control the environment and to be able to cope with changes in the environment. The means to those ends – ways of controlling water loss, or of exerting force, or of maintaining body temperature – are bound to call for sophisticated engineering.

A third factor would be effects of irreversibility in evolution. Following Maynard Smith (1970), Saunders & Ho (1976) argue that in optimising fitness reductions of complexity are less likely to be effective than increases (roughly, because usually anything an organism has got it needs).

We will discuss later, and on a somewhat different tack, how features might tend to become locked-in when they have been present for a long



time in a line of organisms. If, for whatever reason, features become difficult to remove (even when their removal might have been appropriate in adapting to new circumstances) then the only way of changing is to add new features on top.

*Co-operation between parts* Some would say that to be organised at all a system must be made up of co-operating parts. Some overall function is divided into subsidiary functions that are carried out by distinct 'organs' working in collaboration. (The system is literally organ-ised.) Certainly this is what machines are very often like – and it is what organisms are very strikingly like. Both Coleridge and Haldane used this characteristic to define living systems. Descartes had a similar view, by implication, when he likened a living thing to a clock. Let us consider this example.

At the highest level a clock is a device for telling the time. This function arises from a collaboration between a number of subsidiary devices – a chassis, an energy store, a gear train, a governor, a display, and so on. Each of these has further subsystems. For example, the energy store may consist of a ratchet and a spring; the governor of an oscillator and escapement; the display of a dial and hands. At the next level down there are further divisions – the ratchet contains a pawl and notched wheel perhaps. Even this is not the end: the materials must be suitably chosen, the atoms and the way they hold together is important too.

At every level there is room for design modifications, for adjustments to shapes and sizes. Some of these will have little effect on function; others will affect function strongly. (For example, the spoke structure of a wheel in the gear train will be less important than the form of the cogs.) The clock on your mantelpiece could be represented as a point on the surface of a multidimensional functional landscape. (One dimension would represent efficiency in time-keeping while each of the other $n$ dimensions would be concerned with specifying some size, shape, material and so on: together some set of specifications should be sufficient in principle to define the clock on your mantelpiece.) One can even imagine this landscape being explored by successive arbitrary modifications to its components – these either being retained, if they improve the overall function, or being rejected. Thus a better clock might be 'evolved'.

But an imaginative designer is not restricted to such minor adjustments in discovering new and more effective machines. The clockmakers have come up with quite different design approaches to various subsystems. It was found that a pendulum could be replaced with a spring-loaded wheel

in the oscillator, for example. Such different design approaches would be represented by distinct basins in the clock landscape. Either you do it one way or the other: intermediate structures are no use. There seems, then, to be no possibility of going from one to the other through successive small design changes. Yet if it is only a subsystem that has been radically redesigned the new basin will still be in a similar part of clock space.

Of course there are also quite different clock designs – sand clocks, water clocks, sundials, piezeoelectric clocks have little in common structurally. Intermediates between these are even less likely to be satisfactory.

Turning to organisms, we can consider the ultimate co-operating parts to be small molecules, such as amino acids, etc. There are three features worth noting here.

*1. Co-operation is (to some extent) modular.* The small molecules do not all interact with each other on an equal footing: very often groups of molecules interact strongly with each other and then the group as a whole has some functional relationship with other molecules or groups. For example, the amino acids that make up a protein molecule interact strongly and in a complicated way with each other in making a folded structure. This machine may then have some rather simply defined function – for example it catalyses reaction A → B. The situation might be compared to a hi-fi system which consists of modules, such as amplifiers and speakers, of the most advanced design which any fool can plug together. A consequence of modular design is that it allows sections to be interchanged with others whose detailed working might be somewhat different. A module is interchangeable with another that has the same function.

Modular construction of organisms makes sexual reproduction possible. Here, in effect, half the components of one pre-existing machine are mixed in with half from another. This is only possible because gene products are interchangeable modules – at least within a given species.

*2. Co-operation is (to some extent) hierarchical.* To be modular is already to be hierarchical to at least one level, but there may be many levels for some parts of the organisation of an organism. For example, enzymes are often made from several subunits: ribosomes are made from many proteins and RNA molecules. More complex organelles, such as mitochondria, are made up of large numbers of proteins, lipids, and other components. For multicellular organisms a structural hierarchy extends upwards through cell, tissue and organ. It seems to be a general principle of efficient and complex organisations that they tend to be structurally somewhat hierarchical (Simon, 1962). Also the module idea appears at all levels – the kidney, for example, is practically a plug-in unit.



*3. Co-operation is between standard micro-parts.* Life on Earth now embodies another idea that is familiar to engineers: it keeps to a standard kit of (small molecule) components. The real design work only starts at the next level up: not in the choice of micro-components – because there is very little choice – but in the way in which these components are arranged.

*Why co-operation between parts is characteristic of life* Efficient machines (clocks, organisms) are often organised in a strongly co-operative way. It is a feature, and a snag, about such organisations that their components have to be critically made and put together exactly. A deck chair is a more co-operative organisation than a heap of stones – and more comfortable: but the latter has the advantage that its components can be put together almost any way and still be something to sit on.

So far, in considering evolutionary landscapes, we have been using the convention that downhill is fitter. But for this part of the discussion the normal 'uphill' convention is, I admit, better. One then imagines the highly co-operative machines that organisms are, as being perched precariously on high, narrow ridges in the functional landscape.

The question of how organisms could ever get to such places will be the subject of the next section; but we can see already how it is that only organisms among naturally occurring systems could negotiate such hazardous regions of object space. Where one false move can be fatal all will be lost sooner or later; so you do not expect to find among non-living objects very many whose survival depends on being highly co-operative. If organisms can survive and indeed continue to evolve in spite of this it is because for them one false move does not matter. If there are many copies around moves of all sorts can be made without the plan being lost – provided the moves are not too frequent and provided the viable systems continue to reproduce. It is life's way of surviving in the long term, not by just being there but by continuously making copies, that allows that most characteristic feature of the evanescent systems through which it persists in the long term. Haldane's comment that organisms are systems whose parts co-operate is not strictly a definition, but it points to the most characteristic aspect of the organisation that is generated by evolution. But how is it generated?

## The evolution of functional interdependence

The question 'which came first, nucleic acid or protein?' is only one of the chicken-and-egg questions that can be asked about central

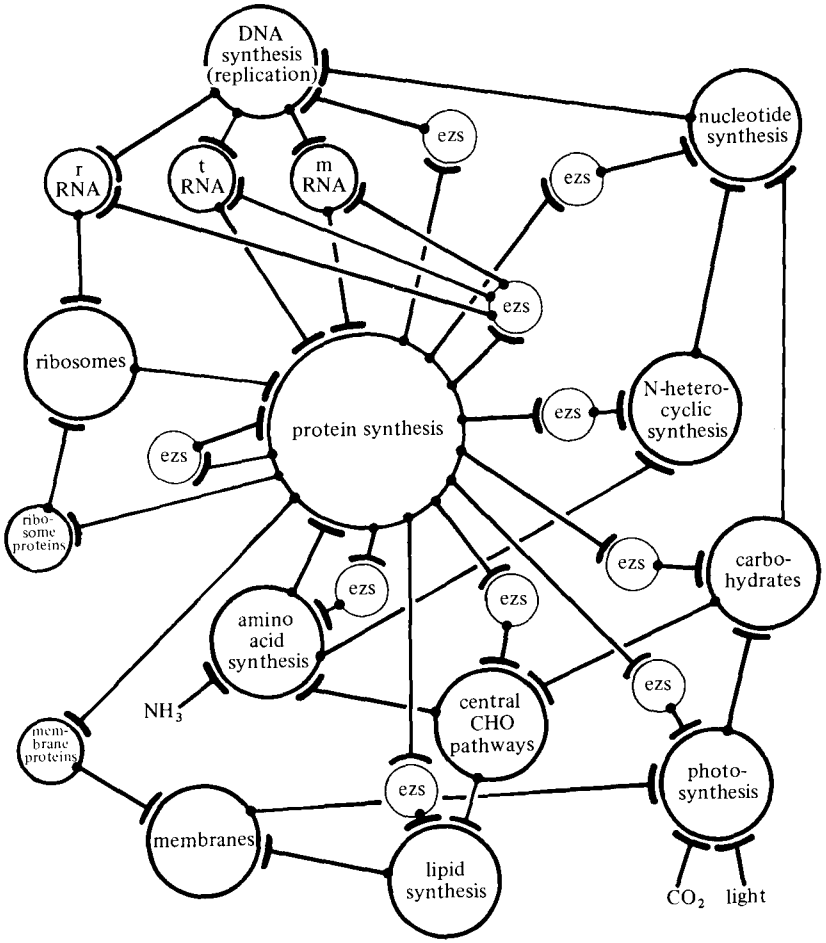



Figure 3.2. Some of the interdependences of central biochemistry. A dependence is represented by a 'crutch' symbol. ezs, enzymes.

biochemistry. Quite simply, in central biochemistry everything depends on everything. A sketch of just some of these interdependences is given in figure 3.2. How could such a situation have evolved? This question is the more pertinent when we consider that co-operation between parts seems to be the most characteristic feature of the kind of organisation generated by evolution. Yet it is not immediately obvious how that main engine of evolution – natural selection – could have produced it. One can see how, once it is there, such an organisation can be locally optimised, and kept like that, through natural selection. But how are these interacting machines



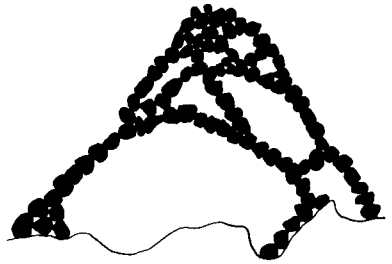


Figure 3.3. How could a structure like this be built if you were only allowed to touch one stone at a time?

set up in the first place when the significance of each depends on the existence, already, of all the others – when 'the whole is pre-supposed by all the parts', as Coleridge put it? The ball may be able to find the lowest point in its basin, but, with so many parameters having to be right, the mouth of the basin in the relevant evolutionary landscape seems too tiny ever to have been hit on. So many things have to be there before a machine built along the lines of modern organisms could start to work at all. How can you arrive at a situation in which everything depends on everything, through a succession of small modifications to some initial system that was presumably not so critically made?

*Getting stones to co-operate* Let us take an arch of stones, as in figure 3.3, as a model. The integrity of this structure depends on everything being simultaneously in place. It is a highly interdependent structure, like the modern biochemical system (cf. figure 3.2). Would it be possible to arrive at such a structure through a series of small steps? Suppose that we were to build an arch such as that in figure 3.3 using only stones that we found lying about and using only one hand; and we were restricted in our building work to touching only one stone at a time. At first sight this task might seem impossible – and it would be impossible if the stones in the final product were the only ones allowed to be used. But there was nothing in the rules to say that: some add-and-subtract procedure such as that shown in figure 3.4 could provide a solution.

One might imagine a situation in which highly arched structures like this were generated preferentially. Imagine a community of primitive people who worship heaps of stones. They see an accidental heap as the work of a god architect worthy of being faithfully copied. Occasionally among the sets of identical cairns that were being made one is a bit different: a stone has been changed, added or removed. This was, of course, carelessness on

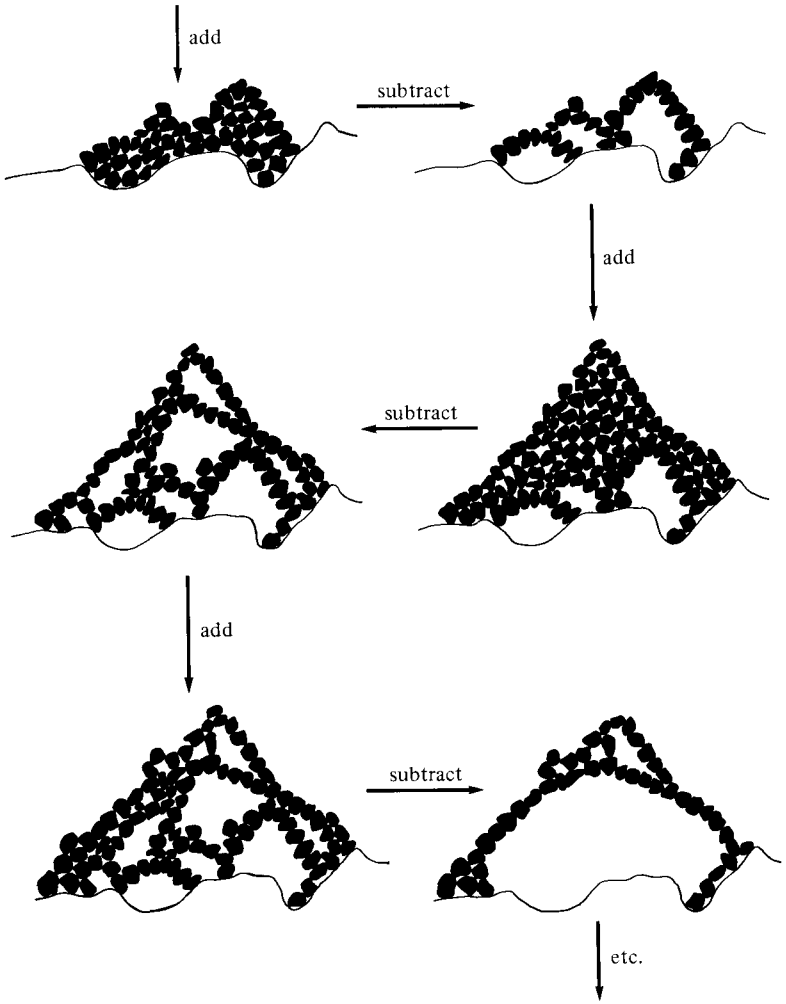



Figure 3.4. A solution to the problem set in figure 3.3. The strongly interdependent structure could have been made one stone at a time if stones had been removed as well as added.

the part of the builders; but the high priests of the society see in it the hand of their god. They ordain that in such circumstances two considerations will determine whether the novel design is worthy of being copied in preference to the old design. First if it is taller and second if it uses fewer stones. Come back in a few thousand years and you would find great arched structures, like cathedrals, all over the place. Two selection pressures on the replicating structures – to perform some function (i.e. to be tall), and yet to do it as economically as possible – would favour designs in which a relatively small number of components were strongly interdependent. Such seemingly paradoxical designs become accessible through evolutionary processes that involve subtractions as well as additions.



Add-and-subtract procedures are common in manufacturing. It may be difficult to see how an object was made because something necessary to hold the parts together during manufacture is no longer there in the final product. How did the ancient Britons put in place those huge lintels in Stonehenge? You can be sure there was some sort of 'scaffolding' which was later dismantled – earth ramps most likely.

The elephant's trunk is a paradoxical structure. How could it evolve from a short snout when structures of intermediate length would seemingly be no good? (Too floppy for nuzzling, not long enough to be properly prehensile.) Maynard Smith (1975) has discussed this evolutionary conjuring trick. It appears to have depended on a feature no longer present – a lengthened lower jaw that allowed a short forerunner of the trunk to be functional (Watson, 1946).

We know that features, subsystems, are subtracted as well as added during evolution – we can see that in later parts of the process that are still visible to us. It would be strange if early biochemical evolution had been different in this respect; it would be an improbable assumption even if there was no evidence for earlier subtractions in the outcome. But there is such evidence: the multiple interdependences of central biochemistry surely call for some kind of 'scaffolding', for other subsystems at one time (no longer present because no longer needed), on which our now interdependent subsystems leaned during an earlier era while they were evolving. Of course this 'scaffolding' was inadvertent, it was not erected *so that* a subsequent elaboration would be possible. Natural selection has no forethought, but it can be efficiently opportunist. If a stage is reached when subsystem X can be dropped to the advantage of the system as a whole, you can be fairly sure that subsystem X will be dropped. But when that happens the remaining subsystems may very well be more strongly interdependent. By analogy with the stone building game, later stones depend on the earlier ones but not *vice versa*. But when some of the earlier stones can be removed because the later stones can now lean on each other, a stage may be reached at which a co-operative set of later subsystems works better than earlier, less critically interdependent, subsystems.

One very critically interdependent set of subsystems is the set of machines used in protein synthesis – the transfer RNAs, the enzymes for attaching

amino acids to them, the ribosome, etc. (These will be discussed further in Chapter 9.) For the reason given at the end of the last section, a strongly co-operative organisation only becomes plausible in Nature in fully reproducing systems. However it happened, the invention of protein (that is, the ability to put amino acids together into long sequences specified by genetic information) was an achievement that only fully reproducing systems could have brought off. However protein synthesis evolved, it did so within organisms that already had highly competent hereditary machinery and that could work well enough without protein. Think of what would have happened, though, when in some line of organisms machinery first appeared that could consistently generate particular amino acid sequences. New ways of making catalysts, new ways of making structures of all sorts would become accessible. With the invention of protein nothing in biochemistry would ever be the same again. Then doubtless earlier ways of catalysing reactions and building structures would have been displaced. But only then: up to that point, while a mark-I protein synthesiser was being perfected, protein could not have had the central place that it now has. What is so evidently a product of evolution could not have been a requirement for evolution to begin with. That only happened when those subsystems went missing that had inadvertently provided the 'scaffolding' to support an otherwise impossible-looking enterprise.

We can perhaps best think of long-term evolution as a combination of many (always small) increases in ordered complexity combined with (sometimes large) decreases (a 'forward creep–back leap' progression; Zuckerkandl, 1975). Paradoxically, perhaps, it is the character of the decreases that imposes a trend to greater complexity in the long term. Think again about the stone building game. You might dismantle a heap of stones in the exact reverse order of its construction – through a succession of stable smaller heaps. In a similar way you might imagine an evolutionary process going into reverse. (Although it is hard to see what selection pressures could bring about such an exact reversal, it is at least possible in principle in the sense that the process would be through a succession of viable intermediates.) The trouble comes when evolutionary pressures favour the deletion, not of the last item that had been added, but of some earlier item(s): when one or several stones lower in the pile can be removed; when the dismantling is not in the exact reverse order of construction; when what had previously been done by an earlier subsystem can be done better by some more sophisticated combination of later subsystems – when an arch of some sort is formed. Further simplification is then blocked. You can not dismantle an arch any more than you can build it by only touching the stones in the arch itself. On the other hand, on the stone building model



there is no such restriction to increases in complexity at any time; to adding a new stone on top.

Clearly a heap of stones, even a partly arched heap is too simple a model; but it is probably fair to say that the well known tendency for ancient design features in organisms to become fixed arises in part from the continual pressure under natural selection to simplify in the short term. One highly rigidifying economy would be where a structure, for example a molecule, is put to many diverse uses (think of glucose or adenine or any of the protein amino acids). Such a structure becomes unchangeable, not because some modification might not be beneficial for one of its functions under some circumstances, but because the same modification could hardly ever fail to be disastrous for at least one of its other functions. Note, though, that multifunctionality could only be the gradually acquired outcome of evolution. A structure that is indispensable now would have been less so at a much earlier time. Earlier still it may not have been there at all.

In the most general terms we might say that a subsystem will tend to become fixed during evolution when other subsystems come to depend on it. At first sight you might suppose that later subsystems will consistently tend to fix earlier ones. This might be true if the organisation of organisms had a simple 'heap' structure – if later additions always depended on earlier ones but not *vice versa*. But that is not quite so. Organisms have an 'arched' organisation to a very considerable extent, in that very many subsystems depend mutually on each other. As we have already seen, this implies that, as well as there being subsystems present now that were not there in the past, there were subsystems in the past that are not there now.

We might think of a feature that has recently appeared in an evolving line as having the status of an 'optional extra' (and likely to be variable, as Darwin noted). If it remains for long it may become a necessity (and less variable) because other subsystems have come to presuppose it – or perhaps because the organism changes its way of life.

Sometimes, subsystems become redundant. This is less likely to happen if it has acquired many diverse functions, if many other subsystems lean on it. But it will never be quite safe. An altogether more sophisticated 'technology' may oust it in each of its roles, as the invention of protein no doubt displaced previously established means of carrying out organic reactions (Chapter 9). On the stone building model, even an arch can be dismantled through a new arch formed *above* it (figure 3.4).

*Evolutionary trees* Darwin pointed out that the familiar fairly ready classification of organisms into groups and subgroups was evidence for evolution – that the present structure of relationships could be under-

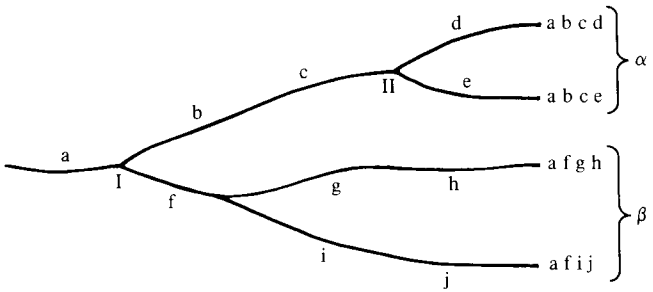



Figure 3.5. Each of the four species on the right-hand side of this evolutionary tree shares a characteristic, a, that became fixed before the last common ancestor, I. The subgroup $\alpha$ has two characteristic features, b and c, that were fixed between I and II. For the subgroup $\beta$ only f is a defining characteristic.

stood in terms of descent from common ancestors (figure 3.5). As a rule the characteristics used to define some group of organisms should have been present in the common ancestor of that group, and yet would probably not have been in the line that led to the common ancestor for too long – otherwise it is unlikely to be characteristic. On the other hand, it should have been an old enough idea, by the time the common ancestor appeared, to have been already fixed – otherwise the feature would probably not be maintained in subsequent independently evolving lines (see figure 3.5 and its legend).

Symbiotic relationships may complicate things. In the origin of eukaryotic cells, for example, mitochondria look like degenerate forms of previously symbiotic bacteria, which before that were independent (Margulis, 1970; Richmond & Smith, 1979). Clearly we cannot rule out the possibility that early biochemical evolution involved such collaborations. Figure 3.6 illustrates an evolutionary tree that contains a fusion of previously divergent lines. There is an ambiguity here as to whether A or B is the proper last common ancestor of, say, D and E. But the ambiguity does not seriously affect the idea that all life on Earth (because of its biochemical near-identity) must be derived from some common ancestor. If D and E have features in common that are rather detailed and seemingly arbitrary so that convergent effects can be discounted (for example if they have settled on exactly the same set of amino acids), then B is the last common ancestor at least as regards these features. All these features must have been in B, whether or not they were in A.

Sexual reproduction also mixes genes – to allow beneficial design modifications to come together in the same individual. Such fusions are, however, between very close branches of the tree – usually between members of the same species. If we take a long enough view, then, and see the branches of the tree as representing species, this kind of gene mixing does not change the general picture.



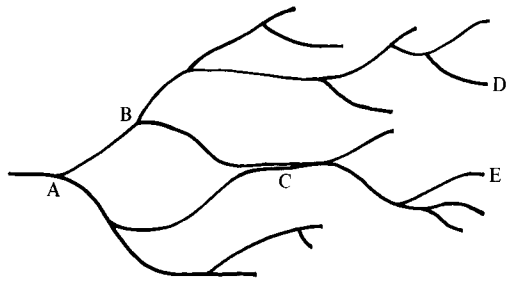



Figure 3.6. Compare figure 3.5. Because of the symbiosis, and subsequent fusion at C, E may have inherited features that became fixed between A and C, and that were in neither B nor D. But any features that are shared by D and E (and which do not have a convergent explanation) must have been present in B. B is then the effective last common ancestor of D and E at least with respect to these features. It is in this sense that we may say that the universality of so much of present-day biochemistry indicates that all life now on Earth is descended from a common ancestor, whether or not symbioses were involved.

In the construction of evolutionary trees there are various levels of anthropomorphism to be guarded against. First there is the gross anthropomorphism of thinking of ourselves at the end of a line that started with some ultimate ancestor, some first ancestor able to evolve under natural selection. Of course, in a rough sort of way, and forgetting about symbiotic effects, there is a line – and it raises a marvellous image. There you are standing beside your father, your grandfather, your great grandfather . . . your (great)$^n$ grandfather. By the time the line is about 200 km long your (great)$^n$ grandfather's knuckles are touching the ground. By 7000 km the appearance of the company would be less like a military parade than a fishmonger's slab. Seen in this way, as a line, evolution would not only be comic but incomprehensible. Yet sometimes it seems almost to be taken for granted that early biochemical evolution ought to be seen as a line; that the tree of evolution was like a standard apple tree (figure 3.7), straight to begin with (knowing where it was going perhaps?) and then bursting into a flourish of alternatives.

The tree in figure 3.8 is of the sort in which species that no longer have

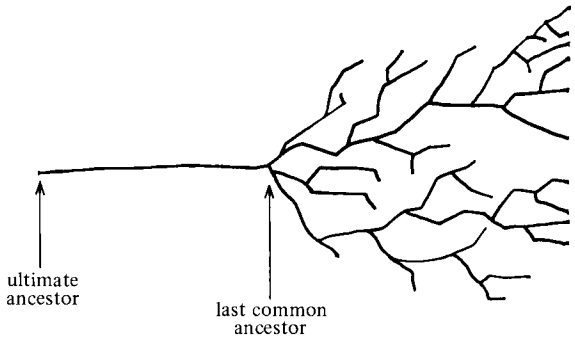



Figure 3.7. It is often implied that before the last common ancestor of all life now on Earth, evolution followed an unbranched course, it being felt, perhaps, that this is the most economical hypothesis. But branching – experimentation with different design modifications – is an essential part of the mechanism of evolution as we now see it. The most economical hypothesis should be that the tree has always been branched. This picture is too much like a standard apple tree – another human artefact.

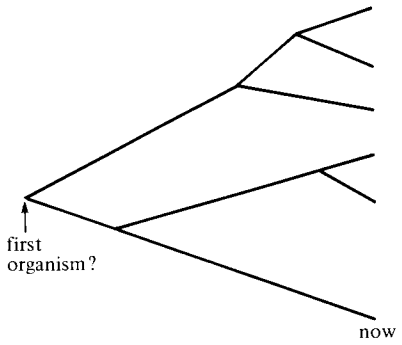



Figure 3.8. Thinking only about organisms that are alive today one may be persuaded that their common ancestor must have been the first organism. This is not so.

living descendants are left out. Such trees are very common in discussions of the evolution of proteins, and understandably so since only the proteins of living species are generally accessible. Although useful, such trees are clearly incomplete and they can be misleading. From such trees it is easy to get the impression that the last common ancestor is to be identified with the ultimate ancestor of life on Earth.



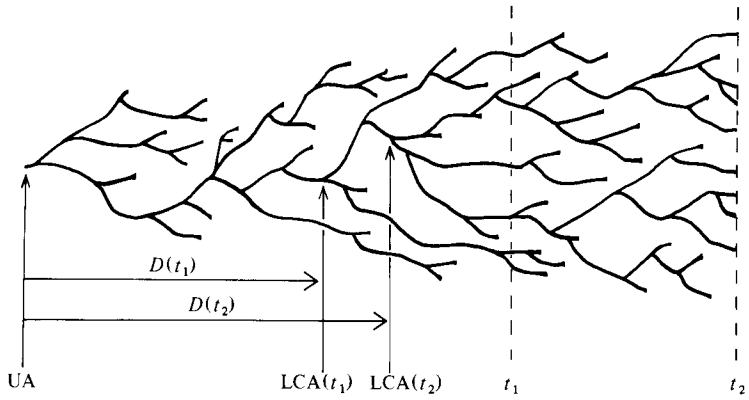



Figure 3.9. During evolution later forms have often displaced earlier ones entirely: the vast majority of species that have ever lived are extinct. So the evolutionary tree was something like this – heavily pruned. At a given time (say $t_1$) all living species could be referred to some last common ancestor, LCA($t_1$). But at a later time, $t_2$, there is a new common ancestor, LCA($t_2$). Any such change can only increase the distance between the ultimate ancestor, UA, and the LCA (from $D(t_1)$ to $D(t_2)$). That our last common ancestor was evidently highly sophisticated biochemically suggests that, for $t$ = now, $D$ is very considerable.

Darwin saw clearly the effect of extinctions on the form of the evolutionary tree. Towards the end of the chapter on natural selection in *The Origin of Species* he says: 'Of the many twigs which flourished when the tree was a mere bush, only two or three, now grown into great branches, yet survive and bear all the other branches; so with the species that lived during long past geological periods, *very few now have living and modified descendants*' (my italics).

One might add that of the species living at a given time the number that will have living and modified descendants at increasingly distant times in the future can only become less. The last common ancestor of all life at any given time can only increase its distance from the ultimate ancestor (see figure 3.9). That we all share such a complicated biochemistry indicates that in Darwin's tree of life on Earth this distance is now very considerable. What is decidedly not indicated is that the last common ancestor had the only kind of biochemistry there ever was. This would be like supposing, on the basis of an examination of a few mammals, that all life must have hair. Hair was fixed in the common ancestor of what is now a large and successful group of organisms. Other features, such as a backbone, were present

and fixed in earlier progenitors and so are common to still larger groups of organisms still living. The more fundamental the design idea, the longer ago it was fixed and the fewer alternative ideas there now are. It is quite in keeping with this that at the most fundamental levels the number of design approaches should have fallen to just one. The unity of biochemistry is an artefact generated by our point of view: the unity to be seen at the base of life at any given time reflects those parts of the organisation that happened to be frozen into the last common ancestor of life at that time (see again figure 3.9 and its legend).

To conclude this section: it is typical that features become invariant in an evolving line. There is no reason to suppose that our now central biochemistry was any different in this respect – that it was born invariant. Surely, like other kinds of invariance that have appeared at all levels, our biochemistry had invariance thrust upon it. Part of that invariance was the set of small molecules that we ambiguously call 'the molecules of life'. If the invariance was a product of evolution so too, presumably, was the choice of those molecules. They *became* 'the molecules of life'.

### How radical was very early evolution?

If it is true that the multiple interdependences of our biochemical subsystems is an outcome of subtractive episodes in early evolution, what was it that was subtracted? Before there were enzymes and lipid membranes, for example, how were organic reactions controlled? By simpler kinds of peptides and simpler kinds of amphiphobic molecules? Not necessarily at all. The evolutionary mechanisms that we have been thinking about do not operate wholly through gradual transformations of subsystems through continuous successions of intermediate forms. These mechanisms have been takeovers. Here there is no necessary implication even of similarity between the new structures and the structures they replace. It was rather functions that were continuous: the new structures had to perform some similar task to the old ones – and presumably rather better – but they did not need to be made of the same stuff, or for that matter to operate at all similarly. You do not have to use stones as the scaffolding for a stone arch. Later in this book we will try to identify more closely the precursors of our now central biochemicals. In this present section we will be concerned only with general expectations – that if later evolution and the properties of machines generally are anything to go by, the basis of the first biochemistry would have been not only simpler than ours but of a different kind.



We are so used to the idea that evolution must proceed through continuous successions of small modifications to organisms – which is broadly true – that it may seem a contradiction to suggest that the evolution of the subsystems of organisms could be different. In so many discussions of the origin of life it seems to be taken as a necessary feature of the Darwinian idea that if, say, a complex protein is now used for some vital function, then the original way that function was carried out must have used some very simple kind of 'protein' (a couple of amino acids joined together perhaps, or a semi-chaotic polymer). Yet evolution at the higher levels is so evidently not just the elaboration of a pre-existing set of subsystems – many new things have been invented – that it is naive to suppose that during the evolution of our biochemical subsystems radically new inventions were not allowed.

There is no inconsistency in the idea that a succession of systems may be continuously related to each other through small modifications while the subsystems are not. In the preview we considered a rope as a model for this: a continuous rope may be made up of discontinuous fibres. A rope might even have different kinds of fibres in it, so that at one end it was hemp and at the other nylon, with no sharp changeover anywhere.

To be a more realistic model for an evolving line of organisms, we should think of a rope in which the fibres making up more central strands are longer than others. The fibres that together make up the most central strand of all stretch back as far as we have any sight of the rope – to some last common ancestor. Only the more peripheral fibres are easily modified – with sometimes new ones appearing while others peter out. That is to say, the locus of active evolution is mainly in the more peripheral zones. Darwin makes this point in chapter 5 of *The Origin of Species*: that the features in organisms that are most variable are typically those that have been acquired most recently.

Depending on that now most unvarying strand of central biochemistry, there are other layers of strands which are also well hardened in. If you think of higher organisms, say the land vertebrates, there is not so much about them that is variable. Not only is the biochemical ground plan common, as it is for all life on Earth, but at much higher levels conservatism is more remarkable than variety – for example in tissues, organs, in the way these make up still higher order systems such as the digestive system or the circulatory system. Even numbers of bones often seem hard to change.

Of course there are differences between a mouse and a tortoise; but these differences are not all that far in. And the differences between a mouse and an elephant are still more superficial. If this is not immediately evident it is

because we tend to be over-impressed by the latest design modifications. Again, as Darwin pointed out, it is on (often elaborately developed) recent features that an organism may depend for its way of life. And that is because, in adapting to circumstances, it is the more recent features that are more easily changed. (On the stone building analogy, stones that have just been added, such as the top stones in figure 3.3, can easily be adjusted since the integrity of the organisation as a whole does not depend on their exact arrangement.)

Keratin was probably invented by early land vertebrates as a material for scales to cut down the evaporation of water (Fraser, 1969). In any case keratin is a comparatively recent idea – and it has a large effect on our impression of variety in land vertebrates. Different groups have put keratin to different uses. The reptiles have stayed with scales, the birds prefer feathers, the mammals hair. Different subgroups use these features in different ways – the scales of the tortoise are not just for keeping the water in; feathers are not always for flying; hair may be a secondary sexual characteristic, and so on. In hooves, claws, fingernails or horns, we see some of the multitude of individual uses to which this amazing material has been put. At a superficial level the uses of keratin are still evidently somewhat variable under evolutionary pressures. But at the level of actually changing, say, feathers into something else, birds already seem to have passed a critical point. They can seemingly only modify the form of keratin they have got – not change it in a fundamental way.

So it is, on the whole, only the more superficial strands of the rope, corresponding to comparatively recent features, that are the loci of active evolution. Deeper strands go further back to be common to larger groups of organisms – and the deeper they are the larger the groups of living organisms that share them. It is not only because organisms are descended from common ancestors but also because such strands become fixed that organisms can fairly readily be classified into a hierarchy of groups and subgroups (cf. figure 3.5). But the strands *become* fixed. Of course they were not fixed to begin with: they were not even necessary to begin with. Even such a widely well established idea as a backbone must have been hardly more than an optional extra when its first rudimentary forms made their appearance. On the most plausible extrapolation the now universally fixed strands of (now) central biochemistry were not always fixed. What was biochemistry like when protein was an optional extra? To begin to answer that sort of question we should look at mechanisms of radical change in evolution as they operate still – if mainly now at high levels of organisation.



*Preadaptation and functional ambivalence*

Most of the time natural selection is a conserver, keeping organisms in some local basin in the evolutionary landscape; keeping them adapted to their environment. If environmental pressures change, natural selection may operate as an improver in relation to the new conditions – appropriately retuning sizes, shapes and other parameters to optimal performance: adjusting the pre-existing strands.

One of the most interesting questions in evolution is how natural selection could be an inventor: how new strands originate. The main part of the answer seems to be this: through the discovery of alternative functions for pre-existing structures. Such structures are said to be **preadapted** to the alternative functions. Preadaptation is strictly fortuitous, but by no means implausible when we come to consider particular examples. It arises from a general property of machines of all kinds – what we might call **functional ambivalence**.

This book, you will find, is a tolerable flower press. (Not heavy enough to be really good, but the slightly water absorbent paper is helpful, and so are the numbers on the pages for recording the location of specimens.) The amateur handyman is familiar with the idea that a use can often be found for an object that was designed with no thought for that use. Indeed the idea of functional ambivalence is too familiar to need much elaboration: the chair is a rudimentary step-ladder, the rolling pin an offensive weapon, the pocket calculator a paper weight, and so on. Indeed the difficulty might be to think of objects that did not have endless alternative uses.

At all levels organisms can be seen to cash in on the functional ambivalence of structures. The glucose molecule might seem an ideal sort of water-soluble fuel; but then it is very well adapted too to be the unit for a highly water-insoluble polymer for making plant cell walls. And cellulose is not only used for that: in the cotton plant, for example, it puts wings on seeds.

The functions of protein are too many to enumerate, and most of these must have been accidentally hit on long after protein synthesis had been perfected. The immunoglobulins, that are so characteristic of vertebrates, were probably nevertheless derived from a class of invertebrate proteins (Barker, McLaughlin & Dayhoff, 1972). It seems to be common, indeed, for protein classes to be older than particular protein functions (Zuckerkandl, 1975).

Even a given protein may turn out to have diverse uses – keratin for example. Enzymes almost always show some activity over a range of sub-

strates in the laboratory, and this may be important *in vivo*. Jensen (1976) has discussed the likely importance of 'substrate ambiguity' for the evolution of the enzymes for new metabolic pathways. Apparently this happens through 'recruitment' of enzymes that are already present in the organisms, catalysing analogous reactions. Provided an existing enzyme can catalyse a reaction, even if only incompetently to begin with, if that catalysis confers an advantage on the organism then it is likely that subsequent mutations will be selected that will improve competence so that, in effect, a new enzyme has been discovered.

Jensen lists a number of families of proteins with similar amino acid sequences and yet more or less divergent functions; for example, the serine proteases (Stroud, 1974) and the enzymes that work with pyridine nucleotides (Bennett, 1974).

At a somewhat higher structural level, the microtubule provides another example of a structure that has been put to many uses (Watson, 1976; Dustin, 1980). Microtubules are cylinders that may be many micrometers long and have a hollow core of about 14 nm diameter, the wall being about 5 nm thick. They are made up of proteins with molecular weights of 55000 and 57000. One function for microtubules is as a sort of internal skeleton for cells, allowing them to assume highly non-spherical shapes – as for example in nerve cells. Restricted to eukaryotic cells the microtubules are also involved in that most characteristic eukaryotic function of mitosis, where they make up part of the machinery that draws the chromosomes apart. They are also the main structural elements of cilia which contain twenty microtubules apiece and provide engines of locomotion for many single-celled organisms (Satir, 1974). But then, as you might have guessed, that is not all that cilia are used for. Our lungs, for example, are kept clean through a continual upward flow of mucus powered by the concerted beat of cilia attached to the linings of the air passages.

Then again tissues have diverse functions. The main function of bone may be as a load-bearing material, but latterly mammals found that bone is useful too to transmit sound – in the middle ear. And it is hard to think of an organ that does not have several functions. The lungs, you might say, are gas exchangers – but in the opera house they also have a noisier function. Toothed jaws may be weapons or tongs as well as food processors, and so on and on.

*Why is functional ambivalence so common?* It is perhaps easier to comprehend the functional landscape for an organism than for a subsystem of an organism. For an organism there is one overall function that



matters – fitness – and hence only one functional dimension to be added to the multitude of structural dimensions. But, in contributing to fitness, a subsystem may, as we have seen, perform several functions. The ideal set of structural specifications for one of these functions is unlikely to be ideal also for the others. The design of a multi-purpose object is almost bound to be a compromise. There is not just one hypersurface but many, and a local lowest point on one is unlikely to correspond exactly with lowest points on the others. Indeed, thinking in wholly abstract terms, one might be inclined to say that any correspondence at all would be pure chance; and that for functions that depend on highly specified objects the probability of such a correspondence is too low to be worth considering – because the basins for highly specified objects are, relatively speaking, very tiny. But if we think more realistically this pessimism is evidently not justified: quite often a highly specified object – that appears to have been designed in some detail for some particular function – turns out to be preadapted, and often in some detail, for something else. (A book really is not a bad flower press, and the crocodile's mouth is a rather good playpen for its children.) When we say that functional ambivalence is a factor to be reckoned with in evolution, we mean that there is, after all, a tendency for basins on seemingly disparate landscapes to overlap; and that this principle applies to the machines in organisms as to machines in general.

The reason seems to be something like this. Much of the detailed specification needed to make a machine (i.e. functional object) are in its subsidiary components. A tube of constant diameter, for example, might be what is required to make an ideal X – although a makeshift X might be possible with something less perfect. So under selection pressures generated by X's usefulness the means of manufacturing the required tube is perfected. But tubes are very generally useful things, they are major or minor components of other machines that now become possible – some perhaps for the first time because, for them, a makeshift, inaccurate tube would be no use at all. Speculating on the origin of microtubules one might perhaps guess that the skeletal function came first, because so long as they had some rigidity they would do. But if they were of constant diameter they would be better because they could pack and align better. Then cilia, for which packing and alignment is more critical, become possible.

That particular story may very well be wrong, but the general idea seems understandable – that the perfection of a structure for one function may make another function possible for the first time. The first function may correspond to a basin with a wide mouth that is easy to find; but near its lowest point it overlaps with another basin, on an interpenetrating land-

scape, that would have been too small to have been found *de novo*. Then one of two things may happen. The original structure is further modified to be an optimal compromise for the new and old function(s), or two (or more) structures diverge, each specialising in one of the functions (for example in the appearance of new proteins through gene doubling, or in the evolution of the lung as separate from the oesophagus).

Referring to protein evolution, Zuckerkandl (1975) says: 'Any new functional development in an old molecule will benefit from the accumulated "capital" of optimisation of the old molecule with respect to the various general functions that any protein is bound to carry out.' Simply to have a definite tertiary structure is quite an achievement already; and not to interfere with other processes – not to misbehave – is another. Much the same could be said about books: a number of alternative uses is opened up by their simply being flat and chemically stable.

Technological analogies are often used in discussions of evolution (for example by Pirie, 1951, 1957, 1959; Rensch, 1959; Maynard Smith, 1975), and they seem particularly appropriate to this part of the discussion. In evolution, as in the development of our present civilisation, new 'technologies' are discovered from time to time. Some technique of fabrication is developed first in one area, perhaps for quite trivial reasons, and then turns out to have a very general usefulness. Was fire first invented to keep warm, or to fend off large animals, or to tenderise their meat? Perhaps it was at first a toy; but in any case its invention let in other technologies – adding metal and glass to the range of available materials. Perhaps in turn these materials were used for beads and trinkets to begin with. It is hard to say; as it is hard to say exactly when and for what purpose bone-like materials were first used in animals. But, for large classes of animals, once the technology of being able to control the crystallisation of apatite into hard, rigid masses had been perfected, nothing was ever to be the same again. Of course we are inclined to say that bone was invented for load-bearing functions. That was doubtless the reason for its success; but we cannot be sure that the first partly controlled apatite precipitations would have been strong enough to be useful in that way. Perhaps they were simply a means of storing calcium and phosphate – still one of the functions of bone.

Beginnings are often special. If some new technology, some new manufacturing process, is to catch on then there must first be a use for poorly made products. That is to say, on one of the many interpenetrating landscapes corresponding to different conceivable functions for these new kinds of products there must be a catchment area wide enough to be found. As



evolution proceeds other basins are discovered which may be much narrower. Some of these will also be shallower than the original basin. But there is no necessary correlation here. Narrow basins may very well be deep.

It is the discovery of deep but narrow basins that constitutes the kind of technological breakthrough that is such a familiar part of the recent history of our civilisation – in the discovery of the means of making mild steel, or PVC, or transparent glass; or, at higher levels, insulated wires, polymer films, photographic emulsions; or at higher levels still, printed circuits or internal combustion engines. Such techniques are highly sophisticated, depending on detailed know-how. Yet they are very generally useful and our whole civilisation has come to depend on them.

In organisms there are equivalent pieces of high technology – in the techniques of manufacture of, say, lipid membranes or microtubules; or of bone or keratin; or of neurone-based data processing systems. But beneath all these, at the base of everything, there is a particularly intricate piece of engineering – that hi-fi equipment on which protein synthesis depends. As I have already remarked, we cannot suppose that the early development of this technology was with any sort of view to the functions that were eventually to become possible.

### *Functional overlap*

If a given structure can often have several functions, a given function, at least of a general sort, can often be achieved through quite different structures. We might call this **functional overlap**. Examples from everyday life are obvious and numerous – such as the various kinds of clocks that we talked about. Among organisms too there is a great variety of structural means to similar ends. Such general animal functions as breathing, locomotion or defence may be carried out by subsystems that are different in detail or very different indeed. At a deeper level structurally dissimilar enzymes may bring about the same reaction; for example, Campbell, Lengyel & Langridge (1973) found a $\beta$-galactosidase different from the usual one in a strain of *E. coli* that lacked the normal gene.

Within organisms too there are often areas of overlap in the functions of subsystems. The most obvious are again at the highest levels – a bird may walk or fly to catch the worm. Such an internal functional overlap between strictly alternative modes would be pointless if it was perfect. Many birds have the option of walking or flying to get about; but that only makes sense because, although there may be times when either would do,

there are other times when one or the other is preferable or obligatory. The overlap here is incidental to a means of extending the range of the general function of locomotion – it is the regions of non-overlap that matter.

More usually functional overlaps are not strictly between alternatives. To keep warm a cat may move to the fire, or fluff up its fur, or shiver; but when it is really cold it may very well do all three.

These 'and' and 'or' aspects of functional overlap can be seen too at the level of protein molecules. The multiplicity of transport proteins in *E. coli*, for example, provides this organism not only with the ability to pick up a wide range of useful molecules, but it allows it to cope with the same molecule present under a wide range of external concentrations (Jensen, 1976; Lin, 1970). There may be more general as well as more specific transport proteins: for example, there are two aspartate transport systems in *E. coli*, one of which operates on a wide variety of $C_4$-dicarboxylic acids while the other is more fussy (Kay, 1971). Similarly with enzymes: the digestive enzymes trypsin and chymotrypsin hydrolyse overlapping but different sets of peptide bonds, while at the same time the trypsin-susceptible set is smaller. Then again, *Aspergillus nidulans* has an enzyme that will hydrolyse amide links in general as well as having one that is specific for acetamides and another for formamides (Hynes, 1975).

There is no need to continue this catalogue: it is both general and understandable that it may be advantageous for an organism to be able to carry out an important general function in a number of different ways. This may allow a wider range of situations to be coped with. At all organisational levels one can think of alternative or back-up subsystems that can be switched in; and some of these latter are effective precisely because they work differently from the systems whose functions they are supplementing.

### *Takeovers in evolution*

Functional ambivalence, functional overlap and, a third element, redundancy, can together create the most radical transformations in evolution. They can bring about **takeovers**. Here some general function that had previously been carried out by one set of structures comes later to be carried out instead by some different set of structures.

A takeover is to be inferred from the evolution of the lungs (as described by, for example, Rensch, 1959; Schmalhausen, 1968). The stage was most probably set here by a functional ambivalence of the food canal of early fish – it can hold air as well as food. This means that the food canal might act as a buoyancy device and also as an alternative means of absorbing



oxygen. Selection could then operate to improve either or both of these functions without interfering too seriously with the primary function of the food canal. It seems that a swelling or sac formed off the upper oesophagus which gradually ballooned out to form, in some lines, swim bladders, as in many modern fish; and in other lines lungs, as in the modern lungfish. It would be difficult to disentangle the exact sequence of events or the early roles of buoyancy and respiration in providing selection pressures – or indeed to be sure that there were not other functions involved in the earliest stages. But the overall story seems clear enough: part of the upper food canal of early fish evolved into an elaborate organ of respiration through successive small modifications set in train through an initial functional ambivalence. This led to bimodal organisms with a strong functional overlap between gills and lungs. Finally, for later land animals, the gill mode became redundant, the gills themselves disappearing – at least in the fully developed animals.

It is typical that this takeover was only of a general function – acquiring oxygen and disposing of carbon dioxide. In detail the function changed in that the external medium changed. This was a genuine takeover nevertheless: the general function was always vital and always maintained; and the transformation, although through a continuous succession of organisms, was not through a continuous succession of subsystems. One strand of the rope gradually petered out while another became stronger.

A takeover can be seen too between the amphibian and reptilian way of preventing the loss of water from the skin – in the change from secreting mucus to keratinised scales. And there have been many other examples of radical design changes in evolution: in, say, defence mechanisms, in modes of communication, or of protection of the growing embryo. Again it is not necessarily the case that the new means evolved from the old.

In keeping with our earlier discussions it is now mainly at the higher levels of organisation that radical takeovers are most easily effected. It is here that evolutionary experimentation is most active, where subsystems are provisional enough to come and go. It is here that the strategies of active evolution are now most easily to be seen – and it is with such strategies in mind that we should try to extrapolate back to that much earlier era when our central biochemistry was truly in the making.

Not that biochemical innovation is finished by any means. Secondary metabolism can be highly individual. For example, new enzymes are being invented still, and being discarded still, in the pathways leading to such things as flower colouring materials. And there must be some biochemical basis for such higher order evolutionary transformations as the change in

the shape of a limb, even if little is known at present of the detailed chemistry that controls such things.

Even in primary pathways there are alternatives, often, to suggest that some redesigning of the details of central biochemistry is still possible. And there can be redundancies in primary metabolism. That there are 'essential' amino acids, and vitamins, illustrates our incapacity now to synthesise vital components that were made by our remoter ancestors.

### *Efficiency range*

Let us move somewhat from 'how' questions to 'why' questions. If you have a way of making machines that work perfectly well, why change?

Because, of course, they do not work perfectly well – nothing does. In particular you do not expect that the first general way (design approach) that becomes possible for carrying out a given function will remain the best way. In any case the design approach with the greatest **efficiency range** (that is, potential for improvement in the long term) cannot be recognised as such by natural selection – not to begin with.

In evolution new opportunities for new design approaches arise. There are both external and internal opportunities to be considered. For example, the evolution of the lung was pushed, latterly, by the opportunities inherent in the conquest of the land. But this evolution depended on internal things too. And these were more than a convenient oesophagus. One can think of any subsystem within an organism existing in a 'niche' created by the rest of the organism. The subsystem and its 'niche' co-evolve. The gills, for example, co-evolved with a heart and circulation – a prerequisite for effective lungs. Thus the lungs might be said to have taken over a 'niche' that had been created in part by the gills. The deviousness of the evolution of the lung, then, was not some haphazard changing of plan (there was no plan of course) but a necessary general way of going.

On chance alone you would expect that, sooner or later through evolution, some improvement to an initial design approach would turn up – you are more likely to draw an ace the more cards you take. But the expectation of improved opportunities is stronger than this. As discussed earlier, as 'technologies' in organisms evolve – for example the ability to make tubes precisely – opportunities may appear of a kind that could not have been available to begin with, that depend absolutely on techniques of fabrication that have to be closely controlled. As any technology evolves machines of higher efficiency range come into view.

For example, there are more ways than one to kill a rabbit. A stone is a



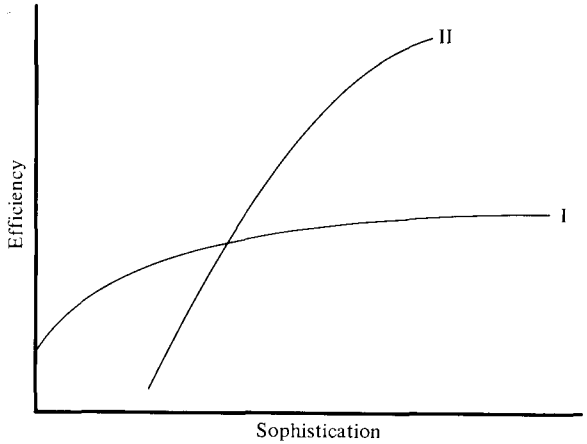



Figure 3.10. The law of diminishing returns usually applies to a particular approach to an engineering problem and can be expressed as some sort of concave-down curve of efficiency *vs* sophistication. As technology as a whole advances new design approaches to engineering problems emerge and usually some of these have a greater efficiency range than initial, 'zero-technology', systems. For example, I might be the curve for stones and II for shot guns as rabbit killers. A central theme of this book is that our nucleic acid–protein based biochemistry has a curve like II: there is no 'zero-technology' version of it that would work. Rather our system took over, beyond some crossover point, from a system with a curve like I. This earlier system was structurally as unlike our biochemistry as stones are from shot guns. Its virtue was that it could start from scratch (that is, its curve cut the *y*-axis).

zero-technology rabbit killer, an object to be found in the environment, that is fortuitously preadapted. This approach can be improved on: for example a spear as a sharpened stone attached to a stick. But the shot-gun is much more efficient although its invention depended on the development of a surrounding technology that was not always or primarily pursued with weapons in mind. There is hardly a field of engineering where there has not been a similar leap-frogging progress; where later devices were often more than improved versions of earlier ones; where alternative design approaches of increasing efficiency range have superseded each other (see figure 3.10).

It may be more difficult for that sort of thing to happen in evolution, and eventually the more central subsystems may become fixed; but we should expect such things to tend to happen – and indeed where evolution is still active they do.

Life on Earth started with zero-technology devices, some set of micro-

systems fortuitously preadapted to evolve under natural selection. We know from the sophistication of the eventual outcome that there must have been an extended period of active evolution between those first systems that could start to evolve and those nucleic acid–protein–lipid machines that were eventually to inherit the Earth. From chemical considerations, given in Chapters 1 and 2; from the highly co-operative character of present biochemistry; by analogy with active evolution as it is still visible at higher organisational levels; and from the consideration of the efficiency range of machines in general, there is no reason to suppose that those zero-technology devices that we seek were made in the way that organisms now are made.

Looking at our central biochemistry; at its 2000-odd reactions remaining unconfused, each controlled by a separate protein; at the interlocking allosteric circuits within that system; at the synthesis of these proteins which depends on a technology that seems to presuppose them; at a technology all the more ingenious in being based on a small group of small chemically accessible molecules, it is perhaps no wonder if we are inclined to identify 'life' with protein and then, by extension, to imbue amino acids with a touch of vitality – at least to the extent of labelling them as 'molecules of life' and seeing in their abiotic production a step towards life.

Yet this reaction is precisely the opposite of what it should be. It might have been appropriate rather if our biochemistry was simple or simply an elaboration of a demonstrably zero-technology machine. But it is not. This is no heap built only by one-at-a-time additions and adjustments. This is a cathedral of arches built now on each other, depending now on each other with almost every stone a key-stone in a dozen dimensions. To build these arches there were presumably earlier forms, less rigid, less ingeniously interdependent, that were to be the inadvertent discardable scaffolding for the later design.

Even Occam might have agreed that a proliferation of hypotheses beyond that of a mere adjusting and heaping up of subsystems is needed to understand the making of our biochemistry – particularly when takeovers cry out to be considered. And this is hardly a proliferation. Takeovers and the elements of takeovers – preadaptation, functional overlap and redundancy – can be seen operating in later evolution as the means of radical invention. We should doubt whether any of the molecular structures in our present biochemistry were there at the start, whether any first subsystems would have had the efficiency range to have stayed the course.



**Conclusion**

What is there to hang onto then: what has been invariant through evolution from the start?

The search for some constant underlying substance through change is natural perhaps to our way of thinking – it is an inheritance from Aristotle who distinguished *substance*, the root of constancy in the universe, from *form* which is changeable. If a form persists – for example the form of a piece of sculpture – that is thanks to the persistence of the material on which that form is impressed (so the story went).

Clearly, though, the persistence over a billion years of, say, cytochrome *c* is not like this. Even on the much shorter time scales of an individual life, molecules do not persist as such – they are being continually remade. It is not cytochrome *c* that persists within organisms and between generations of organisms; it is the ability to make cytochrome *c* molecules. That depends on other molecules – in the end on DNA molecules. But neither do these persist in the long term between generations in the kind of way that the Venus de Milo has persisted through the centuries. DNA molecules too are remade. What is inherited in the long term – indeed perhaps all that is now inherited in the long term – are particular *forms* of DNA molecules, particular base sequences. Only genetic information provides the long-term lines of continuity through evolution, to allow the repeated reappearance of phenotype structures – such as cytochrome *c*. Here it is form that lasts, not substance.

If we ask why certain phenotype forms now recur so reliably, the answer is to be found in the highly co-operative character of the machinery of which these forms are a part. As we discussed, such co-operativeness, and hence the invariances that it imposes, can only be a *product* of evolution. There is no reason to expect that this form-invariance was original.

But perhaps DNA itself is an exception? One might suppose that if it is only genetic information that provides the long lines of continuity in evolution then at least the kind of material that holds that information must be invariant. (RNA, say, might have changed to DNA because these can read each other's messages, but you might suppose that really radical changes would be impossible.) That might be so if the evolution of a genetic material was restricted to successive small modifications to that material – if takeovers could be discounted. Then indeed the evolution of a genetic material would be particularly hemmed in. But if, as we suspect, those now fixed central strands of biochemistry were not always fixed (and not always central) then takeovers could have been a part of the story of

the invention of our biochemistry as they have been part of the story of much later evolution. In particular takeovers of the genetic material, through organisms using more than one kind of genetic material, might have provided mechanisms for change that would be no more restricted in principle than takeovers of other kinds: a **genetic takeover** might be between wholly different kinds of materials. In Part II we will use this idea to shift our view of the problem of the origin of life.

# PART II

# A change of view

Figure 4.1. A formalised scheme for a genetic takeover. $G_1$s are sets of genes – initially 'naked genes' – that evolve to control their immediate environment: they make surrounding phenotypes. These are symbolised as protective boxes which become more substantial and elaborate. $G_2$s are another class of genes – secondary genes – that can only work within a rather sophisticated box: but in that context they are a more efficient type. The secondary genes at first assist the primary genes in a minor way. Later they take over the means of production of phenotypes which have become transformed in the meantime to the exclusive service of the secondary genes. Letters A–E denote stages referred to in the text.

# 4

# Genetic takeover

In the last chapter we came to the conclusion that, most probably, the components of our central biochemistry are 'high-technology' replacements of earlier designs – ultimately of 'zero-technology' designs with which evolution must have started. Takeovers, analogous to those visible in more recent evolution, could have provided the means for such redesigning; and, as we saw, the multiple interdependence of the present set of subsystems is itself evidence that here we are looking at the outcome of an evolution in which takeovers played a part. This is consistent with the conclusion that we had arrived at already in Chapter 2, that while a genetic material is a prerequisite for evolution through natural selection, nucleic acid could not have been the first genetic material. The question now is whether takeovers could have operated also at the very centre as a means of radically changing genetic materials. Is a genetic takeover possible?

An overall mechanism for a genetic takeover is suggested in figure 4.1. Referring to ideas developed in the last chapter I have here represented the earliest systems, A, as 'naked genes'. The evolution of such organisms would be expected to proceed nevertheless through the elaboration of phenotypes (as in systems B) in branching lines occupying various niches. In particular, in more difficult niches more elaborate phenotypes would be necessary go-between structures.

*A sketch* I will leave until later in the book speculations as to workings of primary organisms – that is, organisms based on primary ('self-starting') genetic materials. These speculations will depend on inorganic crystal chemistry to be developed in Chapter 6. And I will leave also till later some of the more detailed speculations as to how our biochemical system might have caught on. In the meantime here is a rough speculative sketch, based on the scheme in figure 4.1.

We must suppose that within the phenotypes of some of the evolving primary organisms (by stage B) metabolic pathways were being gradually built up that created consistent supplies of new molecules of various sorts with functions relevant to these primary organisms. In particular, new kinds of organic polymers were invented by some of the more highly evolved species. The original functions for such molecules would not have depended on particularly accurate specification – they were used for such things as glues, gels or fibres perhaps. Nevertheless these functions were often improved by improved specification. From time to time such polymers turned out to be preadapted to other functions; sometimes to more sophisticated functions so that new selection pressures arose for more precisely made, and often more complex, materials. To be a good structural polymer, for example, it would often be important for the monomers to be uniform and regularly arranged – as in, say, cellulose – and if these monomers were chiral then the chirality too should be controlled.

Now we might suppose that among the polymers being used for rather sophisticated structural purposes there was a class that contained a hydrophilic backbone to which were attached more hydrophobic groups. The structure-forming technique depended on these more hydrophobic side chains tending to cluster together. Under selection pressures to define these structures more accurately, more specific associations of the side groups were encouraged through modifications that allowed hydrogen bonding between them. Eventually something like RNA had been made because of useful effects, gradually discovered, resulting simply from its ability to twist into defined objects. The base sequences in these RNA-like molecules were perhaps rather simple – say stretches of poly(A) and poly(U) – and the information for specifying them was contained in the primary genetic material.

The genetic function of this RNA-like polymer was a fortuitous preadaptation: because it so happens that one way to make a structure-forming polymer is to have side groups on the polymer such that they key with each other – and that is also part of the design for a replicating molecule. But this functional ambivalence could only have appeared at a late stage in the evolution of organisms that had solved all sorts of much more fundamental problems: how to make ribose and distinguish it from its isomers; how to bring about energetically uphill condensations, and much more besides. Such a concatenation of expertise would be inconceivable except in an evolved organism. But to arrive at this new kind of replicating molecule the primary organisms need in no sense have had that goal in view. Each of the 'technologies' required would have been of a general usefulness – providing graded selection pressures for their evolution.



Even the replication of a molecule might have had a purely phenotypic role to begin with. For example, for some structural purpose it might have been important that stretches of different RNA strands locked with each other. It would not matter then what the sequences were so long as the sequence in one strand was the complement of whatever sequence happened to be in the other: and the way to ensure that would be by forming one strand on the other.

By supposing that at first replication had a phenotypic role, we can perhaps overcome one of the great problems of imagining the genesis of a genetic material: that unless replication is pretty accurate it is no use, information being quickly lost between generations. Continuing our speculations, if the phenotypic structures being made were fairly crude – for example for some sort of capsule – the replicative technique for matching up stretches could still have been useful even if it was somewhat haphazard. And there would be no information to lose. (The matched stretches, you remember, only had to be complementary – they did not have to be anything in particular.) But now there would be evolutionary pressures for improving the fidelity of replication – to make less crude products. Sooner or later, following the evolution of efficient catalysts for assisting replication, particular sequence information would be being passed on indefinitely between generations. This would allow much more intricate intermolecular and intramolecular twist structures to reappear in successive generations. (This idea will be elaborated further in Chapter 9.)

By now the organisms would be genetically bimodal (i.e. at stage C in figure 4.1) since part of the information needed to specify the common phenotype – that is, the information specifying the twisting of RNA molecules – was now being transmitted directly between these molecules. Given that these RNA structures remained useful to the organisms, then mutations would be selected – in either of the genetic materials – that tended to improve the conditions for this new technique of RNA polymerisation. In particular the 'replicase' catalysts would be improved.

If, perhaps, nucleic acid first got a toe-hold in some evolved primary organism through its structure-forming prowess, I doubt whether that would have been sufficient in itself to complete a takeover. Continuing our story, we might now suppose that twisted RNA structures came to interact with amino acids. The amino acids did not come from an external soup – which would have been far too messy – rather they were metabolites of the organisms and their synthesis had been perfected previously over long periods. They were there because they were useful as such – they were metal chelators perhaps. Dipeptides might also have been similarly useful and hence any technique for their synthesis. Among the techniques chanced

on was one that depended to some extent on twisted structures made by RNA. The RNA molecules only had to help here, to provide some specificity. They formed structures, perhaps, in which more or less specific pairs of activated amino acids were orientated for reaction. Under new selection pressures created by this and by the usefulness of dipeptides (and then, later, higher peptides) the RNA-controlled technique of peptide synthesis was perfected, until eventually a Mark-I protein synthesiser emerged.

However it happened, the invention of the means of specifying long polypeptide sequences would have been a long haul, and surely only possible within an already quite highly evolved organism. Even when this was complete, the primary genetic machinery must still have been there. But now the breakthrough had been made, perhaps, to a new technology, to an altogether faster and more accurate way of making structures and controlling organic reactions. The protein revolution was imminent (i.e. C → D → E in figure 4.1). One by one reactions could now be taken over by catalysts made from this new material – until eventually all the reactions needed to make the parts of the new system were under the control of this system itself. Then, only then, was the necessarily highly evolved 'scaffolding' kicked away to leave an 'arch' of subsystems to surprise us.

*Two points in particular* Extracting two salient points from the foregoing discussions we may note that $G_1$ and $G_2$ genes were both pre-adapted to their genetic functions – but under very different circumstances. $G_1$ was a mineral of some sort – it was necessarily a geochemical product. Its preadaptation was wholly fortuitous. $G_2$ was a biological product: its preadaptation depended on their being common design elements in polymers with genetic and other simpler, but biologically relevant functions. In view of these very different circumstances of birth you might have expected $G_1$ and $G_2$ to have been structurally very different. In any case the mechanism of genetic takeover allows them to be different *because $G_2$ appears, not from a modification of $G_1$, but from molecules that $G_1$ is controlling.*

Whatever the details, $G_1$ (or for more complex cases *some* pre-nucleic acid genetic material(s)) had to support an extended evolution during which the 'technologies' were established that were eventually to allow the synthesis of (at least) nucleic acid molecules. Genetic takeover thus allows specifically what in any case seems to make sense that *our central biochemical machinery was invented by quite highly evolved organisms,* by systems that would have clearly conformed to both the genetic and the teleonomic definitions of life discussed at the start of the last chapter. Whatever else,



we know this much about $G_1$: it must have been capable (eventually) of holding and replicating considerable amounts of information.

### Objections to genetic takeover

I will devote the rest of this chapter to considering objections to the idea of genetic takeover as they might be raised by someone who has picked up the general suggestion that evolution started with some kind of genetic material radically different from DNA and then changed over. We might imagine a coffee room discussion between this critic and an advocate of genetic takeover.

*Dr Kritic:* I can't say I think much of this idea that our kind of life took over from some earlier kind. It seems to me to have all the qualities of a bad theory – it is unnecessary, speculative, vague, complicated, defensive, not fitting the known facts, and, anyway, impossible.

*Dr Advo:* But apart from that it's O.K.? May I pick you up first on the *unnecessary* charge? You have to be very satisfied with current ideas on anything to say that alternatives are unnecessary. Can we agree that we are looking for the most probable general sequence of events in the light of the evidence available?

*Dr Kritic:* Sure. But genetic takeover cannot be part of such a sequence: it is *impossible.* You see, evolution is concerned with the elaboration of genetic information – in the end that is what it consists of: DNA molecules containing messages that contrive, by specifying phenotypes, to bring about their own propagation. But the whole thing keeps going only because a message in one DNA molecule can get copied into a new one. That would not be possible between radically different kinds of genetic material. Also the reading equipment, by which genetic messages are translated into an effective phenotype, would only work with one or a very restricted range of genetic materials. With a change to a new genetic material all the information in the old one would become illegible – it would be lost. That's not really a takeover: it's simply a waste of time.

*Dr Advo:* Nature can be enormously wasteful: she has certainly wasted time during evolution if you like to think of it that way – laboriously evolving one structure only to have it displaced by others. Yet often this is an essential part of invention. (Darwin's central idea should really have been called Natural Rejection since mostly this is what happens. And the 'amount of organisation' in an organism (whatever that means) is anyway related to the size of the pile of design modifications rejected on the way.) We might ask whether a life based on a non-nucleic acid genetic system

could have provided some sort of booster rocket to get nucleic acid life off the ground. Then, even if the booster later falls away, it may have been part of the evolutionary history of the winning system without which that history cannot properly be understood.

*Dr Kritic:* Mmmmmm.

*Dr Advo:* Now you have been assuming that the only help that one kind of life could be for another would be to provide it with genetic information. That is not so; at least not in the sense of the first kind of life handing over messages in its own admittedly illegible script.

Think about it this way. The optimal design for an organism depends on what the environment is like. That would have been true at the most fundamental levels to begin with. The environment must provide the units out of which the first genetic material is made. First life is thus limited by what primitive geochemistry can do: it starts the best way it can with some genetic material that can put itself together. Once this life evolves, however, it changes the available environments. Its evolved phenotypes – possibly a variety of them in different niches – now provide a range of new opportunities for new kinds of life; if only by providing consistent supplies of particular complex molecules not previously available, or only available in very impure form. It is no novelty for one set of organisms to create the possibility for another set of organisms (plants made animals possible, for example). A first life system might well increase the probability of a completely new system simply by providing new kinds of foods. Suppose that it made monomers for an alternative genetic material . . .

*Dr Kritic:* Are you trying to say that a primary life would have evolved the ability to make nucleotides? What selection pressures, may I ask, would have brought about that remarkable circumstance?

*Dr Advo:* I will avoid asking you in return what selection pressures would have caused the primitive Earth to make nucleotides, since that might sound like bickering: but really it is you, not me, that is falling into the anthropomorphic trap. Primary life, as you so aptly call it, would not have been in any sense looking for new genetic units, it would simply have been evolving the ability to do organic chemistry because that is a generally useful ability. As it evolved this competence it would happen on a great variety of molecules – for making phenotypic polymers and so on. This would increase the chances, over the chances that had been there for the unevolved Earth, that some alternative units would become available for some alternative genetic material. Maybe several were produced in this way but, among them, one species of the more primitive life discovered nucleotides. It was subsequent events that were to make that discovery so significant for us.



*Dr Kritic:* Let me see then, so primary life made nucleotides and there was this secondary life waiting in the sidelines to gobble them up?

*Dr Advo:* Not exactly. I am asking you in the meantime to think along analogous lines to those who believe that nucleic acid molecules evolved on the primitive Earth: the only difference is in the source of the units.

*Dr Kritic:* But the trouble with ideas that depend on nucleic acid molecules forming on the primitive Earth is not only that the nucleotides might be hard to come by, but that a polymerase substitute would have been needed to make them join up and copy pre-existing sequences. That problem would still be there?

*Dr Advo:* Yes – according to this line of thought.

*Dr Kritic:* But there are now new problems. This source is not a general environmental pool, but a set of micro-environments provided by the set of individuals of some particular species or group. That would drastically reduce the available places where a second kind of life could catch on; also it would be very dependent on primary life. And how would the nucleotides get out of the primary organisms? How would primary life hand them over, and how would secondary life pick them up?

*Dr Advo:* Secondary life would need nucleotides before it could *begin*. (It might be molecules like nucleotides, but let's say nucleotides for the sake of argument.)

*Dr Kritic:* I see; this is the 'naked nucleic acid gene' idea; but that was never a very good one. Even given activated nucleotides and a pre-existing nucleic acid strand, this strand would not by itself select the appropriate new monomers – single nucleotides do not bind at all well to nucleic acid strands. Even if they were emplaced and polymerised before coming away again as a completed complementary strand, how would this sequence of events be controlled? Even if somehow it was, all you would have would be replicating nucleic acid molecules unable to do anything except replicate.

*Dr Advo:* Much of what you say is true, and it suggests that the role of primary life was not simply as a provider of suitable molecules. But you make two mistakes. First when you complain that secondary life would depend on primary life. Well, it would to begin with, and perhaps the amounts of the key molecules being made by primary life would be small. But that wouldn't matter so long as the supplies were consistent and clean. It can be said in favour of a two-stage take-off that primary life does not build up stock-piles of essential nutrients for secondary life – it creates a means of production. The problems inherent in stock-piling molecules are thus largely removed. Your second mistake is in supposing that replicating nucleic acid molecules could not evolve: in the right circumstances they

could – as Spiegelman has shown. But I would agree that there is difficulty in imagining the circumstances under which 'naked' nucleic acid molecules would replicate.

*Dr Kritic:* Not the only difficulty by any means: to become independent eventually, secondary life has to learn to make those molecular bits that primary life had provided for it. Now how is it going to do that? It can't read the genetic messages in primary life.

*Dr Advo:* Something like Horowitz's mechanism might work (figure 1.12): not only some vital product like a nucleotide is provided, let us say, but also the immediate precursor metabolite. A secondary organism that can use that too will have an advantage, and it will be able to use it if it evolves the means of catalysing the last step. Similarly all the way back, step by step, it copies the metabolic route in the primitive organism. It never has to read the genetic messages: the metabolic intermediates have told it how to go.

*Dr Kritic:* Very neat. But the trouble with this sort of mechanism is that many metabolic intermediates are unstable: the question of how the goods got across (which you never answered incidentally) would be very acute here. Still worse, this whole scheme presupposes an ability in the secondary organisms to evolve a variety of catalysts at will (as it were). How is it going to do that until it has invented protein? And what selection pressures are going to lead to the invention of protein when the key function of specific catalysis could not be expected to emerge until a late stage? In any case the invention of protein is going to call for some very well made catalysts. I see only two conceivable ways of getting out of this: either nucleic acids themselves can be catalysts – RNA molecules, say, twisting up into objects with catalytic grooves in them – or an obliging primary organism makes and hands over the required catalysts. I suppose that twisted tRNA-like molecules might do, but I cannot see them having the kind of universal ability to evolve a given catalytic function 'on demand' that the Horowitz mechanism requires. And as for primitive organisms making and handing over such things as tRNA activating catalysts – well, why should they? Anyway you admitted that the whole idea had fallen through with the need to start from replicating nucleic acid molecules. By making a consistent supply of nucleotides, a primary life might have made this less impossible, but nucleotide supply was not the only problem: the idea is still impossible.

*Dr Advo:* I agree.

*Dr Kritic:* Then I rest my case.

*Dr Advo:* What case?



*Dr Kritic:* The case against genetic takeover.

*Dr Advo:* But we haven't been talking about genetic takeover.

*Dr Kritic:* What?

*Dr Advo:* No, all I was wanting to do was to show that by providing supplies of molecules not previously available a primary life could at least make the origin of our system 'less impossible'. I was also wanting you to clarify some of the remaining problems: it helps one to see how to adjust the hypothesis towards something that might actually work. The necessary adjustment, I think, is to go from the kind of saprophytic idea that we have been discussing to something more like an endosymbiotic one: to suppose that the secondary genetic systems started *within* primary organisms. (*Dr Advo draws figure 4.1 on the back of an envelope and explains the idea in more detail.*)

*Dr Kritic:* Well I can see now how the primary system could transfer metabolites to the secondary system, but your account of the evolution of protein synthesis was sketchy to say the least; and you are still supposing that accurate nucleic acid replication was achieved long before protein enzymes would have been possible. Is this really much of an advance?

*Dr Advo:* It is not just that with secondary genes evolving within the same organism as primary genes there would be an easy means of transferring nucleotides. There would also be a reason for maintaining supplies. Even before the new genetic material started working, nucleotides, and also nucleic acid-like polymers, were useful to the primary organisms, and the means of production of these molecules was maintained only for that reason. Similarly, this newly discovered trick of direct information transfer between the nucleic acid-like molecules would be continued only if it was useful to the organism as a whole. That is the *sine qua non* of any further progress: but, given that, there would be no limit in principle to the assistance that the evolved and evolving 'technologies' of (a still mainly primary) organism might provide for the younger partner – or should I say younger part. There is no need to imagine any sort of genetic altruism: these primary genes would have been as selfish as any other kind; but it would have been in their short term interest to help the secondary genes if the secondary genes in any way helped them. You could say the same about any two genes cohabiting any organism. Of course nowadays all genes are made of the same stuff and work the same way – organisms now are 'homogenetic': but I see no reason why there should not be 'heterogenetic' organisms. Indeed the account I gave of the evolution of protein synthesis was sketchy – I do not know, for example, how the genetic code evolved. I could speculate on that if you like or refer you to several interesting

speculations on this question in the literature. (To be discussed in Chapter 9.) But that is not the main issue here. All such speculations that I have come across are evolutionary – they talk of the gradual perfection of this and that subsystem. But there is only one engine for the evolution of ingenious competence that I know of and that is natural selection. To evolve, the subsystems have to be part of an organism of some sort. Now there might be no need to postulate an earlier kind of life if some minimum nucleic acid–protein system could be conceived of as having formed spontaneously on the primitive Earth. But I do not see such a system as conceivable. You say yourself that naked nucleic acid genes are no good, and anything else would be more complicated – nucleic acid plus something else. I see no alternative to postulating some other kind of starter life to provide the milieu within which our kind of life system began its evolution.

*Dr Kritic:* But do we not have the same problems all over again in trying to understand how starter life got off the ground?

*Dr Advo:* Oh, we have a whole lot of problems: but they are not the *same* problems. There is nothing to say that starter life relied on an interdependent set of complex subsystems, as our life does. Indeed one of the things that we can definitely say about primary organisms is that they were not built that way.

*Dr Kritic:* If they were that simple how did they come to be so good at doing organic chemistry (you say in effect that they became nucleotide chemists, for goodness sake).

*Dr Advo:* There is a distinction to be made between having to be complicated before you can start to evolve, and being able to become complicated through evolution. But I agree there is a whole nest of problems here. There must have been at least one pre-nucleic acid genetic material with considerable information capacity, and there must have been some way in which this information could specify organic reactions. I would see these questions as constituting the main part of the problem of the origin of life: what was the primary genetic material and how did it work?

*Dr Kritic:* I'm not so sure. Even if there were such 'self-starting' organisms I doubt whether 'heterogenetic' organisms would ever catch on. They sound ungainly beasts to me. Two different genetic systems in the same organism? A recipe for confusion.

*Dr Advo:* People usually try to think of an alternative genetic material as being something like the one that they know; but the takeover mechanism does not require primary and secondary genes to be similar, and indeed it might be very much better if they weren't, for the reason you mention. If the genetic control systems were very different chemically there might be no confusion. But I daresay a genetically bimodal organism is not the ideal type.



*Dr Kritic:* Why then would they have appeared?

*Dr Advo:* Because a primary organism is even less likely to be the last word in biochemical competence. The primary genetic material was surely a nasty compromise between what would replicate well enough and what units were around. New genetic materials might well have been a good idea when they became possible, because of limitations on what the primary material could do. Working through protein our system is very versatile, but before that Jack-of-all-trades appeared there might have been a case for having different means for making different kinds of phenotype structures.

*Dr Kritic:* O.K. I will concede that genetic takeover is possible logically; even that some earlier different kind of life might have been relevant to the early evolution of our life. But this whole idea is so *speculative*. You seem to think that the world is full of genetic materials – actual ones on the primitive Earth, and then potential ones for subsequent evolution to stumble on. Genetic materials don't grow on trees you know. Within biochemistry there are only two and they are very alike. And there are no known replicating molecules outside biochemistry – no one has ever made one.

*Dr Advo:* Well actually I think that secondary genetic materials *did* grow on trees – evolutionary ones. I think that primitive life acted in effect as a *search device* for secondary genetic materials – by setting up the means of production of various different sets of small molecules in various evolved lines; by developing generalised techniques for organic synthesis; by creating uses for various polymers and the means of producing them; and then later by making higher order structures that depended on precise lining up between polymers. All such evolutionary advances would have edged towards the discovery of a new genetic material. As for speculative, surely this whole subject is speculative?

*Dr Kritic:* Yes, but genetic takeover is much more so: and you admit it is *vague*?

*Dr Advo:* What I have said so far is. But genetic takeover is meant to be a background hypothesis. It is meant to be vague – in the sense that it is a general frame within which to make more specific speculations. Chemical evolution is like that as well: it says that matter formed into atoms, then small molecules and then bigger ones; that these then came together to make higher order structures; that some of these could reproduce, and so on. Put at its baldest it's pretty vague. But I would not criticise it on that

count. The question is whether the frame is correct – can it hold more specific speculations that are at the same time plausible? Now I do not think that there is any picture of the origin of life that fits the frame of chemical evolution. On the other hand, the genetic . . .

*Dr Kritic:* But the genetic takeover frame is more *complicated*, and that's always bad for a hypothesis. If you make a hypothesis complicated enough it can be made to fit anything.

*Dr Advo:* If it is more complicated it is less vague – but let that pass. Life is complicated – our life anyway: a simple explanation for its evolution would be bound to be wrong. Not that the frame itself is much more complicated (is figure 4.1 very complicated?). It is only the picture that is bound to be complicated. My criticism of the chemical evolution frame is that it does not easily let in a certain kind of complexity that should be in the picture. I mean those evolutionary processes that gave rise to the functional interdependence of the subsystems, and led to the choice of such a non-primitive-looking genetic material as DNA.

*Dr Kritic:* Maybe; but there is an openness about chemical evolution and it *fits the facts*: maybe nucleotides were not in the primitive oceans but Miller's experiment, and others like it, have shown that many of our biochemicals could have been there. That cannot be a coincidence.

*Dr Advo:* No, I'm sure it is not a coincidence that our biochemistry uses many easily made molecules. But if you remember that biochemistry is a kind of chemistry you should not be so surprised at that. The question is *when* the choice of the particular set of molecules now common to life was made. If it was made wholly before life started then all the present basic parts should be easy to make. If on the other hand the choice of biochemicals was made at some later stage of evolution, as part of the invention of a secondary genetic system and its means of control, then you might expect only some of the component molecules to be easy to make: because by that time the already well evolved organisms had the means to make complex molecules if need be. But there would always be a preference, nevertheless, for easily made molecules if these would do. Well, what are the facts? Our biochemistry is made from some basic molecules that are easy to make (amino acids, sugars) and others that are not (nucleotides, lipids). Genetic takeover fits the facts better.

*Dr Kritic:* Yes, if you insist on either a 'wholly before' or a 'wholly after' origin for our biochemicals. But no chemical evolutionist would insist that every biochemical was premade. Some, I daresay, did come in later.

*Dr Advo:* But the late starters include two of the most important types. And, anyhow, who is it that is making *ad hoc* hypotheses now?



*Dr Kritic:* You are, for goodness sake! I have never heard of a greater bit of ad hoccery than this primary life that you talk of. Even that I might not mind if the whole idea was not so obscure. *Defensive* is perhaps the best word: failing signally to put its neck out. The explanation for our biochemistry is to be put out of sight, is it? Neither where it can be checked by experiments simulating primitive Earth conditions nor by extrapolation backwards of the life-system that we have direct knowledge of. The whole secret is hidden in some long lost 'self-starting' ancestor. And all you can tell me about it is that it might have been 'very different'. Your idea of an initial booster rocket that fell away is appealing but useless. Dimly we see the preparations for the launching of the rocket: but the launch itself takes place in a mist that gets thicker, if anything, as the rocket rises. When the mist clears we find that the rocket is powered by a motor that we are familiar with, but you tell me that this was not the original motor – which fell away before the clouds dispersed?

*Dr Advo:* Roughly speaking yes, but the rocket that emerged was not there at all on the launch pad: it was actually assembled in the clouds (I think this analogy is feeling the strain).

*Dr Kritic:* But you see what I mean by obscure and defensive: all the interesting stuff is out of sight – so as not to hamper speculation. Can you think of an experiment that could conceivably disprove genetic takeover? That is the real test: if there is no conceivable way in which a hypo . . .

*Dr Advo:* Oh come off it. You know how hard it is to disprove background hypotheses of any sort. 'Pictures' can easily be disproved perhaps, but it is more difficult to disprove a 'frame' – to show that there is no 'picture' that could fit it. The best you can hope for are considerations that will either increase or decrease their plausibility. Taken together such considerations can amount to virtual disproof. But you will be lucky to find one neat experiment. So my answer to your question is, no, I cannot think of a single experiment that could disprove genetic takeover any more than I can think of one that would disprove chemical evolution. But I think chemical evolution has become less likely as a result of experiments, observations and ideas from a number of fields. (Discussed in detail in Chapters 1 and 2.) The same could very well happen to genetic takeover: for example, if life that was closely similar to ours was detected elsewhere in the Universe, or if a one-step simulation experiment generated nucleotides, or if in spite of determined searching no simpler kind of genetic material than nucleic acid was found experimentally; or if it could be shown theoretically that nothing simpler than nucleic acid could work.

As for the charge of deliberate obscurity, it is the subject that is obscure

– or rather remote (although I confess to finding remote subjects interesting). Certainly if you shy away from remote subjects they will remain that way. The corpuscular theory of matter of Newton's time was far from sterile even if, at that time, it was hard to get experimental confirmation of the idea that matter is made up of particles too tiny to see. That was an obscure idea, about remote entities, that I have no doubt raised sarcastic comment at the time. But if you are prepared to live with a plausible but unproved idea, formulating experiments and further speculations on the provisional supposition that it is true, then its plausibility may grow or shrink. Of course you do need data, as directly relevant as possible. But even the most indirect data and the most tentative inferences can add up.

Now the evolution of our biochemistry is not by any means lacking in relevant data, even if direct evidence is lacking. We know much about what atoms and molecules can do and this places limits on conceivable or likely hardware for any kind of life, given also what we can deduce as to minimum conditions for evolution. We may still be very uncertain about conditions on the early Earth, but we can already place limits, and these should get tighter with the present activity in planetary and Precambrian research. We should then be able to draw up a general specification for a first organism on the basis that it had to appear spontaneously and evolve on the early Earth. And the abiotic syntheses of biochemicals are relevant here too. One would like as complete a list as possible of all the kinds of molecule that are easy to make from materials and energy sources that might have been available to early organisms. Such lists are being built up. Nor is the hi-fi system that finally emerged irrelevant to that starter life that we seek. The original genetic control apparatus had to come to control organic reactions, as I have said, but of course in particular it had to be able to handle at least some of our now central biochemicals. And then the general way in which our biochemistry is put together – what leans on what – can help us, perhaps, to see a sequence for its original construction; and in the details of its interdependences let us see, perhaps, just where things went missing and what sorts of things they must have been.

No, there are plenty of data: if genetic takeover seems more obscure than 'straight' chemical evolution that is because it has not been thought about enough, and experiments have not been designed with that possibility in mind. You scorn the thought that the secret of the origin of life is in a first link that is now missing and was made differently from us. Perhaps you feel that the rules of the game demand that we stick to the pieces that we can see. But that rule is not in the parsimony of Nature, it is in the paucity of the imagination.



*Dr Kritic:* All right, suppose that you have put two and two and two together and from these various considerations you have managed to come up with what you think is the answer. How are you going to know? What sort of experimental test would be possible? How do you recognise an unevolved primary organism if you found one – or made one? If you don't have a million years to spare how do you know that it can evolve at all, never mind that it is going to invent nucleic acid? It's not the paucity of the imagination, it's the paucity of experimental approaches that worries me.

*Dr Advo:* That sounds so like the sceptics of atomism who could see no way of verifying the totally invisible; or the ancients for whom the stuff of the stars was unknowable; or Mendel himself who did not believe that the material basis of heredity would ever be discovered. Not being able to see experimental approaches is part of the paucity of the imagination. First we would have to put two and two and two together. Then we might see.

*Dr Kritic:* Go ahead then.

*Dr Advo:* It might take half a book . . .

# 5

# On the nature of primary genetic materials

There seems to be little in the last resort that a gene can be but a fragment of a specific macromolecular pattern, capable of being copied in growth by the laying down of fresh molecules in conformity with the same design. *C. N. Hinshelwood (1951)*

## Introduction

In the last chapter, we distinguished between primary genetic materials, which can arise without a previous biological evolution, and secondary genetic materials which can only appear within already evolved organisms. It seemed clear enough that nucleic acid was a secondary genetic material, in which case there must have been at least one truly primary genetic material in our evolutionary history – and at least one such material in the world to be found. In this chapter, we start looking.

If genetic takeover obscures the view of first life from the standpoint of present-day biochemistry, it by no means sets us adrift. The design specifications for any system able to evolve indefinitely under natural selection are quite restrictive even if these specifications do not tell us immediately what kinds of materials must be involved. Indeed, the specifications for a primary organism – one that can evolve from scratch – may seem at first to be so restrictive as to exclude everything that is physically plausible. Our approach here will be to take full advantage of the freedom in choosing initial materials that genetic takeover provides, and try to find *anything* that might have been able to start to evolve under natural selection.

The nature of primary genetic material(s) is at the centre of this problem. We will consider first the possibility that the material(s) we seek might have been something like DNA, and then pursue a line of thought that leads away from that, away from linear encoding of information, away from the



organic chemist's idea of a polymer – and indeed, away from organic chemistry altogether. Even through such a journey, the essential requirements for a genetic material can be retained. In Chapters 6 and 7, we will become concerned mainly with inorganic crystal chemistry and mineralogy, since, as I will try to show, it is during mineral precipitation processes that primary genetic and other control structures would be most likely to appear. In Chapter 8, we will discuss how an evolving mineral-based life could 'learn to do organic chemistry', evolving to the point at which metabolic pathways developed – some of these perhaps similar to our present central metabolic systems. Chapter 9 will be concerned with the appearance of organic genetic polymers within some advanced primary organisms; with the subsequent evolution of our nucleic acid–protein system; and with the eventual establishment of this system as the basis of the only life form on Earth.

### *Genes and chemostats*

Following our discussions in Chapters 2 and 3, we can guess quite a bit about our ultimate ancestor. It was able to reproduce in a hereditary manner. It belonged to a very large set of possible structures related to each other by minor modifications – there was evolutionary landscape open to exploration. It was made from units that were available in the primitive environment. Unlike us, it could be explained without reference to pre-existing genetic information. Its hereditary machinery was based on structures that could template-replicate fairly accurately from the start, or which could rapidly evolve accuracy of replication. This primary genetic material, although containing initially no genetic information, had a large potential capacity: it could contain any of a very large number of possible replicable patterns that could exert some control on their immediate environment so as directly or indirectly to affect their future prevalence – by increasing the efficiency either of replication or of the distribution of the products of replication. Also, we know something about where our ultimate ancestor lived. Like all organisms, it must have been an open system and hence, to have been maintained indefinitely, it must have lived in another open system – in a matter–energy stream of some sort.

Indeed, the easiest environment for life is the microbiologist's chemostat (figure 5.1) in which identical conditions are maintained, or at least recur reliably. It is this, not the petri dish, that should be our model for the ideal environment in which a delicate early life might thrive. An oceanic soup of organic molecules, built up in some previous era, looks too much like a

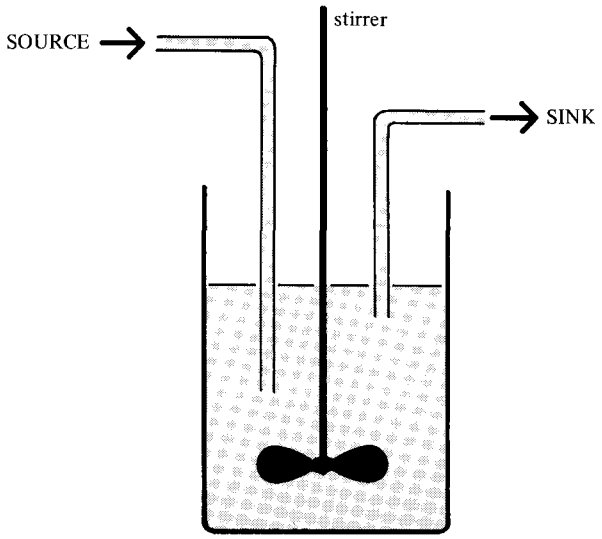



Figure 5.1. A chemostat provides constant conditions for the culture of micro-organisms. These are removed in suspension at the same rate as fresh nutrient solution is supplied.

gigantic petri dish; too liable to be altered by, and to the detriment of, any life that might appear in it. Organisms now are not set up like that: instead they are locked into geochemical and biological cycles which ensure the long-term maintenance of their environments. Indeed, every stable biological niche is a chemostat if you like to look at it in that way: there must be an endless supply of essential feedstock and an endless provision for disposing of waste products and corpses (see Bull, 1974).

Without carrying this discussion any further for the moment, let us formalise the ideal situation for those simplest kinds of evolving entities that we discussed in Chapter 4 – 'naked genes'. Such genes are represented by the black boxes in figure 5.2*a*. To operate, a black box must be provided with suitable component molecules – 'genetic units' – premade by the environment and out of which the box makes another like itself. There must be many ways in which the units can, in principle, be put together to make different boxes and, occasionally, in the manufacture of a new one, there must be a mistake in copying that allows for the exploration of this field of possibilities. Furthermore, some boxes must be, for various reasons, more efficient at having offspring than others. Such a system will evolve, with more efficient boxes emerging as time goes on. In practice, the environment must also provide other materials – if only solvent molecules – and to go



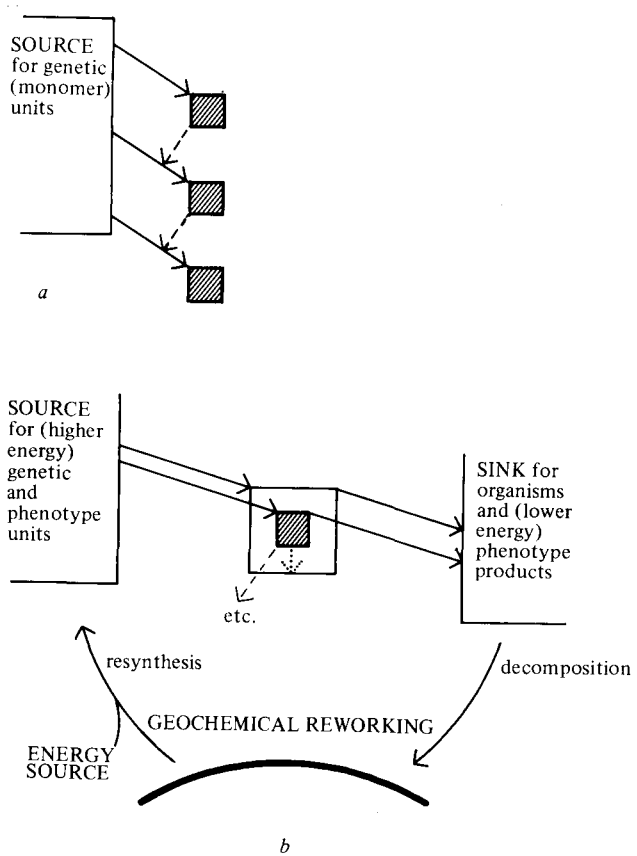



Figure 5.2. (*a*) The simplest evolving system would consist of 'naked genes' (hatched boxes) in an environment providing its monomer units (solid arrows). The synthesis of a new gene also requires a transfer of information from a pre-existing gene (dashed arrows). (*b*) Early organisms most probably lived in open systems, the environment maintaining sources of nutrient and sinks for waste products. These organisms would have consisted of genes whose immediate environment had been more or less modified as a result of information in the genes (dotted arrow). Such a modified environment would protect or assist the replication of the genes, that is, it would be a phenotype (outer white box).

very far, and in any case to set up a situation in which a genetic takeover could occur, the black box must transform its immediate environment as part of its technique of efficient survival and replication. The evolution of such 'clothed genes', i.e. genes + phenotype, will almost certainly call for other environmental molecular supplies. A minimum 'chemostat' that

would be necessary for the long-term evolution of such an organism, is illustrated in figure 5.2*b*.

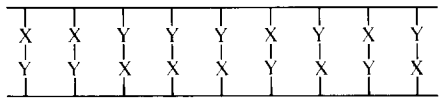
This chapter, and the next two chapters, will be an attempt to explicate figure 5.2.

### Some design principles for replicating molecules

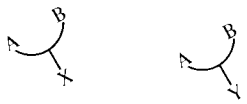
First, let us try to prise open that black box to see something of the machine inside. 'What is it made of?' may seem the most pressing question, but we must start nearer the known functional requirements with 'how did it work?' In particular, let us concentrate on its replicative function through the question: 'among physiochemical systems, what are the easiest imaginable microscopic information-replicating machines?'

We might seek a *self*-replicating polymer molecule; by which I mean a molecule that could hold information and replicate it without pre-evolved machinery. It should be necessary only to add copolymer with a given sequence of monomer units to a mixture of free monomers in solution for these monomers to link up, forming copies of the original copolymer sequence. This should go on with little or no help. There is no indication that DNA is like this – it seems to need a great deal of catalytic machinery, as we will be discussing in a few pages. But let us start with the assumption that the primitive genes we seek were molecules something like DNA.

The simplest DNA-like replicating molecule would be a binary ladder copolymer, for example:




made from trifunctional monomers:




such that groups A and B can form irreversible bonds with each other, while groups X and Y can associate reversibly. Figure 5.3 illustrates one possible scheme for the self-replication of such a molecule. The problem is to find a monomer pair with physical and chemical properties that conform to a set of design specifications. Let us consider some of the features that would be needed.



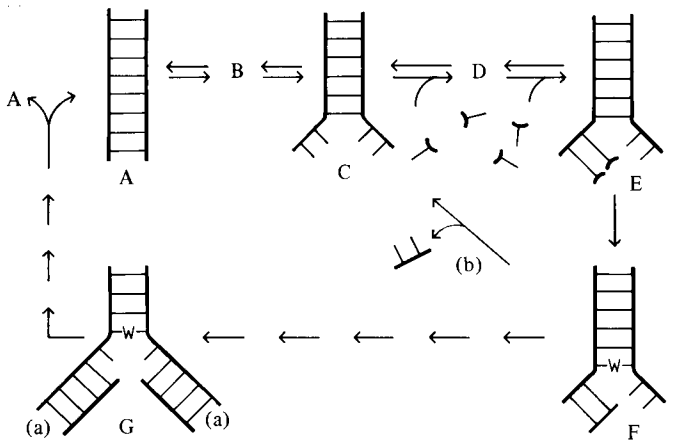



Figure 5.3. A scheme for a self-replicating ladder copolymer (A). Rung bonds must be sufficiently reversible for A occasionally to 'fray' at one end (B, C), allowing monomers to bind (D, E) and to polymerise irreversibly (E → F → G) so that the chains are forced apart (G → 2 × A). W represents a weakened rung, (a) partly made ladders and (b) a short single chain. (After Cairns-Smith & Davis, 1977.)

*1. Solubility*

Monomers, ladder and intermediates must be soluble: if the ladder is to be a primitive genetic material, then, presumably, soluble in water. It may be necessary to have solubilising groups (e.g. charged groups) for this purpose.

*2. Stereochemical feasibility*

(*a*) The three functional groups in each monomer must be so arranged that, with the A and B groups joined up to form a polymer chain, the side groups X and Y can point in such a direction that they can mate with the side groups of another similar chain to make a ladder. This constraint eliminates, among others, any very small molecules for the monomers – there must be at least three atoms on the main chain between the side groups if the latter are to be arranged parallel to each other on the same side of the chain. (In a double helical ladder, successive side groups are at an angle to each other but, in that case, the length of chain between them is longer.)

(*b*) The various intermediate structures, such as those in figure 5.3, must be stereochemically possible.

(*c*) All transition states should be (at least) stereochemically possible.

*3. Specificity*

(*a*) X and Y must either cross-pair or self-pair, but not both.

(*b*) A and B must be able to react quantitatively to form main chain bonds, but there must be no other irreversible reactions between A, B, X or Y, or any other (e.g. solubilising) groups.

(*c*) A and B must be relatively unreactive in the free monomers which should neither polymerise spontaneously in solution nor tend to extend existing chains.

*4. X–Y association equilibria*

Considering the particular scheme in figure 5.3, the various X–Y association constants must have suitable values, such that 'fraying' at one end of a ladder occurs occasionally to give structures such as C. Also, the concentration of free monomers must be high enough for structures such as E to form (monomers might tend to pair with each other). Such initiation steps must not happen too readily, however, or partly made ladders (e.g. (a)) would start to replicate prematurely leading, often, to incomplete new ladders. Also, if the rung-bonding is too weak, short single chains would be generated (b).

Ideally, the two end-rungs of A, and any monomer pairs, should have low association constants: those attaching the monomers in E should be higher. It would be useful if after the irreversible main bond-forming reaction, E → F, the next rung was weakened, due to stereochemical or electronic strain in the product F, so that the subsequent propagation processes, F → G → 2 × A, could be relatively fast, discouraging side reactions that would lead to partly formed chains.

*5. A–B reaction rates*

These, too, must be fast in the propagation stages to prevent side reactions (although inhibited in other circumstances – 3(*c*) above).

Alternatively, one may imagine self-replication schemes in which chains separate completely, as in formalised diagrams of DNA replication. But here an essential problem of self-replication is even more acute: that under the same conditions, rungs of ladders must tend to break while rungs between monomers and single strands tend to form. This is the opposite of what you would expect on entropy grounds. Also, the sequencing problem is still there: one stage must be inhibited until the previous stage is complete. Figure 5.4 illustrates a conceivable scheme of this kind.



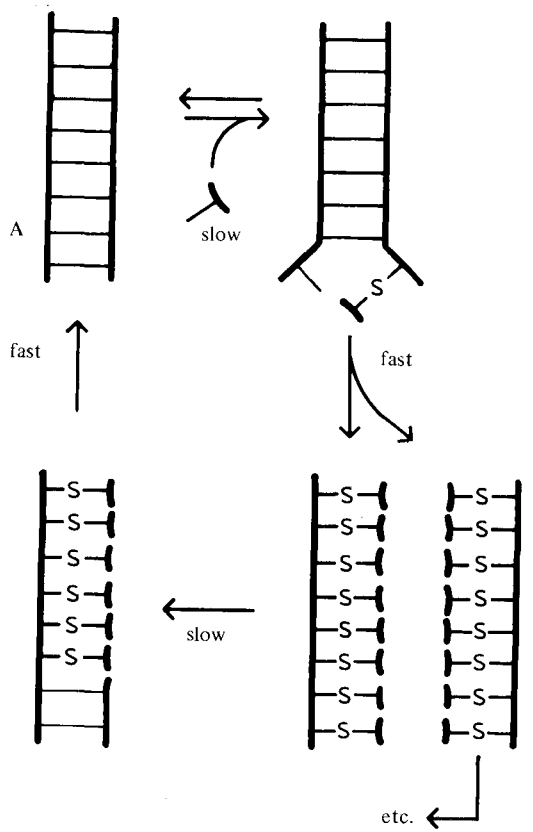
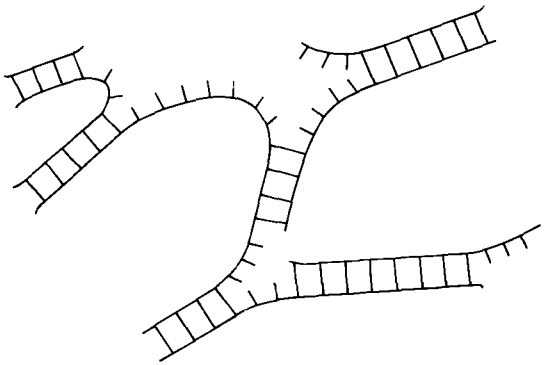


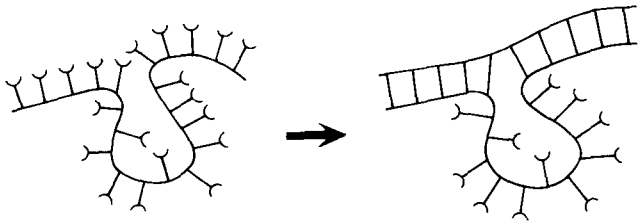

Figure 5.4. An alternative conceivable mechanism for a self-replicating ladder copolymer (cf. figure 5.3). Here the monomers themselves pull the ladder (A) apart. The monomers are present in excess and form particularly strong bonds (indicated by the letters S) with single strands. Slow and fast processes alternate as shown to maintain a correct sequence of stages and so prevent the formation of incomplete chains.

*6. Stereochemical controls*

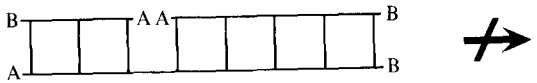
Under *2*, we discussed stereochemical features needed to allow appropriate processes: there are others that would be required to prevent inappropriate ones, even assuming the functional group specificities given in *3*. Mainly this new class of problems arises because soluble organic polymers are generally flexible. There would always be a danger of insoluble tangles:



And some of the attached monomers might be left out during the polymerisation stage; for example:



Then, again, the direction of polymerisation has to be constrained in some way. Otherwise, the formation of daughter chains might become blocked; for example:




*Chance and engineering*

We have discussed in all about twenty largely independent design considerations for a self-replicating molecule of an apparently rather simple type. None of these considerations is, by itself, particularly taxing: one can find water-soluble molecules, potential A–B pairs that link quantitatively and irreversibly with each other, even perhaps X–Y pairs that associate sufficiently specifically. And one could find an X that would not react with an A, or a Y that would not react with a B, or itself, and so on. The problem is to find a monomer pair that combines all twenty or more desirable features. The trouble with such molecular engineering is that one cannot adjust or machine molecules: if a bond angle or a van der Waals



radius or an activation energy barrier is inappropriate, that is just too bad – one must choose another molecule. The scope for engineering is thus firmly tied to the number of molecules available in principle to choose from.

Here is an analogous situation. Suppose you want to buy some object that can only be had ready-made – let us call it a snark. You go into a shop that sells snarks and ask the assistant if he has a green one. The shop has about 1000 of these things in stock and, generally speaking, about one in ten snarks that come in from the suppliers are green. The assistant is confident: 'Certainly sir, I will go and fetch one.' 'But, wait a minute,' you say, 'it must be the three-wheeled type.' Again, as it happens, one in ten, or so, snarks have three wheels. There should be no problem: the chances are that of the 1000 snarks in stock, about ten will be *both* green and have three wheels. But then you remember that it has to fit into the cupboard under the stairs – the snark must be no longer than 2 metres. Again, 10 % or so are the short kind – and you might be lucky. But you only have to add a few more conditions like these (each, perhaps, easy in itself) for the assistant to start pursing his lips and saying that no one has ever asked for *that* before.

To return to the question of monomers for self-replicating polymers. Suppose that you have a very large box of miscellaneous organic molecules. You sift through them and put aside all those that are soluble in water. You then select, from these, molecules that have a possible A-group. From these, you now select those with *also* a B-group that will allow polymers to form: and you proceed like this to smaller and smaller subsets by applying successively the various design constraints that we discussed. Given that some of these constraints depend on such features as activation energy barriers, bond angles, etc. being rather accurately suitable, you would be lucky if, on average, these siftings would leave you with more than one-tenth of your initial set at each stage. Even on this rather optimistic reckoning, to have a reasonable chance of being left with one molecule worth testing at the end of your search you would need about $10^{20}$ different ones to choose from to begin with. That is more than there are in *Chemical Abstracts* (by a factor of around $10^{14}$).

Such numbers can be out by factors of billions without affecting the main argument: a self-replicating polymer of the kind discussed would be difficult enough for an organic chemist to design with the whole field of organic molecules to choose from. As a primary genetic material, it is wholly implausible. Apart from tending to prefer stable and perhaps water-soluble molecules, I can see no reason why previtral organic reactions should have tended to converge on some particular combination of func-

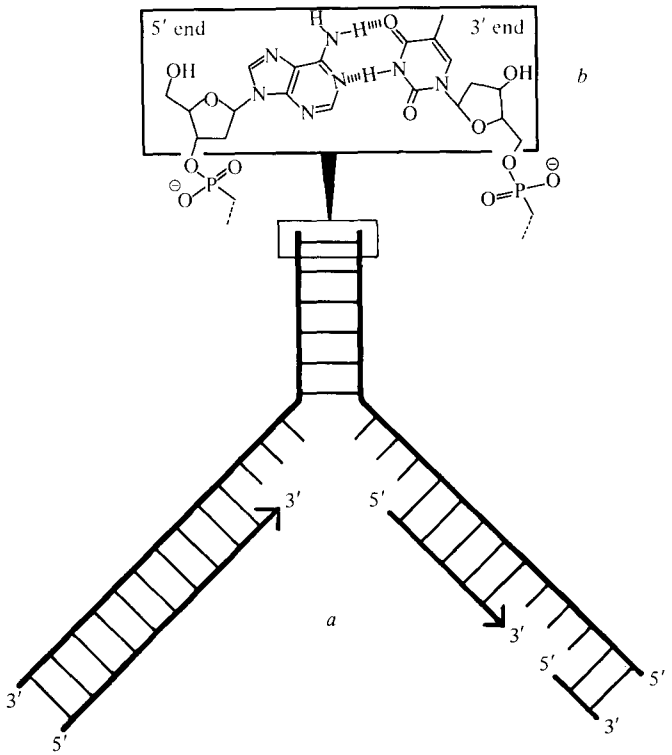



Figure 5.5. (*a*) During the replication of DNA the strands of the parent molecule are peeled apart. The single strands act as templates for new synthesis. But the strands run in opposite directions, and synthesis on the 5′-ending strand takes place discontinuously. (*b*) Closer view of 5′- and 3′-ends, showing in this case an adenine side group paired with a thymine.

tions required for such a molecular machine. To make amino acids is one thing, to make self-replicating organic ladder polymers – whether nucleic acid or anything else – is quite another.

### *Assisted replication of organic polymers: (a) catalytic assistance*

*A case in point* As I said earlier, there is no indication that DNA is a self-replicating molecule in the strict sense that I have been using that term. Organisms today use complex enzyme systems to catalyse DNA replication. For the light that it may throw on our immediate discussion and because I will be returning to the question of nucleic acid replication in Chapter 9, let us consider this matter in some detail.



The simplicity of Watson & Crick's original scheme has given way to an amazing complexity of the still incompletely understood reality (see, for example, Watson, 1976; Alberts & Sternglanz, 1977; Cozzarelli, 1980). There is not just one enzyme involved but a multi-enzyme complex or **replication apparatus**. In *E. coli*, at least thirteen proteins operate in the replication fork. And the mode of operation of this machinery makes figure 5.3 look childish. In *E. coli* there is a 'gyrase' that uses ATP energy to put negative twists into the DNA double helix: this makes it easier for another enzyme to unwind the DNA into single strands which are then stabilised by binding of yet other proteins. The single strands then act as templates for synthesis of new double helix. These unwound strands necessarily run in opposite directions: synthesis of new DNA however, by a DNA polymerase, only takes place in one direction – towards the 3′-end. As a result, one of the DNA molecules has to be made in a series of short stretches (figure 5.5): this is the more complicated because the polymerase apparently cannot start a new chain of DNA on a single strand. Another primer enzyme is needed for this and it puts on a piece of RNA. With a (hybrid) double strand thus started, the DNA polymerase can get to work extending (as proper DNA) towards the previously placed segment. The RNA piece that had been on the end of this segment has to be removed (with another enzyme) before the final sealing of the gap with yet another enzyme – DNA ligase (figure 5.6).

Much of this seemingly cumbersome manufacturing procedure can be understood in terms of the overriding need for fidelity in DNA replication. (It has been estimated that *E. coli* makes, on average, one mistake in replication for every $10^9$–$10^{10}$ base pairs (Drake, 1969): that is about one mistake for every *metre* of DNA processed.) It would seem that starting a new strand on an existing strand is particularly liable to errors. Presumably, there is insufficient stereochemical control until a fair sized piece of double helix has been formed. Using RNA for starting is a clever way of labelling the dubious stretches for subsequent identification and removal. But why make so many starts, by insisting on a 5′ → 3′ direction of new strand synthesis? Because, among other things, this direction makes easier an error-correcting mechanism that is built into the DNA polymerase. A DNA polymerase is really two enzymes in tandem – a polymerase proper and an exonuclease. Faced with a mismatched base pair at the growing end of a new strand, the polymerase part delays the addition of another deoxynucleotide, allowing the exonuclease part to hydrolyse off the misplaced deoxynucleotide, and continues in this manner until a correct base pair is reached. It then stops, or at least slows down to a lower rate than that at

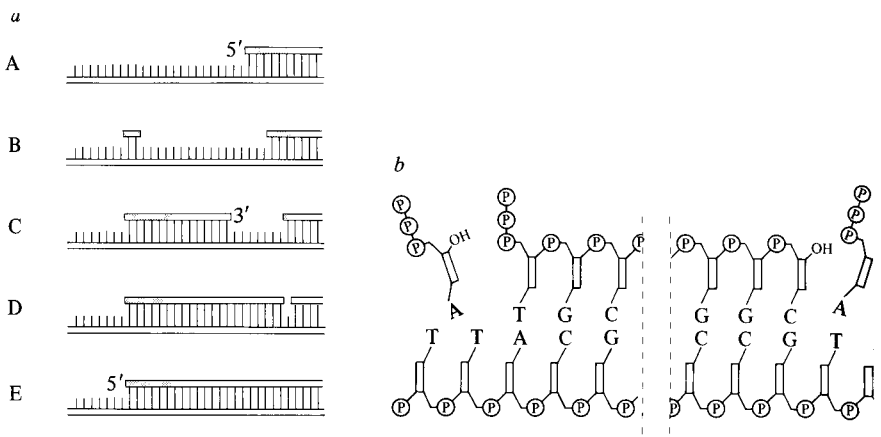



Figure 5.6. (*a*) Overall synthesis of a new DNA strand in the 5′ direction (between A and E) actually takes place with the strands growing in the 3′ direction (as indicated in figure 5.5). Since DNA synthesis can only take place by adding to an existing strand, the many new starts implied by this back-to-front mechanism each requires initiation through emplacement of a section of RNA (hatched) that is subsequently removed. Many enzymes are involved in these processes. (*b*) On the right we have a closer view of a daughter chain growing in the (normal) 3′ direction. Growth in the 5′ direction (left) would be precarious – because liable to be blocked by accidental hydrolysis of the activating pyrophosphate group attached to the growing chain.

which the polymerase part can now proceed to add new deoxynucleotides – the exonuclease always poised to undo future errors.

One can see, from figure 5.6*b*, why such a device would not work so easily in the 3′ → 5′ direction. In that direction, the growing chain would hold the activating pyrophosphate needed to make the main chain phosphoester linkages. This would be a nuisance in any case since accidental hydrolysis of the pyrophosphate group would prevent further synthesis until the 5′-phosphate had been reactivated. Removal of a misplaced nucleotide by the exonuclease would have a similar effect – the growing chain would have to be reactivated each time.

No doubt the present machinery for DNA replication has been the product of an extended evolution. No doubt there was much simpler machinery to begin with. And it is very plausible that RNA, whose replication seems to need less machinery than DNA, was the earlier genetic material – we will be assuming as much later. But RNA replication still needs evolved machinery, as far as anyone knows: neither DNA nor RNA have been



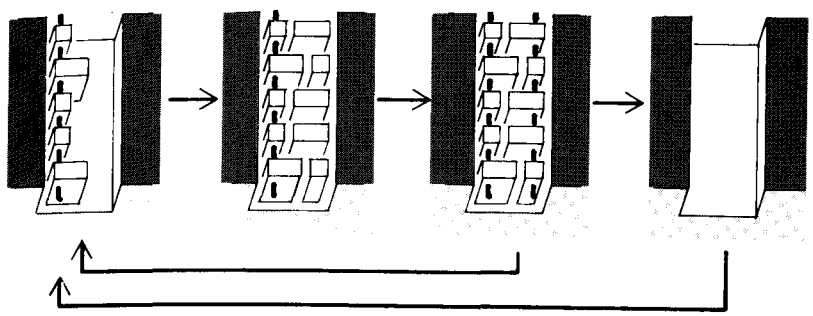


Figure 5.7. A grooved surface might assist in the alignment of monomer units during the replication of a copolymer consisting of units of different sizes.

shown to replicate at all effectively without enzyme assistance (Orgel, 1979; Orgel & Lohrmann, 1974). It is an extrapolation that does not yet fit the experimental evidence that nucleic acid could ever replicate without sophisticated catalysts. Nor does such an idea fit with theoretical expectations – remember how many constraints there are on the design of self-replicating polymers generally. It looks very much as if the nucleotide, which is far too big to be a plausible probiotic product (Chapter 1), is, at the same time, far too small to be a monomer for a *self*-replicating polymer: there are not enough possibilities to choose from with molecules of this sort of size.

An advantage of the idea of genetic takeover is that a secondary genetic material, such as DNA, need never have been a *self*-replicating molecule: it would have evolved in a context that was already biologically evolved. As touched on in the last chapter, and as we will discuss in more detail in Chapter 9, nucleic acid replication could have been a function gradually arrived at through a number of preadaptations. Part of that gradual evolution would have been the evolution of assisting catalysts (probably not protein to begin with).

But perhaps the environment might have provided a catalyst for the replication of some DNA-like replicating polymer. One might think of a grooved crystal surface orienting a large–small copolymer so as to create sites with a complementary specificity (see figure 5.7).

Such a mechanism dispenses with specific X and Y functional groups. On the other hand, a fortuitously adapted environment would be needed. The groove must be of the right shape: the polymer must tend to adsorb in it; then all the monomers must shuffle into place; then they must completely polymerise; then they must come away. A neat balancing of energies would still be required, and many of the problems discussed before

are still there. For example, some effect must stop premature polymerisation or, alternatively, premature desorption of incomplete chains. Something must make double occupancy of a groove unstable. There must be no poisoning of the groove with strongly binding impurities. And so on.

*Assisted replication of organic polymers: (b) environmental programming*

A second form of assistance for replicating organic polymers would be through macroscopic manipulation of environmental conditions. This is the way that an organic chemist would approach the problem. Looking at any of the schemes in figures 5.3, 5.4 or 5.7, the problems of monomer design would be greatly relieved if one could take batches of molecules synchronously through different stages by suitably changing temperature, pH, etc. One might imagine a machine, rather like an automatic peptide synthesiser, that could take the molecules through blocking, activating, condensing, and deblocking stages to mitigate what is perhaps the worst of all the problems – inappropriate cross-reactions and side-reactions. (For discussions of the design of laboratory replicating polymers, see Cairns-Smith & Davis, 1977; and Stillinger & Wasserman, 1978.)

Could the primitive Earth have been in any sense such an automatic machine? Kuhn (1976) has suggested something of the sort: that small environments could have had characteristic programmed sequences of change, some of which might be suitable for assisting nucleic acid replication. An example which Kuhn gives is of a number of rocks in the sun: as the day proceeds, shadows will fall in different places and in complicated ways, creating different characteristic diurnal temperature programmes in different sites among the rocks. On a seashore, the tides might add another sequence of wetting and drying; or morning dew might have a similar effect. Whether such an idea is feasible or not depends on how reliable and how particular a programme would be needed.

But the basic problem remains. If a number of independent considerations have to coincide for a given circumstance to arise, then the probability of that circumstance rapidly approaches zero as the number of considerations increases. Even for individually plausible considerations, a few becomes too many and several becomes absurd. And that seems to be the trouble with imagining a circumstance that would allow the consistent replication of any organic polymer that is analogous to DNA. That sort of thing looks too ingenious, however we may try to distribute the ingenuity between the design of monomers, or of assisting catalysts, or of a fortuitously preadapted environment. Penrose (1959a and b) in his designs for macroscopic replicating devices, found that, here too, what appears to be quite a simple problem turns out to be rather complex when it comes down to the mechanical details.



**Replication through crystallisation**

The difficulties discussed in the last section arose because there are insufficient reasons to see why a combination of properties and conditions required for ladder replication should be generated purely by physicochemical processes. Too much has to be put down to coincidence. What we seek are models of genetic replication in which such coincidences are reduced: where not only individual aspects of design are physicochemically possible, but where their appropriate combination is also somehow explainable in physicochemical terms. To illustrate what I mean, think again about the hypothetical grooved crystal acting as a jig for a copolymer replication (figure 5.7). One of the difficulties was that the jig had to be just the right size for the polymer. This need not be such a coincidence if the crystal that acts as the jig is the polymer itself. Crystals often have grooves on their surfaces that fit the molecules out of which they are made.

I will not take that idea any further because there is a better related idea illustrated in figure 5.8. Here, the pairing of chains is dispensed with altogether, and the grooves are templates as well as jigs – that is, they specify the sequence of incoming monomer units as well as holding them in a position suitable for polymerisation. I have chosen a short–long copolymer sequence in this illustration, but there are several other possibilities that might be imagined – for example, with wide–narrow, or anionic–cationic copolymer sequences that aligned in a complementary way in the crystal. In such models, crystal packing forces – repulsive as well as attractive – take the place of the mutually recognising X and Y groups in the ladder model. Instead of one polymer chain growing on another single chain, the polymers grow on crystalline bundles of chains. The chains in a given bundle would each have the same sequence (or alternatively, for some models, the bundle would contain an equal number of complementary sequences).

How do the chains eventually separate? They don't. These 'crystal genes' would not be separate molecules but crystallites of colloidal dimensions. The genetic information would be highly redundant, it is true, repeated perhaps thousands of times in the chains of a given bundle; but this

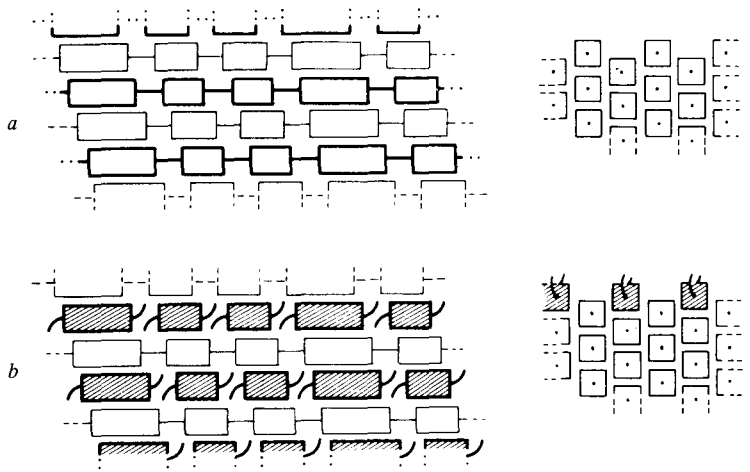



Figure 5.8. (*a*) Top and end views of a hypothetical copolymer crystal showing surface grooves. (*b*) Unpolymerised units (hatched) are located and aligned in the grooves in a manner similar to that in the polymer. (From Cairns-Smith & Davis, 1977.)

could make the process of replication more reliable: a given template groove is specified by at least three adjacent chains and more weakly by more distant chains that together make that groove. A mistake in one chain might thus be overlooked by a growing polymer which conforms, as it were, to a majority opinion. To complete a replication cycle, it would not be necessary that the individual chains separate, but the crystallites would have to break up from time to time and so maintain a population of colloidal dimensions.

Notice, particularly, that the sequencing of the various processes would not be so critical as in the corresponding processes for ladder replication. It would not matter if the monomers began to polymerise before a complete row was in place, nor would it matter if the crystallite broke up while there were still incomplete chains attached to it. The whole thing can be more casual just because it is, from the point of view of information density, less efficient. But the density of information that a genetic material can hold is much less important than the fidelity with which it can transmit that information to offspring.

There would still be plenty of things to go wrong though. There would be the possibility that spontaneous polymerisation in solution would compete with the surface reaction. And the surface reaction would have to be sterically possible. And how good would these grooves really be in recognising and aligning the monomers? Strictly the grooves will fit the polymer out of which they are made – not the corresponding monomer row: that would be misfitted to the extent that the monomers have a different shape and electronic structure from the polymer. In any case, would the binding forces really be strong enough for small monomers?



In general, the misfitting of the monomers in the grooves would be less for rather *large* monomers since there would be less proportionate change in structure on polymerisation. Also, the binding would be stronger. The surface polymerisation might even be specifically preferred, from the following consideration. Suppose that the monomers in figure 5.8 were really oligomers of two different kinds and lengths. If they were long enough they might bind quite well in the corresponding crystal grooves due to favourable interactions along most of their length – in spite of misfitting at their ends (designated by the bent sticks at the ends of the blocks in figure 5.8*b*).

One might compare this with the binding of the lysozyme substrate in its groove in the enzyme: the catalytic activity of lysozyme is thought to depend on its substrate fitting well in the enzyme groove except in one place where it is strained – a strain that is relieved by the substrate molecule passing along the reaction co-ordinate (Phillips, 1966). In the model in figure 5.8, the end-strains would be relieved by polymerisation which might thus be specifically catalysed by the grooves.

However interesting as a project for polymer chemistry, I doubt whether we are yet talking about a primary genetic material. We would seem to need rather large molecules as monomer units – so that they would 'crystallise' well on the polymer crystal surfaces. But (the old problem) primordial clean supplies of large molecules – even of molecules of the size of glucose – are particularly difficult to imagine since such molecules are likely to belong to a very large set of similar molecules that would form under similar conditions and interfere with crystallisation processes.

We seem to have arrived at another dead end. Not, I think, because we have been trying to rely too much on crystallisation processes, but because we have still not made enough use of them. The simple picture of units locating and then irreversibly locking into place fits our intuition about what macroscopic manufacturing processes should be like, rather than what most crystallisation processes really are like. The astounding overall accuracy of crystallisation depends on local reversibility: a site on the surface of a growing crystal does not have to be all that selective – a newly inserted molecule can usually come away again if it turns out later to have been misplaced. By making some of the bonds in our model irreversible, we threw away that key to the accuracy of crystallisation. We had to rely

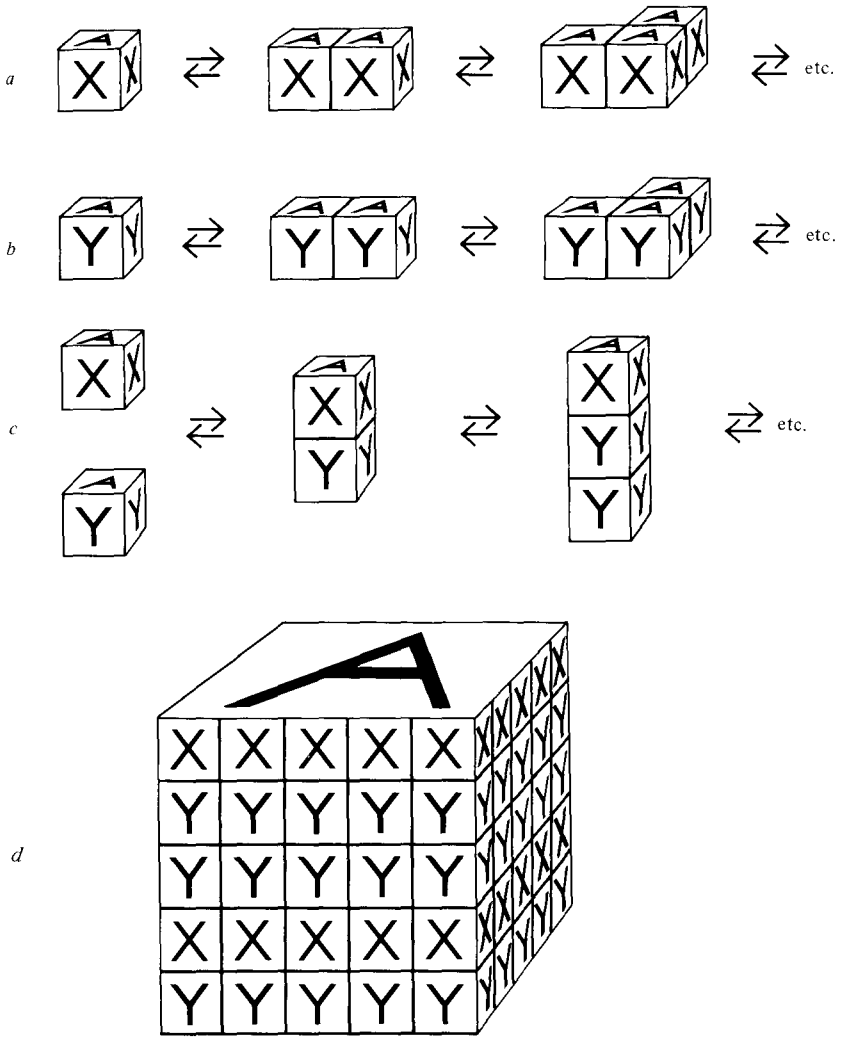



Figure 5.9. Suppose that there are two kinds of molecular cubes, each with a pair of opposite A-faces, the other faces being either all X or all Y. Suppose also that like faces associate and unlike faces do not. X–X and Y–Y associations would themselves tend to give two classes of platelets (*a* and *b*). A–A associations would give any of an immense number of possible linear copolymers (*c*). The combination of all three types of association would tend to lead to crystals like (*d*). This could be a genetic crystal provided that, in addition, (i) A-face growth was usually inhibited, and (ii) the crystals periodically cleaved, but only normal to the A-faces. Such would be a 'type-1' genetic crystal.



instead on accurate surface–monomer recognition and hence on implausibly large units. The question sharpens into this: how can you reliably retain genetic information in polymers that have reversible main bonding? For crystal-gene models this is not, as it may sound, a contradiction in terms.

*Models for two types of genetic crystals*

Consider the following pair of units:

Let us suppose that the faces of these cubes tend to stick together reversibly such that A–A, X–X and Y–Y interactions have lower energies than any mixed interaction. Furthermore, these energies are determined solely by the faces in contact. A supersaturated solution of a mixture of these units would give rise to disordered crystals with two opposite 'A-faces' and four 'X/Y-faces', as in figure 5.9. Each crystal would contain its own 'information' in the form of identical X–Y sequences being replicated through growth on any of the X/Y-faces. (Growth on the A-faces should be inhibited, since this would lead to new 'information' being added in a random manner.) In the model in figure 5.9, information is being held in one dimension and being replicated through crystal growth in the other two dimensions. (This was true also for the model shown in figure 5.8.) I will call such things **type-1 genetic crystals**. A **type-2 genetic crystal** would be one in which information is held in two dimensions and replicated in the third. A block model for a type-2 genetic crystal is illustrated in figure 5.10.

Figures 5.11 and 5.12 illustrate replication cycles for type-1 and type-2 genetic crystals respectively.

*Crystal growth and error correction* We have moved, it seems, very far from DNA. Yet, curiously, in one crucial respect we are closer to DNA in these last models than in any of the previous ones – because these models have a 'proof-reading' mechanism built into them. Think about the growth of a crystal face by successive surface nucleations. In figure 5.13*a*, we have the first difficult step of such a nucleation with a single unit precariously attached to a site. Under only slightly supersaturated conditions the chances are that that unit will come away again before others

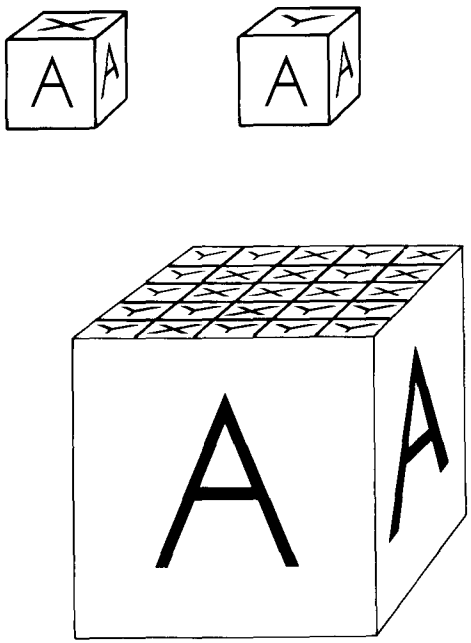



Figure 5.10. (cf. figure 5.9.) A 'type-2' genetic crystal might be based on cubes with four A-faces and either two opposite X-faces or two opposite Y-faces. Such a crystal would contain information in two dimensions.

come alongside. Eventually, by a chance fluctuation, a stable island forms (figure 5.13*b*). Even then, new units will come and go many times: the difference between crystal growth and crystal dissolution is not absolute under such circumstances: both happen, it is a question of which is marginally faster. And mistakes are frequent – perhaps usual. That is no matter since a crystal with a mistake in it is less stable – more soluble – than a more perfect crystal: the crystal will tend to redissolve until the mistake has been eliminated. Trial and error – and error correction – are the secrets of success. It is perhaps significant that the two most accurate processes that we know of in the universe – the faultless replication of a metre of DNA and the growth of a hand-sized crystal – each depend on a continuous product control. For crystal growth, this 'proof-reading' is probably much the most important element: only when a molecule is already snuggled into a lattice can the lattice 'feel' whether it is really the right molecule, rightly orientated. (Is it generally true, perhaps, that it is easier to know when something is wrong than to get it right in the first



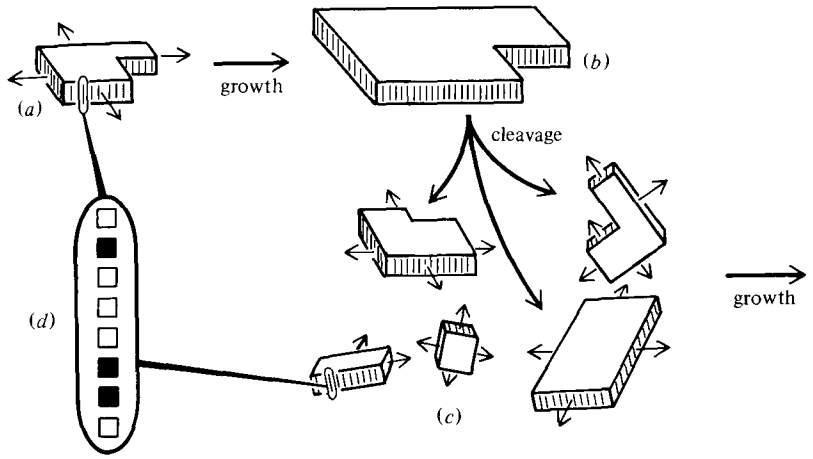



Figure 5.11. A replication cycle for a type-1 genetic crystal. (*a*) is a crystallite consisting of a bundle of identical sequences (*d*). Growth takes place on faces, as shown, by the incorporation of monomers such that polymer chains are not extended. This gives a platey morphology (*b*). (*b*) Shatters into smaller fragments by cleaving parallel to the growing faces ($b \rightarrow c$). Each of the crystallites (*c*) still contains only the copolymer sequence that was in (*a*). Through this combination of crystal growth and crystal cleavage, replication continues indefinitely, maintaining a population of crystallites with a roughly constant distribution of sizes.

place?). With DNA replication too, it is likely that product control is at least as important for fidelity as initial accuracy in the placement of incoming nucleotides – perhaps more so (Hopfield, 1974; Ninio, 1975; Topal & Fresco, 1976). The exonuclease is one of the main keepers of our genetic inheritance through the element of 'reversibility' that it (and other repair enzymes) effectively introduce into the main bonding of DNA. But this 'reversibility' must be very specific: faced with a misfit, the enzyme must know which strand to believe, it must distinguish between parent and daughter. (It is hard to see how this could happen in a primitive ladder polymer without yet more design constraints to be added to an already daunting list. Yet after-the-event control is, perhaps, a necessary part of tolerably accurate molecular replication.)

In genetic crystals such as those shown in figures 5.9 and 5.10, the genetic information could be retained in spite of surface main bond reversibility, because most of the copies of the information are safely inside the crystal. If there is a misfit, it is easy to see which sequence is wrong – the one that is different from all the others. Usually, this will be near the

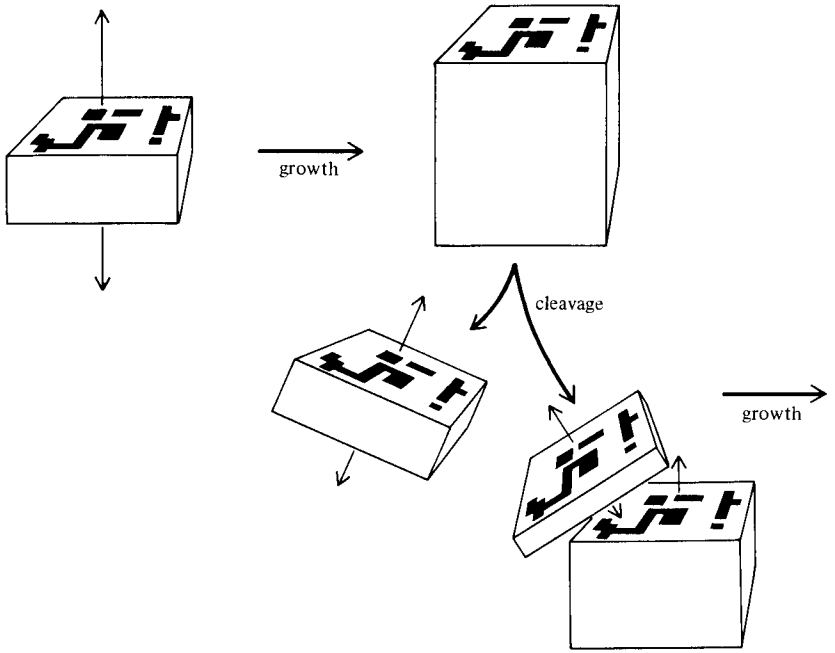



Figure 5.12. A replication cycle for a type-2 genetic crystal (cf. figure 5.11). Here information is held as a two-dimensional pattern that replicates indefinitely through exclusive growth on two opposite faces and cleavage only along planes parallel to the growing faces.

surface and so will tend to be removed automatically through local dissolution resulting from lattice instability caused by the misfit.

*An energy problem* These models involve an assumption that is arguably unlikely, however: there is no significant energy difference between different X–Y arrangements. Suppose there was a big difference – let us say that in figure 5.9 an alternating X–Y sequence is much more stable than anything else. Then misfits during growth that increased the extent of this alternation would be less disfavoured. Indeed, if the energy drop resulting from alternation within a polymer was greater than the energy rise due to misfitting between the polymer and its neighbours, then any initial information would be quickly wiped out. The initial 'imperfect' crystal would have learnt to be 'perfect'. In any case, even a small energy difference would lead to eventual rubbing out of initial information – here thermodynamics would be on the side of finding that unique informationally boring solution.



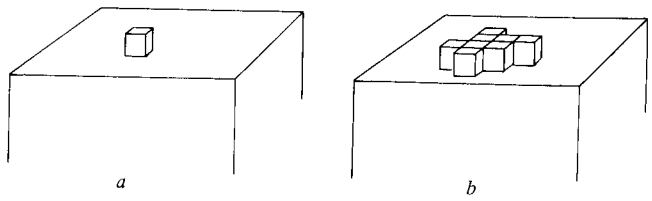



Figure 5.13. Surface nucleation of a new layer during crystal growth (see text).

*A question of supersaturation* This is the negative side of reversible main-bonding, but is by no means a fatal objection: there are many crystals that contain partial irregularities that have very little effect on stability. Provided the energy difference between a disordered and ordered structure is small enough, the disordered structure could still be preferred. The crucial question is one of balance between rate of erasure and rate of selection of a functionally significant disordered structure. I will return to this topic in Chapter 7. But, we can already see this much about what is required for the replication of crystal genes: the environment must provide a solution of suitable supersaturation. If the solution is too supersaturated, then error correction mechanisms will too often be overtaken and crystals will be formed that are heavily mutated with inappropriate disorders; if it is kept only barely supersaturated, then even small differences in energy between the informationally disordered and 'blank' regular structures could lead to the latter being preferred in the long term. Only an open system could maintain an appropriate level of supersaturation. Over the next two chapters, we will consider the question of how the primitive Earth could have set up open systems of the kind required.

### *Generalised genetic crystals*

At the most general level, a genetic crystal would be a small crystal, probably of colloidal dimensions, that contained some irregularity of structure that was accurately copied in growth by the laying down of fresh molecules in conformity with the pre-existing irregularity. That is close to Hinshelwood's pre-Watson–Crick description of what a gene must be like (quoted at the beginning of this chapter). Our last models were both substitutional crystals, with genetic information held as arrangements of chemically distinct units, thus reflecting, still, an aspect of DNA's struc-

ture. But even this homage to DNA is not necessary. As we will discuss in Chapter 7, the irregularity in question might be purely physical and not chemical – a pattern of dislocations, perhaps, or a complex form of twinning.

*Big writing to begin with?* We should bear in mind the possibility of another kind of redundancy, apart from the multiplicity of copies inherent in the genetic crystal idea: primitive genetic information might have been 'writ large', the irregular features (the information units) being bigger than the units out of which those features were built (the genetic monomer units). Think of the reproduction of a printed page by photography: here, the information units are the letters on the page, but these are formed from silver grains that are several orders of magnitude smaller. Where fidelity of replication is a major problem – and this would be so right at the start of evolution – we should not assume that the genetic information was written as it is now, at the level of the genetic monomer units, even if subsequent evolution might lead to systems operating more intricately. In terms of the block model in figure 5.10, the information might consist of relatively large patches – domains – of X and Y co-operatively more stable and more easily copied. We will return to this idea in Chapter 7 when we have more particular examples to consider.

### Inorganic genes

Under cover of abstractions, we have now, I think, moved well into inorganic chemistry. Our course was set in that direction as soon as we started to think about replication through crystallisation.

Mainly it is a question of bonding. The genetic units, we said, should be smallish, water-soluble, yet able to form colloidal crystallites, stable when dispersed in water. This suggests strong bonding throughout the crystallites. Yet all this bonding has to be quantitatively reversible at (some) crystal surfaces. A typical organic polymer crystal is held together by a combination of secondary forces, which are reversible but not strong, and covalent bonds that are strong but not reversible.

This would be too neat a dismissal of any conceivable organic polymer crystal gene: perhaps hydrogen bonding could provide strong enough template forces. And the main bonding is reversible in some organic polymers (e.g. polyoxymethylene). But it would seem that, in organic polymerisations, activation energies are usually too high, and/or reaction trajectories too sterically demanding, for controlled polymerisation on the surface of



preformed polymer crystals to happen very often. Usually, monomers *first* polymerise into a chain which *then* crystallises. This is the wrong order of events for the crystal structure to have much chance of controlling the character of the polymerisation. Indeed, usually organic polymer crystallisation is only partial, zones of aligned chains being embedded in chaotic tangles. This can be seen as a likely consequence of polymerisation and crystallisation being separate events taking place in that order.

If we cannot dismiss completely the possibility of a primary organic crystal gene, I think it fair to say nevertheless that the main bonds of organic polymers are generally too difficult to form, and too stable when formed, for interchain forces to exert a decisive control on polymerisation.

Inorganic polymers present a far more promising field – in particular, those kinds of polymers that underly the majority of minerals. Most of these are essentially crystalline – they do not exist free in solution. Many are colloidal, their crystallites having formed at ordinary temperatures from dilute aqueous solutions barely supersaturated with small units such as silicic acid and hydrated cations. These crystalline polymers are far more complex than our idealised models: but often they contain features of the bonding arrangement required. There are often bonds of different kinds arrayed in different directions which may give rise to perfect cleavage in certain planes and preferential growth in certain directions. The bonds between the units are generally strong, ranging between electrovalent and polar covalent types. Inside, the crystals are thus stable at ordinary temperatures. In contact with water, however, at the surfaces of the crystals, these bonds are more labile – allowing perhaps the kind of error correction during crystal growth that we have talked about. In any case, the extensiveness of crystallinity typical of many colloidal minerals makes it clear that, however it is done, the crystal structures themselves must largely control the way in which their covalent bonds are made. Furthermore, mineral crystals are frequently substitutional, and possess indeed a variety of irregular features that might store genetic information.

#### *Three objections*

Later, we will come to the idea that not only primary genes but early phenotypes might have been purely inorganic – that minerals might have been the materials out of which an early primary life was made. Our journey will be via chemistry that may be unfamiliar to some readers. Perhaps we should check that we are not obviously off course. I can imagine three objections that Dr Kritic of Chapter 4 might make at this stage.

First, he might make a general point to doubt the validity of the journey so far: 'Long serial arguments are always unsatisfactory – where the conclusions from argument A are used as a basis for argument B, whose conclusions are then used for C and so on. Even if individually the arguments are plausible, the string of them taken as a whole may be quite implausible.'

Dr Kritic would be quite correct in saying this and, indeed, all of us interested in very early evolution should have his remark written on our wall. But the discussion so far has not been an extended argument series. It has been, as I said, a journey; a sightseeing tour through the landscape of possible primary genetic materials to show what a barren sort of place it is without attempting to visit everywhere. If there seems to be an oasis in the region of inorganic crystal chemistry, I cannot quite prove that there are no other oases elsewhere. And the sceptical arguments suggesting that the regions around DNA are indeed barren do not depend on considerations in series: they depend on considerations piled on top of each other. Unlike serial arguments, which become weaker as they become more numerous, parallel arguments become stronger when taken together. (Another thing to write up on the wall.) Having arrived at inorganic crystal chemistry, there is only one way to check that there is an oasis here and not a mirage: we must take a closer look. There is nothing for it but to continue the journey, to find specific examples in inorganic crystal chemistry and see if these can be worked out to give a detailed plausible story. Only after that, with detailed considerations in mind, can we hope to design clearly relevant experiments.

There is another kind of objection that I have sometimes come across – of the 'you're-not-going-to-tell-me' kind. For example: 'you're not going to tell me that life started as a kind of mud'.

A similar stone wall might have been put up in earlier times to refute the suggestion that moulds are alive, since they too are macroscopically unprepossessing. That kind of long view can be decidedly unhelpful. The critical questions concern structures in the size range visible under the electron microscope, and the most intricate details of crystalline architecture. At these levels, colloidal inorganic minerals – 'muds' – can be far richer in structure than mineral organic materials like raw petroleum. (The microstructures of 'muds' will be considered in detail in the next chapter.)

A more sophisticated objection is along these lines: 'life must be founded on organic molecules because only then is the necessary variety of molecular machinery possible – indeed, there must be strong faithful bonds between atoms to make proper molecular machines at all. That means that there must be hard covalent bonds of the sort that form typically between first-row elements. In particular, carbon is implicated because of the usual stability of the C–C bond under aqueous conditions and because of the high valency of carbon – and hence the unique richness of variety in its compounds.'



This argument cannot be dismissed so easily. Life is nothing if not naturally occurring machinery. But it makes the opposite mistake to the previous one: it takes too close a view – it assumes that an atomic grain size is essential for living machinery. This is not even true of all components of present-day life. (I doubt whether the exact position of every calcium atom in your teeth was genetically determined or really matters very much.) Some living machines can work very well without being specified down to the last atom in the way that, say, an enzyme is. There are structures specified at various levels in life now. Who is to say that primary life used any truly molecular machines? We have indeed started to answer this argument in pointing out that a primary genetic material might work better if its information was not written so finely as genetic information is written today. When we come to discuss primitive phenotypes, I will pursue this question of grain size further. We will then think of the ultimate machines of primary life as having been made on a slightly larger scale than ours.

*The necessary and the efficient* Returning to an earlier theme, we should always be careful to distinguish the necessary from the efficient. Indeed, it is efficient to be able to make the most intricate possible machinery. But, if this is a more difficult way, and if it is not necessary for making evolvable systems, then it is not to be expected for primary life. On the other hand, once clumsy first organisms had started to evolve, we can predict selection pressures tending to make them less clumsy: a continual drive towards finer and finer control through increasingly fine-grained machinery. Genetic takeover provides a mechanism for even the most central machinery to give in to such pressures through radical changes. The result should be predictable: machines will arise eventually of the finest possible grain size with atoms as nuts and bolts and molecules as working components. Arguments about the excellence of carbon are thus true but misdirected: they do not tell us what to expect of first life: they tell us what to expect of biochemically highly evolved life. And, of course, all life now on Earth (that we know of) is firmly in this latter category.

# 6

# The first biochemicals

## Introduction

By assuming that at least one genetic takeover was involved in the origin of our biochemical system, we are released from any need to conform to present biochemical ground rules in attempting to formulate likely structures and *modus operandi* of a primary life. In the last chapter, we took full advantage of this freedom in trying to arrive at the kind of stuff that primary genes were made of. What processes, occurring in the non-biological world, might inadvertently conform to specifications for a self-operating primary genetic material? That was the central question, and our tentative answer was 'crystallisation'. This seems to be the only non-biological process that would be accurate enough to sustain an extended evolution. It was possible to be somewhat more specific about what kind of crystallisation we should be thinking about and what kinds of crystals: and we were pushed into contemplating inorganic materials (inorganic in the chemist's sense).

### *The change of view*

This change of view from organic to inorganic has to be radical, I think, to be satisfactory. It is not just a question of imagining a mineral to begin with instead of DNA. Primitive biochemistry should be seen as having been quite different from present biochemistry in all but the most abstract essentials. The change of view that I suggest is altogether away from the idea that life started with some class of organic Earth products (which the Earth is no longer making), to the idea that the materials of first life, all of these materials, were inorganic minerals – presumably of



kinds still being produced by the Earth. To understand the beginnings of life, then, it is present geochemistry, not present biochemistry, that is most germane.

This chapter is a long pause for information. Some readers will become impatient with it, wondering why so many minerals are being brought in. The reason is partly that we are engaged in a search: first of all we should survey the field of possibilities; we should try to get some feel for the variety and variability of the materials that the Earth can most easily produce. And we will be at least touching on a substantial fraction of the possibilities. I will grant that it may not be important to know, say, all the names and formulae of the zeolites given on pages 189–91. But one should have some notion of the variety of zeolites, of their crystal structures and morphologies, and of how they are formed in Nature. Then, again, I do not imagine that at first reading all the subtleties of layer silicate chemistry that we will be discussing will be retained by the reader who is not already familiar with them. But it should at least be noted that there are such subtleties: that silicate layers contain, very often, fixed charges; that they are often flexible; that they are highly articulated structures that can transmit effects through them; that many layer silicate types can form from dilute solutions at around ordinary temperatures. It may be that the most important thing is to get a general sense of what these minerals are like; but even that can only be arrived at by considering some examples in detail.

In any case, the search is not for one mineral – the true stuff of first life. It will be part of the view to be developed in Chapters 7 and 8 that early organisms were assemblages of minerals of different sorts that had in common the ability to crystallise from solutions derived from the weathering of rocks.

Such materials are, broadly speaking, **clays**; and it will be convenient to define clay for our purposes as any microcrystalline material (particle size less than 10 $\mu$m or so) that can form from aqueous solution near the Earth's surface. On this definition, a clay is very often, but not necessarily, a layer silicate. Framework silicates, metal oxides, hydroxides and sulphides are among other types that can be 'clays' within the above definition.

*Other minerals of interest* In this chapter, we will consider a number of minerals that are not themselves clays but whose structures are similar to major clay types and are known in greater detail, for example muscovite. In other cases, the mineral may be of interest as a source of

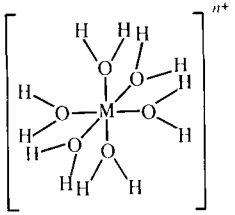
weathering solutions: olivines and pyroxenes are discussed briefly for this reason, and because they also help to illuminate general aspects of silicate structures.

*Two main themes* Two of our main concerns in this chapter will be with the structures and conditions of synthesis of clays. These themes will be returned to in later chapters to provide the basis for more detailed discussions of crystal genes, to be given in Chapter 7; and also for the development of ideas, in Chapter 8, about how minerals organisms might have started to photosynthesise and to evolve a competence in organic chemistry.

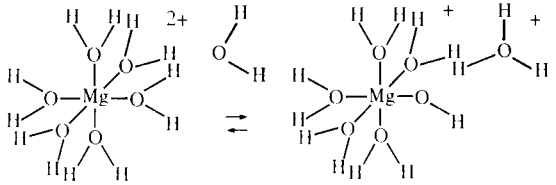
But, in this chapter, I will refrain from such speculations. First, let us study the materials: then we will be ready to think about a putative stone-age biochemistry at the beginning of our evolution.

### Cation polymerisation

Metal cations in aqueous solution are invariably hydrated, the water molecules in their immediate vicinity being more or less regularly arranged. An inner shell of octahedrally co-ordinated water molecules is common:



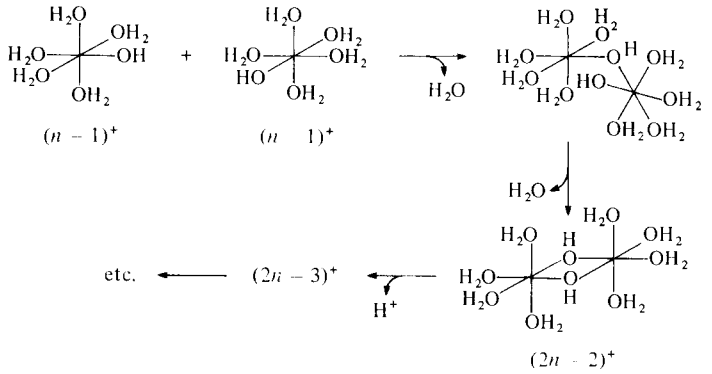
Such an object – an aquo ion – is a kind of molecule: the co-ordination bonding between oxygens and metal is very strong, even if activation energy barriers are low enough for water exchange to take place rapidly. $Cr(H_2O)_6^{3+}$ and $Rh(H_2O)_6^{3+}$ are exceptions here with exchange rates on the time scale of hours, but the usual time scales are from seconds to nano-seconds (Cotton & Wilkinson, 1972). These molecules are acids, they can donate a proton to a water molecule, for example:



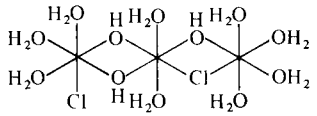
The conjugate base on the right hand side is called a hydroxo complex.



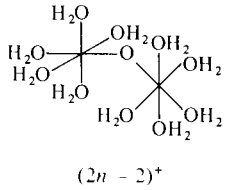
Almost all hydroxo complexes undergo **olation** to give, in the first place, dimers joined by one or two hydroxo bridges. Further deprotonation and olation can lead to higher oligomers. For example:



This may be further complicated by anions displacing water molecules, and perhaps also hydroxo groups, from co-ordinating positions, for example in oligomers such as:



Another way in which positive charge on polymers may be reduced is by **oxolation** where oxo bonds are formed through loss of protons, for example:



As the pH of a solution containing a given cation is increased, so is the tendency to oligomerisation. A precipitate of high polymer, for example an hydroxide or oxide, usually appears within 1 or 2 pH units from the onset of oligomerisation.

Although the structures shown above are vaguely reminiscent of oligo-saccharides, there is one important respect in which they are usually quite different from organic structures. These cation oligomers are seldom isol-

able, being in a dynamic equilibrium that shifts fairly quickly on dilution, change of pH, etc. $Cr^{3+}$ is one interesting exception – a blue dimer and a green trimer can be separated from solutions of chromium (III) that have been refluxed and cooled (Laswick & Plane, 1959). But even here the kinetic stability is less than that of, say, a sucrose molecule in neutral aqueous solution.

According to Baes & Mesmer (1976), cations are generally monomeric at sufficiently low concentrations with only a small number of kinds of oligomers predominating as concentrations increase. Very often, from one to three of the following types will be found:

$$M_2OH^{(2n-1)+} \quad M_2(OH)_2^{(2n-2)+} \quad M_3(OH)_3^{(3n-3)+}$$
$$M_3(OH)_4^{(3n-4)+} \quad M_3(OH)_5^{(3n-5)+} \quad M_4(OH)_4^{(4n-4)+}$$

(Co-ordinated water molecules are ignored in these formulae.)

Between the generally small inorganic oligomer and the crystalline precipitate of high polymer, which becomes accessible to X-ray structural analysis, there is much ignorance. Particularly with highly charged cations, metastable materials may persist for a considerable time at ordinary temperatures. For example, Rausch & Bale (1964) made $Al^{3+}$ solutions in which platelets (*ca* $1 \times 30$ nm) formed at first, and this gave rise to a gel. The gel reverted to a sol after a few days. In another study, Turner & Ross (1970) found that a precipitate was formed first which contained chloride. This then dissolved while the concentration of oligomers in solution increased: these in turn decreased as the stable crystalline form of $Al(OH)_3$ appeared after a few weeks.

The polymerisation of $Fe^{3+}$ at ordinary temperatures provides a useful example to consider in more detail. Murphy, Posner & Quirk (1976a) investigated the reaction of partly neutralised ferric nitrate solutions using an electron microscope. The first products identified were small discrete spheres of apparently amorphous polymer (diameter 1.5–3.0 nm, mol. wt 3000–30000). On ageing, the spheres joined up into rods of two to three spheres. With solutions of higher ionic strength and/or $Fe^{3+}$ concentration these rods remained the same length and aggregated into extensive raft-like arrays. The spheres slowly coalesced within the rods and the ferric oxide-hydroxide polymorph goethite was formed. With lower ionic strength and/or $Fe^{3+}$ concentration the rods increased in length without forming rafts but again crystallised into goethite. After extended ageing, some of the rods had coalesced to give lath-like goethite microcrystals. Figure 6.1*a* is an electron micrograph taken after ageing for one year. Laths, and single and double mature rods are visible, but also, still, single



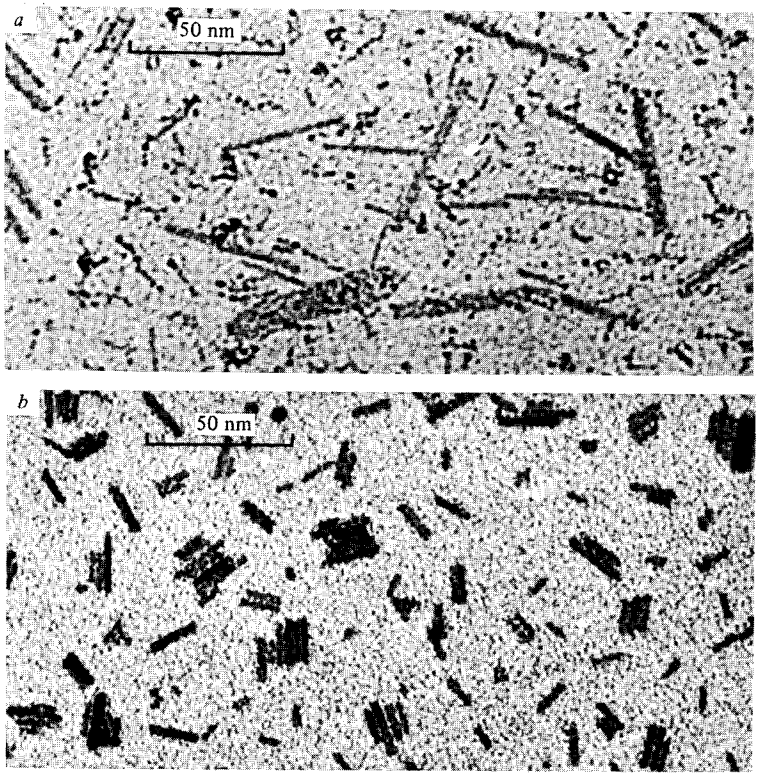



Figure 6.1. Transmission electron micrographs of products of polymerisation of ferric solutions for 1 year. (From Murphy, Posner & Quirk, 1976a and b.) (*a*) 0.0165 M $Fe(NO_3)_3$, OH:Fe ratio 2.08. *ca* 3 nm diameter spheres as well as rods and laths are visible. (*b*) 0.0165 M $FeCl_3$, OH:Fe ratio 2.10. Rods, and rafts made from these rods, are visible. The rods often seem to be made up of two *ca* 2 nm thick rods side-by-side.

spheres and bundles of joined up spheres. Note especially the size range of these objects.

Changing the anion to chloride (Murphy, Posner & Quirk, 1976b) did not affect the initial spheres but the subsequent development was different (figure 6.1*b*). Here $\beta$-FeO(OH) was the crystalline product. With perchlorate as anion (Murphy, Posner & Quirk, 1976c) laths of yet a third polymorph of ferric oxide-hydroxide were formed – lepidocrocite – as well as goethite rods. (The crystal structures of these polymorphs will be described shortly.)

Spiro *et al.* (1966) had previously noted polymer spheres of about 7 nm

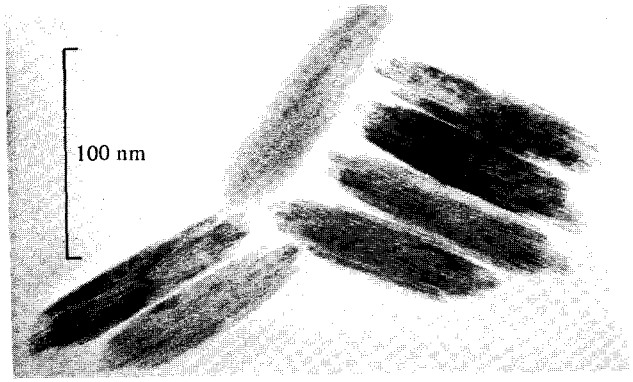



Figure 6.2. Transmission electron micrograph of $\beta$-FeO(OH) 'somatoids' formed at room temperature from 0.01 M $FeCl_3$ solution after 14 days. (From Gildawie, 1977.)

diameter (mol. wt about 240000) in similar experiments with ferric nitrate neutralised with bicarbonate. When separated from other solution components, these spheres were apparently indefinitely stable at ordinary temperatures. In a subsequent study (Brady *et al.*, 1968) the iron atoms in the polymer were found to be tetrahedrally co-ordinated with oxo as well as hydroxo bridges between them. In another series of studies, Dousma & de Bruyn (1976, 1978) describe the formation of initial particles (again of a remarkably definite size – here around 4 nm) and discuss mechanisms for their formation through olation and oxolation, their subsequent clustering and the linking of these clusters into chains.

Matijevic & Scheiner (1978) prepared a number of sols of haematite ($\alpha$-$Fe_2O_3$) with variously shaped particles (*ca* 200 nm) by ageing ferric solutions for 1–3 days at 100 °C. Again, both shapes and sizes were very uniform within a given preparation. Preparations of 'monodisperse' sols from aluminium, titanium, chromium, iron, copper and thorium are described by Matijevic (1976).

Micromorphologies of $\beta$-FeO(OH) have attracted attention for some time (Mackay, 1964). Spindle-shaped crystals obtained from dilute ferric chloride solutions are particularly striking (figure 6.2). These 'somatoids', as they are called, vary from 100–500 nm long and 20–100 nm diameter according to the conditions of preparation, but within a given preparation they are again very uniform in size and shape. The somatoids tend to aggregate into sheets with their long axes perpendicular to the sheets.

Nucleation by crystalline phases can have a profound effect on the course of ferric ion polymerisation (Atkinson, Posner & Quirk, 1977). The



number of initial nuclei can determine whether an indefinitely stable sol or a murky precipitate forms (Hsu, 1972). Silica can inhibit the formation of $\gamma$-FeO(OH) during oxidation of ferrous chloride solutions (Schwertmann & Thalmann, 1976). And organic molecules can have an important effect: oxalate can accelerate the crystallisation of haematite from 'amorphous' iron hydroxides (Fischer & Schwertmann, 1975). These authors suggest that chelated ferric ions act as templates for crystal growth. On the other hand, traces of inorganic cation impurities in minute early haematite crystals were found to inhibit further growth as a result of lattice distortions (Nalovic, Pedro & Janot, 1976).

What is true of ferric ions is true very generally in that the hydroxides, oxide-hydroxides and oxides that are the typical products of cation polymerisation in aqueous solution may seem often 'amorphous' or 'cryptocrystalline'. But such words are expressions of ignorance of what are complex but far from chaotic colloidal products.

### *Hydroxide crystal structures*

*Brucite* (magnesium hydroxide, $Mg(OH)_2$) is an important archetype structure. Here, two-dimensional polymers are stacked on top of each other. A ball-and-spoke representation of the structure of one of these polymer layers is shown in figure 6.3*a*. One can see it as a final outcome of olation of $Mg(H_2O)_6^{2+}$ (cf. scheme on page 167), where all the hydroxyl oxygens are co-ordinated with three magnesium atoms. The hydroxyl bonds are normal to the plane of the layers which thus have knobbly surfaces. These key together in the stacked layers – the hydrogen atoms on adjacent surfaces are in contact in such a way that any hydrogen atom in any layer is nested into the surface niche created by a triplet of hydrogens on an adjacent layer. Alternatively, a brucite layer can be imagined as two close-packed sheets of spheres, representing the hydroxyl groups, with the much smaller magnesium atoms filling all the octahedral sites thus created (figure 6.3*b*). Hydroxides of $Ca^{2+}$, $Mn^{2+}$, $Fe^{2+}$, $Co^{2+}$ and $Ni^{2+}$ have essentially the same structure as brucite. $Fe^{2+}$ and $Mn^{2+}$ frequently replace $Mg^{2+}$ in the lattice to the extent of about one or two per thousand. (Cation substitution in minerals is so general that I will not remark on it in the following unless it is particularly extensive.) Brucite most commonly occurs as a metamorphic alteration product of *periclase* (MgO) in contact-metamorphosed dolomites and limestones. It is also formed **hydrothermally** (i.e. with water at high temperatures and pressures) in serpentine veins.

*Gibbsite*, or hydrargillite (aluminium hydroxide, $Al(OH)_3$), has a similar

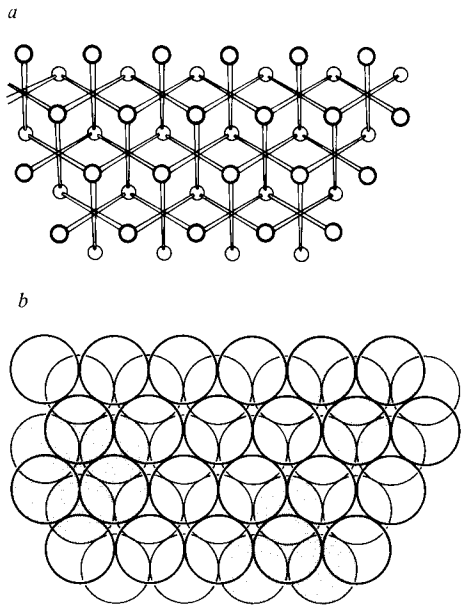


Figure 6.3. Two representations of a brucite layer. (*a*) Upper and lower planes of spheres represent hydroxyl groups. Magnesium ions are located at spoke intersections. (*b*) Two closest packed planes of spheres have **octahedral sites** between them, that is, niches with six surrounding spheres (like those shaded on the left). In brucite the spheres represent hydroxyls, and magnesium ions fill all the octahedral sites. Closest packing of spheres also generates two kinds of somewhat smaller sites surrounded by four oxygens (like those shaded on the right), but none of these **tetrahedral sites** is occupied in brucite.

structure to brucite except that only two of the three octahedral sites are occupied (figure 6.4). This can be seen as an end result of olation in which hydroxyls are only co-ordinated twice to metal atoms. Also, the stacking of the layers is different from that in brucite: the oxygen atoms at the base of one layer are directly above those on the top of the layer beneath. Only half the hydrogen atoms point towards the adjacent layers so that complete interlayer hydrogen bonding is possible. Chromic oxide-hydroxide has a somewhat similar structure except that the octahedral sites are here fully occupied and the hydrogens, of which there are only half as many, all form interlayer hydrogen bonds.

Along with boehmite and diaspore, discussed below, gibbsite is a major constituent of bauxites. It is often a late weathering product of alumino-silicate minerals and found particularly in zones that have been subject to



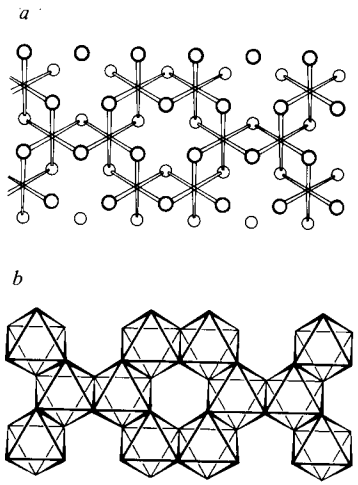


Figure 6.4. Two representations of the ideal gibbsite structure. (*a*) A ball-and-spoke model (cf. figure 6.3*a*). Here only two out of three of the octahedral positions are occupied by aluminium ions. Otherwise similar to brucite. (*b*) Alternatively a gibbsite layer can be represented as a flat network of octahedra joined through shared edges.

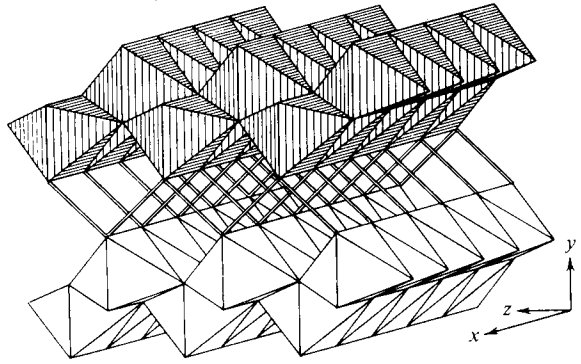



Figure 6.5. Representation of boehmite ($\gamma$-AlO(OH)) using solid octahedra (cf. figure 6.4*b*). (From Deer, Howie & Zussman, 1962, after Ewing, 1935a).

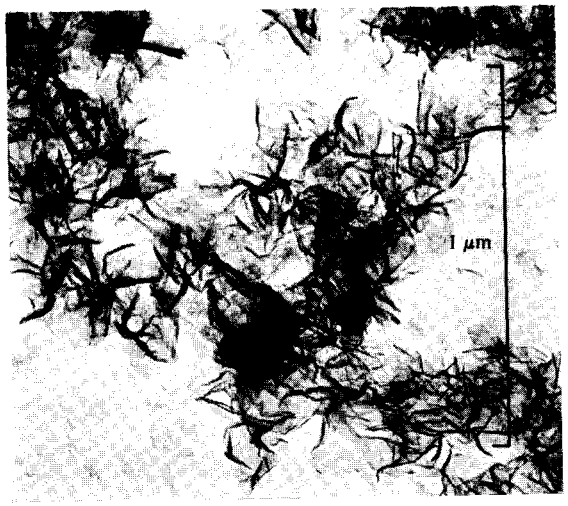



Figure 6.6. Transmission electron micrograph showing crumpled and rolled up sheets in boehmite that had been prepared by refluxing 0.002 M $Al(NO_3)_3$ solution for 24 days and then ageing for 11 days at room temperature. (From Gildawie, 1977.)

intense weathering. (Roughly speaking, the silica has been washed away.) On the other hand, gibbsite may be a transitory early weathering product too. Busenberg found that feldspar suspended in distilled water under 1 atmosphere of $CO_2$ gave a solution saturated with gibbsite within 6 minutes (see Clemency, 1976): also, Smith & Hem (1972) have shown that cryptocrystalline gibbsite can form fairly rapidly in certain solutions becoming microcrystalline in one or two months.

*Boehmite* (aluminium oxide-hydroxide, γ-AlO(OH)) has a different kind of layer structure. The aluminium atoms are located in a double sheet of distorted octahedral sites within the layers which are stacked on each other through oblique hydrogen bonds (figure 6.5). This structure can be seen as a product of both olation and oxolation. The layered structure of boehmite is vividly suggested by the crumpled and rolled-up crystals shown in figure 6.6.

*Lepidocrocite* (ferric oxide-hydroxide, γ-FeO(OH)) has a similar structure to boehmite. It occurs under oxidising conditions as a weathering product of iron-bearing minerals.

*Diaspore* (aluminium oxide-hydroxide, α-AlO(OH)) is different again. This is not a layer structure but consists of double strips of octahedrally co-ordinated aluminium atoms running in the *z*-direction (figure 6.7). All



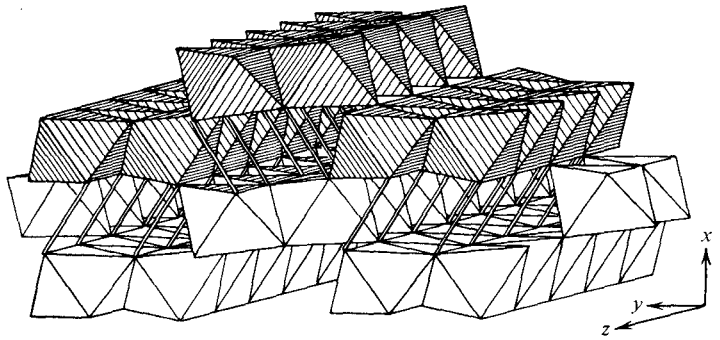



Figure 6.7. Representation of diaspore (α-AlO(OH)) using solid octahedra. (From Deer, Howie & Zussman, 1962, after Ewing, 1935b.)

the oxygens are involved in hydrogen bonds which are in paired rows between the strips.

*Goethite* (α-FeO(OH)) has a similar structure to diaspore. Goethite is very common and occurs as a weathering product of iron-bearing minerals. As already discussed, it can be obtained artificially from ferric salts. Goethite appears to be stable in water between 0 °C and 100 °C.

Other minerals with a diaspore-like structure are *groutite* (MnO(OH)), *montrosite* ((Fe, V)O(OH)), and *para-montrosite* ($VO_2$).

*Akaganéite* (β-FeO(OH)) is uncommon in Nature, having only comparatively recently been identified as a mineral (Mackay, 1964). Its main interest lies in its frequent formation from ferric chloride solutions as already discussed. Its crystal structure (Mackay, 1960) is similar to diaspore in that it contains chains of double octahedra. The chains are arranged differently, however, leaving channels containing chloride ions and water molecules (figure 6.8.).

*δ-FeO(OH)* is yet another polymorph of ferric oxide-hydroxide. It is obtainable by rapid oxidation of ferrous hydroxide (Francombe & Rooksby, 1959).

The structures and interrelationships of various iron oxides and hydroxides have been discussed in detail by Bernal, Dasgupta & Mackay (1959) and by Misawa, Hashimoto & Shimodaira (1974).

### *Metal oxides*

Almost all metal oxides are densely bonded three-dimensional crystalline polymers. Many of the common mineral oxides can be repre-

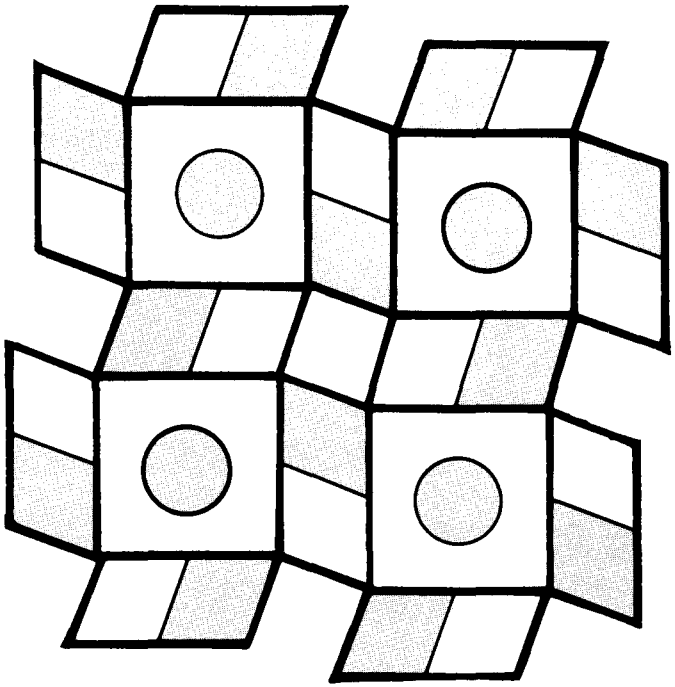


Figure 6.8. $\beta$-FeO(OH) has a similar structure to $\alpha$-$MnO_2$ which is shown here. (Data from Bystrom & Bystrom, 1950.) Strings of paired octahedra (cf. figure 6.7) are represented as parallelograms. These form channels that hold chloride ions and water (large circles).

sented as a succession of close-packed layers of oxygen spheres with cations occupying some of the octahedral sites (but between *all* the layers, thus differing from true layer structures such as brucite). *Corundum* ($\alpha$-$Al_2O_3$) and the similar *haematite* ($\alpha$-$Fe_2O_3$) are like this. So are the complex and various families of *spinels* (e.g. spinel itself, $MgAl_2O_4$) except that here tetrahedral as well as octahedral sites are partly occupied by divalent and trivalent cations respectively. The arrangement is shown in figure 6.9. *Inverse spinels* have trivalent cations in the tetrahedral sites and an equal mix of divalent and trivalent ions in the octahedral positions. *Magnetite* ($Fe^{2+}Fe^{3+}_2O_4$) is an important member of this class.

One can formally imagine metal oxides as final oxolation products from the polymerisation and cross-linking of aquo cations. And, indeed, some of the above mineral metal oxides may sometimes form in this way, for example haematite and magnetite. And many metal oxides such as $Cu_2O$, CuO, $Ag_2O$, AgO, HgO, SnO, PbO and $PbO_2$ are obtainable as products



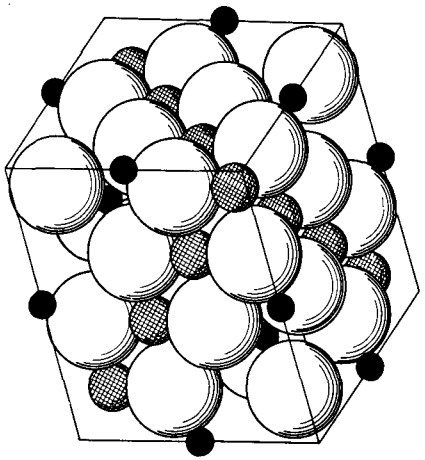


Figure 6.9. Spinel structure. Black spheres are in tetrahedral and hatched spheres in octahedral sites between the larger oxygen spheres. (From Verwey & Heilmann, 1947.)

of cation polymerisation from aqueous solutions in the laboratory (Feitknecht & Schindler, 1963). Many such products are initially colloidal, metastable and 'amorphous' or 'cryptocrystalline' – which means that, from our point of view, they look interesting.

There is a great variety of manganese oxides, and many occur in soils and deep sea sediments, and are typical constituents of clays (Chukhrov, Gorshkov, Rudnitskaya, Beresovskaya & Sivtsov, 1980). Little is known about how they are formed in Nature. Schwertmann (1979) has discussed this question in a short review of the genesis of mineral oxides, oxyhydroxides and hydroxides under ambient conditions.

Schwertmann considers also titanium oxides associated with clays – a topic of particular interest for our later discussions on photosynthesis. Among $TiO_2$ forms, *anatase* appears to be the main weathering product. The high-temperature polymorph *rutile* is derived mainly from metamorphic rocks, although it does sometimes crystallise from solution (see also Force, 1980). The mixed oxide *pseudorutile* ($Fe_2Ti_3O_9$) occurs as a weathering product, and a whole range of mixed iron and titanium coprecipitates can be formed in the laboratory and matured under quite mild conditions (70 °C for 70 days at pH 5.5) to give iron-containing rutile and anatase phases (Fitzpatrick, le Roux & Schwertmann, 1978).

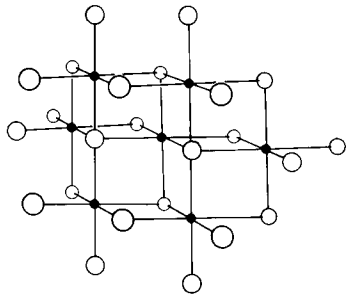


Figure 6.10. Ball-and-spoke model of the paramolybdate anion (open circles, oxygen).

*Metal sulphides*

Rickard (1969) discusses the formation of iron sulphides at low temperatures and pressures. All six of the known mineral forms can be synthesised via $Fe^{2+}$ in aqueous solutions with sulphide, polysulphide, or sulphur. These six minerals are *pyrite* and *marcasite* ($FeS_2$); *pyrrhotite* ($Fe_{(1-x)}S$, where $x$ lies between 0 and 0.25); *smythite* and *greigite* ($Fe_3S_4$); and *mackinawite* $FeS_{(1-x)}$. Pyrite and pyrrhotite are the stable phases, but metastable phases, especially mackinawite, are easily formed (Taylor, 1980). Iron sulphides, like many other sulphides, readily form colloidal suspensions.

In Chapter 8, we will take up the idea that sulphides might have taken part in electron-transfer processes in primitive forms of photosynthesis (Granick, 1957, 1965) and that perhaps colloidal iron sulphides were the ancestors of present day iron–sulphur redox proteins (Hall, Cammack & Rao, 1974).

**Anion polymerisation**

Whereas aquo cations tend to polymerise when the pH's of their solutions are raised, a number of oxyanions form polynuclear species in solution when the pH is lowered. Molybdenum and tungsten have a rich oligo-anion chemistry. $MoO_4^{2-}$, for example, gives, under a wide range of conditions, the paramolybdate anion $Mo_7O_{24}^{2-}$ with little sign of intermediate species. The structure of this ion is shown in figure 6.10. The paratungstate ion, $W_{12}O_{42}^{12-}$, has a similarly compact structure.

Germanium (IV) in aqueous solution can form oligo-oxyanions containing eight or so germanium atoms. On the other hand, tin (IV) is monomeric in alkaline solution but yields the very insoluble stannic oxide ($SnO_2$) in the neutral pH ranges (8.5–4.0) (Johnson & Kraus, 1959).



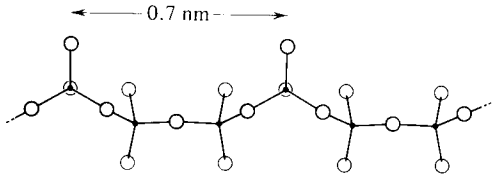



Figure 6.11. Conformation of the polyphosphate chain in Maddrell's salt. (From data in Thilo, 1962.)

Boron and aluminium are somewhat similarly related: borate oligomerises to a limited extent in aqueous solution while the aluminium oxyanion (probably $Al(OH)_4^-$) remains essentially as monomer in alkaline solution. Like tin (IV), aluminium precipitates insoluble products (this time hydroxides) in the middle pH range.

*Phosphate*

A number of salts contain infinite polyphosphate chains in the solid, for example Maddrell's salt II (figure 6.11). Others, for example Graham's salt, contain branched chains; still others cyclic structures (Van Wazer, 1958). Also phosphorus pentoxide exists in the solid either as the oligomer $P_4O_{10}$ or as infinite three-dimensional polymer. These materials are formed under anhydrous conditions, however; none is thermodynamically stable in water. Here they all tend to depolymerise to orthophosphate. Branched phosphate chains are the most quickly degraded, but straight chains may, under suitable conditions, remain metastable for a considerable time in aqueous solution. (This property is made use of by all modern organisms in the triphosphate chain of ATP; and there are bacteria that store phosphorus as polyphosphate.)

Table 6.1 lists factors that influence the rate of depolymerisation of condensed phosphates. Figure 6.12 illustrates in particular the effect of pH on the hydrolysis of the linear trimer. Figure 6.13 gives an idea of the time scales involved in the degradation of high-polyphosphate polymers. Even after three weeks in this experiment, only about 50 % of the polymer was degraded as far as orthophosphate. That was at 60 °C, pH 5. With no catalysts, and at room temperature and neutral pH, the rate of hydrolysis of polyphosphate chains is usually to be measured in years.

This ability of polyphosphate to persist kinetically against the thermodynamic odds is quite exceptional for polyoxyanions (Van Wazer, 1958, p. 452); indeed, it is exceptional for aqueous solutions of inorganic polynuclear species generally. The polyphosphates are the nearest things to

Table 6.1. *Factors influencing the rate of hydrolysis of condensed phosphates (from Van Wazer, 1958, p. 453)*

| Factor | Approximate effect on rate |
|---|---|
| Temperature | $10^5$–$10^6$ times faster from freezing to boiling |
| pH | $10^3$–$10^4$ times slower from strong acid to base |
| Enzymes | As much as $10^5$–$10^6$ faster |
| Colloidal gels | As much as $10^4$–$10^5$ faster |
| Complexing cations | Very many-fold faster in most cases |
| Concentration | Roughly proportional |
| Ionic environment in solution | Several-fold change |

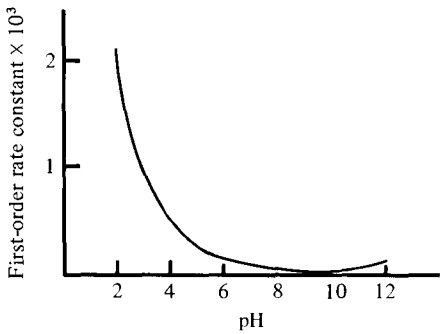



Figure 6.12. Effect of pH on the rate of hydrolysis of sodium triphosphate. (From Crowther & Westman, 1954.)

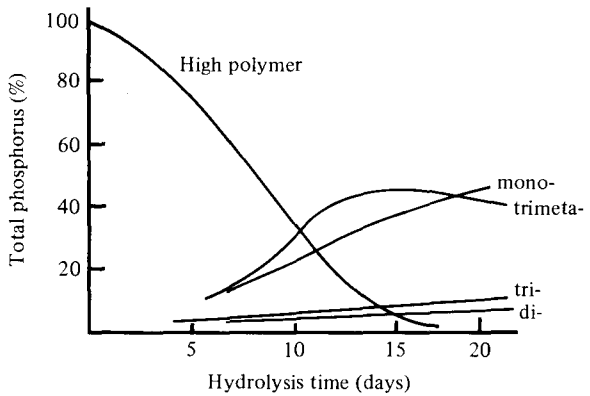



Figure 6.13. Degradation of high polyphosphate polymers at pH 5. (From Thilo & Wieker, 1957.)



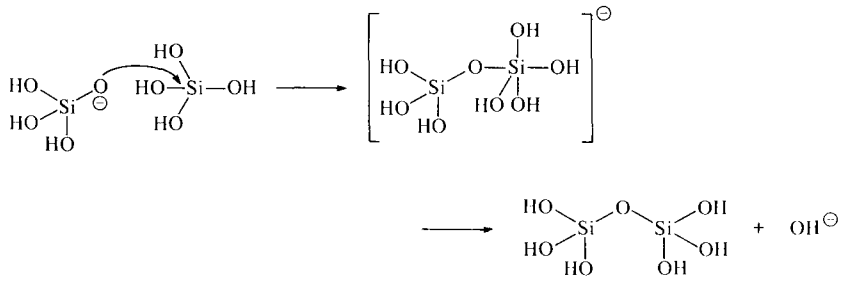
organic biopolymers that we have yet come across in this chapter. Polyarsenates do not share this property of indefinite metastability in water. (Nor do arsenate–phosphate copolymers – we have yet to come across a soluble inorganic polymer that might be able to hold much information.)

### *Silicate*

Like the corresponding phosphates, sodium polysilicate $(Na_2SiO_3)_n$ is prepared under anhydrous conditions. It dissolves in water to give viscous concentrated alkaline solutions. On dilution, smaller oligomers are found, and below a concentration of about 1 mM the species present are essentially monomeric. On acidification, a solution of sodium silicate gives the hydrated oxide – as silica sol or silica gel. Since this process can be reversed, it would appear that such polymeric and oligomeric species as exist in solution are reversibly formed and degraded (Van Wazer, 1958, p. 449). Silicate is unlike phosphate both in precipitating an oxide from aqueous solution (one can hardly imagine $P_2O_5$ forming this way) and having a more dynamic population of polynuclear species in solution. Phosphate remains exceptional.

Again in a way that is becoming familiar, it is at the next structural levels that metastability appears. At room temperature, silica sols may remain metastable for a considerable time: often they develop into gels of $SiO_2 \cdot nH_2O$ that mature further losing some of the chemically bound water. But, at these temperatures, and on laboratory time scales, they never arrive at the thermodynamic ideal – quartz.

Although the extent of sodium silicate polymerisation increases with reducing pH, the rate has a maximum around pH 8–9 (figure 6.14). This is close to the pH (around 9.5) at which silicic acid and its mono anion are present in equal concentrations: it is therefore reasonable to postulate a bimolecular nucleophilic mechanism for the polymerisation (Gimblett, 1963). One possibility would be:



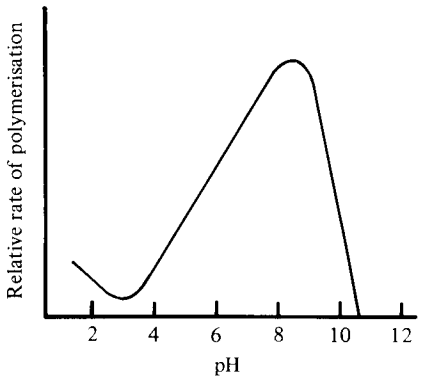



Figure 6.14. The influence of pH on the rate of silicate polymerisation. (From Gimblett, 1963.)

Unlike the polyphosphates, branched polysilicates are not specially disfavoured. This, and the tetravalence of the silicic acid unit, allows complex, densely cross-linked polymers. Presumably, the network becomes so rigid that reactions such as the one above can no longer take place because they are too sterically demanding. Thus unreactable Si–OH groups remain inside the structure.

Even at neutral pH and ordinary temperatures, silica is appreciably soluble in water. For example, Wey & Siffert (1962) found that different forms of amorphous silica in contact with water gave solutions of 120–140 parts per million (p.p.m.) of silica within a week or so. The silica was present in solution as monomeric silicic acid (Alexander, Heston & Iler, 1954). Silica is far more soluble at around neutral pH than most of the oxides, oxide-hydroxides and hydroxides that we discussed under cation polymerisation. This has geochemical implications that we will return to. Figure 6.15 shows how the solubility of silica and of alumina changes with pH.

We can perhaps understand why silicic acid monomer is the dominant solution species in quasi-equilibrium with silica. A mature silica particle will be highly cross-linked so that even surface silicic acid units will be multiply bonded and thus much more difficult to detach than they would be if singly bonded (one can think of analogies with the chelate effect or the relatively much greater stability of multiply base-paired oligonucleotides). Monomeric silicic acid units will nevertheless come away from time to time, but dimers or trimers would require greater numbers of simultaneous bond breakages to do likewise. Hence, at dynamic equilibrium, the



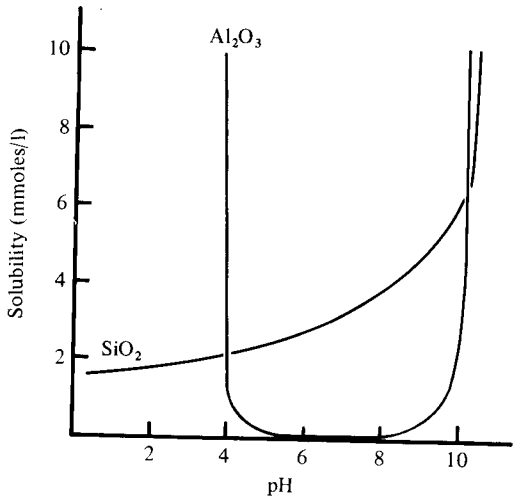



Figure 6.15. Effect of pH on solubility of silica and alumina at ordinary temperatures. (From Mason, 1966.)

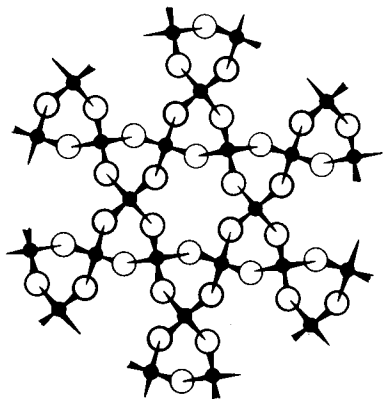


Figure 6.16. Plan of the structure of $\beta$-quartz. (From Wells, 1975.) Small black circles represent silicon atoms. The oxygen atoms lie at different heights above the plane of the paper, those nearest the reader being drawn with the heaviest lines. Each atom is repeated at a certain distance above (and below) the plane of the paper along the normal to that plane so that the $Si_3O_3$ rings in the plan represent helical chains.

supply of solution species will be almost exclusively of monomers. Some of these, it is true, will dimerise, trimerise and so on; and there will thus be a small concentration of small oligomeric species. But the bigger these become the more easily will they be mopped up by mature silica particles with which they can the more easily make multiple bonds. This argument would not apply to ideal linear polymers with no kind of cross-linking or folding since then, say, a 100-mer might be as likely to break into a 61-mer and a 39-mer as a 99-mer and a monomer. Here a smoother equilibrium distribution would be expected.

The crystalline forms of silica are considerably less soluble – about 27 p.p.m. for cristobalite (Fournier & Rowe, 1962) and about 6 p.p.m. for the more common stable form, quartz, at 25 °C (Morey, Fournier & Rowe, 1962). Each of these is a three-dimensional network of $SiO_4$ tetrahedra, linked through all the corners. In the quartz structure (figure 6.16) one can distinguish helices of linked tetrahedra: in a given crystal the helices of a given type are either right-handed or left-handed but not both. Crystals of quartz are thus chiral. An interesting two-dimensional crystalline polymer of silicic acid has been described by Kalt, Perati & Wey (1979).

### Non-layer silicate minerals

Most of the minerals that make up the rocks in the Earth's crust are silicates. As a group they combine structural features that we have discussed under cationic and anionic polymerisation – although there are also many new features.

Silicates are generally classified according to the state of polymerisation of the formally anionic silicate component. This may extend in zero, one, two or three dimensions. We will consider all but the two-dimensional type in this section.

#### *Orthosilicates*

These are the simplest class from the above point of view: here the silicate units are unpolymerised. *Olivines* are examples. They have the general formula $M_2SiO_4$ where M is a divalent metal. For example, *forsterite* is $Mg_2SiO_4$, and *fayalite* $Fe_2SiO_4$. Between these there is a whole series $(Mg, Fe)_2\ SiO_4$ with different proportions of $Mg^{2+}$ and $Fe^{2+}$. Manganese, calcium and small amounts of many other metals are also found in olivines. The idealised structure is given in figure 6.17. The divalent ions occupy octahedral sites between silicate tetrahedra. Alternatively, the



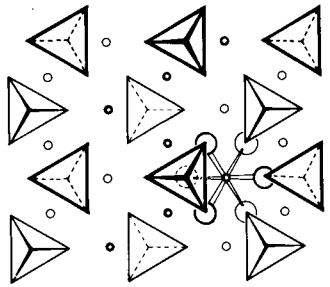


Figure 6.17. Structure of olivine as determined by Bragg & Brown (1926). $SiO_4$ groups are represented as tetrahedra. Here two sets are shown with their bases at levels 1/4 and 3/4. $Mg^{2+}$, $Fe^{2+}$, or other divalent cations are the small circles. Two sets of these ions are shown here at levels 0 and 1/2. The octahedral co-ordination of one of these ions is shown in more detail: here the large circles are oxygen atoms, and the dashed circle is an oxygen atom of a tetrahedron below the one shown.

structure may be seen as an infinite array of approximately hexagonally close packed oxygen spheres with silicon in one-eighth of the tetrahedral niches and metal in half of the octahedral niches.

Olivine is a major constituent of the volcanic basalt rocks. It is among the most easily weathered primary minerals. That the silicate units in it are not self-polymerised may have something to do with olivine's relative water solubility, but it is clearly not the whole story since other orthosilicates, for example the *garnets*, $M_3^{2+}M_2^{3+}(SiO_4)_3$, are very stable to weathering. So is *zircon*, $ZrSiO_4$. In spite of the separation of the silicate tetrahedra these orthosilicates are really three-dimensional polymers if we remember their metal–oxygen bonding. We may note that the oxides of tri- and tetravalent metals are typically very much less soluble in water, much less inclined to depolymerise, than silica.

#### *Chain silicates*

The *pyroxenes* are the commonest iron–magnesium rock-forming silicates. Their structures contain linear siloxane chains with a two-repeat conformation (figure 6.18*a*). The chains have a truncated triangular cross-section (figure 6.18*b*) and are held together by cations between them (figure 6.18*c*). For example, in *diopside*, $CaMg[Si_2O_6]$ there are magnesium and calcium ions in six-fold and eight-fold co-ordination respectively.

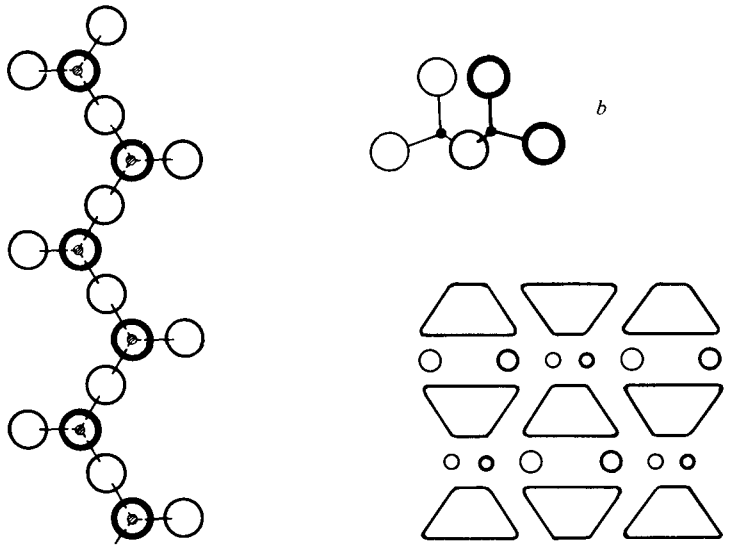



Figure 6.18. Pyroxene structure. (*a*) Top view of a pyroxene chain. Large circles are oxygen atoms at two different levels and the small shaded circles are concealed silicon atoms. (*b*) End view of a chain (silicon atoms as small black circles). (*c*) Packing of chains viewed as in (*b*) with cations between them in eight-fold and six-fold co-ordination (larger and smaller circles).

In keeping with their fiery name, the pyroxenes are high-temperature minerals occurring in almost every type of igneous rock. Many, such as diopside, also occur in metamorphic rocks. Although diopside may be formed hydrothermally in rock fissures there is little evidence for pyroxenes being synthesised from aqueous solutions at ordinary temperatures and pressures.

*Band silicates*

*Amphiboles* represent another group of minerals that are widespread in igneous and metamorphic rocks. These structures are based on pairs of fused pyroxene chains (figure 6.19). In addition to the formal polysilicate anions there are hydroxyls (or fluorides) that fit into the gaps created by the linking of the two chains. Tremolite, $Ca_2Mg_5[Si_8O_{22}](OH, F)_2$, is an example. Another is riebeckite, $(Na_2Fe^{2+}_3, Fe^{3+}_2)[Si_8O_{22}](OH, F)_2$, indicating the variety of cation patterns possible. The hornblende group are still more complex, for example *common hornblende*: $(Ca, Na, K)_{2-3}(Mg, Fe^{2+}, Fe^{3+}, Al)_5[Si_6(Si, Al)_2O_{22}](OH, F)$. Amphiboles often occur as fibrous forms: for example a fibrous form of riebeckite – *crocidolite* – otherwise known as blue asbestos. There is some evidence for formation of this mineral at moderate temperatures and pressures (Deer, Howie & Zussman, 1962, Vol. II, p. 348). On the whole, amphiboles form under a wider range of conditions than pyroxenes and may sometimes crystallise from aqueous solution, but as in the case of the pyroxenes, there is little evidence for their formation at ordinary temperatures and pressures.



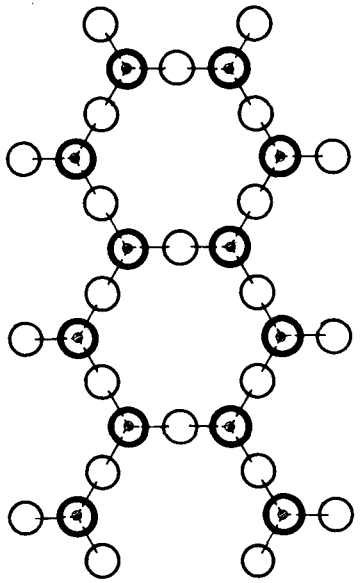


Figure 6.19. Top view of an amphibole band (cf. figure 6.18).

*Framework silicates*

*Quartz* and *cristobalite*, which we touched on earlier, are members of this group of silicates. Quartz and cristobalite each consist of an uncharged siloxane scaffolding. More usually the three-dimensional frames are anionic, containing also aluminium, and they imprison, more or less effectively, alkali or alkaline earth counterions.

*Feldspars* In *feldspars* the counterions are held most firmly. There are two main classes: the *alkali feldspars*, $(K, Na)[AlSi_3O_8]$, and the

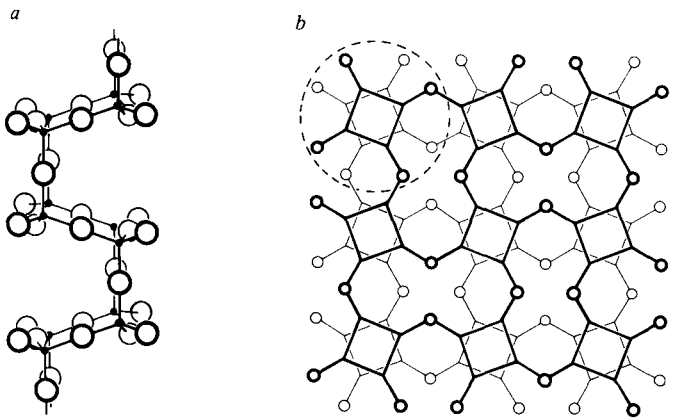



Figure 6.20. The feldspar framework is made out of $SiO_4$ and $AlO_4$ tetrahedra, the latter giving the framework a negative charge. The structure can be thought of as consisting of rings of four tetrahedra joined to each other in columns which are in turn connected laterally. (*a*) Is a side view of an idealised column showing tilting of successive rings (small black circles, Si *or* Al). (*b*) Encircled is a simplified top view of a column of the type shown in (*a*) revealing also a stagger between successive rings. The whole picture shows the numerous sideways connections between the columns that completes the three-dimensional frame, and the relatively large cavities that hold ions such as $K^+$, $Na^+$ or $Ca^{2+}$ that neutralise the negative charges of the frame. (After Taylor, 1933.)

*plagioclases* which include *albite*, $Na[AlSi_3O_8]$, and *anorthite*, $Ca[Al_2Si_2O_8]$, as end members of a complete series with intermediate compositions. Always the number of aluminium atoms in the frame corresponds to the number of charges on the imprisoned cations, since each aluminium substitution creates one negative charge in the frame and there are no other anions present. The aluminium substitutions may be ordered as in *microcline*, $K[AlSi_3O_8]$, or disordered as in its high temperature polymorph *sanidine*. The general structure of feldspar is described in figure 6.20 and its legend.

Feldspars are the major constituents of most igneous rocks and are by far the most abundant minerals in the Earth's crust. They occur also in metamorphic and sedimentary rocks. In the latter, they may be detrital or they may have crystallised from pore solutions – sometimes at ordinary pressures and temperatures.

*Feldspathoids* will tend to crystallise from a magma, in place of feldspars, where there is a deficiency of silica. Examples are *leucite*, $K[AlSi_2O_6]$, and *nepheline*, $Na_3K[Al_4Si_4O_{16}]$. In some members of this group, anions as well as cations are held in the alumino-siloxane framework: for example in *sodalite*, $Na_8[Al_6Si_6O_{24}](Cl_2)$. The cavities are larger in the feldspathoids than in the feldspars and they intercommunicate. The ions within them can thus exchange without disrupting the framework (figure 6.21).



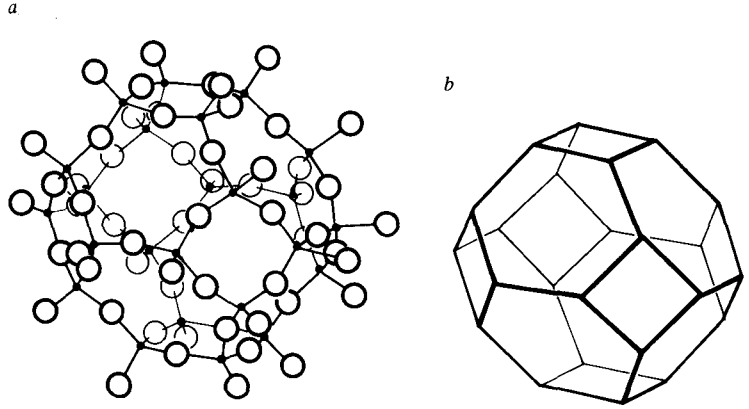



Figure 6.21. (*a*) The building unit of the sodalite framework. (After Bragg, 1937.) Large circles are oxygen atoms; small circles silicon *or* aluminium atoms. (*b*) A simplified view. Here lines are Si–O–Si bonds (etc.) with silicon (or aluminium) atoms at corners. The structure can be thought of as consisting of such bricks closely packed.

*Analcite* (*or analcime*), $Na[AlSi_2O_6]\cdot H_2O$, is something between a feldspathoid and a zeolite. Its framework structure is illustrated in figure 6.22.

*Zeolites* The *zeolites* have a still more open structure than the feldspathoids and always include water molecules in the channels which, like the cations, are mobile and can be freely exchanged. Indeed, zeolites are commercially important ion-exchange materials.

Although three-dimensional polymers, the zeolites frequently have fibrous or platey morphology, visible either macroscopically or microscopically, and which is related to their crystal structure on the one hand and conditions of formation on the other.

The *natrolite group* are characteristically fibrous. The reason for this can be seen from the structure of natrolite itself, $Na_2[Al_2Si_3O_{10}]\cdot 2H_2O$. This consists of multiply bonded polymer chains with relatively few cross-links between them (figure 6.23). Although there are narrow channels in the

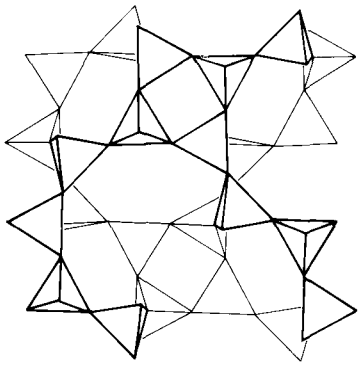


Figure 6.22. Pattern of connections between tetrahedra in the framework of analcite. (From Deer, Howie & Zussman, 1962.)

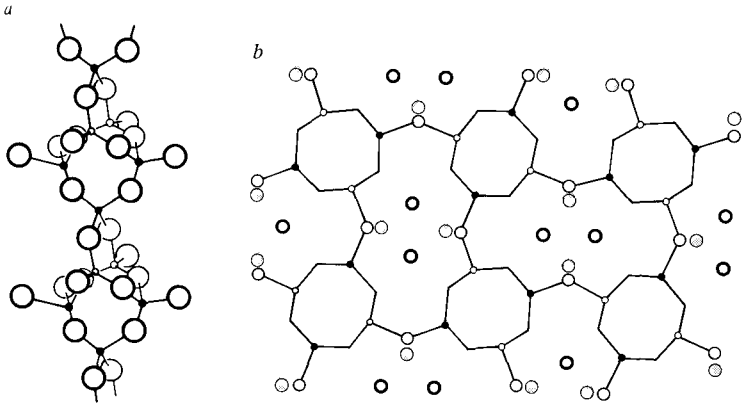


Figure 6.23. Like feldspar (figure 6.20), natrolite can be thought of as a set of polymer chains cross-linked laterally. The unit chain of natrolite is shown in (*a*) (after Bragg, 1937). Small black circles are silicon atoms and open circles aluminium atoms. (*b*) Shows the (relatively few) sideways connections and the locations of sodium ions (heavy circles) and water molecules (hatched circles).

fibre (*z*) axis, ion exchange in natrolite is more likely to take place through larger channels that run in directions perpendicular to *z* (Meier, 1960).

The alumino-siloxane frames in these minerals are evidently fairly flexible, the dimensions of the unit cells depending on the kind of cations present and on the state of hydration (Hey, 1932).

The closely related pair, *heulandite* and *clinoptilolite*, (Ca, $Na_2$)[$Al_2Si_7$



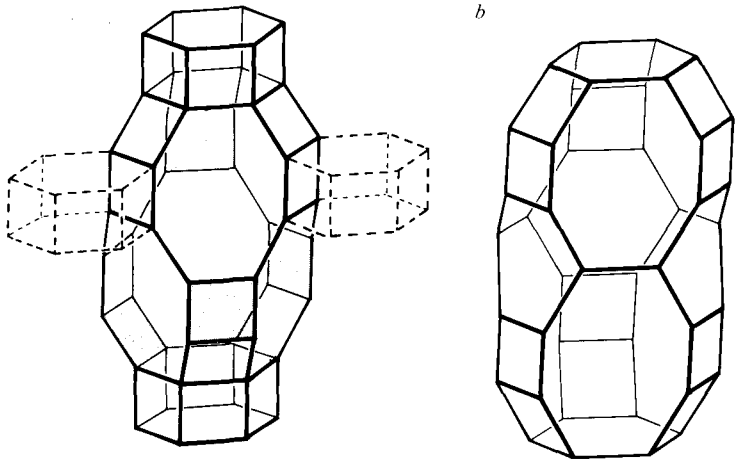


Figure 6.24. (*a*) Chabazite building brick (cf. figure 6.21) showing mode of connection with two of the adjacent units. Similar connections are made with four other units via shaded squares. (After Dent & Smith, 1958.) (*b*) Building brick of erionite frame. (After Barrer & Kerr, 1959.)

$O_{18}$]·$6H_2O$, are examples of platey zeolites. Two-thirds of the (Si, Al)$O_4$ tetrahedra in the frame are linked in nets of six-membered rings that are joined via the remaining tetrahedra to give a three-dimensional structure which has two other sets of channels between the nets.

Like analcite, *phillipsite*, ($\frac{1}{2}$Ca, Na, K)$_3$[$Al_3Si_5O_{16}$]·$6H_2O$, *harmotome*, Ba[$Al_2Si_6O_{16}$]·$6H_2O$, *chabazite*, *gmelinite* and *levyne*, (Ca, $Na_2$)[$Al_2Si_4O_{12}$]·$6H_2O$), have more even distributions of bonding. They are sometimes described as molecular sieves on account of their ability to hold small molecules within their channels. Thus phillipsite picks up $NH_3$ and $CO_2$ while chabazite absorbs straight-chain but not branched paraffins. *Erionite* ($Na_2K_2Ca$, Mg)$_{4.5}$[$Al_9Si_{27}O_{72}$]·$27H_2O$ is similar to chabazite but has a more fibrous morphology. Figure 6.24 illustrates the cage-like structural units of chabazite and erionite which are typical of this group.

*Mordenite*, ($Na_2$, $K_2$, Ca)[$Al_2Si_{10}O_{24}$]·$7H_2O$, is also frequently fibrous but with a different kind of structure based on chains containing five-membered rings of tetrahedra. It has a system of channels with a 0.66 nm free diameter that runs parallel to the fibre axis. There are also smaller channels in this direction and others (0.28 nm) that run at right angles (see figure 6.25).

*Faujasite* (approx. ($Na_2$, Ca)$_{1.75}$[$Al_{3.5}Si_{8.5}O_{24}$]·$16H_2O$) is still more impressive, containing channels that have a minimum diameter of 0.9 nm.

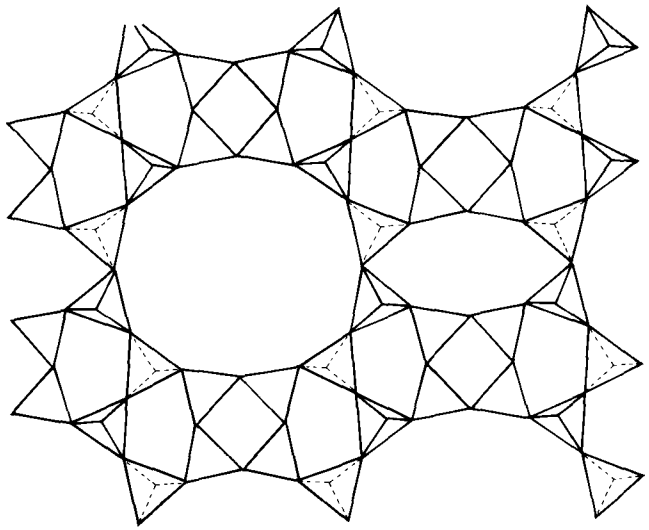


Figure 6.25. Mordenite has large (0.66 nm diameter) channels which run parallel to its fibre axis. These, as well as smaller channels, can be seen in this projection. Other (0.28 nm) channels intersect at right angles – they would run vertically in this picture. (After Gottardi & Meier, 1963.)

Not only normal paraffins, but molecules such as cyclohexane and benzene can enter its structure. A detailed account of the sorption of ions and molecules by zeolites is given by Barrer (1978).

### *Synthesis of framework silicates from aqueous solutions*

Zeolites are comparatively easily synthesised in the laboratory. In a typical procedure an alumino–silica gel is heated under pressure with alkaline aqueous solutions containing ions such as $K^+$, $Na^+$, etc. In this way, many natural and a number of quite new zeolites have been made. In Nature too, hydrothermal synthesis of zeolites at comparatively low temperatures is common. Often they are found in fissures and solidified gas bubble cavities of basic volcanic rocks, having crystallised there from aqueous solutions at elevated temperatures.

Quite rapid laboratory syntheses of zeolites at somewhat lower temperatures have also been described – for example at 130 °C (De Kimpe, 1976) and 90 °C (Barrer & Sieber, 1977). Indeed, an inadvertent synthesis of phillipsite and chabazite at still lower temperatures was reported more

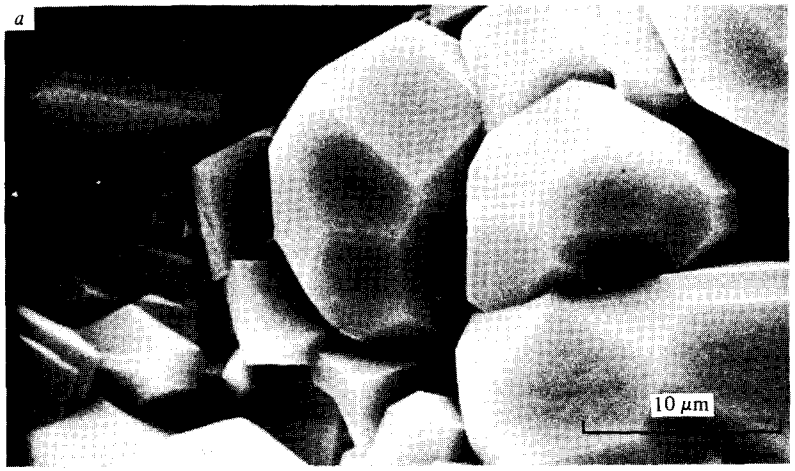


than a hundred years ago – in the wall work of Roman baths where hot springs were emerging at a maximum temperature of 70 °C (Daubrée, 1879).

Zeolites form in vast quantities within sedimentary rocks. (Minerals that have formed *in situ* like this are said to be **authigenic**). Figure 6.26 illustrates the varied morphologies of some zeolites that crystallised naturally in this way. Authigenic synthesis of zeolites can occur at ordinary temperatures and pressures and sometimes, by geological standards, very rapidly. Time scales for the development of massive zeolitic deposits range from a few hundred to tens of millions of years. High concentrations of $K^+$ and $Na^+$ and high pH are important factors favouring rapid synthesis. (For a review of zeolites and zeolitic reactions in sedimentary rocks, see Hay, 1966.)

Sheppard & Gude (1973) describe in detail a typical situation for low-temperature zeolite synthesis – in tuffs in the *ca* 30 square mile 'Big Sandy Formation', Arizona. (A tuff is a rock formed originally from compacted volcanic ash – here in an ancient alkaline saline lake.) Although the glassy texture of the starting material was often preserved, the material itself was completely altered into zeolites (mainly analcite, clinoptilolite, erionite and chabazite, with smaller amounts of phillipsite, mordenite, and harmotome) as well as potassium feldspar, layer silicate clays and silica (for example quartz and the non-crystalline silica mineral, opal). Authigenic layer silicate clays were found almost everywhere: montmorillonite was particularly abundant and consistently the earliest mineral to form. Analcite and potassium feldspar were in some places the only framework species, but most zeolitic tuffs consisted of two or more zeolites. The zeolites had evidently formed by solution and recrystallisation of the original volcanic glass – except for analcite which seems to have formed from zeolitic precursors. The potassium feldspar also crystallised indirectly – from analcite as well as from other zeolites. Differences in the mineral compositions of different regions is interpreted in terms of differences in pore water composition, arising in the first place from differences in conditions under which the original tuffaceous rock was deposited in the ancient lake. Thus nearest to the lake margin conditions were less saline and alkaline, and non-analcite zeolites were favoured in soils formed here; nearer the centre a rising salinity and pH (*ca* 9+) encouraged rocks that transformed to analcite and potassium feldspar. All of the above processes must have taken place at or near ordinary temperatures and pressures since none of the tuffaceous rock had been deeply buried.

Hay (1966) lists, according to age, 30 locations where authigenic zeolites



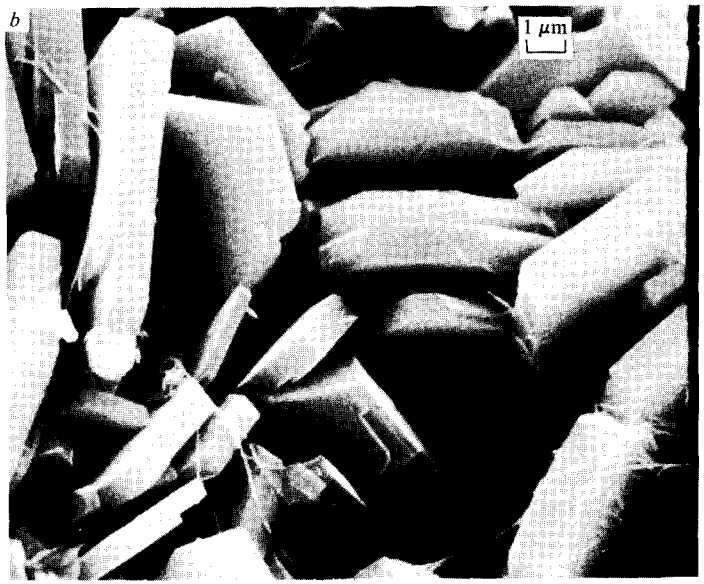
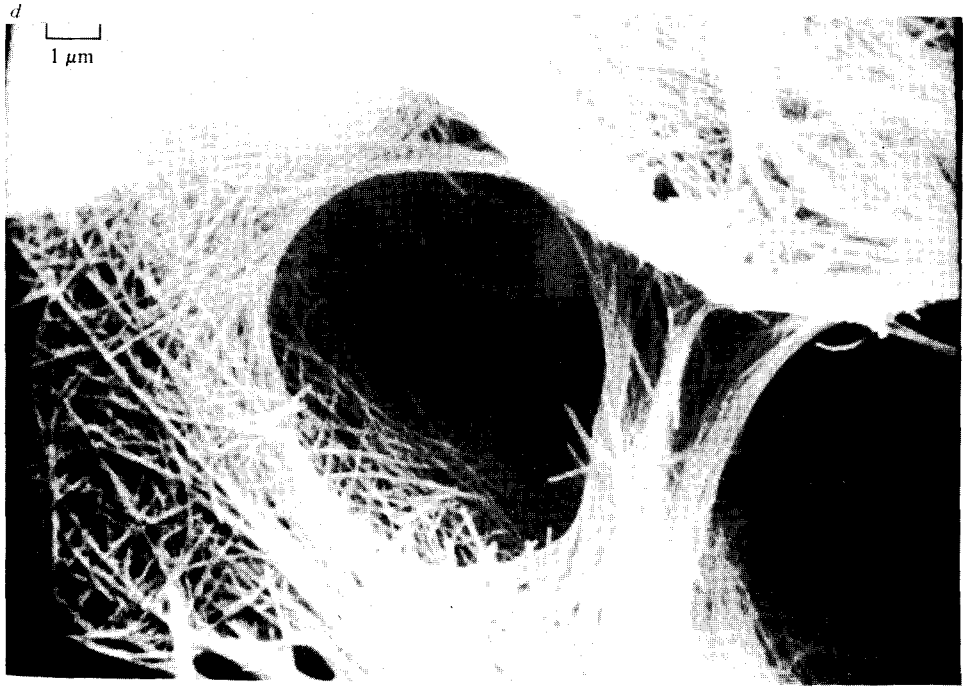




Figure 6.26. Scanning electron micrographs of zeolites in sedimentary rocks. (From Mumpton & Ormsby, 1976.) (*a*) Analcite blocks. (*b*) Clinoptilolite plates with a few threads of mordenite. (*c*) Erionite bundles. (*d*) A 'rat's nest' of interlaced mordenite fibres.



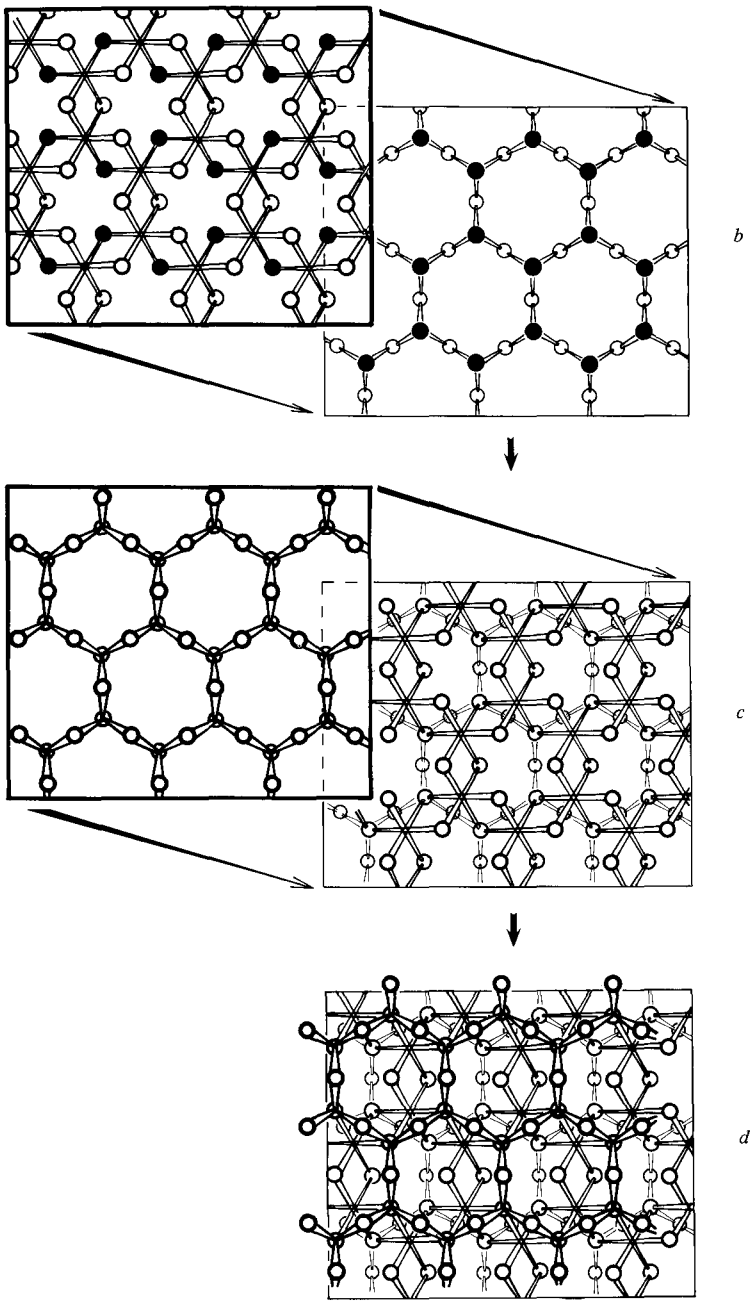



Figure 6.27.



and associated silicate minerals have been found in deposits of saline alkaline non-marine environments. He lists also 89 marine and fresh-water locations. Tuffs are common but not invariable host rocks. The youngest marine occurrence reported is in the gulf of Naples (Muller, 1961) in the top 10–15 cm of sediment thought to have been deposited over the last 500 years, and consisting of volcanic ash and its alteration products – which make up about 50 % of the material. Some of this consists of very fine-grained analcite, and (usually microscopic) authigenic quartz crystals, but the best crystallised of these products are not framework silicates at all but layer silicates, including kaolinite crystals of up to 1 mm long, and mica with short pseudo-hexagonal crystals as much as 1.5 mm maximum diameter. Which brings us to the next topic.

### Layer silicates

Like chain and band silicates, the layer silicates are a fusion of two polymer types – here two-dimensional metal oxy-hydroxy and siloxane types. The mineral silicates that we have considered so far have been three-dimensional polymers with covalent bonds of one sort or another extending throughout the structure. In the layer silicates, the covalence is usually restricted to two dimensions. As a result, the layer silicates contain clearly defined, more or less separable two-dimensional macromolecules.

Two of the main structural types of silicate layers can be imagined as being built up, as shown in figure 6.27, by fusing an extended siloxane net either onto one side or onto each side of a gibbsite sheet. This would give the ideal (1 : 1) kaolinite and (2 : 1) pyrophyllite layer structures respectively. These are described as dioctahedral layer silicates because, as in gibbsite, only two of the three octahedral sites are occupied. (Side views, showing the stacking of kaolinite and pyrophyllite layers, are given in figure 6.28.)

Figure 6.27. Formal construction of silicate unit layers. Starting with a gibbsite sheet (*a*) remove two-thirds of the hydroxyls (shaded) from the lower side and plug in instead the apical oxygens of an extended amphibole-like siloxane network. Such a network is shown in (*b*) with the apical oxygens shaded. This would give the 1:1 dioctahedral (kaolinite) layer type (*c*). Performing a similar operation also on the upper surface of the gibbsite sheet would lead to the 2:1 dioctahedral (pyrophyllite) structure (*d*). Repeating these operations, but starting from brucite instead of gibbsite, would lead to the corresponding trioctahedral archetypes (serpentine and talc). In these models the balls are either oxygens or hydroxyls and cations are located at spoke intersections. The real structures are considerably distorted from these ideals.

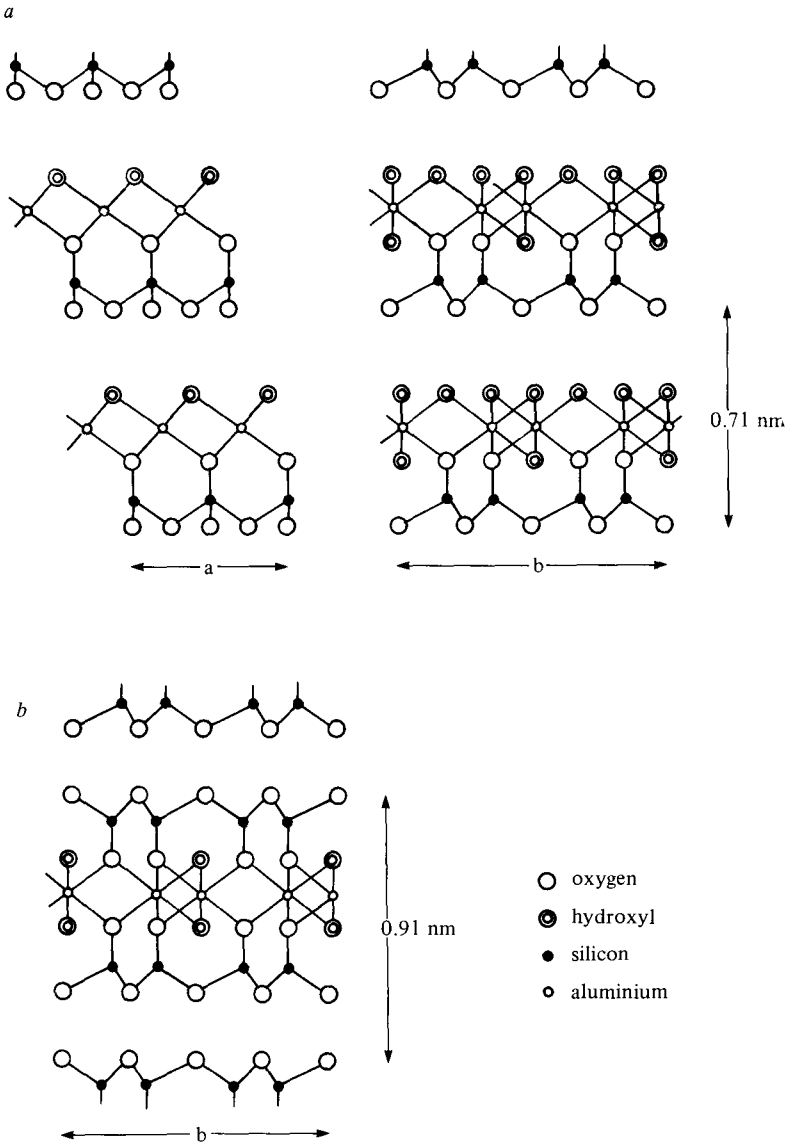



Figure 6.28. (*a*) The stacking of ideal kaolinite layers as seen along the *y*- and *x*-directions respectively. (After Brindley, 1951.) (*b*) The stacking of ideal pyrophyllite layers as seen along the *x*-direction.



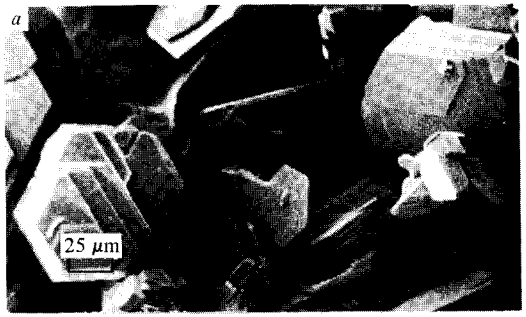
Table 6.2. *A classification of layer silicates (see text)*

| | DIOCTAHEDRAL | TRIOCTAHEDRAL |
|---|---|---|
| 1:1 | A<br>*Kaolinite clays:*<br>kaolinite<br>dickite<br>halloysite | B<br>*Serpentines:*<br>chrysotile<br>lizardite<br>antigorite<br>amesite<br>greenalite<br>berthierine<br>cronstedtite |
| 2:1 | C<br>pyrophyllite<br>muscovite mica<br>margarite | D<br>talc<br>phlogopite mica<br>biotite mica |
| | | E:<br>chlorite |
| F: Vermiculite, illite and smectite clays | most illites<br>*Smectites:*<br>montmorillonite<br>beidellite<br>nontronite | vermiculite<br>some illites<br>*Smectites:*<br>saponite<br>hectorite<br>sauconite |

The corresponding trioctahedral silicate layers, based on brucite, are those of serpentine and talc. These four major structural classes of layer silicates are:

1:1 dioctahedral
1:1 trioctahedral
2:1 dioctahedral
2:1 trioctahedral

Table 6.2 lists examples of these types. A typical layer silicate crystal consists of a stack of such units either directly in contact with each other (e.g. kaolinite itself or talc); or with intervening water molecules (halloysite); or with intervening cations (e.g. micas), or both (e.g. montmorillonite). Alternatively, a brucite-like layer may regularly interleave with 2:1 unit layers (chlorite). In addition, different kinds of silicate layers may





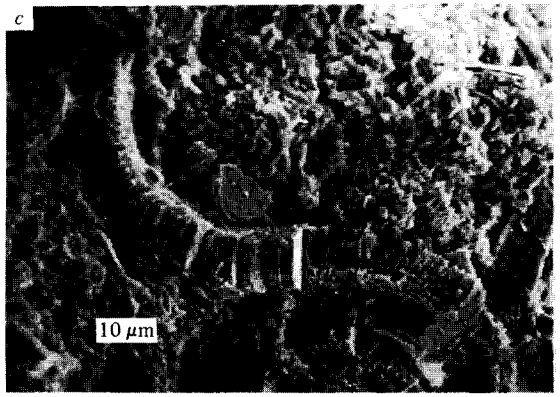
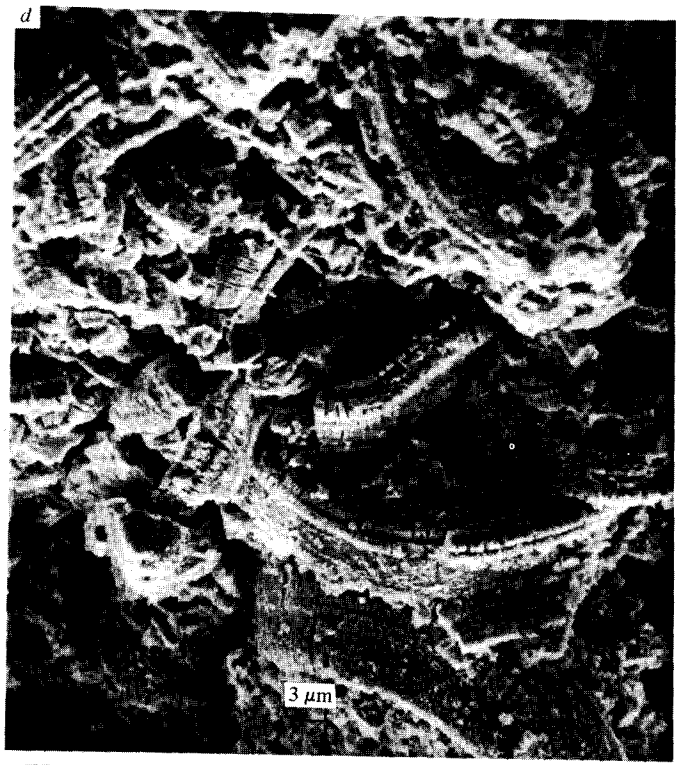
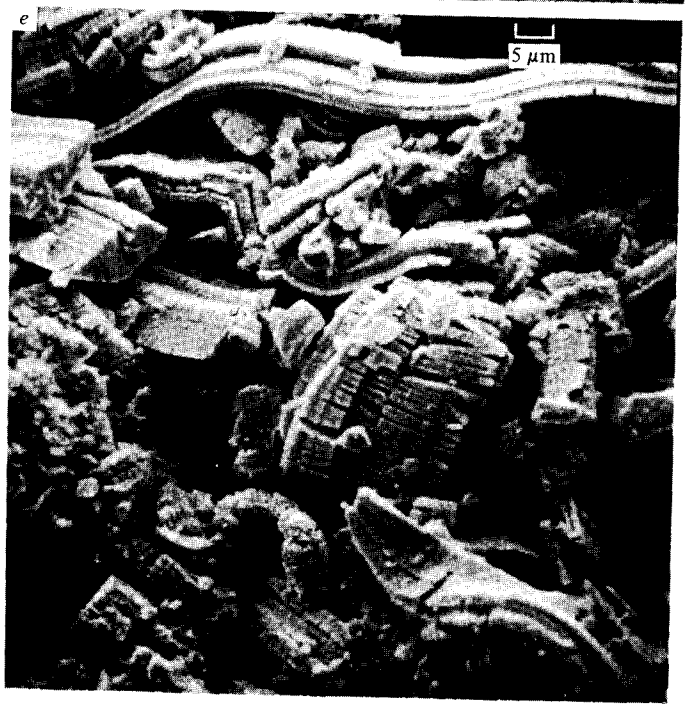



Figure 6.29. Scanning electron micrographs of kaolinite minerals. (From Keller & Hanson, 1975, and Keller, 1976, 1977.) (*a*) Dickite from Anglesey, Wales. (*b*) Kaolinite from Keokuk, Iowa. (*c*) S-shaped 'book' or 'vermiform' of kaolinite from Hodges mine, Georgia Kaolin Co. (*d*) Grooved vermiforms of kaolinite from Jacal, Mexico. (*e*) Kaolinite vermiforms and books showing complex grooving.





interleave, either regularly or irregularly. Even where the layers are identical, they may stack in different ways. And real layers are always distorted out of the ideals of figure 6.27. Add to all this the possibilities inherent in cation substitutions within the layers as well as between them – and the possibilities, often, of organic molecules between the layers – and one has an impression of a complex subject of study. This impression is correct.

The layer silicates in Table 6.2 are divided into the four main structural classes as well as showing a cross-group (E) and a cross-group (F). We will follow the sequence A, B, C (except F), D (except E and F), E, F.

### A. *The kaolinite group of clays*, $Al_4[Si_4O_{10}](OH)_8$

*Kaolinite* itself is the major constituent of kaolin (china clay) where it occurs as massive deposits in a limited number of localities. But it is also widely dispersed in Nature and clays of the kaolinite group are possibly the commonest of all clay minerals. Kaolinites form either hydrothermally or as a weathering product of other silicates, particularly feldspars. Typical examples are shown in figures 6.29*a* and *b*, illustrating well made microcrystalline plates. 'Book' and 'vermiform' (worm-like) morphologies are often found in low-temperature products (Keller & Hanson, 1975) (see figures 6.29*c*, *d*, and *e*). These pictures vividly reflect a crystal structure consisting of molecular layers stacked on top of each other. The thinnest plates in figure 6.29 contain about 200 unit layers of the kind shown in figures 6.27*c* and 6.28*a*. The unit layers are held together by hydrogen bonding between hydroxyls on one side and siloxane oxygens on the other.

The unit layers in kaolinite clays are considerably distorted from the ideals of figures 6.27*c* and 6.28*a*. There are a number of reasons for this. For one thing, the apical oxygen atoms in an unconstrained siloxane net would be somewhat further apart than the corresponding array of oxygen atoms in an ideal gibbsite sheet – that is to say, the two sets of shaded atoms in figure 6.27*a* and *b* would not match up exactly for size. As a result, the siloxane net is compressed. This can happen quite easily through alternate contrary rotations of the tetrahedra (we will return to this idea). But the array of connecting oxygen atoms on the gibbsite sheet would not only be undersized, it would be distorted. The gibbsite structure shown is an idealisation: in fact the aluminium-occupied sites are considerably smaller than the vacant sites. As a result, some of the oxygen atoms are actually too far apart. Pairs of tetrahedra connected to such pairs of oxygen atoms are tilted towards each other. These effects can be seen readily in the kaolinite mineral, *dickite* (figure 6.30).



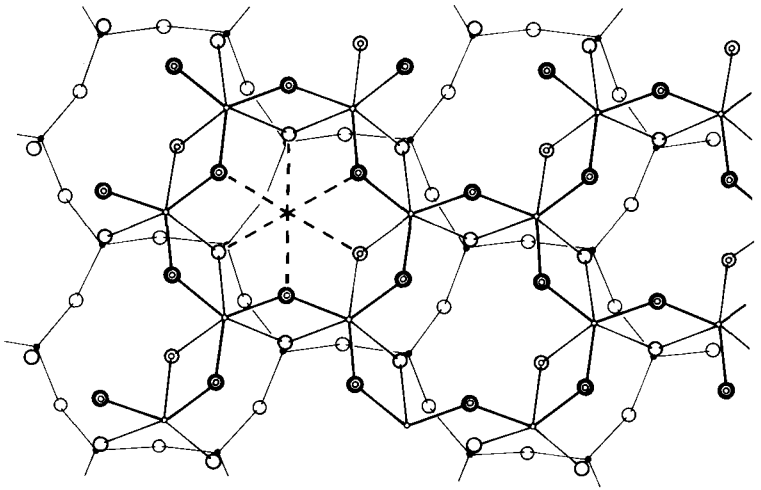
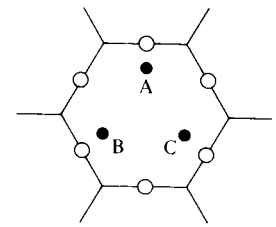


Figure 6.30. Dickite illustrates divergences from the ideal kaolinite layer shown in figure 6.27*c*. Note the distortions of the octahedral environments of the aluminium atoms. One of the (much larger and more symmetrical) octahedral vacancies is indicated by dashed lines. There are both rotations and tiltings of the $SiO_4$ tetrahedra. (After Newnham, 1961.) Oxygens and hydroxyls are the larger open and double circles; aluminium and silicon atoms are smaller open and black circles.

The distortions in dickite and in kaolinite itself are somewhat different, the tetrahedra being rotated by 7.3° and 11.3° respectively. Also, while in dickite the surface hydroxyls are about normal to the layer, in kaolinite they are more like the hydroxyls of gibbsite, that is, one in three lies almost in the layer plane (Giese & Datta, 1973; Rouxhet, Samudacheata, Jacobs & Anton, 1977). The hydroxyls countersunk in the other surface of the layer are tilted towards the vacant octahedral sites.

Kaolinite itself and dickite are described as polytypes; that is, they differ in the way in which the unit layers are stacked on top of each other: more precisely, it is a difference in the relationship between the positions of the octahedral vacancies between adjacent layers. The dioctahedral character of the kaolinite group produces, inevitably, directionality in the unit cell: the vacancy may be in any of the positions A, B or C:



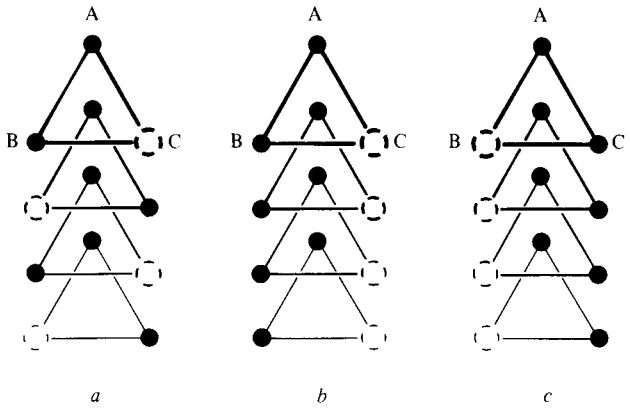



Figure 6.31. Kaolinite polytypes differ in the positions of the octahedral vacancies in successive layers. The vacancies alternate in dickite (*a*) whereas in kaolinite itself they are either all on the left (*b*) or all on the right (*c*) of the direction of the stacking off-set (see text). (Modified from Bailey, 1963.)

In kaolinite itself, the vacancies all point the same way; in dickite, they alternate. Although there are, in principle, very many stacking modes with different rotations and displacements of adjacent layers that would allow complete interlayer hydrogen bonding (Newnham, 1961), the stacking turns out to be the same in both kaolinite and dickite – successive layers being displaced along the $x$-axis by $-a/3$. By convention, they are said to be stacked in the direction of the A site (figure 6.31). This makes the three sites different – which raises new questions. Which site is vacant in kaolinite itself, and between which two do the vacancies alternate in dickite? For dickite at least there is a straight answer: between B and C (figure 6.31*a*). Of the three possibilities for kaolinite, two – all-B and all-C – are enantiomers, while the third – all-A – is achiral and would be expected to have a different energy (Newnham, 1961; Bailey, 1963; Wolfe & Giese, 1973). It is generally assumed that most samples of kaolinite consist of a mixture of enantiomeric crystals.

Unfortunately, it has not so far been possible to demonstrate the chirality or otherwise of single kaolinite crystals – crystals available are too small for present techniques (Wolfe & Giese, 1973). The investigation of a large crystal might still be inconclusive since an occasional changeover between B and C vacancy positions could make a large kaolinite crystal achiral although it contained chiral domains. 'Dickite stacking' has indeed been found in kaolinite crystals (Plançon & Tchoubar, 1977) and highly disordered kaolinites are common.



Ideally, polytypes would correspond to different ways of stacking otherwise identical layers. With kaolinite itself and dickite the situation is not as simple as this since the layers are distorted in different ways. Seemingly this difference is a cause as well as a consequence of the different stacking modes: Wolfe & Giese found that the energy was lower if a crystal that had started one way – as dickite, or as either enantiomer of kaolinite – continued the same way, although the commonness of disordered structures indicates that this is not an overriding factor in many cases.

*Halloysite* is another mineral with a kaolinite layer. Here, the layers are less well ordered in relation to each other, less firmly stuck together: indeed there is very often a sheet of water molecules between the silicate layers. This is readily detected by X-ray diffraction through an increase in the interlayer spacing from the normal 0.7 nm to 1.0 nm. With the electron microscope halloysite is often seen to be tubular. Originally thought to be a simple consequence of the misfitting of siloxane and gibbsite sheets, the origin of the tubular morphology in halloysite no longer seems so clear since contrary rotations of siloxane tetrahedra provide such an easy way of relieving the misfit. It is clear, however, that unit layers can be flexible and hence are able to roll up.

Figure 6.32 (*a* and *b*) shows electron micrographs of a tubular halloysite from Wagon Wheel Gap, Colorado, investigated by Dixon & McKee (1974). The tubes are loosely swiss-rolled packets of unit layers stacked without water between them. Narrower tubes of circular cross-section may have about five layers per packet; the larger tubes with polyhedral cross-section have about 35. Spaces between the packets can be seen in figure 6.32*a*. The internal void diameters of nine tubes measured were between 7 nm and 38 nm.

These authors also studied a halloysite from Minnesota consisting of spheroids (figure 6.32*c*). These appear often to be hollow with an internal void of about 150 nm diameter and a patchy exterior of kaolinite plates.

Kirkman (1977, 1981) prefers the term 'multifaceted squat cylinder' to describe similar halloysites from New Zealand: the 'spheroids' appear to have a more or less definite axis. This became more obvious on treatment with alkali which produced objects resembling stuffed olives. (Indeed, figure 6.32*c* shows some spheroids like this.) Among the most striking objects in Kirkmam's papers are the spiral discs shown in figure 6.33. These are very flat (perhaps as little as 20 nm thick) and were found securely jammed between silica flakes where they had evidently grown (probably over a period of some thousands of years). Often several of the discs are intergrown (figure 6.33*b*).

Nagasawa & Miyazaki (1976) describe halloysites obtained from a


Figure 6.32. Transmission electron micrographs of tubular and spheroidal halloysite. (From Dixon & McKee, 1974.) (*a*) Tubular form from Wagon Wheel Gap. (*b*) Replicas of tubes. (*c*) Replicas of spheroidal halloysite from Minnesota showing flat kaolinite plates (P).

variety of Japanese sites: some having arisen from hydrothermal alteration of volcanic glass in pumice beds, others by weathering of feldspar in granitic rocks, still others by alteration of clayey matrix of sand by circulating ground waters. Various morphologies – tubes, laths, balls, scrolls, and plates with rolled edges – are among the forms described and illustrated. All of these had interlayer water. Whatever the kind of location, the water



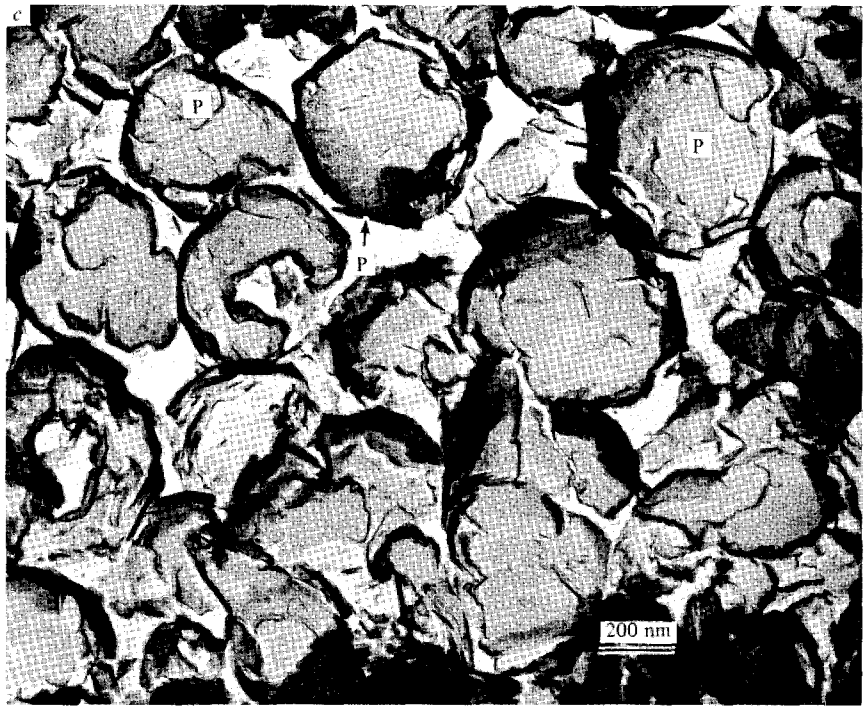



Figure 6.32(*c*).

was more easily removed from the older samples, suggesting a tendency to transform towards kaolinite itself with age. In any case, kaolinite itself was commoner as an associated mineral in older locations.

Parham (1969) simulated tropical weathering of potassium feldspar with water (at about 78 °C) in a soxhlet extractor for 140 days. Surface replicas from the feldspar were then examined under the electron microscope. Already in this time, flame-like membranes were seen to be attached to the feldspar. These were taken to be 'primitive halloysite' on account of their morphological identity to features found in naturally weathered feldspars where halloysite is the product. These membranous projections often occurred in lines suggestive of underlying crystallographic features such as dislocations. In some cases, tubes appeared to be growing out of pits in the feldspar (see also Kirkman, 1981). Similar morphological features were found by Parham on feldspar crystals that had been exposed in a road cut in the warm rainy climate of Hong Kong for less than five years.

From these results and from a survey of localities of recent weathering in which halloysite has been identified, Parham concludes that halloysite

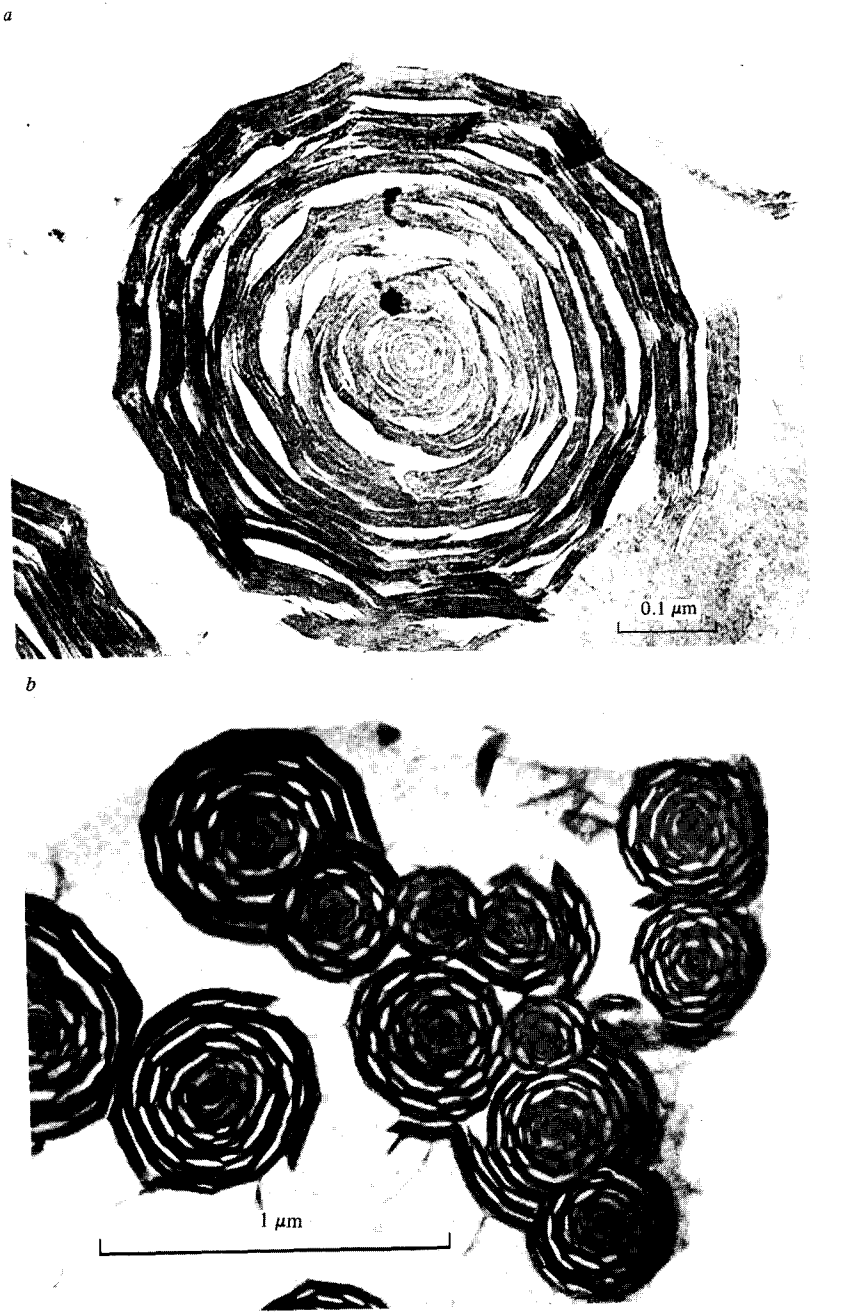



Figure 6.33. (*a*) and (*b*) Transmission electron micrographs of spiral discs of halloysite from Pahoia Tuff, New Zealand. (From Kirkman, 1977.)



tends to precede kaolinite as a weathering product. This is in line with the conclusions of Nagasawa & Miyazaki referred to above, and also with those of La Iglesia & Galen (1975) who studied the transformation of halloysite into kaolinite at room temperature in the presence of the chelating agents, oxalic acid and EDTA. Initial halloysite first recrystallised, then degraded – and then kaolinite itself appeared. Kaolinite itself can also be transformed into halloysite under low-temperature hydrothermal conditions (150 °C) (Eberl & Hower, 1975).

A beautiful illustration of the growth of clays – probably halloysite – is shown in the series of scanning electron micrographs of Keller (1977), four of which are given in figure 6.34. (See also Keller, 1976, 1978, and Keller & Hanson, 1975). Clearly these crystals are not simply bits of disintegrating feldspar – they have grown from solution.

### B. *Serpentines*

Serpentines are 1:1 trioctahedral layer silicates. The magnesium serpentines, formally derived from brucite, have a composition approximating to $Mg_3[Si_2O_5](OH)_4$. These minerals are formed in Nature principally from the alteration of other magnesium-bearing silicates, such as olivine, by migrating hot gases and solutions. Aumento (1970) discusses typical examples: serpentinisation of ultrabasic intrusions in Canada and on the mid-Atlantic ridge. A partly serpentinised rock will contain a variety of serpentines in cracks and fissures. A detailed example is given by Rimsaite (1973).

*Chrysotile*, also called 'white asbestos', and the best known of the serpentine minerals, often occurs in veins as silky fibres several inches long. This fibrous character reflects a structural relationship to the microcrystalline halloysite rather than the macroscopically more similar looking amphibole, 'blue asbestos', that we discussed earlier. Like halloysite, chrysotile has a swiss-roll structure. Here the reason is clearer than for halloysite: the octahedral sheet, this time an oversized brucite sheet, does not comfortably fit its tetrahedral siloxane net, and contrary tetrahedral rotations cannot provide a means of relief. As a result, the unit layer tends to curl – with the siloxane net on the inside.

*Lizardite* is a platey serpentine whose characteristically small particle size is possibly due to misfitting strains.

*Antigorite* has an ingenious means of distributing strain illustrated in figure 6.35. This kind of structure accounts for the great variation in the *a*-dimension of the unit cell which may be as large as 10 nm. It accounts

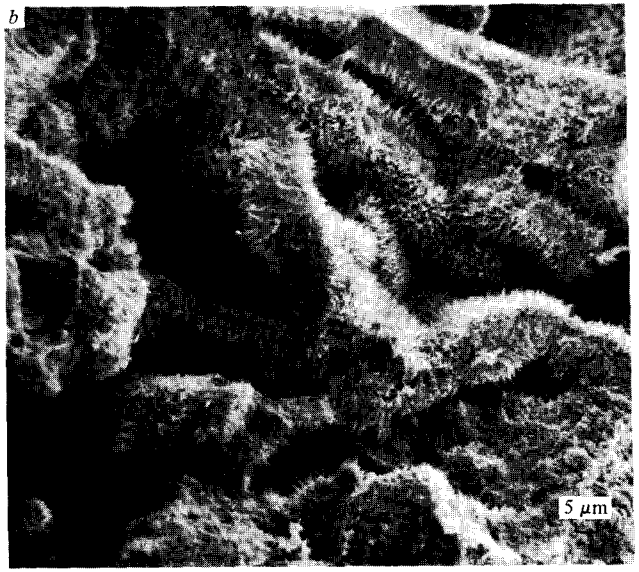
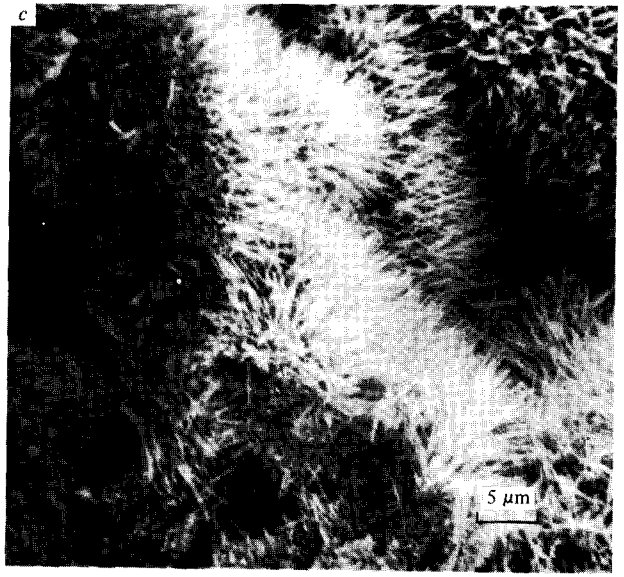




Figure 6.34. Scanning electron micrographs of clay minerals of the kaolinite group growing from solution. (From Keller, 1977.) (*a*) Clay beginning to form along microjoints in Sparta granite. (*b*) Feldspar fragments almost completely coated by elongated clay crystallites. (*c*) Closer view of (*b*). (*d*) Long clay crystals clumping like wet grass.





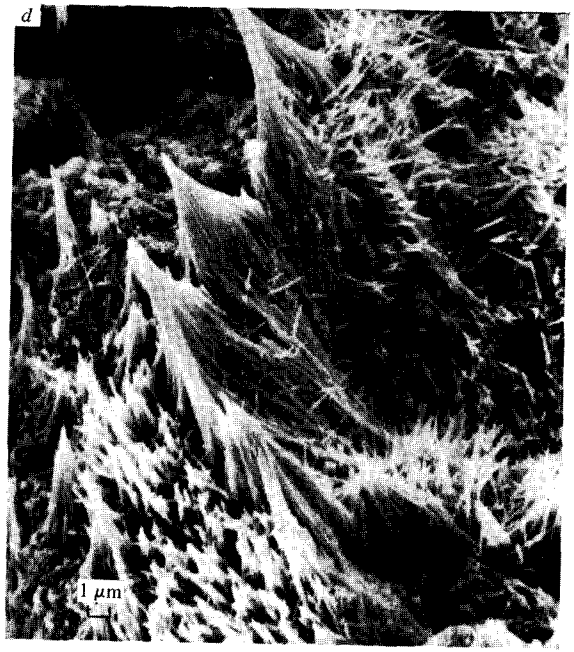




Figure 6.34.

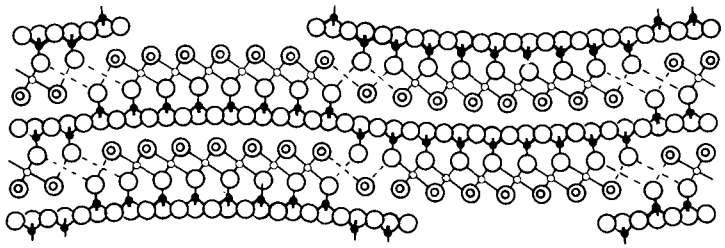


Figure 6.35. Structure of antigorite viewed along the *y*-axis. The curl imposed on the layers by an oversized octahedral sheet is accommodated by periodic inversions so that on a larger scale the sheets are more or less flat. (After Kunze, 1956.) Oxygens and hydroxyls are larger open and double circles; magnesium and silicon atoms are smaller open and black circles.

also for the existence of fibrous antigorites which are elongated in the *y*-direction. (Chrysotile, like halloysite, prefers to be rolled up along the *x*-axis.)

The whiteness of white asbestos underlines another similarity between it and the kaolinite group: there is seldom much substitution of cations in tetrahedral or octahedral positions – this in spite of the ultrabasic rocks from which serpentines are derived usually having plenty of coloured transition metal ions in them.

Exotic serpentines can be synthesised hydrothermally (at *ca* 500 °C) in the laboratory. Germanium may replace silicon for example. Interestingly, this variety is never tubular, the larger germanium atoms relieving the intersheet misfit. Non-tubular serpentine can arise also where magnesium in the octahedral sheets has been suitably partly substituted with (the smaller) aluminium (Roy & Roy, 1954).

Tubular chrysotiles have been made in which nickel, cobalt, or iron partly or completely replace magnesium. (See Jasmund, Sylla & Freund, 1976.) Serpentines with different amounts of nickel in them occur in Nature as a series between end members (Brindley & Maksimovic, 1974).

*Amesite* is a highly substituted 1:1 trioctahedral mineral with an ideal formula of $(Mg_4Al_2)[Si_2Al_2O_{10}](OH)_8$. The octahedral aluminium atoms generate positive charges that are balanced by the negative charges generated by the tetrahedral aluminium atoms. These opposite charges will tend to come near each other in the crystal structure, both within and between layers, while at the same time like charges keep their distance. There are a number of ways in which this might happen. In an amesite sample from Antarctica, Hall & Bailey (1979) found that aluminium-rich tetrahedral



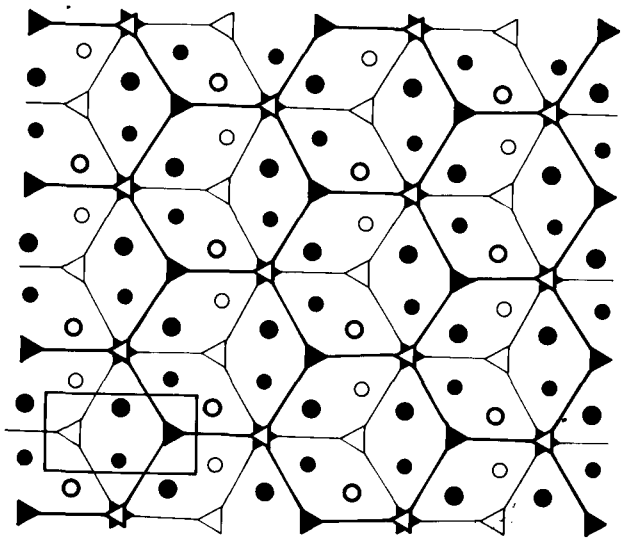


Figure 6.36. Simplified view of two layers of an amesite from Antarctica. The two alumino–siloxane nets are represented by heavy and light lines connecting triangles which are the tetrahedral positions (white, aluminium-rich and black, silicon-rich). The magnesium-rich octahedral sites are black circles with those at the lower level slightly smaller. The octahedral aluminium-rich sites at the two levels are heavy and light open circles. The rows of aluminium atoms here run in one direction. If we call this 'six o'clock' then there are five other possibilities – at 'eight o'clock', 'ten o'clock', 'twelve o'clock', etc. (From data in Hall & Bailey, 1979.) The rectangle in the lower left-hand corner indicates a class of octahedral and tetrahedral aluminium-rich sites in an amesite from the Urals described in the text.

and octahedral sites of adjacent layers were arranged in slightly zig-zag rows in one of six directions as shown in figure 6.36. These crystals were often hexagonal prisms made up of six twin sectors. It seems clear that these sectors correspond to the six different ways of arranging the rows of aluminium atoms. (We will be discussing twinning of crystals in some detail in the next chapter.)

Another kind of interlayer ordering of aluminium substitution is described by Anderson & Bailey (1981) who studied an untwinned portion of an amesite crystal from the Urals. The general structure was the same as in figure 6.36, except that here the aluminium-rich sites were arranged in left-handed helical columns parallel to the *z*-axis. In projection normal to the layer plane, the aluminium atoms lie at the corners of a diamond. The

rectangle in the lower left corner of figure 6.36 shows four sites which, for the Urals amesite, would be occupied by aluminium – as would the other identically placed sites. But again, as with the amesite from Antarctica, there are alternative geometrically equivalent arrangements – this time there are three similar ordering patterns. (There are two other similar positions – at $\pm 120°$ – for rectangles like the one in figure 6.36.)

Anderson & Bailey (1981) see this sort of arrangement as consistent with the much more complex twinning typical of the Urals amesite as compared with the Antarctic sample: there would be less lateral distortion to the crystal structure, and it would thus be easier for there to be a good lateral fit between domains that had these more subtly different arrangements.

Deer, Howie & Zussman (1962) give typical analyses for three iron-bearing 1:1 trioctahedral layer silicates:

*Greenalite:* $Fe^{2+}_{4.5}$, $Fe^{3+}_{1.0}[Si_4O_{10}](OH)_8$

*Berthierine:* (previously '7 Å chamosite'):
$Fe^{2+}_{3.6}Al_{1.6}(Mg, Fe^{3+}\text{ etc.})_{0.8}[Si_{2.6}Al_{1.4}O_{10}](OH)_8$

*Cronstedtite:* $(Fe^{2+}, Fe^{3+}, Al)_{4.6}[Fe^{3+}_{2.0}, Si_{2.0}, O_{10}](OH)_8$

Greenalite appears to be something like a ferrous analogue of chrysotile or lizardite, although in a study of synthetic hydrothermal products containing different proportions of magnesium and iron, Jasmund, Sylla & Freund (1976) found that (orthorhombic platey) greenalite was distinct from (monoclinic tubular) chrysotile in the sense that there was no single series of intermediate structures between them. Iron was unlike nickel or cobalt in this respect. Berthierine can be seen as somewhat like an amesite with $Fe^{2+}$ in place of $Mg^{2+}$: whereas in cronstedtite, iron ($Fe^{3+}$) stands in for tetrahedral $Al^{3+}$ as well.

Greenalite, cronstedtite and particularly berthierine, have characteristics that justify their being considered in a class somewhat apart from the 'classical' serpentines, chrysotile, lizardite and antigorite. Chemically, they are similar to chlorites; physically their very small particle size pushes them towards the category of clays. And they are more like clays too in the way in which they appear to be formed in Nature – often as low-temperature weathering products.

That at least is the conclusion arrived at by Harder (1978) from experiments on low-temperature synthesis (20 °C and 3 °C) of iron layer silicate minerals in which, among other things, greenalite-like and berthierine-like products were formed. Harder has found indeed that iron-bearing layer silicate minerals form rapidly from hydroxide precipitates and dilute silica



solutions if the conditions are right. The following are favourable factors: (1) Reducing conditions. ($Fe^{2+}$ must be present; reducing agents, dithionite or hydrazine were used.) (2) High pH, generally 8–9. (3) Presence of $Mg^{2+}$. (4) Low $SiO_2$ concentrations (10–20 p.p.m.). (5) A stoichiometric proportion of adsorbed silica in the hydroxide precipitates. In such circumstances, iron layer silicates became detectable by X-ray diffraction in a matter of days.

The formulae of these iron-bearing minerals are enough to indicate a considerable complexity – the partly dioctahedral character of greenalite, for example, and the necessarily strongly negatively charged tetrahedral sheets in berthierine and cronstedtite. Further insight into berthierine's structure has been obtained by Mossbauer spectroscopic studies by Yershova, Nikitina, Perfil'ev & Babeshkin (1976). They found iron in octahedral sites of two different kinds with different degrees of distortion from cubic symmetry: 65 % of the iron was in the less distorted set of sites. This iron was more readily oxidised on heating.

### C. *Dioctahedral micas*

*Muscovite*, $KAl_2[Si_3AlO_{10}](OH, F)_2$, rivals feldspar and quartz in ubiquity, if not in abundance: it can form and survive under a wide range of conditions and is found in igneous and metamorphic rocks, and in sediments either as detrital particles or having crystallised *in situ*. We will discuss the structure of mica, particularly muscovite, in some detail; other 2:1 layer silicates can then be described largely through alternative or progressive moves from this 'archetype'.

Let us recall first that the kaolinite unit layer is derived formally by fusing a siloxane net onto one side of a gibbsite sheet (figure 6.27*c*). Many of the characteristic features of the kaolinite group of clays arise more or less directly from the asymmetry of this arrangement and through the possibility of extensive interlayer hydrogen bonding. By having a siloxane net on both sides of a gibbsite sheet, as in pyrophyllite (figure 6.27*d*) these effects are eliminated – the resulting material has (more or less) flat layers only weakly held together by van der Waals forces. Pyrophyllite has indeed a somewhat slippery feel reminiscent of graphite. It is not really typical of 2:1 layer silicates however. Usually, there are quite strong interlayer forces of a kind different from the direct hydrogen bonding interactions in the kaolinite group. Muscovite shows these (electrovalent) forces at work in a relatively straightforward way.

In muscovite, as in many micas, about one in four of the silicon atoms of

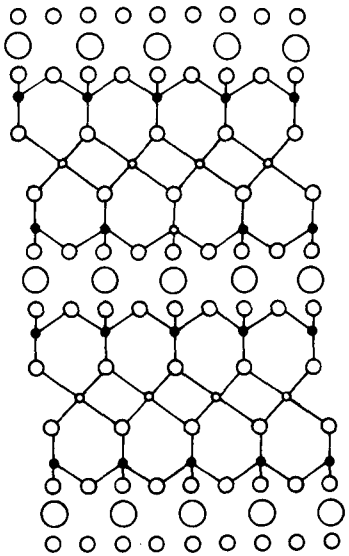


Figure 6.37. Side view of mica along the *y*-axis. Large, middle and small open circles are respectively potassium ions, oxygens, and and octahedral cations. Black circles are silicon *or* aluminium atoms in tetrahedral positions.

the siloxane nets are substituted by aluminium. As in the framework silicates each substitution creates one negative charge. The muscovite layers are thus strongly anionic. They also have arrays of indentations in the six-ring siloxane nets that cover their surfaces. As it happens, the radius of the potassium ion is such that it can nest fairly well between two such indentations in layers above it and below it. The muscovite crystal thus consists of 2:1 unit layers held together by potassium ions between them. The extent of substitution of aluminium for silicon is such that, with the interlayer sites fully occupied, the negative charge in the layers is exactly balanced. An idealised side view of mica layers is given in figure 6.37: in muscovite, two of the three octahedral positions are occupied by $Al^{3+}$.

As a result of the keying effect of the potassium ions, the upper siloxane net of one layer lies almost exactly opposite the lower net of the layer above it – there is no *inter*layer displacement in the ideal structure and little in real structures. There is, inevitably, an *intra*layer displacement however: it is not possible to put siloxane nets exactly opposite each other on each side of a gibbsite sheet. The intralayer displacement (of $a/3$) can be seen in figures 6.37 and 6.38.



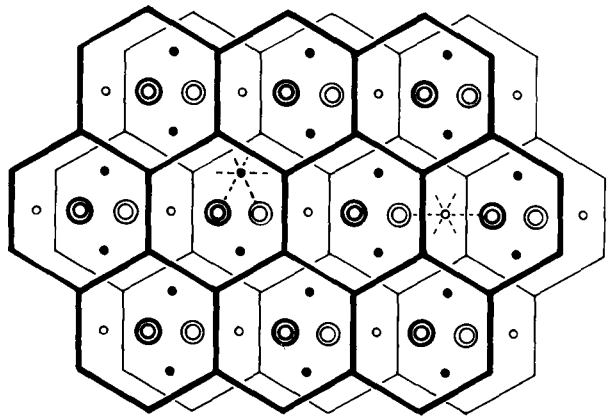


Figure 6.38. Simplified top view of a mica layer showing displacement, by $a/3$, of the upper from the lower alumino-siloxane nets (to the left in this picture – but there are six possibilities). Double circles are hydroxyls. The central plane of octahedral sites is represented by small circles of which there are two kinds, designated M(1) and M(2) (open and black circles respectively). (After Smith & Yoder, 1956.)

*Polytypism in micas* This intralayer displacement immediately provides possibilities for polytypism. The unit layer of a mica can be said to have an arrow in it pointing in the direction of the displacement of the upper siloxane net in relation to the lower net (see particularly figure 6.38). The question arises as to how the arrows in a stack of layers are related to each other. If the keying requirements of the potassium ions was the only factor to be optimised there would seem to be an immense number of possibilities for a multilayer crystal. In the stacking of ideal layers each can be placed equally well in one of six orientations, obtained by rotations of multiples of 60°. A crystallite with, say, 1000 layers in it would be one of $6^{999}$ (i.e. about $10^{777}$) possibilities – all of equal energy. In the event, most muscovite crystals are found to be stacked in only one of a few ways – most often in only one way.

Whatever the explanation for this regularity of stacking, it is reasonable to look for rules. Smith & Yoder (1956) assumed that the angle between the 'arrows' would most likely remain the same in successive layers. On the further assumption that these angles are either *always* in the same sense or *always* alternate, Smith & Yoder showed that there would be six preferred stacking modes for a crystal such as muscovite, however many layers it might contain; that is, there are six particularly simple polytypes. Figure 6.39 displays these possibilities. It has turned out that regularly stacked 2:1 layer silicates very often fall within this set.

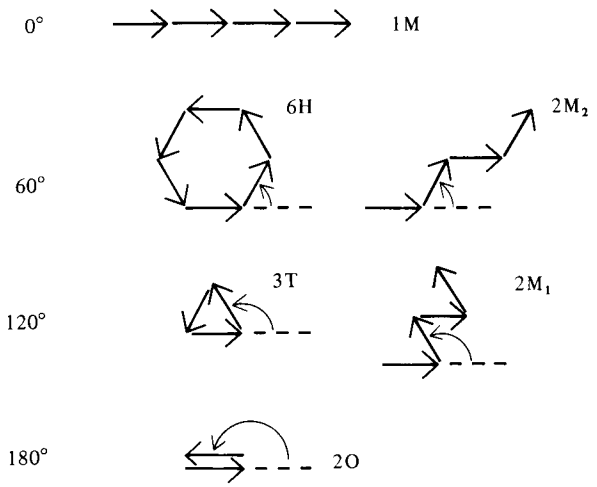



Figure 6.39. The six simple polytypes of mica according to Smith & Yoder (1956). They are derived from different rotations between successive layers. The numbers in the designations ('1M', '6H', etc.) refer to the number of layers required to repeat the original orientation. The letters refer to the crystal classes, monoclinic, hexagonal, trigonal, and orthorhombic.

The commonest polytype for muscovite is $2M_1$ (see figure 6.39) and this is the stable form (Velde, 1965). 3T and 1M are also found as are more randomly stacked crystals (Ross, Takeda & Wones, 1966). The 1M polytype may be kinetically favoured by crystal growth mechanisms: it appears first during laboratory synthesis of muscovites (Yoder & Eugster, 1955). Bailey (1966) in reviewing this topic concludes that, generally, stable polytypes of layer silicates occur nearly to the exclusion of others in natural environments in which sufficient thermal energy or pressure had been available. Metastable polytypes may persist otherwise. Thus when muscovites are found in sediments, crystals of the 1M polytype are more likely to be authigenic (formed *in situ*) and $2M_1$ crystals detrital (carried from elsewhere).

The mutual displacement of the siloxane nets within a single layer of micas generates two kinds of octahedral sites between them: those in the direction of the displacement and those on either side. This is illustrated in figure 6.38. The sites are designated M(1) and M(2). That they are different can perhaps be most easily seen by inspecting the arrangement of the hydroxyl groups in relation to them. The single M(1) site has two '*trans*' OH groups; the pair of M(2) sites have their OH groups '*cis*'. It is in the



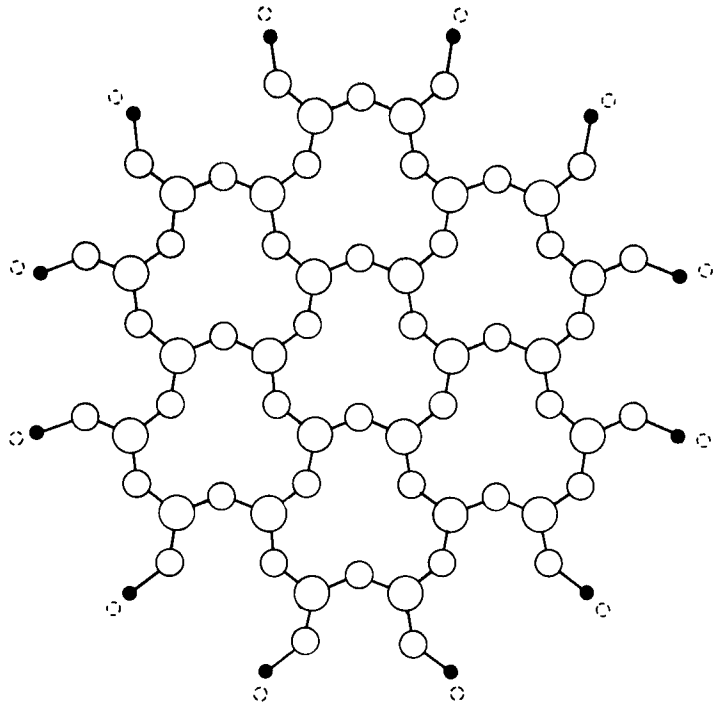


Figure 6.40. Alternate contrary rotations of the tetrahedra in a net such as that shown in figure 6.27*b*. Oxygens are shown as open circles with the nearer apical ones larger. Silicon and aluminium atoms are concealed except round the outside. The extent to which a tetrahedral sheet is contracted is indicated by the shift in the positions of the outermost (exposed) silicon and aluminium atoms.

latter that the octahedral aluminium atoms are located in muscovite. The vacant M(1) site is larger than the other two.

As in the kaolinite group, the siloxane sheets in muscovite are contracted through symmetrical contrary rotations of alternate tetrahedra. This has the effect of lowering the symmetry of the surface oxygens from hexagonal to ditrigonal (see figure 6.40) and it changes the co-ordination of the potassium ions from twelve-fold to six-fold – three of the oxygen atoms in each of the surface indentations will come closer and three will move away from the potassium ions. Furthermore, instead of six orientations for adjacent layers giving identical environments for the potassium ions, there will now be two sets of three. In the 1M polytype, for example, the potassium ion finds itself with triads of oxygen atoms above and below staggered in relation to each other, that is, octahedrally arranged. On rotating the

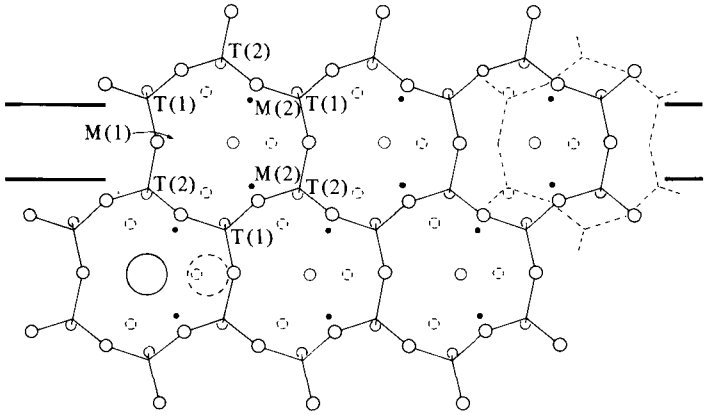



Figure 6.41. Muscovite $2M_1$ drawn from a unit cell given by Güven (1971). The alumino-siloxane net on the underside is omitted except at the top right of the picture where it illustrates a direction of intralayer off-set from right to left. Only one upper and lower potassium ion is shown (lower left). The tetrahedral rotations are such as to bring the bridging oxygens closer to the aluminium atoms. The tetrahedra are also tilted to create shallow grooves that lie horizontally in this picture. (The groove direction is indicated by the solid bars.) The tilting arises from the vacant octahedral sites (M(1)) being bigger than the occupied sites (M(2)). The two classes of tetrahedral sites are designated T(1) and T(2).

layers by 60°, the oxygen triads are now opposite each other at the corners of a prism. A further 60° rotation regenerates an octahedral environment, and so on. The octahedral arrangement is presumably the most stable – hence, presumably, the preference is muscovite for 1M, $2M_1$ or 3T polytypes (see figure 6.39).

As remarked earlier, the contrary rotations of the tetrahedra in a siloxane net can never lead to a perfect match between its apical oxygen atoms and a gibbsite sheet, since the latter is insufficiently symmetrical. In muscovite, as in dickite, the tetrahedra are tilted. Pairs of oxygen atoms on the same side of the gibbsite sheet that straddle vacant M(1) sites are further apart than the others and two out of the six Si–O–Si groups in each siloxane ring are actually stretched. As a result, the bridging oxygens are depressed by some 0.02 nm below the other basal oxygens so that the surfaces of muscovite's unit layers are corrugated. The corrugations run in the direction of the intralayer offset between the siloxane nets. The structure of a $2M_1$ muscovite layer including these features is shown in figure 6.41.

Figure 6.42 shows how this grooving must disturb the local symmetry of the $K^+$ sites. Instead of layer rotation generating three equivalent arrange-



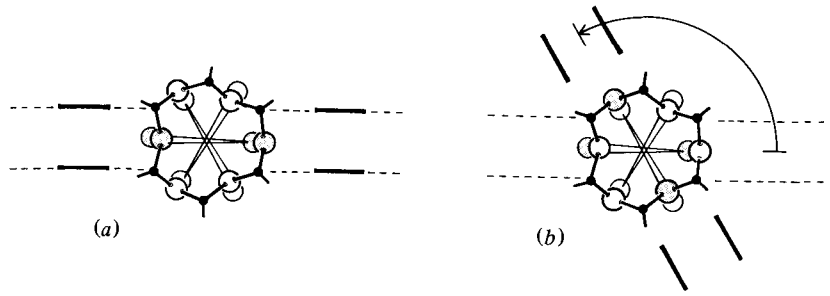



Figure 6.42. (*a*) The arrangement of oxygens around interlayer $K^+$ in 1M muscovite. Tetrahedral rotations cause six of the twelve oxygens to be nearer the $K^+$. This octahedral environment is not quite symmetrical, however, because the tetrahedra are also tilted so that two of the six co-ordinating oxygens are slightly further from the $K^+$ (that is, those in the grooves; see figure 6.41). These oxygens (shaded) are orientated '*trans*'. (*b*) By rotating the top layer by 120° an octahedral arrangement of oxygens is recreated, but with the more distant oxygens '*cis*'. This is the arrangement for the $2M_1$ or 3T polytypes.

ments, there are now two distinct possibilities according to whether the 'special' ligand oxygens, i.e. those in the grooves, are '*trans*' (0°) or '*cis*' (120°). Indeed, there are two possibilities anyway if one looks beyond the basal oxygens: the OH groups and especially the octahedral $Al^{3+}$ ions are differently related between layers rotated by 0° on the one hand or ± 120° on the other (cf. figure 6.41).

It would have been more immediately understandable, perhaps, if the stable polytype of muscovite had been 1M rather than $2M_1$ because in 1M muscovite the grooves of successive layers are parallel. In the event, we are left with two questions: why is 120° rotation favoured and why do the rotations alternate + and – ?

It would appear that the answer to the first question lies in parallel grooves being energetically *dis*favoured. Parallel grooves and ridges cannot nest into each other (because the potassium ions prevent the necessary offset): rather the ridges clash all along their lengths (Güven, 1971). A cross-ridged arrangement, although still not perfect, is a better compromise in reducing the total energy of interaction between adjacent negatively charged surfaces. (As a model for this, one might imagine two ridged air beds on top of each other: they would stack slightly closer if the ridges were angled rather than directly opposed.) Without this effect, the 1M type might indeed have been favoured – as it generally is for the ungrooved trioctahedral micas.

We are left with the perplexing question of selectivity between + and –

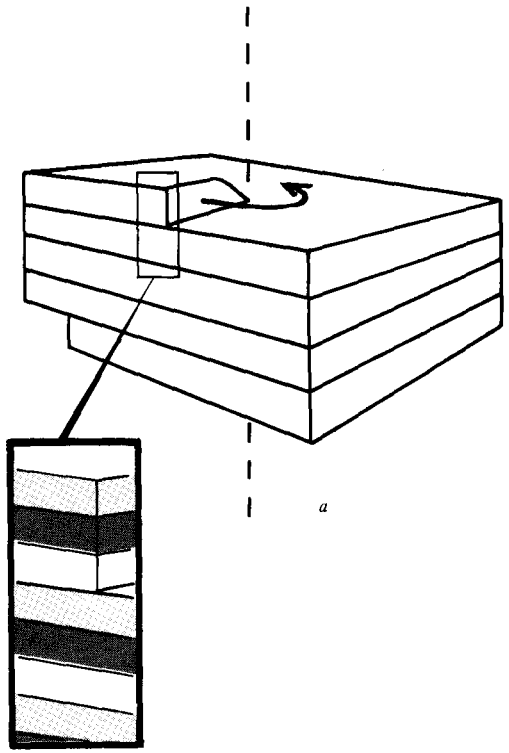



Figure 6.43. (*a*) A screw dislocation (cf. figure 7.1*d*) creates a continuous ramp spiralling around a central line of maximum disturbance – the screw dislocation line (dashed). Crystal growth at a screw dislocation may be particularly easy because new units can add to an advancing step (arrow) which is never eliminated. Hence the difficult process of surface nucleation (cf. figure 5.13 on page 159) is avoided once growth through a screw dislocation is under way. (*b*) A step caused by a screw dislocation may have several layers in it. If in addition there are different possible orientations for the layers (polytypism) then a complex repeating pattern may be generated.

rotations which seem to be structurally equivalent. To alternate, layer $n$ must know which way layer $n-2$ is lying. How can it feel that through layer $n-1$? (These layers are a nanometer apart – too far probably for direct electrostatic interactions to be effective.)

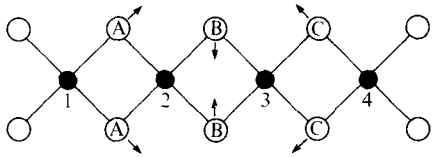
A multilayer spiral growth mechanism can explain the existence of polytypes with complicated long-range repeats, as shown in figure 6.43. Smith & Yoder (1956) suggested that such effects could account for polytypism in micas. (This might give a neat answer to our question: layer $n$ knows about layer $n-2$ because layer $n$ *is* layer $n-2$ – reappearing at the next turn of the growth ramp.) One has to assume only that once started each layer retains its orientation. No transfer of information vertically between layers is needed. This cannot, however, be a complete explanation where a polytype is a thermodynamically favoured form. Whatever its mode of birth, a $2M_1$ crystal of muscovite is a special lowest energy arrangement. If you were to rotate layer $n$ in the middle of such a crystal, the energy would rise and only return to the lowest energy state at 360°. Layer $n$ still knows which way it should be. There must be a vertical transfer of information across a distance of at least 1 nm. We are left with the question: 'how does layer $n$ know which way layer $n-2$ is lying?'



*A mechanical model* Megaw introduced an idea into the discussion of silicate structures that may be useful here. She thought of the structure of the feldspar anorthite in the sort of way that an engineer thinks about a framework of girders – in terms of stresses and strains which may be transmitted over considerable distances (Megaw, Kempster & Radoslovich, 1962). How and how far an effect can be transmitted like this will depend on the compliance of the 'building elements'. (Some are more compliant than others: for example, bond angles at oxygen are more compliant than angles at silicon or aluminium.)

Look again at the siloxane network in figure 6.40. Suppose that you were forcibly to rotate one of the tetrahedra. Then the whole set would embark on a dance of alternate contrary rotations. Such an effect would diminish with distance according to the compliance of the bonds in the structure.

More generally, any kind of stress applied to a silicate layer will produce strains spreading sideways to some extent within its nicely articulated structure. Such a transmission of effect can be imagined, for example, within a set of adjoining octahedral sites:




Suppose that a large cation is put into site 1 in place of a smaller cation. As indicated by the arrows, this would push the A-oxygens outward and tend to compress the other oxygens attached to the cation in site 2. Hence the B-oxygens would tend to move inward a little. The C-oxygens around the cation in site 3 would now have more space to move away from each

other – outwards. As a result of this site 4 would become better adapted for a larger cation. All this because there was a larger cation three sites away.

Vertical effects of this sort can be imagined that might account for the transmission of information *through* silicate layers. For example, the disposition of the surface oxygens of a siloxane net would be directly affected by the repeated forces exerted on these oxygens by the adjacent layer. But the disposition of the apical oxygens that join the siloxane net to the octahedral sheet would be affected too. For example, if one siloxane net in a 2:1 layer was stressed as in our imaginary experiment on figure 6.40 so as to counter-rotate the tetrahedra still further from the 'ideal' arrangement, this would compress *one* of the hexagonal arrays of apical oxygens. The octahedral cations would then be pushed closer together, and hence the *other* array of apical oxygens would be compressed too. The other surface oxygens would have to accommodate – they would tend to adjust through contrary rotations in a manner similar to the directly stressed oxygens on the opposite surface of the layer.

Radoslovich (1963) was thinking in 'mechanical' terms when he said of potassium ions in muscovite that they occupy 'an equilibrium position determined by a complex balanced system of interlocking strong bonds reaching right through the adjacent layers to K's at the next level above and below'.

*Allostery* There is a good analogy between this kind of transmission of effects in articulated silicate structures and **allosteric** effects in proteins. Haemoglobin provides the classical example of allostery. This molecule has four oxygen binding sites some 3–4 nm away from each other. The binding constant of any site increases substantially as the other sites are occupied. How do the sites know about each other's state of occupancy? The answer seems to be through concerted pushings and pullings of an articulated intervening structure – somewhat like a system of 'levers' (Perutz, 1978). More precisely, pairs of identical subunits are related through an axis of symmetry which tends to be preserved in spite of the distortion produced on oxygen binding: thus the imposed distortion in one site is copied in the others (rather as opening or closing one end of a pair of lazy-tongs correspondingly opens or closes the other end).

A symmetrical structure was an essential part of allostery as originally formulated (Monod, Wyman & Changeaux, 1965). Soon, it was realised (Koshland, Némethy & Filmer, 1966) that symmetry is not essential for long-range transmission of information through folded protein molecules: a distortion in one site may produce a different or even an opposite effect



on a similar site elsewhere. It all depends on the detailed characteristics of the intervening articulated network.

For really complex networks – like proteins or layer silicates – one would expect a given structure to produce diverse distant effects according to detailed structural considerations.

*Other examples of 'allosteric' effects in layer silicates* We have already come across what seems to be an 'allosteric' effect, in the polytypism of the kaolinite group – where incidentally the layers are not symmetrical. Here, the layers are apparently distorted by the way they stack in such a way as to favour further stacking as before.

Takeda & Ross (1975) found that 1M and $2M_1$ polytypes of the trioctahedral mica biotite with very similar chemical constitutions had differently distorted layers, and they proposed that once a stacking sequence starts it tends to persist through a similar train of cause and effect.

What appears to be an example of 'negative allostery' in a silicate is provided by the altered micas discussed by Norrish (1973). Here, binding in one place inhibits a similar binding elsewhere. $Na^+$ may sometimes replace $K^+$ between *alternate* layers: it seems that a plane of sodium ions alters the orientations of the hydroxyl groups within the layers above and below so as to make it more difficult for potassium ions to be removed from the opposite sides of the layers above and below. Hence the regular alternation – with the $K^+$ planes 'knowing' that there are $Na^+$ planes next door.

*Chemical effects* Much of our discussion so far may give an oversimplified impression of the processes through which layer silicate crystals are formed. Almost certainly they do not form by simply stacking premade unit layers: these two-dimensional macromolecules are probably synthesised at the same time as they crystallise (at least the ones that we have been talking about so far). This lets in another possible mechanism for control of layer stacking – through chemical effects; in particular, through effects arising from different arrangements of cation substitutions. The stacking of a given layer, on the layer below it, may predispose it to a particular arrangement of cations which in turn predisposes the next layer. In that case, transmission of information might be partly chemical and partly physical.

A 3T muscovite whose structure was determined by Guven & Burnham (1967) may be a case in point. Its gross composition was:

$$K_{0.90}Na_{0.06}Ca_{0.01}Ba_{0.01}(Al_{1.83}Fe^{2+}_{0.04}Fe^{3+}_{0.04}Mg_{0.09}Ti_{0.01})[Si_{3.11}Al_{0.89}O_{10}](OH_{1.98}F_{0.03})$$

This is very similar to a typical $2M_1$ muscovite analysed by Radoslovich (1960):

$$K_{0.94}Na_{0.06}(Al_{1.84}Fe^{3+}_{0.12}Mg_{0.06})[Si_{3.11}Al_{0.89}O_{10}](OH)_2$$

Yet the layers of the 3T polytype are substantially different from the $2M_1$ type, not only physically (having shallower grooves and a different space group), but chemically. For one thing, the aluminium substitutions were ordered into only one of the two classes of tetrahedral sites – unlike other muscovites where the substitutions are equally distributed. Also, the minor octahedral constituents were concentrated in one of the two aluminium sites. Such a chemical asymmetry might well help to transmit the one-handed style of the 3T polytype stacking. In any case, the chemical asymmetry within the layers is ordered (repeated) between the layers, presumably concomitantly with the crystal growth processes that determined the 3T stacking mode.

Ordering *within* layers is another important topic. As already discussed, dioctahedral layer silicates invariably show long-range ordering of octahedral vacancies. Also, as we have just seen, more elaborate octahedral ordering is possible as is the ordering of aluminium in tetrahedral sites. It is surprising that long-range tetrahedral ordering is not more common in 2:1 silicates since the effect of replacing a silicon atom with an aluminium atom must be particularly powerful: Al is bigger than Si and, most important, it introduces a negative charge. For example, for potassium ions to be locally neutralised, exactly three of the atoms in the twelve nearest tetrahedral sites above and below a potassium ion should be aluminium. That consideration alone makes a random distribution of tetrahedral substitutions most unlikely. Bailey (1975) has pointed out that seeming disorder in tetrahedral substitutions may be a result of assumptions made in crystal structure determinations. Tetrahedral ordering in domains was suggested by Gatineau (1964) from diffuse X-ray scattering evidence. This idea involves zig-zag chains of aluminium substitutions arrayed in one of three different directions in different domains. Most interestingly, in Gatineau's model, aluminium-rich domains face similar aluminium-vacant domains across the $K^+$ planes. Gatineau's specific suggestion has been viewed with caution by Bailey (1975) on the grounds that it violates a general principle that aluminium atoms avoid being beside each other as far as possible in silicate structures, and also because since 1964 other kinds of domaining have been found that could produce diffuse scattering without Si, Al ordering. This matter remains unsettled, but short-range ordering of some sort seems likely.



In the brittle micas, there is about twice as much aluminium substitution as in the normal micas, to balance out the calcium ions that mainly fill the interlayer spaces. Margarite is the closest type to muscovite: in this case there is nearly complete alternate ordering of the aluminium substitutions in a specimen of composition:

$$Ca_{0.81}Na_{0.19}K_{0.01}(Al_{1.99}Mg_{0.03}Fe_{0.01})[Si_{2.11}Al_{1.89}O_{10}](OH)_2$$

In addition, there was a slight difference between the two tetrahedral sheets of any layer, one containing somewhat more aluminium than the other (Guggenheim & Bailey, 1975).

### D. *Trioctahedral micas*

The simplest trioctahedral 2:1 layer silicate is *talc* with the ideal formula $Mg_3[Si_4O_{10}](OH)_2$. The unit layer is a brucite sheet with siloxane nets fused on each side. Talc, like its dioctahedral counterpart pyrophyllite, is atypical of the group as a whole having only weak, van der Waals, interlayer bonding.

*Phlogopite* is the better archetype for trioctahedral micas corresponding to muscovite for the dioctahedral class. Its ideal formula is $K(Mg_3)[Si_3AlO_{10}](OH, F)_2$, but this is seldom approached in natural samples where $Fe^{2+}$, $Fe^{3+}$, $Al^{3+}$ and $Ti^{4+}$ in particular may occupy a substantial proportion of octahedral positions. Also, $Na^+$ and, to a smaller extent, $Ca^{2+}$, are found standing in for $K^+$ in interlayer positions. While muscovite tends to be short on tetrahedral aluminium, in phlogopite micas there is usually more aluminium than the ideal.

*Biotite* is a very common mica of igneous and metamorphic rocks. It is somewhat arbitrarily distinguished from phlogopite in having a Fe:Mg ratio of more than 1:2. Figure 6.44 illustrates the wide range of compositions of these two micas in relation to four extreme structures. Even this underestimates the possibilities, though, as the formula of a particular biotite illustrates:

$$K_{0.9}Na_{0.07}Co_{0.01}(Fe^{2+}_{1.42}Mg_{0.85}Mn_{0.01}Al_{0.41}Ti_{0.13}Fe^{3+}_{0.09})[Si_{2.8}Al_{1.2}O_{10}](OH_{1.7}F_{0.02})$$

This biotite, chosen from 32 examples given in Deer, Howie & Zussman (1962), is from a garnet mica schist from Morar in Scotland (Lambert, 1959). It is not untypical in being short on (OH, F) and having 0.1 octahedral positions out of three vacant.

The hydroxyl groups in micas are well placed to be spectroscopic probes

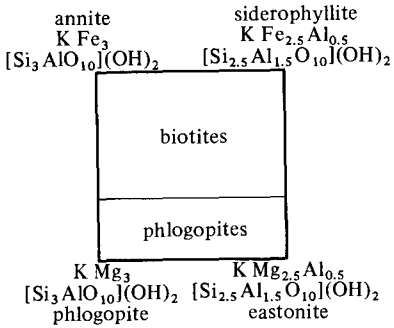
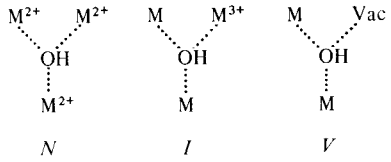



Figure 6.44. The phlogopite–biotite composition fields. Most phlogopites and biotites lie within these fields. The division between them is arbitrarily chosen to be where Mg:Fe = 2:1. (From Deer, Howie & Zussman, 1962.)

of octahedral cation distribution. The OH stretching frequency is sensitive to the cations (usually three in trioctahedral micas) to which OH is co-ordinated. A glance at the formula of the Morar biotite is enough to realise the large number of possible sets of nearest neighbours that any one OH group might have. The OH stretching bands are understandably complex (not helped by the influence, also, of interlayer cations). Vedder (1964) distinguished nevertheless three general classes of OH groups – 'Normal', 'Impurity' and 'Vacancy':




which gave distinctive composite bands. More detailed examination of the bands can give clues as to which ions are involved (Vedder, 1964; Wilkins, 1967; Farmer *et al.*, 1971; Rousseaux, Gomez Laverde, Nathan & Rouxhet, 1973; Rausell-Colom, Sanz, Fernandez & Serratosa, 1979). An adjacent vacancy has the effect of altering the orientation of the OH groups from approximately normal to the layer plane to approximately in the plane – the hydrogen atom then pointing towards the vacancy as in dioctahedral silicates. Using polarised infrared radiation, Chaussidon (1973) was thus able to distinguish OH groups lying '*trans*' from those lying '*cis*' across vacancies. Both types were present in comparable amounts indicating both M(1) and M(2) vacancies. This is in contrast to dioctahedral micas where the vacancies are ordered in the M(1) site.



In all, one has an impression of considerable freedom in the arrangement of octahedral components in the trioctahedral micas. A review by Hazen & Wones (1972) of synthetic products confirms also a considerable freedom in choice of kinds of ions – provided they are small enough and divalent they will pass as octahedral guests. Even small trivalent ions, the tetravalent $Ti^{4+}$, and the monovalent $Li^{+}$, are tolerated. Indeed, divalent ions may sometimes be quite absent, being replaced by a balance of $Li^{+}$ and $Al^{3+}$ (in the *lepidolite* micas).

The variety of possible ways of arranging the octahedral cations in trioctahedral micas may be very great, yet clearly there are restraints over and above the requirements of small ionic size and overall charge balance. The number of vacancies, for example, seldom rises much above 0.3 for every three sites. If these vacancies are not so obviously ordered as in the dioctahedral micas, the cations are ordered in many cases, with larger cations generally preferring M(1) sites (Bailey, 1966, 1975). This goes to emphasise that a layer silicate structure only partly resembles a rigid array of pigeon holes within which ions can be distributed. A cation will control the size of its pigeon hole and that will affect the preferred size of adjacent pigeon holes. Lateral 'allosteric' as well as electrostatic effects can be expected. Veitch & Radoslovich (1963) have discussed the theoretical grounds for octahedral ordering on the basis of both size and charge. The 'allosteric' effect in particular helps us to understand the fairly clear distinction that exists in layer silicates between dioctahedral and trioctahedral types: the structural restraints induced by one or other of these forms of organisation seem to be self-perpetuating. Dioctahedral domains within a trioctahedral structure, or vice versa, might be electrostatically balanced but are likely to be sterically strained and hence disfavoured unless some other compensating factor is at work.

Although trioctahedral sheets are bigger than dioctahedral sheets, they are still usually smaller (in 2:1 silicates) than the tetrahedral sheets – because the latter are also enlarged by substantial substitutions of aluminium for silicon. As a result, tetrahedra are alternately rotated as usual; but since M(1) and M(2) sites are more similar in size, there is little corrugation of the surfaces. Hence, 1M polytypes are favoured for trioctahedral micas.

### E. *Chlorite and vermiculite*

*Chlorites* occur in igneous, metamorphic and sedimentary rocks – in the latter both as detrital and authigenic particles. There is a wide range of sizes for the crystals: they may be found as large blocks or as clay-sized particles.

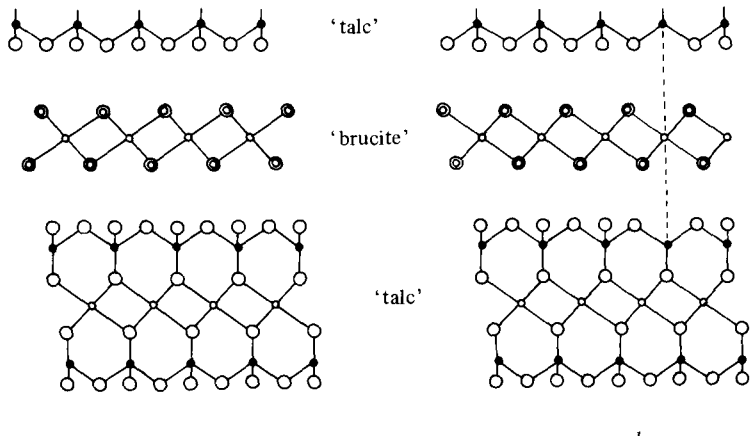



Figure 6.45. (*a*) and (*b*) A typical chlorite crystal is a stack of alternating talc-like and brucite-like units. Two of the numerous stacking modes are shown. In (*b*) one of the sites in the 'brucite' interlayer lies exactly between tetrahedral sites in the 'talc' layers above and below. (After McMurchy, 1934, and Bailey, 1975.) Larger open circles, oxygens; double circles, hydroxyls; small open circles, octahedral cations; and small black circles, tetrahedral cations.

An ideal chlorite is generally taken to consist of talc and brucite layers regularly interleaved (figure 6.45). Thus, a 2:1 structure is combined with the kind of interlayer hydrogen bonding typical of the kaolinites and serpentines. Real chlorites, however, are far from this ideal, the 2:1 unit being more like biotite than talc in the variety and extent of its cation substitutions. Aluminium for silicon substitutions occur within the range $(Si_{3.5}Al_{0.5})$–$(Si_2Al_2)$. The negative charges thus created are compensated by octahedral substitutions – generally $R^{3+}$ for $R^{2+}$ – which may be either exclusively in the brucite layers or also in the talc layers. $Al^{3+}$ is usually the main source of these positive charges although $Fe^{3+}$ commonly contributes a substantial fraction. Iron is indeed an important component of most chlorites – mainly as $Fe^{2+}$ standing in for $Mg^{2+}$. A typical formula would be:

$$(Mg, Al, Fe)_6[(Si, Al)_4O_{10}](OH)_8$$

with minor amounts of other ions such as $Mn^{2+}$, $Cr^{2+}$, $Ni^{2+}$ and $Ti^{4+}$ in octahedral positions.

As might be imagined, the greater complexity of chlorites as compared with micas multiplies the possibilities for polytypism. Brown & Bailey



(1962) have derived twelve possible one-layer polytypes and Lister & Bailey (1967) no less than 1134 two-layer polytypes on the basis of a classification that distinguishes different ways in which a brucite layer can be placed on a talc layer with maximum hydrogen bonding, different ways of superposing these, and different directions of intralayer offset in the talc subunits. Most chlorites have a semi-random stacking due to arbitrary displacements ($\pm b/3$) between adjacent chlorite units, but the relationship between 'talc' and 'brucite' subunits tends to remain fixed.

Steinfink (1958a) determined the crystal structure of a monoclinic chlorite of composition:

$$(Mg_{2.6}Fe_{1.7}Al_{1.2})[Si_{2.2}Al_{1.8}O_{10}](OH)_8$$

From differences in oxygen–tetrahedral cation distances, it was clear that the tetrahedral aluminium was concentrated in one of the two kinds of site. The distribution of octahedral cations was also far from arbitrary. Judging from bond lengths and electron densities, of the three octahedral 'talc' sites the first contained magnesium, the second magnesium + iron (about 3:1) and the third iron. The octahedral aluminium was in two of the three interlayer sites which contained also about 25 % iron. The third of these 'brucite' sites contained magnesium with about 25 % vacancies. A triclinic chlorite also studied by Steinfink (1958b) showed no ordering in the 'talc' part of the structure, but again aluminium was concentrated in the interlayer, with signs of ordering there.

Brown & Bailey (1963) determined the crystal structure of a chromium chlorite from Erzincan, Turkey, which had the following composition:

$$(Mg_{5.0}Fe^{2+}_{0.1}Cr^{3+}_{0.7}Al_{0.2})[Si_{3.0}Al_{1.0}O_{10}](OH)_8$$

In this particular polytype, the trivalent cations in the 'brucite' layers come opposite the set of tetrahedral sites of the 'talc' layers within which the aluminium ions are concentrated. Thus the centres of excess positive charge in the 'brucite' layers lie exactly between the centres of excess negative charge in the 'talc' layers above and below. The arrangement is illustrated in figure 6.45*b*.

Most chlorites are trioctahedral both in the 2:1 and interlayer components. Donbassite is the name given to doubly dioctahedral chlorites, that is, those based on pyrophyllite and gibbsite rather than talc and brucite. The vacancy in the gibbsite interlayer is partly occupied to make the required balance of charges. In a Ia donbassite:

$$(Al_{4.1}Fe^{3+}_{0.04}Fe^{2+}_{0.01}Mg_{0.08}Li)_{0.26}[Si_{3.14}Al_{0.86}O_{10}](OH)_8$$

Aleksandrova, Drits & Sokolova (1972) found both tetrahedral ordering of aluminium and an ordering between gibbsite and pyrophyllite components analogous to that in the Erzincan chlorite. Here, however, it is the least charged, that is, partly vacant site of the gibbsite interlayer that lies over the aluminium-rich tetrahedra.

Cookeite is a still more curious chlorite with a dioctahedral 2:1 layer and a trioctahedral interlayer. Bailey (1975) reports on a preliminary study of a Ia cookeite of approximate composition:

$$(Al_{4.02}Li_{0.86})[Si_{3.02}Al_{0.97}O_{10}](OH)_8$$

The lithium is ordered in one of the positions in the interlayer and the aluminium is apparently in only one of the tetrahedral sheets of the 2:1 layer.

It is a feature of the micas that their layers are firmly locked together. The chlorites are similar in this respect, and they are similar too in having a high level of tetrahedral substitution which locates negative charges close to the 2:1 layer surfaces. It is the interlayer junction that is quite different in the two cases, a simple plane of cations in the micas being replaced by a hydroxide sheet. As a result, the characteristic distance between 2:1 layers changes from about 1 nm in the micas to about 1.4 nm in chlorites.

Going from chlorites to *vermiculites*, we move two steps further from the micas. The negative layer charge is still dominated by aluminium for silicon substitutions, but the charge is generally somewhat less than that in either micas or chlorites due to excess positive charge in the central octahedral sheets of the 2:1 layers. These are still about 1.4 nm apart, but the junction between them is significantly different from the arrangement in chlorite. The hydroxide interlayer is replaced by two (incomplete) planes of water molecules with divalent cations, usually $Mg^{2+}$, octahedrally hydrated between them. The arrangement of octahedral sites is similar to that in brucite, but they are much less heavily populated: here (unlike brucite) each magnesium ion contributes two net positive charges to the interlayer – and occupation of between one in six to one in nine of the sites is generally enough to compensate the 2:1 layer charges – and there is no need for higher valence interlayer cations, although these are sometimes present. An analysis of vermiculite from Llano, Texas (Foster, 1963) illustrates how the charges may be balanced:

| | $Mg_{0.48}K_{0.01}$ | $(Mg_{2.83}Fe^{3+}_{0.01}Al_{0.15})$ | $[Si_{2.86}Al_{1.14}O_{10}]$ | $(OH)_2$ |
|---|---|---|---|---|
| | interlayer | 2:1 octahedral | tetrahedral | |
| Charges: | +0.97 | +0.14 | −1.14 | |



Mathieson & Walker (1954) had suggested that the interlayer cations in vermiculite would, as a result of mutual repulsion, tend to be distributed within a hexagonal network of sites. Shirozu & Bailey (1966) defined the situation more precisely in a crystal structure determination of a vermiculite from the Llano location. Here 0.41 magnesium atoms are concentrated in one of the three available octahedral interlayer sites. There is also an ordering of aluminium between aluminium-rich (T1) and aluminium-poor (T2) tetrahedra. The magnesium-rich interlayer sites lie exactly between the aluminium-rich tetrahedral site in a manner analogous to the Erzincan Cr-chlorite (see figure 6.45*b*). The (double) charges on the magnesium ions can thus be locally balanced.

The partial occupancy of the interlayer sites in vermiculite still leaves open the possibility of super-ordering. Where, as is common, only a third of the possible sites are occupied by magnesium, Bradley & Serratosa (1960) suggested that these might be arrayed in a regular manner over three unit cells – that is, as far away from each other as possible. An arrangement like this has been found by Alcover, Gatineau & Mering (1973) in a vermiculite from Kenya. Here, interlayer magnesium is present in domains. In each domain only one of three sets of sites is occupied, but different sets can be occupied in different domains. This ordering is two-dimensional, there being no relation between successive interlayers. These authors suggest that there may be a corresponding ordering of the aluminium tetrahedral substitutions to balance the charges locally, and Bailey (1975) has pointed to the possibility that the Llano vermiculite might combine a similar short-range domaining of aluminium substitutions within tetrahedral sheets with the long-range intersheet ordering implied by Shirozu & Bailey's study (1966).

Much, if not most, vermiculite in Nature is formed by alteration of trioctahedral micas. The main part of this process would appear to be simply the replacement of interlayer potassium by magnesium and water – a reaction that can indeed be simulated in the laboratory by leaching biotite or phlogopite with MgCl (Barshad, 1948; Caillère & Hénin, 1949). The Llano vermiculite referred to above was evidently a product of weathering of a phlogopite that was found unaltered just below the surface. There is a puzzle here, however, since the ordered arrangement of aluminium substitutions deduced as being in the vermiculite could not have been there in the supposed parent phlogopite (Shirozu & Bailey, 1966).

F. *2:1 clays*

The *Illites* are an abundant if not very well defined group of clays that are structurally quite close to the micas – generally to muscovite. Being clays, their crystallites are small. They are also less well crystallised than muscovite, a somewhat lower layer charge contributing, perhaps, to a less accurate stacking of layers. In addition to having less aluminium for silicon substitution there is some replacement of interlayer potassium by, for example, $Ca^{2+}$, $Mg^{2+}$ and $H^{+}$. There is, however, little interlayer water. An idealised formula is:

$$K_{0.5-0.75}Al_2[Si_{3.5-3.25}Al_{0.5-0.75}O_{10}](OH)_2$$

Glauconite found in marine sediments is an iron-bearing illite in which $Fe^{3+}$, $Fe^{2+}$ and $Mg^{2+}$ occupy a substantial proportion of the octahedral sites.

The *smectites* are again related to the micas but now more remotely. The interlayer cations are here mainly $Ca^{2+}$ and $Na^{+}$, rather than $K^{+}$, and they are readily exchangeable. The crystallites are smaller than those of the illites and usually have a very irregular outline. The layer charges are low enough to allow the most characteristic property of the group – that the unit layers readily intercalate water molecules. Smectites are the main class of *expanding clays*. Calcium smectites generally have two sheets of interlayer water but the sodium smectites often pull in many more. Indeed the unit layers of smectites can come more or less completely apart in aqueous suspension to give, in effect, a solution of two-dimensional macromolecules. These are quite flexible. Often the appearance under the electron microscope is less like a conventional crystal than a piece of crumpled linen (figure 6.46). Even when the layers of a smectite crystal are apparently stacked neatly on top of each other there may be little regularity in the interlayer relationships. Regular crystalline organisation may be largely confined to two dimensions – and even this may be quite short-range.

Often, too, a clay crystallite consists of layers of different kinds. Illite and montmorillonite, for example, are frequently interstratified – indeed kaolinite layers may be included too. Such interstratification may be either regular or irregular.

In spite of the potential and actual structural variability of smectites, it has been possible to obtain from certain localities clays of a more or less consistent type. *Montmorillonite* is one of these. It can be given the following idealised formula:

$$(\tfrac{1}{2}Ca,\ Na)_{0.33}(Al_{1.66}Mg_{0.33})[Si_4O_{10}](OH)_2\cdot nH_2O$$



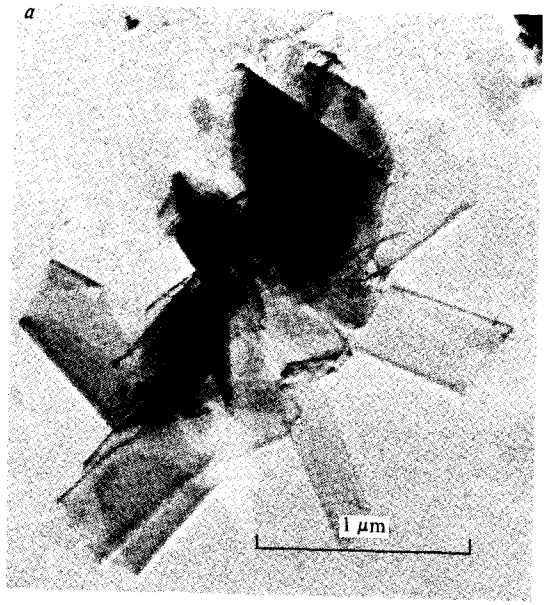



Figure 6.46. Aggregate of back-folded lamellae in Tatatilla montmorillonite after dispersion. (From Grim & Guven, 1978.)

The ease of expansion of montmorillonite is sometimes attributed to the layer charge arising mainly in the octahedral sheet – some distance from the compensating interlayer cations.

In *beidellite* however, the layer charges arise mainly from tetrahedral substitutions:

$$(\tfrac{1}{2}Ca,\ Na)_{0.33}(Al_2)[Si_{3.66}Al_{0.33}](OH)_2\cdot nH_2O$$

*Nontronite* is similar except that $Fe^{3+}$ replaces octahedral $Al^{3+}$ in the idealised formula.

*Saponite* is the trioctahedral equivalent of beidellite with a unit layer thus similar to that of vermiculite:

$$(\tfrac{1}{2}Ca,\ Na)_{0.33}(Mg_3)[Si_{3.66}Al_{0.33}](OH)_2\cdot nH_2O$$

There may be little if any distinction between a high-charge saponite and a low-charge vermiculite (Suquet, Iiyama, Kodama & Pezerat, 1977).

Completing the set, *hectorite* is the trioctahedral equivalent of montmorillonite, the octahedral charge here being provided by substitution of some $Li^{+}$ for octahedral $Mg^{2+}$:

$$(\tfrac{1}{2}Ca,\ Na)_{0.33}(Mg_{2.66}Li_{0.33})[Si_4O_{10}](OH,\ F)_2\cdot nH_2O$$

A *bentonite* is a mixture of (mainly) smectite clays – in particular montmorillonite and beidellite.

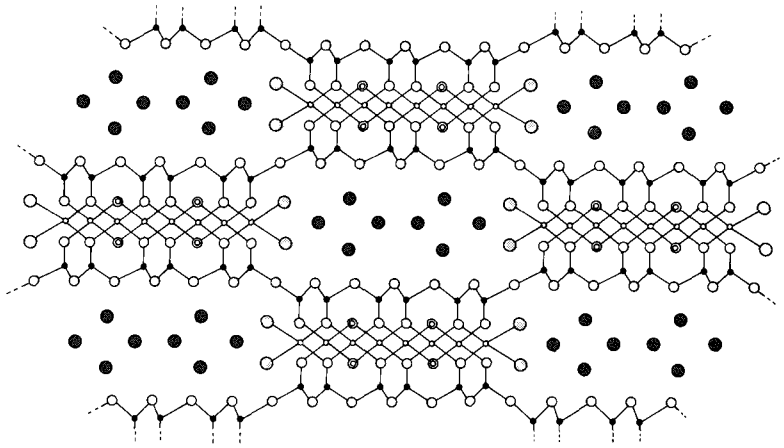


Figure 6.47. The structure of sepiolite (after Brauner & Preisinger, 1956) showing channels containing both mobile or 'zeolitic' water (dark shading) as well as more strongly fixed water (lighter shading). Small open circles, magnesium atoms; small black circles, silicon atoms.

*Other clay minerals*

*Sepiolite* has a crystal structure consisting of narrow talc-like strips joined at their corners, as in figure 6.47, so that channels are formed. The crystals are typically fibrous as shown in figure 6.48*a*. Note the groove that is about 10 nm wide running along the length of the larger of the crystals shown here. Figure 6.48*b* is a view across such a sepiolite crystal at higher magnification. The channels of the ideal crystal structure are clearly visible here – these are about $0.7 \times 1.3$ nm. There are also tubular pores of between 2 nm and 20 nm across.

*Palygorskite* is somewhat similar, but with narrower strips and a more complex pattern of octahedral cations.

Although structurally close to the layer silicates, these two are really framework minerals, similar to zeolites. A discussion of their surface properties has been given by Serratosa (1979). Sepiolite and palygorskite tend to form when magnesium-rich waters evaporate.

*Imogolite* and *allophane* are hydrous alumino-silicates. Imogolite can be seen under the electron microscope to consist of long thin tubes with an outer diameter of about 2 nm and an inner pore of about 1 nm – like the finest imaginable capillary tubing (figure 6.49). Cradwick *et al.* (1972) have proposed a structure that is like kaolinite in so far as silica units are fused on one side of a gibbsite sheet, but the units are inverted: here three of the




Figure 6.48. Electron micrographs of sepiolite (from Rautureau & Tchoubar, 1976). (*a*) Grooved sepiolite laths. (*b*) Cross-section at high magnification showing channels in the crystal structure (cf. figure 6.47) as well as larger holes running through the crystal.

oxygen atoms of each $SiO_4$ unit are shared with the gibbsite sheet leaving one as a hydroxyl group. There is no Si–O–Si bonding – the $SiO_4$ units are separate as in an orthosilicate (figure 6.50*a*). This kind of bonding imposes a powerful distortion on the gibbsite sheet which thus curls into a tube with ten to twelve –Si–OH units on the inside circumference (figure 6.50*b*).

Allophane particles are also visible in figure 6.49. Henmi & Wada (1976)

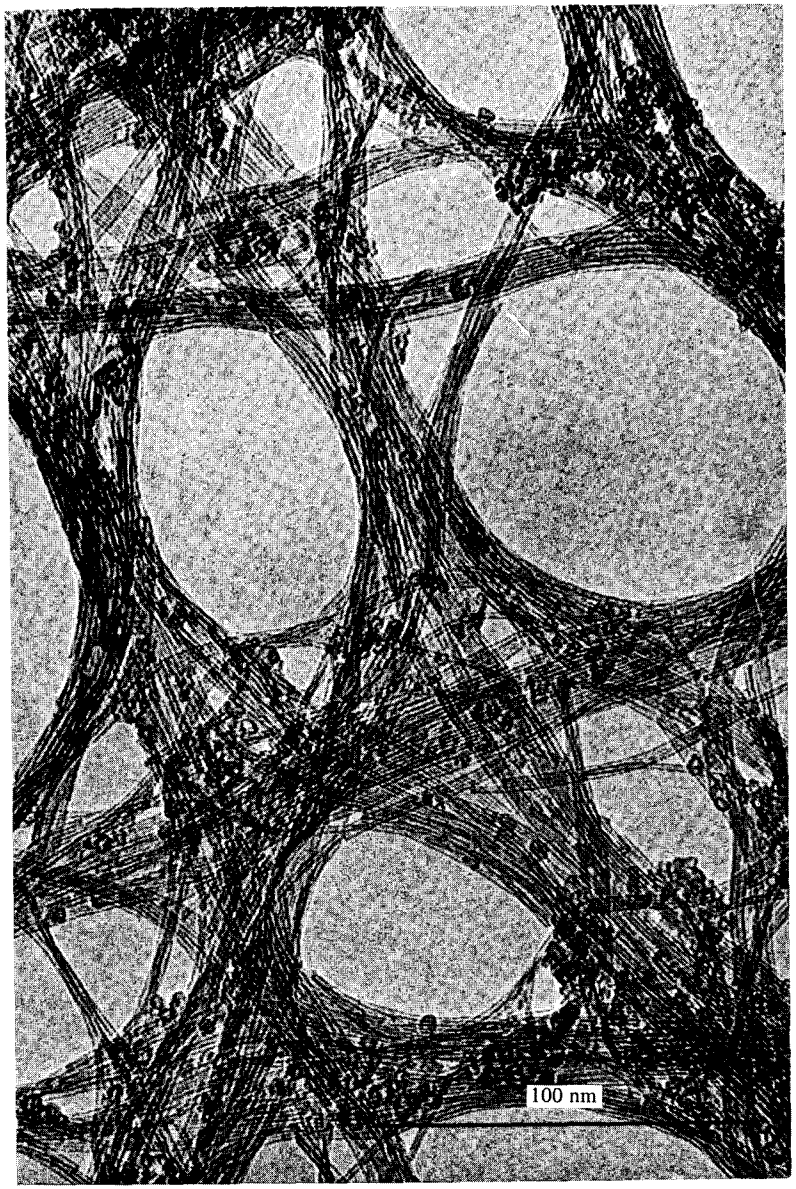



Figure 6.49. Transmission electron micrograph of imogolite tubes and allophane pods from Kurayoshi. (From Nakai & Yoshinaga, 1978.)



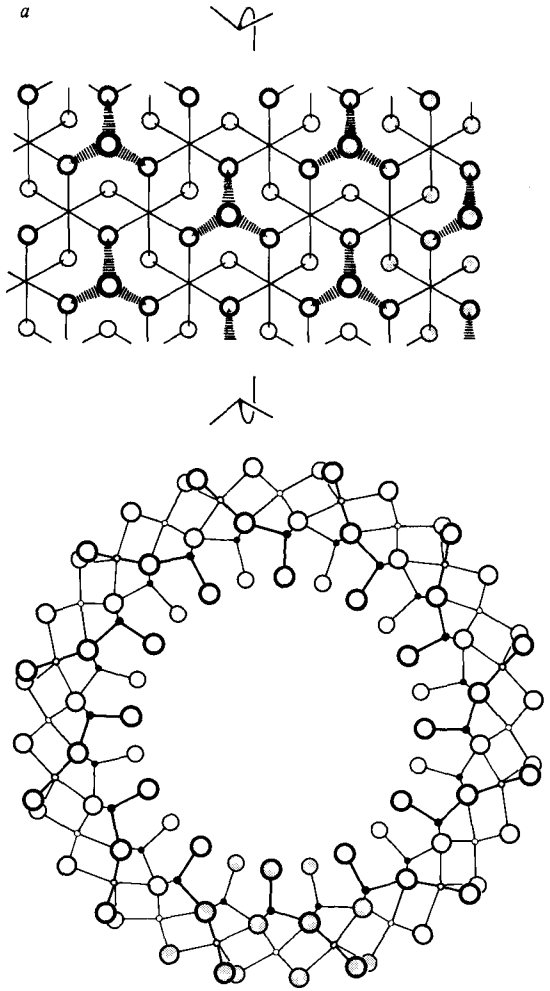



Figure 6.50. (*a*) According to Cradwick *et al.* (1972) imogolite consists of a gibbsite sheet with separate silicate units fused to one side (these are shown with 'spring' bonds). The flat structure shown would be strained – the 'springs' would be under tension – because silicate units are not big enough. So instead the structure rolls up as indicated by the arrows. (*b*) A view down an imogolite tube showing the 1 nm diameter pore. This view is as from the right-hand side in (*a*). Corresponding atoms visible in both (*a*) and (*b*) are shaded.

have proposed that such particles are hollow spheres with an outer diameter of 3.5–5.0 nm and a wall thickness of about 1 nm or less. The walls seem to have pores in them that can let water molecules through (Wada & Wada, 1977; Wada, 1979), and are sites for phosphate binding (Parfitt & Henmi, 1980).

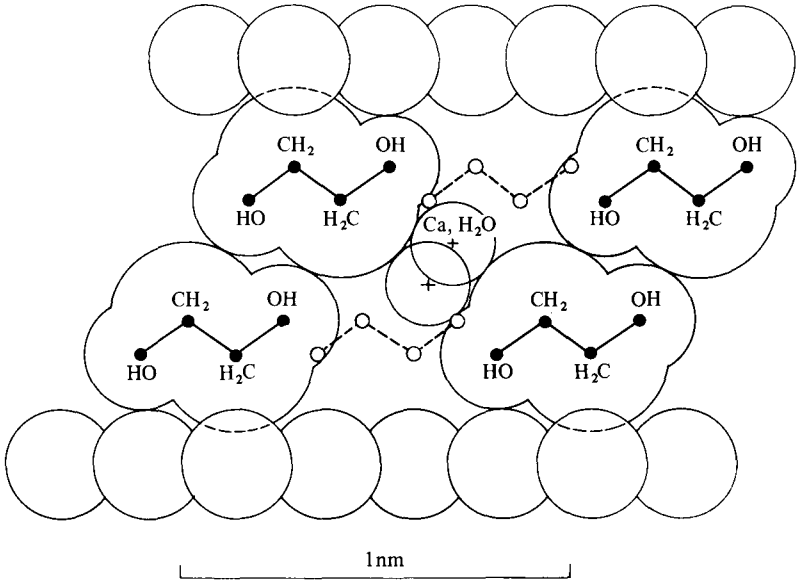



Figure 6.51. Ethylene glycol molecules between the layers of a calcium montmorillonite. (After Reynolds, 1965.)

Both imogolite and allophane can apparently form quite quickly under suitable conditions. A 'proto-imogolite' can be made by mixing dilute solutions of aluminium salts and monomeric silicic acid at pH 4.5; on heating (40–100 °C) this gave tubes very similar to natural imogolite (Farmer, Fraser & Tait, 1977; Farmer & Fraser, 1979). Allophane precipitates from the $CO_2$-charged waters of 'Silica Springs' in New Zealand (Wells, Childs & Downes, 1977). These minerals are particularly important constituents of soils derived from the weathering of volcanic ash, but their occurrence is widespread – they can form from plagioclase feldspar (Tazaki, 1979a). As they are difficult to detect by X-ray methods, both imogolite and allophane may have been underestimated as soil components.

Parfitt & Henmi (1980) conclude that the allophane membrane can often be regarded as a defective sort of imogolite; but they distinguish such 'proto-imogolite allophane' with an Al:Si ratio of *ca* 2 from other forms, such as the Silica Springs material which has an infrared spectrum more like felspathoid.



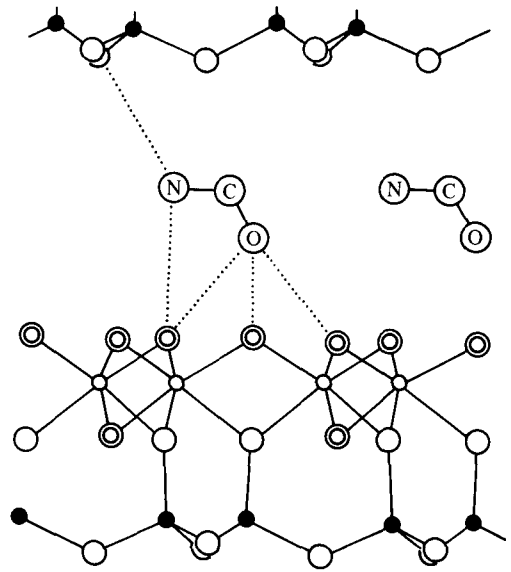



Figure 6.52. Dickite–formamide intercalate viewed along the *x*-axis. Dotted lines represent hydrogen bonds between the nitrogens and carbonyl oxygens of the formamide molecules, and the hydroxyls below and siloxane oxygens above. (After Adams, 1978.)

### *Clay–organic interactions*

Smectites, particularly montmorillonite, have long been known to take up organic molecules of various kinds. Often this is between the silicate layers and in that case changes in the interlayer spacing can easily be detected by X-ray diffraction. For example, ethylene glycol can replace most of the interlayer water molecules in calcium montmorillonite to form highly organised complexes such as that shown in figure 6.51. Many other kinds of polar organic molecule can form such organo–clay complexes – for example, alcohols, ketones, ethers and amines may come in, instead of water, between the layers of smectites and vermiculites. To be adsorbed from dilute aqueous solutions polar molecules generally have to be fairly big ($> C_5$) if they are to compete well enough with water (Hoffmann & Brindley, 1960).

Organic cations may replace inorganic interlayer ions in many 2:1 layer silicates. *N*-Alkyl ammonium ions and ionised heterocyclic compounds such as pyridinium are particularly effective.

Small highly polar molecules are favoured types for intercalating kaolinite layers – for example, formamide intercalates in a regular way in dickite (figure 6.52). Urea, hydrazine, and dimethyl sulphoxide are other examples.

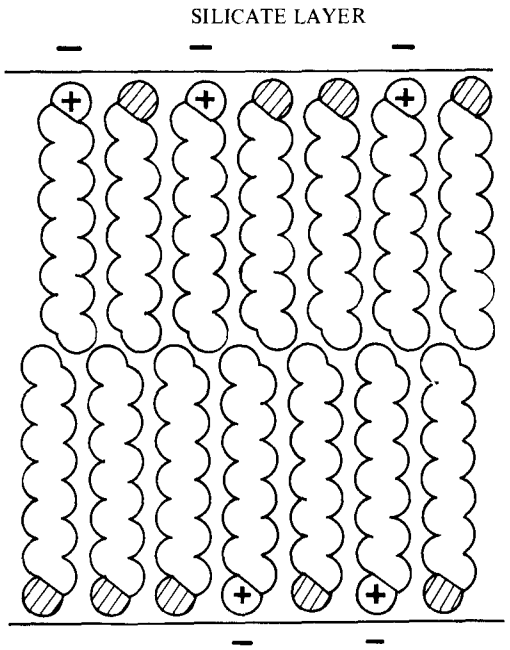



Figure 6.53. Packing of (positively charged) long-chain alkyl ammonium ions together with corresponding long-chain alcohols between mica-like silicate layers. (After Weiss, 1963.)

Rather high concentrations are needed, however, and the interlayer species are often rather easily removed in contact with water.

This whole topic of clay–organic interactions has been thoroughly reviewed in a book by Theng (1974). Here, particular attention is paid to molecules of current biological importance: not only 'biochemicals' but pesticides, antibiotics, alkaloids and other classes, illustrating the great variety of organic molecules that will bind to clays of some sort in some way or another. Kaolin has been used medically for a long time as a means of taking up various toxic materials – although 2:1 clays are generally the most effective in forming organo–clay complexes.

Generally, too, it is basic organic molecules that are most readily taken up – in the neutral to acid pH range. Nucleotides and their heterocyclic components have been studied particularly (Lailach, Thompson & Brindley, 1968a, b; Lailach & Brindley, 1969; Thompson & Brindley, 1969). Most clays do not readily adsorb the major biomonomers at the kind of pH ($7.1 \pm 1$) thought to have existed in the primitive ocean, but the binding



of nucleotides is enhanced in the presence of transition metal cations (Odom, Rao, Lawless & Oró, 1979). Porphyrins, too, are more readily bound along with such metal ions (Cady & Pinnavaia, 1978).

Fatty acids may be removed from aqueous solutions by clays (Meyers & Quinn, 1971), but complexes between fatty acids and montmorillonite are not formed very easily. Edges rather than interlayers are likely to be the adsorption sites (Sieskind & Ourisson, 1971). Triglycerides, too, tend to give rather indefinite complexes with 2:1 layer silicates unless these already have organic interlayer material in the form of suitable long-chain alkyl ammonium ions: then triglycerides may be taken up between the layers in a regular way (Weiss & Roloff, 1966). A similar effect had been noted by Weiss (1963) with the co-adsorption of long chain alcohols and corresponding alkyl ammonium ions (figure 6.53).

Simple sugars such as glucose do not bind well to clays. Apparently they are too highly hydrated; methylated sugars complex more easily (Greenland, 1956).

Not only layer silicates, but many other inorganic materials with layer crystals can intercalate organic molecules – for example niobates, titanates and molybdates (Lagaly, 1981).

### The genesis of clay minerals

Conjurors make use of our tendency to take mental short cuts. The egg broken in the teacup cannot be the same egg as the perfect one found later in the lady's handbag; but the clever conjuror persuades us that it is. Your perception (if not your intellect) can be persuaded of the most preposterous notions on the basis of a principle that normally works very well in everyday life. We cannot be for ever wondering if what seems to be the same object really is: if the green book on the table this morning is the same green book that was left there last night, or if the train emerging from one side of a distant tunnel is the same train that we saw entering the other side a minute before. Such suppositions are common sense but they are not logically secure – and it is on this distinction that conjurors make their living.

Nature cannot be credited with the conjuror's guile, but it sometimes seems to play tricks on us by producing the same kinds of molecules, or the same or similar crystal structures in different circumstances; and we may be persuaded to draw causal lines between them like the spurious causal line between the egg in the teacup and the egg in the handbag. One of these tricks we have already discussed at length: amino acids and purines are

found (for example) in meteorites. They are also found in us. Therefore, (short cut thinking tells us) our amino acids and purines came in the first place from meteorites (or alternatively from thunderstorms or whatever). The synthesis of clays in Nature provides another interesting example of an often false set of causal connections that have been quite hard to disentangle and that have likewise their origins in thermodynamics – here in the tendency for certain types of crystal structure to reappear under different circumstances.

Consider a piece of granite. It contains feldspars, quartz and some layer silicates – biotite probably and perhaps some muscovite. Now granite is strong stuff, but it does not last for ever; it is worn down by the weather and we can often see it disintegrating. We can see little pieces being broken off, being carried by rainwater to rivers to be further transported, further broken up and dumped in the sea. This grain of sand from the sea shore may well have originated as a quartz crystal in granite.

Much of the products of rock weathering that are carried by rivers is clay, very much finer than our quartz grain and structurally somewhat similar to the micas of such igneous and metamorphic rocks as granite. It is natural to suppose that here we are seeing another component of the original rock very finely ground down now, modified perhaps, but still essentially the same stuff. Examining a tiny illite crystal from a marine sediment, we might be persuaded that, with its structure so like that of muscovite, it must be a piece of an originally igneous or metamorphic muscovite crystal.

Such a conclusion would be far too hasty. Clays are not just bits knocked off hard rocks. For one thing, much clay is derived from parent minerals that are structurally dissimilar. Kaolinite is a frequent major weathering product of granite. Its structure is quite different from the feldspars that may provide most of its component atoms. As illustrated now so vividly with the scanning electron microscope, kaolinite clays grow from solution (figures 6.29 and 6.34). Often, this is through a series from less to more crystalline phases. In weathering there is not only decomposition but resynthesis to be reckoned with.

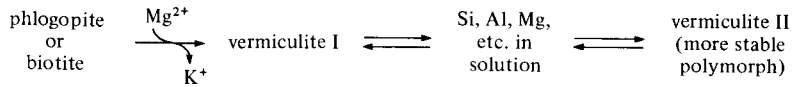
Again, in considering a succession of clay phases, we can easily be misled by short cut thinking. If kaolinite is like halloysite from which it often seems to be derived, one's first idea might be that the rolled up layers of halloysite straighten out, sweep off their water molecules and stack up into kaolinite. That might seem a sensible way of avoiding the trouble of making new layers. But Nature is not concerned with saving trouble: it is clear from the pictures that the kaolinite stacks are too neat to have been formed in that way. For silicate crystals forming at low temperatures it is not



always, or even usually, a question of one crystal gradually transforming into another: it is a question of one crystal dissolving to provide the nutrient for another (similar or quite different) crystal. Where the structures are similar (for example where halloysite changes to kaolinite itself) the second crystal may not so much remember the structure of its immediate predecessor as rediscover that structure in more perfect form. Seeding by the old phase or epitaxial growth on it may perhaps transfer information to the new phase, but general similarity of crystal structure (for example when both are in the kaolinite group) is no good evidence for this. Millot (1973) comments on such series that they 'testify the transition from disorder to order and seem to necessitate successive recrystallisations rather than unlikely thermodynamic transformations'.

Or consider the abundant illites of marine sediments. Their structure is close to that of muscovite but not necessarily because they are derived from it by minor transformations. More usually, indeed, they are derived from smectites (Jonas, 1976) which are much less like muscovite. Again, though, we must not suppose that to be 'derived from smectite' implies that the smectite layers stacked up on each other, changed sodium and calcium to potassium, some silicon to aluminium, and so on. It would be better to think of the smectite as possibly creating initial seeds but being mainly a reservoir of ionic units that were to reform via solution into brand new illite crystal (Jonas suggests as fringes on the smectite seeds).

True, there are sometimes direct successions, where layers are altered without being remade. You expect the product in such cases to be less well organised than the starting material – as perhaps sometimes when montmorillonite forms from micas. In the making of vermiculites, too, there can be, at least to begin with, a direct inheritance of major structural elements from a parent mineral: laboratory studies suggest that biotite or phlogopite can have their $K^+$ exchanged for $Mg^{2+}$ (aq.) without very seriously disturbing the silicate layers. But that may not be all that happens in Nature over long periods of time. Maybe a particular polymorph (with a particular arrangement of tetrahedral substitutions perhaps) that suits $K^+$ as the interlayer cation is no longer the ideal for $Mg^{2+}$ (aq.). If that is so, the resulting crystal will be slightly metastable compared with some more ideal polymorph. The first formed vermiculite, that is, will be slightly more soluble than this putative ideal. Eventually, given a seed of the more perfect form recrystallisation to it will be inevitable, that is:



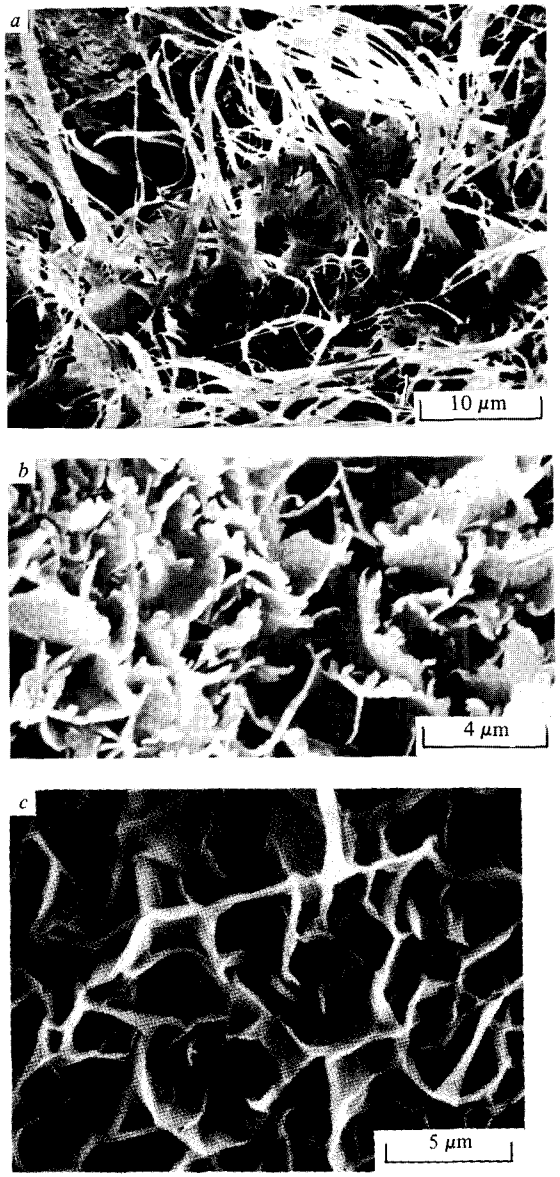



Figure 6.54. Scanning electron micrographs of some clays that have grown *in situ* in sandstones. (From Wilson & Pittman, 1977.) (*a*) Illite. (*b*) A mixed-layer smectite–illite. (*c*) Smectite. (*d*) Smectite with kaolinite. (*e*) Web-like smectite. (*f*) and (*g*) 'Honeycomb' and 'cabbagehead' chlorite. (*h*) Enlargement of (*g*). (*i*) Cliachite. (*j*) Cristobalite. (*k*) Iron oxide (?) grains with kaolinite.



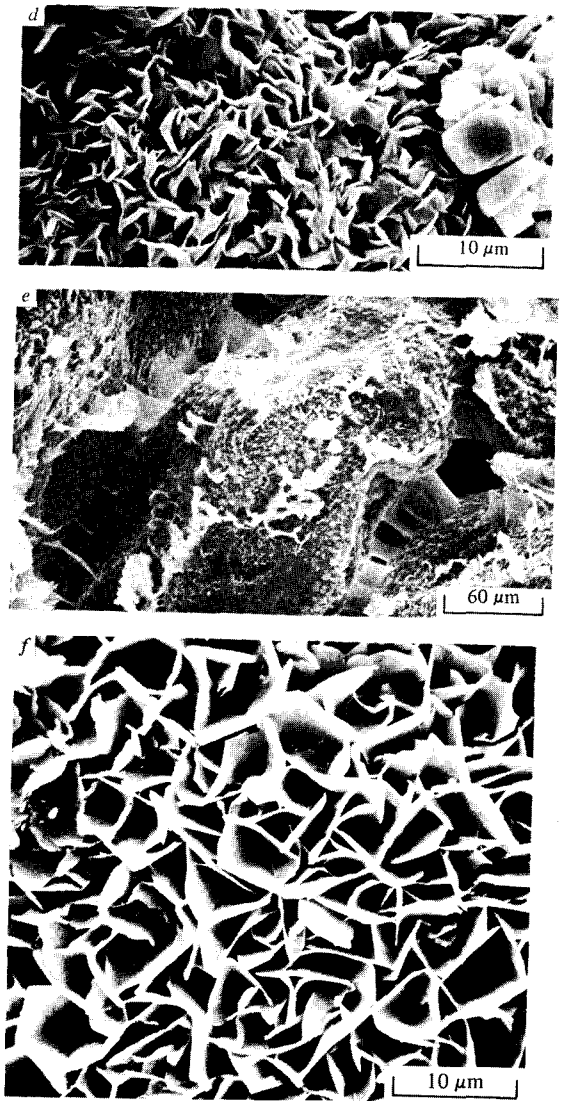



Figure 6.54.

Such a recrystallisation of an initial vermiculitised phlogopite might perhaps account for the paradoxical situation that Shirozu & Bailey (1966) noted in the Llano vermiculite and which we discussed earlier: this did not have the same detailed structure as the parent phlogopite – in particular, the arrangement of aluminium substitutions in the tetrahedral sheets was

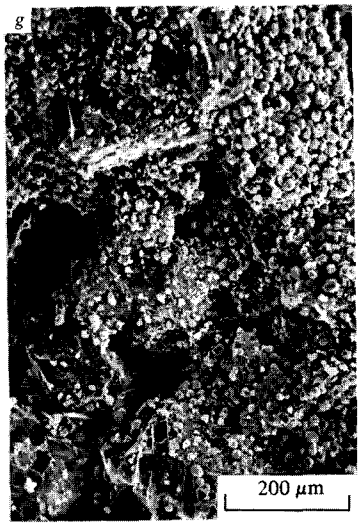


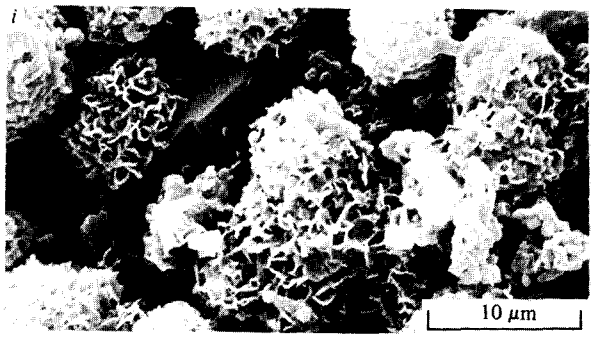


Figure 6.54 (*contd.*)

quite different. These authors remark that 'it is difficult to believe that sufficient energy would be available during surface weathering to order an original disordered Si, Al distribution'. Then again, Meunier & Velde (1979) have found that some vermiculitic mineral derived from the weathering of biotites in granites represents entirely new recrystallised material.

In sum, then, we may at least suspect that even a change in siloxane organisation that seems quite small, if it takes place at low temperatures,





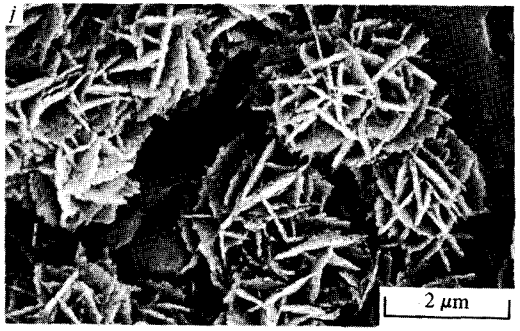





Figure 6.54 (*contd.*)

will call for a radical reconstruction because there is little if any scope for 'annealing' (e.g. through aluminium and silicon atoms hopping from one site to another). Only those units at surfaces and edges are accessible to reorganisation: 'Ostwalt ripening' seems to be the only way of creating new kinds of long-range order – that is to say, defective crystals or less stable crystal structures dissolve at the expense of more perfect crystals or more stable polymorphs. As with a piece of knitting, one can only put things right at the edges: to correct a faulty plain–purl arrangement in the

middle, there is nothing for it but to undo the structure and remake from scratch.

From an extensive study of sandstones, Wilson & Pittman (1977) found that authigenic clays are very common, indeed usual constituents. Of 785 samples from 55 formations, 91% were judged to contain clay crystals that had grown in the pores of the sandstone after the sandstone had been formed. The electron micrographs shown in figure 6.54 illustrate a rich variety of microcrystalline forms. The lath-like illite shown in figure 6.54*a* is attached like seaweed to a detrital grain. Figure 6.54*b* is a clay composed of both smectite and illite layers. The membranous character of smectite, and the tendency for authigenic smectite to form a compartmentalised or cellular structure, is illustrated in figure 6.54*c*. Another typical growth habit of authigenic smectite is as a crinkly coating on detrital grains. This is illustrated in figure 6.54*d* which shows also some kaolinite crystals. The membranes of smectite bridging detrital grains in figure 6.54*e* are particularly convincing indicators of authigenic origin.

Like smectite, chlorite can create a highly compartmentalised structure – as with the chlorite crystals coating a sand grain in figure 6.54*f*. The crystallites here are more clearly individual than with smectite. Figure 6.54*g* shows a barnacle-like growth of complex chlorite structures ('cabbageheads') shown at higher magnification in figure 6.54*h*. In addition to stacked plates there were occasional vermicular forms of kaolinite (cf. figure 6.29).

Those ubiquitous framework minerals potassium feldspar and quartz were also found to crystallise *in situ*, here adding new material to existing detrital grains. Sometimes there were irregular spheroidal aggregates of curled flakey material identified as cliachite – a form of aluminium hydroxide (figure 6.54*i*) and there might be minute rosettes of cristobalite (figure 6.54*j*) or rounded flakey aggregates (figure 6.54*k*) (thought to be a form of iron oxide).

There can be little doubt that these objects formed *in situ* from solution, although exactly when, over what period, and at what temperatures and pressures may be more difficult to judge. As such clays are so common, it seems unlikely that the conditions for their formation are particularly critical.

Velde (1977) says of soils that 'most clay minerals are found to develop in one soil profile or another'. Wilson & Pittman (1977) make a similar comment in relation to sandstones: 'all the major varieties of clay minerals are known to form authigenically in sandstone either as a direct precipitate from formation waters (neoformation) or through reactions between precursor materials and the contained water (regeneration)'. This constant



reappearance of a fairly limited group of crystal types within the porous matrices of the Earth's upper crust must have, in the end, a thermodynamic explanation. The typical minerals of the igneous and metamorphic rocks, particularly those formed at the highest temperatures such as olivine, are, on the whole, the least stable when subjected to conditions so different from their conditions of formation. They tend to dissolve in water and recrystallise into more stable phases. Among the most stable are microcrystalline materials – clays – although exactly which clays appear where, may depend in complicated ways on conditions that change from place to place and with time.

### *Laboratory studies*

I have referred already to a number of microcrystalline minerals that can be made fairly easily in the laboratory at ordinary temperatures, for example ferric oxide-hydroxides, gibbsite and iron-bearing 1:1 clays. Kaolinites, illites and smectites have proved more difficult in spite of being the dominant clays in Nature, formed mainly by weathering under surface conditions (Harder, 1977). A main reason for this difficulty would appear to be that, at ordinary temperatures, clay synthesis is only successful from very dilute solutions. Furthermore, the reactions are slow, often needing periods of months or even years. These studies were initiated mainly by French chemists (see, for example, Caillère & Hénin, 1948, and papers in Hocart, 1962; an account of recent developments is given by Siffert, 1979).

It was soon evident that clays containing aluminium – by far the most important class – were particularly tricky. Reviewing this topic, Siffert & Wey (1973) saw the main problems in the strong hydration of aluminium and in the insolubility of aluminium over the near-neutral pH range. Wey & Siffert (1962) had managed partly to overcome these problems in the synthesis of 'protokaolin' by decomposition of the aluminium oxalate complex – to supply suitably hexa-co-ordinated aluminium in solutions containing monomeric silica. Fulvic acid – a complex mixture of organic molecules obtained from peat – was also effective as a complexing agent for aluminium in low-temperature kaolinite synthesis (Linares & Huertas, 1971; La Iglesia & Martin-Vivaldi, 1973). Kaolinite was also obtained using ion-exchange resins (La Iglesia & Martin-Vivaldi, 1975) or feldspars (La Iglesia, Martin-Vivaldi & Lopez Aguayo, 1976) to generate a slowly changing pH as a means of maintaining a suitable and homogeneous level of supersaturation. In this last case, micas and traces of smectite were also obtained.

Harder (1977) discusses another way of creating conditions suitable for

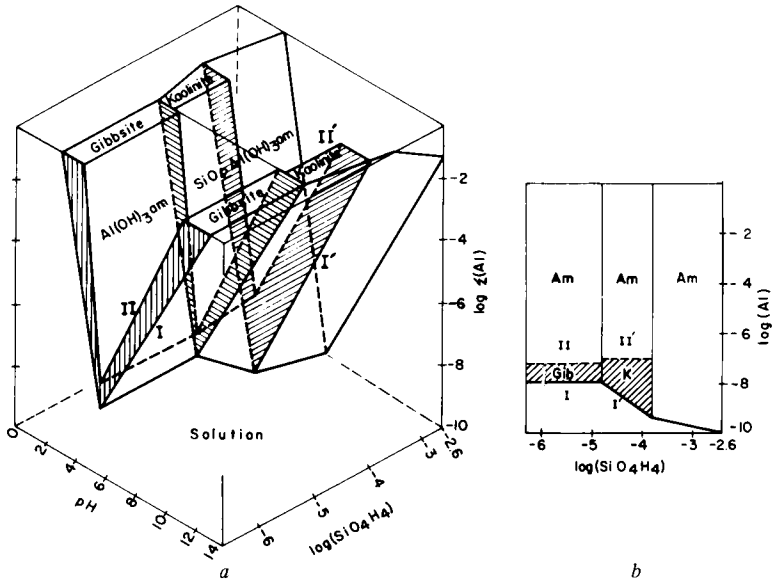



Figure 6.55. (*a*) Stability diagram for the system silica–aluminium–water at 25 °C and 1 atm. The solid phases considered are gibbsite (I), amorphous aluminium hydroxides (II), kaolinite (I′), and the corresponding amorphous silico-alumina gel (II′). (*b*) Cut at pH 5.6 from (*a*). Hatched zones correspond to the crystalline phases gibbsite (Gib) and kaolinite (K). Am corresponds to the different amorphous phases indicated in (*a*). (From La Iglesia & Van Oosterwyck-Gastuche, 1978.)

clay synthesis. When hydroxides are precipitated from solutions containing silica the latter is strongly chemisorbed and mixed silica-hydroxides are formed. Many of these start to generate clays after a few days ageing at room temperature. It is necessary that silica concentrations in the initial solutions are low – undersaturated with respect to amorphous silica (less than 100 p.p.m. at 20 °C). Also precipitation should be slow. This can be achieved by the kinds of homogeneous precipitation methods mentioned above – slow change of pH or decomposition of complex ions – or alternatively by oxidation (e.g. $Fe^{2+} \rightarrow Fe^{3+}$; cf. Harder, 1978). Also, the overall composition of the precipitate should be similar to that of the clay mineral being made. Not all insoluble hydroxides are effective. The best results are obtained where the radius of the cation is between 0.083 nm and 0.078 nm (as for $Mg^{2+}$, $Ni^{2+}$, $Co^{2+}$, $Zn^{2+}$ and $Fe^{2+}$). With a radius of 0.057 nm, aluminium is seemingly too small, but even this can be incorporated in solid solutions with the more easily made clays. Harder lists about two



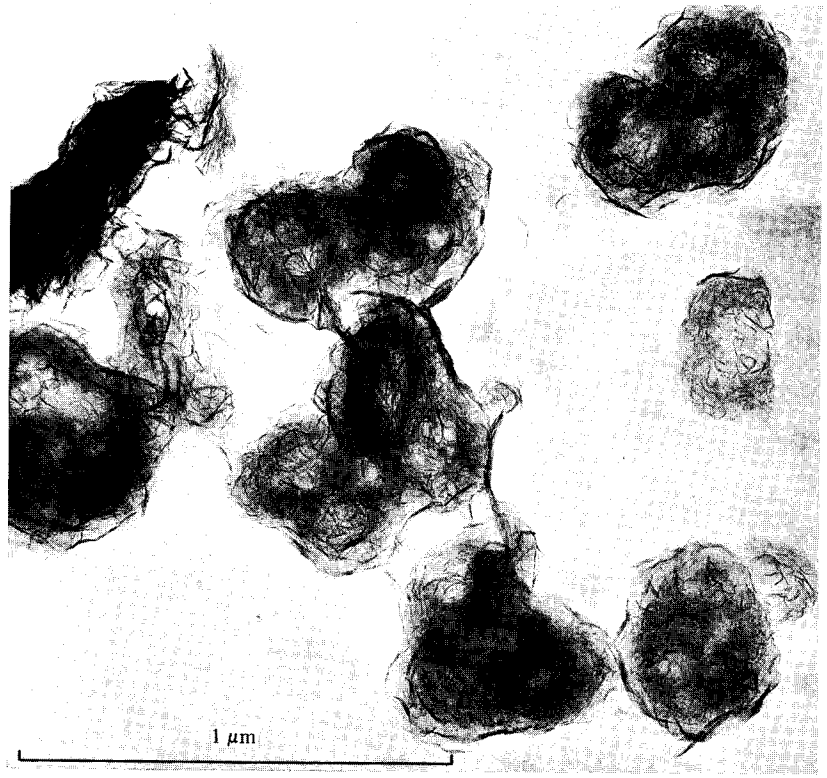



Figure 6.56. Synthetic hectorite. Transmission electron micrograph showing partly transformed brucite crystals, 'doughnuts', enclosed in hectorite 'bags'. (From MacKenzie, 1971.) This stage was reached after 6 days reflux of a slurry of composition $SiO_2$:MgO:LiF = 0.01:1.0:0.25 in pyrex glass.

dozen clays of all four layer types that have been made in this way. Kaolinite is the most notable absentee.

La Iglesia & Van Oosterwyck-Gastuche (1978) have discussed thermodynamic restraints on conditions for kaolinite synthesis. They stress the need for very dilute solutions. Figure 6.55 shows furthermore how narrow is the region of concentrations and pH for kaolinite synthesis. Even when thermodynamic conditions are satisfied it is clear that other factors can stand in the way of synthesis – in particular, suitable nuclei for crystal growth may have to be present. In another paper, Van Oosterwyck-Gastuche & La Iglesia (1978) review laboratory syntheses of kaolinite at various temperatures and stress that it is the laws of crystal growth rather than Arrhenius kinetics that must be considered.

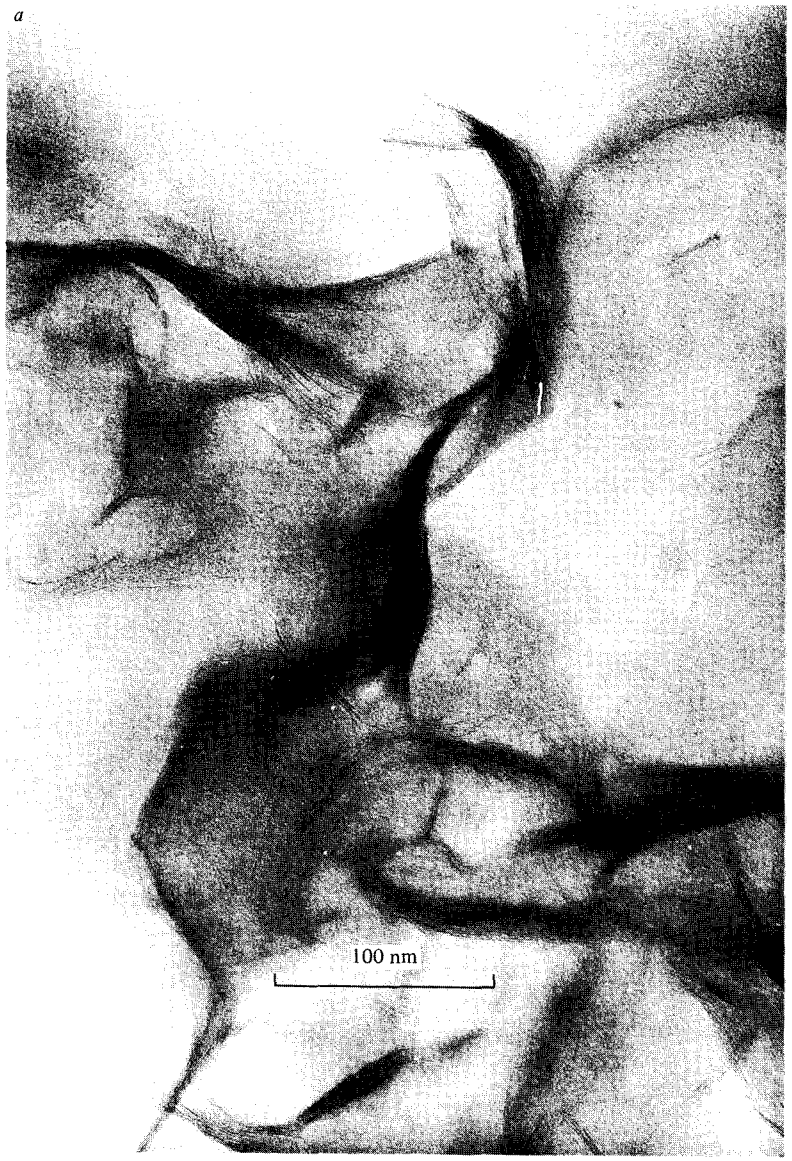



Figure 6.57. (*a*) A synthetic hectorite. The typical smectite appearance is similar to that in figure 6.54*c*. (From Baird, Cairns-Smith, MacKenzie & Snell, 1971.) Conditions: heating for 7 days at 90-95 °C in polypropylene: $SiO_2$:MgO:LiF = 2.0:1.0:2.5. (*b*) At a magnification of a million, 2:1 layers are clearly visible when they happen to lie edge-on. This is part of the picture (*a*) in which the typical smectite fabric can be seen as arising from flexible layers that have become connected together through regions in which they are stacked on top of each other.



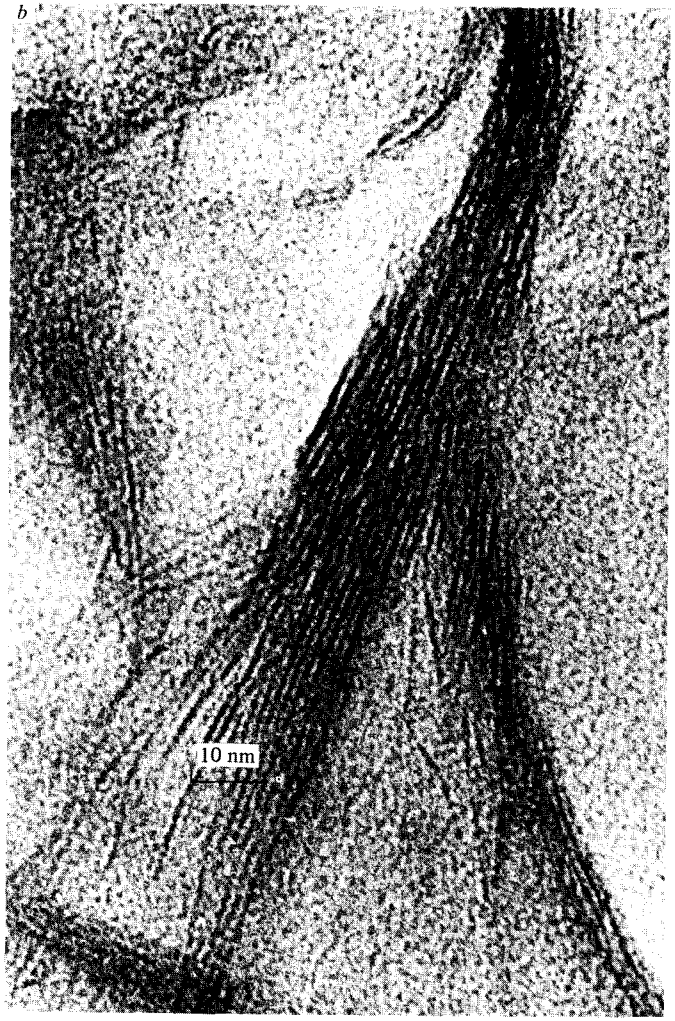



Figure 6.57.

Baird, Cairns-Smith & MacKenzie (1973) made an electron microscope study of the efficient and rapid synthesis of hectorite that had been discovered by Granquist & Pollack (1959) – by boiling slurries of $Mg(OH)_2$, $SiO_2$ and LiF. In this reaction, brucite crystals are quickly attacked by silica to give an intermediate material with some of the properties of smectite. The membranous hectorite that appeared after about 24 h was much more extensive than the initial brucite crystals (figure 6.56). It had evidently crystallised from solution, and seemingly at the expense of the 'pseudo-smectite' which had partly redissolved. A more mature product is shown in figure 6.57.

Such intermediate materials as 'pseudo-smectite', and the similar amorphous silicified hydroxides used by Harder, can be thought of as reservoirs of suitable units for subsequent crystallisation from solution. The units themselves may be monomeric, or possibly they are small oligomers particularly suited to the synthesis of clays – hence perhaps the importance of the conditions of formation of the intermediate materials. Siffert & Wey (1973) have suggested that aluminium clays are made from oligomers such as:

$$(HO)_3\text{–Si–O–}\{Al(OH)_2(H_2O)_3\}$$

or

$$(HO)_3\text{–Si–O–}\{Al(OH)\,(H_2O)_3\}\text{–O–Si–}(OH)_3$$

to give respectively 1:1 and 2:1 clays. But whatever the nature of the units, the burden of the laboratory studies has been that clay synthesis from aqueous solutions takes place through unit-by-unit crystal growth mechanisms.

It is generally true that for accurate crystal growth a solution should be only barely supersaturated with respect to the perfect crystal, so that at the same time it remains undersaturated with respect to imperfectly formed crystal. This is the basis of the 'error correction' mechanism that we talked about in the last chapter. If levels of supersaturation are raised then the number of kinds of imperfect organisation that can persist increases: the resulting crystals are increasingly flawed – until eventually there is more flaw than crystal and we describe the precipitate as 'amorphous'.

On the whole, it is the more insoluble materials that are most inclined to give amorphous precipitates – and clays such as kaolinite are very insoluble at ordinary temperatures. For clay synthesis at around 20 °C, it is evidently quite a short step between solutions being too dilute for anything to come out and too concentrated to allow the appearance of anything recognisably crystalline.

But we should remember, too, that making a clay is not just a crystallisation – not just an orderly emplacement of units as is, say, the crystallisation of sucrose. It is also a polymerisation. Covalent bonds have to be made and broken in the process with several water molecules having to be eliminated as each unit adds. Although little is known here about the details, it is generally the case that to make and break covalent bonds rather particular orientations of groups are required and activation energy barriers have to be surmounted. In these circumstances, we can imagine metastable polymers forming and persisting – as in the polymerisation of silicic acid to silica gel (discussed on pp. 181–2). We can imagine the orderly unit-by-unit processes of crystal growth easily becoming blocked by tangles of polymer attached to the crystals. Only at concentrations low enough (at least) to prevent spontaneous net polymerisation in solution would you expect such tangles to be unstable – for the crystal surfaces to be cleared to expose those relatively few sites at which orderly growth can proceed.



*Complex systems in Nature* In view of the rather narrow range of conditions often needed for low-temperature clay synthesis, we might wonder how it is that Nature so often seems to get it right. There are perhaps two general answers to this question. First, the weathering solutions from which clays form are rather dilute, so the conditions are likely to be roughly right in that respect at least. Then, secondly, Nature only has to get it roughly right – at least in the first place. The most stable phases – for example well crystallised kaolinite – may only be arrived at after a succession of other minerals that are simpler or less fussy about the requirements of orderly growth. Millot (1973) gives allophane → halloysite → disordered kaolinite → ordered kaolinite as a typical sequence.

To a first approximation, then, we can perhaps consider the synthesis of clays in Nature as taking place as follows. The materials that come out of solution from freshly weathered rocks tend to be metastable and poorly organised; these eventually come into quasi-equilibrium with their surrounding solutions which are now at a lower level of supersaturation than before. New, more orderly phases crystallise from these solutions, so reducing their concentrations still further so that still more orderly phases become kinetically possible. Each new phase reduces the level of supersaturation of the surrounding solutions to just above that at which it can grow – a level at which the previous phase dissolves. More stable, more insoluble, more orderly phases can thus be arrived at. Thermodynamics eventually gets its way.

Indeed, that must be too simple a picture: solutions in contact with minerals in Nature are not necessarily in quasi-equilibrium with those minerals. In open systems, species in solution may be added or subtracted so that the new phase that appears may have a different chemical composition. To make matters more complicated, solid particles may be being transported. Also, there may be seasonal variations – alterations of high and low silica levels in solution, changes of pH, and so on. Thus, different minerals may be favoured in the same place at different times. In discussing conditions for kaolinite *versus* smectite, Johnson (1970) points out that, while a general environment might favour one, there could nevertheless be micro-environments that were favourable to the synthesis of the other. In any case, smectite and kaolinite are often found together in spite

of supposedly different conditions needed for their synthesis. We should remember that while 'pure' clay deposits may catch the eye, clay minerals occur mainly in complex assemblages.

As yet a further complication, the growth of a new phase may very well affect the environment in which it is forming. For example, a clay may clog up the pores of a sandstone within which it is forming, altering the patterns of solution flows and hence the patterns of variation of species in the solutions: if they stop moving, pore solutions will come nearer to equilibrium with the surrounding rocks. Hence, and in other ways, long-term variation can be superimposed on shorter term fluctuations: a mineral assemblage may 'evolve' in a manner that is perhaps analogous to the 'evolution' of an ecological system – in so far as we can imagine a succession of species, with each predisposing the conditions for the next. This may go on indefinitely or alternatively the system may stabilise if, by chance, an assemblage is arrived at which happens to stabilise the environment for itself. A rather simple case in which that may have happened was found by Clemency (1976). He found two different kinds of clays alongside each other, a smectite and kaolinite, each apparently self-maintaining. Clemency suggests that the relatively impervious smectite would favour a slow percolation of water and so a higher concentration of silica, thus stabilising smectite, while the more porous kaolinitic rock would be in equilibrium with the more dilute solutions passing more rapidly through it. 'The pore water characteristics within each rock type thus impresses its own particular equilibrium constraints on the clay minerals forming within each rock type.'

### Conclusion

With this reminder that the conditions for synthesis of clay minerals in Nature are often rather complicated, and that the products are often mixtures of many different crystal types, we come to the end of this chapter. My main objective has been to illustrate the variety and complexity of clay mineral types; to persuade those of my readers who might have needed persuading that *mud* can be interesting, highly structured stuff. In Part III, I will argue that several particular features of clay mineral structures correspond to what is really needed for the components of organisms that could generate spontaneously, evolve under primitive conditions and then, through photosynthesis, begin to invent an organic biochemistry.

# PART III

# A new story

# 7

# First life

## Introduction and review

In this chapter we will start to pick up some of the threads from earlier chapters to try to arrive at some more specific speculations about how life on Earth might have originated. This will be to illustrate the general idea that our first ancestors might have been inorganic minerals, rather than with much hope of the details being correct. Like the fictional details of an historical novel that help to make more real some long past way of life, it is perhaps only when one tries to see how a mineral life might have worked, in terms of what we happen to know now about the details of mineral structures and synthesis, that such a way of life becomes thinkable.

Even just to think about thoroughly non-nucleic acid–protein life styles may be a help – because it was the main conclusion of Part I of this book that first life would almost certainly have been of an altogether different kind from ours: that it would have been made from different basic units that were put together differently and worked differently. However unpalatable Dr Kritic and others may find this idea, it seems to me that an assertion of original biochemical *dis*similarity is an essential starting point from which to try to understand the origin of life. We must detach our fingers from one guide rail before we can grasp others that can lead us much closer to the answer.

As we saw, this 'unpalatable' proposition is in any case based on common sense. Whatever else, organisms are machines: life is natural engineering – and engineering principles apply. You do not expect first ways of doing things to be like more sophisticated ways – not in mechanical details anyway. First life would have been different from us, not so much because

conditions on the early Earth would have been particularly different from now, but because the specifications for a first organism are quite different from the specifications for an advanced organism. The first organisms had to start up a new enterprise with no pre-existing technology. Later, evolved techniques of fabrication could be presupposed.

As we saw also in Part I, a too facile view of the way evolution works stands in the way of a general acceptance of this common sense. It is often supposed without argument that the unity of our central biochemistry makes a powerful case for the absolute invariance of certain small molecules for life, or at least for life on Earth. Alanine, glucose and so on are seen as invariants through evolution, analogous perhaps to the atoms that are invariant through chemical reactions. Amid all the changes that evolution can bring about, at least these basic small molecules can be relied upon – so it is said.

That may be true now, but there is no more an absolute law of the conservation of alanine in biochemistry than there is a law of the conservation of 1 cm right-handed star screws for automobiles. The invariance of our central biochemicals arises from their multiple and intricate interdependence in our present biochemical machine. Biochemical invariance is evidently a system invariance; it would have appeared as the system evolved. But there would have been no rules till that happened. And no foresight either – no way of knowing what was going to be just the thing in the remote future.

It is not as if mechanisms for radically updating original design approaches cannot be imagined through the early evolution of our biochemistry as they can be seen still in operation at higher organisational levels. Evolution is not always restricted to the optimisation of some initial set of subsystems. New subsystems have evidently been invented. As discussed in Chapter 3, this arises through switchings between structures and functions (as well as redundancies) in conjunction with the adaptation of structures to functions. That is the main inventive technique of evolution, the way quite new design approaches can be arrived at. Eventually some elaborately interdependent set of subsystems becomes frozen in, and the period of radical invention is over. If our central biochemistry may seem always to have been frozen, this is an artefact of our point of view: the last common ancestor of all life on Earth now happens to be more recent than that time at which the central machinery of biomolecular control became rigid.

As for the argument that the components of our life were all ready and waiting on the primitive Earth, it is not clear in the first place that any of our central biochemicals would have been present on the Earth in reasonable amounts. And in any case the interpretation of experiments supposed to show this is ambiguous. Evolving organisms would tend to follow biochemical pathways of least resistance, and these would tend to be like abiochemical pathways; and stable molecules would tend to be preferred biochemically and abiochemically. This can explain why some of our smaller biomolecular components keep on turning up all over the place (inside and outside organisms). On the other hand, some very critical structures that are central to our biochemistry – nucleotides and lipids, for example – are difficult to make; do not turn up in plausible simulations of conditions on the primitive Earth; would, as far as one can see, need an evolved biochemical system to produce them consistently or polymerise them in an organised way. Needless to say the evolved biochemical system that synthesised such molecules in the first place did not absolutely depend on these same molecules – they must have been useful to begin with rather than vital. The evolved biochemical systems that we must postulate were organisms of another kind, with takeover mechanisms, in particular genetic takeover, providing the continuity between them and us. The continuity of evolution can be compared with that of a long rope where no one fibre stretches from one end to the other. Or we might see it as resembling the continuity of a business enterprise where staff, premises, even the nature of the main product might have changed over the years without any sudden point at which everything changed.



Or we might use the ancient idea that life is a kind of fire, to illustrate the central theme of this book. How do you set fire to coal? Often by setting fire to paper. First you have to get *something* burning. Evolution has been a kind of spreading fire and the problem of the origin of life can be seen as a problem about how something started to evolve, to transform matter into 'survival machines', in Dawkins' phrase, and to build up a 'technology'. Then, inevitably, first ways would be improved on and displaced. Sooner or later the coal would be alight with no trace left of the paper.

There are many other examples of processes for which beginnings are special and may involve mechanisms that are then dispensed with (think of booster rockets, starter motors, detonators, eggs). Particularly for any sort of 'evolutionary' process – where one thing leads to another – the process itself can easily be transformed by its own effects (think of the progress of a game of chess or a piece of music). This all seems to be so much common sense that we should surely stop insisting in the name of Occam or anyone else that life would have started in anything like the way in which it was eventually to be most successful. It would not. It would not have known how to. Let us insist rather that there must have been *organisms* that in-

vented our biochemistry: that the origin of our kind of molecular life is to be distinguished from a remoter origin on which that depended. Darwin's idea must be taken further back, I think, and away from present biomolecular hardware to a different region of chemistry altogether. And there, somewhere there, is to be found the true origin of life.

That was the gist of Part I of this book. In Part II we started looking for the appropriate region of chemistry. That search was started in earnest in Chapter 5, although Chapter 1 had already suggested that organic minerals might have been inappropriate starting materials. We concentrated attention on the hardest part of the problem, on the one essential piece of machinery for any organism able to take part in long-term processes of evolution. Whatever else, if the 'fire' was to keep alight through millions of years, if 'know-how' was to be retained and improved on so that ultimately our nucleic acid–protein–lipid machine could be made, there had to be some sort of accurate information-printing machinery in those very early organisms. The main clue to the nature of this first machinery is that it must have been able to put itself together without any evolved machinery of any sort already there. From such considerations, and by thinking about bonding characteristics in organic and inorganic molecules and crystals, crystals with covalent bonds in them seemed a much better bet than organic polymers. Our course was thus set towards inorganic chemistry – towards colloidal minerals that could crystallise from aqueous solutions at around ordinary temperatures. We arrived at clays, defined rather broadly.

Chapter 6 provided more detailed information and ideas which will form the basis of a further discussion in this chapter on the nature of the primitive genetic material. The accounts of mineral structures given in Chapter 6, particularly the electron micrographs, also served to illustrate that even without genetic information the common colloidal minerals often have the forms of membranes, tubes and such things – that here we have, perhaps, potential phenotypic apparatus for first life, a clue to a rough but ready means of making some of the hardware of early organisms. We will also take this idea further – in this and in the next chapter. But since, to be relevant, the synthesis of any such hardware must have been to some extent under genetic control, it is with possible genetic structures that we must start.

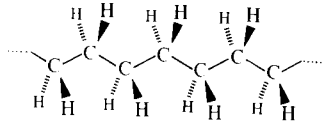
### Information storage

In looking for a primary genetic mineral we can apply first the constraint that, whatever it was, it must have been able to store information. An object can in principle store information if it can be in one of a

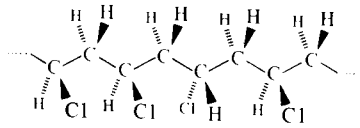


large number of alternative, reasonably stable configurations. A book stores information because there are many ways of arranging letters, etc. on pages, and a particular book is just one such way. Similarly, a DNA molecule stores information through having one of a very large number of possible base sequences. On the other hand, a perfect crystal could not store information because there is only one way of being perfect; nor does a drop of water, because although there are many ways of arranging molecules in water none of them lasts.

The slightest feature can make the difference between a structure having zero information capacity, and its having a very large capacity. For example, you might suppose that there would be no way of storing information in a straight polyethene molecule:



In practice that may be so, but in principle you could store more information in a polyethene molecule than in the base-pair sequence of a DNA molecule of the same weight. Since one in every 90 or so carbon atoms is $^{13}C$, information could be stored as a particular arrangement of these atoms on a long polyethene chain. Similarly, polyvinyl chloride (PVC) could hold much information if the chlorine isotope sequence could somehow be specified and read. But here there is another way that does not depend on there being any two-letter or *n*-letter alphabet. A sequence of As could hold information if some of the As were printed upside down. Similarly, (atactic) PVC could hold much more information than the same weight of DNA – through a particular left–right sequence of chlorines along the chain:



#### *Crystals as information stores*

There are very many ways in which real crystals (as opposed to the text book ideals) could hold information. Real crystals almost invariably contain a multitude of defects of various sorts. Over and above its regularly repeating **crystal structure** any given crystal will have various irregular

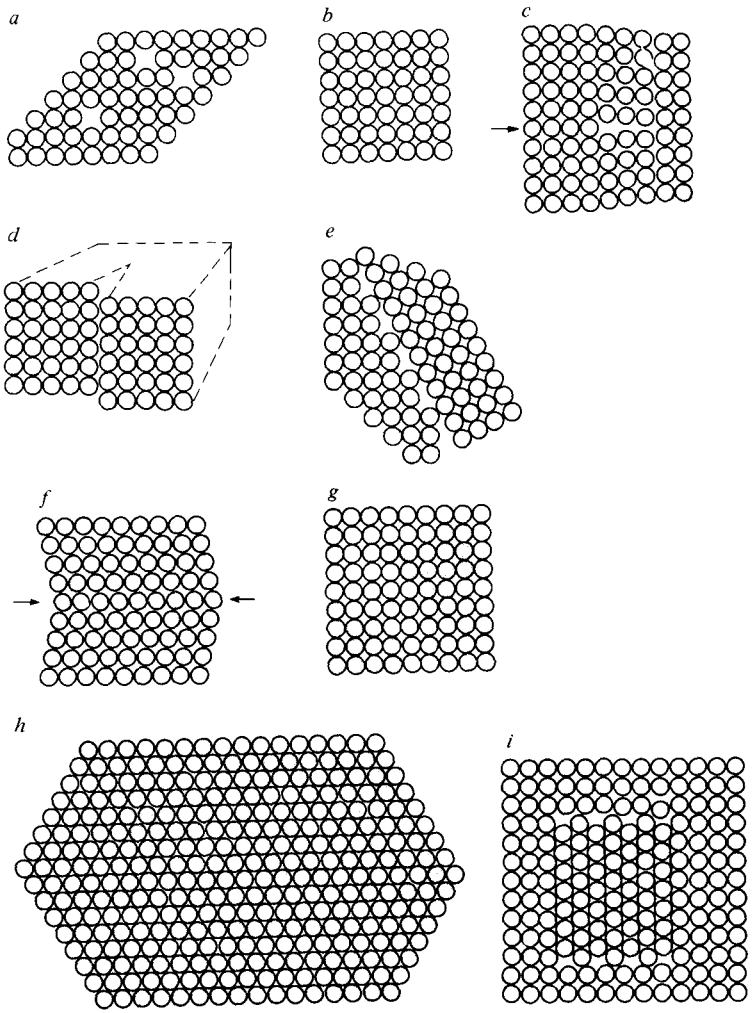



Figure 7.1. Common types of defect structure in crystals. (*a*) Vacancies. (*b*) Substitutions. (*c*) An edge dislocation showing an extra plane of units. (*d*) A screw dislocation (cf. figure 6.43). (*e*) A grain boundary where similar but differently oriented lattices meet. (*f*)–(*i*) show more intimate forms of crystal intergrowth. (*f*) As in (*e*) except that the misorientation is such that there is a common plane of units at the junction (the composition plane, arrowed). Such crystals are twinned. (*g*) As with (*b*) except that the substitutions are in domains. (*h*) Another kind of domain structure: here the domains have the same overall composition but differ in the alignment of units. (*i*) Different crystal forms of the same material may be intergrown, or alternatively the intermingled phases may be chemically distinct. Here there is a domain of one phase within another.



features – it will have some sort of **defect structure**. It may contain point and line defects such as **vacancies**, **substitutions** and **dislocations** (figure 7.1*a–d*). The positioning of such features could in principle represent specific information analogous to a DNA base sequence or to the arrangement of letters on a page. Alternatively, information might be held on a somewhat larger scale. A crystal usually consists of a mosaic of at least slightly misaligned blocks; and very often crystals have regions, or **domains**, that are grossly misaligned (e.g. figure 7.1*e*). In **twinned crystals** there is at least one major misalignment, such that at least one plane of atoms – **the composition plane** – is shared by adjacent domains. Twinning is very common, particularly in minerals. Twinning and other forms of **crystal intergrowth** are illustrated in figure 7.1*f–i*.

Crystal defect structures may sometimes be visible to the naked eye – for example twinning is often visible. But such structuring goes down to a very small scale indeed, as the high-resolution electron microscope is revealing increasingly. It is possible to see minute heterogeneities in most sorts of crystals that have been examined – a row of atoms missing, a slight misalignment, and so on. Very often there is a profusion of such tiny imperfections. Such features can be described as **ultrastructures**. (For a review of the ultrastructure of minerals see Hutchison, Jefferson & Thomas, 1977.)

Domaining can be seen to be very common – and is often on a minute scale. Some of the very best pictures have so far come from high-temperature materials. In the 1 MV electron microscope image of tungsten–niobium oxide ($4Nb_2O_5 \cdot 9WO_3$) shown in figure 7.2*a*, rows of atoms can be seen, as can the changing directions of the rows in different minute domains (figure 7.2*a*).

In figure 7.2*b* one can see a misalignment of blocks in a $\beta$-FeO(OH) 'somatoid' that was formed at low temperatures. (Look also at the picture of a sepiolite crystal in figure 6.48*b*.) Another low-temperature mineral, beidellite, reveals a complicated mass of dislocations (figure 7.2*c*; compare figure 8.7). Very numerous microtwin boundaries can be seen in the pyrrhotite crystal shown in figure 7.2*d*. Zeolites too are beginning to reveal a rich ultrastructure (Bursill, Thomas & Rao, 1981).

Zeolites, and more especially feldspars, can have complicated twin structures. Feldspar crystals are generally highly twinned and various types of twinning have been classified according to the orientation of the composition plane and the symmetry relations between the parts of the crystal on either side of the plane. Empirical 'twin laws' for feldspars were given a crystal structural interpretation by W. H. Taylor: across a composition plane there is some change in the structure of the alumino–siloxane frame

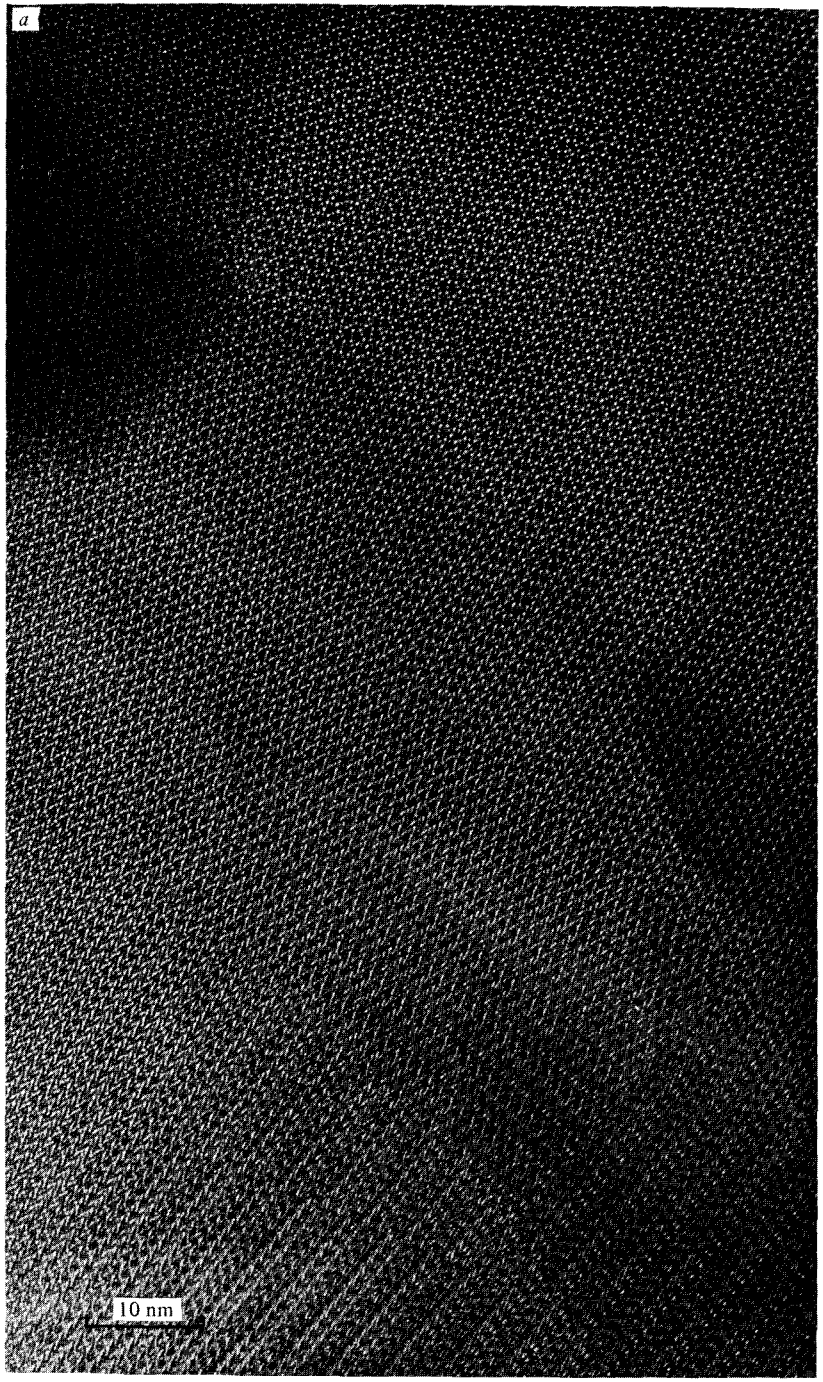



Figure 7.2*a*. *For explanation, see caption on p. 270.*



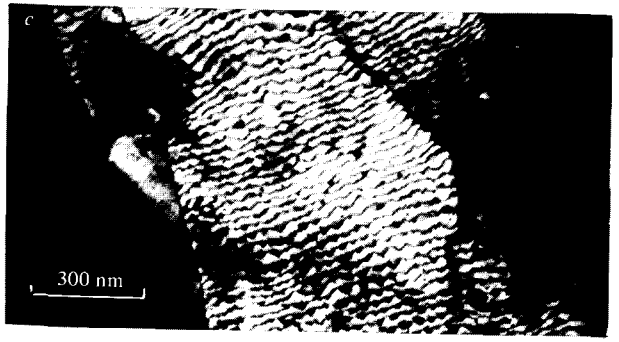




Figure 7.2*b* and *c*. *For explanation, see caption on p. 270.*

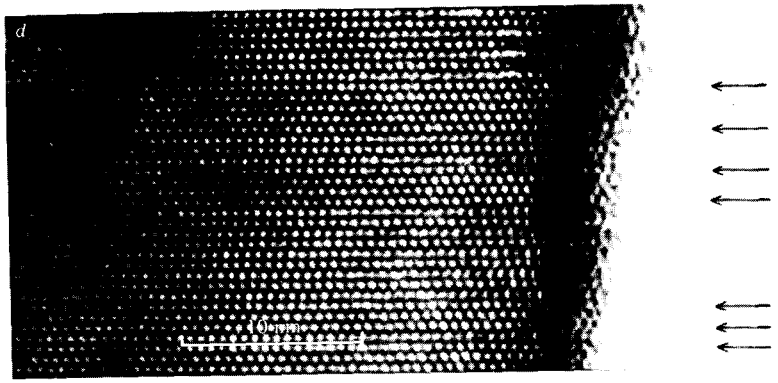



Figure 7.2. Electron micrographs showing small-scale defects in crystals (see text). (*a*) $4Nb_2O_5 \cdot 9WO_3$ (from Horiuchi, 1978). (*b*) Cross-section of $\beta$-FeO(OH) crystal (from Galbraith, Baird & Fryer, 1979). (*c*) Dislocations in a beidellite crystallite (from Guven, Pease & Murr, 1977). (*d*) Microtwin boundaries in pyrrhotite (arrowed) corresponding to changes in the arrangement of iron atom vacancies (from Nakazawa, Morimoto & Watanabe, 1975).

which, although a major change in the sense that a wholesale reconstruction of the lattice would be needed to eliminate it, nevertheless involves little distortion of covalent bond angles and lengths. There are several such subtle structural modulations that can be correlated with the classical 'twin laws'. Interpretations of the 'Manebach law' and the 'Baveno law' as given by Taylor, Darbyshire & Strunz (1934) are illustrated in figure 7.3.

We have already discussed the idea, in Chapter 5, that *chemical* defect structures might have provided the basis for a crystalline genetic material. As we saw in the last chapter, it is very common indeed for there to be some choice about which ion is present at some particular site in a mineral lattice. And for many minerals, especially silicates, such a chemical heterogeneity is an essential feature. Montmorillonite would not have negatively charged layers that can disperse in water if there were not a suitable degree of substitution of cations within the layers – in that case it would simply not be montmorillonite.

*Information and metastability* To store information the 'defects', whatever they are, must stay put: they must not move about inside the crystals as the vacancies and dislocations in metals easily do. (Metals are malleable for this reason.) There are signs, though, that the (much less malleable) silicates and metal oxides are not nearly so internally mobile as metals – that at ordinary temperatures the main Si–O and M–O bonding



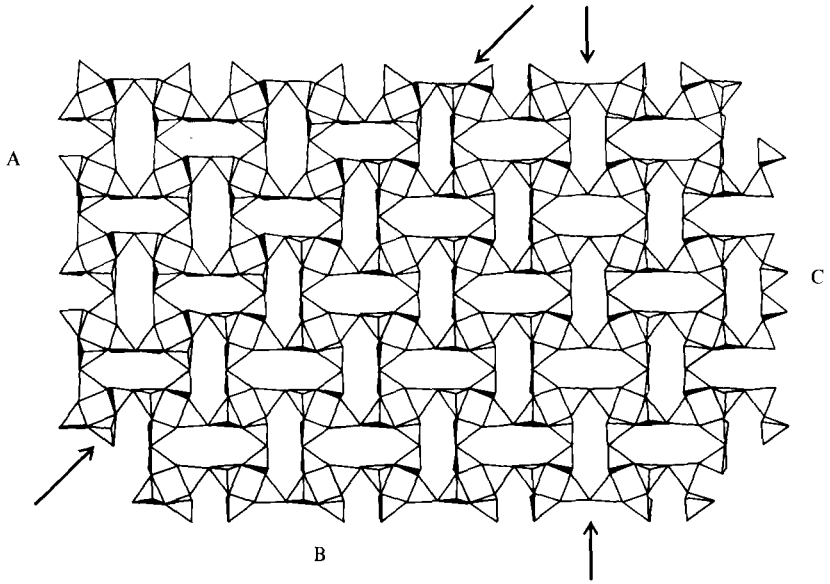



Figure 7.3. Two of the kinds of twinning in feldspar. 'Baveno twins' are related as between sections A and B; 'Manebach twins' are related as between B and C. Arrows indicate the composition planes.

can be maintained within crystals over very long periods of time. One sign is the common persistence of metastable silicate and oxide phases at ordinary temperatures. The technique of thermoluminescent dating also depends on metastable features in crystals persisting indefinitely at ordinary temperatures. This technique depends on the accumulation of radiation damage within silicate and oxide crystals. The damaged crystals cannot anneal except at somewhat elevated temperatures when they lose their stored energy in a burst of heat and light.

Tracks left by high-energy particles in micas and many other minerals represent minute metastable features which can persist for millions of years at ordinary temperatures (Fleischer, Price & Walker, 1975).

As we saw in the last chapter, mineral crystals can reform, via aqueous solution, at ordinary temperatures: but that is no bad thing since this kind of limited reformation could have provided a mechanism for correcting errors in the replication of information in crystals without necessarily losing the information itself (Chapter 5). What is important is that deep inside the crystals there is no wholesale mobility of ions – that at least some of them can stay put for long periods.

From considerations of the site energies of different ions in mineral

crystals it seems clear that at least $Al^{3+}$ and $Si^{4+}$ are very difficult to move at ordinary temperatures (Dowty, 1980). As an indication of this, Sipling & Yund (1974) found that the activation energies for $Al^{3+}/Si^{4+}$ disordering in some potassium-rich feldspars lay between $378 \pm 42$ and $483 \pm 55$ kJ $mole^{-1}$. The disordered forms of these feldspars become stable at about 500 °C; but for the rates to be conveniently measurable the experiments were performed at around 1000 °C. This is typical: high temperatures – hundreds of degrees – are generally needed for wholesale, diffusive movement of atoms in silicates, as in many other kinds of oxide crystals.

Although it is a sign that a certain kind of crystal might be able to hold information if it can remain highly metastable over long periods of time; and although it is true that only metastable structures can hold information, it is not true that there is any quantitative relationship between how metastable a structure is and how much information it might contain. There is very little difference in energy between all the ways possible of arranging letters on a page. A good information store is like that: energy is not an important factor. To be sure, there will be one fixed configuration that will have a lower energy than any other (there has to be a best somewhere) but it should only be minutely more stable and there should be vast numbers of configurations that are, energetically speaking, virtually just as good. If poly(AT), say, was thermodynamically far more stable than any other arrangement of bases in DNA, selection would have had to fight against a continuing tendency for stretches of DNA to 'revert' to poly(AT): one reason why DNA is a good information store is that the energy is not greatly affected by the base-pair sequence. So we might take it as a promising sign for a crystal information store if different configurations of the features that might hold the information have similar energies.

The idea of solid-state information storage is not new to the electronics industry. A magnetic tape, for example, contains information in the arrangement of crystal domains. The problems here are not so much in finding oxides, etc. that could somehow hold information, as in the techniques of writing it and reading it. In looking for a solid-state genetic material the corresponding problems are of how the information could have replicated and how it could have exerted a control on its environment. The first of these questions can already give us a very tight set of constraints.



**Replication**

We decided in Chapter 2 that only template replication would been an efficient enough means of transferring information between generations of primary organisms. DNA replication is like this, although the templating process is strongly assisted by proteins. In Chapter 5 we discussed more particular problems associated with template replication. There are four general requirements whether polymers or crystals are being thought of: suitable units must be provided; there must be some kind of matching-up mechanism between units and template; there must be a way of locking the units together, and a way of separating the new copy (or copies) from the original template. So we can divide the question of primitive genetic replication into four subsidiary questions.

*1. How were the units supplied?*

The task for the environment was to maintain a supply of units that were energetically uphill from a genetic polymer or crystal. New genetic material must have tended to form spontaneously (although it must not actually have formed unless there was a template of genetic material already there). We saw how difficult it was to imagine the primitive Earth creating an environment that was maintained 'supersaturated' with respect to a DNA-like polymer, that is, that maintained a supply of activated nucleotides. On the other hand, it is very easy to imagine such a provision for clays. As we discussed in the last chapter, rock weathering processes create regions where slightly supersaturated conditions are maintained with respect to clays, which can then form through orderly unit-by-unit crystal growth mechanisms. And we saw how it was that appropriate levels of supersaturation could be achieved where, as often happens, the concentrations of units in the mother solutions were being controlled by the dissolution of other similar minerals, or of the same mineral in a less well organised form. And of course, unlike activated nucleotides, the units in question would have been simple, achiral, and plentiful.

*2. What was the matching-up mechanism?*

A matching-up mechanism is at the heart of template replication – as the means of transferring information into newly forming structures. The details of this process must depend on the way the information is stored. But there has to be something *analogous* to base-pairing. The

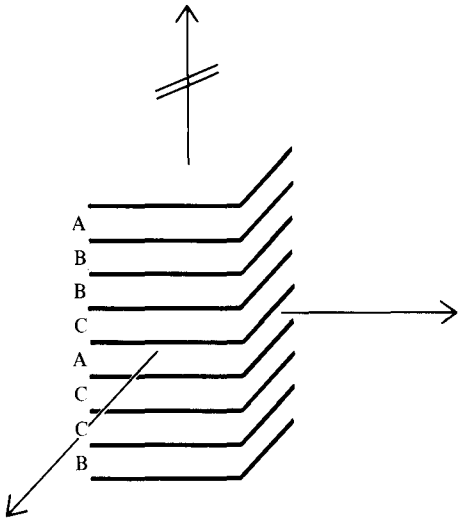



Figure 7.4. A crystal of a disordered polytype could be said to contain information in the form of a particular stacking sequence, that is, a particular sequence of relationships between adjacent layers. Such 'information' would be amplified by exclusive sideways crystal growth (and replicated through subsequent crystal cleavage as in figure 5.11).

genetic information, whether as a pattern of chemical substitutions, or of dislocations, or of crystal domains or whatever, would have to direct the formation of new material 'in conformity with the same design' to use Hinshelwood's phrase.

For a crystal gene the matching-up mechanism would make use of the orderly addition of new units that takes place as a crystal grows. As discussed in Chapter 5, the 'come and go' aspect of this operation is an essential part of it: there has to be a way of putting mistakes right, the bonding between units and crystal has to be reversible as long as it is at the surface.

One can see in all this a particularly tight constraint. The matching-up mechanism must be able to correct certain kinds of defects – mistakes in matching; but it must leave other kinds of defect alone – those defects whose patterning constitutes the genetic information. Indeed these kinds of defects it must copy.

*Polytypism as a key phenomenon?* A matching-up of one-dimensional information in the form of a particular stacking sequence would take place in a disordered polytype that grew exclusively sideways while



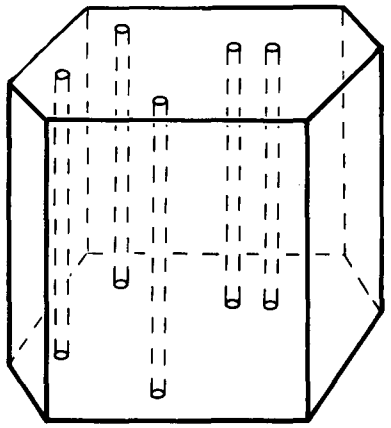


Figure 7.5. The disposition of a set of parallel screw dislocation lines in a crystal might be said to constitute information that would be amplified by crystal growth that was exclusively in the direction of the dislocation lines. (It would be replicated through subsequent crystal cleavage as in figure 5.12.) Such a genetic crystal would be analogous to a stack of punched cards each with the same pattern of holes.

retaining its initial stacking sequence (figure 7.4). If the thin laminar crystals thus produced were to break up suitably this could provide the basis for a genetic crystal of the kind given in figure 5.11.

*The screw dislocation as a key phenomenon?* A screw dislocation is an example of a defect that is often replicated through crystal growth (compare again figure 6.43). This kind of defect is faithfully copied, while other kinds, arising from initial misplacement of adding units, etc., can be put right by local reversals.

A crystal containing many parallel screw dislocation lines might conceivably replicate information in the form of a particular disposition of these lines (cf. figure 5.12). This would be an example of a two-dimensional information store, like a stack of identical punched cards with their lines of superimposed holes analogous to the dislocation lines (figure 7.5).

*Interlayer ordering as a key phenomenon?* Alternatively, coming closer to the formal model shown in figure 5.12, information might be held in the form of a particular arrangement of cations – a specific irregularity within a silicate layer – that was copied as the crystals grew through some layer-on-layer mechanism. In such a way very complicated two-dimensional

information might conceivably be printed off. The resulting crystals would be 'semi-ordered' like the stack of punched cards referred to above: within the layers there would be some element of disorder but between the layers this disorder would be repeated in an orderly way.

In Chapter 6 we discussed several examples of interlayer ordering. There was the possible ordering of domains across $K^+$ planes as suggested by Gatineau for muscovite: there were indications of ordering of aluminium-rich sites between the layers of Ia chromium chlorite from Erzincan, and for vermiculites more generally. The orientation of substitutions was clearly ordered between layers in amesites. The ordered forms of polytypism that we discussed in Chapter 6 are further examples of, often, quite subtle ordering between layers. It would not be surprising, and indeed it seems sometimes to be the case, that a substitutional feature of one layer can induce a similar feature in the layer that grows on top of it.

What kind of substitutional feature might be the most promising? One might think first of a very fine-grained information store where the information units – the letters – were individual cations within layers. One might imagine, say, a 2:1 layer silicate with information written as a particular cation-by-cation arrangement of $Al^{3+}$ and $Si^{4+}$ in tetrahedral positions. And we might suppose that the negative charge pattern resulting from this would locate counterions, such as $Mg^{2+}$, which in turn determined the locations of $Al^{3+}$ and $Si^{4+}$ in the lower tetrahedral sheet of the newly forming layer. Then we might suppose that there was a similar control of the upper tetrahedral substitution pattern through electrostatic or distortion effects operating through the newly forming layer. This whole process would then repeat until there was a stack of layers each containing the identical very complex pattern of cation substitutions.

Such a device could have a marvellous information capacity – more than the same weight of DNA. But there are a number of difficulties. There is a purely geometrical problem in that the intralayer offset between the lower and upper tetrahedral sheets (figures 6.37 and 6.38) creates an ambiguity – any one tetrahedral site is equally close, in the ideal structure, to two sites in the opposite sheet of the same layer. We saw, when discussing polytypism, that it must indeed be possible for information of some sort to be transmitted through 2:1 silicate layers, and the mechanism suggested – via mechanical distortions – might operate here. But in normal polytypism what is transmitted through a layer is not an individual piece of information about some particular atom, it is information about sets of atoms – for example about the direction in which grooves are aligned.

Information held as the complex irregular patterning of individual cations might seem too fragile in any case – too liable to be wiped out by mistakes in copying. And the tendency for intralayer ordering of cations – sideways ordering – would positively tend to wipe out the information: it would be like having a typesetter with an obsessive preference for regular repetitions of letters. (That would be worse than having a typesetter who was just careless sometimes.)



Another difficulty would be that the very irregularity that constituted the information would mitigate against its accurate replication because it would create a rather distorted kind of crystal with, probably, a weak discrimination for incoming units.

A good genetic crystal would have to be rather well organised, with tight discrimination – with far more order in it than disorder. Although some disorder is necessary to provide the capacity for information, capacity would not be the main problem: it would be much more important to begin with that information was transmitted accurately even if its amount was rather small.

We might ask then whether *domains* of a particular cation arrangement might be units of information so that what is copied between layers is not the positions of individual cations but the nature or alignments of more extensive two-dimensional arrays of cations. Such domains pre-existing in a growing crystal might exert a more decisive control so as to be copied more accurately.

*Crystal intergrowth as a key phenomenon?* More generally we might see a genetic crystal as intergrown – with genetic information held as some sort of patterning of domains. Twinning is a particularly pure and prevalent form of crystal intergrowth so let us start with this question. How would the composition planes be disposed in a genetic crystal that held its information in the form of a particular twin structure?

A twin in crystallography is Siamese. It consists of two or more lattices joined in an abnormal but understandable way (see figures 7.4*f*, 7.3 and 7.6). Buerger (1945) has considered the question of how twinned crystals can arise through the processes of crystal growth. During early stages of growth, when crystal nuclei are poorly organised and when solutions are likely to be more highly supersaturated, 'mistakes' in the growing crystals are more easily made and more likely to remain. This is particularly true of twin structures. Here the units that are misplaced are misplaced in relation to their next-but-one neighbours across a composition plane (see figure 7.1*f*); but the nearest neighbour relationships, which are energetically most important, are correct. Twinning is a particularly persistent kind of defect

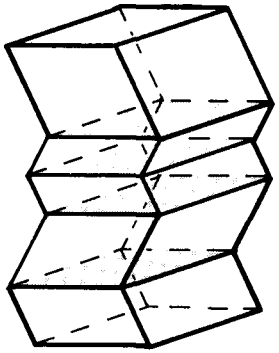


Figure 7.6. Multiple twinning often reveals itself in grooves and ridges parallel to the composition planes (here shaded). Compare figure 7.1*f*.

structure because once the mistake has been made growth beyond the composition plane (for example vertically in figure 7.6) soon becomes quite normal.

But if twinning, and the persistence of twin structure through crystal growth, are to be the basis of a replicable genetic information store, then the twinning must be multiple and microscopic; and the composition planes should not only be retained but extended during crystal growth. Furthermore, the number and mutual alignments of these composition planes must not change during crystal growth.

We have then some geometrical constraints on the morphologies of genetic crystals of this sort. If a patterning of composition planes is to remain constant while being extended, then the composition planes must be parallel to each other in the direction(s) in which they are being extended. There are only two general ways in which this could be so. The first is an example of a type-1 genetic crystal (page 155) in which the information is held in one dimension and amplified through growth in two (cf. figure 5.11). For example, if we look at the twinned crystal in figure 7.6, the disposition of composition planes is a vertical, one-dimensional 'message' that would be amplified through exclusive sideways growth.

In the second possibility, an example of a type-2 genetic crystal (cf. figure 5.12), growth would only be in one dimension. The composition planes would have to be parallel to that and hence, in this direction, to each other; but they need not otherwise be parallel to each other. The crystal that we are now thinking about is somewhat similar to the model in figure 7.5; only here it is planes rather than lines that intersect the layers within



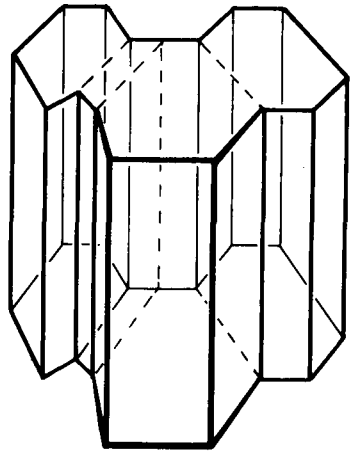


Figure 7.7. In a multiply twinned crystal like this the disposition of a set of vertical planes (the internal composition planes and the sides of the crystal) might constitute information that would be amplified through growth exclusively in the vertical direction (cf. figure 5.12).

which the information is held. In place of the pile of punched cards we might imagine instead a stack of cards on each of which there is an identical, complex, irregular, crazy-paving pattern (figure 7.7). Each multidomained layer would have the same shape (a more or less complex polyhedron, probably) giving an overall tabular or columnar morphology (possibly grooved or fluted as shown in figure 7.7). Since growth must not take place sideways in this case, a constant cross-section should be a characteristic feature.

The Antarctic amesite crystals studied by Hall & Bailey (1979) (see page 213), have the sort of twin structure required except that the twinning here was too regular and on too large a scale to be able to hold much information. The Urals amesite (Anderson & Bailey, 1981) is a better example, with its more complex twinning.

Although for the sake of discussion I have taken the simplest and most ideal form of intergrowth – twinning – the main arguments have been purely geometrical: any kind of intergrowth would do that conformed to the general requirements, and we might be able to identify a potential genetic crystal from its morphology – without knowing what sort of intergrowth was involved.

Kaolinite vermiforms of the kind shown in figure 6.29 (pages 200–1) are of interest in this connection. Some of the crystals, particularly in

figure 6.29*e*, almost look as if they had been *extruded* from an irregularly shaped hole. Presumably this was not so and the constant cross-section arose from the processes of crystal growth. But in the complex environment in which they grew it is surprising that sideways growth should have been so uniform, if indeed these crystals were growing sideways. An alternative interpretation would be that, after some initial phase, growth was exclusively at the ends extending the vermiforms which then, from time to time, broke into shorter lengths (cf. figure 5.12).

Grooves in crystals are often a sign of twinning, or of some other form of crystal intergrowth (figures 7.6 and 7.7), and the grooves that characteristically run along kaolinite vermiforms suggest something like twinning here, with composition planes lying parallel to the main (or exclusive) direction of growth (cf. figure 7.7). (Some such internal defect structure is suggested particularly by the broken crystals in figure 6.29*d*.)

So we might take these electron micrographs as providing visible evidence for the replication of information through processes of crystal growth – information at least in the form of a complex cross-section but probably also as some kind of crazy-paving domaining of the individual layers. These pictures also provide visible evidence for another feature that would be necessary for a genetic crystal: cleavage preferentially in the planes in which the information is held and across the direction of growth (figure 5.12). This is not what you normally expect: normally twinned crystals tend to cleave along composition planes, not across them.

But then kaolinite is not at all a simple sort of crystal. A study by Mansfield & Bailey (1972) of large kaolinite vermiforms emphasises the subtlety of kaolinite crystallography and it may provide the clue as to the nature of this putative domaining in Keller's (much smaller) crystals. Mansfield & Bailey's crystals were divided into domains characterised by different positions, among three possibilities, for the orientation of the vacant octahedral sites (see page 204). The intersection between three such domains is illustrated in figure 7.8. From the sharpness of the diffraction spots, Mansfield & Bailey reckoned that the domains were at least a few nanometres across. This might be compared with the mosaic patterns in replicas of etched kaolinite surfaces that were observed by Williams & Garey (1974). From their published electron micrographs the surface mosaic units here appeared to be mainly in the 100–1000 nm range. (The grooves along vermiforms in figure 6.29*d* and *e* suggest that here domains are commonly about 1000 nm across.)

According to Mansfield & Bailey the 'pseudotwinning' of kaolinite would most likely have arisen in the first place through an early growing



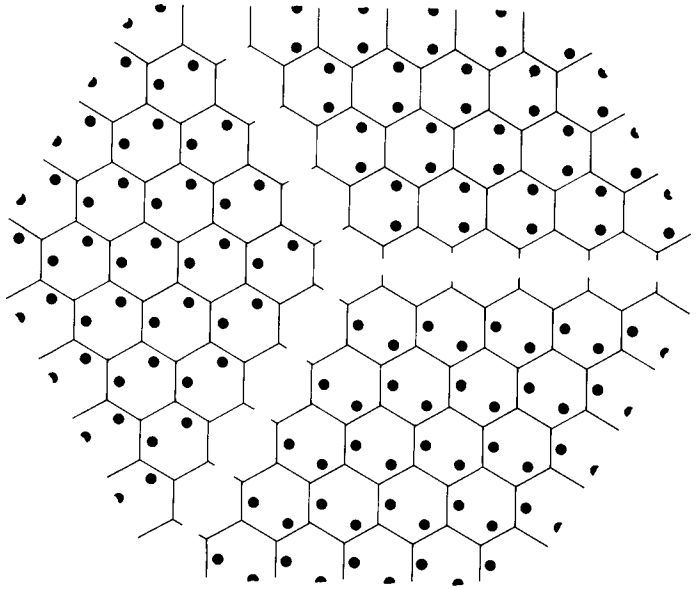


Figure 7.8. Domains in kaolinite (after Mansfield & Bailey, 1972): see text.

together of initially separate crystals that happened to have started with different vacancy orientations. The lattices of such crystals could never match up with each other perfectly because they would be distorted in different ways: the domain boundaries must be rather complicated – more than just planes across which octahedral vacancy positions change (hence the term 'pseudotwinning').

In view of the common tendency for mistakes in the stacking of layers in kaolinite crystals, one might suspect that in these vermiforms, vacancy orientations might not always be accurately copied. (They must *tend* to be, otherwise the crystals would not be kaolinite.)

Mansfield & Bailey found also another kind of stacking fault, although not in all of the crystals. This was a true twinning where successive packets of layers were turned upside down. (Somewhat like figure 7.6, making grooves parallel to the kaolinite layers.) Occasional switching of vacancy orientations (or layer inversions even) might not matter because the information is seen to be primarily in a pattern of domain boundaries. What is important is that these boundaries do not wander as the crystal grows. But as indicated already we have some direct evidence for fidelity here from the morphology of some of those microscopic vermiforms – from

their 'extruded' appearance – whether the more detailed structural considerations are correct or not.

*Crystals as replicators* We can see, then, a variety of possible ways in which defect structures of various sorts might be copied through crystal growth processes. Each of these ideas depends on the notion that crystal growth mechanisms, although reversible, are only locally reversible under conditions of supersaturation. The mechanisms have no foresight, they cannot see the most perfect way of placing all the units even if they can, in a sense, 'feel' the best way of placing one or a small number of units. They can follow some easiest path. Often the easiest way of proceeding, even the most orderly way of proceeding in the circumstances, may be not to eliminate a defect but to *repeat* it: to extend a screw dislocation or a disordered layer stacking sequence; or to repeat a pattern of domains. That way there may be less local disturbance. Suppose, for example, that a preexisting layer consists of domains each with one of three possible orientations for rows of units (as in figure 7.1*h*). What would be the most orderly way of making another layer? With all the rows, this time in the same direction? It would take foresight to see that that was the way to perfection because locally there would be a severe mismatch: two-thirds of the new rows would now be misaligned with respect to the rows underneath. Even to shift a domain boundary a little will involve some higher energy intermediate, especially if the interlayer forces are strong and depend on the matching of orientations for that strength.

Mansfield & Bailey suggested that in kaolinite the 1:1 layer type, and the adoption of the same vacant site in each layer, produces an asymmetric and rather strained structure – and that that is why kaolinite crystals usually remain very small. The pseudotwinning, and the inversions of packets of layers, they see as ways of 'alleviating the strain by redirecting the distortions on a domain scale'. We saw something broadly similar to this in antigorite with its curious periodic lateral inversions of 1:1 serpentine layers as a means, here, of alleviating the considerable strain inherent in the mismatching of tetrahedral and octahedral sheets in that structure.

It is an interesting thought that in layer silicate minerals the perfect structure may very often not be the structure of lowest energy: structures with various forms of domaining may be more stable with the energies of these various forms very similar. After all it would be sheer coincidence if there was no mismatch at all between the elements of an asymmetric layer, such as the 1:1 silicate layer. The chances are that on its own, *in vacuo*, it



would tend to curl one way or the other. To that extent there must be some energy penalty, however small, in having the layers flat in the crystals. As a crude model you might imagine an old painting which has developed a tendency to curl because the paint film has contracted. If the painting is kept on a stretcher the paint film will craze.

We can take this analogy one step further. I daresay that for a painting there would be one craze pattern that was of minimum energy – some regular array of hexagons I suppose. But who ever saw that on a painting, or exactly the same such craze pattern twice? Similarly, a silicate layer could make a very good information store, following our earlier discussion (page 272), if some sort of 'crazing' is better than none at all, but if at the same time no particular craze structure – no special regular superlattice – is much more stable than countless others.

### 3. *What was the locking mechanism?*

For DNA replication the locking mechanism is quite distinct – it is in the formation of phosphodiester bonds. The security of this lock depends on the relatively high activation energy barriers that protect such bonds from hydrolysis. This bonding is not normally reversible.

As we saw in Chapter 6, linear silicate and metal-oxy polymers in aqueous solutions are not able on their own to maintain a complex metastable pattern – their bonding is too readily reversible. For crystal genes, however, the security of the information would not depend on the integrity of bonds exposed to water in the same way: it would depend rather on there being many copies of the information with most of the copies inside the crystal. In the growth of silicates or metal oxide crystals from aqueous solution there is local surface reversibility. That is important for the matching-up mechanism. But because water participates in those bond-making and bond-breaking processes which are necessarily involved in the growth of crystals of this sort, this reversibility will generally be restricted to the surfaces. Eventually the information in, say, a layer of a growing layer silicate becomes locked in through the growth of further layers on top.

For crystal genes, then, the matching and locking mechanisms would be less distinct than for DNA with its two kinds of bonding for these two functions – reversible hydrogen bonding (mainly) for the matching, and irreversible covalent bonds for locking. For crystal genes of the kind that we have been thinking about the strategy of holding and replicating information may be similar, but the tactics are here not the same at all. In

particular the matching and locking mechanisms are not so distinct from each other, and the bond types used mainly are like neither of the bond types used mainly in DNA.

### 4. How did master and copy separate?

One rather unexpected difficulty in the idea of an organic polymer as a primary genetic material was with the question of how, having synthesised one chain on another, the chains could be made to come apart (Chapter 5). The trouble was that, if the matching forces were strong enough to hold the units reliably in place, then they would together be too strong to let the chains separate. There were various ways in which this problem might be overcome, but they all seemed to need too much contrivance to provide the basis for a primary genetic material. The corresponding final stage of the replication of information in growing crystals looks rather easier – it would be through the break up of the crystals in such a way as to expose the information on new surfaces.

Crystals very often break up as they grow. A flask containing a supersaturated solution may remain for days with nothing happening – and then quite suddenly be full of thousands and thousands of crystals. The initial delay showed that the rate of spontaneous nucleation was far too slow for thousands of independent nucleations. It is more likely that what happens in such circumstances is that one crystal nucleates spontaneously and then grows quickly and breaks off pieces to make new nuclei which in turn grow and break up. This kind of thing is called 'breeding' (with no implication of course that any very complex information is necessarily being propagated in the process). But we can at least say this about a working crystal genetic material: it must 'breed' in this unsophisticated sense as a necessary but insufficient condition. More precisely, it must have a strong preference for cleaving across the direction or directions of growth and in the planes or plane within which the information is held (*i.e.* as in figures 5.11 and 5.12).

These last considerations make a layered structure attractive since there is here a strong preference for cleavage in the plane of the layers and, as discussed earlier, there are ways in which information could be held within the plane of the layers. Nor is it asking too much that crystal growth must not take place in this same plane – growth must be exclusively layer-on-layer. Highly elongated habits are common enough in crystals, indicating that there can often be a strong preference for one direction of growth. The



kaolinite vermiforms, indeed, can most easily be explained as having arisen from a crystallisation process that was mainly in one direction – in the correct direction for our purposes. And of course these long fragile vermiforms often break up as they should – in the plane of the layers, across the preferred direction of growth.

It seems too good to be true that kaolinite, the commonest clay on Earth perhaps, might also be a primary genetic material. But it combines nevertheless many of the characteristics that you would want. Maybe the first genetic materials were something like kaolinite.

### Primitive genetic control

We come now to the third basic characteristic of genetic information: it must not only be held and replicated, it must have a meaning. We might say that the meaning of genetic information depends on its exerting some sort of control that tends to increase its own prevalence. To be genetic information it must tend more or less indirectly to be self-propagating.

The indirectness is important – one of the characteristics of highly evolved organisms. Physical scientists naturally tend to seek the explanation for some persistent or recurring patterning of atoms in terms of the relative stability of that pattern, or at least of the more or less immediate processes involved in its formation. But evolved genetic patterns are not to be explained away like that. If you found, let us say, that part of a mammal's genome read ATTGCGTAGCGTAAGTCG, you should not expect to find any direct physicochemical explanation for it. Your efforts to show that this sequence was especially thermodynamically stable or easily formed would be unsuccessful. More likely the sequence specifies the positions of a few amino acid side chains in a protein molecule – helping to make an enzyme, perhaps, that helps the animal to see better, that makes it aware more quickly when there is a cat in the vicinity. That would be a very indirect reason for ATTGCGTAGCGTAAGTCG.

We might say, then, that while life depends on replicating structures, it depends too on there being circumstances in which particular complicated structures are preferred. This in turn depends on there being possibilities for indirect and yet successful control of the environment by replicating patterns. Perhaps, indeed, evolution can be described most simply as being the appearance of increasingly indirect ways in which replicating patterns can show their mettle.

We will take it, then, that the first genetic materials evolved along such

lines – that, although the patterns they contained exerted a rather direct kind of control to begin with, a major early evolutionary trend was towards more indirect modes.

*Direct effects of millions and millions of gene copies*

To begin with, the replicating defect structures in a crystal genetic material might have operated simply through the effects of such defects on the bulk properties of that material – via effects on crystallite shapes and sizes – and on the way the crystallites interacted with each other.

Van Olphen (1977) discusses in detail various factors that can influence the bulk properties of clay suspensions. Quite small differences in the composition of the solutions within which clay particles are suspended can make the difference between, say, a drilling mud being very runny or far too stiff; or it might make the difference between a badly drained and a well drained soil, or between a place where you might build a house or be very ill advised to. Such differences are to be explained in colloid chemical terms, in terms of effects operating at the microscopic level. If adding a relatively small amount of lime or gypsum to a soil greatly improves the drainage, this is probably because negatively charged clay crystallites have had their charges reduced or neutralised so that the particles now adhere to each other, via the calcium ions to form separate flocs. But, depending on the shapes of the crystallites present, on their sizes, on their electric charges – and often too on details of the distributions of these charges – a given flocculating agent, like $Ca^{2+}$, may operate more or less efficiently or it may even operate in reverse. Perhaps the particles are positively charged anyway, but they have special sites on them with a strong selective affinity for $Ca^{2+}$. Then they would become more highly charged and more likely to deflocculate when the calcium concentrations of the surrounding solutions increased.

For clays, anions can act as deflocculators by cancelling positive charges on the edges of clay platelets – where there are exposed intralayer cations. Polyanions are particularly good at this: indeed polymetaphosphate can even reverse the edge charge (see figure 7.9). This is thought to be effective in destroying a kind of 'card-house' association between the clay particles that can produce apparently very solid gels (see figure 7.10; compare figure 6.54*f* on page 247).

Not only can small changes in the environment of a given clay suspension make a big difference to porosity, ion exchange properties, etc., but between different clay samples small differences in structure may also be



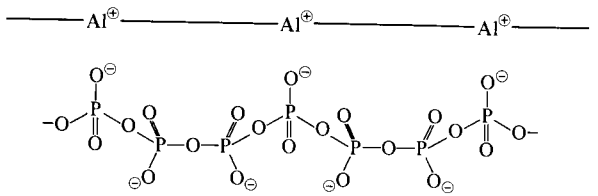


Figure 7.9. Partly exposed cations at clay edges can lead to positive charges there. Polyphosphate can reverse such charges, possibly, as suggested by Van Olphen (1977), through co-ordination of some of the anionic groups of the polyphosphate leaving others to build up a net negative edge charge.

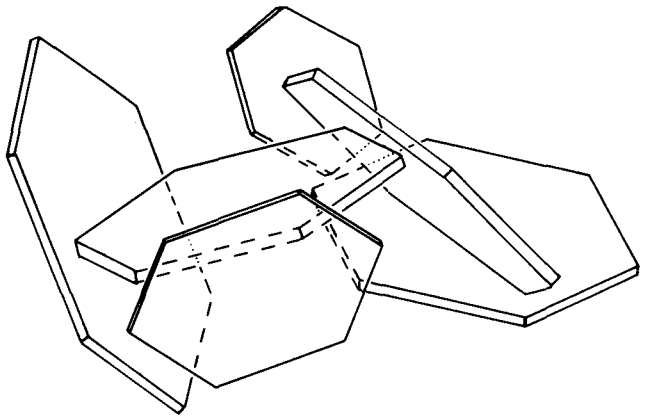


Figure 7.10. Interactions between positively charged edges and negatively charged faces of clay particles can give rise to 'card-house' associations.

important. In particular the shapes and sizes of the crystallites and their defect structures are likely to be important – the latter because they affect the distribution of permanent charges in the crystallites and the locations of special adsorption sites. On such features may depend, for example, the stability of a given 'card-house' to changes in the environmental solutions, or whether an impervious gel-like association is preferred or a very porous lumpy flocculation.

Some genetic control of some quite pedestrian bulk properties would be enough to allow us to imagine the beginnings of a Darwinian evolution. Figure 7.11 illustrates the kind of situation that we touched on in Chapter 3 when discussing how 'naked genes' might work. Here we suppose that

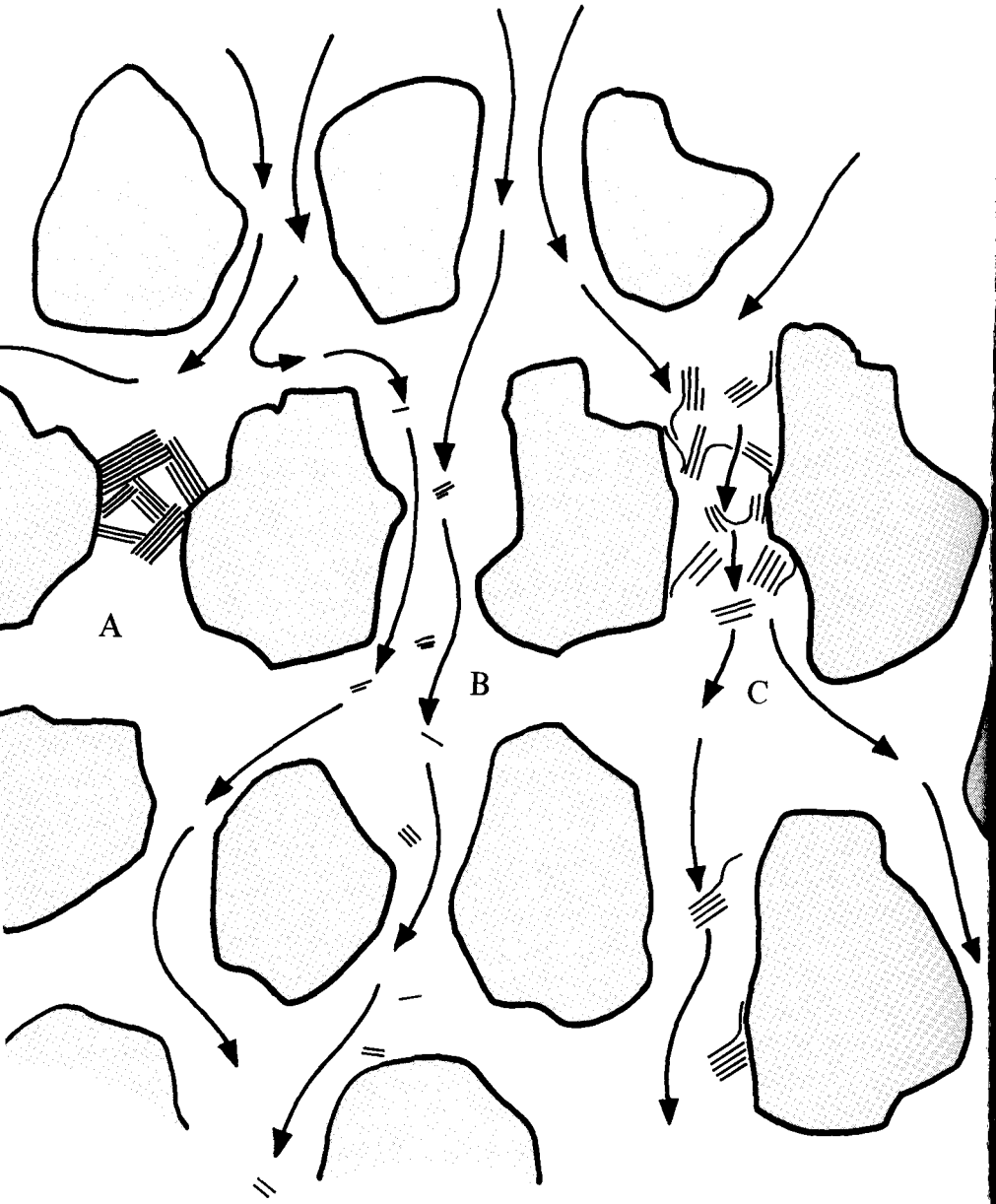



Figure 7.11. We imagine here replicating clay crystallites forming in a sandstone through which nutrient solutions are moving. In zone A the clay being formed is too impervious – it deflects the nutrient flow. In zone B the crystallites do not adhere sufficiently and they are being washed out of the sandstone. The clay type forming in zone C is the most successful of the three: it is porous and yet not easily dislodged wholesale. Sometimes, however, a piece breaks off to seed a similar colony downstream.



clays are growing in a sandstone (cf. figure 6.54) and we suppose that one clay is printing off copies of a defect pattern that makes crystals that associate with each other strongly to form a plug. This stops the local flow of nutrient solutions, and the clay crystallites stop growing there. The pattern of defects that makes that kind of crystal is a failure. On the other hand, it would not be a good idea either to be too loose. A defect structure that produced crystallites that did not stick at all to each other or to the walls of the pores would also be selected against because the crystallites containing such patterns would be washed away. The best compromise might be to stick to the walls but let the nutrient solutions flow through by having a rather open porous texture.

Because it is a consequence of interactions between millions and millions of gene copies, you might say that the consistency of such a clay would be a phenotype. Even this seemingly commonplace property, provided it could be influenced by (usually) reliably replicating patterns, could provide the basis for an evolution through natural selection. Inevitably the clays in a location for which consistency was important would evolve an appropriate one. No wildly improbable event would be needed to start such a process off because nothing very clever is required. Of all the crystallite shapes, etc. that might be being replicated perhaps half would be adequate as starters, because they stick (at least a little) and do not stop the flow (at least not completely). Of that large fraction of all the possibilities that were adequate, there would be a subfraction that were a bit better, and a sub-subfraction that were better still. In line with the discussion of optimisation procedures given in Chapter 3, evolution would be expected to proceed from patterns belonging to larger subsets into those belonging to smaller subsets (cf. figure 3.1), that is, from structures that could have been almost anything to those that were rather special. How far this would go, how small the subset that could eventually be arrived at, would depend partly on how intricate the control of consistency through defect structure could be. It would depend also on whether a very particular consistency made a perceptible difference to survival chances or replicative efficiency in the available environments. That is to say the evolutionary potential would depend on the characteristics of a relevant functional landscape: to get started there must be a large catchment area provided by some easily achieved function, but to go very far there must be fairly deep valleys somewhere.

*How odd might the properties of genetic crystals become?* If we had no experience of highly organised protein molecules made from perfectly

specified amino acid sequences, if the only long polypeptides that we had ever come across had been copolymers of the sort that arise from normal chemical polymerisation procedures, we would have no conception of the subtle pieces of machinery that a well made protein can generate. For one thing we would not even have come across a sample of material in which all the polypeptide chains were identical. We would have had experience only of mixtures and hence of visible properties that were the statistical outcome of mixed assemblages. (We would be lucky, indeed, to have come across a sample of long polypeptide chains in which any one chain was the same as another.) Blends are likely to be boring, average things concealing special properties of rare, or even not-so-rare, individuals.

Straight away, then, you might expect the visible properties of a material composed of replicating crystallites to be quirky: at least now there is the possibility that characteristics arising from individual defect structures, rather than from mixtures of certain general classes of such structures, could show up. But of course the possibility of evolution through natural selection inherent in (nearly perfectly) replicating structures leads to a still more interesting thought: that exceedingly rare individual structures might be discovered in time, with properties that would have been very odd indeed for the normal average sort of material. Indeed in proceeding from set to subset to sub-subset, in exploring an evolutionary landscape, natural selection can quite quickly arrive at subsets that are so small that you can discount the possibility that that subset could have been hit on in any other way (cf. pages 87–8).

Let us not suppose, then, that microcrystalline minerals, such as clays, could never have very interesting properties because the stuff at the end of the garden spade does not look very interesting. (And let us not be too put off by the unfortunate term 'defect' – a 'defective' silicon crystal can be a microprocessor – it is a question of *what* 'defects' and where they are.)

The ultimate limit to a crystal's performance as a genetic material would be set by its maximum genetic information capacity. In an idealised case this would be given by:

$$I_{max} = \log_2 w \text{ (bits)}$$

where $w$ is the number of *a priori* equally likely ways in which the replicable features could be disposed. As an example, suppose that there was complete freedom about the disposition of the screw dislocation lines in the hypothetical genetic crystal shown in figure 7.5, and suppose that there were 100 such lines in a crystal 1 $\mu$m across. If such a crystal was a layer silicate it would have about $10^6$ unit cells in a layer, with about that number



of positions for an intersecting dislocation line. There would then be about $10^6 \times 10^6$ ways of putting two dislocation lines – and about $10^{600}$ ways of putting 100 such lines, giving a maximum information capacity of about 2000 bits. By comparison, an average sort of globular protein with 200 amino acids is one of $20^{200}$ possibilities, which means that the information capacity here is about 860 bits.

There are dubious assumptions in calculations of this sort – the independences of the screw dislocations and the accuracy of replication of their disposition are the most serious here – but we can see all the same that quite a moderately complex defect structure could give rise to microcrystals – objects still of colloidal dimensions – that could in principle be as ingenious as protein molecules (that is to say selected from a similarly large set of possibilities). If the information was in the form of a pattern of 100 replicating domains the capacity could be very much greater than for a screw dislocation array (cf. figure 7.7). (Kaolinite vermiforms may combine each of these ideas – if each of the columnar domains implied by the electron micrographs contains one or more screw dislocation lines in the direction of growth.)

*How to 'focus' information* Although the information content of such genetic crystallites might be similar to that of a protein molecule, the information density would be very much less. Does this limit the effects of such information to effects on consistency and other such bulk properties? In particular, does it rule out anything really clever – like the intricate control of organic reactions? Would such information be altogether too coarse-fingered ever to be able to manipulate individual molecules – ever to arrive at that point that we know early evolution did arrive at where a competence in protein and nucleic acid chemistry had been achieved?

There are four things to be said here. *First*, there is nothing in the definition of life to say how large or small genetic information should be written. These are practical, not theoretical questions. *Second*, coarse fingers, human fingers, can in effect manipulate molecules: if 100 ml flasks and such things can control organic reactions when suitably used, I do not see why contrived (evolved) micrometre-scale apparatus might not be effective too. There is nothing in the definition of life either to insist that a protein-like technique of fine molecular discrimination should have been the basis of the first forms of biomolecular control. (I will come back to this topic in the next chapter.) *Third*, modern genetic information is not in any case all that fine-grained: it takes about 100–1000 nm of DNA to specify a protein molecule. There are two stages in the 'focussing' of this

rather diffuse information so that some precise molecular activity can be specified. The information is translated into a somewhat more compact kind of molecule with a higher information density – a protein – and then this molecule folds up. The information in a protein is written still rather coarsely – as a sequence of choices of *ca* nanometre-sized objects; but, because of the folding, a globular protein can discriminate at a much finer level. As we discussed in Chapter 3, the positioning of the few side chains at the active centre is the consequence of many, if not most of the other side chains in the molecule, with the choice of remoter side chains allowing 'fine adjustment' of the positioning at the active centre. The effect of the choice of perhaps 100 amino acids is thus concentrated at the active centre. Each choice may seem rather coarse but 100 one-in-twenty choices represents enough information to allow a very detailed specification in principle.

This kind of focussing of information originally spread out in a linear sequence depends absolutely on the beautifully articulated quasi-crystalline structure of the folded protein molecule. Real crystals too can transmit effects over a distance through concerted pushings and pullings, as we discussed, in particular, for layer silicates (pages 223–5). And information of a sort can thus be focussed in real crystals in the sense that characteristics of a given site may be determined by the cumulative effects of atoms distributed within a quite extensive adjacent region. So the *fourth* point to make is that relatively diffuse information in a crystal could very well exert effects that were concentrated or expressed at particular sites. With reference to the suggestion made by Mansfield & Bailey (1972) that a large kaolinite domain becomes unstable because of the accumulation of distortions created by the unsymmetrical unit cell of kaolinite, we might speculate that the detailed character of a given domain boundary, the extent to which it represented a misfit between differently distorted lattices, might depend on the sizes and shapes of the adjacent domains. (The bigger they were, the more out of step they would be with each other.) In such a case a given crazy paving pattern of domain boundaries would be chemically a very interesting structure, it would be a particular array of high-energy sites with particular adsorptive and catalytic properties. The edges of layers, and the grooves created by stacks of layers, might also represent replicable active features.

### *Threats and opportunities*

To imagine how more particular and intricate genetic patterns might catch on we should try to imagine more complicated situations than



that illustrated in figure 7.11. Suppose for example that the flow rate was variable – and in the rainy season it increased while at the same time the ion concentrations dropped. The successful clays would be those whose crystallites would cohere in solutions of low ionic strength. Perhaps indeed the most successful would adhere to each other more strongly under those circumstances, so that they formed an impervious plug to stop the flow locally, to reduce the rate at which they were being redissolved by the now undersaturated solutions. When the ion concentrations increased again, in the dry season, the crystallites would 'know' to separate because conditions were no longer dangerous – it would be time to be printing off some more crystals.

Although abnormal, perhaps, such behaviour for a clay would not be absurd. But anyway we are not talking about normal clays. We are talking about clays of a sort that we have no direct experience of (well, probably not, but I will return to this at the end of the book). Suppose that we really could tailor-make a clay particle; suppose we could place substitutions precisely, we could arrange lattice tensions so that they were just so; we could contrive shapes and sizes exactly; suppose that we had the techniques to fabricate clay unit layers in the kind of way that we can fabricate microprocessors – only at a still much finer level. Then we would have little trouble, I think, in engineering much more ingeniously adapted clays than the one imagined above. You could make materials with properties unlike those of any natural clay, with specially engineered crevices, perhaps, that had an affinity for very particular ions or molecules. Or you might make a clay that self-assembled into a mansion of a card-house with rooms of just certain sizes interconnected with each other in certain definite ways. Can you doubt that if we could fabricate silicate layers with atomic precision we could produce materials that to the layman would have magical properties? Such a material might seem alive. It would not be alive because it would be a human artefact, its multiple ingenuities built in by hand. But suppose instead natural selection had been the engineer. Then it would be alive.

What for a normal physicochemical system would be an extraordinary coincidence becomes almost an expectation for a structural feature that is subject to evolution through natural selection. Among such features are those that help to propagate the genetic patterns more widely. (Because those genetic patterns that had in the past been adept at publishing themselves are bound to be the most prevalent now.) A simple way would be for a clay to be of such a consistency that flocs occasionally broke off and seeded further growth downstream (figure 7.11*c*). But in a finite open system

there is a limit to the amount of spreading that is possible downstream. So perhaps clays might be selected that on drying made a friable powder that was easily spread by the wind to seed further growth of similarly adapted material elsewhere. Or perhaps, like the salmon, the clays would swim upstream to lay their eggs. Ridiculous? Not altogether: we might imagine clays forming in groundwaters that occasionally reversed the direction of flow (because of an overflowing river or a very high tide or something). Then the clever thing would be to break into flocs when the flow was 'upstream'. For a clay that had been engineered by natural selection such a feature would be quite understandable because it would be based on a suitable conditional property that would not in itself be un-claylike. For example when the concentration of $Na^+$ reached a certain level (a sign of back-flow) loose flocs would be formed.

Nor is there anything very extraordinary in a material having 'conditional properties' – all materials have them (for example, *if* I heat this piece of plastic *then* it will soften). What would seem extraordinary about an evolved genetic clay would be the detailed appropriateness of the conditional behaviour. It would be a teleological way of speaking to say that such a material had learned to cope with particular habits of the environment; as it is a teleological way of speaking to say that the crocuses know when the spring has arrived. But such teleological talk becomes justifiable (at least as kind of shorthand for longer-winded mechanistic explanations) when talking about products of natural selection – and sooner or later it would become justifiable when talking about the behaviour of replicating mineral crystallites. If the total behaviour of the crystallites was seemingly highly improbable; if it depended on complex individual features of the crystallites being just so; if these features were replicable or produced under the control of replicable structures – and anyway if their configuration that had generated the strange behaviour was indeed a product of natural selection – why then it would be plain silly not to call such minerals alive.

### Early phenotypes

The main oversimplification in the foregoing discussion of modes of genetic control has been in thinking about pure genetic crystals. Pure mineral deposits are not at all common and nor is it necessarily so that primitive genetic crystals could catch on most easily if theirs was the only kind that was crystallising out in a given locality. Let us consider what might happen if both genetic and non-genetic crystals were being formed in the same place.



To begin with you could think of the non-genetic crystals as simply part of the environment of the genetic crystals – something the latter had to put up with, perhaps. But sometimes, by chance, the non-genetic crystals might be an advantage. For example, they might tend to offset the main snag about a genetic crystal – that to work most effectively it must be in a solution of just the right level of supersaturation (cf. page 159). We might imagine non-genetic crystals acting as a buffer for the units out of which the genetic crystals form: they might be some relatively quickly crystallising material – such as imogolite or allophane. These would hold concentrations down during periods of plenty preventing a too rapid growth of the genetic crystals (that would lead to too many mistakes in replication). On the other hand, when the solutions flowing into the regions became undersaturated the non-genetic crystals would dissolve to maintain supplies. Or perhaps non-genetic crystals could act as other kinds of stabilisers, as pH buffers or flow controllers.

The point at which such non-genetic minerals might be said to be part of the phenotype of the primary genetic crystals would be when the genetic crystals evolved characteristics that encouraged favourable non-genetic minerals to form. It is only one step away from helping yourself to help someone else who helps you. But how might the genetic crystals affect the formation of some other, perhaps quite different kind of material?

There are several ways. The genetic crystals will have effects on flow rates, for example, and this is bound to alter the compositions of the surrounding solutions and influence the sets of minerals crystallising from them. Not only the types, but the morphology of a given mineral type can be strongly influenced by conditions of synthesis. Newly formed halloysite appears to be particularly sensitive here, showing a great variety of micromorphologies that appear to result from differences in such things as local permeability, pH of waters and grain size of primary minerals (Tazaki, 1979b).

Control through epitaxy is another possibility. One crystal can nucleate another's growth and it is typical that such nucleation is at defect sites (see for example Parham (1969) on the growth of halloysite on feldspar). So there might be a marvellous scope here for a critical control by highly contrived defect structures at the surfaces of genetic crystals. Special defect sites on evolved genetic crystals might be able to nucleate the growth of mineral crystals that would not normally have formed – minerals for which nucleation tends to be the difficult step.

Even if only the morphologies of the newly forming minerals was being controlled by the genetic crystals this might still have a profound effect on

the local physicochemical environment of the genetic crystals. Look again at the varied morphologies of zeolites (figure 6.26 on page 194) or at the various forms of ferric oxide-hydroxides (figure 6.1 on page 169) or at the variety of forms and types of clay minerals in sandstones (figure 6.54 on page 246). If genetic crystals could control the kinds and forms of such minerals, even if they could only exert a general kind of control by helping one kind of growth process or hindering another, they would be controlling indirectly such things as local porosity, local ion-exchange characteristics, local catalytic effects and so on. They would have a more indirect and complicated means of affecting factors that affected their own survival and propagation.

### *Advantages of specialist phenotype structures*

Given reliably replicating mineral crystallites we can, then, begin to see simple mechanisms for indirect self-help. But we can see also a driving force – a reason for a long-term trend in that direction. There are positive advantages for genes to operate through other structures. The ways in which a genetic material can directly control its environment are bound to be limited by design constraints imposed by the overriding need for such a material to be able to replicate information accurately. Crystals that were good at pattern printing would probably have to be rather rigid with quite strong inter-unit forces so that crystal growth proceeded in a very orderly manner. On the other hand, although some control structures could be made from rigid materials, others might be more suitably made from flexible ones. The control of more intricate and heterogeneous micro-environments might become possible if, say, tubes and vesicles were present as well as plates and grooved rods. It is, I suppose, more difficult to template-replicate a tube than a plate. A tube might be reproduced without being template-replicated so long as the characteristics of the tube could be specified by information in, say, a plate – and so long as that information was replicable. Modern organisms use this device of dividing the roles of information replicator on the one hand and main control structure on the other between quite different kinds of molecules – between DNA and the much more versatile and flexible protein.

Embarking yet again on the hazards of more particular speculation we might imagine a primary organism with an analogous device. We might imagine that from time to time, when groundwaters were fairly concentrated, the layer-on-layer growth of domained genetic kaolinite crystals was interrupted by a few layers of halloysite instead, these layers becoming



unstable eventually and rolling up to form independent tubes whose exact swiss-roll structure was affected to some extent by domain patterns which they had inherited. The maze of tubes thus made then helped to create an appropriate consistency perhaps, or was a pH buffer, or bound certain heavy metal ions that would otherwise interfere with kaolinite synthesis.

### *The general technique of gene action*

To improve the general analogy between the way in which primary genes might have been effective and the way in which modern genes work it is worth emphasising what genes of any sort do and what they do not do.

In the first place genes can only operate passively because they are not themselves sources of energy – they can only redirect flows or potential flows in the environment. And genes do not specify a phenotype through specifying a brick-by-brick construction in the kind of way that you might physically build a house. The construction procedure is more like the way an architect builds a house – relying on the known tendency for others to work in certain ways. As remarked in Chapter 3, most of the explanation for the arrangement of the atoms in an organism is to be given in physico-chemical terms. Genes only cue the outcome. That is all they could possibly do since they contain far too little information for them to be able to specify the positions of every atom in the phenotype. Even in the making of a protein molecule the final functioning object is not intricately made by the genetic information: most of the explanation for why a protein folds the way it does is to be found in books on physical chemistry. There is only one factor that you will not find there – why the amino acid sequence is as it is. (If the primary structure of a protein really was the sole determinant of its tertiary structure you would in principle be able to make any adjustment to the position of any atom in the folded molecule by suitably altering the sequence. This is clearly not the case, only some adjustments to the tertiary structure are physicochemically possible.)

Although in the details of how they would work and what they were made of primary mineral genes would have been quite different from DNA genes, they would have been similar in that they too would have worked by 'cueing' the environment, by triggering events. The main difference would have been that the kinds of events triggered would have been closer to the ground in all senses – nearer to what tends to happen anyway.

*Vital mud*

As for the size of primary organisms, they were not the neatly packaged boxes of genes that micro-organisms are now – with roughly one gene of each kind needed to run and remake the localised 'survival machine' that contains them. The microscopic size of the modern bacterium should be seen as one of its most notable achievements rather than any sign of evolutionary infancy. The bacterium is a highly concentrated package of ingenuity. Before that sort of thing became possible there would have been clumsier, bulkier ways towards our biochemistry (as I will discuss in the next chapter).

The first primitive organisms that we have been trying to conjure up were not really microscopic objects at all. The control of the environment by primitive genes depended, not on the individual acts of individual genes, but on effects depending on millions and millions of copies of them. There was thus no need for anything as neat as a cell. If you are a gene very close to the ground, if your modes of preferential survival and propagation depend on deflecting somewhat, and to your advantage, processes that are going on in any case around you, there is no need to be so cordoned off. Indeed it is better not to be. It is only much later, when you are a gene made of an unnatural sort of stuff having no direct connection with the usual products of the Earth, that you must have your life support systems immediately around you. Then you *have* to be in the kind of cell that has now become the identifiable unit of life – but only then.

For a picture of first life do not think about cells, think instead about a kind of mud, an assemblage of clays (including perhaps, zeolites, iron sulphides, oxides, etc.) actively crystallising from solutions. It might not have seemed to be very prepossessing stuff, this vital mud, if you had gone back to the primitive Earth to look; but it would have had some strange properties if you had taken the trouble to examine it carefully. It is not just that it would have tended to alter its properties (consistency, ion-exchange characteristics, etc.) according to circumstances; this substance would have seemed to know about the world, because its behaviour would have been appropriate in quite complicated ways to its survival – and to its spreading.

*A sequence of places to live*

Where on the primitive Earth would you have looked for these earliest forms of life? You would have looked in open systems – you would have looked for conditions for continuous culture where food



supplies were assured and where conditions for growth remained reasonably constant over the long term. For these mineral organisms that would have meant somewhere between weathering and sedimentation. You might have looked in the porous weathering crust of a piece of granite, or in a volcanic tuff, or in precipitates from groundwaters far from the origin of the ions in those waters. Or you might have looked in streams or rivers or lakes, or in the sea and its sediments.

It is well understood that the descendants of an organism that evolved in one place may now live somewhere else. We do not live in the trees any more, and yet the design of our hand and eye owes much to an arboreal way of life. Going further back, our circulatory system was basically invented by animals that lived in the sea, while cell components such as mitochondria were first put together by micro-organisms with some now unknown way of life. Scene changes have been an important part of evolution – it is almost as if the machine had to be sent to different workshops to have different components at different levels of organisation perfected.

I think that there would have been such scene changes at the start of evolution because, whether you believe a mineral story or not, there were very special problems for organisms at the beginning. Life would have started, I think, in a special workshop that happened to be set up to produce a first working genetic material. An unevolved first mineral life, for example, would have been restricted to particular regions, where conditions happened to be precisely right for orderly replicative crystal growth – a particular region of a particular sandstone perhaps. But then, with the evolution of indirect modes of control, with the cued collusion of other materials, the genetic information could have been preserved and propagated in a wider range of places.

So the first workshop would have been somewhere stable, deep down probably, below ground or near the bottom of the sea. To begin with, security would have been more important than opportunity. But adjacent to the easy places where evolution could start there would have been more difficult places to provide a clear field for some of those with more elaborate ways of surviving. As part of this adventure we might suppose that some kinds of mineral organisms found places to live at the surface, in rivers or lakes perhaps; that they found places in the sunlight – and another kind of chemistry.

# 8

# The entry of carbon

Figure 8.1 gives an outline of the plot of our story – about how the last common ancestor of life on Earth came into being. In the last chapter we were concerned with the first three stages shown in this figure, arriving at the idea that the first organisms were, quite simply, made of clays; and that *life* would soon have become a wholly appropriate form of description for these systems, since they would only have been plausibly explicable in terms of a Darwinian evolutionary history.

Everything about this story rests on the supposition that there was some class of colloidal inorganic crystalline minerals forming on the primitive Earth that could replicate more than trivial amounts of information accurately. From our discussions of conditions on the primitive Earth; of how microcrystalline minerals form today; of what is needed for accurate replication – and from considerations of bonding characteristics in organic and inorganic materials – it seems altogether more plausible that growing mineral crystallites rather than organic polymers should have been the primitive genes. We cannot yet put our finger on one mineral and say that it combines all the required features, but we can say that clays show between them all the features required. We can say, even, that one kind of clay – kaolinite – is a conceivable candidate. In any case this question is open to further clarification through observation and experiment; we will come back to this in the coda.

The details of the life-styles of the earliest organisms are much less accessible to experiments and observations. But they are less important. In the latter part of Chapter 7 we saw a number of ways in which that engineer natural selection might have taken hold if reliably replicating structures were there. A more important point was the more general one – that a



| | | |
|---|---|---|
| **STAGE 1:** **unevolved crystal genes** | **STAGE 2:** **evolved crystal genes** | **STAGE 3:** **complex 'vital muds'** |
| Clay minerals on the primitive Earth include some with complex replicable defect structures | Some patterns of replicating defects become more common because they confer on assemblages of clay crystallites properties (including large-scale properties) that favour clay synthesis | Replicating defect patterns evolve more indirect modes of control: complex assemblages are formed in which genetic and nongenetic clays promote each other's synthesis and propagation |
| **STAGE 4:** **photosynthesis** | **STAGE 5:** **metabolic pathways** | **STAGE 6:** **structural organic polymers** |
| In a number of lines, evolved clays provide photochemical machinery for making a few metastable molecules, especially polyphosphates, very simple C, H, O compounds and ammonia | Elaborately specified clays make the apparatus for multi-step organic syntheses | The availability now through biosynthesis of particular chirally specified monomers provides the basis for the manufacture of precisely made organic polymers for various structural purposes |
| **STAGE 7:** **molecular genes** | **STAGE 8:** **protein synthesis** | **STAGE 9:** **organic organic-synthetic machinery** |
| In one line of organisms a class of interlocking structural polymers (protopolynucleotides) can replicate as an alternative mode of synthesis creating a secondary minor genetic material within the organisms | The secondary genetic material proves to be more versatile: it comes to control some steps in the synthesis of organic molecules, especially the joining of amino acids into peptides | With the ability to make protein, and the (late) discovery of enzymes and of lipid membranes, alternative, more efficient techniques of organic synthesis become available: all the clay machinery is dispensed with |

Figure 8.1.

trend towards indirect action by a primary genetic material might be expected. Our questions for this chapter are on why and how organic molecules became involved in this trend. We will be concentrating on the middle three of the nine major stages suggested in figure 8.1.

### From inorganic crystals to organic molecules

If an inorganic crystal life is easier to start up, no doubt an organic molecular life is more efficient in the end. So given a mechanism (genetic takeover) you would expect early evolution to change from one kind of life to the other. You might say that evolution started with inorganic materials not because it wanted to, but because it could not have started in any other way – practical organic chemistry is altogether too difficult for unevolved organisms. In organic synthesis there are not the error correction mechanisms of crystal growth processes. The main virtue of organic chemistry – the fidelity of intramolecular bonding – is thus also at the root of its difficulty: the outcome of a loosely controlled organic synthesis is usually a very complicated mixture of metastable molecules – a tar of some sort.

While semi-chaotic organic mixtures can be useful – and we will be thinking about some possible primitive biological uses for such mixtures later – it is hard to see very precise higher order structures being put together from organic molecules on the primitive Earth. From our discussions in Chapter 5 it is especially difficult to see how information could be replicated through semi-chaotic organic mixtures.

### '*Self-assembly*'

Looking at our present biochemical subsystems we can see organic molecules specifying more or less precise higher order structures. A membrane of a particular thickness and fluidity will put itself together from a suitable collection of lipid molecules (figure 8.2). A groove with a particular catalytic effect may be specified by a folded polypeptide sequence. These and other forms of 'self-assembly' are a crucial part of the whole technique through which modern genetic information is expressed – through which a message is converted into action. Given the competence to specify a suitable set of covalent bonds, such objects as membranes and catalysts may need no further specification. In terms of an idea introduced towards the end of Chapter 3, organic chemistry has evidently an enormous efficiency range for several key biochemical functions. But that is not to say that organic molecules could have carried out these functions to begin with. I would suggest that for most biochemical functions devices made from organic molecules would have a curve like curve II in figure 3.10: great potential but no competence at low levels of preorganisation.

Even to make a lipid membrane, for example, the molecules should be amphiphilic and at least roughly of the right shape (Israelachvili, 1978; see



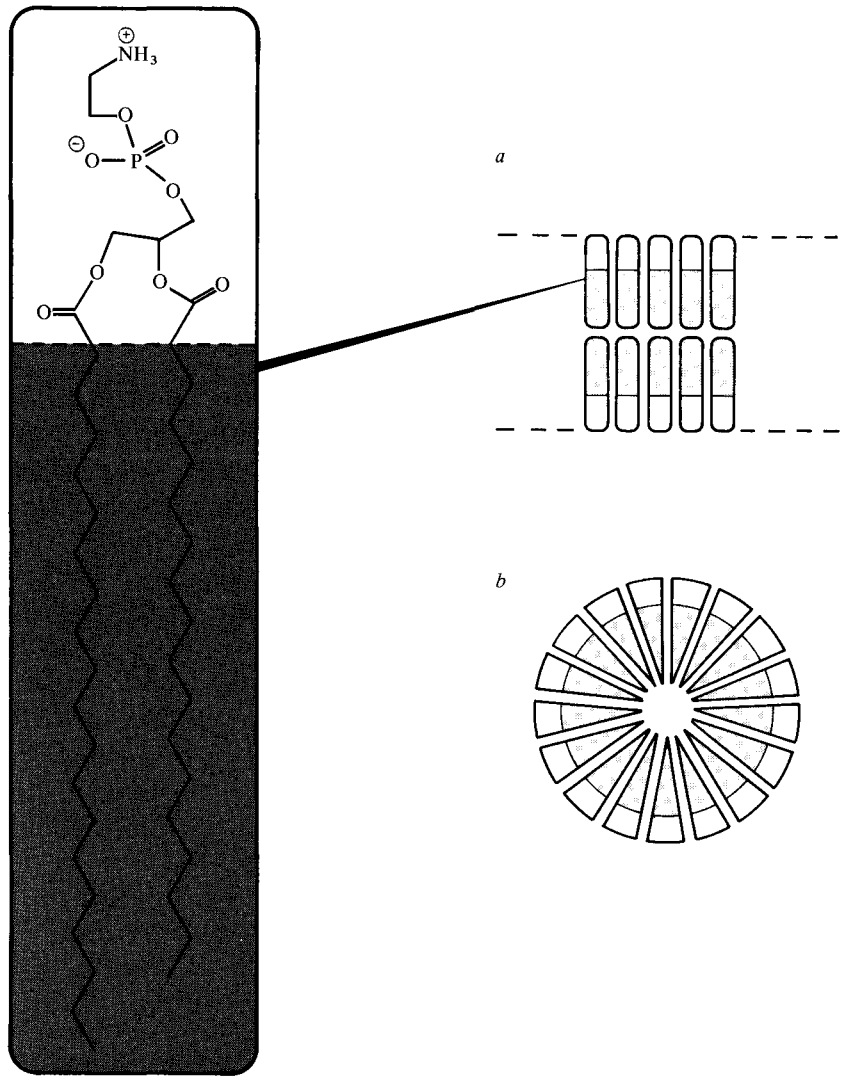



Figure 8.2. (*a*) The pair of hydrophobic chains in phosphatidylethanolamine has a similar cross-section to the (hydrated) hydrophilic head group. Such lipid molecules may 'self-assemble' into an extensive membrane structure. (*b*) Single-chained lipids are effectively wedge-shaped and tend to form finite micelles instead.

legend to figure 8.2). And much more prearrangement is needed for a protein to fold up consistently so as to make, say, a catalytic groove. Generally speaking organic molecules have to be preorganised to a considerable extent before they will organise themselves at all precisely.

We can see this in the familiar phenomenon of crystallisation, which is an example of a 'self-assembly'. In order for an organic material to crystallise it is usually necessary that the material should be fairly pure.

*Requirements for a 'self-assembly'* We can list three very general requirements for a 'self-assembly' from molecular units:

1 There should not be too many kinds of molecular units present.
2 There should be a sufficient cohesion between the units.
3 The bonds between the units should be reversible.

Now it is difficult for a loosely controlled organic product to satisfy conditions 1 and 2 at the same time. And for this reason: the forces between organic molecules are intrinsically weak, so these units have to be fairly large if they are to hold together. But if they are fairly large, and products of poorly controlled reactions, they are likely to consist of a plethora of structural types. It is no coincidence that organic geochemical products are typically tars.

Why are inorganic minerals so often crystalline? The reasons are that here the units are held together by strong bonds (polar covalent mainly), and because this bonding is reversible under the conditions of crystal growth. Hence the effective units are atoms rather than molecules. The units thus belong to a far more restricted set of possibilities. All three of the above conditions can thus be satisfied with much less preorganisation.

It might very well be argued that if the units for crystal life are to be atoms there would be too little scope for variety. It is all very well, you might say, to have a very limited set of units that can thus organise themselves more easily; but if the units are that simple they could only give rise to a very limited number of kinds of higher order structures. With only the choice between a handful of kinds of atoms (rather than between the astronomical numbers of organic molecules) how could really specific objects ever be engineered?

To answer this we must note again the distinction between the X-ray crystallographer's 'crystal structure' and the real structure of a crystal – the electron microscopist's structure. It is only the idealised crystal structure that is truly inherent in the collection of units out of which the crystal forms: the defect structure of a real crystal – its arrangement of dislocations, twin boundaries, substitutions, and so on – is something over and above. And of course it was just such features that we saw as constituting the genetic information of crystal life, and through such features that its phenotypic characteristics could arise. A defect structure is a metastable patterning that is not itself makeable by a 'self-assembly'. *In crystal life the defect structures of crystals are the nearest analogues of the covalent structures of molecules in molecular life.*



A defect structure is one shade more abstract than a molecule – it is not so much a thing in itself as a distortion or patterning of another thing, a form of writing in an otherwise blank crystal structure. That more abstract form of organisation of matter is more easily replicated I think; and when it comes to the specification of control structures, such as membranes or specific adsorptive or catalytic sites, it has a head start over a collection of small molecules, which have a great deal of highly co-ordinated joinings and foldings ahead of them before they can be built into a precise coherent structure large enough to exert a control on others. The task is easier when it is through the modification of a structure that is at least large enough, and coherent, in the first place. One might say that crystal life is a modified perfection, while molecular life is a tamed chaos. The second is the more creative approach and has greater potential, but the first is easier to bring off.

*The origin of chiral discrimination (continued from Chapter 1)*

The essential sophistication of the molecular way of life can be seen at its sharpest in the problem of the origin of chiral discrimination. The gist of the problem, as we discussed it in Chapter 1, is this. Most organic molecules, beyond the very simplest, are chiral: so there would be severe limitations on any molecular life that had to stick with achiral units. But if you use chiral components to make specific sockets, or indeed machines of any sort, you must be able to distinguish between enantiomers. If you were to build up some molecular machine you would soon begin to see this: one molecule or side chain would not fit with another because the chiral relationship between them was wrong. And as more units were built in, the number of relationships and number of ways in which these could be wrong would increase very rapidly. More particularly, the functioning tertiary structure of a protein molecule could not be specified at all by an amino acid sequence unless the chirality of the amino acids was specified. This would be true for the specification of a socket of any sort, never mind a chiral one. It is not that you might get a mixture of D- and L-sockets, you would simply never get the same socket twice because each protein mole-

cule would contain a different mix of right-handed and left-handed amino acid units. (All the protein molecules would be chemically distinct diastereomers with their own ways of folding.)

The problem of how to handle chiral molecules is pressing and immediate for *de novo* molecular life – or at least for anything like our kind of molecular life. Yet the problem is not there at all at the beginnings of an inorganic crystal life – because that kind of life would not use chiral units.

The information in a crystal life is built in at the next level up from its units – exclusively in the way the units are arranged. And this would often include chiral information. One kaolinite domain, for example, must be either 'left-handed' or 'right-handed', and generally this handedness is copied between layers. Indeed the transmission of chiral information in growing crystals is very well known – even quartz can do it. This is the one piece of information that the X-ray crystallographer's ideal crystal structure can convey – whether to be right-handed or left-handed. But the defect structure could also, independently, be chiral: defects such as domain boundaries would almost certainly create chiral surface features – replicable perhaps if the speculations of the last chapter have any truth in them. In any case, a screw dislocation cannot help being of one chirality and, in any case, often replicates as a crystal grows.

*Organisms would have been needed* It must be stressed that this way of approaching the problem of the origin of chiral discrimination depends on the idea that the crystals that we have been talking about – or rather their defect structures – were products of natural selection. Otherwise there are too many objections of the sort that we discussed in Chapter 1.

One of these objections to the notion that minerals first injected chirality into life on Earth was that for every 'right-handed' crystal structure or crystal defect there should be, by and large, another 'left-handed' one. A universal bias one way, although quite conceivable, has not been clearly demonstrated and would be weak anyway. Chiral discrimination by the environment would either be localised or very weak; and multiple chiral discrimination – where the environment was making some appropriate *set* of chiral choices – could only be hoped for in restricted, probably evanescent, environments. In any case multiple chiral discrimination would rapidly become quite implausible as the number of members of the set increased. But this objection only holds for non-biological systems.

The extraordinarily non-statistical character of the outcome of evolutionary processes is one of the main distinctions between biology and the



rest of physics and chemistry. This arises from the repeated selections of selections of selections. It might not take very long, on a geological time scale, for an entire species to share a genetic modification that originated in one individual. There are admittedly a few examples in the physicochemical world where, seemingly, a single molecular event triggers a macroscopic effect: the whole pattern of frost flowers on a window may have depended on just where and how one crystal nucleus became stable enough to grow. But on the whole the physicochemical world is more statistical than that. What one molecule does doesn't matter. But what is quirky and unreliable at best in the physicochemical world is part of normal procedure in the evolution of organisms. Here the effects of a few key molecular events can be retained and amplified indefinitely because there exists the hereditary machinery to do it. Just one mutation in one gene in one individual of one species can become established to represent, eventually, the main form of that gene throughout a group of organisms spread worldwide.

In the kind of crystal life that we were thinking about at the end of the last chapter the crystals that were forming continuously in the environment had, like other kinds of organisms, been evolving the means to stabilise their own production, to improve the rate and fidelity of the printing of the genetic information and to propagate that information more widely. In very easy environments the spontaneous generation of (unevolved) organisms might complicate matters; but in more difficult environments only organisms that had phenotypes that were evolved to some extent could survive. In such places there would be no more spontaneous generation and 'common ancestor effects' of the kind discussed above could be expected: you could explain the universal occurrence of, say, a particular left-handed socket within a certain species of the primary life in the same way as you might explain the (near) universality of our genetic code – because all the extant systems are remote progeny of a single ancestral system that happened to be like that.

The general explanation for the origin of chiral discrimination, then, would be as follows. First life, a crystal life, did not use chiral units and did not need to be chirally discriminating. As it evolved, some of the species of this crystal life came to use organic molecules as phenotype components. These molecules were at first 'optional extras', useful but not vital, not part of central biochemistry which was still wholly inorganic. To begin with, the organic components were probably ill organised mixtures with unsophisticated functions. But in some biological niches there were advantages to organisms that had slightly better organised mixtures. Then, later

still, when polymers and multimolecular structures became important, there were advantages to organisms that could make more chirally uniform products. Gradually, with the invention of increasingly sophisticated organic phenotypes, chiral discrimination improved through surface crystal sites becoming more closely engineered to interact with particular enantiomers of particular molecules.

Chiral discrimination would have emerged gradually, then, and as part of the emergence of a more general competence in recognising and handling molecules. In so far as molecules in the primary biochemistry were recognised through interactions with sockets of some sort, the recognition of an enantiomer would not be qualitatively different from other acts of recognition: from the point of view of a socket the wrong enantiomer of the 'plug' may be a very poor fit – like a right hand in a left glove. Once an organism could distinguish between, say, isoleucine and valine it would probably have been able to see the difference between D and L. But I do not think that either of these functions was a very early evolutionary achievement.

### *First uses for organic molecules (including polyphosphate)*

So far in this chapter we have been enlarging on arguments introduced earlier as to why life would have started with inorganic control machinery (because that was easier) and then would have moved to organic machinery (because that would have been better). In the very broadest terms we can see why. But first life would not have been able to foresee that organic chemistry was to be the coming thing. So we must ask a more short range 'why?' Why were organic molecules first brought into organisms?

One can only speculate. But the question is not particularly perplexing so long as one does not try to relate the uses of organic molecules then to the uses of similar molecules now. Early organisms would not only have been less competent at handling organic molecules but they would have had different needs from organisms now. Thinking about the kinds of clay organisms that we discussed at the end of the last chapter one can see how quite crude mixtures of organic molecules might have affected the flocculation behaviour of clay assemblages through their influence on inter-particle associations.

Organic anions, particularly oligo- or polyanions can have major effects on the rheological properties of clays by acting as peptising (deflocculating) agents. They stick to the edges of clay layers cancelling net positive charges there and hence breaking up 'card-house' structures. Such effects can be



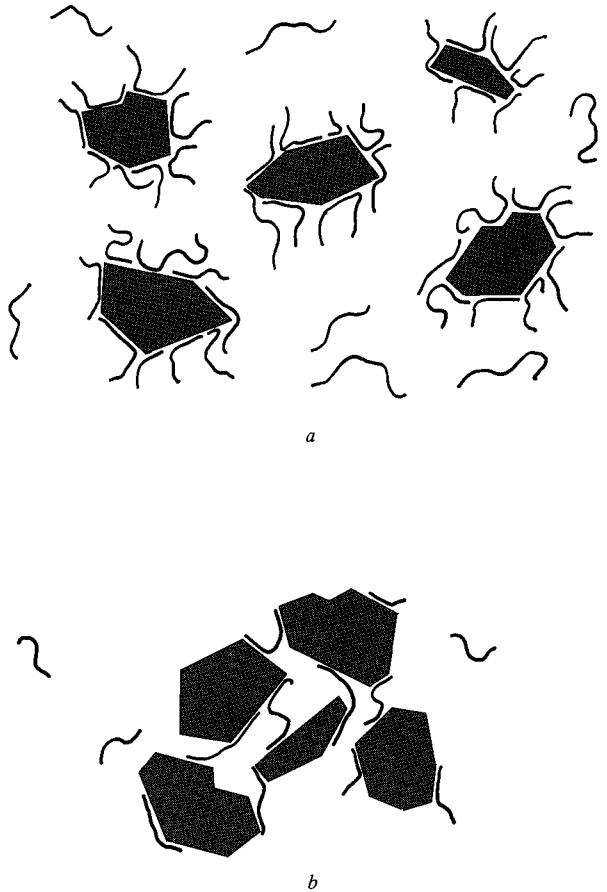


Figure 8.3. (*a*) Clay crystallites being kept apart by a charged hydrophilic polymer. (*b*) Smaller amounts of the same polymer may flocculate the crystallites.

quite complicated. Small amounts of the polyanion carboxymethyl cellulose can act as a peptiser for montmorillonite while *very* small amounts have the opposite effect, helping instead to join the particles together (Van Olphen, 1977; see also figure 8.3). Tannins – which are phenolic oligomers – are also effective peptisers and are used in the oil industry to control the rheology of drilling muds. Polyphosphates are potent peptising agents that probably act by co-ordinating to exposed cations along the edges of clay layers (figure 7.9). A firm lump of clay may liquify and flow if treated with 0.5 % of its weight of a polyphosphate with an average chain length of fifteen units (Corbridge, 1978).

Again physical characteristics of clays can be greatly affected by the intercalation of organic molecules between the layers – for example organic cations standing in for interlayer metal cations, or polar organic molecules for interlayer water.

Apart from effects on rheological properties, flocculation behaviour, and so on, organic molecules may be catalysts for clay synthesis. As discussed towards the end of Chapter 6, oxalate as an aluminium chelator may encourage kaolinite synthesis. Even a crude mixture such as fulvic acid may be effective. More generally one might see it as a major early function for organic molecules, that were serving a fundamentally inorganic kind of life, that they should be effective ligands for metal cations.

We can see, then, a number of reasons why it might have been useful for clay organisms to have been able to synthesise (or somehow acquire) certain classes of organic molecule. Speculating further one might guess that Krebs acids and amino acids were originally incorporated for their metal complexing prowess; that heterocyclic bases of various shapes and sizes were designed originally to intercalate clay layers; that oligophosphate chains like that on ATP were originally means of attaching organic molecules to clay edges. I will come back to more specific speculations of this sort later. But there are more serious questions, 'how' questions of a general sort, that we must tackle first. How did primitive organisms acquire such expertise in organic chemistry that they could reach the point at which molecules of the complexity of nucleotides were being made? How was organised organic chemistry possible without either organic chemists or enzymes?

### Organic chemistry without enzymes

The first techniques for controlling organic reactions would not have been determined by any tendency to anticipate protein. Clays would have controlled organic molecules in ways that made appropriate use of inherent properties of this kind of material. There would have been similarities, I dare say, between some aspects of the first and later techniques. For example, chiral discrimination was perhaps achieved through specific binding sites for clay life as it is for ours. But we are probably being too optimistic if we imagine that any inorganic crystal surface, however contrived by specified defects, could achieve the selectivity of proteins, or seriously mimic the selectivity and catalytic ability of a protein enzyme.

Another way of controlling organic reactions is the organic chemist's way – with larger scale apparatus. Here you do not have specific catalytic



sites that can pull particular molecules from a complex mixture so as to specify particular reactions and sequences of reactions. You do not manipulate molecules as individuals. Instead you manipulate the conditions to which groups of molecules are subjected. But operating that way you will not be able to carry out different operations at the same time in the same reaction vessel. In the absence of highly efficient selective catalysts, sequences of reactions must generally be controlled through sequences of operations: solutions must be mixed, conditions adjusted, side products removed, and so on. In those earlier organisms that we are trying to imagine there must have been, I think, in place of protein, some quite elaborate chemical automation. At its simplest, to keep the wrong molecules apart there had to be 'glassware', and to put the right molecules together there would have had to be 'tubes' of some sort – and 'pumps' as well, I dare say.

Not that the modern cell is entirely without such techniques – it is not just a bag of enzymes by any means. Organisms today do not rely solely on the specificity of their enzymes to keep control. Even the prokaryotic cell is quite highly structured. Referring to the internal morphology of cells generally, Mahler & Cordes (1971) remark: 'These complex structures, we are beginning to suspect, hold much of the secret of how cellular processes are controlled in both space and time: the secret may consist, at least in part, of isolating and maintaining the different cellular constituents – mainly enzymes, together with their substrates, products and modifiers (activators, inhibitors) – in different compartments; sometimes allowing, sometimes denying mutual accessibility.'

An indication that early approaches to organic chemistry might have had to put much more reliance on compartmentalisation comes also from the consideration, from Chapter 1, that modern enzymes are on the whole only discriminating enough to be able to work with the sorts of mixtures that they normally come across. Within cells these mixtures are of a fairly limited number of kinds of molecules. Much more complex mixtures would have existed if the control had been weaker in the first place – if the transport proteins and enzymes had been less competent. Where a system goes very well (in a certain way), provided it is going very well (in that way), you naturally look for some other way for it to have started. In this case you look for another way, I think, that relies much less on sockets, because not only is it hard to see such particular things as sockets being made from very complicated mixtures of organic molecules, but it is doubtful whether socket control would work on very complex mixtures. You look for something that is more like the organic chemist's kind of practical chemistry.

*Elements for an organic synthesiser*

So let us start to think about very early biochemistry in terms of what sorts of things are needed to bring about sequences of organic reactions automatically – if you are not allowed to use enzymes or other such high-fidelity microscopic control devices. Let us first draw up a list of necessary or desirable hardware. After that we can try to see which if any clay materials might be structurally 'preadapted' for such elements, and under what general circumstances evolutionary processes might have, in effect, assembled them, first into rudimentary and then into increasingly sophisticated automatic organic-synthesising machines.

*1. Basic 'glassware'* There would have to be functional equivalents of the chemist's flasks, tubes, T-junctions, and so on, for controlling solutions. There would have to be the means of 'sometimes allowing, sometimes denying mutual accessibility' as Mahler & Cordes put it when describing the membranous 'glassware' in modern cells. It is not essential that anti-mixing devices should be membranes. Different populations of molecules might be kept apart by being adsorbed on separated surfaces, or by being immobilised in a gel, with mutual accessibility being controlled in this case by diffusion processes in the gel. Also much of the 'glassware' might be in the environment – for example if the pores of a sandstone acted as containment devices for very early clay organisms so that they needed no boundary membranes of their own. But, all the same, if the 'glassware' is eventually to be highly contrived by genetic information, some kind of inert, manipulable, membranous material would seem to be at least highly desirable.

*2. Energy transducers* This term can cover a wide range of devices in which there is a change in a form of energy. For example, a furnace converts chemical energy into heat; a steam engine converts heat into mechanical motion, and so on. Within organisms there are many energy transductions. In photosynthesis, for example, light energy collected from the environment goes through several forms before being used or stored as a fuel. As we will discuss in more detail later, one of these forms is as a proton potential across a membrane. Other potentials are generated through various kinds of pumps in modern cells – cation pumps, anion pumps, water pumps, molecular pumps of many sorts. Such forms of active transport now depend on cleverly asymmetric lipid–protein assemblages (Lodish & Rothman, 1979) that are very far from primitive in their means of operation.



But pumps of one sort or another would have been needed, I think, long before the present pumping techniques were invented. What a serious design restriction it would be for an automatic organic synthesiser that it should always have to run down osmotic gradients. Sooner or later a solution would have to be concentrated or somehow 'unmixed', and you cannot reasonably expect the environment to provide evaporating pools just when needed except perhaps for some especially simple synthetic operations. One might imagine that chemical operations were sequenced, in the pre-protein era, through 'production lines' of some sort, in which case at the very least there would have to be the means to move solutions. Again one might imagine first ways of using flows in the environment (groundwater movements, tides, etc.) but again there would be great limitations in being so dependent: there would be strong selection pressures, if there were pressures to do organic chemistry at all, for more positive forms of the control of the means to do it. An automatic organic synthesiser without pumps may be conceivable, but I doubt whether it would work very well.

*3. Reagents* Condensation reactions are encouraged by conditions under which water is being continuously removed, and so one can imagine an automatic system in which a water pump helps to bring about a condensation reaction by creating a dehydrated region. Similarly, an electron pump can in principle bring about a reduction by providing electrons, or an oxidation by removing them. But 'dehydrating power' and 'reducing power' do not have to be tied to energy transducers in this way; they can be made portable in the form of soluble reagents. ATP and NADH are biochemical devices of this sort and of course the organic chemist has many dehydrating agents, reducing agents, and so on, on his shelves. There are also more specialised group-transferring agents – acetylating agents such as acetyl-coenzyme A or the chemist's acetic anhydride, or indeed ATP, again, as a phosphorylating agent. We could predict, I think, that quite soon in the evolution of a competence in organic synthesis, early organisms would have hit on the idea of using some kind of energy transducer to make some kind of high-energy reagent(s). There would not have to be many such reagents greatly to enlarge the field of accessible organic molecules. (There are not after all so very many kinds of reagents – coenzymes – in our biochemistry; see pages 376–8.)

*4. Catalysts* Homogeneous catalysts such as acids, bases and metal ions would be useful not only to accelerate reactions but to exert

control: A and B react *when* catalyst C is present, for example. Similarly, catalytic sites on solid surfaces might exert control by determining just *where* some thermodynamically favourable reaction took place. And I dare say there would always be some choice about which molecules reacted. But, as we have already discussed, we should not expect anything approaching absolute specificity for any of the catalysts in a pre-protein life. It would be reasonable to suppose only that members of some (fairly large) class of molecules would react at a given type of site, and that within that class there would be differences of rate.

*5. Binding sites* Various kinds of binding sites are inevitable when solid surfaces are present, and these can be affected by defects, such as dislocations, that intersect the surfaces. We discussed in Chapter 7 how genetic information in crystals might have been 'focussed' to create special high-energy sites.

*6. Collectors* We discussed at the end of Chapter 1 the very great difficulties in the idea that primitive life forms could select molecules from a complicated environmental 'soup'. Today's heterotrophic micro-organisms select molecules by means of exceedingly sophisticated lipid–protein transport machinery. Even if one can imagine how a primitive organism could have coped with the tarry organic materials that might have been present in the primitive environment, it is not clear whether indeed such material would have been present except at best in rather restricted locations. But there must have been carbon feedstock(s) of some sorts for those organisms to evolve the techniques of organic synthesis. Where and what would these feedstocks have been? How would they have been collected?

*7. Separators* It is easy to mix things, but in most organic syntheses there are separations to be done as well. In organisms, excretion involves separating what is not wanted from what is, and similarly the working-up of an organic reaction involves the 'excretion' of unwanted side products. As remarked before the work-up is perhaps the most difficult part of practical organic chemistry. How do you 'demix' the components of solutions when you lack very specific binding sites? As I said, pumps of some sort are likely to be needed for automatic demixing operations. Conceivably, given sufficient 'glassware', early organisms could have used some kind of chromatography, that is, some technique based on relatively small differences in the affinities of static binding sites for components in moving solutions.



*8. Sequencing mechanisms* Even if early organisms had evolved the means of constructing all the types of apparatus that we have referred to already, there would still be problems of master control. Procedures would have to be specified. There would have to be somehow written into the assemblage of apparatus such instructions as: 'after this has happened, do that'. A production line is one way of building in such instructions, but you might need other sequencing techniques as well. It would be at least desirable to have timed procedures in some cases, as with Merrifield's automatic peptide synthesiser (figure 1.11), that is, to have a set-up that was more like a washing machine than a conveyor belt – in which different things happen in the same place at different times. Diurnal and other environmental cycles might be able to impose a programme on early organisms as Kuhn (1976) suggested (see page 150). The world might provide a set of time switches. But it would be much better all the same to have a more positive kind of control; not to have to depend on the outside world in this way. And it becomes less and less likely that more complicated procedures could be made to fit in with what the world happened to be doing. Built-in means of performing such operations as periodically reversing flows, or opening and closing control valves, would be very much better.

*9. Feedback control* Even programmed time sequences are not enough for the more sophisticated forms of automatic control. In a washing machine, for example, rather than have instructions of the sort 'pump water in for 30 seconds' or 'heat the water for 2 minutes' it may be better for the instructions to be along the lines of 'pump water in to such and such a level' or 'heat the water to such and such a temperature'. Where water pressures and ambient temperatures are variable there are obvious advantages in this more intelligent way of working – where sometimes goals are specified rather than all the details of the procedures needed to arrive at these goals. But to work like this a mechanism must contain some kind of sensor to see how things are going, and then some means of communicating this information to an effector that can initiate, or maintain, or adjust, or stop an appropriate action. Such mechanisms can be fairly simple. For example, in a ballcock valve the ball is the sensor which communicates information about the water level to an effector, an inlet tap, through a connecting rod. A self-controlling mechanism of this sort is said to contain a feedback loop – information about an output is fed back to control an input.

It is almost what you mean by an automatic device that it usurps the main function of the human operator in that it can see when things are

wrong and make appropriate adjustments. So we might guess that to replace the considerable skills needed to perform a multi-stage organic synthesis one would need automatic machinery with many feedback loops built in. I doubt if the synthesis of, say, a nucleotide could be controlled wholly by a blind sequencing of operations. Apart from inevitable variations in conditions – in the composition of the feedstock, in the ambient temperature, and so on – the machinery itself would be time-variable: often a tube would become slightly clogged or a catalyst slightly poisoned. Sooner or later something would go out of synchronisation.

In any case feedback loops are part of industrial techniques of automatic control of chemical operations, although there is nothing approaching the automatic synthesis of a nucleotide here. The only chemical factory with that sort of expertise is the modern living cell. It is also well endowed with mechanisms for feedback control. For example, the concentrations of certain molecules are held at appropriate levels through their acting as inhibitors for enzymes involved in their production. This control is generally allosteric (see page 224). One site (or set of sites) acts as sensor – it binds the inhibiting molecule. The information that the molecule is bound is transmitted by allosteric means – through the articulated intervening structure of the enzyme protein – to the effector which is the active centre of the enzyme. Often the reaction which is being inhibited is several steps back along the production sequence, so there may be very little structural relationship between the molecule that is acting as inhibitor and the molecule whose synthesis is being directly inhibited.

It is typical of a clever control device that it makes a contrived causal connection like this. Why should a rise in temperature cause an electric current to flow in a motor winding? That happens in a refrigerator because the thermostat makes the connection. In modern organisms proteins are almost invariably the means of making contrived connections: they know what to do and when to do it. They know to shut off the catalysis of this reaction, or let in that sodium ion, or turn on the synthesis of the other messenger RNA molecule. (Looked at one at a time such control techniques might seem to be mere refinements: but taken together I suspect that they make our kind of biochemistry *possible* in much the same way that the various feedback control devices in an airliner make that way of flying possible.)

So we might add to our shopping list of components for pre-protein organic-synthesising machinery, the elements needed for some sorts of feedback control devices – sensors, communicating links, and effectors – without for the moment being very clear as to exactly what these objects might be like, except that they could not have been protein. Perhaps when we get into the shop we will see some things that might do.



### Minerals and the primary synthetic apparatus

I am not suggesting that to start to do organic chemistry organisms would have needed all nine of the types of apparatus discussed in the previous section. But it seems to me to be a common sense, based on observations of how organic chemists go about things, that many if not most of such items would have to be included in practicable machinery that was to bring about automatically the consistent clean manufacture of particular moderately complex organic molecules. As discussed in Chapter 1, nucleotides and lipids are really very difficult molecules to make and even a particular set of a few amino acids of a particular chirality would not be easily contrived. And yet such supplies would be the minimum, it would seem, before our kind of biochemistry could even start to evolve. What is more, as discussed in Chapter 5, the general requirements for *any* kind of spontaneously replicating organic polymer are such that its monomer units could not be very simple. Then, as discussed in this chapter, the 'self-assembly' of organic molecules into any very definite kind of structure would call for an ability, already, to specify a set of covalent bonds, that is, a certain minimum competence at organic chemistry. Perhaps vesicles could be made from semi-chaotic organic mixtures, but that would be far from enough to direct the sequences of operations needed to do organic chemistry properly – to arrive at the stage at which molecules were being produced competently enough to provide the machinery on which competent molecular synthesis could be based. There is a clear vicious circle here that cannot be broken, I think, within the ambit of organic chemistry itself. Just how vicious the circle is can be seen from the following consideration. Even with explicit knowledge of organic chemistry and the ability to plan sequences of reactions; even with supplies of pure and often quite complex organic molecules; even with the ability to choose and set up appropriate apparatus, and to choose an appropriate solvent, temperature, pH, and reaction time; even with the ability to monitor reactions, to make adjustments if need be, and then eventually to separate out a desired product; even with all such things on his side the organic chemist is hard put to it to make molecules that would form themselves into the kinds of microscopic apparatus that can control the making of other organic molecules.

It is not, it seems, in the nature of *simple* organic molecules, or of semi-

chaotic mixtures, to constitute the elements of organic-synthetic machinery: as I will try to make more explicit in the rest of this chapter, inorganic minerals seem much nearer the edges of the necessary competence.

*Virtues of inertness* The membranous character of so many clays, and their predisposition to form vesicles, tubes, etc. has already been commented on, and illustrated in many of the electron micrographs of Chapter 6. Apparatus made from clays would often be relatively inert with respect to organic molecules. These minerals do not readily react covalently *with* organic molecules under aqueous conditions. And the catalytic effects of clays in aqueous solutions and at ordinary temperatures are quite limited – indeed clays may often stabilise organic molecules (see for example Van Olphen, 1966).

In any case it is not usually the general clay structure that shows catalytic activity but special sites – typically Lewis or Bronsted acid sites, or oxidising or reducing sites, on the surfaces or edges of layer silicates (Solomon, Loft & Swift, 1968a, b; Theng, 1974; Van Olphen, 1977). In looking for the ideal kind of control material for an organism (cf. protein) one positively does not want a material with promiscuous catalytic effects: one wants a material which shows catalytic effects when suitably contrived by genetic information. For crystal life that means, probably, that the effect should depend on the defect structure of the (crystalline) control material. And that seems to be the case, very often, with clays.

Because clays do not readily form covalent bonds with organic molecules the controllers and the controlled could be more easily kept apart. By contrast, a semi-chaotic mixture of organic polymers and oligomers would be a poor sort of material for a reaction vessel if, as would seem likely, it became embroiled with the reactions that it was supposed to be containing. Organic chemists are not just being old fashioned in preferring, very often, glass to plastic for reaction vessels – even for low-temperature reactions. In making a complex chemical circuitry it would be as important to have chemically inert materials to hand as it is important to have insulators in making electronic circuitry.

On the other hand, the term 'living clay' might not be altogether inappropriate as a description of this microscopic chemical apparatus that we are trying to imagine. As we discussed already in the last chapter, an evolved clay, even a slightly evolved clay, might have some queer properties. The question now is this: from what we know about the properties of ordinary clays can we indeed imagine apparatus being specified by clays with defect structures that had been contrived by natural selection? Or put



another way: suppose that you could engineer the sizes and shapes of clay crystallites, and suppose that you could build in more or less particular arrangements of cation substitutions, could you thus make crystallites that would be, or would self-assemble into, components of organic-synthetic apparatus?

*Card-houses and bundles of sticks* One possible construction system we have discussed already – the 'card-houses' that result from edge–face associations between clay platelets (figure 7.10). A collection of clay crystallites of defined shapes, sizes and charge distributions might come together to make a complex defined microporosity where again and again there were similar suites of 'rooms' similarly interconnected.

The electron micrographs of zeolites shown in figures 6.26 and the lath-like sepiolite of figure 6.48 suggest another way in which crystal habit might have been able to define compartments and connections between them. Not only do these minerals have channels in them as part of their crystal structures but the packing together of their crystallites, and especially the packing of grooved rods into bundles, could define pores – perhaps quite precisely.

*Mixed layers and micro-origami* Smectite layers that have been separated in solutions of low salt concentration will reform multilayer crystallites when the salt concentration is increased. Where mixtures of differently charged layers are present these may come together selectively to give two kinds of crystals, or there may be a segregation into vertical domains, or regular or irregular mixed layers may be formed (Frey & Lagaly, 1979). These possibilities are illustrated in figure 8.4*a–d* respectively.

This is a kind of 'self-assembly'. There is even an element of specificity about the mode of assembly in all but the disordered case. The layers must be recognising each other to some extent to arrange themselves as they do.

How this happens is not fully understood. But it is not altogether mysterious when one considers how subtle can be the dependence of interlayer forces on the structures of clay layers. Our discussions on polytypism in Chapter 6 illustrated this. To begin with, the strength of the interlayer force is greatly affected by the charges in the layers. Giese (1978) calculates that the energy required to separate (by 0.9 nm) the layers of the highly charged mica margerite is about 750 kJ $mole^{-1}$, whereas for muscovite, with half as many charges, the energy is about 125 kJ $mole^{-1}$. For (uncharged) pyrophyllite layers the separation energy is only about 25 kJ $mole^{-1}$.

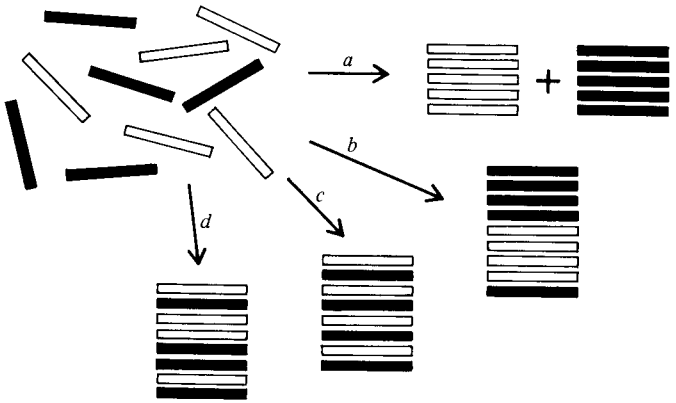



Figure 8.4. Alternative modes of assembly of low- and high-charged smectite layers. Interlayer cations are not shown. (After Frey & Lagaly, 1979.)

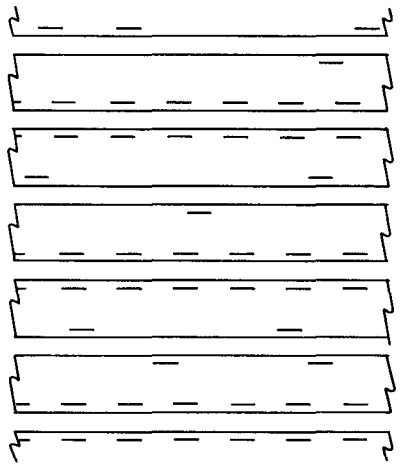


Figure 8.5. Arrangement of polarised layers in rectorite. Interlayer cations are not shown. (After Lagaly, 1979.)

The stronger forces depend on negatively charged silicate layers being held together by an intervening sheet of cations. This will only work as a cohesive mechanism if both the layers are charged – and presumably it is strongest when there is a similar charge density on each of the layers. Anyway there is often a tendency for highly charged layers to prefer each other's company and for less highly charged layers to do likewise (Frey & Lagaly, 1979; Lagaly, 1979).



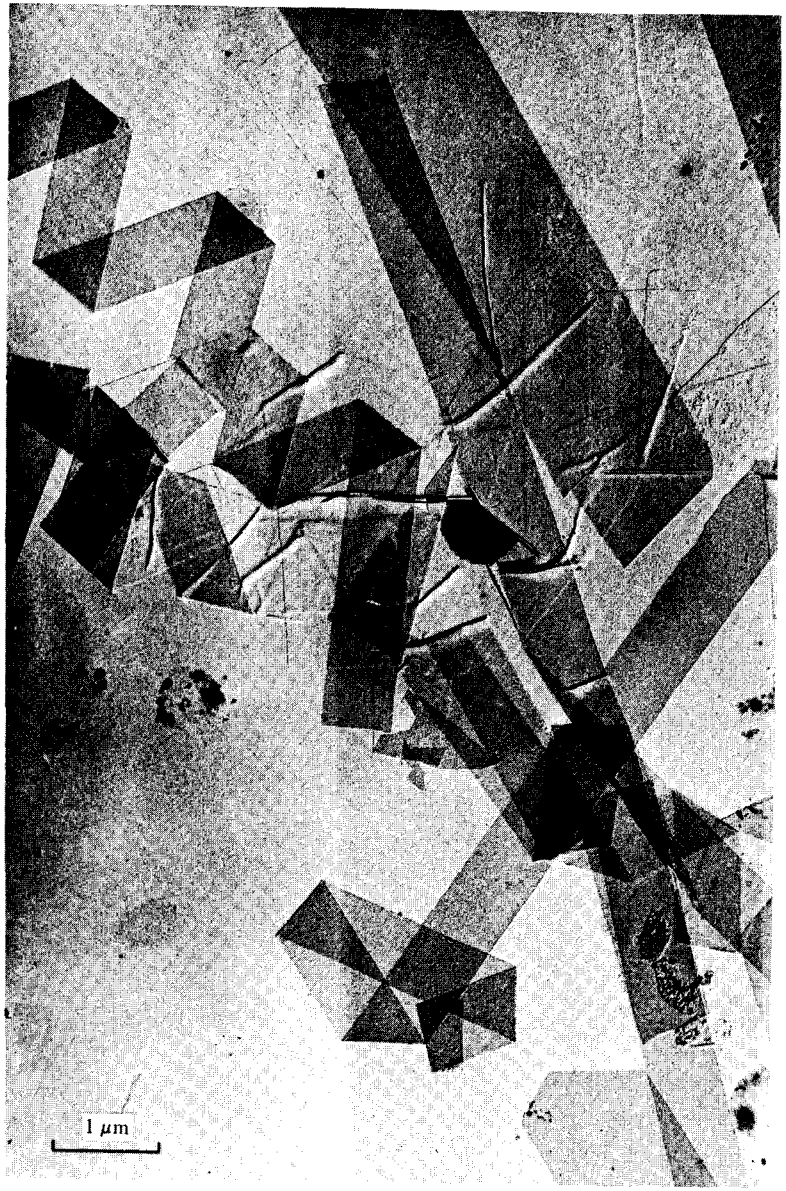



Figure 8.6. Rectorite from Allevard, France. (From Weir, Nixon & Woods, 1962.)

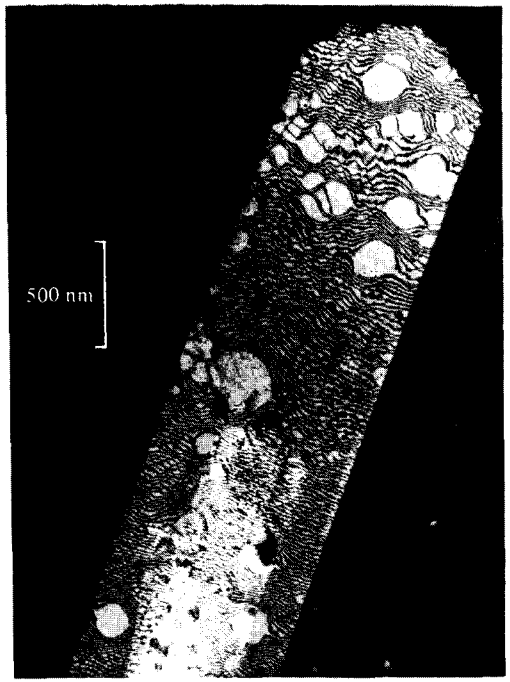



Figure 8.7. Beidellite crystal showing dislocations. (From Guven, Pease & Murr, 1977.)

More strictly it is the charge densities of facing tetrahedral sheets that are important. 2:1 layers are often quite highly polarised with aluminium substitutions concentrated in one of the two tetrahedral sheets. Rectorite is a case in point. In this structure mica-like and smectite-like interlayers alternate, but there is only one kind of (polarised) layer (figure 8.5). The mica-like junctions are very firmly glued, but the smectite-like junctions are not – with the result that rectorite may easily separate into double 2:1 layers. The thinner plates and ribbons in the electron micrograph of rectorite shown in figure 8.6 are such double layers.

Apart from the charge densities on adjacent tetrahedral sheets, the lateral distribution of charges within the sheets must have some effect. For example, if the charging is patchy there will only be strong forces between the layers where the patches on facing tetrahedral sheets at least roughly correspond. And then there is the question of being in register. Substitutions alter bond lengths as well as, often, introducing charges. Now mica-like cohesion depends on the interlayer cations being located in the hexagonally arrayed indentations of the layer surfaces, and these will only lie



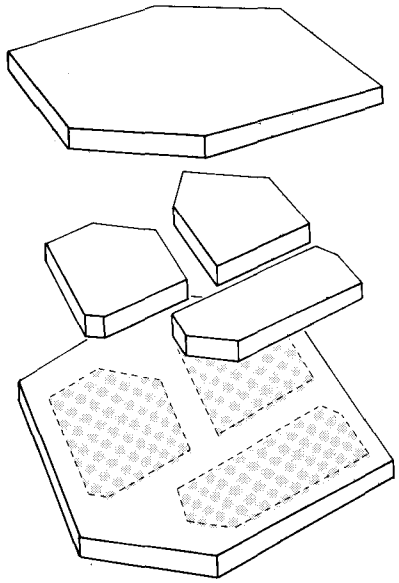


Figure 8.8. If small layers, or packets of them, are located at particular positions between more extensive layers then a complex maze of compartments might be defined (see text).

exactly opposite each other over an extensive area if the two opposing layers have exactly the same dimensions. If either layer is distorted then the other should be similarly distorted if they are to stick together well. The electron micrograph of a beidellite crystal in figure 8.7 shows how smectite layers may be in register in small patches but largely out of register, as indicated by the profusion of dislocation lines. That the layers are indeed firmly glued together where they are in register is suggested by the greater density of dislocation lines around the clear patches.

We can begin to see in all this how tetrahedral and other substitutions in smectite layers might affect the ways in which such layers would 'self-assemble'. Now think about what might be possible if substitutions, especially tetrahedral substitutions, could be contrived. The ways in which 2:1 layers come together could then also be contrived. For example, a 2:1 layer with a complex pattern of high-charge patches would have a specific affinity for another layer with a corresponding (mirror image) pattern of patches – particularly if the pattern was rather complicated and irregular; and particularly too if the patches were exactly in register. Layers with such prearranged mutual cohesion might be of different lateral extents and hence come together to specify very complex interconnected compart-

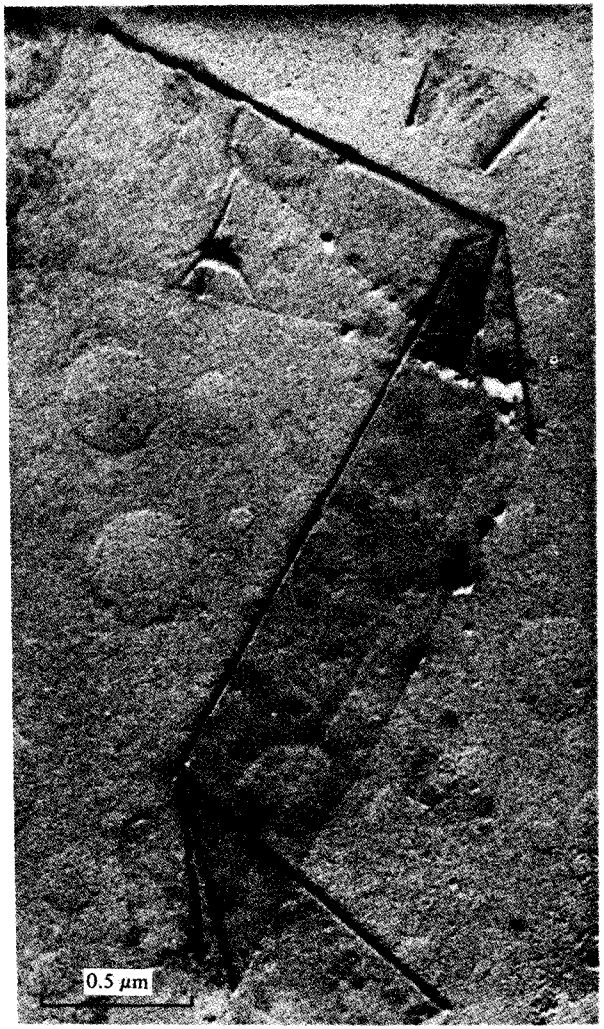



Figure 8.9. Electron micrograph of a folded beidellite ribbon from Black Jack Mine. (From Grim & Guven, 1978.)

ments. A simple example of such a 'self-assembling maze' is imagined in figure 8.8.

But perhaps the most interesting possibility arises when we come to think about a two-dimensional analogue of the globular protein's construction system – when we come to think about forms that might be defined by folding up a layer in a way that was specified by intralayer interactions. The toughness and flexibility of silicate layers or small packets



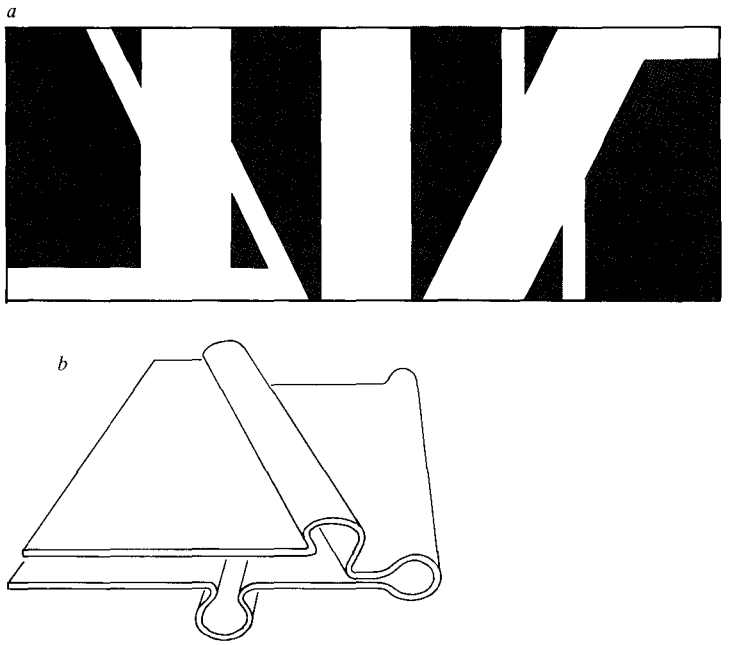



Figure 8.10. Defining a complex object through folding a flexible strip with self-cohesive patches (hatched in (*a*)). See text.

of them is vividly illustrated in many of the electron micrographs in Chapter 6, and especially in rectorite (figure 8.6) and in the crystal of beidellite in figure 8.9. These structures can fold profusely without breaking.

We seem to have the raw materials here for a sort of micro-origami. (Origami: the Japanese art of making objects by folding paper.) Imagine an extensive 2:1 layer, or a small packet of such layers, and imagine that, due to specific substitutions in the tetrahedral sheets, there are pairs of charged patches that key with each other. A specific folding mode might thus be prearranged. (It would be as if one had a piece of paper with areas marked out on it and with instructions such as 'stick area A to area A′, B to B′', and so on – and then when you have blindly followed these instructions you find that you have a box or a toy aeroplane or something.) Figure 8.10 shows an example of the kind of thing that I mean. The patches on the strip shown in figure 8.10*a* can overlap each other perfectly in only one way – and that requires making a pair of intersecting rucks as well as a 180° fold. In such ways tubes of different dimensions might be specified as well as complicated connections between them.

Of the construction systems that we have been thinking about, two –

'card-houses' and 'bundles of sticks' – might have been suited to direct expression of genetic information by 'naked genes'. That is to say, genetic crystals themselves might have come together in such ways to make higher order functional objects.

We decided in Chapter 7 that flexible structures would probably be unsatisfactory for primitive genetic materials – they would not be copied sufficiently faithfully. So construction systems that depend on thin flexible layers would have to be based on other materials that were somehow being produced by the genetic material. One way would be for individual layers sometimes to separate from a genetic crystal to provide what, from then, would be separate phenotype structures. Another possibility, that we considered earlier, would be for layers of some other clay to grow epitaxially on the surfaces of genetic crystals. (Rather like DNA making an RNA strand rather than another DNA one.)

*'Printed circuits' and 'integrated circuits'* Pure metal oxides, and silicates such as mica and kaolin, are generally insulators at ordinary temperatures; but many minerals, such as ZnO, $TiO_2$ and $Fe_2O_3$, are **semiconductors** when they contain suitable defects. Small amounts of cations in a higher or lower valency state than the norm can be particularly effective in creating either ***n*-type** or ***p*-type** semiconductors respectively. In *n*-type materials conduction arises from the electrons introduced by the impurities that are over and above those required for the covalent bonds in the structure. In *p*-type semiconductors there is a deficiency of electrons but these 'positive holes' can also act as charge carriers.

In so far as semiconductivity is a defect-dependent property one might speculate that patterns of electrical connectivity, 'printed circuits', might be particularly readily specifiable by genetic information that was itself in the form of a crystal defect structure. More subtle solid state devices that depend on patterning of *n*- and *p*-type semiconductive zones might also be imagined – some kind of 'integrated circuit'.

*More indirect specification of micro-environments* One can see a number of shells of influence exerted by DNA in modern organisms. DNA controls its surroundings via RNA and then protein, and then more remotely through molecules, such as lipids, that proteins help to make. In higher organisms the chains of command are still more stretched – in the specification of tissues and organs for example. We saw this trend towards indirect genetic control as perhaps the most significant long-term trend in evolution (page 285).



And we have been supposing that this sort of thing would have been happening during the early evolution of clay organisms. With respect to organic reaction control, we have been seeing genetic crystals as having possibly exerted primary control – as 'naked genes'. But control via flexible structures would have been a more indirect secondary control – at the same level as the folded RNA molecule. And it seems likely that there would have been still higher order effects on organic reactions – further shells of genetic influence.

For example, any of the pieces of apparatus that we have been discussing would have tended to create new micro-environments which would have influenced subsequent mineral precipitations – whatever effects there might have been on organic reactivity. Perhaps a zeolite tends to form because a piece of (secondary) apparatus makes a rather alkaline pocket with an inner surface feature that nucleates that zeolite. And then perhaps that zeolite is a catalyst, or an adsorbent (or whatever) for organic molecules in the surroundings. This would then be tertiary control – the level at which protein operates in modern organisms.

At a much later stage, once organic molecules were already being synthesised under genetic control, a whole range of new techniques for the further control of organic reactions would become accessible. There would be new ways of joining clay particles together – via specially designed molecules with sections that would adhere to layer surfaces or edges. And new kinds of organophilic environments could be made between and around clay layers. In particular, it would become possible to specify hydrophobic regions through the synthesis of suitable amphiphilic molecules. For example, silicate layers that have been suitably propped open by long-chain alkyl ammonium ions may then take up the corresponding alcohols (figure 6.53). Here one set of molecules – the organic cations – is collaborating with the clay to create a set of micro-environments that has a more or less specific affinity for the other set of molecules – the long-chain alcohols. In so far as the cations will locate near the negative layer charges one can see perhaps how the disposition of such micro-environments might come under genetic control.

*Osmotic construction* Apparatus specified at primary, secondary and tertiary control levels might often include compartments that developed different solute concentrations and that were separated by semipermeable membranes. This would create various osmotic effects – and the possibility of another kind of construction system: that of the 'silica garden', where a vesicle containing a high salt concentration may grow or

embark on a complicated budding. This happens because not only is there an osmotic pressure tending to swell or burst the vesicle, but the material of the semipermeable membrane is continuously being precipitated.

It is, of course, one of the necessary conditions for a putative 'vital mud' that it should be actively precipitating (although not necessarily very quickly) in its environment. In some respects ordinary muds and soils are models for the stuff that we have been trying to imagine. Here also, very often, clays are actively forming and there may be hierarchies of effect. As we discussed at the end of Chapter 6, the minerals that form at a particular place and time may depend, among other things, on the minerals that have already formed. The difference is that among the mineral species of a 'vital mud' there would have to be at least one that could hold and replicate potentially large amounts of information. That, of course, would make all the difference – because, through it, the mud could develop an evolutionary memory, of how to survive and propagate.

In this section I have been trying to show how, given the means to accumulate the know-how, there might have been a great variety of techniques available for handling organic molecules, via precipitating inorganic mineral structures. In the next section we will start to think about the sources of power needed to run this putative apparatus.

### Power supplies

Radioactive heating inside the Earth is the main power source for the tectonic engine that reforms rocks at high temperatures and pressures inside the Earth and then pushes them up to the surface where they are then metastable. For an inorganic mineral life no other power supplies would have been needed to begin with. Given the slightly supersaturated solutions derived from the weathering of high-energy rocks, the units for the very first organisms could have put themselves together without any further inputs of energy. It would only have been when organic molecules became involved that energy would have become a problem. Or perhaps we should say that it was only when the mineral organisms started to trap energy from the environment that synthetic organic chemistry became possible. If we have abandoned the idea of a primordial soup as a supplier of useable organic fuels, and if we see the primordial carbon source as having been $CO_2$, then, presumably, some primitive technique of photosynthesis was at the beginnings of organic biochemistry.

According to the classical way of looking at it, the first such photosyntheses would have been strictly non-biological, contributing to a primordial soup. But there is also a 'neo-classical' view, as proposed by Granick (1957), Sillen (1965) and Hulett (1969): first photosynthesis was still non-biological, but in place of a general primordial soup, one should imagine instead numerous 'microsoups' being generated continuously by the action of sunlight in localised regions. We will come back to Granick's particular suggestion that organic molecules were synthesised around impure magnetite particles acting as light transducers.



There is a great advantage already in seeing organic molecules as having been produced locally like this (where required, as it were). The organic molecules did not then have to be concentrated from very dilute solutions; they did not have to survive for millions of years, and they could have been made in relatively minute amounts. And a rather inert atmosphere would be a distinct advantage – inert atmospheres are often used by organic chemists to minimise unwanted side reactions.

Mineral organisms that occupied areas close to localised sources of organic molecules might then have become adapted to use these molecules in the kinds of ways that we have already discussed. They might even have evolved mechanisms for entraining photoactive particles (magnetite or whatever).

On our story, however, we can go a step further and imagine that the phototransducing minerals were not simply found in the environment but were *made* by early inorganic organisms with the nature of the active minerals, the shapes and sizes of the particles, and especially their defect structures, contrived by natural selection.

As we shall discuss shortly, mineral phototransducers are very conceivable, but to be at all competent they would have to be contrived to a considerable extent, with a number of components connected together appropriately. I will be suggesting that mineral organisms would have been preadapted to the discovery of photosynthesis because structures that would have been typical for such organisms – especially mineral membranes and semiconductors – are precisely the components needed. I will be assuming, with Hartman (1975a), that the first organisms to use organic molecules were autotrophs with respect to organic molecules, that is, these organisms synthesised their organic molecules from $CO_2$, etc. (These organisms were heterotrophs with respect to their inorganic constituents – they depended on supersaturated inorganic solutions.)

*Catching the energy of sunlight*

Let us first consider briefly four ways in which solar energy might have been converted into chemical energy on the primitive Earth.

*1. Photothermal effects* The simple heating effect of sunlight, by causing water to evaporate, might encourage condensation reactions in ways such as those investigated by Lahav, White & Chang (1978) discussed in Chapter 1 (page 48). One might imagine such effects being contrived by inorganic organisms: for example, a 'vital mud' that had light and dark coloured patches would develop transverse thermal gradients in the sunlight. Such gradients might be useful, if only to drive water pumps.

*2. Photochemical fuels* There is no great difficulty in principle in imagining the localised photochemical production of high energy organic materials on the primitive Earth, or in primitive organisms. All that might be needed would be a reaction such as:

$$A+B \underset{\longleftarrow}{\overset{h\nu}{\longrightarrow}} C+D$$

such that C+D was a metastable combination whose energy could be recovered by catalysing the (usually slow) back reaction. The prospect of making such solar fuels is a topic of considerable current interest that has been reviewed by Bolton (1978). The technical problems are, however, very considerable (and, incidentally, no commercially satisfactory reactions have yet been discovered). One of the main difficulties is in preventing the back-reaction from taking place at once. Quite high activation energy barriers would be needed, or alternatively there would have to be suitable secondary reactions to remove one or other of the immediate products from the scene. In either of these cases, though, some of the light energy would be lost as a fee for the security of that energy. As always, for organic molecules, there are also problems associated with side-reactions. Then again it is necessary that the back-reaction should be somehow harnessable to a useful synthetic process.

*3. Photovoltaic effects* A typical solar cell for converting sunlight into electricity contains a junction between an *n*-type and a *p*-type semiconductor. On making such a junction in the first place, some electrons would spread across the junction from the *n*- to the *p*-region, creating a localised potential difference which would soon stop further flow. Energy



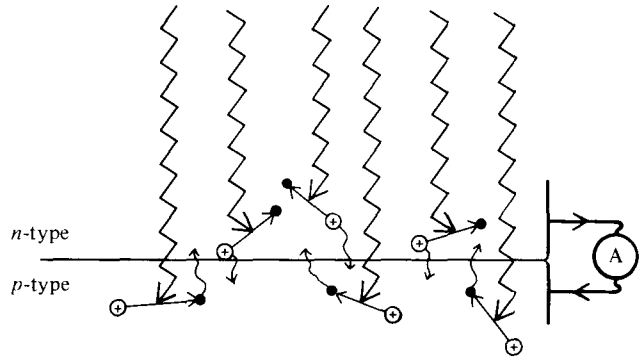



Figure 8.11. Solar cell design. Photons make new mobile charge carriers – electrons and positive holes – near a junction between semiconductors. A localised potential gradient across such a junction tends to separate the charge carriers, opposing immediate recombination and allowing instead recombination via an external circuit (right).

could not be drawn from this potential – the system would now be in equilibrium. On illuminating such a junction, electrons are excited so that new mobile electrons and positive holes are created. The standing electric field serves to separate these immediate photoproducts to some extent – electrons preferentially cross the junction into the *n*-region while positive holes go the other way. Hence energy can be drawn through electrons flowing back in an external circuit (figure 8.11).

*4. Photoelectrochemical devices* We can take a step nearer to the photosynthetic apparatus of modern plants by considering another approach to the trapping of solar energy that has attracted interest recently (e.g. Bard, 1980) – through photoelectrochemical devices that might be said to combine the photochemical with the photovoltaic approach.

If an *n*-type semiconductor is immersed in a solution, some electrons may move from the semiconductor solid into the solution so that at equilibrium there is a potential gradient between the bulk of the solid and the solution. If the semiconductor is illuminated with sufficiently energetic photons, electron–hole pairs will be generated – and they will tend to separate because of the potential gradient, the mobile electrons moving into the bulk of the solid while the positive holes migrate towards the surface. As in the photochemical and electrovoltaic approaches this separation of the products of the primary photo-event is a crucial part of the process of catching the photon energy. But to continue indefinitely

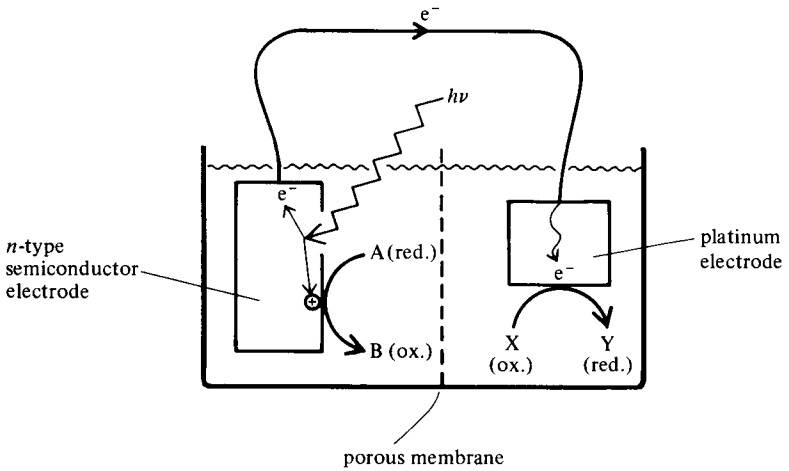



Figure 8.12. An *n*-type semiconductor photoelectrochemical cell in which a redox reaction, $A+X \rightarrow B+Y$, is driven against a thermodynamic gradient. The free electrons and holes produced by photoexcitation tend to separate due to a localised potential gradient, the electrons moving away from the surface of the semiconductor electrode to be transferred to the electrode on the right.

there must be somewhere for the electrons and holes to go. A suitable sink can be provided for the holes if the solution contains some electron-donating (reduced) species which can fill the holes, in becoming oxidised. The electrons, on the other hand, can be taken via, say, a platinum wire to another electrode some distance away (which need not be a semiconductor and need not be illuminated) where a suitable reduction reaction takes place. A formalised photoelectrochemical cell of this sort is illustrated in figure 8.12. An analogous cell can be set up using a *p*-type semiconductor.

It is often necessary in such devices to use an externally applied electrical bias in which case the redox processes are assisted rather than driven by the light. Halmann (1978) reduced $CO_2$ to formic acid, formaldehyde and methanol using an electrical bias and ultraviolet radiation. In one experiment the bias was provided by a solar cell to make thus a more complicated but wholly photosynthetic device. Inoue, Fujishima, Konishi & Honda (1979) have reported another such system.

These are particularly interesting cases in that they involve not only the trapping of energy but the fixation of carbon from the atmosphere. Nitrogen too can be fixed by inorganic photosynthetic means, as Schrauzer



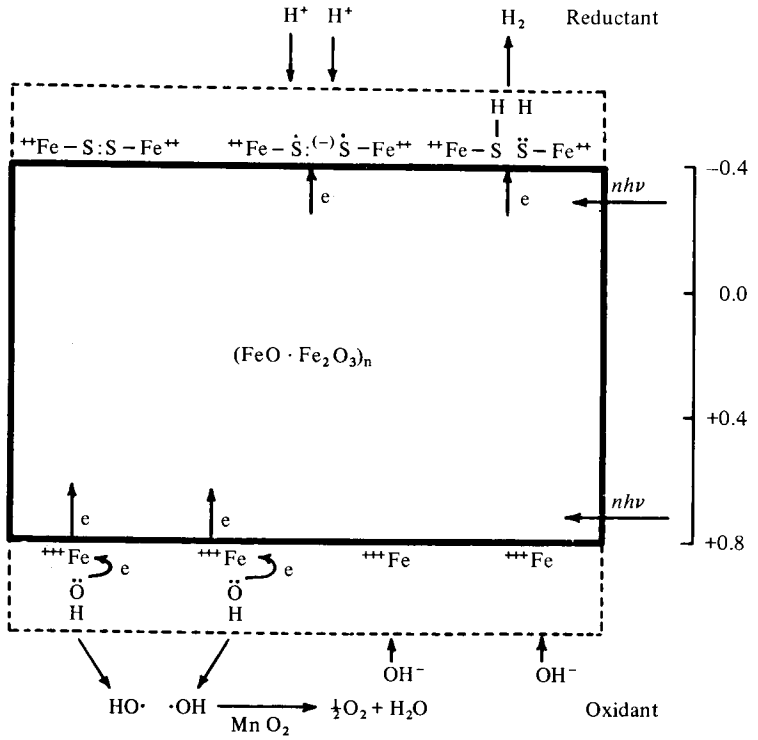



Figure 8.13. Granick's hypothetical device for a primitive photosynthesis consists of a very small impure crystal of the *n*-type semiconductor magnetite. Reduction and oxidation reactions take place on different surfaces. In this case water is being split in two stages: disulphide impurity on one surface mediates the transfer of photo-excited electrons to protons. On the other surface electrons are taken off hydroxyl groups in a monolayer of ferric hydroxide. The OH radicals are then converted to $O_2$ with $MnO_2$. (From Granick, 1957.)

& Guth (1977) have shown: ammonia is produced when damp iron-doped $TiO_2$ (rutile) is irradiated with ultraviolet light:

$$N_2 + 3H_2O + n \text{ photons} \rightarrow 2NH_3 + 1\tfrac{1}{2}O_2$$

Smaller amounts of hydrazine are also formed under these conditions.

*The ideal arrangement* It is a serious limitation of many photoelectrochemical systems that they only work, or only work well, with ultraviolet photons. While it is very possible that there was much more

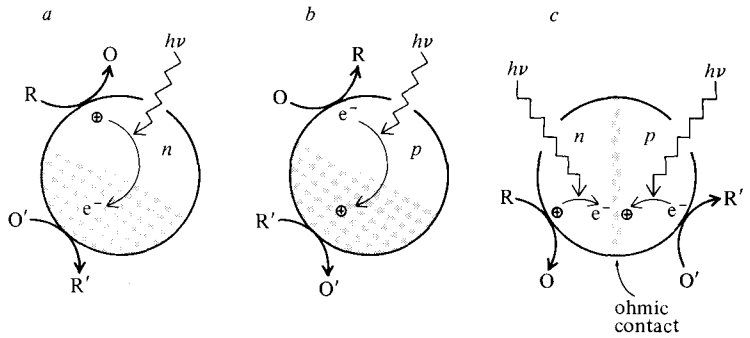



Figure 8.14. Bard's models for semiconductor particles acting as photosynthetic devices. (After Bard, 1979.) (*a*) Mobile electrons in an *n*-type semiconductor move from an illuminated surface into the bulk of the particle. Oxidations may then take place at the illuminated surface while reductions take place elsewhere. (*b*) A similar scheme for a *p*-type semiconductor. (*c*) A two-photon system with *n*- and *p*-type semiconductors in contact.

ultraviolet light on the primitive Earth than now, availability would not be the only problem for a high-energy photosynthesis. Ultraviolet photons are very destructive – it would have been difficult for a well organised organic biochemistry to have evolved in organisms that had to live in the ultraviolet sunlight. One can imagine ways round this problem – say by having only part of the organism exposed to the light – but there is no doubt that modern plants enjoy a double advantage when they use the softer photons of visible light. They have the keys to unlock an abundant source of chemical energy which, without those keys, is comparatively harmless. That seems the ideal arrangement.

The plant's technique is to combine the effects of two or more lower energy photons. Similar synthetic systems are being considered (Balzani *et al.*, 1975; Bolton, 1978; Bard, 1980). One of the ideas is to use an *n*-type and a *p*-type cell back to back – to have two illuminated semiconductor electrodes each giving an uphill push to the electrons. Bard (1979) has discussed a 'dual *n*-type semiconductor model' and compares this to plant photosynthesis. Granick's hypothetical photosynthetic magnetite particle was a similar idea similarly comparable to modern photosynthesis in its two-stage operation (see figure 8.13). Bard also considers some much simpler particulate semiconductor systems that are also possible in principle. Examples are given in figure 8.14. The great recurring practical difficulty with such schemes is in the prevention of back-reactions – the



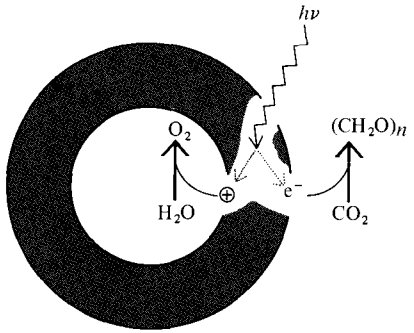



Figure 8.15. A minimal view of the reduction of $CO_2$ in plant photosynthesis. Photons create electrons and positive holes within the membranes of closed vesicles (thylakoids). The electrons eventually reduce $CO_2$ to carbohydrate, while the holes oxidise water.

oxidised and reduced species that are being made always tend to recombine. Let us see how modern plants overcome such problems.

### *The thylakoid membrane*

The specialised organelles of plant photosynthesis, the chloroplasts, contain numerous membrane-bound vesicles called thylakoids. The thylakoids are the engines of photosynthesis and the thylakoid membranes the main sites of activity. Here photons generate electrons and positive holes. These are separated, the reducing effect of the electrons and the oxidising effect of the holes being directed to opposite sides of the membranes. The thylakoid membranes are highly structured and necessarily asymmetric – and they serve as barriers against back-reactions. A first very simplified view of photosynthesis is given in figure 8.15.

A photon is caught in the first place by an array of a few hundred chlorophyll and carotenoid molecules (figure 8.16*a* and *b*). These are located in the thylakoid membrane, and act as an antenna. The light energy is rapidly passed between the molecules by successive re-emissions and absorptions until it arrives at a special chlorophyll molecule attached to a protein, where an excited electron and positive hole are created. In $O_2$-producing photosynthesisers there are two classes of such reaction centres, designated I and II, with somewhat different arrays of light-catching pigments.

The following is an account of what happens next in terms of Mitchell's

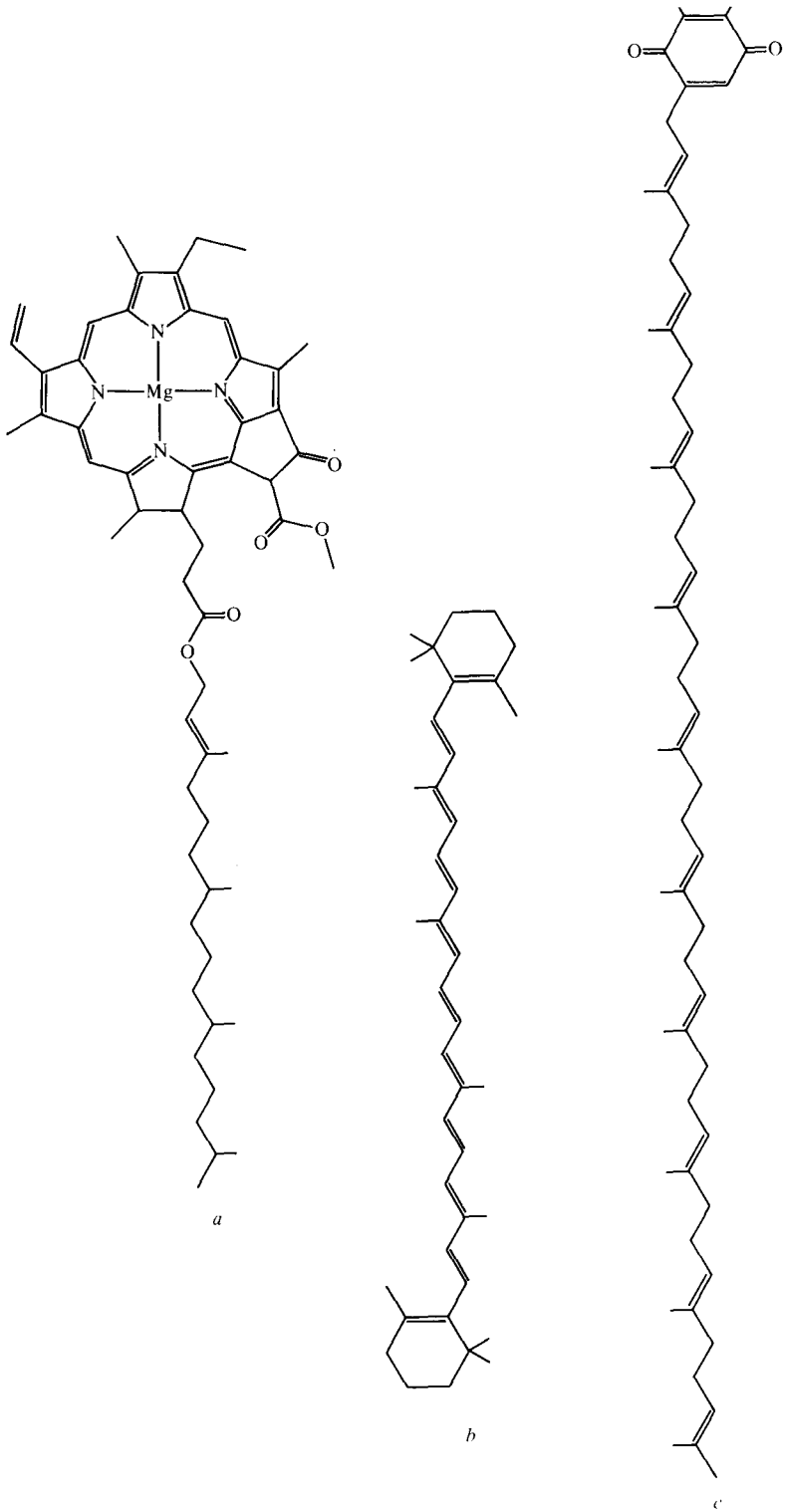


Figure 8.16. Three molecular types involved in photosynthesis. (*a*) Chlorophyll *a*; (*b*) β-carotene; (*c*) plastoquinone.



now generally accepted chemiosmotic hypothesis. (See Hinkle & McCarty, 1978; and Miller, 1979, for more detailed accounts.)

Positive charges that are generated at a type-II reaction centre oxidise water and thus appear on protons – inside the thylakoid:

$$4\oplus + 2H_2O \rightarrow 4H^+ + O_2$$

The high-energy electrons are taken through the thylakoid membrane to its outer edge, and react there with a hydrogen carrier, plastoquinone (figure 8.16*c*):

$$4e^- + 4H^+ + 2PQ \rightarrow 2PQH_2$$

picking up protons from outside the thylakoid. Plastoquinone then moves through the thylakoid membrane to the inner surface to give up its reducing power now to an electron carrier, the iron-containing protein cytochrome *f*, depositing protons inside the thylakoid:

$$2PQH_2 + 4 \text{ cytochrome } f \text{ (oxid.)} \rightarrow 4 \text{ cytochrome } f \text{ (red.)} + 2PQ + 4H^+$$

The electrons have by now lost much of their original energy and pass via the copper-containing protein plastocyanin to fill the positive holes of a type-I reaction centre, allowing it to keep up a supply of the somewhat higher energy electrons generated by light energy in this type of reaction centre. These newly excited electrons run through a 'wire' of two further iron-containing proteins in the membrane – iron sulphur protein and ferredoxin – arriving at the outer surface where they reduce the hydrogen carrier FAD (for structure see figure 9.1):

$$4H^+ + 4e^- + 2FAD \rightarrow 2FADH_2$$

taking up protons from the outside to do this, although two protons are put back in the final transfer of reducing power to the hydride carrier NADP (for structure see figure 9.1):

$$2NADP^+ + 2FADH_2 \rightarrow 2NADPH + FAD + 2H^+$$

Hence light energy from several photons is used to make a molecule of a reducing agent. It seems a long haul. Why so many transfers? Partly, no doubt, to help to prevent back-reactions by physically separating oxidised and reduced species. But the chemiosmotic hypothesis gives also a more subtle explanation: the picking up and dropping of protons concomitant with the changing back and forth between electron and hydrogen carriers, and the way in which these carriers are arranged in the membrane, allows protons to be picked up mainly on one side of the thylakoid membrane and deposited on the other. Much of the energy of the excited electrons thus ends up as a proton potential difference across this membrane. ATP is then

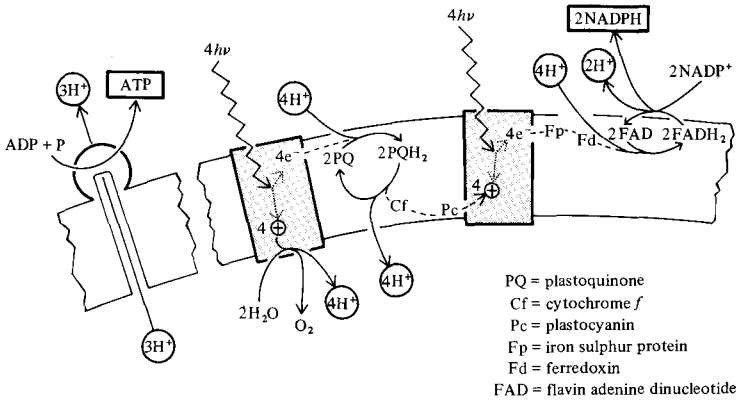



Figure 8.17. Production of NADPH and ATP by thylakoids. The membrane of the thylakoid consists of a lipid bilayer within which many proteins and other species are located in a definite way. The shaded boxes represent pigment systems II and I that are connected in series. Electrons are pumped outward, eventually making NADPH, while protons are pumped inwards. Backflow of protons makes ATP at separate sites in the thylakoid membrane.

made in separate locations by allowing protons to flow back down this potential gradient. (See figure 8.17.) The thylakoid thus makes two high-energy reagents – NADPH and ATP – which elsewhere in the chloroplasts are used in making carbohydrate from $CO_2$. The production and use of these reagents in photosynthesis are summarised in figures 8.17 and 8.18.

In discussing the primary quantum conversion apparatus of present day photosynthesis, Calvin (1974) remarked that it 'in some ways simulates a solid state device in which impinging light separates charge – positive and negative, electrons and holes'. This has been a recurring theme: Granick (1957, 1965); Krasnovsky (1974); Bard (1979, 1980), have made similar comments and each of these has made a suggestion along the lines that semiconductor minerals might have provided a primitive stand-in for the antenna–reaction centre combination of the modern photosynthetic apparatus. The important role of iron in the electron transport function of the modern system has also been remarked on as possibly having some very early evolutionary significance.

The barrier membrane is clearly a crucial component of those energy transducers that depend on chemiosmotic effects. Even chlorophyll is a less important part of the modern technique of photosynthesis. The membranes of the Halobacteria transduce energy very effectively with a quite different pigment molecule (Stoeckenius, 1976). But, like thylakoids, they depend absolutely for their action on being closed vesicles bound by an



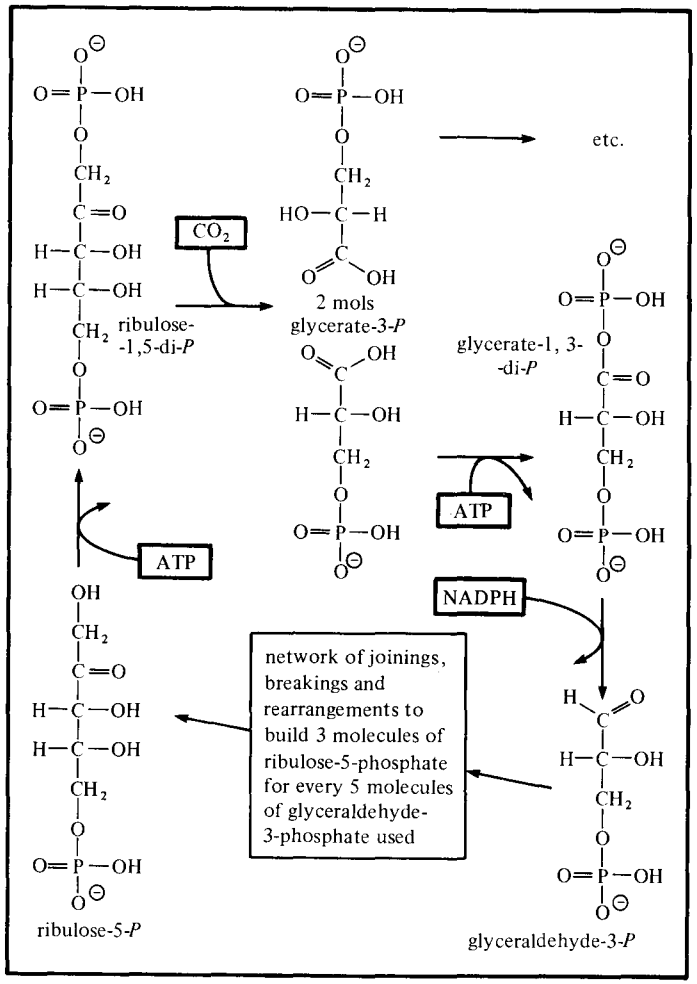
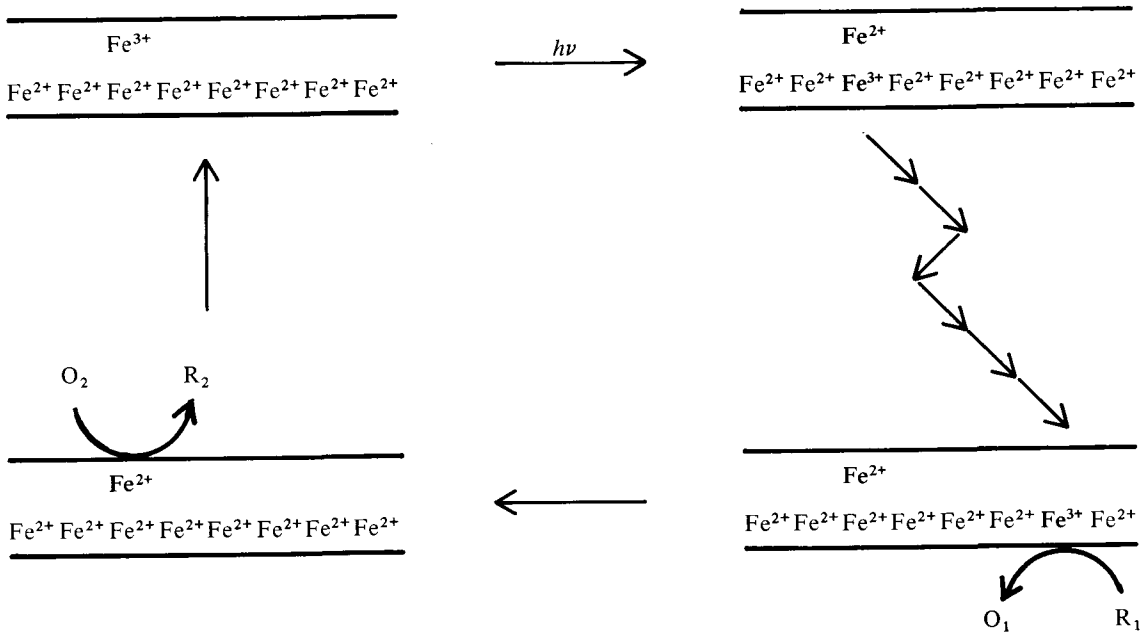



Figure 8.18. Use of NADPH and ATP in the production of carbohydrate from $CO_2$ (the Calvin cycle).

asymmetric membrane across which develops a proton potential. Mitochondria are similar, although they use fuels rather than photons to generate the potential that makes the ATP. *E. coli* too, that supposedly humble creature, has in its cell membrane fuel-driven proton pumps – that push protons out – and a marvellous collection of machines driven by a backward proton flow. These include transport proteins that actively pull in molecules such as particular sugars and particular amino acids, as well as ion pumps. Even *E. coli*'s flagella are powered by the proton flux (see Hinkle & McCarty, 1978).



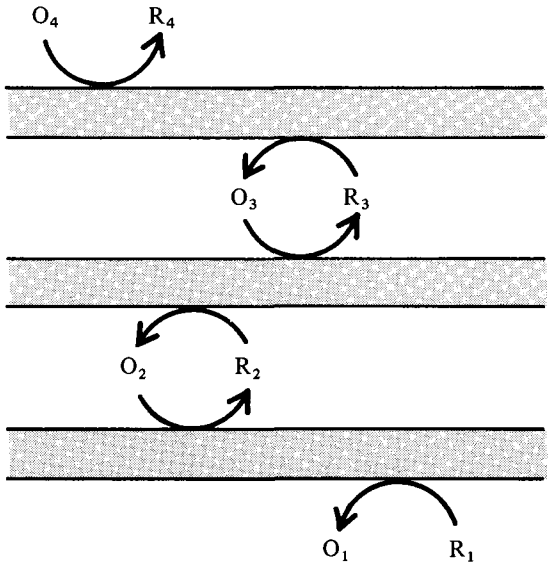




Figure 8.19. (*a*) A hypothetical photoredox machine based on a 1:1 layer silicate which has a few tetrahedral $Fe^{3+}$ ions and many octahedral $Fe^{2+}$ ions. After a light-induced charge transfer, an $Fe^{3+}$ 'hole' migrates among the octahedral sites until it brings about an oxidation. The unstable $Fe^{2+}$ is a reducing agent. (*b*) A more powerful multi-photon device with photomembranes and redox pairs at different redox levels. $R_4$ is a powerful reducing agent and $O_1$ a powerful oxidising agent.

*In place of the thylakoid membrane*

If mineral semiconductors seem likely elements of the very first photosynthetic apparatus, mineral membranes seem at least as likely. I have argued already for an important role for membranous materials for very early organisms; but of all places it is in the design of primitive energy transducers, and most particularly for phototransducers where excited photoproducts must be quickly chemically separated, that membranes – asymmetric membranes – are most called for.

The modern thylakoid membrane is decidedly a post-protein invention. There have to be proteins to structure the membrane, to lay out the charge conduits, to hold the right metal ions in the right places, and to maintain the essential asymmetry. Without proteins how could all this be arranged in an organic membrane? How could genetic information have set it up?

The case for clay membranes for the first photosynthetic apparatus depends partly on the whole idea that, anyway, first genes and first organic-synthetic apparatus would have been best made from clays: but it depends especially on the naturally membranous character of so many clays, whose structures are often necessarily asymmetric. Such mineral membranes could hold transition metal cations to catch light and conduct charge – as well as acting as inert barriers to separate photoproducts.

One might think particularly of clay minerals such as smectite, halloysite, imogolite or allophane, which often form membranes, tubes and vesicles. Iron atoms are common in clays, and these might have served both as light catchers and as charge conduits. Nontronite with $Fe^{3+}$ in it – a yellow smectite clay – develops a dark blue-green colour when the iron is partly reduced chemically. This is attributed to charge-transfer transitions which then become possible between adjacent iron atoms in the lattice (Anderson & Stucki, 1979), and it indicates that, at least in the sunlight, charges could move along rows of iron atoms embedded in the unit layers of smectite, or any suitably 'doped' 2:1 layer clay.

Even 2:1 silicate layers are sometimes asymmetric. There can be a difference in the charges associated with each of the tetrahedral sheets – as in rectorite (figure 8.5). A 1:1 unit layer is asymmetric anyway and I have used such a layer in the speculative photoredox machines illustrated in figure 8.19. It is worth noting here that structural $Fe^{3+}$ in 2:1 layer silicates can be a particularly effective oxidising agent both for benzidine (McBride, 1979) and for adsorbed $Fe^{2+}$ (Gerstl & Banin, 1980).

The stability of the oxidation state of a transition metal atom is easily affected by the field of ligands around it (Nyholm & Tobe, 1963). This is in line with the observations of Yershova *et al.* (1976) on differently oxidisable iron atoms in berthierine (page 215). So we can take it that the redox potential of a given iron atom in a clay layer would be influenced by the pattern of charges and lattice strains generated by the cations in its immediate surroundings. In any case, different transition metals could provide alternative redox levels for the membranes in a device such as that shown in figure 8.19*a*. A more powerful multi-photon redox machine might then be envisaged, as shown in figure 8.19*b*.



*Making use of phosphate*

It has often been suggested (e.g. by Lipmann, 1965) that simple condensed phosphates such as pyrophosphate, triphosphate or polyphosphate might have been the forerunners of ATP and other energy carriers. Various possible primitive environmental sources of condensed phosphates have been considered (Griffith, Ponnamperuma & Gabel, 1977; Ponnamperuma, 1978), but it must be said that condensed phosphates are unusual minerals on the Earth today and that even monophosphate is generally present in oceans and freshwaters in concentrations of only around $10^{-7}$ M (Halmann, 1974).

Schwartz (1971) has suggested an ingenious way of overcoming the problem of low environmental phosphate concentrations. The basic problem is the ubiquity of calcium in the environment and the insolubility of the main phosphate source, apatite ($Ca_{10}(PO_4)_6(F,Cl,OH)_2$). Noting that many biochemical molecules, particularly the di- and tricarboxylic acids of the Krebs cycle, are excellent complexing agents for calcium – and that indeed a widely used analytical method for rock phosphate depends on the solubilisation of apatite with citrate – Schwartz suggests that such organic molecules might have served to increase local phosphate concentrations by removing calcium from solution. Although not now a 'biochemical', oxalate would have been particularly effective because of the great insolubility of calcium oxalate.

Here we are trying to avoid using probiotic sources of organic molecules, but we might surmise that if phosphate was important for early organisms, then so might calcium-complexing agents – and that this provided selection pressures for the first synthesis of Krebs acids.

But was phosphate important for very early organisms? We might note first that nowadays phosphorus is an inconvenient element for organisms – unlike other major biological non-metals it cannot be cycled in volatile form through the atmosphere and it often becomes the limiting nutrient

(Emsley, 1977). Of course we can now see many reasons why our biochemistry is committed to phosphorus; but when did it become so? In view of the central part played by phosphorus in the energy transactions of present-day life, and the initiating role, according to our story, of energy-transducing mechanisms for the organic part of early biochemistry, one might suppose that phosphate, and the ability to make condensed phosphate, came in early.

A second point to note is that although concentrations of phosphate are generally low in bulk environmental waters, phosphate is adsorbed by clays, and may be present in interstitial waters of the sediments of lakes, estuaries and oceans at concentrations that are three orders of magnitude higher (Halmann, 1974).

*Were polyphosphates among the first 'organic' biochemicals?* A third point is that phosphates, and particularly polyphosphates, might have been useful for an early clay-based life – as a means of preventing sideways growth of genetic crystals, or as metal-complexing agents or as strings to hold clay platelets together. And phosphate itself would have been a much cleaner starting material for making simple water-soluble polymers than abiotic organic molecules, such as sugars, which if they had been there at all would have been present in a tar of some sort. And at least to make phosphate polymers you do not need to inject energy to make the monomers. It is only in condensing the phosphate units together that you begin a kind of 'organic chemistry', and build persistent metastable molecular structures. Perhaps, then, pyrophosphate provided one of the first hinges between the crystal-inorganic and the molecular-organic way of life: perhaps it was among the very first 'organic' biochemicals.

But is it plausible that wholly inorganic organisms could have driven the formation of phosphoanhydride bonds using some available energy source? Could one imagine such machinery being made from, say, clay?

In thinking about this question we may note first that the amount of energy needed to form phosphoanhydride bonds is not exceptional. As discussed by Mahler & Cordes (1971) pyrophosphate and ATP are 'energy-rich' – but then so are many other molecules. For example, at pH 7 and 25 °C the standard free energies of hydrolysis of pyrophosphate and ATP (to ADP) are about 34 and 31 kJ $mole^{-1}$ respectively. This is rather less than for the active ester *p*-nitrophenyl acetate (55 kJ $mole^{-1}$), about the same as for acetyl-coenzyme A (32 kJ $mole^{-1}$), and not so very different from ethyl acetate (20 kJ $mole^{-1}$).

Nor is the origin of pyrophosphate energy particularly mysterious even if it has turned out to be difficult to pin-point. Among the contributing factors that have been suggested are restricted delocalisation of electrons in condensed phosphates and differences in solvation between simple and condensed phosphates. Such factors could be greatly affected by binding on a mineral surface. For example, binding on the edges of a layer silicate with exposed positive charges (figure 7.9) would tend to localise the electrons on monophosphate as well as pyrophosphate, and the solvation energy difference would also be expected to change. Hence when bound to a suitable mineral surface the condensed phosphate might very well be stabilised.



Such an effect would not in itself be sufficient to provide a means of altering the position of the equilibrium between monophosphates and condensed phosphates in solution: it could not have provided a means of production of pyrophosphate – that would be getting energy for nothing (compare figure 8.21*a*). But for a surface in the sunlight one could imagine a number of ways in which photon energy could be used to alter the position of equilibrium favourably. For example, photons might in effect knock the pyrophosphate groups off the surface through photo-induced charge transfers in the mineral, temporarily reducing the edge charges.

*Energy-coupling devices* Given a source of condensed phosphates the energy of their hydrolysis might be coupled to synthetic reactions that would normally be energetically unfavourable. For example Rabinowitz, Flores, Krebsbach & Rogers (1969), starting with a solution of glycine and 0.1 M sodium trimetaphosphate, made diglycine in 31% yield after 359 h at room temperature and pH of between 7 and 8. A mechanism of the kind suggested for this homogeneous condensation is shown in figure 8.20: here the condensed phosphate is acting as a coupling agent. The simpler diphosphate, sodium pyrophosphate, was much less effective (0.7% yield) although Burton & Neuman (1971) have used this reagent successfully in the presence of apatite crystals. They suggest that the reaction takes place in grooves on the crystal surfaces.

Heterogeneous reactions for using the energy of condensed phosphates would seem to offer the best prospects for control, particularly if the solid surface in question is a clay whose detailed structure has been engineered by natural selection. As remarked earlier, coupling agents are not very selective in themselves – the selectivity has to come from somewhere else. One might imagine a number of coupling mechanisms via coadsorption.

The three-stage cycle shown in figure 8.21*a* would *not* work, because although it is quite feasible that the equilibrium for the reaction should be

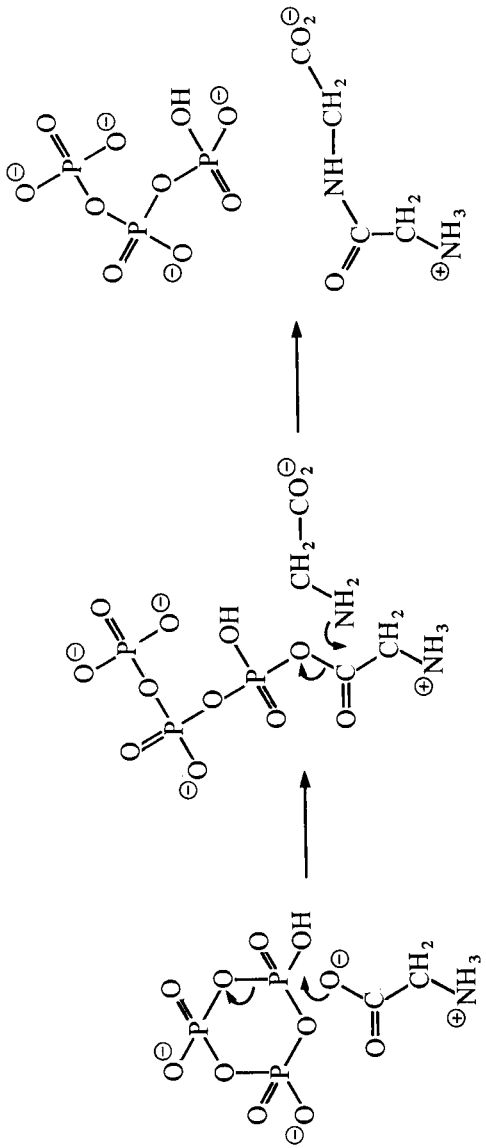


Figure 8.20. Simplified mechanism for a condensed phosphate acting as a coupling agent for amino acids (cf. figure 1.10).



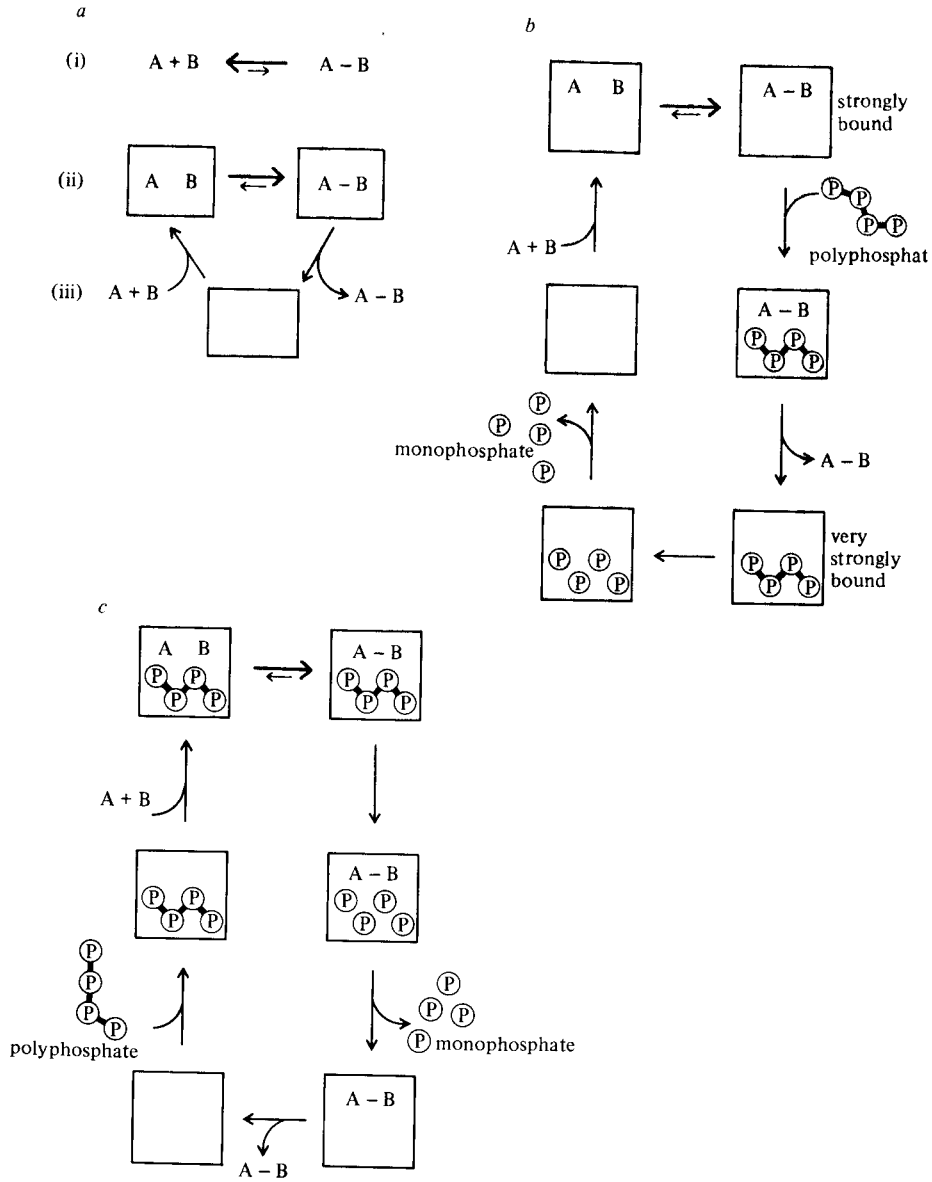



Figure 8.21. (*a*) If (i) represents the position of an equilibrium between species in solution and (ii) the situation when these species are adsorbed on a surface, then the cycle shown in (ii) and (iii) could not operate spontaneously. (*b*) Elaboration of the cycle in (*a*) in which condensation of A and B is coupled with hydrolysis of polyphosphate. Such a cycle could in principle operate spontaneously. (*c*) Similar to (*b*) except that A–B is bound more strongly, not less strongly, when polyphosphate is bound nearby.

shifted in favour of synthesis when the reacting species are adsorbed, it would be necessary in that case that the product should be more strongly bound than the starting materials. (To keep the cycle going you would have to supply energy to remove A–B from the mineral surface – otherwise you could make a fuel for nothing.) In the scheme shown in figure 8.21*b* the product A–B is, in effect, displaced by polyphosphate which must now be at least as strongly bound. So energy is still needed to get the polyphosphate off – and is provided in this scheme by the hydrolysis of the polyphosphate *in situ* to monomers which are less strongly adsorbed.

Notice that it would not be necessary that in displacing the synthetic product the polyphosphate should compete for the same site; all that is needed is that when polyphosphate binds, the binding of the product is weakened. Such an effect might well operate *through* a clay membrane. Suppose, for example, that the binding of some (anionic) product was at an anomalously positively charged region, resulting from local high-charge octahedral cations. Then adsorption on the opposite side of the clay membrane by a phosphate polyanion might very well serve to displace the synthetic product. In that case, reagent and reactant would never need to mix and 'work-up' would be that much easier.

Figure 8.21*c* is a similar scheme more suited to cationic reaction products – I will leave the reader to see how the energy of polyphosphate hydrolysis is coupled to the synthesis in this case.

Figure 8.22 represents conceivable forms of energy coupling that again might operate through a clay membrane. This scheme depends on the idea that the redox potential of a structural transition metal cation would generally be altered by the proximate binding of (multiply negatively charged) polyphosphate. This might be a direct electrostatic effect, or it might operate through 'allosteric' distortions of the geometrical arrangement of the ligand atoms around the cation. Since R and O would be a metastable pair it might be as well if they were generated on opposite sides of the coupling membrane, with perhaps also the charge on $M^+$ causing a local 'allosteric' adjustment to the crystal structure that made $M^+$ less accessible from the upper surface.

The machine in figure 8.22 is a sort of electron pump, and one might try to imagine other kinds of pumps similarly driven by polyphosphate hydrolysis. One might think of an asymmetric clay membrane with pores – clusters of octahedral vacancies – that open when polyphosphate binds . . .

Of course all these designs that we have been considering are highly speculative – the more so as we do not even know in detail how the modern equivalents work. It is too easy, you may say, to make such paper schemes.



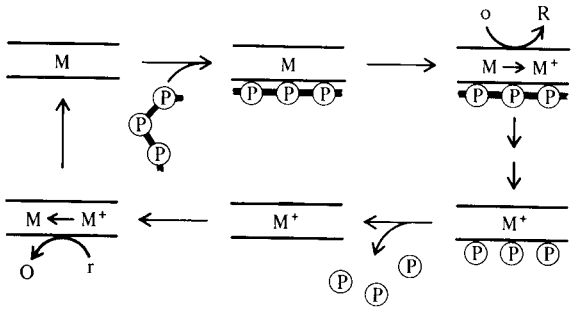



Figure 8.22. M represents a structural transition metal cation in its normal valency state in a crystal – for example an octahedral $Fe^{2+}$ in a trioctahedral layer silicate. M becomes a reducing agent for o when (highly negatively charged) polyphosphate binds. (M 'wants' to become more positively charged now.) Polyphosphate, which is now very strongly bound, must fall apart to come off again (as in figure 8.21*b* and *c*): when it does so, $M^+$ is left as an oxidising agent for r. Hence the reaction $o + r \rightleftharpoons O + R$ is driven to the right against the thermodynamic gradient at the expense of condensed phosphate bonds.

Yet really that is my point. In clays, only somewhat contrived clays, you have the elements for many kinds of energy-transducing machines that would otherwise be difficult to come by.

### Evolving biochemical pathways

Like any material with an engineering purpose, clay minerals would have had their strengths and weaknesses as materials of construction for organic synthesisers. As well as for making energy-transducing machinery, the clay membrane might have been particularly suitable for feedback control devices – depending again on the 'allosteric' properties of silicate structures. Their ability to create interconnecting compartments might have been particularly useful too in making chemical oscillators (for example as in figure 8.23) for time sequencing devices.

Following from earlier discussions in this chapter, the main weak point for clays would have been in molecular recognition. This would have been much less effective before there was protein to provide well engineered sockets. To mitigate this lack of discrimination I would suggest three rules for organisms embarking on organic synthesis.

The first rule would be to use a simple feedstock, not a complicated soup of molecules. The ideal environment would be one that had only a few

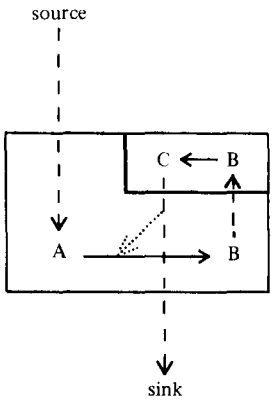



Figure 8.23. A chemical oscillator. Reaction A → B is self-inhibiting, indirectly, via C which is generated in a separate compartment. Because of the delays introduced by diffusion into and out of the subcompartment, the concentrations of A, B, and C do not reach a steady state, but oscillate instead. (Solid arrows, chemical reactions; dashed arrows, diffusion or other form of translocation; dotted arrow, inhibitory control.)

things to offer as potential organic feedstock – perhaps just water, carbon dioxide, nitrogen, phosphate and sulphide. There would be much less recognising to do in an environment like that.

The second rule would be to be tidy, to prevent as far as possible different types of molecule that were being made from getting arbitrarily mixed up with each other. As far as possible everything should be kept in its proper place so that a molecule would be recognised by where it was. (That is, after all, the chemist's day-to-day way of recognising compounds: he knows that he has, say, acetic anhydride, partly because it looks like acetic anhydride and smells like it; but mainly because it came out of the acetic anhydride bottle.)

The third rule would be to be unambitious to begin with – to specialise in the production of just one kind of molecule or of some limited set. After all, in starting the adventure into organic chemistry, inorganic organisms would have had one huge advantage: they did not really need organic molecules at all. The first organisms that made those first simple organic products would have done so under selection pressures that added somewhat to their viability – that increased, perhaps, the range of the environment that could be occupied. But whatever they were for, these organic products must have been 'optional extras' to begin with.



*Two modes of metabolic evolution*

*1. Symbiosis* A favourable site for the start of organic biochemistry would have been at some junction between the ancient elements: where there was not only Earth and Water to provide the main materials, but also Air, and the Fire of sunlight, to provide materials and energy for the new evolutionary adventures.

For this part of our story we might imagine the kind of site that Darwin (1871) imagined for the origin of life: 'some warm little pond with all sorts of ammonia and phosphate salts, light, heat, electricity etc.' We might translate Darwin's image to a later stage of evolution and think of Darwin's pond as an ecological system consisting of a community of well evolved clay-based organisms living in shallow waters exposed to the sunlight.

In this pond there are various species that have recently evolved photosynthetic competences – independently and for different reasons. They are strictly speaking plants: they are organisms using sunlight to make metastable molecules.

One species is a *phosphate-condensing plant* of the kind that we discussed earlier. Polyphosphates are a relatively simple class of linear or cyclic polymer that would have profoundly modified the immediate environment of clay particles that were making them (see page 309). Such control could have been all the more effective since polyphosphates are relatively unstable polymers. This is an important general point. For a biological control molecule to be effective it is usually necessary that the molecule is not only being made, but also being destroyed or otherwise disposed of. One might think of messenger RNA molecules, or neurotransmitters, or hormones – it is no good if such molecules simply go on accumulating. In terms of our primitive example, the consistency of a clay–polyphosphate assemblage would be controllable by controlling the relative rates of synthesis and hydrolysis of the polyphosphate molecules.

We might also suppose that there were *carbon dioxide reducing plants* in Darwin's pond – using the 'electricity' that Darwin saw as part of the story. Not lightning, but those photovoltaic and photoelectrochemical effects which must happen when the sun shines on any assemblage of clay particles. Usually the energy from the potentials that are generated will be quickly converted to heat, in a chaos of tiny currents, before there is any lasting chemical effect. But that is because usually the photoelectrochemical cells in an ordinary patch of mud are arbitrarily arranged with electrical effects cancelling out and any redox products quickly recombining. But

here we are talking about clays designed by natural selection to make the kinds of machines that were discussed in the last section – with localised but powerful oxidising and reducing sites and the means to prevent back reactions.

One can think of a number of modes of control open to a clay organism that reduced carbon dioxide to simple organic molecules. In so far as there were changes in acidic and basic species present in solution there would be changes in pH – and this might have provided a means of optimising pH for clay synthesis even if the organic materials had no other catalytic use. Or perhaps the carbon dioxide was simply a sink for reducing power. In this way, perhaps, $Fe^{2+}$ in solution could be converted to $Fe^{3+}$ so that a ferric oxide-hydroxide membrane could be made, for example:

$$\tfrac{1}{2}CO_2 + H^+ + e^- \rightarrow \tfrac{1}{2}HCO_2H$$
$$Fe^{2+} \rightarrow Fe^{3+} + e^-$$
$$2H_2O + Fe^{3+} \rightarrow FeO(OH) + 3H^+$$

*Nitrogen-reducing plants* might have accounted for that ammonia in Darwin's pond – perhaps through the photoreaction found by Schrauzer & Guth (1977), if titanium dioxide had been among the minerals in some patch of vital mud (page 333).

One can see simple ways in which the ability to control the production of ammonia might have been useful: again for pH control or as a sink for reducing power; but also to make a new kind of interlayer cation to replace $Na^+$, etc., and so adjust the cohesion between clay layers. Here would be a new kind of exchangeable cation that could be picked out of the air.

Then again there might have been various *sulphur plants* driving redox reactions one way or the other between sulphide, disulphide and higher oxidation levels.

Control of the oxidation level of sulphur in solution is part of the means of controlling the crystal forms of iron sulphides (Rickard, 1969 – see page 178). This might have provided sulphur plants with part of their machinery for photosynthesis: Hall, Cammack & Rao (1974) have suggested that mineral iron sulphides were precursors of ferredoxins, comparatively simple proteins whose role in modern photosynthesis has already been referred to. A ferredoxin molecule can be thought of as a minute 'crystal' of a few iron and sulphur atoms (often four of each) that is stabilised by being held in a folded polypeptide frame.

Different photosynthetic organisms living in the same pond could hardly help interacting with each other – if only through kinetic effects on pH and on $E_h$ (oxidation–reduction level). In addition, each would be a source of



more or less long-lived metastable molecules. Communities of organisms would tend to persist if the effects of the products of the members were on the whole mutually beneficial.

For example, a carbon dioxide reducing plant that was making a smudge of Krebs-type carboxylic acids might help to dissolve apatite in the environment (page 343) – to the benefit of phosphate plants, perhaps – while the phosphate plants' products would be sources of energy for organic synthesis that could be used in the dark by yet other organisms that were deeper in the mud at the bottom of the pond. Similarly, sulphur plants might inadvertently be exporting redox energy in soluble form or as a metastable colloidal iron sulphide.

In so far as two or more organisms were mutually beneficial there would be advantages if they were held together in some more positive way than simply sharing a small pond. More particularly, for more open waters we might imagine small heterogeneous colonies enclosed by a mutually contrived barrier of some sort. These organisms would now have their own tiny 'pond', much more sensitively contrivable to the mutual advantage of the organisms in it.

For a suitable barrier material we might think of minerals that readily precipitate from solution as a result of pH and/or $E_h$ shifts. Such are the materials of chemical sediments: for example silica and metal carbonates, sulphides, oxides and oxide-hydroxides (Krumbein & Garrels, 1952). We might see such materials being deposited at the interface between a general solution, containing appropriate metal and other ions, and the solution entrained by a floccule of actively photosynthesising organisms which through their activities were shifting their localised pH and/or $E_h$ (figure 8.24*a* and *b*).

Such a barrier would tend to remain permeable to small molecules and ions – its deposition would depend directly on some of the species being able to diffuse through. Also the pH and $E_h$ gradients would only be maintained so long as photosynthesis, and clay crystallisation, etc., was continuing within the colony: if the nutrient flow was cut down too much the barrier would start to redissolve – and continue to do so until it was thin enough to let in sufficient nutrients.

From these considerations we might predict that such a barrier would grow to the point at which it just let in the essential nutrients. In that case it might very well not let out larger photosynthetic products such as polyphosphates or the larger organic molecules. A positive osmotic pressure would then develop inside to provide a driving force for growth and reproduction in the manner of a 'silica garden' (page 327) – see figure 8.24.

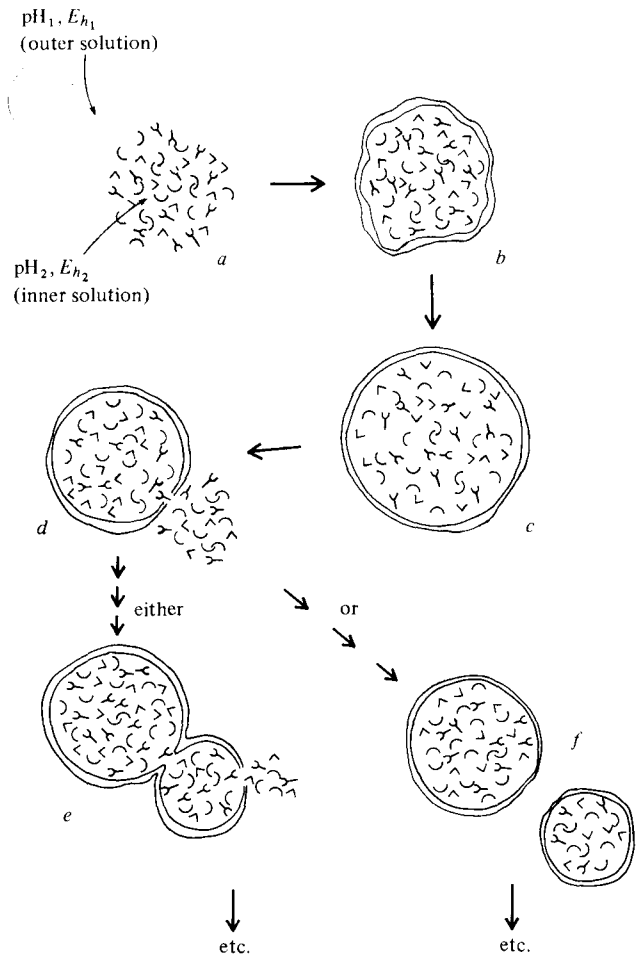



Figure 8.24. (*a*) A group of inorganic plants create a localised pH and $E_h$ different from that in the surroundings. (*b*) This precipitates ions present in the outer solution and makes an interfacial barrier. (*c*) Photoproducts generate an internal osmotic pressure which (*d*) bursts the wall, leading either to (*e*) multicompartments, or to (*f*) 'cell division' (or both).

Not only would the potential for more elaborate organic syntheses be improved now that separately evolved subsystems were in closer collaboration, but this near-cellular form of life would also produce new reasons for making metastable molecules. There would be jobs to be done in controlling the osmotic pressure; in suitably structuring the internal solutions (a gel consistency might be a good idea for the explosive budding stage of



figure 8.24*d*); in lining the inside of the barrier to adjust its permeability and pliability; or to nucleate the kinds of crystal growth in the barrier membrane that would optimise such properties.

*2. Forward extension of pathways* Another form of metabolic elaboration would have been from within, through a kind of exploratory process based on metabolites whose production had already been perfected. This would have been an earlier enactment of a general evolutionary process that is still visible to us. I am thinking of the way in which secondary metabolic routes are built up in modern organisms. Whole classes of secondary plant products are made by joining and modifying very limited sets of standard molecules – the morphine alkaloids are derived from tyrosine for example. For a primitive organism, 'experiments' based on a pre-existing metabolite would usually have to be performed in a separate compartment to prevent the various products and intermediates of the new reactions from becoming embroiled with the pre-existing reactions. So the outcome from a wholly internal origin of a metabolic path might be a set of suborganelles connected together. Such a production line would have to include purification stages, in keeping with the general rule for multi-step organic syntheses that at least some of the intermediates must be purified to stem the slide into tar.

I am suggesting, then, that a metabolic pathway would be built up *forwards* through a series of experimental episodes with each episode being more or less completed before the next was embarked on. The very first episode in a series would start from a simple feedstock from the environment, but each episode in the series would begin with the making of some new crude product through a mutation in the clay-held information. Perhaps a new catalytic site was generated; or a new interconnection made between compartments that brought materials together that had not previously been mixed. The next long stage would be a gradual refinement of the new product through natural selection operating on its means of production: specialist organelles were perfected which contained suitable sub-compartments, catalytic sites, adsorption sites, pumps, etc. suitably connected up. The original smudge of reactions was thus gradually transformed into just one reaction or one short clean sequence of reactions. This general mode of evolution of a biochemical pathway is illustrated in figure 8.25.

In part this progression would have been through improved design of apparatus for controlling the conditions for the reaction – improved pH control, catalyst specificity, and so on – but partly too it would have been

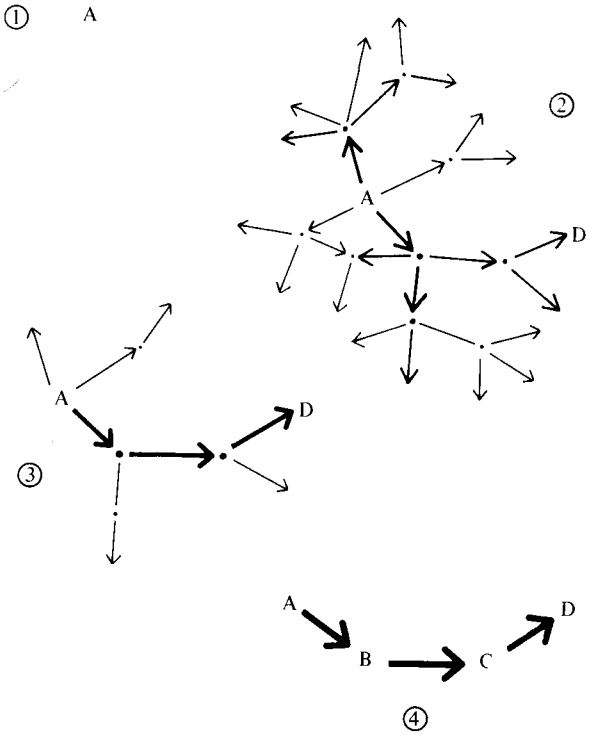



Figure 8.25. An episode in the forward evolution of a metabolic pathway. 1. Some established metabolite, A, is induced to react in a variety of ways – for example because of the introduction of a new catalytic site, or because it becomes mixed with some high-energy reagent. The result is a semi-chaotic mixture of products. 2 and 3. Selection pressures favouring mixtures rich in one of the products (D) improve the conditions for reactions leading to that product. Other reactions tend to be suppressed. This mechanism is only plausible for short sequences, but the intermediates B and C would need to have no function in themselves. Longer (straight or branched) pathways would become possible through further episodes starting with D, B, or C.

through improvements in purification procedures (work-up). Yields would fall short of 100% so that in addition to necessary waste products there would be other unwanted materials to be dealt with. Much of this would be insoluble tar. For organisms that were actively synthesising clays, intractable wastes might be dealt with by being enclosed in clay membranes and eventually dumped. Or, looking again at figure 8.24, suppose that intractable materials tended to stick to the outer wall: they would then be left out of the newly budding 'cells'.



*Fixing carbon dioxide*

A carbon dioxide molecule is only properly fixed when its carbon atom has been joined to another carbon atom. Just to make formic acid or formaldehyde or methanol is not quite good enough, because the next step of putting two such things together can still be difficult.

*Maintaining $C_1$ receptors* Modern organisms go out of their way to avoid joining two $C_1$ units. In the fixation of carbon dioxide in plant photosynthesis, for example, the $C_1$ unit, carbon dioxide, joins to a $C_5$ sugar – and the Calvin cycle can be seen as a device for keeping up the supply of the appropriate receptor molecule (figure 8.18).

The reductive carboxylic acid cycles of bacterial photosynthesis similarly regenerate carbon dioxide receptors which have more than one carbon atom in them (Buchanan, 1972). The short reductive carboxylic acid cycle, a reversal of the Krebs cycle, is shown in figure 8.26.

Interestingly, the mechanism of formaldehyde uptake in the formose reaction involves the same idea: once the formose reaction has started there is a network of reactions (pages 28–31) which continuously regenerate sugars that can fix formaldehyde. The Breslow cycle (figure 1.7) is thus analogous to the carbon dioxide fixing cycles in organisms.

*The evolution of biochemical cycles* The formose reaction may provide us with a model of more primitive, more chaotic versions of the carbon-fixing cycles of modern biochemistry. The real formose reaction is far, far more complex than Breslow's minimum mechanism. There are dozens, perhaps even hundreds, of formaldehyde-fixing sugars involved. It is formose as a mixture that fixes formaldehyde to produce more of the mixture. One might surmise that the first carbon dioxide fixing systems evolved from some such smudge of molecules to be refined later as catalysts became more specific. Such an evolutionary progression would be a cyclic version of the mechanism shown in figure 8.25: one or two closed cycles selected from what had originally been a dense network (Cairns-Smith & Walker, 1974).

*Formaldehyde as an early biochemical?* Although we will have doubts about the idea later, let us pursue for the moment the suggestion that the formose reaction might have provided a primitive form of metabolism for clay-based organisms that could photoreduce carbon dioxide down to formaldehyde.

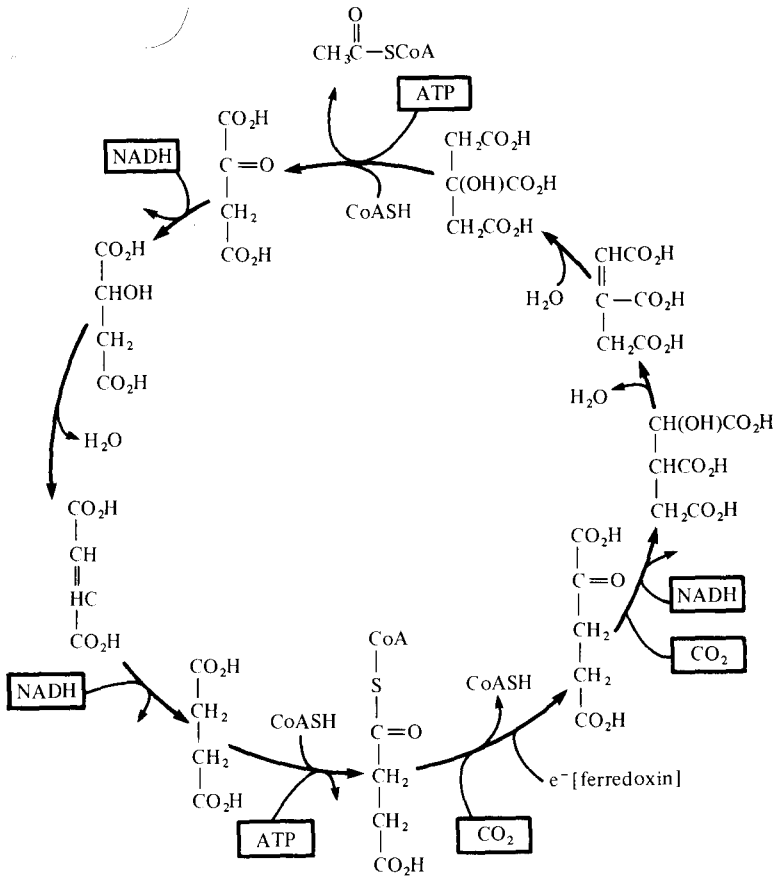



Figure 8.26. The short reductive carboxylic acid cycle of bacterial photosynthesis. This is the Krebs cycle driven in reverse by high-energy photoproducts ATP, NADH and reduced ferredoxin.

There are micro-organisms on the Earth today that can handle formaldehyde and some of these may help us to bridge the gap between a putative formaldehyde-based metabolism and the $CO_2$-based metabolism of modern plants (Quayle & Ferenci, 1978). But formaldehyde is hardly a convenient biochemical now. It is too reactive, especially with amines and phenols: apart from anything else it resinifies protein. But before protein, formaldehyde might have been an easier reaction intermediate. Hence the seemingly curious feature might be explained that a molecule so inimicable to our present biochemistry should be so closely related to sugars – because there was a time when formaldehyde was a key biochemical, the internal precursor of sugars.



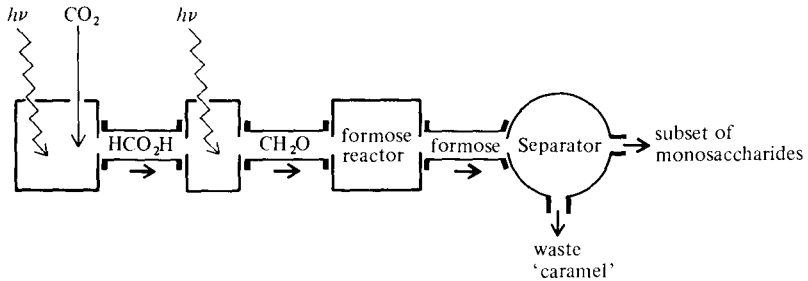



Figure 8.27. Hypothetical organelle for making mixtures of monosaccharides (see text).

One might then imagine the forward evolution of an organelle for making monosaccharides (figure 8.27). First there was a formic acid plant photoreducing carbon dioxide. Some of the product was then further reduced to formaldehyde in an adjoining compartment to give, more or less at once, a sticky caramellising mess – formose. This was useful (as a glue or something) and so formose production became more efficient through the evolution, now, of another attached suborganelle – a new compartment or set of compartments which provided improved conditions for the reaction. In the surroundings of such formose reaction vessels there would be other compartments (typical of any clay assemblage) with flows and counterflows and weak binding sites that had slightly different binding constants for different monosaccharides. As a result there were, from time to time, accidental chromatographic effects with more or less particular sugars accumulating in certain regions. Hence separators began to evolve and with them different kinds of formose mixtures became available suited to somewhat different functions.

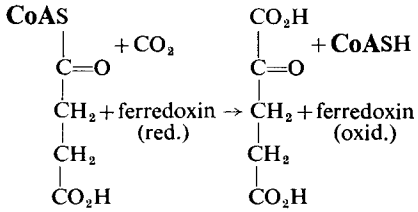
In view of the complexity and instability of formose it is doubtful if such a system could ever have generated particular sugars efficiently, if at all. Yet it might have provided a starting point for the evolution of something better. As a variant of the system in figure 8.27, one might imagine that some carbon dioxide was fixed in the formose reactor to give now somewhat more carboxylated mixtures. These would have to be reduced sooner or later to maintain a $C_1$-fixing capacity – which would require a powerful, presumably photoproduced, reducing agent. But it would be a way of getting rid of formaldehyde – and a step towards our present system (cf. figure 8.18).

*Carboxylic acids first?* Alternatively, the set of carbon dioxide fixing molecules might have been more oxidised than sugars – a mixture

of (mainly) carboxylic acids, a disorganised version of the reductive carboxylic acid cycles of bacterial photosynthesis.

Buchanan (1972) has argued that this bacterial photosynthesis is the more primitive kind, and Hartman (1975a) would see clay organisms as having picked up carbon dioxide with carboxylic acid receptor molecules.

Carboxylations with carbon dioxide are chemically more difficult than hydroxymethylations with formaldehyde. This might be seen as a point against a reversed Krebs cycle for early carbon fixation. Any such cycle, even a semi-chaotic one, would have to be driven. α-Carboxylations are an essential part of carboxylic acid cycles and are particularly difficult. For example, the reaction:




is being driven by the powerful reducing agent, reduced ferredoxin. Also the succinyl group is in a reactive form as a thioester.

But such objections are by no means clear. Referring back to those rules for organisms embarking on organic chemistry, there would be something to be said for reaction types that were not too easy – and could be better controlled for that reason – and for first organic products that belonged to a smallish set of possibilities. And we know that mixtures of carboxylic acids ('fulvic acid') can catalyse clay synthesis (page 251) – so we can see an immediate use. And in any case we are imagining clays that are quite capable of generating powerful reducing sites: α-carboxylations might thus have been quite possible (perhaps via the photogenerated radical anion $\cdot CO_2^-$ – compare Getoff's reaction sequences on page 32). If it is so, as Hall, Cammack & Rao (1974) suggest, that the ancestors of ferredoxins were iron sulphide minerals, then the critical role of ferredoxin in the reductive carboxylic acid cycle could be taken as further evidence in support.

Yes, I think there is much to be said for the idea that our present Krebs acids are examples of the very earliest carbon compounds in our biochemistry. Such molecules might have been the more convenient as they are mainly achiral. Sugars were perhaps too unstable and various to have been easily dealt with at the start.



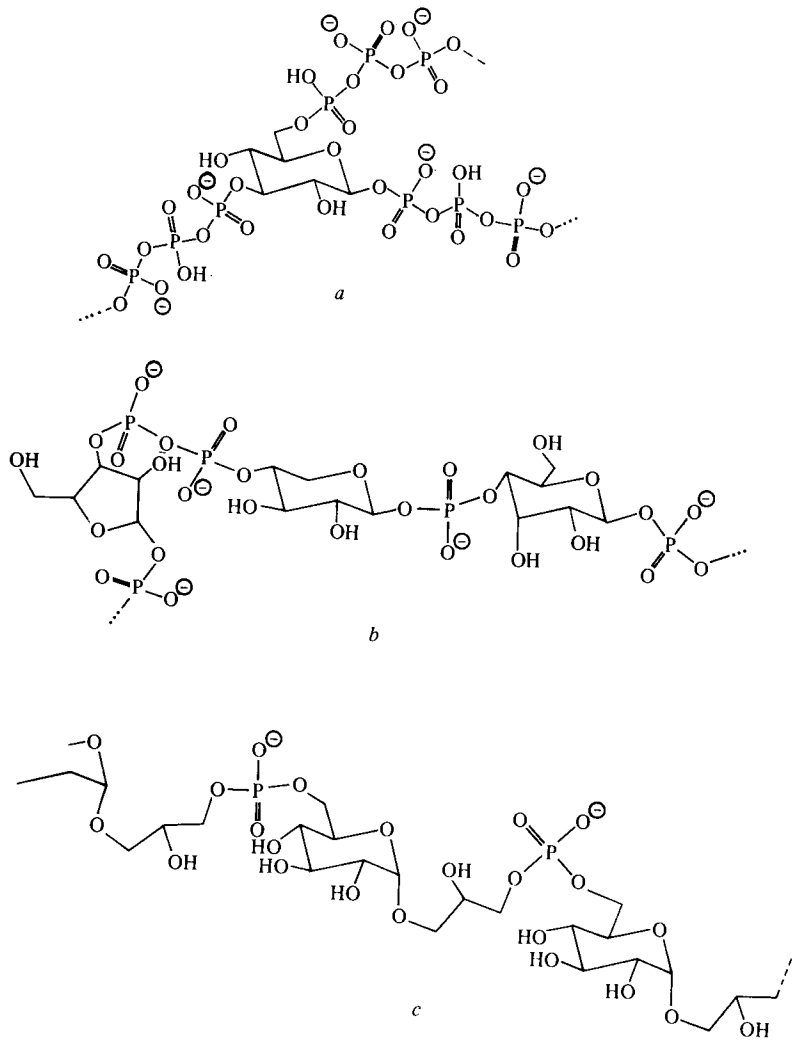



Figure 8.28. (*a*) A possible early use for sugars: to give a few branch points in mainly polyphosphate polymers. (*b*) A somewhat random teichoic acid-like polymer. (*c*) A modern teichoic acid: polyglucosyl glycerol phosphate.

*Uses for sugars* Eventually, though, some pre-nucleic acid organisms must have made sugars and been able to generate particular enantiomers of particular sugars. Only then would the subsequent consistent synthesis of nucleotides have become feasible.

Whether via formaldehyde or not, the first sugary products would have

been semi-chaotic mixtures with functions of the sort that semi-chaotic mixtures can perform, for example as an anchor, diffusion controller, or osmotic pressure controller. Then there would be selection pressures to make different mixtures for different purposes and to be less chaotic, to specialise in mixtures that were richer in more effective components.

This pressure might have been strongest where monosaccharides were being used to make polymers. For example, the properties of polyphosphate that were discussed earlier could be greatly modified if sugars were to be incorporated as connectors. Then branched structures would become possible (figure 8.28*a*) as well as more evenly balanced sugar–phosphate copolymers – teichoic acid-like molecules (figure 8.28*b* and *c*). Such molecules could have provided new ways of controlling viscosity, holding clay particles together and generally structuring the aqueous environment around the clay particles. And they could also have had a more definite structural use. Perhaps, like the modern teichoic acids, such molecules helped to structure the outer cell walls.

Once monosaccharides were being incorporated into polymers – and particularly where these polymers had structural functions that involved polymer chains aligning to any extent with each other – then there would be strong selection pressures on the control of the synthesis of the monomers. It would matter now if all the sugars in a polymer chain were the same or not – and chiral uniformity, or lack of it, would matter too. Polysaccharides would often be particularly sensitive. Much of the point of a polysaccharide is lost unless it is well made. For example, cellulose owes its excellent inertness to having lath-like molecules that stack compactly. But that shape depends on cellulose being made up from glucose molecules of a definite chirality $\beta$-linked through 1- and 4-hydroxyl groups. There are also more subtle properties of polysaccharides that depend on the shapes of the molecules. As we will discuss in the next chapter, three-dimensional gel structures can be formed by precise intertwinings and other interactions between some kinds of polysaccharide molecules. As with cellulose, one mistake – one wrong sugar or a wrong connection between two sugars – could be very disruptive.

### *The entry of nitrogen*

Whether you put sugars or Krebs acids at the very start, and wherever you put lipids, there are reasons for thinking that nitrogen came in relatively late. This was the view expressed by Lipmann (1965) who, on the basis of the present relationships between the molecules in our biochemistry, suggested the general sequence: carbon dioxide, formate, carbohydrate, acetate, fats and lipids, amino acids and then nucleotides.



There are two ways of looking at the place of nitrogen in our biochemistry. By paying attention to the now central genetic–catalytic machinery – the nucleic acids, the proteins, the coenzymes – one sees nitrogen as essential. Nitrogen is everywhere at the centre of control. But it is not at the centre of the biochemical pathways. Nitrogen is absent from *this* centre. Leaving aside coenzymes, there is no nitrogen in the glycolytic pathway or the Krebs cycle. And carbohydrates and lipids represent great tracts of the biochemical map that are virtually nitrogen-free. What exceptions there are here – for example glucosamine and ethanolamine derivatives – can easily be seen as later elaborations.

This apparently ambivalent position of nitrogen can be explained in terms of the idea that many of our metabolic pathways were invented, or at least sketched in, under the control of clay genetic–catalytic machinery, in organisms which to begin with need have contained no nitrogenous molecules. We might suppose that nitrogen had yet to be incorporated by the time Krebs acids and carbohydrates were already quite well established.

It is not only formaldehyde and simple amines that might have been difficult to deal with together, but even ordinary sugars and amino acids tend to misreact, especially through the Maillard or 'browning' reactions. It would seem that, generally speaking, special measures have to be taken if both sugars and amino acids are to be handled within the same system. It is perhaps significant that reactions between sugars and amino acids are early signs of deterioration of biological materials – this is one of the first things to go wrong as biological organisation breaks down (Feeney, Blankenhorn & Dixon, 1975).

Abelson (1966) had pointed to the difficulty that these reactions present to the whole idea of a primordial broth. An evolving organism would also be constrained by this mutual reactivity which might indeed be the explanation for the absence of nitrogen from the centre of the biochemical map – it might have been necessary to keep clear of amino acids until a C, H, O and phosphate organic biochemistry was well under control.

*A first set of amino acids* But eventually, we must suppose, it became possible to incorporate nitrogen – perhaps through a symbiosis with a simple ammonia plant of the kind that we were thinking about earlier. In any case, ammonia is the effective form in which nitrogen is incorporated into biochemical molecules today. In only a few steps, eight of the twenty protein amino acids are derived from central non-nitrogen

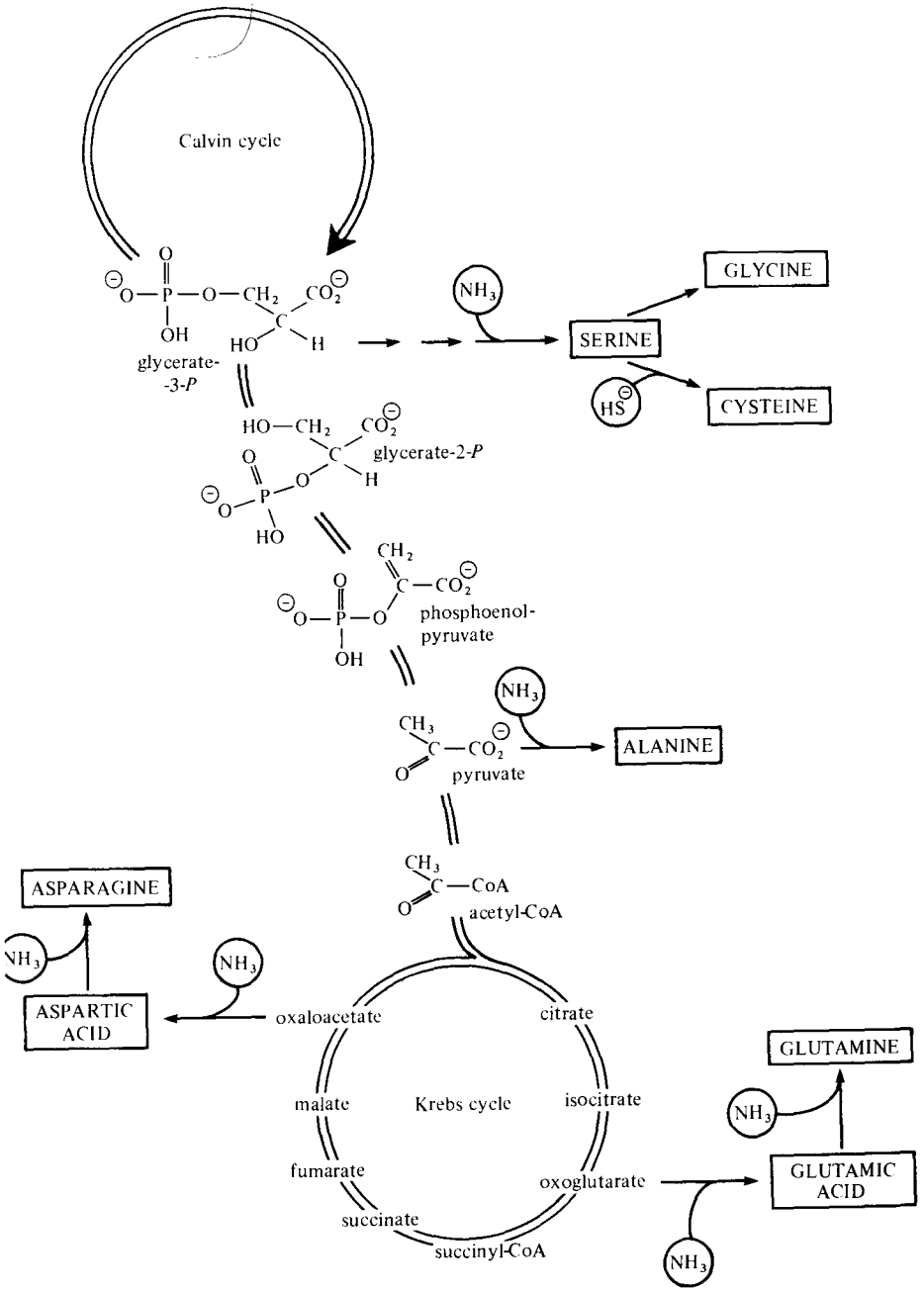



Figure 8.29. Eight of the twenty protein amino acids lie close to the central non-nitrogen pathways.



metabolites (figure 8.29), which suggests that amino acids, or at least some of these amino acids, were among the first organic nitrogen compounds.

### *The metabolic onion*

Following on ideas of Bernal (1960), Lipmann (1965) and Szent-Györgyi (1972), Hartman (1975a) would see our biochemistry in terms of concentric shells that formed on top of each other, each new shell tending to fix the shells underneath ('the onion heuristic'). Looking at figure 8.29 it is easy to see those eight amino acids as part of a new shell that was added after the centre was established.

We can see some uses for amino acids from the point of view of a clay organism. As zwitterions they would be a new class of highly water-soluble molecules that were reasonably stable. They would act as buffers, and as a new range of ligands for metal cations. And their interactions with clay would be strongly conditional on such factors as pH and metal ion concentrations. Clay genes would extend their control of their immediate aqueous environment if they could control the production of molecules like these.

One might predict that an advanced organo–clay biochemistry would have found uses too for some heterocyclic nitrogen compounds, since these too are often both water-soluble and stable, and since they often interact with clays (Chapter 6). Pyridines, pyrimidines and purines, for example, commonly intercalate 2:1 layer silicates serving as alternative exchangeable cations or (un-ionised) in exchange for interlayer water. Hence, complex organophilic micro-environments might be contrived in regions between clay layers.

*Purines and pyrimidines* Heterocyclic nitrogen compounds are, of course, a key class for modern biochemistry – with purines and pyrimidines at the very centre of the modern control machinery. But to judge from their positions on the metabolic map, heterocyclic nitrogen compounds were comparatively late additions. The purines and pyrimidines are at least the next shell out from those eight amino acids in figure 8.29. Their present biosyntheses presuppose, respectively, glycine and aspartic acid as structural units.

The modern biochemical route to purines is shown in figure 8.30. In line with the forward evolutionary strategy, we would see this pathway as a final refinement of a maze of earlier routes to purines and other molecules like them. To begin with there would have been many paths to many

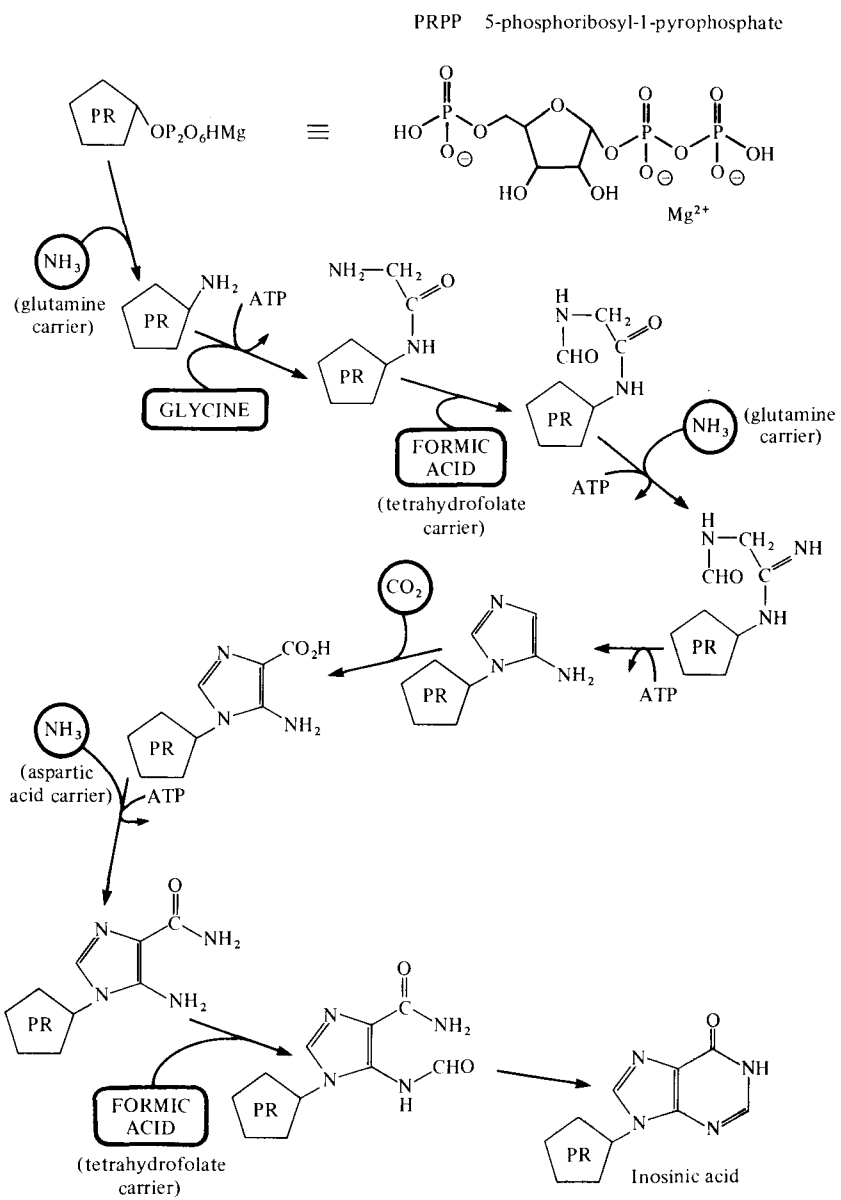



Figure 8.30. Contemporary purine biosynthesis is integrated with the synthesis of nucleotides. The activated RNA nucleotides adenosine triphosphate (ATP) and guanosine triphosphate (GTP) are derived from inosinic acid in four further steps each.



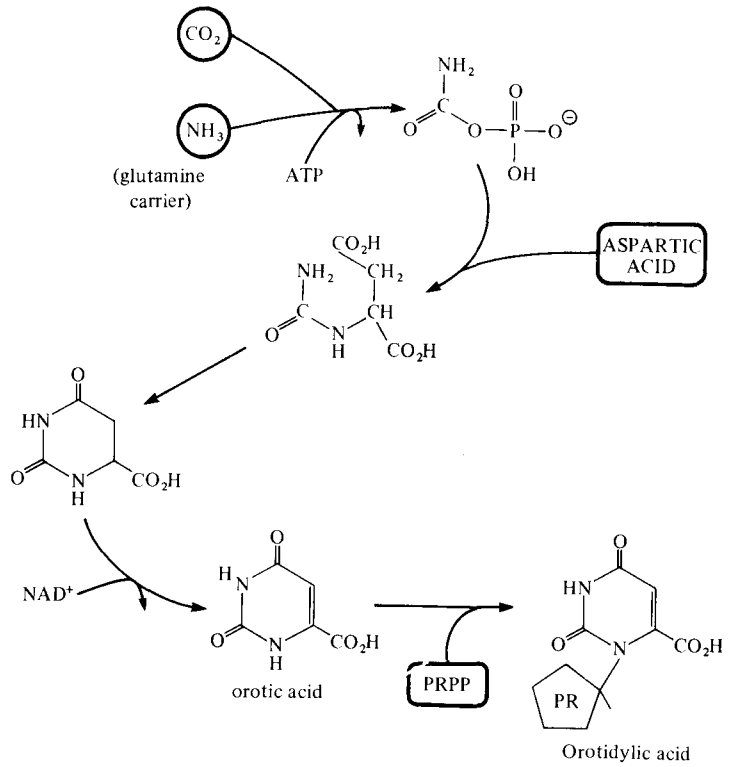



Figure 8.31. Pyrimidine biosynthesis (as far as orotic acid) is separate from nucleotide synthesis. Orotidylic acid is converted into the activated RNA nucleotide uridine triphosphate (UTP) in three further steps. Then one step more is needed to make cytidine triphosphate (CTP).

products. That would have been so not only *within* organisms that could only at first make semi-chaotic products, but it would be so also *between* different organisms, where there would be differences of style. Just because these adventures were peripheral they would be varied – in line with Darwin's generalisation (page 105).

But I will assume that the modern purine pathway still retains some information about the conditions under which its first semi-chaotic versions evolved. It looks as if a rather sophisticated sugar phosphate biochemistry was already established to provide by far the most complicated component, PRPP (figure 8.30, top), which is at the start of the synthesis. One might surmise that the original evolutionary experiments were in making nitrogen derivatives of sugars and sugar phosphates, using materials that were

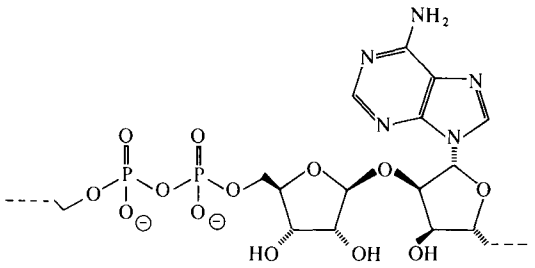


Figure 8.32. An example of a modern nucleic acid-like biopolymer, poly(ADP-ribose).

available at the time. Most of these materials were very simple – only glycine has more than one carbon and/or nitrogen atom in it. Two other 'first-shell' amino acids – glutamine and aspartic acid – are also involved in the modern synthesis as carriers for ammonia. This may also be a reflection of the molecules available early on, although it is fairly clear that the formate carrier, tetrahydrofolic acid, would have been a much later invention. (The structure of this coenzyme is given in figure 9.1.) The ATP reagent would also have to be seen as a later replacement.

The modern biochemical route to pyrimidine nucleotides is shown in figure 8.31. It is more ambiguous as to whether the sugar or heterocyclic unit should have come first since these units only join together after the pyrimidine unit has been made. But again the only component molecules with more than one carbon and/or nitrogen atom in them are PRPP and an amino acid – this time aspartic acid.

Why attach molecules with nitrogen in them to sugars? Surely it was not so that a new kind of genetic material would be found: replicating molecules were still far in the future. The reasons would have been more immediate – to make positively charged polysaccharides; to make polymers that would stick between clay layers; to make ultraviolet filters – you can think of many unsophisticated reasons. Even today there are molecules that are somewhat like nucleic acids but which seem to have much less sophisticated functions, for example teichoic acids (figure 8.28*c*) or poly (ADP-ribose) (figure 8.32).

*Outer shells* Several of the more complex amino acids are at much the same distance from the central pathways as the nucleotides. On Nicholson's (1974) metabolic map it takes about ten steps to make phenylalanine or tyrosine, for example, and about fourteen for tryptophan. Histidine seems the most recent of all – in a shell beyond the nucleotides – about 26 steps from the central pathways.



Many of the functional group types for coenzymes are also near the outer edges of the biochemical map. Among these, even the simple pyridine nucleus is apparently difficult to put together. For example, pyridine-2,3-dicarboxylic acid is six steps beyond tryptophan from which it is derived; and there are then four and five more steps to get to NAD and NADP respectively. Many of the coenzymes (figure 9.1) are evidently later than the RNA nucleotides of which they are elaborations.

*Two ambiguous classes* Porphyrins and other tetrapyrrols are more difficult to place since although they are derived from glycine and succinate – molecules that would have been available early on – tetrapyrrols are not themselves on the routes to other types.

A forward evolution is nevertheless strongly implied from the biosynthetic relationships between different tetrapyrrols, as Granick (1957) pointed out. Chlorophylls, haems, cytochromes and vitamin $B_{12}$ are all derived from a single tetrapyrrol intermediate (Battersby, Fookes, Matcham & McDonald, 1980). The nucleotide portion(s) of vitamin $B_{12}$ coenzymes implies a post-nucleotide origin for this subgroup, but the porphyrins cannot be placed more accurately than 'after glycine and succinate' unless other considerations are introduced.

Lipids are similarly ambiguous being derived from a central molecule, acetate, and yet being a more or less self-contained class. Towards the end of the next chapter I will suggest a late entry for both lipids and porphyrins.

## *Conclusion*

While I think that it is broadly correct that our metabolic pathways were built up from the centre towards the periphery, we should not imagine that all the details of the present pathways reflect detailed sequences of events. No doubt there were some major revisions. For example, the present Calvin cycle involves not only aldol condensations but also ketol transfers – which are much less normal reactions. These depend on the particular action of a coenzyme – thiamine pyrophosphate. Similarly, pyridoxal phosphate mediates reactions that would not otherwise have been expected. There would have been a great deal of updating possible once such 'magicians' had come onto the scene: and then again when enzymes had appeared.

So no doubt an evolved clay-based life would have used significantly different tactics in the design of pathways, in view of the very different means of organic-molecular control available to it. Perhaps cyanide, or its more accessible hydrate formamide, was used in those early days rather

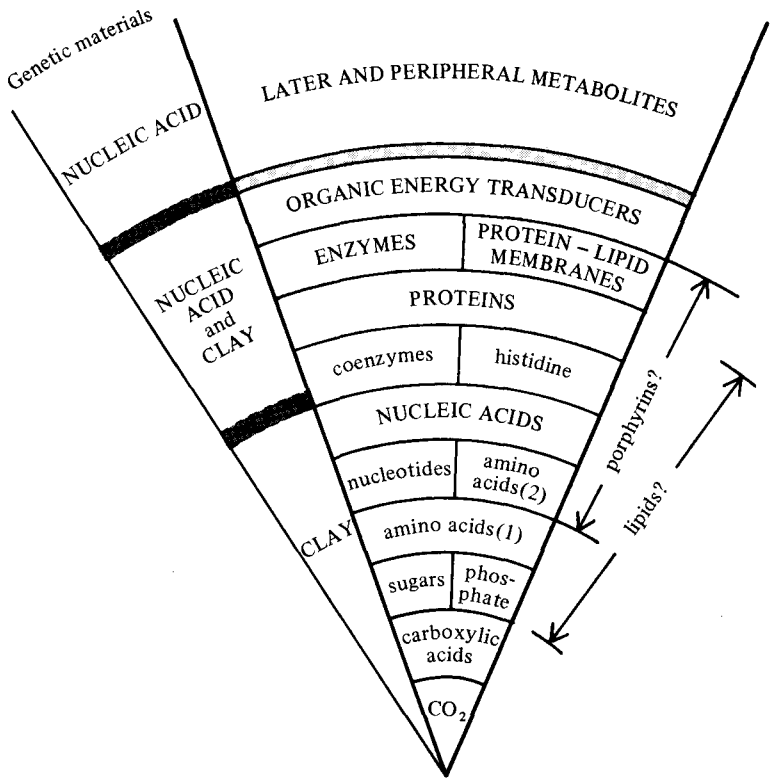



Figure 8.33. From the ways in which they now depend on each other, most of the components of our central biochemistry can be seen as having been added in a succession of layers (central sector). The classes shown here have been greatly simplified and there would have been a considerable spreading of some of them, such as the coenzymes, into layers above and below. Lipids and porphyrins are especially ambiguous, there being a wide range between the time when the components for their synthesis were available and the time when these classes would themselves have been needed for subsequent developments (right). A model like this only makes sense on the supposition that there was a takeover of genetic control. The timing of the genetic takeover, assumed to have been a one-step takeover between clay and nucleic acid, is indicated in the sector on the left.

than the less reactive ammonia and formic acid – as conceivably formaldehyde had an earlier biochemical role as discussed above. In that case our nitrogen pathways must have been quite radically redrawn. The question is very open, but I am inclined to think that the earliest successful syntheses of such molecules as adenine were not so different from present biosyntheses: because it is such a good idea, if you want to keep things under control, *not* to use building blocks that are too reactive – especially ones that react multifariously with themselves.



But even the broadest view of the biochemical sequence of events that can be read into the structure of today's pathways provides an argument for genetic takeover. So much had to evolve before our genetic material could have been made that there must have been other genetic material(s) to have supported that evolution. If the evolution of our biochemical pathways had depended from the start on nucleic acids – and if the onion idea is correct – then nucleic acids should be near the metabolic centre as well as at the control centre of present life of Earth. The present position of nucleic acids in the metabolic scheme can most easily be interpreted as that of an usurper that had come late onto the scene.

Figure 8.33 summarises the broad view of biochemical evolution given over the last few pages. I have added a few extra layers to the onion in anticipation of the next chapter.

# 9

# Revolution

In the last chapter we were trying to see why and how clay-based organisms might have evolved a competence in organic chemistry – as part of a trend towards a more extended control of the environment by (clay-held) information. According to our overall story, the stage was now set for revolution. Control would pass from the old centre of government to the provinces. At least one new genetic material was to appear in the organic parts of at least one species of evolved clay-based organisms; and this new material was to take over entirely the control of the means of its own reproduction.

According to the story so far, clay-based life has evolved a very considerable competence in organic chemistry. It can fix carbon dioxide and nitrogen. It can discriminate between enantiomers and can carry through quite long reaction sequences. It can use the energy of sunlight to make reagents to drive energetically uphill reactions. Clay life has arrived at a level of technological development at which the consistent synthesis of particular molecules of the complexity of nucleotides has become possible. It is expert at small-molecule organic chemistry – and is indeed now exploring the uses of polymers such as polysaccharides. Perhaps even small peptides – dipeptides, tripeptides – are being put together. But such molecules do not yet carry genetic information in the kind of way that our nucleic acids and our proteins may be said to. Nor do they yet directly contrive the synthesis of other molecules. Mainly the functions of these polymers and oligomers are structural – making a suitable gel with suitable conditional properties perhaps: in general helping to provide the milieu for metabolism rather than an intricate molecular control.

That, I think, is how the stage was set for the invention of our biochemical control machinery. This invention was to require the appearance of some polynucleotide-like replicating polymer – and then this polymer



was to come to control peptide synthesis. Clearly there are several more 'how' questions ahead. But at least such an eventuality becomes more plausible within an evolved biochemistry of some sort, as we discussed in Chapter 3 and 4.

But *why*, I hear Dr Kritic say. Why should natural selection ever stumble on such an extraordinary thing as a replicating molecule? If you remember the load of design restraints on any such thing (Chapter 5), it hardly looks like the kind of molecule that you come on by accident. Yet how else? Natural selection could not have *aimed* at molecular replication. It may be true that a replicating molecule is much more likely to have appeared in an evolved phenotype than a primordial soup, but is not such an accident still wildly improbable all the same?

If you recall the inventive strategies of evolution that we discussed in Chapter 3 you may see that such a conclusion is much too hasty. How was the lung stumbled on, or the inner ear, or the ability to fly? There was no forethought there either. The answers to such questions lie very largely in the terms 'preadaptation' and 'functional ambivalence' (Chapter 3). An object designed for one purpose – or evolved under one set of selection pressures – often turns out to have other uses. Indeed when we concluded that a replicating polymer would be more likely to appear in an evolved phenotype than in a primordial soup, we were making use of this idea. It is just because evolved apparatus, coupling agents, and so on, can have many uses that this would be so. Organic chemistry is generally easier in a well stocked laboratory even if the particular piece of research now being embarked on was not in mind when the laboratory was set up.

So our question is not a new one in kind. It is a question of whether a more specific preadaptation is plausible: whether some of the more detailed subsidiary functions needed for molecular replication could have been arrived at and perfected under selection pressures for functions that were initially more accessible. I will be suggesting in this chapter that a preprotein system for locating and connecting together organic molecules provided preadaptations for molecular replication – and that this same expertise in joinery was also to predispose organisms to the discovery of protein synthesis.

## The seeds of revolution; nucleic acid

### *Locating and connecting*

Clays could have controlled the locations of organic molecules in ways that we have already discussed – by controlling their synthesis and hence where they were produced, and by providing more or less specific

edge and surface sites where organic molecules could congregate. Hence genetic clays could have controlled organic molecules either directly or more indirectly via secondary clays.

The next shell of influence would start to form when, to some extent, organic molecules were communicating with each other – when, for example, molecules A and B, made in different places, could find each other and connect up without the further intervention of clays.

Today, proteins are very much involved in intermolecular communications. An enzyme brings molecules A and B together to react; or the protein subunits of a microtubule or muscle fibril 'self-assemble'. This is because, with protein, molecular surfaces can evolve which will key specifically with other molecular surfaces. A protein can be engineered to choose a molecule from its surroundings by having a socket that fits; or a collection of proteins can put themselves together in a way that is specified by their structures – a jigsaw puzzle shaking itself together by kinetic motion.

But before there was protein – before there was a generalised technique for forming the inverse image of another molecule – how did organic molecules recognise each other? How did they communicate? How could an organic molecule that had been made in one place be located precisely somewhere else?

Let us take that last, more specific question. The location of components is clearly an important part of the building of machines of any sort – and that includes biomolecular machinery. Think of an enzyme, for example, or a photosynthetic membrane. Several or many functional groups, molecules and ions have to be appropriately arranged. Indeed the arrangement, you might say, is everything.

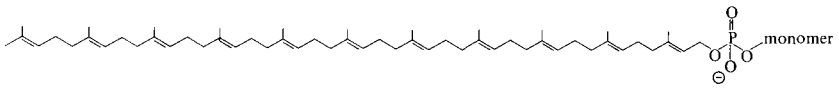
We might get some insight into pre-protein locating devices for organic molecules by thinking about some of the non-protein devices that are in our present biochemistry. Some of these may reflect techniques that were more prominent in earlier times.

*Lipophiles* The clustering together of amphiphilic molecules is perhaps the most obvious example of where molecules seem to know how to arrange themselves. To some extent at least their structure tells them. If clay organisms had been able to make such molecules, then, by specifying the synthesis of some particular set, they would have been able to control the making of higher order organic structures – micelles, membranes, vesicles perhaps.

Then, given a system that has hydrophilic and lipophilic regions in it,



other molecules would tend to partition themselves according to their hydrophilic or lipophilic character. In modern organisms the location of a molecule may be determined largely by its possessing a suitable 'handle'. For example the (lipophilic) phytyl chain in chlorophyll (figure 8.16*a*) is a locating device – it helps to hold the molecule in the thylakoid membrane. A similar device is used in the synthesis of bacterial cell walls. There, large and mainly hydrophilic monomer units have a long polyisoprenoid chain attached to them while they are being elaborated:



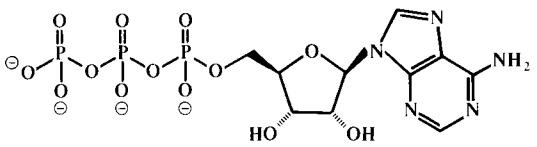

*Coenzyme 'handles'* There seem to be 'handles' too on most coenzymes (figure 9.1). Many of these molecules are much bigger than they would seem to have to be to account for their chemical functions. The seemingly non-functional parts can be understood as devices that help the enzyme to hold some key group securely without blocking its functionality. They can be understood as 'handles', that is, or in some cases as 'handles plus tie-lines' (for example in coenzyme A).

That coenzyme 'handles' now interact with proteins does not preclude them from the category of possible pre-protein locating devices. It is very striking how many of our coenzymes have nucleotide or nucleotide-like 'handles'. This led Orgel & Sulston (1971) to suggest that some of our coenzymes were designed originally to link up with polynucleotides. White (1976) has suggested that perhaps a metabolic system composed of nucleic acid enzymes existed before the evolution of ribosomal protein synthesis. Our coenzymes were part of this system: so were precursors of our transfer or tRNAs, which can be seen as large coenzymes participating in the group transfer of amino acids. These are attractive ideas.

*Labels* There is a certain ambiguity between handles and labels. A fancy handle may help in identification – and you might say that the tRNAs are like this. They turn an amino acid into an object that can be recognised by nucleic acid – the point being that a stretch of polynucleotide is very good at recognising another (complementary) stretch of polynucleotide, but it is not very good at recognising anything else. So one might surmise that when coenzymes were associating with nucleic acids rather than with protein enzymes – before there was protein to do most of the recognising – the coenzyme handles had to be more explicit about what

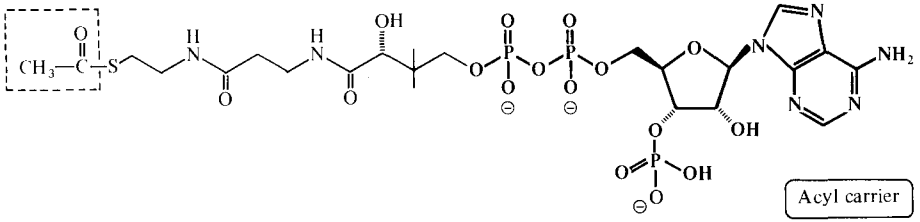
MAIN OR TYPICAL FUNCTIONS

*Adenosine triphosphate (ATP)*



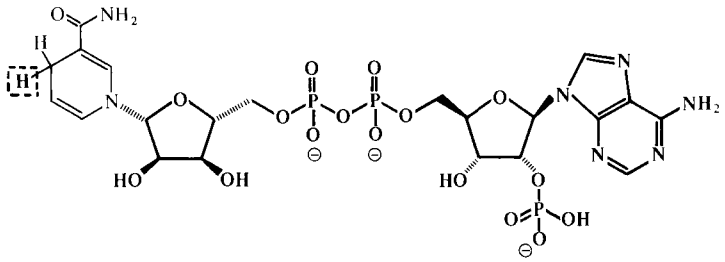
An energy carrier: drives synthetic reactions against thermodynamic gradients

*Coenzyme A (CoASH)*, acetyl derivative



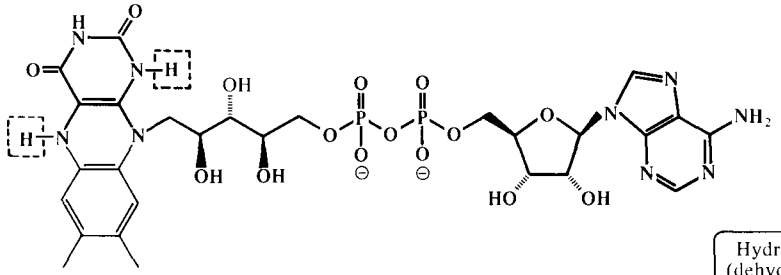
Acyl carrier

*Nicotinamide adenine dinucleotide phosphate (NADPH)*
(NADH lacks the phosphate side group)



Hydride carrier

*Flavin adenine dinucleotide (FAD)*
(In FAM, flavin adenine mononucleotide, there is only a phosphate group in place of the right hand adenosine diphosphate part)
Hydrogenated form ($FADH_2$):

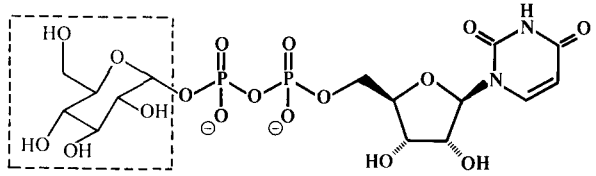


Hydrogen carrier (dehydrogenations)

Figure 9.1. The main coenzymes, emphasising RNA nucleotide or RNA nucleotide-like parts.

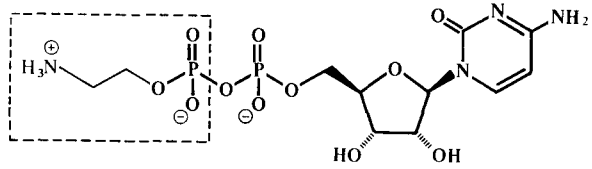


*Uridine diphosphate coenzymes*, e.g. UDP-glucose



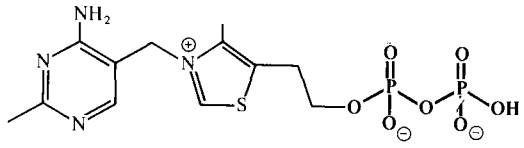
glycosyl carriers (e.g. in polysaccharide synthesis)

*Cytidine diphosphate–alcohols*, e.g. CDP-ethanolamine



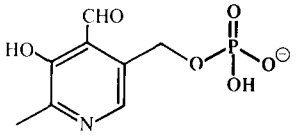
alkyl phosphate carriers (e.g. in phospholipid synthesis)

*Thiamine pyrophosphate (TPP)*



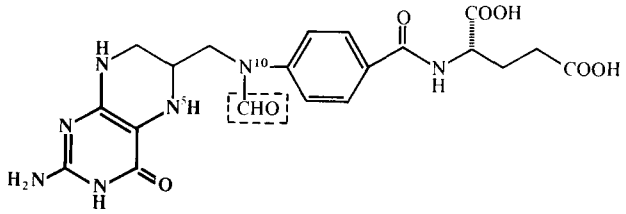
decarboxylations of α-keto acids: $-C(=O)-CH_2OH$ group carrier in the Calvin photosynthetic cycle (figure 8.18)

*Pyridoxal phosphate*



various transformations of amino acids

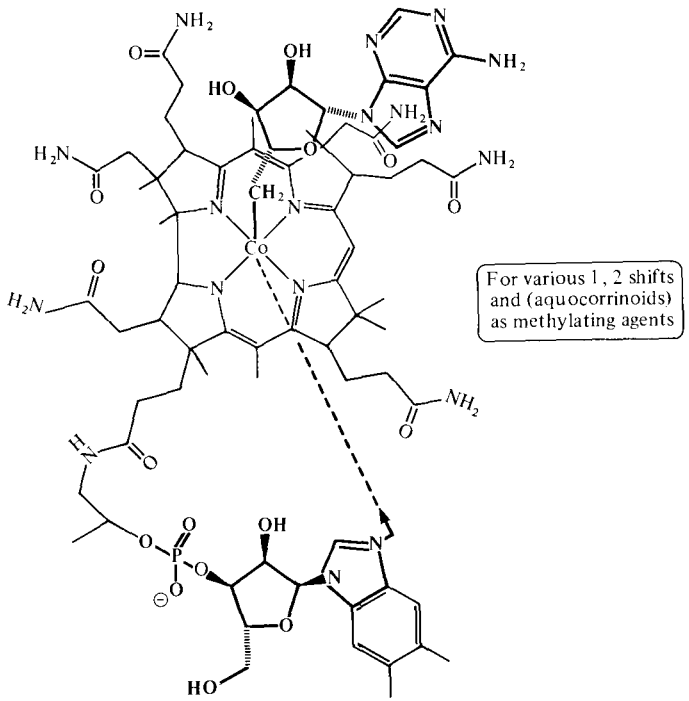
*Tetrahydrofolic acid ($FH_4$)*, $N^{10}$ formyl derivative



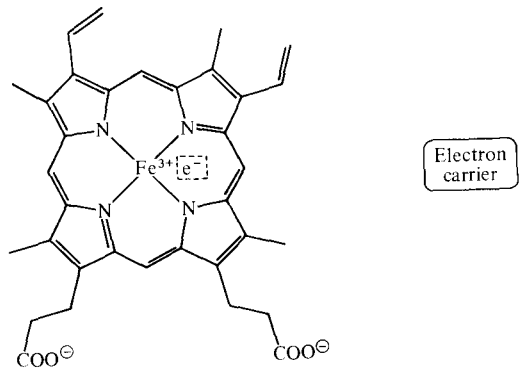
$C_1$ carrier – of formyl (as shown) or hydroxymethyl, or methyl groups; using $N^5$, $N^{10}$ or both

Figure 9.1 (*contd.*)

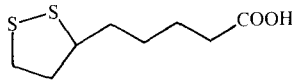
*Vitamin $B_{12}$ coenzymes*, e.g. Co-5′-deoxyadenosylcorrinoid




*Haem*




*Lipoic acid*




*Biotin*

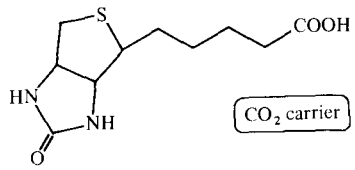



Figure 9.1 (*contd.*)



they were carrying. We now see, perhaps, only the stub ends of earlier much fancier devices: in place of the adenosine or other nucleotide 'handle' in ATP, coenzyme A, etc., there were longer identifying oligonucleotide 'labels'. In our present tRNAs, then, we are perhaps seeing examples of what all coenzymes used to be like.

*Carbohydrate systems* Carbohydrates are involved in a number of identifying and locating devices in modern organisms. Plant glycosides, such as saponins, are at least more water-soluble by virtue of their monosaccharide or oligosaccharide attachments, even if the sugar units have no more precise locating functions.

Oligosaccharides attached to the outer surfaces of cells are important for communications between cells – in recognition and cohesion. The multiplicity of possible monosaccharide units and of the ways in which they can be linked together to make straight or branching chains makes it possible in principle for a great deal of information to be conveyed in the choice of just a few kinds of oligosaccharide – although proteins are also involved in these very sophisticated intercellular interactions.

Interactions between polysaccharide molecules may give us a better indication of pre-protein techniques. Polysaccharide molecules can often connect specifically with each other, if they contain like chains of monosaccharide units. This idea can be used to explain the properties of several gel-forming polysaccharides (Rees, 1977; Rees & Welsh, 1977; Winterburn, 1974).

ι-Carrageenan is one of the gel-forming polysaccharides in agar, the intercellular matrix material of red seaweeds. The ι-carrageenan molecule contains stretches of alternating D-galactose 4-sulphate and 3,6-anhydro-D-galactose 2-sulphate which are interrupted every so often with a D-galactose 2,6-disulphate, or 6-sulphate, unit in place of the anhydrogalactose (figure 9.2). The regular stretches twist together to form double helices which are well stabilised by regular hydrogen bonding between the chains – through the two hydroxyl groups of the galactose sulphate units. Hence the molecules become tied together into an open three-dimensional network (figure 9.3). A great deal of water is entrained in such networks – a typical agar gel is more than 99% water.

The double helical tie-points in agars can be 'melted' and reformed by heating and cooling. With alginates – major components of brown seaweeds – the cation environment controls the state of gelation. Here there is a rather different way in which the polysaccharide chains lock together. The alginates are copolymers of D-mannuronic acid and L-guluronic acid.

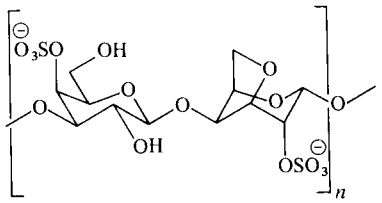


Figure 9.2. A regular sequence such as this in an ι-carrageenan molecule will form a double helix with a similar stretch in another molecule.

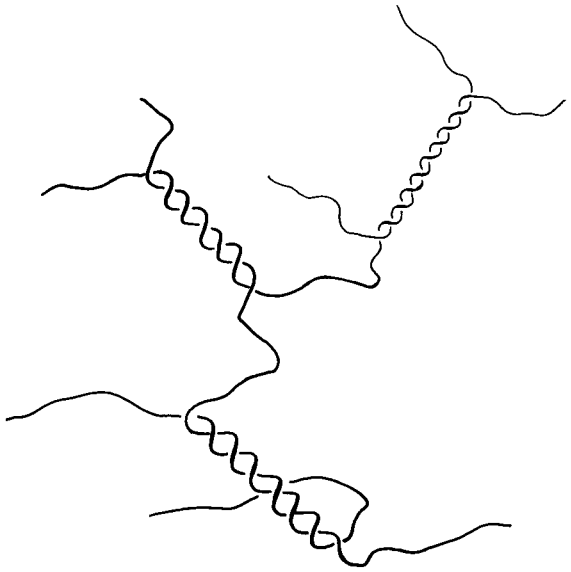


Figure 9.3. ι-Carrageenan molecules joining together through stretches of intermolecular double helix.

Stretches of pure L-guluronic acid are self-cohesive in the presence of calcium ions and it is suggested that the metal ions are located as shown in figure 9.4, in 'egg boxes' between the, rather rigid, puckered chains (Grant, Morris, Rees, Smith & Thom, 1973).

The cellulose fibre owes its rigidity and stability to the locking together of the identical flat polyglucose molecules that make it up. Albersheim (1975) has suggested that at a higher level, in the organisation of the fibres in a primary plant cell wall, there are matchings of a more complicated



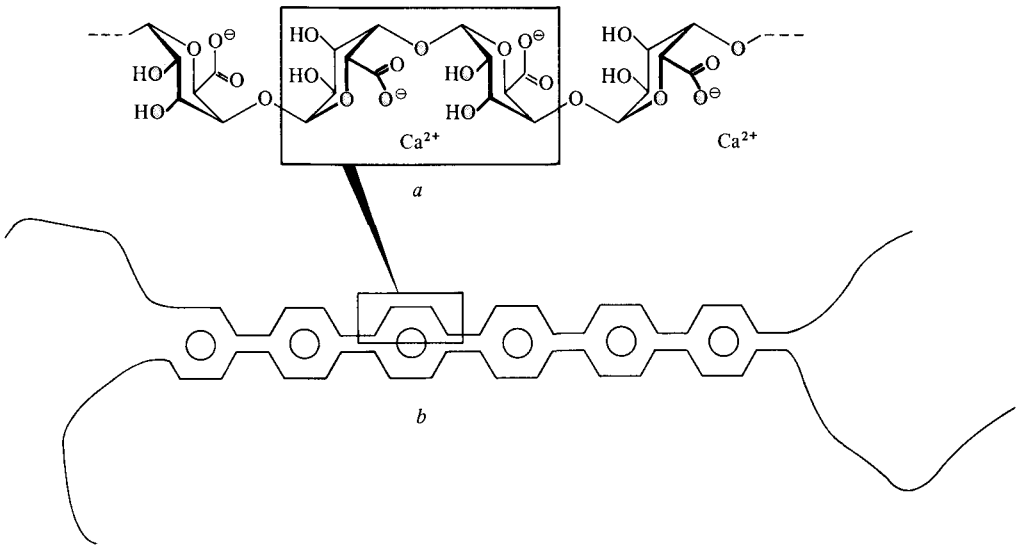


Figure 9.4. (*a*) A stretch of 1,4-linked α-L-guluronate showing a mode of co-ordination with calcium ions (co-ordinating oxygens shaded), as suggested by Rees (1977). (*b*) Longer view of a pair of such chains held together by calcium ions in ten-fold co-ordination. (Redrawn from Rees, 1977.)

kind. A complex branched polysaccharide in the cell wall contains stretches of 1,4-β-linked glucose (as in cellulose itself) and it is thought that these stretches lock onto the cellulose fibres, holding them in place.

### *The invention of nucleic acid*

I am not imagining that the very polysaccharides that we now find in our seaweeds (or anywhere else) were present as components of pre-protein locating mechanisms. What these studies of modern polysaccharides show is the range of control devices that could have become available to primitive organisms that had arrived at the ability to make particular monosaccharides and put them together in particular ways. We might suppose that molecules like polysaccharides – big molecules that were rather regular, water-soluble, and not too flexible – were among the first that could lock with each other at all precisely.

In the last chapter we were imagining an era of early evolution when a number of species of organisms were exploring a variety of water-soluble polymers – for directly connecting clay particles together. Now we might

suppose that clay organisms could make polysaccharides with suitably contrived self-cohesive stretches in them. The organisms could now specify a pattern of association that did not directly involve clay particles – they could make, now, a new set of micro-environments among the twisted and twisting molecular threads. Here would be a new kind of matrix with places in it. Other molecules could be located in this matrix if they had oligomers attached to them that were of the right sort.

Before protein appeared, to do away with most of them, there might have been quite a number of kinds of organic locating mechanisms. Polynucleotide was perhaps only one of the more sophisticated ideas that were around. No alarm bells rang when heterocyclic side chains first started to be attached to various sugar-phosphate polymers. There was, I daresay, some duly pedestrian reason for it – to make some polymer stretches positively charged, perhaps, to hold them between clay layers, or to provide an ultraviolet filter.

Nitrogen heterocyclic molecules often stack into micelles in aqueous solutions (for example acridine orange, cynanine dyes) and, in any case, rigid and somewhat hydrophobic side groups would tend to cluster together within a gel matrix. Interesting hydrophobic regions might be made like this through inter- and intramolecular associations that might be further stabilised by hydrogen bonds between the heterocyclic side groups.

For a molecule readily to fold on itself, the rather rigid polysaccharide backbone would be less satisfactory than one that included more flexible links, such as open-chain alcohols or phosphate groups. The modern teichoic acids have both these kinds of flexible groups in them (figure 8.28*c*).

But there is a snag about flexibility here. It may allow self-folding but it will weaken the tendency for stretches of polymer to lock with each other. So an increasing flexibility – arising, say, through an increasing frequency of interruptions by phosphate in polysaccharide chains – might call for side groups that fitted into each other – pairing bases for example. Hence a molecule with a flexible hydrophilic backbone plus more hydrophobic side chain that could pair with each other might be seen as a sophisticated way of getting the best of both worlds – for making a particularly effective structure-forming system in which flexible molecular segments could nevertheless be located precisely. In addition, other molecules that had suitable (polynucleotide-like) tails attached to them could also be precisely placed.

These molecules were not yet nucleic acids. Structurally they were getting close, but functionally they would have been more like our proteins. With purely structural roles to begin with, they were to lead to that era dominated by the 'nucleic acid enzymes' that White talks about. Not all the side



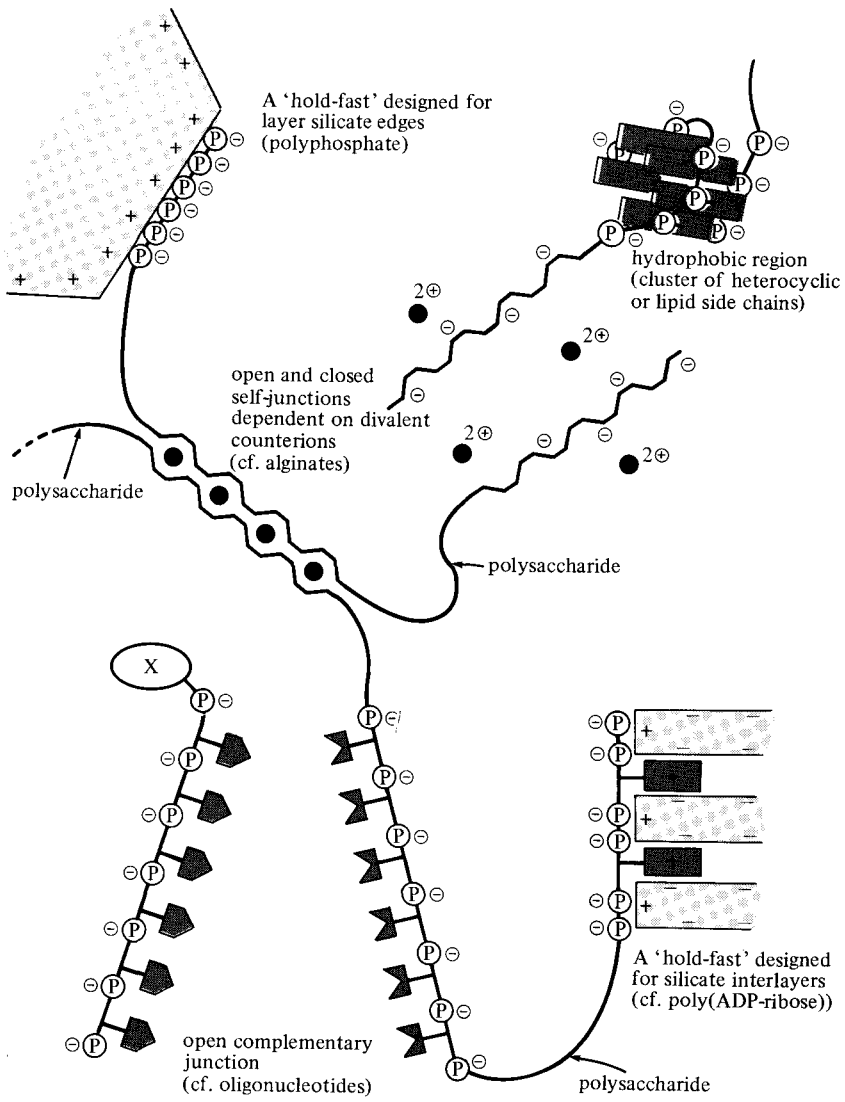



Figure 9.5. Some elements in an imaginary form of 'molecular connective tissue' for clay-based organisms. There are three main functions for this material: to hold mineral components together, to structure the intercrystalline aqueous environment, and to locate particular smaller organic components in particular places within that structure – for example, lipophiles might concentrate in polymer-micelle regions (top right), while a special functional grouping (X) might be located by virtue of an attached oligomer tail.

chains would have been heterocyclic bases; and, of the bases, not all would have been pairing types. Cohesion between side chains was only part of the structure-forming technique – and no one had yet whispered the word 'replication'. A sketch of the sorts of molecules that we might be thinking about now is given in figure 9.5.

*Questions of synthesis* One might imagine clay organisms polymerising organic monomers by lining them up in specially engineered grooves and tubes. But even with such apparatus it is questionable whether it would have been possible to specify a long copolymer sequence unit-by-unit. In Chapter 7 we decided that genetic information would have been written on a rather larger scale in clays than it is written in nucleic acids. It would be especially difficult to see how a clay template could be sufficiently finely contrived to specify the positioning of individual amino acids in a polypeptide chain since the information density here is so high. But copolymers mainly made of blocks (with an occasional special group perhaps) – molecules like carrageenans or alginates – might have been much more accessible.

With such molecules the genetic information is expressed firstly in the kinds of blocks and special groups used. For primitive organisms this would have been determined by the nature of the clay synthetic apparatus. Secondly, the genetic information is expressed in the arrangement of the blocks and special groups.

Referring to the hypothetical molecules in figure 9.5, we might imagine a clay organism that has one organelle specialising in the production of (something like) hexa(U), another specialising in the production of teichoic acid-like material, and so on. These organelles are connected up with a final one that has in it a template for aligning the pieces appropriately before the final assembly. That template would contain part of the information needed to make the final products. But because most of the units being handled were fairly large, this information could still be quite coarse grained – it might be, for example, an arrangement of grooves defined by twin composition planes with occasional special sites at edges or intersections.

Given the biochemical machinery for making stretches of polymer that stick to each other, the discovery of an organic replicating molecule would have come much closer. Oligomers such as hexa(U) would, on their own, tend to stick to complementary regions of a pre-existing polymer. This – part of their *raison d'être* as structure-formers – would also provide an alternative mode of alignment for the final assembly of special connector molecules that were made up entirely from such oligomer units. A molecule like this we might now call a 'protopolynucleotide'. (It would be like nucleic acid but with, perhaps, different pairing bases and/or a somewhat different backbone.)



The 'classical' hypothesis for the replication of nucleic acid involved the idea of single nucleotides locating on a preformed strand and then polymerising (Watson & Crick, 1953). It is doubtful if such a system based on single nucleotides would work. Single nucleotides do not stick well to polynucleotide strands. But long enough stretches do and indeed the kind of mechanism that we have been imagining has been demonstrated to the extent that hexa(U) segments have been joined together on a poly(A) template without enzymic assistance (Naylor & Gilham, 1966).

So that is one way, perhaps, in which organic replication might have eased itself into an evolved clay-based life. One aspect of one kind of organic polymer was now directly transmissible between the organic polymers themselves: a copolymer block structure was a new kind of genetic information that was independent of clay. The clay templates previously needed to specify such block structures could now be dropped. This was only a small devolution, no doubt, since most of the genetic information needed to make these copolymers would have been involved in the specification of the necessarily complex apparatus needed for the synthesis of the monomers and for making the blocks. The assembly of these blocks was a comparatively trivial affair. But it was a start.

We discussed in Chapter 4 another possible easy route to organic genes – through non-genetic replication – where it was important only that two sequences were more or less matching, but where it did not matter much what the sequences were. We were thinking of structure-forming systems, but one might also imagine labelling systems that were like this. If you want to write an oligomer label that will seek out a particular stretch of polymer then it might be an idea to build the oligomer on the polymer so that it has a complementary sequence. The label will then remember where to come back to.

A third easy route that might be imagined would be through organic genes that, to start with, held no sequence information. For example, suppose that ABAB . . . is a self-cohesive polymer, then it might be best synthesised on a pre-existing strand – so that the *length* of the daughter strand is defined. Length is the only genetic information being transmitted in this case, since the polymer is always ABAB. To begin with, perhaps, only this sequence is replicable although evolving replication apparatus subsequently allows other sequences, eventually any sequence, to be replicable.

When for any reason there was a selective advantage in being able to

copy copolymer blocks, or features of any sort, there would have appeared selection pressures to have improved the efficiency of that replication. New kinds of catalytic grooves and tubes might be expected to appear now and be modified by selection. Such a piece of apparatus would no longer be a template transferring information to a phenotype polymer. It would be a jig, designed mainly to hold: to locate daughter units, and to help to join them up – indifferent to the features (e.g. sequences) being copied. In short, *any* use for organic replication would be a reason for the evolution of some sort of 'replication apparatus'.

To begin with, and from the clay's point of view, replication was simply an alternative way of finishing off the synthesis of a certain rather limited class of phenotype polymer. The technique would have been limited too in that it would first have appeared in only one species (with perhaps rather odd requirements). These polymers would in any case have been part of the 'secondary', or outer, metabolism – useful no doubt, but specialist structures adapting a particular organism to a particular life style. In this respect they would have been like our modern carrageenan, or lignin, or melanin.

Indeed up to now all the organic chemistry of the evolving clay organisms was 'secondary metabolism' in the sense that it was only very indirectly concerned with the central task of transmitting genetic information to the next generation. This has been an important part of our story. So placed, the organic part of primitive metabolism could be variable under evolutionary pressures. A great variety of products and possibilities could be explored – as they are explored still in the vast numbers of secondary metabolic products of present-day life.

These new organic genes would have been peripheral to begin with (in spite of a slight smudging now between primary and secondary metabolic categories). They were recent elaborations – optional extras if you like. Only some organisms had them at all. In line with more general evolutionary considerations (pages 93–104) you might have expected that soon there would have been a considerable structural variability between the organic replicating polymers in different organisms that were occupying different niches. There would be room, still, for evolutionary experimentation with these new materials – and a chance to discover a new molecular biology.

In fact the main work had already been done. It might have seemed a small step to have introduced those first crudely replicating polymers, but it was fatal to the whole clay system. In any competition between organic and inorganic control systems, the organic ones would be almost bound to



win in the end – because metastable structures can be engineered more finely with the use of organic molecules. Anything clay could do organic molecules would eventually come to do better.

*Improvements* There was a long road ahead, though, even if it was downhill overall. First of all the new genetic information itself would have to come to be more finely written and more faithfully copied – say through copolymer blocks becoming shorter, and single units being included more often, till eventually information was replicated unit-by-unit. This would be possible in principle now that the templates were the protopolynucleotides themselves rather than crystal surfaces.

To begin with we might imagine 'replication apparatus' made from clays. Positively charged grooves on the edge surfaces of layer silicates, or grooves along sepiolite crystals, might have been particularly well suited (look at figures 5.7, 6.29, 6.48 and 7.9). There would also by now be some quite sophisticated organic structures to be brought in – if they happened to be of marginal assistance – and then improved under natural selection. Later, protopolynucleotides might collaborate, or even take over as the main material for the 'replication apparatus'.

Of the various control polymers now around these protopolynucleotides would be star performers, their flexibility and high information density allowing them to fold into functioning objects with particularly well specified tertiary structures. But these new and versatile machines were to set the stage for another more dramatic breakthrough. This would have happened through the normal tendency for evolving genetic information to extend its control. Rather as clay genes had extended their control by assisting the growth of secondary clays, so these new organic genes were to come to cause the formation of another kind of organic polymer – only this time the secondary control material was to be very precisely specified – more intricately indeed than the genetic material itself. With this, the invention of protein, control was to shift decisively in favour of the protopolynucleotide genes. Protein was the way that nucleic acid won.

## The agents of revolution; protein

### *Amazing machinery*

Our understanding of protein synthesis in modern organisms has developed rather similarly to that of DNA replication (described in Chapter 5). In each case formalised schemes were drawn up early on in which monomer units were seen to line up against a premade polymer

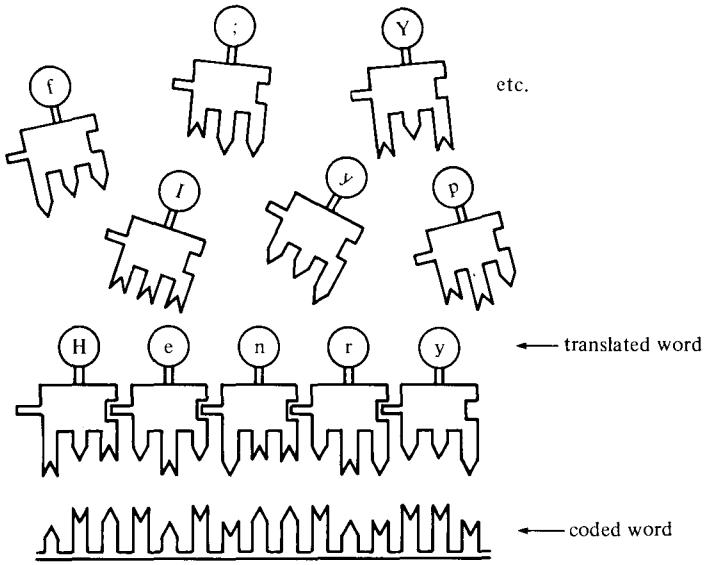



Figure 9.6. Mechanical translation of a message written in a four-symbol code can be imagined that depends on adaptors that fit sets of symbols – here triplets – in a message. Each adaptor that fits a particular triplet has a characteristic symbol of a much longer alphabet attached. The set of adaptors thus constitutes the 'code book' for the translation. There would be 64 kinds of adaptors possible in this case, corresponding to 64 possible kinds of symbol for the translated message. (This might be, for example, upper and lower case letters, plus eight punctuation marks, spaces, etc.) Such a system would correspond quite closely to the way in which (four-symbol) genetic messages are used to determine the sequences of the twenty different kinds of amino acids in proteins. In protein synthesis, however, no more than two adaptors are put in sequence at the same time (see figure 9.7). With only twenty amino acids, together with 'stop' and 'start' punctuation marks, the coding capacity appears to be underused by the protein-synthetic machinery.

acting as template. For protein synthesis, schemes of this sort were necessarily complicated by the lack of correlation between the number of kinds of monomers in the supposed template (DNA with four kinds of nucleotide units), and in the polymer that was supposedly being produced (protein, with twenty kinds of amino acid units). It was difficult in any case to see why amino acids, with their diverse side chains, should tend to line up against a polynucleotide template.

Crick's adaptor hypothesis, and the idea of a triplet code embodied in a set of adaptors, was a neat solution. With such an idea a line-up mechan-



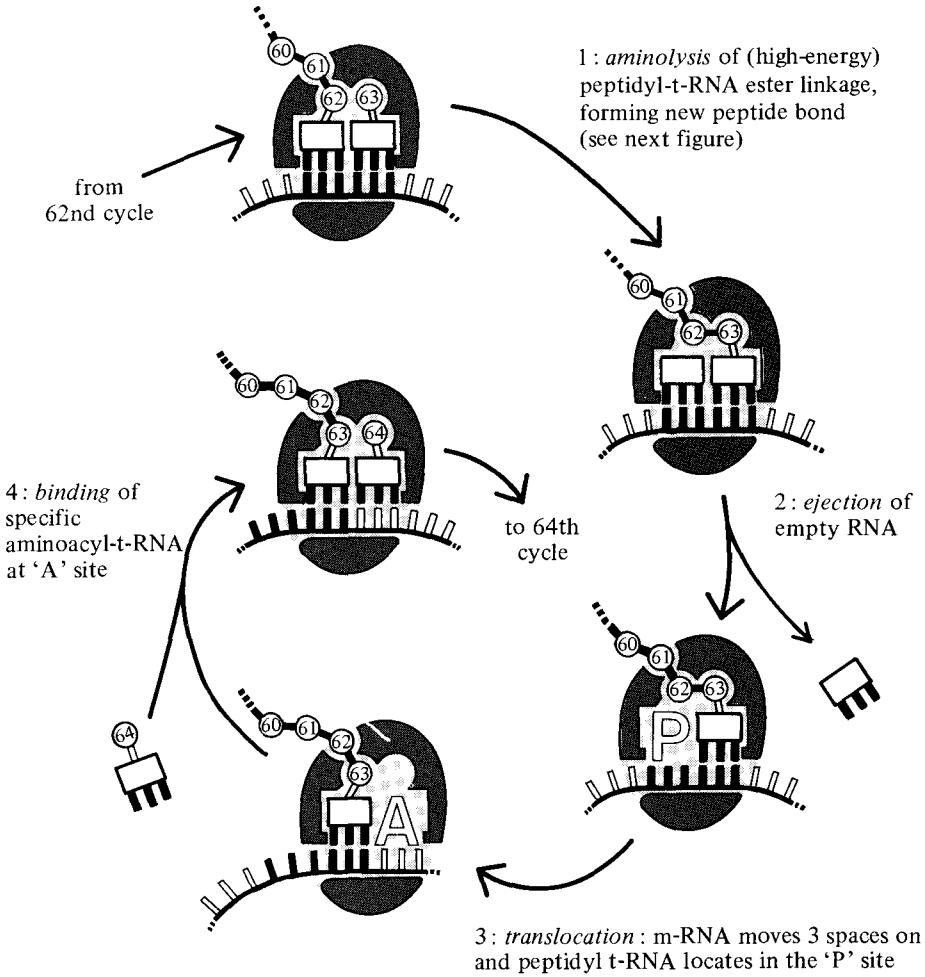



Figure 9.7. A formalised and highly simplified view of events in the elongation of a peptide chain by a ribosome. Whereas the amino acid units being put together contain 7–24 atoms, the stretch of message tape needed to specify one amino acid has 80–90 atoms in it. The whole ribosome (stippled) is a machine built from some 170000 atoms. In addition several soluble proteins are needed for this machine to work.

ism can be formulated for translating messages written with a four-symbol alphabet into messages written in an alphabet of up to 64 symbols (figure 9.6). The postulated adaptors were soon identified with a set of relatively small RNA molecules of about 80 nucleotide units each – the tRNAs.

For protein synthesis, as for DNA replication, a line-up mechanism,

although correct at a certain formal level, is not literally correct. Here again the true mechanism, although still not completely understood, is a sequential unit-by-unit one: it is more like the operation of a typewriter than that of a printer's block. And it is not exactly DNA, but offprints of one of its strands that act as templates – the messenger or mRNAs.

The third kind of RNA – ribosomal or rRNA – is the major constituent of the ribosome. This is the device that brings the other two kinds of RNA together. Again ideas about it have changed somewhat. At one time the ribosome was thought of as a static sort of jig. Now it is seen rather as a GTP-powered coupling machine of great sophistication.

All the ribosomes in all the organisms on Earth seem to work in much the same way. Ribosomes are large and complex structures. The *E. coli* ribosome, for example, contains three single-stranded RNA molecules with about 3000 nucleotide units in the biggest of them, and about 1600 and 120 in the other two. There are also 55 protein molecules – all of them different from each other and mostly very different. There are about 8000 amino acid units altogether in these structural proteins.

In operation, GTP and several non-structural proteins are needed usually – for initiation, for chain elongation, and for termination. Each of these stages has special requirements. A formal picture of the main synthetic operation – the cycle of chain elongation – is given in figure 9.7.

Looking at Merrifield's machine for peptide synthesis (figure 1.11), or indeed reading the experimental section of any paper describing a laboratory synthesis of a long peptide, we should not really be surprised that the protein-synthetic equipment in organisms is so elaborate. It is (quite simply) *difficult* to put amino acids together into long defined sequences.

*Preadaptations?* How did such expertise evolve? There was no forward planning. The ribosome was not designed first and then evolved towards. It was arrived at, as usual, through successive small modifications or new combinations of pre-existing machines – but machines that had not necessarily been concerned, to begin with, with anything like protein synthesis.

Already, according to our story, some key pieces for a molecular translating machine have been invented. These organisms are attaching labels to molecules as part of a generalised technique for putting things in their proper place. Particular enantiomers of particular amino acids are among the molecules that are being made consistently – and they are being located in this way. So something very like the tRNAs is already there. For a long time there have been selection pressures to perfect the means of attaching amino acids (among other things) to their particular RNA-like oligomers.



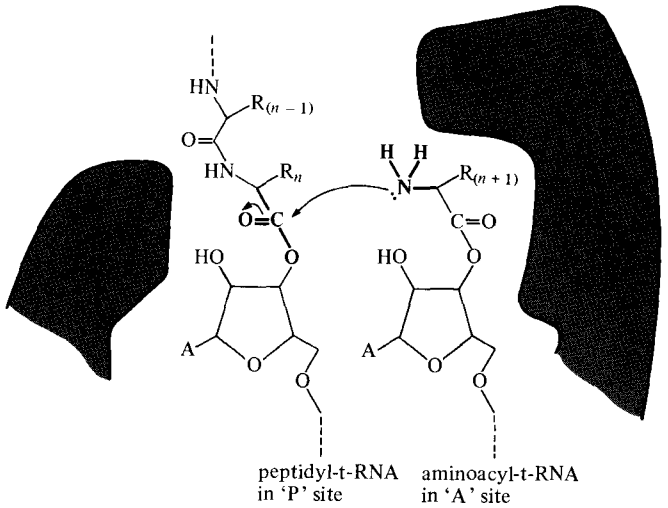



Figure 9.8. Close-up of the first event in peptide bond formation (in stage 1 of figure 9.7).

One might suppose that sometimes the purpose of locating a particular molecule somewhere was to bring it next to another similarly located molecule with which it was to react – that this was part of the technique of the 'nucleic acid enzyme', and a means whereby dipeptides, among other things, could be conveniently made. Then, somehow, this technique was extended to allow the synthesis of polypeptides of indefinite length.

However this happened, the site where the amide links were forming had to be appropriately made. It had to be accessible to the reactants – the amino acids with their identifying labels – and yet kept dry to avoid the ever present threat of hydrolysis in place of amide-bond formation. And the site should have been such as to favour the un-ionised form of the amino group that was reacting (see figure 9.8).

Following our earlier discussions, clays might have been particularly useful in helping to provide suitable reaction sites. That is to say, clays might have been components of protoribosomes.

Speculations about protoribosomes would be less wild, perhaps, if we knew more about the details of how the modern ribosome works. But we can see already one most characteristic feature of this huge 'enzyme', which can be used to illustrate yet again how the pre-existence of a well evolved organism can soften the edges of the most difficult looking puzzles.

Translocation, the shifting of the message tape in relation to the site in the ribosome where the peptide bonds are formed, is especially what allows this 'enzyme' to judder on indefinitely – so as to make a polymer

instead of releasing its product after each cycle of operations as most enzymes do (see figure 9.7). There is a mechanism here equivalent to the gate in a cine camera; or the mechanism that shifts the carriage of a typewriter between letters. It seems very contrived and mechanical. And yet, can we not imagine preadaptations for that sort of thing?

Here is one possibility. Although the subsidiary energy carrier, GTP, is generally used to help to drive the cycle of chain elongation in peptide synthesis, there appear to be some cases where GTP is not absolutely required (Spirin, 1978). The energy derived from the aminolysis (figure 9.8) is seemingly sufficient to drive the rest of the cycle, including translocation. So perhaps translocation came first – the aminolysis (or hydrolysis) of active esters of tRNA-like molecules being used simply to *move* RNA molecules.

We will touch again on the intriguing question of the evolution of the ribosome. But for the rest of this section we will be concerned mainly with two related questions that are perhaps even more critical (and which we skated over rather). These are questions about how amino acids were attached to the correct adaptors in the early days, and how the genetic code evolved.

### *Matching adaptors to amino acids*

One of the tangles in the problem of the origin of our protein-synthetic machinery is that, now, to make a protein there must be many proteins already made (cf. figure 3.2). Yet so much of the present day protein-synthetic machinery is RNA, in one form or another, that it is tempting to take a simple line and suppose that there were earlier versions of our protein-synthetic machinery that did not use proteins – that were made entirely of RNA-like molecules or of RNA-like molecules in conjunction with other non-protein materials (Rich, 1965). That way we can dispense with some of the twists in the tangle. We do not then have to suppose that there was ever any other way of making protein.

Of all the present-day involvements of protein in protein synthesis the most critical is in that large set of large and highly sophisticated enzymes that select amino acids, activate them, and then join them to correct tRNA adaptors – the aminoacyl-tRNA ligases. This is precisely the point in protein synthesis today at which the amino acids are distinguished from each other. It is the only point: once a tRNA molecule is loaded up there are no further checks. If it has not been properly loaded then a wrong amino acid will duly be incorporated in some growing polypeptide chain.



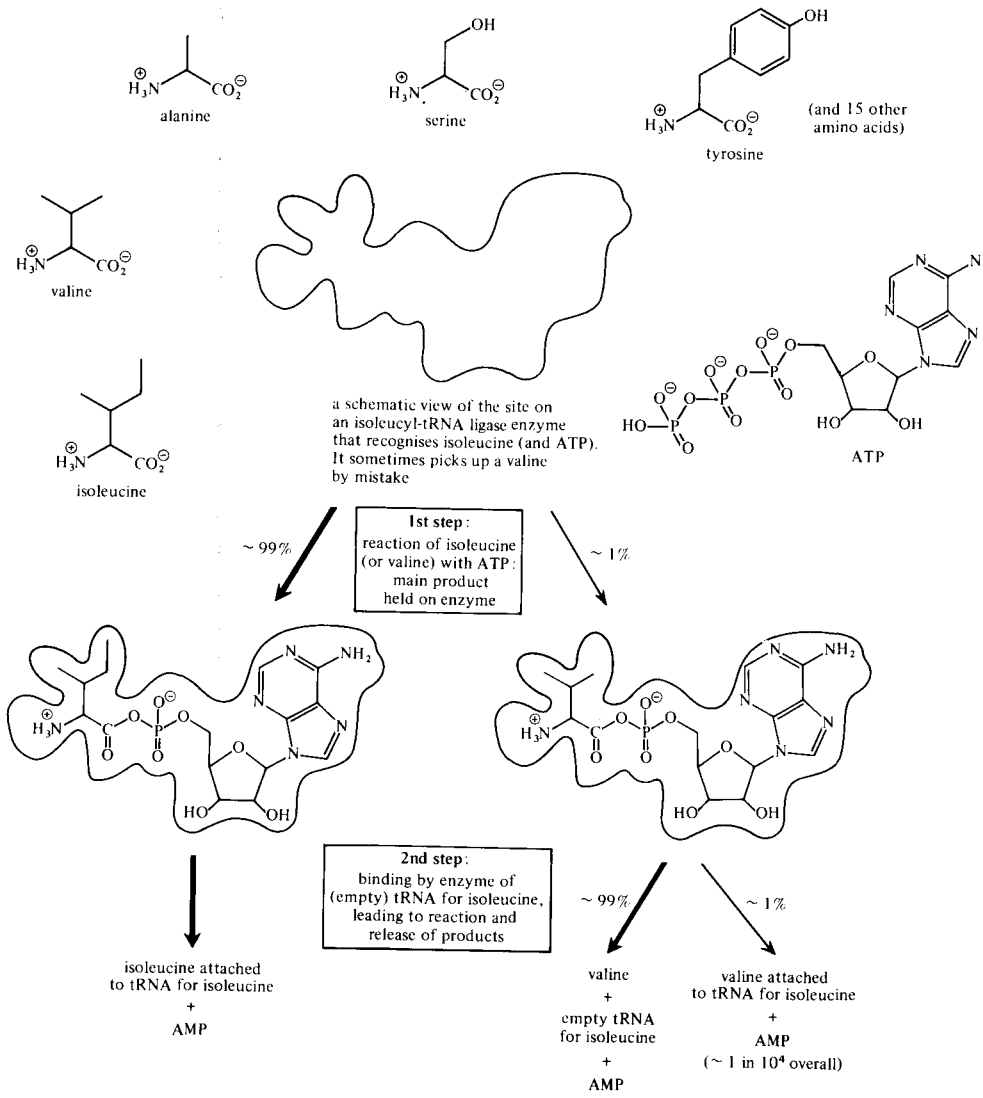



**Figure 9.9. Using two steps to attach a tRNA molecule to an amino acid improves the overall discrimination of the process. The second step acts as 'proof reader'.**

One might try to imagine aminoacyl-tRNA ligase enzymes made out of RNA. There should be no great problem about recognising the RNA adaptors in that case. The difficulty would be to imagine an RNA device that could also distinguish one amino acid side chain from another. It might be that a suitably folded RNA-like molecule could have made a sufficiently discriminating socket (Smithies, in Crick, 1968). In the folding of a long copolymer, information may become, in effect, concentrated in a small active region. But the overall information density in RNA-like molecules is much less than for proteins. And even protein seems only just good enough in some cases. There is a rather elaborate two-step mechanism for joining amino acids to their adaptors which appears to be needed to achieve sufficient discrimination. The isoleucyl-tRNA ligase, for example, makes a mistake every 100 or so times and picks up a valine instead. The second step acts as 'proof reader', almost always hydrolysing a wrongly activated amino acid rather than completing the operation by forming an ester bond with the isoleucine-coding tRNA (figure 9.9).

Perhaps pre-protein 'ligase machines' were similar in this respect to the modern enzymes in using series of events of relatively low discrimination to achieve a sufficiently high discrimination overall. This would be conceivable for evolved organisms of the kinds that we have been thinking about that can set up production lines, pump solutions through tubes, and so on. But one can also think of other ways in which a pre-protein biochemistry might have contrived to attach the correct labels to particular molecules such as amino acids. Here are four possibilities (that would not have been mutually exclusive).

*1. Evolved specificity* This would have been a forerunner of the way in which today's proteins acquired their discriminatory powers – through successive fine adjustments to a socket of some sort. Genetically determined RNA tertiary structures might have been able to evolve like this to some extent, as remarked on above. In Chapter 7 we imagined evolving crystal defect structures as having possibly been able to engineer binding sites to a limited extent.

*2. Fortuitous specificity* It is quite possible that a surface site on a clay mineral, or a crevice in a folded RNA molecule, might just happen to fit some molecule in the environment – and that this accident was then made use of, whether or not the site was subject to adjustment subsequently through natural selection. It was thought at one time that there might be such a fortuitous relationship between triplets of bases and



corresponding amino acids, but that does not seem to be the case. (It is perhaps too much to suppose that twenty different amino acids should happen to fit exactly with a rather limited set of RNA triplet sequences.) But, nevertheless, one or two fortuitous matchings of some sort may have played a part in the early evolution of the genetic code.

*3. Identification by place* It was an idea introduced earlier that clay organisms would have 'kept tabs' on molecules by keeping them localised. If, let us say, there is an amino acid in this vesicle then it must be glycine because this is the glycine output vesicle. So if glycine is to be attached to a piece of RNA with a particular word written in it then the glycine vesicle should be connected to the vesicle that is making this label. Hence the connection is made, in effect, by the apparatus which need not be capable of individual molecular discrimination (any more than the organic chemist's apparatus). Quite coarse-grained genetic information might thus be able to make the required connections. It might even be possible for the same label (adaptor) to be used over and over. A label could find its way home, perhaps, back to the glycine vesicle, because in or near that vesicle there was a polymer with the appropriate complementary sequence (the original templates for the original labels perhaps).

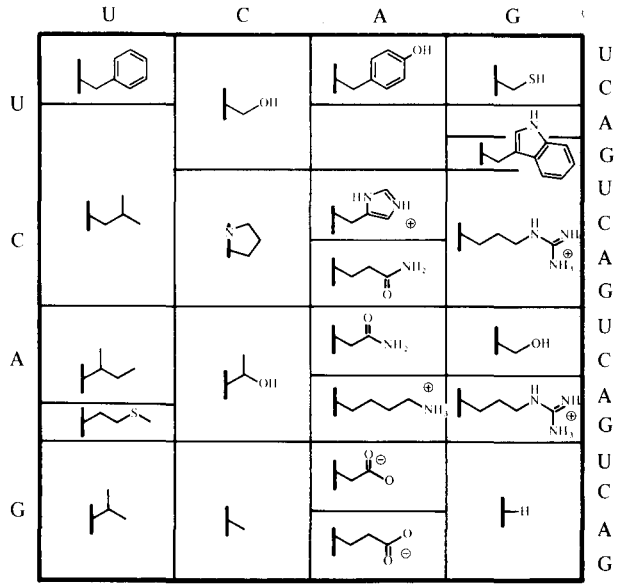
*4. Choice of amino acids* Another way of minimising problems of recognition is by limiting the number of kinds of objects to be recognised – and by making them very different from each other. Of course if you have both isoleucine and valine in your system you may have difficulties telling them apart; but the first systems for making long polypeptides need not have used either such a large or such an *ambiguous* set of amino acids as we now have. Here is another enormous advantage in having well evolved organisms as the source of amino acids rather than a primordial soup. A soup would contain a very ambiguous set of molecules (Chapter 1).

### *Genetic codes*

The present genetic code, that relates particular mRNA triplets to particular amino acids, is shown in figure 9.10. The question of the origin of this code had been well discussed already by the late nineteen sixties (for example by Woese, 1967; Crick, 1968; Orgel, 1968). Several of the ideas put forward then will be used in the following discussion. The main difference will be in the setting that is assumed – not a primordial soup.

| FIRST POSITION | SECOND POSITION U | SECOND POSITION C | SECOND POSITION A | SECOND POSITION G | THIRD POSITION |
|---|---|---|---|---|---|
| U | phenylalanine | serine | tyrosine | cysteine | U |
| | phenylalanine | serine | tyrosine | cysteine | C |
| | leucine | serine | (T) | (T) | A |
| | leucine | serine | (T) | tryptophan | G |
| C | leucine | proline | histidine | arginine | U |
| | leucine | proline | histidine | arginine | C |
| | leucine | proline | glutamine | arginine | A |
| | leucine | proline | glutamine | arginine | G |
| A | isoleucine | threonine | asparagine | serine | U |
| | isoleucine | threonine | asparagine | serine | C |
| | isoleucine | threonine | lysine | arginine | A |
| | methionine (I) | threonine | lysine | arginine | G |
| G | valine | alanine | aspartic acid | glycine | U |
| | valine | alanine | aspartic acid | glycine | C |
| | valine | alanine | glutamic acid | glycine | A |
| | valine | alanine | glutamic acid | glycine | G |

*a*




*b*

Figure 9.10. Two representations of the genetic code. In (*a*) the letters I and T stand for initiating and terminating signals. (*b*) Shows the structures of the side chains of the amino acids, emphasising a non-random distribution of types.



*Which amino acids first?* It is usual to suppose that the earliest amino acids that were coded for would have been 'primordial': that is to say, they would have been from among that set that we discussed in Chapter 1 that are common products of various non-biological syntheses (glycine and alanine most prominently, but aspartic acid, glutamic acid, serine, valine and leucine might be included). The variety of ways in which these molecules form suggests that they are rather generally chemically accessible. Certainly this would be a point in their favour for inclusion at any time. Why not make use of easily made molecules?

To begin with, in the invention of a system for peptide synthesis, there might have been a good reason. Because what is easily made may not be easily recognisable – which, in this yet protein-less biochemistry, might have been much more to the point.

Ease of synthesis and recognisability are two factors: a third would be usefulness. I do not mean usefulness for the individual amino acids – they must already be useful. But there must be a use too for simple polymers that incorporate the amino acids in question.

*Proproteins* We might ask what the simplest useful **proproteins** might have been like – defining a poly-α-amino acid as (at least) a proprotein if it is both sufficiently accurately specified and long enough to fold, or otherwise 'self-assemble', into some distinct higher order structure.

Following Orgel (1972) we could imagine some very simple first proprotein structures based on just two kinds of amino acid, one hydrophobic and one hydrophilic. Orgel suggested that, with such an alternation, coherent β-structures would tend to form, that is, sheets made from aligned polyamino acid chains in which (here) one surface would be covered with hydrophilic and the other with hydrophobic groups (figure 9.11*a*). These β-structures might then be expected to assemble further into water-dispersible bilayers (figure 9.11*b*). Furthermore, there might be a connection between this and a previously noted feature of our genetic code, namely that there is a distinct tendency for more hydrophobic amino acids to have a pyrimidine (especially U) in the middle position of their code words, while hydrophilic amino acids usually have a purine (especially A) (Woese, 1967; Dickerson, 1971). This point can be seen clearly in figure 9.10*b*. The connection that Orgel suggests here arises from studies of the interactions between mononucleotides and polynucleotides. Without any 'replication apparatus' it would be particularly difficult to replicate polynucleotides in which purines were adjacent. (Because single pyrimidine nucleotides would not stick.) Hence primitive nucleic acids might have been alternating pyrimidine–purine copolymers in which case (given a

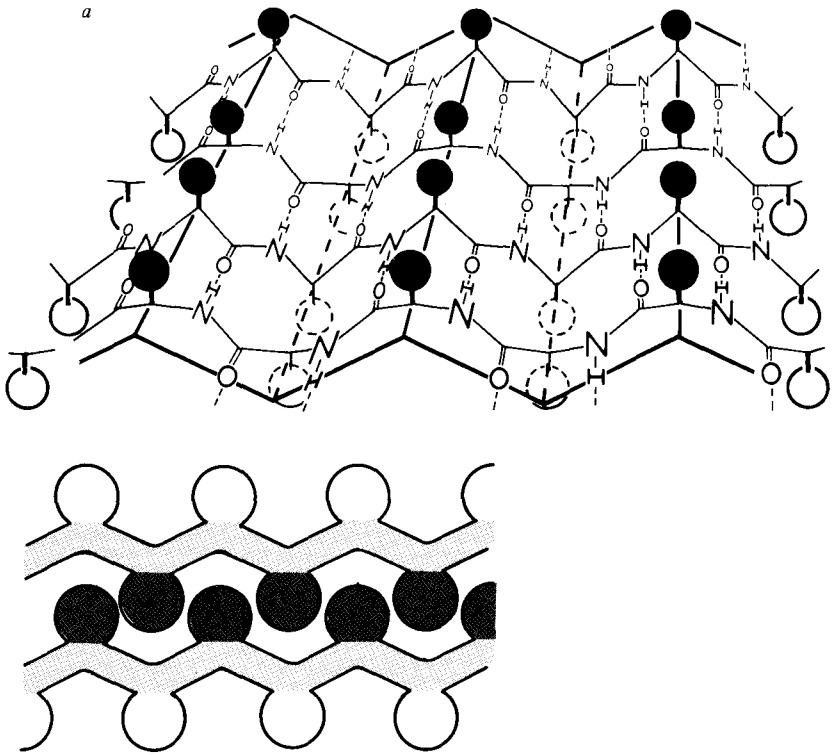



Figure 9.11. (*a*) A $\beta$-structure formed from polypeptide chains arranged parallel to each other and connected by hydrogen bonds. Here the chains are running in opposite directions ('antiparallel'). Alternate side groups point above and below the (pleated) sheet. (*b*) If the black circles and open circles in (*a*) represent hydrophobic and hydrophilic side groups respectively, then pairs of these $\beta$-pleated sheets may form bilayers that are dispersible in water, as indicated in this simplified end view.

similar code to the present one) the early polypeptides would have been predominantly alternating hydrophobic–hydrophilic copolymers.

The tendency for such polypeptides to form $\beta$-structures has been confirmed by a number of studies with synthetic copolymers. For example:

poly(valine–lysine)
poly(leucine–lysine)
and poly(leucine–glutamic acid)

These all form optically clear solutions containing stable, antiparallel $\beta$-pleated sheets – most likely of the bilayer type (Brack & Orgel, 1975; Brack, 1977; Brahms, Brahms, Spach & Brack, 1977). Moderate salt concentrations are generally favourable to making $\beta$-structures, by providing counterions to off-set the mutual repulsion between the charged groups.



Brack reports an interesting effect with montmorillonite and poly-(valine–lysine). The clay becomes hydrophobic, releasing about 90 % of its interlayer sodium ions – the polymer intercalating instead, seemingly in the $\beta$-form.

As expected from the classical $\beta$-structure, these polymers are generally sensitive to substitutions that interrupt the alternation of side chains. And they are sensitive too to chiral purity. Even 5 % of D-isomers in an L-chain noticeably reduces the tendency to form $\beta$-structures. Again this is to be expected from the normal $\beta$-structure where the characteristic up-and-down orientation of the side groups (figure 9.11) depends on the polypeptide being chirally homogeneous.

To our question 'which amino acids came first?' let us take as a provisional and partial answer that there were two – one hydrophobic and one hydrophilic. Let us name them phobine and philine.

*What were the first code words?* Let us accept also what we might call Orgel's First Rule – that the first polynucleotide sequences coding for proprotein were of the general form:

$$\ldots \text{x–Pu–x–x–Py–x–x–Pu–x–x–Py–x–} \ldots \qquad (1)$$

where x is any nucleotide and Pu and Py are respectively A *or* G and U *or* C.

It would simplify our starting point, and make it more realistic perhaps, if we follow the idea (e.g. Crick, 1968) that the very first prototypes of this new way of specifying non-genetic polymers was based on polynucleotides with two rather than four kinds of bases. It is not that the organisms were incapable of producing many different kinds of nucleotides by this stage; but for any novel machine that depends to any extent on a fortuitous coming together of components, the fewer such components the better. Since U-centred and A-centred code words show the hydrophobic–hydrophilic dichotomy most clearly (figure 9.10*b*) we can, for the sake of discussion, choose those two. A typical sequence then becomes:

$$\ldots \text{x–A–x–x–U–x–x–A–x–x–U–x–} \ldots \qquad (2)$$

where x is A *or* U. There are sixteen possible sequences like this, although if we accept for the present Orgel's Second Rule – the preference for Pu–Py alternations – then there would only be one possibility:

$$\ldots \text{A–U–A–U–} \ldots \qquad (3)$$

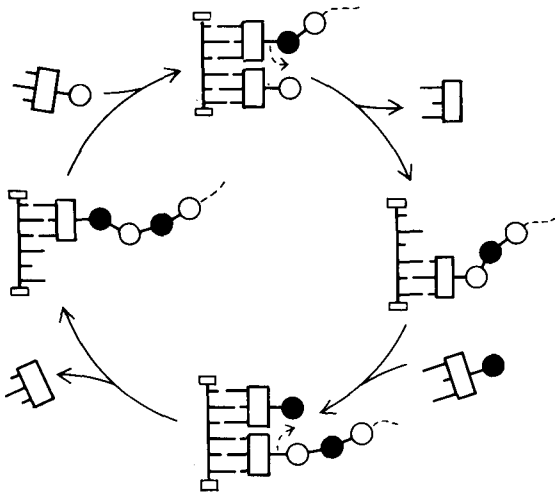


Figure 9.12. A speculative reciprocating mechanism for a (protoribosomal) peptide synthesis avoiding the need for translocation. A very short fixed 'message tape' generates an alternating copolymer of indefinite length.

This is not a very interesting sequence admittedly, and with no potential variability (except in length) it is hardly a 'message' at all. But being generally self-cohesive it could have been useful in itself; and as a first 'message tape', for the synthesis of a first proprotein, it would have had the advantage at least of being unambiguous in two important senses: the same pair of code words alternate wherever you start reading – and whether you read a + or − strand. Furthermore, molecules incorporating such strands would tend to fold on themselves. This might have provided, perhaps, a pair of suitable adaptors for attachment to phobine and philine. With such adaptors and the . . . AUAUA . . . tape, a sequence of operations equivalent to those of the modern ribosome (figure 9.7) would now produce poly(phobine–philine). But the sequence of operations might have been different to begin with. Orgel suggested that peptidyl-tRNAs might sometimes have detached themselves from the message tape and reattached elsewhere on the tape. In that case the tape would still have gone on making the same polypeptide – and quite a short message could have been enough to make a long chain. Alternatively, the rather difficult looking step of translocation might have been avoided to begin with, as illustrated in figure 9.12.

Neither of these devices would be very good for making proproteins of controlled length, so they might soon be superseded; but the evolution of



such simpler modes could nevertheless have helped to perfect some of the basic machinery (tRNAs, etc.) that would be required for those later systems that could make alternating polypeptides of defined lengths by running along given stretches of polynucleotide.

*Why a triplet code?* Our putative table of code word assignments now reads as follows:

| | U | A | |
|---|---|---|---|
| U | – | philine | U |
| | – | – | A |
| | – | – | U |
| A | phobine | – | A |

The dashes stand for 'not applicable' or rather 'not replicable'. We are assuming a triplet code from the outset – for simplicity of discussion more than anything else, because at the very beginning, for certain very simple 'message tapes', one might imagine an easy evolution from one word length to another. For example, our . . . AUAUAUAU . . . can be read not only as an alternating sequence of three-letter words but of words with one letter or five letters – or indeed any odd number. But it is probably fair to say that a triplet code was settled on fairly early. Formally it could have been a singlet code if there were only two amino acids and two nucleotides; but, mechanically, at least triplets might have been needed – simply to get the adaptors to stick to the message tapes. One might guess that a triplet coding system was selected because, with the protoribosomal apparatus available early on (when the decision was made as it were) a triplet of bases was the shortest sequence that would locate properly.

*Code word multiplicity* There would be much to be said for an immutable alternating 'message tape' (of some sort) if all that was required was an alternating polyamino acid. Then no code words other than, say, AUA and UAU would ever appear, to confuse things. Such a set-up could still evolve in so far as the choice of which amino acids were alternating was made by apparatus that was specified by existing genetic systems – the apparatus for making the amino acids and for attaching them to adaptors. More significantly, very much of the expertise now embodied in our protein-synthetic machinery could have begun to appear already at this stage under selection pressures provided by the usefulness of an accurately alternating polyamino acid. For example, the triplet code idea and a

translocating machine of some sort might have been arrived at by organisms that had a use for an accurately alternating polyamino acid of defined length.

But of course we know that the future lay with more interesting sequences. Several things would be needed for such an adventure. The 'replicase apparatus' would have to be competent enough by now to print accurately more complex sequences. This would provide extra coding capacity – and a number of new problems. Every so often mutations would produce non-code words that would block proprotein synthesis because there were no adaptors to fit these words. There would be strong selection pressures at once to make adaptors that would fit any of the words that were now bound to turn up – and to attach each to *some* amino acid to prevent blocking or premature termination of proprotein synthesis (cf. Sonneborn, 1965). The shortest step would be to assign existing amino acids – in our model phobine and philine – to the new adaptors.

*The middle letter* It is clear that in principle (and however long the code word) only one definitive letter in a code word – say the *n*th letter – would be needed to translate between a two-nucleotide message and a two-amino acid proprotein. There might be a convention, for example, that if the *n*th letter is A then it is a philine code word (whatever the other letters are). Similarly, the *n*th letter is a U in any phobine code word. And we might add that it seems to be the case that the middle letter of some set of early (triplet) code words was chosen to be definitive – to account for the hydrophobic–hydrophilic dichotomy in our genetic code referred to earlier (figure 9.10). Perhaps the middle letter was chosen because it was energetically the most critical for locating purposes.

The table of code word assignments would now have read:

| | | U | A | |
|---|---|---|---|---|
| U | | phobine | philine | U |
| | | phobine | philine | A |
| A | | phobine | philine | U |
| | | phobine | philine | A |

The big advance was that now phobine and philine did not always have to be strictly alternating. The polynucleotide sequences were now real messages able to define a more subtle and varied range of phobine–philine proproteins.



*Amino acid multiplicity* Now suppose that a mutation occurs (in the clay genetic information) that causes an organism that has been faithfully producing pure L-philine to make a product that contains L-pseudophiline as well. The adaptor ligase apparatus sees little difference between the two, so this new amino acid becomes more or less randomly substituted for philine in the proprotein being made. For some purposes this might not make much difference – particularly if philine itself is still in large excess. Indeed for certain structural purposes it might be a positive advantage for a proprotein bilayer to be softened up a bit by an occasional anomalous amino acid.

Nevertheless it would always be better to have such things under control: to be able to insert a pseudophiline at genetically determined positions. Such an outcome could be approached gradually through the evolution of an increasing discrimination in the adaptor ligase apparatus. We might suppose, for example, that adaptors that had A-starting, U-centred anticode words would be more easily attached to philine, while adaptors with U-starting, U-centred anticode words would be more easily attached to pseudophiline. That might be due to direct recognition of the newly definitive base in the adaptor. But not necessarily. There could be other more prominent differences between the two kinds of adaptors (remember that we are talking about organisms that could evolve such differences).

The difficult part for the adaptor ligase apparatus would have been in discriminating between the amino acids. This would have been a point in favour of philine and pseudophiline being rather distinctly different if they were to be able to acquire their own code words. On the other hand, as Crick (1968) points out, a similar amino acid would have a better chance of taking over a code word because such an amino acid would be less functionally disturbing to the proteins into which it would be incorporated. The balance between these two contrary factors will be discussed shortly.

*Identifying strands, starts, stops and phases* Now that the ... AUAUAUAU ... tapes were being superseded there would be several new problems of identification. The complementary strands of replicating polynucleotides would not now generally be the same and nor would the messages read the same wherever the reading started. How would the protoribosomes know a message tape when they saw one? How would they know roughly where to begin reading it – and when to stop? How would they know which bases represented the beginnings of words so that the messages could be read in phase?

These are serious problems that are solved nowadays very largely by

means of protein machinery. One might suspect that before there was protein the message tapes had to be a good bit more explicit in themselves – that they contained very clear internal markers.

One way to be clear would be to use another pair of bases exclusively as markers, say G and C. For example, suppose that in the coding strand the third position of the code words was always G. The coding strand would be immediately identifiable and so would the positions of the code words on it. Starts and stops might be signalled by some break in this patterning. It is a well noted feature of our present genetic code that the third position of the code words has a much smaller coding role than the other two positions. Very often it does not matter at all what the third letter is and, even when it does, either of the purines or either of the pyrimidines will usually do (see figure 9.10). It makes some sense to suppose that this now half-used position used to have a vital significance – although not in a coding role.

*Extending the vocabulary* By now our imaginary evolving code table reads as follows:

| | U | A | |
|---|---|---|---|
| U | phobine | philine | G |
| A | phobine | pseudophiline | |

The maximum capacity here would be for four amino acids only, with the third position as a 'semicolon'. But it might not be long before G's and C's started to leak into coding positions on the message tapes; at first this could increase the number of redundant code words for the existing set of amino acids, and then it could let in new amino acids – through some mechanism of the kind already described.

Making the rather arbitrary assumption that G leaked first (and that a code word once assigned never changed) we might imagine the following code table at some stage:

| | U | A | G | |
|---|---|---|---|---|
| U | Leu | term. | Try | G |
| A | Met (init. ?) | Lys | Arg | |
| G | Val | Glu | Gly | |

Such a varied set would fit with the idea that the constraint of recognisability was more important in deciding which new amino acids were to come in than that of not upsetting too much the existing proteins (page



403). (In view of all the assumptions that have been made we should not take this table too seriously. Its main virtue is in its variety. There *is* a great variety among our protein amino acids as well as similarities between many of them: there must have been some way for that variety to have become established.)

After that, on our scheme, C was introduced into the coding strands. (This was now possible because there were now other means of distinguishing between coding and non-coding strands than by their having or not having G's or C's in them.) To begin with, we might guess, C's were mistaken for U's, so that the new C-centred code words were assigned at first to incumbent hydrophobic amino acids. They were later to be differentiated, but by now it was much more important not to upset the existing proteins too much. So this time the differentiation was more marginal with the new amino acids more like those already there. Nevertheless it took place, and under conditions that were still, it seems, quite fluid. Seemingly it was almost completed as far as the 'semicolon' system would allow. At least, to judge from present assignments, there was eventually little redundancy in the table:

| | U | C | A | G | |
|---|---|---|---|---|---|
| U | Leu | Ser | term. | Try | G |
| C | Leu | Pro | Gln | Arg | |
| A | Met/init. | Thr | Lys | Arg | |
| G | Val | Ala | Glu | Gly | |

Finally the possibilities were multiplied by four as the code word marker became unnecessary. The third letter started to acquire a coding significance. In two cases A and G came to be distinguished in this position, and in seven cases a pyrimidine could be distinguished from a purine (cf. figure 9.10). But the evolution of our genetic code never went further than that. For some reason it got stuck.

### *Under what circumstances can a genetic code evolve?*

Other accounts of the origin of our code have been given (for example, Hartman, 1975b; Eigen *et al.*, 1981). A frequent theme is that to begin with message tapes had a low information capacity (but were relatively easily replicated and read), and that they then transformed into tapes with much higher information capacity (but which needed specifically evolved machinery to deal with them).

The crucial point, though, is still more general – and rather generally

agreed. Surely the code *evolved*. It was built up in stages of some sort. *First* there must have been organisms, *then* the evolution of the code.

It has been suggested indeed that our code is designed to minimise the effects of mutations (Dickerson, 1971) and/or errors in protein synthesis (Woese, 1967). In that case, if our genetic code is in any sense optimised, there must have been innumerable trials and experiments with different arrangements. But in any case, in violent contrast to the present staid state of affairs, there must have been a time when code word assignments were more or less fluid. How could that have been?

Crick (1968) insists that code word alterations would only have been possible within organisms that used 'a small number of proteins and especially proteins which were somewhat crudely constructed'. Once there were more than a few proteins present it would become increasingly difficult for a code word to change its meaning – too many proteins being arbitrarily affected. Hence sooner or later the evolution of the code would get stuck.

If we accept this view then the code's evolution must have been virtually complete within a line of organisms for which protein was not nearly so important as now. We have to suppose that even quite late on, by which time the machinery must have been already quite complex, all this machinery was devoted to the production of a relatively small number of proteins. Perhaps some of these became involved in the protein-synthetic machinery during the later phases. But such help could not have been very great: as soon as the protein-synthetic machinery came to depend on more than a very few proteins, the evolution of the code would have stopped. Is it too much to suppose the protein-synthetic machinery was evolved under selection pressures that favoured just a few, perhaps crudely constructed, products?

Not necessarily: a material does not have to have multiple central uses to be vital nevertheless to a particular organism. The colour of a flower's petals, or the stretchiness of a spider's silk, or the olfactory attractions of a female moth might in each case depend on just one kind of molecule that has nevertheless been elaborated under strong selection pressures. To get such things just right may be critical to the particular success of an organism and worth a lot of trouble in synthesis.

So we should not be supposing that protein was a sort of universal invention of all life on Earth, even if it did become universal. It seems more sensible to suppose that it arose in some subclass of organisms, perhaps quite a small subclass with special requirements and for which the elaboration of a few key proproteins was well worth the trouble.



On the line that we were pursuing before, even the earliest proprotein need not have been crudely made in the sense that sequence was unimportant; but it must have been true, if the code was to remain fluid, that the nature of the amino acids present in the proprotein did not affect function too critically. $\beta$-Structure-forming materials seem to be like this. Poly(leucine–lysine) for example has rather similar physical properties to poly(leucine–glutamic acid) (Brack, 1977); and Brack & Orgel (1975) comment on two natural silks – one with an alanine–glycine alternation and one with alanine and glutamine in roughly equal proportions (presumably alternating). Often, it seems, it is the alternation that matters more than the detailed character of the amino acids that are alternating. With materials like that there could be selection pressures to define (simple) sequences very carefully while allowing, sometimes, quite major amino acid changes. The unthinkable for a modern organism – a changeover (in all the proteins) from lysine to glutamic acid – would be quite conceivable for structural materials of this sort, especially if only one or a few proproteins were present in the organism.

It was under such circumstances, perhaps, that the early 'big decisions' were made about the general types of amino acids to be used – hydrophobic, positively charged, negatively charged, large polar, and so on. The recognition problem might have encouraged at this stage a 'contrasty' kind of code book such as that on page 405 (compare discussion on page 397). After that the changes and new introductions were less radical.

There would always be a tendency for lines of organisms to evolve multiple uses for their proteins, and we might be tempted to think that the first organisms to do so were on the way to greater things. But quite the reverse – they would have put themselves out of the race. The code would have stopped evolving in those lines, as it stopped evolving eventually in every line. The future lay with the late developers, with lines that had kept going for a long time with proteins of a limited type, so that a complex code had had time to appear.

### *The new technology*

By now we are imagining that there are many codes that have evolved in different lines of descent from those organisms that first made simple proproteins. In each line the code is fixed, which keeps each line separate. With hindsight we know that only one line is going to win – they all lie before the last common ancestor of all life on Earth (see figure 3.9 on page 103).

The winner would have been one that had gone quite far before its code became frozen in. More precisely it had evolved a rather sophisticated code book that would allow, later, the synthesis of such a variety of proteins that all clay machinery could be dispensed with.

That was the way the revolution was to end, but the final dynamic phase could only have started after the code, our code, was virtually fixed in our ancestral line. It could only have been then that protein began seriously to displace clay – as an alternative material for the membranes, catalysts and so on, on which organic synthesis depended. (This follows from the argument given earlier: when protein started to diversify very much the code would have frozen.)

*Why? How?* The excellence of protein is hardly in doubt – we are all living proof of it. So perhaps it is hardly necessary to argue further for the superiority of protein over clay synthetic machinery. In a nutshell, the new technology would allow a more intricate control.

The 'why' questions raise no great difficulties then. But what of the 'how' questions? How would protein get a foothold? However sophisticated its potential, the first fortuitous unevolved protein version of any piece of machinery would be a poor competitor for the already well evolved clay equivalent – or so it would seem.

*The genetic view* Let us recast our thinking for a moment in terms of genes. We can think of any of our clay–organic organisms, as we can think of any other organism, as a set of genes, a community living in a symbiosis: each gene generates effects that improve the survival chances of the community and hence of the gene itself. In the organisms that we are thinking about, most of the genes are still clay, specifying clay machinery, and through that the synthesis of organic molecules. Some of the genes are nucleic acid. These depend largely on the clay community to maintain monomer supplies, etc.; but they also help the community as a whole – now very largely through this new stuff, protein.

We know what has to happen next. With time, over many generations, the nucleic acid genes in the community must tend to become more numerous while the clay genes become less numerous. It is necessary that, on average, mutations in the nucleic acid genes are more likely to catch on.

In part there is nothing exceptional about such an idea. It is the way evolution works still – genes come and genes go, and some classes of genes are more actively evolving than others. The difference is that for these



organisms there are very different kinds of genes in the community, genes that are made differently and work differently – and there is an overall trend in favour of the more recently invented kind (the kind that can specify proteins).

*Some threads from Chapter 3* There are two ways in which a subsystem, and the genes specifying it, can be radically changed through evolution. The first is the most generally recognized. It depends on a succession of structural modifications. This was how, for example, the fins of fish transformed into the limbs of land animals. Underlying such transformations there must be corresponding successions of genetic change. Clearly such a mechanism would be no good for changing a clay gene into a nucleic acid one. But, as discussed in Chapter 3, there is also an evolutionary mechanism that depends only on functional connections. Such a mode of evolution is not so much a transformation as a replacement or takeover. The gills of fish were replaced by the lungs of land animals. True, there were long successions of small modifications in this – part of the oesophagus transformed into lungs – but, nevertheless, the element of takeover broke the overall connection: gills and lungs needed to have no structural relationships. Clearly *this* is the sort of way that nucleic acid must have usurped clay.

The key ideas in understanding such radical evolutionary processes are preadaptation, functional overlap and redundancy – as discussed at length in Chapter 3.

First consider preadaptation: it would not be at all odd if, say, a $\beta$-structure bilayer that had evolved under selection pressures as a barrier should happen to find other uses – say to orientate acidic and basic groups so as to give some kind of catalytic activity. That sort of thing – preadaptation – is part of normal procedure in evolution. Again and again a feature – an organ, a tissue, a molecule – that has evolved under one set of selection pressures has turned out to have other uses.

Then again functional overlap is commonplace. There can *easily* be a place for an incompetent beginner beside the evolved expert. This may be because the beginner is not actually doing quite the same thing. Part of its function may be similar – there is an overlap perhaps – but there is also an extension of function. As for example when a fish gulps air. This is a very unsubtle way of picking up oxygen when compared with the beautifully designed gas-exchange systems of the gills. But there are times when air gulping is what is wanted. Crude techniques can have their uses – under special circumstances or as a back-up – and hence a crude technique can

get a foothold. Once there, if its evolutionary potential is greater in the circumstances, then the crude technique may transform (this time through a succession of small structural modifications) into a highly sophisticated one that makes the original technique redundant.

*Updating the control of metabolism*

*A step-wise mechanism* Suppose that an organism contains organelles that are specified by clay genes and which bring about the reaction sequence:

$$A \rightarrow B \rightarrow C \rightarrow D$$

There are many of these organelles in the organism – perhaps 100 or so. D is an amino acid that has to be well made – one enantiomer and free of impurities – because it is being used for protein synthesis. But D has other uses as well, and for some of these a racemic, indeed semi-chaotic product that is merely rich in D is still effective. This gives a chance for an alternative design for a percentage of the organelles in which one of the steps – say $B \rightarrow C$ – is catalysed by a Mark-I protein catalyst specified by nucleic acid. The new catalyst is a fairly complex $\beta$-structure that had evolved for another use but which happened to have some $B \rightarrow C$ catalytic activity. Although clumsy, it is perhaps faster or can work at a pH at which the clay machinery tends to dissolve. Hence a new subclass of D-making organelles appear and they are located close to where their product is to be used. The original wholly clay-specified organelles are still the source of D for protein synthesis.

But protein now has a foothold in the organism as a $B \rightarrow C$ catalyst. This catalyst is now improved under natural selection so that C is not only made faster, and under a wider range of conditions by the protein, but it is made more cleanly than ever before. The protein catalyst had been able to evolve a much higher specificity than the clay catalytic surfaces, and most of the clay separator equipment could now be dispensed with. Like a single silicon chip replacing a maze of wires, capacitors, valves, etc., this now perfected piece of protein technology could replace much cumbersome machinery. The engineering, although as complex, was on a smaller scale.

The modified $A \rightarrow D$ organelles could now displace the original ones even as the source of D for protein synthesis. Eventually, through other similar series of events, the whole train of reactions $A \rightarrow B \rightarrow C \rightarrow D$ is taken over and the $A \rightarrow D$ organelle is simplified to hardly more than a bag of three enzymes with primary structures specified wholly by nucleic acid.



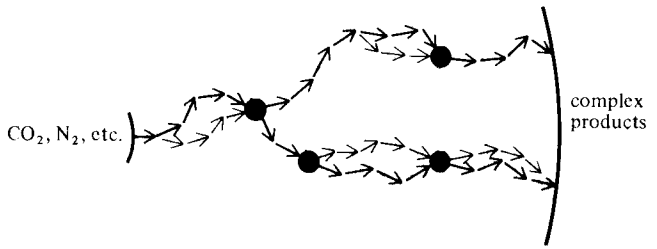



Figure 9.13. See text.

*A stage-wise mechanism* Alternatively, we might imagine evolutionary excursions that started from a quite new set of organelles that were specified by nucleic acid from the start. The metabolite (or environmental molecule) A is being transformed within this new class of substructure to make a variety of (at first) semi-chaotic products. Some of these are useful – and a variety of pathways are built up, in different lines of organisms, according to the forward evolutionary strategy given earlier (figure 8.25).

D is among the molecules that are sometimes being hit on. Since D is already useful in a number of ways, it is particularly likely that this will make a useful contribution as some kind of back-up or alternative mode of synthesis. Selection pressures then further improve the efficiency of the $A \rightarrow D$ reaction train in this new class of organelles – which eventually outstrip the old ones.

This stage-wise story is similar to the step-wise one except that here there would be no need for the whole route $A \rightarrow B \rightarrow C \rightarrow D$ to have been copied. The updated control technique might use a different pathway.

*Starting points, goals and stepping stones* Figure 9.13 (heavier arrows) might represent synthetic activities of an organism using clay-specified machinery to transform a set of environmental starting materials (left-hand side) to a set of end products (right-hand side). In addition there are numerous 'stepping stones' – intermediates, such as glycine and aspartic acid, that are needed for nucleic acid replication and protein synthesis. With the code now fixed such stepping stones would represent fixed points for the future evolution of any line of organisms that was to retain the nucleic acid–protein machinery. These points, as well as many of the end points, would have to remain even if routes between these points were altered (lighter arrows in figure 9.13).

In spite of these restrictions on what could happen there would have been a considerable freedom as to the order of events. Broadly speaking,

it would be an advantage for any of the steps or stages to be updated in any order. Whereas the original pathways under their original control had had to be built up forwards, this takeover of control could have been piecemeal.

*The place of lipids*

Not only enzymes but some kind of membrane-forming material would have been needed before clays could have been dropped entirely. We discussed earlier the suggestion that protein was one of the original organic membrane-forming materials (as $\beta$-structures). Even today protein forms an essential part of the most important kinds of biological membranes (as reviewed by Wickner, 1980). But lipids are also essential components of these modern membranes, and one might guess that the protein–lipid membrane had evolved in its essentials before clay could be dropped.

Again, although the details are difficult to make out, there is no great paradox here. One kind of barrier can stand in for another even if made of a quite different material. One could imagine gradual transformations. A purely inorganic barrier (for example as in figure 8.24) is first lined with polypeptides, etc., and then the inorganic parts become less important and then, in some cases, disappear. Particularly for more subtle functions – for selective barriers, or for well insulated energy-transducing membranes – one might suppose that clay–organic structures would often be better than pure clay structures. Then it would be the old story: because the organic components were subject to more precise specification than the inorganic ones, there would always be a tendency for organic structures to be preferred. And one might guess that it was very largely the invention of lipids that made a final step possible – to purely organic structures.

*Which lipids first?* Of the two main classes of lipids there are some indications that the polyisoprenoids (for example chains like those in figure 8.16), although more complex, are of more ancient origin than the polyketides (for example the straight chains of lipids such as that in figure 8.2).

One of these indications is in the important and somewhat diverse roles of polyisoprenoids in energy transducers: carotenoids are among the photon-catching pigments; plastoquinone is part of the electron-transporting system of photosynthesis, as are the similar ubiquinones in mitochondria; and then the 'handle' on chlorophyll is a polyisoprenoid structure (see figure 8.16). The 'handle' used in bacterial cell wall biosynthesis (page 375) is also polyisoprenoid, and so are the hydrophobic side chains



on some tRNAs. One might suppose that these various structures were invented in the first place to locate in lipid membranes at a time when all lipids were polyisoprenoids. In support of this it has been found that membranes of bacteria that are regarded for other reasons as belonging to ancient classes are often rich in polyisoprenoids (Oró, Sherwood, Eichberg & Epps, 1978; Woese, Magrum & Fox, 1978).

Perhaps polyisoprenoids were, at the time, easier to make. Certainly there is an enormous variety of isoprene-based molecules among modern secondary plant metabolites (Nicholas, 1973), indicating a rather general chemical accessibility. On the other hand, the modern biochemical machinery for polyketide synthesis is particularly sophisticated, involving a multi-enzyme complex and a special carrier protein. It looks as if those seemingly simple straight polymethylene chains are not very easy to put together.

One might suppose, then, that the first organic energy-transducing membranes were built largely from proteins and polyisoprenoid lipids – with perhaps also porphyrins, by now, to help to locate metal ions within the hydrophobic interiors of these membranes.

Again we do not have to assume a sudden switch from one subsystem to another. An organic patch or plug in an otherwise inorganic membrane might contribute in a minor way to the pumping of protons (or whatever) and then gradually increase its role and its extensiveness. Or separately evolved and purely organic organelles might independently contribute to the production of some high-energy molecule, to become later the sole source. After all if the function of organelle A is, say, to use photons to make ATP from ADP, then it can in principle be displaced by organelle B that does the same thing. The inner workings of A and B are not the main point – and might be altogether different. In so far as a subsystem is a module having simple connections with the rest of the system, its replacement can be easily imagined (cf. page 92) – even its gradual replacement, in the kinds of ways that we have discussed.

In figure 8.33 (on page 370) I implied that it was the discovery of organic energy-transducing membranes that allowed, at last, the clay machinery to be left behind. We might now put 'lipids' (meaning isoprenoid lipids) quite late in the sequence, porphyrins later still, and polyketides at some stage after the revolution was all over.

### The conservative consequences of revolution

Sooner or later the circle was complete: organisms appeared that used only the nucleic acid–protein–lipid control machinery to make all the components needed for that machinery. Eventually free living organisms appeared that had no clay in them, and eventually too these included autotrophs. With that the connection was made between very simple environmental atom sources, $CO_2$, $N_2$, etc., and a minimum but inevitably complex set of components – built from what we now call 'the molecules of life'.

There was something quite new about this way of living. Up to then at least some of the monomers for the central control machinery – the units for clays – had been provided by the environment. Now much more complex organic monomers had to be made and activated. But no longer was life tied to weathering solutions, and to the time scales of clay synthesis. These less fettered organisms could now evolve faster and further to fill a whole new range of niches. They were still somewhat tied to the ground, as life still is, in the need for phosphates and metal cations; but mostly the atoms for life were now taken out of the air.

As with so many revolutions there were conservative consequences. Now there was no simple way of living for these organisms and no simplifying evolutionary paths. They had lost the scaffolding on which they had been built. Now at their centre they had a minimum fixed complex machine that was to be inherited by that organism that was to become the last common ancestor of all life on Earth. Biochemistry was still to evolve, but only outward, now, on the basis of what had been fixed when the clay disappeared.

# Coda

## How the problem will be resolved

*The method of overlap* A number of general considerations taken together can often lead to surprisingly particular conclusions. For example in the game of '20 Questions', where the aim is to discover some particular object that is being thought of, it is actually a mistake to start with particular questions. Rather you should start by finding out whether the unknown object belongs to certain large sets. (Through questions such as: Is it in this room? Can you eat it? Is it green? And so on.) By overlapping such sets you can quickly narrow the field.

This way of playing '20 Questions' represents a very general strategy of enquiry. Western science started with the asking of general questions – it started in philosophy. That can still be the right starting territory for new or immature fields, although now there is a vast stock of general ideas available – laws, principles, rules, recurring observations – that have been generated by past scientific enquiry.

For most sciences the game is well advanced, and for most scientists the crucial questions are of a particular sort to be answered by cleverly designed experiments. In the end it must be possible to ask the particular questions; but to insist too soon that only experiment will do as the means of resolving such a remote and complex subject as the origin of life is, to my mind, to misplay the game of '20 Questions' – and quite possibly to miss the chance of solving the problem at all.

This book has been mainly an attempt to look for the place where the answer to the problem of the origin of life may lie – by asking a large number of general questions. There were questions of general biology: what kind of system is it (in the abstract) that can evolve under natural selection? How can evolution give rise to highly co-operative organisation? And so on. At a somewhat more specific level there were chemical ques-

tions – about the most geochemically accessible means of holding and printing large amounts of information, or of making membranes, or of trapping the energy of sunlight in chemical form, or of controlling organic reaction sequences without enzymes. Particular experimental results were often referred to in these discussions, and innumerable observations and experiments lie behind general ideas on, say, the nature of crystal defects, or the characteristics of chemical bonds, or of later evolutionary processes. But our principal preoccupation has been with general considerations, and a region of overlap defined by them.

One of the main conclusions of this book is that the relevant region of overlap is, surprisingly, centred on colloidal inorganic crystalline minerals rather than on 'the molecules of life'. I tried to define this region more tightly by thinking about the sorts of minerals that might have been most suited to genetic and to other functions. As the relevant areas of science advance (as clay structures are known in greater detail, for example) it should be possible to see the critical region more clearly – and more clearly still with the sharper tools of specially contrived experiments.

Alternatively, this region of overlap may be fined away to nothing – if, say, an experiment or a theoretical consideration makes it clear that a crystal gene could not work because no defect structure would be stable enough and at the same time be replicable through crystal growth. These speculations are disprovable in principle.

But perhaps a more likely form of failure for any hypothesis about a remote subject is not through direct refutation, but because the region of overlap fails to sharpen up: the hypothesis remains too vague. Remoteness as such – that is, inaccessibility to the senses – should not prevent such a sharpening with time if the hypothesis is indeed on the right lines. Consider, for example, the question of the existence of atoms. This was a very remote idea for the Greeks, but it was to become a virtual certainty by the end of the nineteenth century – before anyone had in any sense seen an atom. Even hypotheses about remote entities can become very distinct if they continue to work, and especially if they fit in with new and unexpected effects. No, the remoteness of the origin of life on Earth should not prevent an hypothesis about it becoming clearer – although progress may be slow and require the consideration of very many diverse data and ideas.

*Experimental accessibility* In any case it is probably only the historical question that is essentially remote – that is to say the question of how in fact our ultimate ancestor arose. This is a fascinating piece of genealogy, but it is perhaps not the most important question. The important question is rather more general; it is to delineate the set of possible ultimate ancestors. What kinds of systems, among those that could have been generated by geochemical processes on the primitive Earth, could have formed the basis for an indefinite evolution through natural selection? Whatever those systems are like, when you have refined the possibilities sufficiently to be able to make a guess at how they might be made, then your guess should be accessible to experiment. Let me expand a little on this.



We cannot be sure, of course, because God or spacemen or gigantic luck may have put us here. But if you really believe that life arose spontaneously on this planet as a natural outcome of physicochemical processes, then you should expect to be able to re-enact in the laboratory something like the very earliest stages of its evolution. You should expect to be able to make microsystems of some sort that can evolve under natural selection in the fully Darwinian sense.

There are still some provisos. Maybe the experiment would take a million years because some necessary crystal growth process is very slow. Perhaps Nature's original experiments took that long. Perhaps the first organisms had generation times measured in millennia. But Nature had to succeed with materials and conditions that happened to be on the Earth, and commonplace enough to be consistently maintained over very long periods. In the laboratory we can contrive conditions: we can control solution concentrations, adjust temperatures, choose exotic materials for models, and so on. In the laboratory we may lack time, but we can set up and maintain situations whose *a priori* probability in an unevolved world would have been vanishingly small. As discussed in Chapter 5, you do not have to do very much to arrive at the point at which the contrivance of your experimental set-up is far beyond what Nature might have been expected to arrive at (never mind maintain) even given two billion years. Well, Nature *did* contrive the first evolving systems and took less than two billion years to do it.

So I think we should be able to make a primary organism. The difficulties here are not at all like the difficulties of synthesising, say, *E. coli* from scratch. The difficulties in making an organism built along modern lines are technical and enormous – but for *E. coli* we would just about know what to do. By contrast, the technical problems of making something like our ultimate ancestor cannot be very great if Nature overcame them without natural selection. In this case the main problems are theoretical – of knowing *what* to do. When we know that, the experiments may not be so very difficult.

Whenever and however such experiments succeed they will then presumably prove successful under a range of conditions and with a range of somewhat different materials. (To suppose otherwise would be to substitute the postulate of a fluke event by that of a fluke Universe in which there was only one rather precise way of bringing off an origin of life.) So I think that when we are near to the answer we will find many similar answers, and so we will not be able to say exactly how life must have started on Earth. The genealogical problem will not be solved precisely. But the main problem may be effectively resolved all the same – when we have a clear view of a set of possible solutions and experimental demonstrations of some members of that set.

How close are we to this adventure? Weiss (1981) is already well embarked: he reports experiments that clearly demonstrate hereditary effects during the growth of smectite layers through the mechanism of 'intercalating synthesis'. Here new layers crystallise *between* layers of seed crystals, and in doing so they pick up information in the form of (at least) Al for Si substitution densities. By re-seeding new solutions, the process can be repeated for more than 30 generations before the characteristic charge density of the original seed is 'forgotten'. Catalytic and adsorptive properties can concomitantly be transmitted for a similar number of generations.

If it proves possible to transmit clay-held information accurately enough over many generations – with very occasional mistakes – then an experiment along the lines of figure 7.11 (referred to on page 287) becomes imaginable, with selection operating on rheological properties. As discussed in Chapter 7, samples of clay made up of virtually identical particles (because derived from a single seed) might have some very special properties – 'magical' properties almost – if they had been produced through successive selections of replicating, mutating crystallites.

Less direct experiments would still be needed, though, to help to delineate the set of possible primary genetic materials from among the microcrystalline minerals likely to have been forming on the primitive Earth. For reasons discussed in Chapter 7, attention might be concentrated on well crystallised materials – clays such as kaolinite perhaps – to look for elements of order and disorder. Any kind of domaining within layers or ordering between them is of interest, as would be regularities of overall morphology – the appearance again and again, in a sample, of some rather odd complex shape.

The relevant minerals are clays in the most general sense (page 165), including, for example, microcrystalline metal oxides, oxide-hydroxides and sulphides, as well as feldspars, zeolites, or silica where any of these have formed from aqueous solution.



*Continuous crystallisers* Studies could be done on clays made in the laboratory in continuous crystallisers. These are closely analogous to chemostats (see figure 5.1) and are used for producing crystal suspensions under controlled conditions. In a continuous crystalliser, as in a chemostat, constant conditions can be maintained indefinitely through nutrient (supersaturated) solutions flowing in and product – crystals in suspension – flowing out. Such a device might be used fo discover genetic crystals.

Consider first an ordinary crystallisation process for which the difficult step is nucleation (seeding). At first nothing would happen in a continuous crystalliser, the supersaturated solution would simply flow in and flow out. If one seed was added this might still not be enough if all that happened was that that seed grew bigger. Eventually it would be carried out through the waste pipe. To keep going it would be necessary that a seed crystal not only grew but broke up, and furthermore that the rate of production of new seeds in this way was sufficient to maintain the population. It would be necessary that an average crystallite, during its average residence in the crystallising vessel, had grown and broken up into at least two new seeds.

These properties – seed dependence and break-up – would be necessary but insufficient for replicating genetic crystals. We can take one step further by considering an experiment in which there were two different kinds of seeds that were individually replicable – let us suppose that they were two different polymorphic forms of the material that was crystallising out. As in a chemostat containing mutant micro-organisms with slightly different reproduction rates, if one of the polymorphs grew and broke up even marginally faster it would sooner or later take over *completely*. (After every generation time the slower one would constitute a smaller proportion of the population, and that trend would continue however small the amount of the slower reproducer, which would thus, sooner or later, be wiped out.) Under such circumstances kinetic control is everything: it is not necessarily the most stable type that wins in the end, but the one that reproduces most quickly. Such a set-up would be an improved model for a primitive biological system, but still too limited.

Now suppose that there was some immense number, $n$, of possible kinds of seeds that were individually replicable. In this case the individuality of the seeds could hardly be due to polymorphism, to their each having a distinct crystal structure. Here the individuality would have to come from some characteristic defect structure, some kind of patterning superimposed

on the crystal structure. And of course this patterning would have to be replicated as the crystals grew, and be maintained as the crystals broke up (as discussed in Chapter 7). But the same general arguments would apply: if for any reason one of a set of patterns present initially could reproduce even marginally faster than the others, it would eventually displace the others entirely. Occasional mutations would then allow the most successful modification of the most successful modification of the most . . . to catch on. That way, seemingly highly improbable systems could be arrived at. We would be watching a Darwinian evolution. As discussed in Chapter 3, it need not take very long for reproducing, mutating systems to arrive at exceedingly improbable, exceedingly minute, subsets through successive selections (pages 87–8).

*Looking for clay-based life on the Earth today* Nature may be doing experiments like this for us all the time. Look again at those scanning electron micrographs in figures 6.34 and 6.54 which show clays crystallising from solutions percolating through pores in hard rocks. Near the surface of the Earth there are innumerable regions into which barely supersaturated solutions are flowing; within which solutions are being 'stirred' (by turbulence, back-flow, cyclic flows); within which clays are crystallising; and from which clays in suspension are emerging. Nature sets up continuous crystallisers.

An alternative approach, then, would be to look for crystal genes in Nature. In Chapter 7 we discussed in general terms how this might be done by looking for appropriate morphologies: kaolinite vermiforms were conceivable examples. Do such vermiforms breed? Are there conditions where the growth of such crystals is wholly through elongation in the stacking direction plus break-up?

We might look for signs of a limited kind of evolution having taken place. Looking at kaolinite vermiforms, for example (figure 6.29), we might ask if there were selective factors at work that favoured one kind of thickness and cross-section over another, so that one type became predominant. If so (and if the breeding condition was also satisfied) then such vermiforms would be low-level organisms, able to pass on to offspring a rather particular way of surviving and propagating.

There ought to be such 'origins of life' going on now if conditions on the Earth now are broadly similar to conditions on the primitive Earth. Yóu would not necessarily expect starter organisms to be particularly common, in view of the highly stabilised conditions that would be needed to begin with; and you would not expect the higher, photosynthesising forms to be able to compete with modern micro-organisms. But there ought to be at least the beginnings of evolutionary processes to be found somewhere.



Indeed one would expect that such processes would have gone far enough, in some cases, for minerals to be crystallising in certain environments *by virtue* of their being evolved. That is to say because they have some very particular kind of defect structure that was a product of natural selection, they were able to grow at a significant rate in the environment in question.

One might look for this sort of thing in minerals that crystallise readily in Nature, have rather complex morphologies, and yet seem to have been difficult to make in the laboratory. (That would be because in the laboratory conditions had not been set up to allow the gradual evolution of efficiently breeding seeds.) Kaolinite might fit here; it is one of the very commonest clays in Nature, where it comes in a great variety of forms (Keller, 1978); and yet kaolinite is particularly difficult to synthesise (page 253).

More advanced organisms – the kinds of organo–clay systems that we were discussing in Chapter 8 – would have to be sought as fossils. Objects that look like fossils of modern-type cells, or of modern-type colonies of micro-organisms, have been found in some of the most ancient rocks on Earth. The older ones (from about 2500–3500 million years ago) are less sure, and the microstructures have rather abnormal size distributions (Schopf, 1975, 1976). Conceivably these are cells of an altogether earlier form of life.

Any odd-looking structures in very ancient rocks might be interesting, particularly if similar structures appear in a number of different localities within rocks of similar age. They might be fossils of precellular colonies – fossilised vital mud.

*A singular case* Sherlock Holmes preferred what he called singular cases – seemingly baffling ones. 'As a rule,' he said (in *The Red-Headed League*) 'the more bizarre a thing is the less mysterious it proves to be. It is your commonplace featureless crimes that are really puzzling . . .'

The singular aspect of 'The Case of the Origin of Life on Earth' is in the nature of the first genetic material(s). (How on Earth could information printing machinery have generated spontaneously?) But the very difficulty of this puzzle is to be welcomed, as Holmes would have welcomed it: there cannot be so many solutions. If we can find *any* solution it is likely to be near the mark. Hence the concentration in this book on the nature of the primary genetic materials and, in this short chapter, on experimental

possibilities based on ideas in Chapters 5 and 7. There is a promise here of a direct demonstration of what ought to be demonstrable: a synthetic system that is able to evolve under natural selection. (It would not take many generations to demonstrate this ability – compare again page 87).

In direct contrast to the views expressed by Dr Kritic in Chapter 4, it seems to me that the idea of genetic takeover, by widening the initial possibilities, allows us to consider systems that really might be starter organisms. They might seem biochemically unfamiliar, but their mode of evolution would be Darwinian – not some 'chemical evolution' interposed between conditions that one might set up and evolving organisms that subsequently appear.

*Other experimental approaches* The classical experimental approach to the origin of life (through abiotic organic syntheses) is still very relevant. In relation to the later invention of our kind of biochemistry it is important to understand the chemical paths of least resistance. What kinds of organic reactions take place easily at around ordinary temperatures in aqueous solutions?

Reactions in the presence of clays of all sorts are interesting – although not so much general catalytic effects of clays as effects that depend on particular defect structures. (Compare enzymes: a reaction that was catalysed by just *any* polypeptide sequence could have no critical place in our biochemistry.) It is interesting, for example, that only *some* structural $Fe^{3+}$ in a hectorite was effective in oxidising benzidine and that these processes were very sensitive to pretreatment of the clay (McBride, 1979).

Part III of this book suggests many research possibilities: some extending work that has already been done, others more particularly aimed at testing the plausibility of the story being told in Part III. Let us consider very briefly some other questions that we might want answers to.

In Chapter 7 we took it that one of the first ways in which clay genes would have extended the control of their environment would have been through controlling to some extent the conditions for the synthesis of secondary (non-genetic) clays. This gives us another reason for being interested in the processes of clay crystallisation, especially in the ways in which, in mixed assemblages, one clay can affect the growth of another (through epitaxy, stabilising ion concentrations, binding inhibitors, and so on).

We would like to know more about the effects of crystal size, shape, and defect structure on colloidal properties of clays; about how silicate layers 'self-assemble' into mixed layer or card-house structures; how such layers



fold on themselves; how well they stick to each other, and how much that depends on internal cation arrangements, etc. Adsorption of polyphosphate at layer silicate edges, and of organic molecules at edges and interlayers are other topics of interest. How selective are adsorption sites on layer silicates, zeolites – clays of all sorts? Are they ever chirally discriminating? To what extent are they defect dependent?

And we would like to know more about organic catalysts for clay synthesis. How does fulvic acid work? Do any modern plants engineer the kinds of clay around them by producing substances that catalyse clay synthesis? Could one design an organic catalyst for clay synthesis?

How do organic molecules affect the microporosity of clay assemblages and how does microporosity affect distributions of organic molecules? Are there chromatographic effects when solutions flow through clay assemblages? To what extent can (unevolved, uncloned) clays exert a selective influence on the course of organic chemical reactions so as to limit the variety of products? (Mordenite can have such an effect on the formose reaction under somewhat artificial conditions; Weiss *et al.*, 1979.) What are the roles of defect structures and of compartmentalisation? Can one set up oscillations in open clay–organic reactors?

Some of the elements for evolved energy transducers could be looked for in ordinary unevolved clays. There are the proton potentials that can be generated by temperature gradients (Freund & Wengeler, 1978). There is an active interest in clays as photocatalysts and energy stores (Coyne *et al.*, 1981). There is the whole field of semiconductor minerals which include many in the broad category of clays. There are the charge-transfer spectra of transition metals in clays and the effects on redox potential of the crystal environment of such ions. Then there are synergistic binding effects that might be looked for and which would be relevant to some of the formalised energy transducers discussed in Chapter 8. How does the binding of polyphosphate affect the binding of other things?

Already we might begin to think of fabricating energy transducers by orienting asymmetric clay membranes. Can one produce an asymmetric effect – any kind of potential – by illuminating an asymmetric clay membrane?

Related to devices for catching energy there are mineral semiconductor devices for fixing carbon dioxide and nitrogen. Studies on these lines are already well under way.

Further advances in our knowledge of primary and secondary metabolic pathways will help to establish (or tend to refute) the idea that our primary metabolism, like our secondary metabolism, was built from the centre out.

Evolutionary studies at all levels should give us further insights as to general evolutionary mechanisms, for example the roles of symbioses and of takeovers.

Bahadur and his school (see, for example, Bahadur & Ranganayaki, 1970; Smith, Folsome & Bahadur, 1981) have been studying microstructures – complex budding vesicles that Bahadur calls 'jeewanu' – which are generated when certain solutions are precipitated in sunlight (for example solutions containing molybdate, ammonia, phosphate and formaldehyde). Synthesis of a number of our biochemicals has been claimed. It will be interesting to know how these complex structures form. Is this a case of osmotic construction (page 327)? To what extent is their compartmentalised structure relevant to photosynthetic activity? Perhaps the earliest photosynthesising organisms were a subclass of jeewanu, a kind whose structure was guided to some extent by inorganic crystal genes.

By the start of Chapter 9 we were beginning to think about the kinds of polymers that would have been available for structural purposes once the consistent synthesis of particular chirally homogeneous monomers could be presupposed. Insights into possible first uses for homopolymers and regular copolymers can come, as we saw, from studies of polysaccharides, sugar-phosphate polymers, polynucleotides and 'proproteins' – both natural and synthetic. Could we design molecules for holding together clay particles (as in figure 9.5)? Or could we even invent a 'nucleic acid enzyme' (with oligonucleotide coenzymes to go with it)?

In considering first forms of nucleic acid replication and the evolution of our protein-synthetic machinery, we are coming to the final stages of the evolution of our kind of life. But advances in understanding here could still be relevant to the question of genetic takeover. Such advances should tend to favour or refute the contention that the invention of our nucleic acid–protein control machinery could *only* have been made by organisms, highly evolved organisms, that had had to manage without any such machinery to begin with.

*A useful life?* To come back to the very beginning, there are practical reasons for being interested in biogenesis if indeed first forms of life were mineral and could be recreated. Here would be a new technique of fabrication at the otherwise difficult, colloidal, size level – through artificial selection of replicating defect structures. For example, exceedingly finely engineered semiconductor devices would become available, made from the cheapest possible materials. It is such devices that are perhaps most needed to solve the problem of converting cheaply and efficiently solar energy into other forms.



And there is a curious advantage in this form of engineering. If you use artificial selection to perfect your product you do not have to know in detail how it works (any more than a rose breeder has to know in detail how plants work). The replicating pattern with its special properties can be arrived at with an eye only to its performance. And once you have arrived at this valuable structure you can make as many copies of it as you please.

The existence of life on Earth is the best clue to the existence of this undiscovered field of technology.

How did life originate on Earth? That question will be resolved, if never precisely answered, when we have found out how to make primary organisms of various sorts. The problem of the origin of life on Earth will then have lost its mystery.

# References


Abelson, P. H. (1956). Amino acids formed in 'primitive atmospheres'. *Science*, **124**, 935.

Abelson, P. H. (1957). Some aspects of paleobiochemistry. *Annals of the New York Academy of Sciences*, **69**, 276–85.

Abelson, P. H. (1966). Chemical events on the primitive Earth. *Proceedings of the National Academy of Sciences USA*, **55**, 1365–72.

Adams, J. M. (1978). Unifying features relating to the 3D structures of some intercalates of kaolinite. *Clays and Clay Minerals*, **26**, 291–5.

Albersheim, P. (1975). The walls of growing plant cells. *Scientific American*, **232**, no. 4, 80–95.

Alberts, B. & Sternglanz, R. (1977). Recent excitement in the DNA replication problem. *Nature*, **269**, 655–61.

Alcover, J. F., Gatineau, L. & Mering, J. (1973). Exchangeable cation distribution in nickel- and magnesium-vermiculites. *Clays and Clay Minerals*, **21**, 131–6.

Aleksandrova, V. A., Drits, V. A. & Sokolova, G. V. (1972). Structural features of dioctahedral one-packet chlorite. *Kristallographie*, **17**, 525–32

Alexander, G. B., Heston, W. M. & Iler, R. K. (1954). The solubility of amorphous silica in water. *Journal of Physical Chemistry*, **58**, 453–5.

Allaart, J. H. (1976). The pre-3760 m.y. old supracrustal rocks of the Isua area, Central West Greenland, and associated occurrence of quartz-banded ironstone. In *The Early History of the Earth*, ed. B. F. Windley, pp. 177–89. London: John Wiley & Sons.

Allen, F. J. (1899). What is life? *Proceedings of the Birmingham Natural History and Philosophical Society*, **11**, 44–67.

Allen, W. V. & Ponnamperuma, C. (1967). A possible prebiotic synthesis of monocarboxylic acids. *Currents in Modern Biology*, **1**, 24–28.

Anderson, C. S. & Bailey, S. W. (1981). A new cation ordering pattern in amesite-$2H_2$. *American Mineralogist*, **66**, 185–95.

Anderson, W. L. & Stucki, J. W. (1979). Effect of structural $Fe^{2+}$ on visible absorption spectra of nontronite suspensions. In *International Clay Conference 1978*, ed. M. M. Mortland & V. C. Farmer, pp. 75–83. Amsterdam: Elsevier.

Atkinson, R. J., Posner, A. M. & Quirk, J. P. (1977). Crystal nucleation and growth in hydrolysing iron (III) chloride solutions. *Clays and Clay Minerals*, **25**, 49–56.

Aumento, F. (1970). Serpentine Mineralogy of Ultrabasic Intrusions in Canada and on the Mid-Atlantic Ridge. *Geological Survey Canada, paper 69–53.*

Bada, J. L. & Miller, S. L. (1968). Ammonium ion concentration in the primitive ocean. *Science*, **159**, 423–5.

Baes, C. F. & Mesmer, R. E. (1976). *The Hydrolysis of Cations.* New York: Wiley.

Bahadur, K. & Ranganayaki, S. (1970). The photochemical formation of self-sustaining coacervates. *Journal of the British Interplanetary Society*, **23**, 813–29.

Bailey, S. W. (1963). Polymorphism of kaolin minerals. *American Mineralogist*, **48**, 1196–209.

Bailey, S. W. (1966). The status of clay mineral structures. *Clays and Clay Minerals*, **14**, 1–23.

Bailey, S. W. (1975). Cation ordering and pseudosymmetry in layer silicates. *American Mineralogist*, **60**, 175–87.

Baird, T., Cairns-Smith, A. G. & MacKenzie, D. W. (1973). An electron microscope study of magnesium smectite synthesis. *Clay Minerals*, **10**, 17–26.

Baird, T., Cairns-Smith, A. G., Mackenzie, D. W. & Snell, D. S. (1971). *Clay Minerals*, **9**, 250–2.

Baly, E. C. C., Heilbron, I. M. & Barker, W. F. (1921). Photocatalysis. Part I. The synthesis of formaldehyde and carbohydrates from carbon dioxide and water. *Journal of the Chemical Society*, **119**, 1025–35.

Baly, E. C. C., Heilbron, I. M. & Hudson, D. P. (1922). Photocatalysis. Part II. The photosynthesis of nitrogen compounds from nitrates and carbon dioxide. *Journal of the Chemical Society*, **121**, 1078–88.

Balzani, V., Moggi, L., Manfrin, M. F., Bolletta, F. & Gleria, M. (1975). Solar energy conversion by water photodissociation. *Science*, **189**, 852–6.

Bard, A. J. (1979). Photoelectrochemistry and heterogeneous photocatalysis at semiconductors. *Journal of Photochemistry*, **10**, 59–75.

Bard, A. J. (1980). Photoelectrochemistry. *Science*, **207**, 139–44.

Barker, W. C., McLaughlin, P. J. & Dayhoff, M. O. (1972). Evolution of a complex system: The immunoglobulins. In *Atlas of Protein Sequences and Structure*, vol. 5, ed. M. O. Dayhoff, pp. 31–9. Washington, D.C.: National Biomedical Research Foundation.

Bar-Nun, A., Bar-Nun, N., Bauer, S. H. & Sagan, C. (1970). Shock synthesis of amino acids in simulated primitive environments. *Science*, **168**, 470–3.

Bar-Nun, A., Bar-Nun, N., Bauer, S. H. & Sagan, C. (1971). Shock synthesis of amino acids in simulated primitive environments. In *Molecular Evolution I: Chemical Evolution and Origin of Life*, ed. R. Buvet & C. Ponnamperuma, pp. 114–22. Amsterdam: North-Holland.

Bar-Nun, A. & Hartman, H. (1978). Synthesis of organic compounds from carbon monoxide and water by U.V. photolysis. *Origins of Life*, **9**, 93–101.

Barrer, R. M. (1978). *Zeolites and Clay Minerals as Sorbents and Molecular Sieves.* London: Academic.

Barrer, R. M. & Kerr, I. S. (1959). Intracrystalline channels in levynite and some related zeolites. *Transactions of the Faraday Society*, **55**, 1915–23.



Barrer, R. M. & Sieber, W. (1977). Hydrothermal chemistry of silicates, part 21: Zeolites from reaction of lithium and caesium ions with tetramethylammonium aluminosilicate solutions. *Journal of the Chemical Society Dalton*, pp. 1020–6.

Barshad, I. (1948). Vermiculite and its relation to biotite as revealed by base exchange reactions, X-ray analysis, differential thermal curves and water content. *American Mineralogist*, **33**, 655–78.

Battersby, A. R., Fookes, C. J. R., Matcham, G. W. J. & McDonald, E. (1980). Biosynthesis of the pigments of life: formation of the macrocycle. *Nature*, **285**, 17–21.

Beadle, G. W. (1949). Genes and biological enigmas. *Science in Progress*, **6**, 184–249.

Becquerel, P. (1924). La vie terrestre provient-elle d'un autre monde? *Bulletin de la Société astronomique de France*, **38**, 393–417.

Bennett, C. D. (1974). Similarity in the sequence of *Escherichia coli* dihydrofolate reductase with other nucleotide-requiring enzymes. *Nature*, **248**, 67–8.

Bernal, J. D. (1951). *The Physical Basis of Life.* London: Routledge & Kegan Paul.

Bernal, J. D. (1960). The problem of stages in biopoesis. In *Aspects of the Origin of Life*, ed. M. Florkin, pp. 30–45. Oxford: Pergamon Press. Reprinted from Clark, F. & Synge, R. L. M. (eds) (1959). *The Origin of Life on the Earth*, pp. 38–53. London: Pergamon.

Bernal, J. D. (1967). *The Origin of Life.* London: Weidenfeld & Nicolson.

Bernal, J. D., Dasgupta, D. R. & Mackay, A. L. (1959). The oxides and hydroxides of iron and their structural inter-relationships. *Clay Minerals Bulletin*, **4**, 15.

Blout, E. R. & Idelson, M. (1956). The kinetics of strong-base initiated polymerisations of amino acid-*N*-carboxyanhydrides. *Journal of the American Chemical Society*, **78**, 3857–8.

Blum, H. F. (1951). *Time's Arrow and Evolution.* Princeton: Princeton University Press.

Bodansky, Y. S., Klausner, M. & Ondetti, A. (1976). *Peptide Synthesis.* New York: Wiley.

Bolton, J. R. (1978). Solar fuels. *Science*, **202**, 705–11.

Bondy, S. C. & Harrington, M. E. (1979). L-amino acids and D-glucose bind stereospecifically to a colloidal clay. *Science*, **203**, 1243–4.

Bonner, W. A. (1972). Origins of molecular chirality. In *Exobiology*, Chapter 6, ed. C. Ponnamperuma, pp. 170–234. Amsterdam: North-Holland.

Bonner, W. A. & Kavasmaneck, P. R. (1976). Asymmetric adsorption of DL-alanine hydrochloride by quartz. *Journal of Organic Chemistry*, **41**, 2225–6.

Bonner, W. A. & Kavasmaneck, P. R. (1977). Adsorption of amino acid derivatives by *d*- and *l*-quartz. *Journal of the American Chemical Society*, **99**, 44–50.

Bonner, W. A., Kavasmaneck, P. R., Martin, F. S. & Flores, J. J. (1974). Asymmetric adsorption of alanine by quartz. *Science*, **186**, 143–4.

Bonner, W. A., Kavasmaneck, P. R., Martin, F. S. & Flores, J. J. (1975). Asymmetric adsorption by quartz: a model for the prebiotic origin of optical activity. *Origins of Life*, **6**, 367–76.

Bonner, W. A., Van Dort, M. A. & Yearian, M. R. (1975). Asymmetric degradation of DL-leucine with longitudinally polarised electrons. *Nature*, **258**, 419–21.

Brack, A. (1976). Polymérisation en phase aqueuse d'acides amines sur des argiles. *Clay Minerals*, **11**, 117–20.

Brack, A. (1977). $\beta$-Structures of alternating polypeptides and their possible role in chemical evolution. *Biosystems*, **9**, 99–103.

Brack, A. & Orgel, L. E. (1975). $\beta$-Structures of alternating polypeptides and their possible prebiotic significance. *Nature*, **256**, 383–7.

Bradley, W. F. & Serratosa, J. M. (1960). A discussion of the water content of vermiculite. *Clays and Clay Minerals*, **7**, 260–70.

Brady, G. W., Kurkjian, C. R., Lyden, E. F. X., Robin, M. B., Saltman, P., Spiro, T. & Terzis, A. (1968). The structure of an iron core analog of ferritin. *Biochemistry*, **7**, 2185–92.

Bragg, W. L. (1937). *Atomic structure of minerals*. Ithaca: Cornell University Press.

Bragg, W. L. & Brown, G. B. (1926). Die Struktur des Olivins. *Zeitschrift für Kristallographie*, **63**, 538–56.

Brahms, S., Brahms, J., Spach, G. & Brack, A. (1977). Identification of $\beta$-turns and unordered conformations in the polypeptide chains by vacuum ultraviolet circular dichroism. *Proceedings of the National Academy of Sciences USA*, **74**, 3208–12.

Brauner, K. & Preisinger, A. (1956). Struktur und Entstehung des Sepioliths. *Tschermaks Mineralogische und Petrographische Mitteilungen*, **6**, 120–40.

Breslow, R. (1959). On the mechanism of the formose reaction. *Tetrahedron Letters*, **no. 21**, 22–6.

Brindley, G. W. (1951). *X-ray Identification and Crystal Structures of Clay Minerals*. London: The Mineralogical Society.

Brindley, G. W. & Maksimovic, Z. (1974). The nature and nomenclature of hydrous nickel-containing silicates. *Clay Minerals*, **10**, 271–7.

Brown, B. E. & Bailey, S. W. (1962). Chlorite polytypism: I. Regular and semi-random one-layer structures. *American Mineralogist*, **47**, 819–50.

Brown, B. E. & Bailey, S. W. (1963). Chlorite polytypism: II. Crystal structure of a one-layer Cr-Chlorite. *American Mineralogist*, **48**, 42–61.

Buchanan, B. B. (1972). Ferredoxin-linked carboxylation reactions. In *The Enzymes*, vol. 6, ed. P. D. Boyer, pp. 193–216. New York: Academic.

Buerger, M. J. (1945). The genesis of twin crystals. *American Mineralogist*, **30**, 469–82.

Bull, A. T. (1974). Microbial growth. In *Companion to Biochemistry*, ed. A. T. Bull, J. R. Lagnado, J. O. Thomas & K. F. Tipton, pp. 415–42. London: Longmans.

Bungenberg de Jong, H. G. (1932). Die Koazervation und ihre Bedeutung für die Biologie. *Protoplasma*, **15**, 110–73.

Bursill, L. A., Thomas, J. M. & Rao, K. J. (1981). Stability of zeolites under electron irradiation and imaging of heavy cations in silicates. *Nature*, **289**, 157–8.

Burton, F. G. & Neuman, W. F. (1971). On the possible role of crystals in the origins of life. V. The polymerisation of glycine. *Currents in Modern Biology*, **4**, 47–54.

Butlero[v], A. (1861). Bildung einer zuckerartigen Substanz durch Synthese. *Annalen der Chemie und Pharmacie*, **120**, 295–8.

Buvet, R. & Le Port, L. (1973). Non-enzymic origin of metabolism. *Space Life Sciences*, **4**, 434–47.

Bystrom, A. & Bystrom, A. M. (1950). The crystal structure of hollandite, the related manganese oxide minerals, and $\alpha$-$MnO_2$. *Acta Crystallographica*, **3**, 146–54.

Cady, S. S. & Pinnavaia, T. J. (1978). Porphyrin intercalation in mica-type silicates. *Inorganic Chemistry*, **17**, 1501–7.

Caillère, S. & Hénin, S. (1948). Sur la preparation et quelques caractères d'une série d'aluminates hydratés. *Comptes Rendus de L'Académie des Sciences Paris*, **226**, 580–2.

Caillère, S. & Hénin, S. (1949). Experimental formation of chlorites from montmorillonite. *Mineral Magazine*, **28**, 612–20.

Cairns-Smith, A. G. (1966). The origin of life and the nature of the primitive gene. *Journal of Theoretical Biology*, **10**, 53–88.

Cairns-Smith, A. G. (1968). An approach to a blueprint for a primitive organism. In *Towards a Theoretical Biology, I. Prolegomena*, ed. C. H. Waddington, pp. 57–66. Edinburgh: Edinburgh University Press.

Cairns-Smith, A. G. (1971). *The Life Puzzle: On crystals and organisms and on the possibility of a crystal as an ancestor*. Edinburgh: Oliver & Boyd.

Cairns-Smith, A. G. (1974). The methods of science and the origins of life. In *The Origin of Life and Evolutionary Biochemistry*, ed. K. Dose, S. W. Fox, G. A. Deborin & T. E. Pavlovskaya, pp. 53–8. New York: Plenum.

Cairns-Smith, A. G. (1975a). A case for an alien ancestry. *Proceedings of the Royal Society B*, **189**, 249–74.

Cairns-Smith, A. G. (1975b). Ambiguity in the interpretation of abiotic syntheses. *Origins of Life*, **6**, 265–7.

Cairns-Smith, A. G. (1977). Takeover mechanisms and early biochemical evolution. *BioSystems*, **9**, 105–9. Also in *Origin of Life*, ed. H. Noda (1978), pp. 399–404. Tokyo: Center for Academic Publications, Japan.

Cairns-Smith, A. G. (1979). Organisms of the first kind. *Chemistry in Britain*, **15**, 576–9.

Cairns-Smith, A. G. (1981). Beginnings of organic evolution. In *Study of Time*, IV, ed. J. T. Fraser, pp. 15–33. New York: Springer-Verlag.

Cairns-Smith, A. G. & Davis, C. J. (1977). The design of novel replicating polymers. In *The Encyclopaedia of Ignorance: Life Sciences and Earth Sciences*, ed. R. Duncan & M. Weston-Smith, pp. 391–404. Oxford: Pergamon.

Cairns-Smith, A. G., Ingram, P. & Walker, G. L. (1972). Formose production by minerals, possible relevance to the origin of life. *Journal of Theoretical Biology*, **35**, 601–4.

Cairns-Smith, A. G. & Walker, G. L. (1974). Primitive metabolism. *Biosystems*, **5**, 173–86.

Calvin, M. (1969). *Chemical Evolution: molecular evolution towards the origin of living systems on the earth and elsewhere*. Oxford: Clarendon.

Calvin, M. (1974). Solar energy by photosynthesis. *Science*, **184**, 375–81.
Campbell, J. H., Lengyel, J. A. & Langridge, J. (1973). Evolution of a second gene for β-galactosidase in *Escherichia coli*. *Proceedings of the National Academy of Sciences USA*, **70**, 1841–5.
Chaussidon, J. (1973). Le spectre infrarouge des biotites: Vibrations d'élongation basse frequence des OH du reseau. In *Proceedings of the International Clay Conference 1972*, ed. J. M. Serratosa, pp. 99–106. Madrid: Division de Ciencias, CSIC.
Chittenden, G. J. F. & Schwartz, A. W. (1976). Possible pathway for prebiotic uracil synthesis by photodehydrogenation. *Nature*, **263**, 350–1.
Choughuley, A. S. U., Subbaraman, A. S., Kazi, Z. A. & Chada, M. S. (1977). A possible prebiotic synthesis of thymine: uracil-formaldehyde-formic acid reaction. *Biosystems*, **9**, 73–80.
Chukhrov, F. V., Gorshkov, A. I., Rudnitskaya, E. S., Beresovskaya, V. V. & Sivtsov, A. V. (1980). Manganese minerals in clays: a review. *Clays and Clay Minerals*, **28**, 346–54.
Clark, S. P. Jr, Turekian, K. K. & Grossman, L. (1972). Model for the early history of the Earth. In *The Nature of the Solid Earth*, ed. E. C. Robertson, pp. 3–18. New York: McGraw-Hill.
Clemency, C. V. (1976). Simultaneous weathering of a granitic gneiss and an intrusive amphibolite dike near São Paulo, Brazil, and the origin of clay minerals. In *Proceedings of the International Clay Conference 1975*, ed. S. W. Bailey, pp. 15–25. Wilmette, Illinois: Applied Publishing Ltd.
Coleridge, S. T. (*ca* 1820). *Theory of Life*, p. 42. Quoted in Snyder, A. D. (1929). *Coleridge on Logic and Learning*. (1816 and 1823 are alternative suggestions for the publication date of *Theory of Life* – see Snyder, A. D. (1928). Coleridgeana. *Review of English Studies*, **4**, 432–4.)
Corbridge, D. E. C. (1978). *Phosphorus*. Amsterdam: Elsevier.
Cotton, F. A. & Wilkinson, G. (1972). *Advanced Inorganic Chemistry*, 3rd edn. New York: Interscience.
Coyne, L. M., Lawless, J., Lahav, N., Sutton, S. & Sweeney, M. (1981). Clays as prebiotic photocatalysts. *Origin of Life*, ed. Y. Wolman, pp. 115–24. Dordrecht: D. Reidel.
Cozzarelli, N. R. (1980). DNA gyrase and the supercoiling of DNA. *Science*, **207**, 953–60.
Cradwick, P. D. G., Farmer, V. C., Russell, J. D., Masson, C. R., Wada, K. & Yoshinaga, N. (1972). Imogolite, a hydrated aluminium silicate of tubular structure. *Nature Physical Science*, **240**, 187–9.
Cram, D. J. & Cram, J. M. (1978). Design of complexes between synthetic hosts and organic guests. *Accounts of Chemical Research*, **11**, 8–14.
Crick, F. H. C. (1968). The origin of the genetic code. *Journal of Molecular Biology*, **38**, 367–79.
Crowther, J. P. & Westman, A. E. R. (1954). The hydrolysis of the condensed phosphates: I. Sodium pyrophosphate and sodium triphosphate. *Canadian Journal of Chemistry*, **32**, 42–8.
Darwin, C. (1871). In *Life and Letters of Charles Darwin*, vol. 3, ed. F. Darwin, p. 18. London: John Murray.



Daubrée, A. (1879). *Études Synthétiques de Géologie Expérimentale*. Paris. (Quoted in Deer, Howie & Zussman (1962), vol. 4, p. 399.)
Dauvillier, A. (1938). *Astronomie*, **52**, 529.
Dauvillier, A. (1965). *The Photochemical Origin of Life*. New York: Academic.
Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.
De Kimpe, C. R. (1976). Formation of phylosilicates and zeolites from pure silica-alumina gels. *Clays and Clay Minerals*, **24**, 200–7.
Deer, W. A., Howie, R. A. & Zussman, J. (1962). *Rock forming Minerals*, vols I–V. London: Longmans.
Degani, C. & Halmann, M. (1967). Chemical evolution of carbohydrate metabolism. *Nature*, **216**, 1207.
Dent, L. S. & Smith, J. V. (1958). Crystal structure of chabazite, a molecular sieve. *Nature*, **181**, 1794–6.
Dickerson, R. E. (1971). Sequence and structure homologies in bacterial and mammalian-type cytochromes. *Journal of Molecular Biology*, **57**, 1–15.
Dickerson, R. E. (1978). Chemical evolution and the origin of life. *Scientific American*, **239**, no. 3, 62–78.
Dixon, J. B. & McKee, T. R. (1974). Internal and external morphology of tubular and spheroidal halloysite particles. *Clays and Clay Minerals*, **22**, 127–37.
Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *American Biology Teacher*, **35**, 125–9.
Dobzhansky, T., Ayala, F. J., Stebbins, G. L. & Valentine, J. W. (1977). *Evolution*. San Francisco: Freeman & Co.
Dose, K. (1971). Catalysis. In *Theory and Experiment in Exobiology*, vol. 1, ed. A. W. Schwartz, pp. 41–71. Groningen: Wolters-Noordhoff.
Dose, K. (1974). Peptides and amino acids in the primordial hydrosphere. In *The Origin of Life and Evolutionary Biochemistry*, ed. K. Dose, S. W. Fox, G. A. Deborin & T. E. Pavlovskaya, pp. 69–77. New York: Plenum.
Dose, K. (1975). Peptides and amino acids in the primordial hydrosphere. *BioSystems*, **6**, 224–8.
Doty, P. & Lundberg, R. D. (1956). Configurational and stereochemical effects in the amine initiated polymerisation of *N*-carboxyanhydrides. *Journal of the American Chemical Society*, **78**, 4810–12.
Dousma, J. & de Bruyn, P. L. (1976). Hydrolysis-precipitation studies of iron solutions: I. Model for hydrolysis and precipitation from Fe(III) nitrate solutions. *Journal of Colloid and Interface Science*, **56**, 527–39.
Dousma, J. & de Bruyn, P. L. (1978). Hydrolysis-precipitation studies of iron solutions: II. Ageing studies and the model for precipitation from Fe(III) nitrate solutions. *Journal of Colloid and Interface Science*, **64**, 154–70.
Dowty, E. (1980). Crystal-chemical factors affecting the mobility of ions in minerals. *American Mineralogist*, **65**, 174–82.
Drake, J. W. (1969). Comparative rates of spontaneous mutation. *Nature*, **221**, 1132.
Dustin, P. (1980). Microtubules. *Scientific American*, **243**, no. 2, 58–68.
Eberl, D. & Hower, J. (1975). Kaolinite synthesis: the role of the Si/Al ratio

and the (Alkali/H+) ratio in hydrothermal systems. *Clays and Clay Minerals*, **23**, 301–9.

Eigen, M., Gardiner, W., Schuster, P. & Winkler-Oswatitsch, R. (1981). The origin of genetic information. *Scientific American*, **244**, no. 4, 78–94.

Emsley, J. (1977). Phosphate cycles. *Chemistry in Britain*, **13**, 459–63.

Ewing, F. J. (1935a). The crystal structure of lepidocrocite. *Journal of Chemical Physics*, **3**, 420–4.

Ewing, F. J. (1935b). The crystal structure of diaspore. *Journal of Chemical Physics*, **3**, 203–7.

Farmer, V. C. & Fraser, A. R. (1979). Synthetic imogolite, a tubular hydroxy-aluminium silicate. In *International Clay Conference 1978*, ed. M. M. Mortland & V. C. Farmer, pp. 547–53. Amsterdam: Elsevier.

Farmer, V. C., Fraser, A. R. & Tait, J. M. (1977). Synthesis of Imogolite. A tubular aluminium silicate polymer. *Journal of the Chemical Society, Chemical Communications*, 462–3.

Farmer, V. C., Russell, J. D., McHardy, W. J., Newman, A. C. D., Ahlrichs J. L. & Rimsaite, J. Y. H. (1971). Evidence for loss of protons and octahedral iron from oxidised biotites and vermiculites. *Mineral Magazine*, **38**, 121–37.

Feeny, R. E., Blankenhorn, G. & Dixon, H. B. F. (1975). Carbonyl-amine reactions in protein chemistry. *Advances in Protein Chemistry*, **29**, 135–203.

Feitknecht, W. & Schindler, P. (1963). Solubility constants of metal oxides, metal hydroxides and hydroxide salts in aqueous solution. *Pure and Applied Chemistry*, **6**, 125–99.

Ferris, J. P. & Chen, C. T. (1975a). Chemical evolution. XXVI. Photochemistry of methane, nitrogen and water mixtures as a model for the atmosphere of the primitive Earth. *Journal of the American Chemical Society*, **97**, 2962–7.

Ferris, J. P. & Chen, C. T. (1975b). Photosynthesis of organic compounds in the atmosphere of Jupiter. *Nature*, **258**, 587–8.

Ferris, J. P. & Edelson, E. H. (1978). Chemical evolution. 31. Mechanism of the condensation of cyanide to HCN oligomers. *Journal of Organic Chemistry*, **43**, 3989–95.

Ferris, J. P., Donner, D. B. & Lobo, A. P. (1973). Possible role of hydrogen cyanide in chemical evolution: The oligomerisation and condensation of hydrogen cyanide. *Journal of Molecular Biology*, **74**, 511–18.

Ferris, J. P., Joshi, P. C., Edelson, E. H. & Lawless, J. G. (1978). HCN: a plausible source of purines, pyrimidines and amino acids on the primitive Earth. *Journal of Molecular Evolution*, **11**, 293–311.

Ferris, J. P. & Nicodem, D. E. (1974). Ammonia: Did it have a role in chemical evolution? In *The Origin of Life and Evolutionary Biochemistry*, ed. K. Dose, S. W. Fox, G. A. Deborin & T. E. Pavlovskaya, pp. 107–17. New York: Plenum.

Ferris, J. P., Sanchez, R. A. & Orgel, L. E. (1968). Synthesis of pyrimidines from cyanoacetylene and cyanate. *Journal of Molecular Biology*, **33**, 693–704.

Ferris, J. P., Wos, J. D., Nooner, D. W. & Oró, J. (1974). Chemical evolution. XXI. The amino acids released on hydrolysis of HCN oligomers. *Journal of Molecular Evolution*, **3**, 225–31.

Ferris, J. P., Zamek, O. S., Altbuch, A. M. & Freiman, H. (1974). Chemical evolution: XVIII. Synthesis of pyrimidines from guanidine and cyanoacetaldehyde. *Journal of Molecular Evolution*, **3**, 301–9.

Fischer, W. R. & Schwertmann, U. (1975). The formation of haematite from amorphous iron III hydroxide. *Clays and Clay Minerals*, **23**, 33–7.

Fitzpatrick, R. W., le Roux, J. & Schwertmann, U. (1978). Amorphous and crystalline titanium and iron-titanium oxides in synthetic preparations, at near ambient conditions, and in soil clays. *Clays and Clay Minerals*, **26**, 189–201.

Fleischer, R. L., Price, P. B. & Walker, R. M. (1975). *Nuclear Tracks in Solids*. Berkeley: University of California Press.

Force, E. R. (1980). The provenance of rutile. *Journal of Sedimentary Petrology*, **50**, 485–8.

Foster, M. D. (1963). Interpretation of the composition of vermiculites and hydrobiotites. *Clays and Clay Minerals*, **10**, 70–89.

Fournier, R. O. & Rowe, J. J. (1962). The solubility of crystobalite along the three phase curve, gas plus liquid plus crystobalite. *American Mineralogist*, **47**, 897–902.

Fox, S. W. (1956). Evolution of protein molecules and thermal synthesis of biochemical substances. *American Scientist*, **44**, 347–59.

Fox, S. W. (1969). Self-ordered polymers and propagative cell-like systems. *Naturwissenschaften*, **56**, 1–9.

Fox, S. W. & Dose, K. (1972). *Molecular Evolution and the Origin of Life*. San Francisco: Freeman.

Fox, S. W. & Harada, K. (1960). The thermal copolymerisation of amino acids common to protein. *Journal of the American Chemical Society*, **82**, 3745–51.

Fox, S. W. & Windsor, C. R. (1970). Synthesis of amino acids by the heating of formaldehyde and ammonia. *Science*, **170**, 984–5.

Francombe, M. H. & Rooksby, H. P. (1959). Structure transformations effected by dehydration of diaspore, goethite and delta ferric oxide. *Clay Minerals Bulletin*, **4**, 1.

Frank, F. C. (1953). On spontaneous asymmetric synthesis. *Biochimica et Biophysica Acta*, **11**, 459–63.

Fraser, R. D. B. (1969). Keratins. *Scientific American*, **221**, no. 2, 86–96.

Freund, F. & Wenegler, H. (1978). Proton conductivity in hydroxides as related to transport phenomena in soils. *Book of Summaries for the 6th International Clay Conference (Oxford)*, p. 610. London: The Mineralogical Society.

Frey, E. & Lagaly, G. (1979). Selective coagulation and mixed layer formation from sodium smectite solutions. In *International Clay Conference 1978*, ed. M. M. Mortland & V. C. Farmer, pp. 131–40. Amsterdam: Elsevier.

Friebele, E., Shimoyama, A., Hare, P. E. & Ponnamperuma, C. (1981). Adsorption of amino acid enantiomers by Na-montmorillonite. *Origins of Life*, **11**, 173–84.

Gabel, N. W. & Ponnamperuma, C. (1967). Model for origin of monosaccharides. *Nature*, **216**, 453–5.

Galbraith, S. T., Baird, T. & Fryer, J. R. (1979). Structural changes in $\beta$-FeOOH caused by radiation damage. *Acta Crystallographica*, **A35**, 197–200.

Gardner, M. (1976). Mathematical games: On the fabric of inductive logic, and some probability paradoxes. *Scientific American*, **234**, no. 3, 119.

Garrison, W. M., Morrison, D. C., Hamilton, J. G., Benson, A. A. & Calvin, M. (1951). Reduction of carbon dioxide in aqueous solutions by ionising radiation. *Science*, **114**, 416–18.

Gatineau, L. (1964). Structure réelle de la muscovite. Répartition des substitutions isomorphe. *Bulletin de la Société française de Minéralogie et Crystallographie*, **87**, 321–55.

Gerstl, Z. & Banin, A. (1980). $Fe^{2+}$–$Fe^{3+}$ Transformations in clay and resin ion-exchange systems. *Clays and Clay Minerals*, **28**, 335–45.

Getoff, N. (1962). Reduktion der Kohlensaure in wässeriger Lösung unter Einwirkung von UV-Licht. *Zeitschrift für Natürforschung*, **17B**, 87–90.

Getoff, N. (1963a). Einfluss der Ferroionen bei der Bildung von Oxalsaure aus wässeriger Kohlensaure durch UV-Bestrahlung. *Zeitschrift für Naturforschung*, **18B**, 169–70.

Getoff, N. (1963b). Role of iron II ions in carbonic acid reduction in aqueous solutions by means of ultraviolet light. *Oesterricher Chemiker-Zeitung*, **64**, 70–2.

Giese, R. F. (1978). The electrostatic interlayer forces of layer structure minerals. *Clays and Clay Minerals*, **26**, 51–7.

Giese, R. F. & Datta, P. (1973). Hydroxyl orientation in kaolinite, dickite, and nacrite. *American Mineralogist*, **58**, 471–9.

Gildawie, A. M. (1977). *Electron Microscope Studies of Colloidal Metal Oxides*. PhD Thesis, University of Glasgow.

Gimblett, F. G. R. (1963). *Inorganic Polymer Chemistry*. London: Butterworths.

Goodwin, A. M. (1976). Giant impacting and the development of continental crust. In *The Early History of the Earth*, ed. B. F. Windley, pp. 77–95. London: John Wiley & Sons.

Gottardi, G. & Meier, W. M. (1963). The crystal structure of dachiardite. *Zeitschrift für Kristallographie*, **119**, 53–64.

Granick, S. (1957). Speculations on the origins and evolution of photosynthesis. *Annals of the New York Academy of Sciences*, **69**, 292–308.

Granick, S. (1965). Evolution of heme and chlorophyll. In *Evolving Genes and Proteins*, ed. V. Bryson & H. J. Vogel, pp. 67–88. New York: Academic.

Granquist, W. T. & Pollack, S. S. (1959). A study of the synthesis of hectorite. *Clays and Clay Minerals*, **8**, 150–69.

Grant, G. T., Morris, E. R., Rees, D. A., Smith, P. J. C. & Thom, D. (1973). Biological interactions between polysaccharides and divalent cations: the egg box model. *FEBS Letters*, **32**, 195–8.

Greenland, D. J. (1956). The adsorption of sugars by montmorillonite, I and II. *Journal of Soil Science*, **7**, 319–28 and 329–34.

Griffith, E. J., Ponnamperuma, C. & Gabel, N. W. (1977). Phosphorus, a key to life on the primitive Earth. *Origins of Life*, **8**, 71–85.

Grim, R. E. & Güven, N. (1978). *Bentonites: Geology, Mineralology, Properties and Use*. Amsterdam: Elsevier.

Guggenheim, S. & Bailey, S. W. (1975). Refinement of the margarite structure in subgroup symmetry. *American Mineralogist*, **60**, 1023–9.

Güven, N. (1971). The crystal structure of $2M_1$ phengite and $2M_1$ muscovite. *Zeitschrift für Kristallographie*, **134**, 196–212.



Güven, N. & Burnham, C. W. (1967). The crystal structure of 3T muscovite. *Zeitschrift für Kristallographie*, **125**, 163–83.

Güven, N., Pease, R. W. & Murr, L. E. (1977). Fine structure in the selected area diffraction patterns of beidellite and its dark field images. *Clay Minerals*, **12**, 67–74.

Haldane, J. B. S. (1929). The origin of life. *Rationalist Annual* **3**. (Reprinted in Bernal, 1967, pp. 242–9.)

Hall, D. O., Cammack, R. & Rao, K. K. (1974). The iron–sulphur proteins: evolution of a ubiquitous protein from model systems to higher organisms. *Origins of Life*, **5**, 363–86.

Hall, S. H. & Bailey, S. W. (1979). Cation ordering pattern in amesite. *Clays and Clay Minerals*, **27**, 241–7.

Halmann, M. (1974). Evolution and ecology of phosphorus metabolism. In *The Origin of Life and Evolutionary Biochemistry*, ed. K. Dose, S. W. Fox, G. A. Deborin & T. E. Pavlovskaya, pp. 169–82. New York: Plenum.

Halmann, M. (1978). Photoelectrochemical reduction of aqueous carbon dioxide on p-type gallium phosphide in liquid junction solar cells. *Nature*, **275**, 115–16.

Harada, K. & Fox, S. W. (1964). Thermal synthesis of natural amino acids from a postulated primitive terrestrial atmosphere. *Nature*, **201**, 335–6.

Harada, K. & Fox, S. W. (1965). The thermal synthesis of amino acids from a hypothetically primitive terrestrial atmosphere. In *The Origins of prebiological systems*, ed. S. W. Fox, pp. 187–201. New York: Academic.

Harder, H. (1977). Clay mineral formation under lateritic weathering conditions. *Clay Minerals*, **12**, 281–8.

Harder, H. (1978). Synthesis of iron layer silicate minerals under natural conditions. *Clays and Clay Minerals*, **26**, 65–72.

Hartman, H. (1975a). Speculations on the origin and evolution of metabolism. *Journal of Molecular Evolution*, **4**, 359–70.

Hartman, H. (1975b). Speculations on the evolution of the genetic code. *Origins of Life*, **6**, 423–7, and (1978), **9**, 133–6.

Hartmann, W. K. (1975). The smaller bodies of the solar system. *Scientific American*, **233**, no. 3, 142–59.

Hay, R. L. (1966). *Zeolites and Zeolitic Reactions in Sedimentary Rocks*. Geological Society of America Special Paper no. 85, New York: Geological Society of America.

Hayatsu, R., Studier, M. H. & Anders, E. (1971). Origin of organic matter in early solar system. IV. Amino acids: confirmation of catalytic synthesis by mass spectrometry. *Geochimica et Cosmochimica Acta*, **35**, 939–51.

Hazen, R. M. & Wones, D. R. (1972). The effect of cation substitutions on the physical properties of trioctahedral micas. *American Mineralogist*, **57**, 103–29.

Henmi, T. & Wada, K. (1976). Morphology and composition of allophane. *American Mineralogist*, **61**, 379–90.

Hey, M. H. (1932). Studies on zeolites, II: Thomsonite and gonnardite, and III: Natrolite and metanatrolite. *Mineral Magazine*, **23**, 51–125 and 243–89.

Hinkle, P. C. & McCarty, R. E. (1978). How cells make ATP. *Scientific American*, **238**, no. 3, 104–23.

Hinshelwood, C. N. (1951). *The Structure of Physical Chemistry*, p. 452. Oxford: Clarendon.

Hocart, M. R. (1962). *Genèse et Synthèse des Argiles*. Paris: CNRS.

Hoffmann, R. W. & Brindley, G. W. (1960). Adsorption of non-ionic aliphatic molecules from aqueous solutions on montmorillonite. Clay-organic studies II. *Geochimica et Cosmochimica Acta*, **20**, 15–29.

Holland, H. D. (1962). Model for the evolution of the Earth's atmosphere. In *Petrologic Studies; a volume in honor of A. F. Buddington*, ed. A. E. J. Engel, H. L. James & B. F. Leonard, pp. 447–77. New York: Geological Society of America.

Hopfield, J. J. (1974). Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences USA*, **71**, 4135–9.

Horiuchi, S. (1978). Visualising atoms in inorganic compounds by 1 MV HRTEM. *Chemica Scripta*, **14**, 75–81.

Horowitz, N. H. (1945). On the evolution of biochemical synthesis. *Proceedings of the National Academy of Sciences USA*, **31**, 153–7.

Horowitz, N. H. (1959). On defining life. In *The Origin of Life on the Earth*, ed. F. Clark & R. L. M. Synge, pp. 106–7. London: Pergamon.

Hsu, P. H. (1972). Nucleation, polymerisation and precipitation of FeOOH. *Journal of Soil Science*, **23**, 409–19.

Hubbard, J. S., Hardy, J. P. & Horowitz, N. H. (1971). Photocatalytic production of organic compounds from CO and $H_2O$ in a simulated Martian atmosphere. *Proceedings of the National Academy of Sciences USA*, **68**, 574–8.

Hubbard, J. S., Hardy, J. P., Voecks, G. E. & Golub, E. E. (1973). Photocatalytic synthesis of organic compounds from CO and water: Involvement of surfaces in the formation and stabilisation of products. *Journal of Molecular Evolution*, **2**, 149–66.

Hubbard, J. S., Voecks, G. E., Hobby, G. L., Ferris, J. P., Williams, E. A. & Nicodem, D. E. (1975). Ultraviolet gas phase and photocatalytic synthesis from CO and $NH_3$. *Journal of Molecular Evolution*, **5**, 223–41.

Hulett, H. R. (1969). Limitations on prebiological synthesis. *Journal of Theoretical Biology*, **24**, 56–72.

Hull, D. E. (1960). Thermodynamics and kinetics of spontaneous generation. *Nature*, **186**, 693–4.

Hutchison, J. L., Jefferson, D. A. & Thomas, J. M. (1977). The ultrastructure of minerals as revealed by high resolution electron microscopy. In *Surface and Defect Properties of Solids, Vol. 6, Specialist Periodical Report*, ed. M. W. Roberts & J. H. Thomas, pp. 320–58. London: The Chemical Society.

Huxley, J. (1974). *Evolution: The Modern Synthesis*. London: Allen & Unwin.

Hynes, M. J. (1975). Amide utilisation in *Aspergillus nidulans*: Evidence for a third amidase enzyme. *Journal of General Microbiology*, **91**, 99–109.

Inoue, T., Fujishima, A., Konishi, S. & Honda, K. (1979). Photoelectrocatalytic reduction of carbon dioxide in aqueous suspensions of semiconductor powders. *Nature*, **277**, 637–8.

Israelachvili, J. (1978). The packing of lipids and proteins in membranes. In *Light Transducing Membranes*, ed. D. W. Deamer, pp. 91–107. New York: Academic.



Jasmund, K., Sylla, H. M. & Freund, F. (1976). Solid solution in synthetic serpentine phases. In *Proceedings of the International Clay Conference 1975*, ed. S. W. Bailey, pp. 267–74. Wilmette, Illinois: Applied Publishing Ltd.

Jensen, R. A. (1976). Enzyme recruitment in evolution of new function. *Annual Review of Microbiology*, **30**, 409–25.

Johnson, J. S. & Kraus, K. A. (1959). Hydrolytic behaviour of metal ions. X. Ultracentrifugation of lead (II) and tin (IV) in basic solution. *Journal of the American Chemical Society*, **81**, 1569–72.

Johnson, L. J. (1970). Clay minerals in Pennsylvania soils in relation to lithology of the parent rock and other factors – I. *Clays and Clay Minerals*, **18**, 247–60.

Jonas, E. C. (1976). Crystal chemistry of diagenesis in 2:1 clay minerals. In *Proceedings of the International Clay Conference 1975*, ed. S. W. Bailey, pp. 3–13. Wilmette, Illinois: Applied Publishing Ltd.

Kalt, A., Perati, B. & Wey, R. (1979). Intercalation compounds of $KHSi_2O_5$ and $H_2Si_2O_5$ with alkylammonium ions and alkylamines. In *International Clay Conference 1978*, ed. M. M. Mortland & V. C. Farmer, pp. 509–16. Amsterdam: Elsevier.

Katchalski, E. (1951). Poly-α-amino acids. *Advances in Protein Chemistry*, **6**, 123–5.

Kay, W. W. (1971). Two aspartate transport systems in *Escherichia coli*. *Journal of Biological Chemistry*, **246**, 7373–82.

Keller, W. D. (1976). Scan electron micrographs of kaolins collected from diverse environments of origin – I. *Clays and Clay Minerals*, **24**, 107–13.

Keller, W. D. (1977). Scan electron micrographs of kaolins collected from diverse environments of origin – IV. *Clays and Clay Minerals*, **25**, 311–45.

Keller, W. D. (1978). Classification of kaolins exemplified by their textures in scan electron micrographs. *Clays and Clay Minerals*, **26**, 1–20.

Keller, W. D. & Hanson, R. F. (1975). Dissimilar fabrics by scan electron microscopy of sedimentary versus hydrothermal kaolins in Mexico. *Clays and Clay Minerals*, **23**, 201–4.

Kenyon, D. H. & Nissenbaum, A. (1976). Melanoidin and aldocyanoin microspheres: Implications for chemical evolution and early precambrian micropaleontology. *Journal of Molecular Evolution*, **7**, 245–51.

Kenyon, D. H. & Steinman, G. (1969). *Biochemical Predestination*. New York: McGraw-Hill.

Kerr, R. A. (1980). Origin of life: New ingredients suggested. *Science*, **210**, 42–3.

Kirkman, J. H. (1977). Possible structure of halloysite disks and cylinders observed in some New England rhyolitic tephras. *Clay Minerals*, **12**, 199–216.

Kirkman, J. H. (1981). Morphology and structure of halloysite in New Zealand tephras. *Clays and Clay Minerals*, **29**, 1–9.

Kirsten, T. (1978). Time and the solar system. In *The Origin of the Solar System*. ed. S. F. Dermott, pp. 267–346. Chichester: John Wiley & Sons.

Klein, H. P. (1978). The Viking biological investigations: Review and status. *Origins of Life*, **9**, 157–60.

Koshland, D. E., Némethy, G. & Filmer, D. (1966). Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, **5**, 365–85.

Krasnovsky, A. A. (1974). Chemical evolution of photosynthesis. In *The Origin*

*of Life and Evolutionary Biochemistry*, ed. K. Dose, S. W. Fox, G. A. Deborin & T. E. Pavlovskaya, pp. 233–44. New York: Plenum.

Krumbein, W. C. & Garrels, R. M. (1952). Origin and classification of chemical sediments in terms of pH and oxidation-reduction potentials. *Journal of Geology*, **60**, 1–33.

Kuhn, H. (1976). Model consideration for the origin of life: Environmental structure as stimulus for the evolution of chemical systems. *Naturwissenschaften*, **63**, 68–80.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*, 2nd edn. Chicago: University of Chicago Press.

Kunze, G. (1956). Die gewellte Struktur des Antigorits, I. *Zeitschrift für Kristallographie*, **108**, 82–107.

Lagaly, G. (1979). The 'layer charge' of regular interstratified 2:1 clay minerals. *Clays and Clay Minerals*, **27**, 1–10.

Lagaly, G. (1981). Inorganic layer compounds. *Naturwissenschaften*, **68**, 82–8.

Lahav, N. & Chang, S. (1976). The possible role of solid surface area in condensation reactions during chemical evolution: Re-evaluation. *Journal of Molecular Evolution*, **8**, 357–80.

Lahav, N., White, D. & Chang, S. (1978). Peptide formation in the prebiotic era: Thermal condensation of glycine in fluctuating clay environments. *Science*, **201**, 67–9.

La Iglesia, A. & Galen, E. (1975). Halloysite–kaolinite transformation at room temperature. *Clays and Clay Minerals*, **23**, 109–13.

La Iglesia, A. & Martin-Vivaldi, J. L. (1973). A contribution to the synthesis of kaolinite. In *Proceedings of the International Clay Conference 1972*, ed. J. M. Serratosa, pp. 173–85. Madrid: Division de Ciencias, CSIC.

La Iglesia, A. & Martin-Vivaldi, J. L. (1975). Synthesis of kaolinite by homogeneous precipitation at room temperature. *Clay Minerals*, **10**, 399–405.

La Iglesia, A., Martin-Vivaldi, J. L. & Lopez Aguayo, F. (1976). Kaolinite crystallisation at room temperature by homogeneous precipitation III: Hydrolysis of feldspars. *Clays and Clay Minerals*, **24**, 36–42.

La Iglesia, A. & Van Oosterwyck-Gastuche, M. C. (1978). Kaolinite synthesis I: Crystallisation conditions at low temperatures and calculation of thermodynamic equilibria. Application to laboratory and field observations. *Clays and Clay Minerals*, **26**, 397–408.

Lailach, G. E. & Brindley, G. W. (1969). Specific co-absorption of purines and pyrimidines by montmorillonite (clay–organic studies XV). *Clays and Clay Minerals*, **17**, 95–100.

Lailach, G. E., Thompson, T. D. & Brindley, G. W. (1968a). Absorption of pyrimidines, purines and nucleosides by Li-, Na-, Mg-, and Ca-montmorillonite (clay–organic studies XII). *Clays and Clay Minerals*, **16**, 285–93.

Lailach, G. E., Thompson, T. D. & Brindley, G. W. (1968b). Absorptions of pyrimidines, purines and nucleosides by Co-, Ni, Cu-, and Fe(III)-montmorillonite (clay–organic studies XIII). *Clays and Clay Minerals*, **16**, 295–301.

Lambert, R. St J. (1959). The mineralogy and metamorphism of the Moine schists of the Morar and Knoydart districts of Inverness-shire. *Transactions of the Royal Society of Edinburgh*, **63**, 553–88.



Laswick, J. A. & Plane, R. A. (1959). Hydrolytic polymerisation in boiled chromic solutions. *Journal of the American Chemical Society*, **81**, 3564–7.

Lawless, J. G. (1980). Organic compounds in meteorites. In *Life Sciences and Space Research*, vol. 18, ed. R. Holmquist, pp. 19–27. Oxford: Pergamon.

Lawless, J. G. & Levi, N. (1979). The role of metal ions in chemical evolution: Polymerisation of alanine and glycine in a cation-exchanged clay environment. *Journal of Molecular Evolution*, **13**, 281–6.

Lawless, J. G. & Peterson, E. (1975). Amino acids in carbonaceous chondrites. *Origins of Life*, **6**, 3–8.

Lawless, J. G. & Yuen, G. U. (1979). Quantification of monocarboxylic acids in the Murchison carbonaceous meteorite. *Nature*, **282**, 396–8.

Lawless, J. G., Zeitman, B., Pereira, W. E., Summons, R. E. & Duffield, A. M. (1974). Dicarboxylic acids in the Murchison meteorite. *Nature*, **251**, 40–2.

Leach, W. W., Nooner, D. W. & Oró, J. (1978). Abiotic synthesis of fatty acids. In *Origin of Life*, ed. H. Noda, pp. 113–22. Tokyo: Center for Academic Publications Japan.

Lemmon, R. M. (1970). Chemical evolution. *Chemical Reviews*, **70**, 95–109.

Lewontin, R. C. (1978). Adaptation. *Scientific American*, **239**, no. 3, 156–69.

Liecester, H. M. (1956). *The Historical Background of Chemistry*. New York: John Wiley & Sons.

Lin, E. C. C. (1970). The genetics of bacterial transport systems. *Annual Review of Genetics*, **4**, 225–62.

Linares, J. & Huertas, F. (1971). Kaolinite: synthesis at room temperature. *Science*, **171**, 896–7.

Lipman, C. B. (1924). The origin of life. *Scientific Monthly*, **19**, 357–67.

Lipmann, F. (1965). Projecting backward from the present stage of evolution of biosynthesis. In *The Origins of Prebiological Systems*, ed. S. W. Fox, pp. 259–80. New York: Academic.

Lister, J. S. & Bailey, S. W. (1967). Chlorite polytypism: IV. Regular two-layer structures. *American Mineralogist*, **52**, 1614–31.

Lodish, H. F. & Rothman, J. E. (1979). The assembly of cell membranes. *Scientific American*, **240**, no. 1, 48–63.

Loeb, W. (1913). Über das Verhalten des Formamids unter der Wirkung der stillen Engladung. *Berichte der chemischen Gesellschaft*, **46**, 684–97.

Lohrmann, R. (1976). Formation of nucleoside 5′-polyphosphates under potentially prebiological conditions. *Journal of Molecular Evolution*, **8**, 197–210.

Lohrmann, R., Bridson, P. K. & Orgel, L. E. (1980). Efficient metal ion catalysed template-directed oligonucleotide synthesis. *Science*, **208**, 1464–5.

Lohrmann, R. & Orgel, L. E. (1971). Urea–inorganic phosphate mixtures as prebiotic phosphorylating agents. *Science*, **171**, 490–4.

Lohrmann, R. & Orgel, L. E. (1978). Template-directed polynucleotide condensation as a model for RNA replication. In *Origin of Life*, ed. H. Noda, pp. 235–44. Tokyo: Center for Academic Publications Japan.

Lowe, C. U., Rees, M. W. & Markham, R. (1963). Synthesis of complex organic compounds from simple precursors: Formation of amino acid polymers, fatty acids and purines from ammonium cyanide. *Nature*, **199**, 219–22.

Mackay, A. L. (1960). β-Ferric oxyhydroxide. *Mineral Magazine*, **32**, 545–57.
Mackay, A. L. (1964). β-Ferric oxyhydroxide – akaganeite. *Mineral Magazine*, **33**, 270–80.
Mackenzie, D. W. (1971). *Studies on the Synthesis of Layer Silicates*. Ph.D Thesis, University of Glasgow.
Mackinney, G. (1932). Photosynthesis *in vitro*. *Journal of the American Chemical Society*, **54**, 1688–9.
Mahler, H. R. & Cordes, E. H. (1971). *Biological Chemistry*, 2nd edn. New York: Harper & Row.
Mansfield, C. F. & Bailey, S. W. (1972). Twin and pseudotwin intergrowths in kaolinite. *American Mineralogist*, **57**, 411–25.
Margulis, L. (1970). *Origin of Eukaryotic Cells*. New Haven: Yale University Press.
Mason, B. (1966). *Principles of Geochemistry*, 3rd edn., p. 167. New York: Wiley.
Mathews, C. N. & Moser, R. E. (1967). Peptide synthesis from hydrogen cyanide and water. *Nature*, **215**, 1230–4.
Mathieson, A. McL. & Walker, G. F. (1954). Crystal structure of magnesium-vermiculite. *American Mineralogist*, **39**, 231–55.
Matijevic, E. (1976). Preparation and characteristics of monodispersed metal hydrous oxide sols. *Progress in Colloid and Polymer Science*, **61**, 24–35.
Matijevic, E. & Scheiner, P. (1978). Ferric hydrous oxide sols: III. Preparation of uniform particles by hydrolysis of Fe(III) chloride, nitrate, and perchlorate. *Journal of Colloid and Interface Science*, **63**, 509–24.
Maynard Smith, J. (1969). The status of neo-Darwinism. In *Towards a Theoretical Biology, 2. Sketches*, ed. C. H. Waddington, pp. 82–9. Edinburgh: Edinburgh University Press.
Maynard Smith, J. (1970). Time in the evolutionary process. *Studium generale*, **23**, 266–72.
Maynard Smith, J. (1975). *The Theory of Evolution*, 3rd edn. Harmondsworth: Penguin.
McBride, M. B. (1979). Reactivity of adsorbed and structural iron in hectorite as indicated by oxidation of benzidine. *Clays and Clay Minerals*, **27**, 224–30.
McMurchy, R. C. (1934). The crystal structure of the chlorite minerals. *Zeitschrift für Kristallographie*, **88**, 420–32.
Megaw, H. D., Kempster, C. J. E. & Radoslovish, E. W. (1962). The structure of anorthite, $CaAl_2Si_2O_8$, II: Description and discussion. *Acta Crystallographica*, **15**, 1017–35.
Meier, W. M. (1960). The crystal structure of natrolite. *Zeitschrift für Kristallographie*, **113**, 430–44.
Merrifield, R. B., Stewart, J. M. & Jernberg, N. (1966). Instrument for automated synthesis of peptides. *Analytical Chemistry*, **38**, 1905–14.
Meunier, A. & Velde, B. (1979). Biotite weathering in granites of western France. In *International Clay Conference 1978*, ed. M. M. Mortland & V. C. Farmer, pp. 405–13. Amsterdam: Elsevier.
Meyers, P. A. & Quinn, J. G. (1971). Fatty acid–clay mineral associations in artificial and natural sea water solutions. *Geochimica et Cosmochimica Acta*, **35**, 628–32.



Miller, K. R. (1979). The photosynthetic membrane. *Scientific American*, **241**, no. 4, 100–11.
Miller, S. L. (1953). A production of amino acids under possible primitive Earth conditions. *Science*, **117**, 528–9.
Miller, S. L. (1955). Production of some organic compounds under possible primitive Earth conditions. *Journal of the American Chemical Society*, **77**, 2351–61.
Miller, S. L. (1957). The mechanism of synthesis of amino acids by electric discharges. *Biochimica et Biophysica Acta*, **23**, 480–9.
Miller, S. L. & Orgel, L. E. (1974). *The Origins of Life on the Earth*. New Jersey: Prentice-Hall.
Miller, S. L. & Urey, H. C. (1959). Organic compound synthesis on the primitive Earth. *Science*, **130**, 245–51.
Millot, G. (1973). Data and tendencies of recent years in the field 'Genesis and synthesis of clay minerals'. In *Proceedings of the International Clay Conference 1972*, ed. J. M. Serratosa, pp. 151–7. Madrid: Division de Ciencias, CSIC.
Mills, D. R., Kramer, F. R. & Spiegelman, S. (1973). Complete nucleotide sequence of a replicating RNA molecule. *Science*, **180**, 916–27.
Misawa, T., Hashimoto, K. & Shimodaira, S. (1974). The mechanism of formation of iron oxide and oxyhydroxides in aqueous solutions at room temperature. *Corrosion Science*, **14**, 131–49.
Mizuno, T. & Weiss, A. H. (1974). Synthesis and utilisation of formose sugars. *Advances in Carbohydrate Chemistry*, **29**, 173–227.
Monod, J., Wyman, J. & Changeaux, J. P. (1965). On the nature of allosteric transitions: a plausible model. *Journal of Molecular Biology*, **12**, 88–118.
Moorbath, S. (1977). The oldest rocks and the growth of continents. *Scientific American*, **236**, no. 3, 92–104.
Moorbath, S., O'Nions, R. K. & Pankhurst, R. J. (1973). Early archean age for the Isua iron formation, West Greenland. *Nature*, **245**, 138–9.
Moore, B. (1913). *The Origin and Nature of Life*. The Home University Library of Knowledge, No. 63. London: Henry Holt.
Moore, B. & Webster, T. A. (1913). Synthesis by sunlight in relationship to the origin of life. Synthesis of formaldehyde from carbon dioxide and water by inorganic colloids acting as transformers of light energy. *Proceedings of the Royal Society B*, **87**, 163–76.
Morey, G. W., Fournier, R. O. & Rowe, J. J. (1962). The solubility of quartz in water in the temperature interval from 25° to 300 °C. *Geochimica et Cosmochimica Acta*, **26**, 1029–43.
Morgan, T. H. (1926). *The Theory of the Gene*. New Haven, Connecticut: Yale University Press.
Morimoto, S., Kawashiro, K. & Yoshida, H. (1978). The asymmetric adsorption of α-aminopropionitrile on D-quartz. In *Origin of Life*, ed. H. Noda, pp. 349–53. Tokyo: Center for Academic Publications Japan.
Morita, K., Ochiai, M. & Marumoto, R. (1968). A convenient one-step synthesis of adenine. *Chemistry and Industry*, p. 1117.
Muller, G. (1961). Die rezenten Sedimente im Golf von Neapel: 2. *Beitrage Mineralogie und Petrographie*, **8**, 1–20.
Muller, H. J. (1929). The gene as the basis of life. In *Proceedings of the Inter-*

*National Congress of Plant Sciences 1926*, ed. B. M. Dugger, pp. 897–921. Menasha, Wisconsin: George Banta.

Mumpton, F. A. & Ormsby, W. C. (1976). Morphology of zeolites in sedimentary rocks by scanning electron microscopy. *Clays and Clay Minerals*, **24**, 1–23.

Murphy, P. J., Posner, A. M. & Quirk, J. P. (1976a). Characterization of partially neutralised ferric nitrate solutions. *Journal of Colloid and Interface Science*, **56**, 270–83.

Murphy, P. J., Posner, A. M. & Quirk, J. P. (1976b). Characterization of partially neutralised ferric chloride solutions. *Journal of Colloid and Interface Science*, **56**, 284–97.

Murphy, P. J., Posner, A. M. & Quirk, J. P. (1976c). Characterization of partially neutralised ferric perchlorate solutions. *Journal of Colloid and Interface Science*, **56**, 298–311.

Nagasawa, K. & Miyazaki, S. (1976). Mineralogical properties of halloysite as related to its genesis. In *Proceedings of the International Clay Conference 1975*, ed. S. W. Bailey, pp. 257–65. Wilmette Illinois: Applied Publishing Ltd.

Nakai, M. & Yoshinaga, N. (1978). Poster display: occurrence of a fibrous iron mineral in some soils from Japan and Scotland. *Book of Summaries for the International Clay Conference 1978*, p. 653. London: The Mineralogical Society.

Nakazawa, H. & Morimoto, N. & Watanabe, E. (1975). Direct observations of metal vacancies by high-resolution electron microscopy. Part I: 4C type pyrrhotite ($Fe_7S_8$). *American Mineralogist*, **60**, 359–66.

Nalovic, Lj., Pedro, G. & Janot, C. (1976). Demonstration by Mössbauer spectroscopy of the role played by transitional trace elements in the crystallogenesis of iron hydroxides (III). In *Proceedings of the International Clay Conference 1975*, ed. S. W. Bailey, pp. 601–10. Wilmette Illinois: Applied Publishing Ltd.

Naylor, R. & Gilham, P. T. (1966). Studies on some interactions and reactions of oligonucleotides in aqueous solution. *Biochemistry*, **5**, 2722–8.

Negrón-Mendoza, A. & Ponnamperuma, C. (1978). Interconversion of biologically important carboxylic acids by radiation. In *Origin of Life*, ed. H. Noda, pp. 101–4. Tokyo: Center for Academic Publications Japan.

Newnham, R. E. (1961). A refinement of dickite structure and some remarks on polymorphism of kaolin minerals. *Mineral Magazine*, **32**, 683–704.

Nicholas, H. J. (1973). Terpenes. In *Phytochemistry*, *vol. II*, ed. L. P. Miller, pp. 254–309. New York: Van Nostrand Reinhold.

Nicholson, D. E. (1974). *Metabolic Pathways*. Colnbrook, England: Koch-Light Laboratories Ltd.

Ninio, J. (1975). Kinetic amplification of enzyme discrimination. *Biochimie*, **57**, 587–95.

Nissenbaum, A. (1976). Scavenging of soluble organic matter from the prebiotic oceans. *Origins of Life*, **7**, 413–16.

Norrish, K. (1973). Factors in the weathering of mica to vermiculite. In *Proceedings of the International Clay Conference 1972*, ed. J. M. Serratosa, pp. 417–32. Madrid: Division de Ciencias, CSIC.

Nyholm, R. S. & Tobe, M. L. (1963). The stabilisation of oxidation states of the transition metals. *Advances in Inorganic Chemistry and Radiochemistry*, **5**, 1–40.



Odom, D. G., Rao, M., Lawless, J. G. & Oró, J. (1979). Association of nucleotides with homoionic clays. *Journal of Molecular Evolution*, **12**, 365–7.

Oparin, A. I. (1924). *Proiskhozhdenie Zhizny*. Moscow: Izd. Moskovskiĭ Rabochiĭ. (Reprinted in translation in Bernal, 1967, pp. 199–234.)

Oparin, A. I. (1938). *Origin of Life*. Reprinted 1953. New York: Dover.

Oparin, A. I. (1957). *The Origin of Life on the Earth*. Edinburgh: Oliver & Boyd,

Orgel, L. E. (1968). Evolution of the genetic apparatus. *Journal of Molecular Biology*, **38**, 381–93.

Orgel, L. E. (1972). A possible step in the origin of the genetic code. *Israel Journal of Chemistry*, **10**, 287–92.

Orgel, L. E. (1979). Selection *in vitro*. *Proceedings of the Royal Society B*, **205**. 435–42.

Orgel, L. E. & Lohrmann, R. (1974). Prebiotic chemistry and nucleic acid replication. *Accounts of Chemical Research*, **7**, 368–77.

Orgel, L. E. & Sulston, J. E. (1971). Polynucleotide replication and the origin of life. In *Prebiotic and Biochemical Evolution*, ed. A. P. Kimball & J. Oro, pp. 89–94. Amsterdam: North Holland.

Oró, J. (1960). Synthesis of adenine from ammonium cyanide. *Biochemical and Biophysical Research Communications*, **2**, 407–12.

Oró, J. (1965). Stages and mechanisms of prebiological organic synthesis. In *The Origins of Prebiological Systems*, ed. S. W. Fox, pp. 137–71. New York: Academic.

Oró, J. & Kamat, S. S. (1961). Amino acid synthesis from hydrogen cyanide under possible primitive Earth conditions. *Nature*, **190**, 442–3.

Oró, J. & Kimball, A. P. (1961). Synthesis of purines under possible primitive Earth conditions. 1. Adenine from hydrogen cyanide. *Archives of Biochemistry and Biophysics*, **94**, 217–27.

Oró, J., Sherwood, E., Eichberg, J. & Epps, D. (1978). Formation of phospholipids under primitive Earth conditions and the role of membranes in prebiological evolution. In *Light Transducing Membranes*, ed. D. W. Deamer, pp. 1–21. New York: Academic.

Österberg, R., Orgel, L. E. & Lohrmann, R. (1973). Further studies of urea-catalysed phosphorylation reactions. *Journal of Molecular Evolution*, **2**, 231–4.

Paecht-Horowitz, M. (1974). Micelles and solid surfaces as amino acid polymerisation propagators. In *The Origin of Life and Evolutionary Biochemistry*, ed. K. Dose, S. W. Fox, G. A. Deborin & T. E. Pavlovskaya, pp. 373–85. New York: Plenum.

Paecht-Horowitz, M., Berger, J. & Katchalsky, A. (1970). Prebiotic synthesis of polypeptides by heterogeneous polycondensation of amino acid adenylates. *Nature*, **228**, 636–9.

Palm, C. & Calvin, M. (1962). Primordial organic chemistry. 1. Compounds resulting from irradiation of $^{14}CH_4$. *Journal of the American Chemical Society*, **84**, 2115–21.

Parfitt, R. L. & Henmi, T. (1980). Structure of some allophanes from New Zealand. *Clays and Clay Minerals*, **28**, 285–94.

Parham, W. E. (1969). Formation of halloysite from feldspar: low temperature artificial weathering versus natural weathering. *Clays and Clay Minerals*, **17**, 13–22.

Pasteur, L. (1848). Sur les relations qui peuvent exister entre la forme cristalline, la composition chimique et le sens de la polarisation rotatoire. *Annales de Chimie Physique*, **24**, 442–59.

Penrose, L. S. (1959a). Automatic mechanical self-reproduction. *New Biology*, **28**, 92–117.

Penrose, L. S. (1959b). Self-reproducing machines. *Scientific American*, **200**, no. 6, 105–14.

Perutz, M. F. (1978). Haemoglobin structure and respiratory transport. *Scientific American*, **239**, no. 6, 68–86.

Pfeil, E. & Ruckert, H. (1961). Über die Formaldehydkondensation. *Annalen Chemie*, **641**, 121–31.

Phillips, D. C. (1966). The three-dimensional structure of an enzyme molecule. *Scientific American*, **215**, no. 5, 78–90.

Pincock, R. E., Bradshaw, R. P. & Perkins, R. R. (1974). Spontaneous and induced generation of optical activity in racemic 1,1″-binaphthyl. *Journal of Molecular Evolution*, **4**, 67–75.

Pirie, N. W. (1951). In *The Physical Basis of Life*, by J. D. Bernal, p. 71. London: Routledge & Kegan Paul.

Pirie, N. W. (1957). Some assumptions underlying discussions on the origins of life. *Annals of the New York Academy of Sciences*, **69**, 369–76.

Pirie, N. W. (1959). Chemical diversity and the origins of life. In *The Origin of Life on the Earth*, ed. F. Clark & R. L. M. Synge, pp. 76–83. London: Pergamon.

Plançon, A. & Tchoubar, C. (1977). Determination of structural defects in phyllosilicates by X-ray powder diffraction – II. Nature and proportion of defects in natural kaolinites. *Clays and Clay Minerals*, **25**, 436–50.

Ponnamperuma, C. (1965). Abiological synthesis of some nucleic acid constituents. In *The Origins of Prebiological Systems*, ed. S. W. Fox, pp. 221–42. New York: Academic.

Ponnamperuma, C. (1978). Prebiotic molecular evolution. In *Origin of Life*, ed. H. Noda, pp. 67–81. Tokyo: Center for Academic Publications Japan.

Ponnamperuma, C. & Peterson, E. (1965). Peptide synthesis from amino acids in aqueous solution. *Science*, **147**, 1572–4.

Ponnamperuma, C., Shimoyama, A., Yamada, M., Hobo, T. & Pal, R. (1977). Possible surface reactions on Mars: implications for the Viking biology results. *Science*, **197**, 455–7.

Quayle, J. R. & Ferenci, T. (1978). Evolutionary aspects of autotrophy. *Microbiological Reviews*, **42**, 251–73.

Rabinowitz, J., Flores, J., Krebsbach, R. & Rogers, G. (1969). Peptide formation in the presence of linear or cyclic polyphosphates. *Nature*, **224**, 795–6.

Radoslovich, E. W. (1960). The structure of muscovite, K $Al_2(Si_3Al)O_{10}(OH)_2$. *Acta Crystallographica*, **13**, 919–32.

Radoslovich, E. W. (1963). The cell dimensions and symmetry of layer-lattice silicates. IV. Interatomic forces. *American Mineralogist*, **48**, 76–99.

Rausch, W. V. & Bale, H. D. (1964). Small angle X-ray scattering from hydrolysed aluminium nitrate solutions. *Journal of Chemical Physics*, **40**, 3391–4.

Rausell-Colom, J. A., Sanz, J., Fernandez, M. & Serratosa, J. M. (1979). Distribution of octahedral ions in phlogopites and biotites. In *International Clay Conference 1978*, ed. M. M. Mortland & V. C. Farmer, pp. 27–36. Amsterdam: Elsevier.

Rautureau, M. & Tchoubar, C. (1976). Structural analysis of sepiolite by selected area electron diffraction – relations with physico-chemical properties. *Clays and Clay Minerals*, **24**, 43–9.

Rees, D. A. (1977). *Polysaccharide Shapes*. London: Chapman Hall.

Rees, D. A. & Welsh, E. J. (1977). Secondary and tertiary structure of polysaccharides in solutions and gels. *Angewandte Chemie, International Edition*, **16**, 214–24.

Reiche, H. & Bard, A. J. (1979). Heterogeneous photosynthetic production of amino acids from methane–ammonia–water at $Pt/TiO_2$. Implications in chemical evolution. *Journal of the American Chemical Society*, **101**, 3127–8.

Reid, C. & Orgel, L. E. (1967). Synthesis of sugars in potentially prebiotic conditions. *Nature*, **216**, 455.

Rein, R., Nir, S. & Stamadiadou, M. N. (1971). Photochemical survival principle in molecular evolution. *Journal of Theoretical Biology*, **33**, 309–18.

Rensch, B. (1959). *Evolution above the Species Level*. London: Methuen.

Reynolds, R. C. (1965). An X-ray study of an ethylene glycol–montmorillonite complex. *American Mineralogist*, **50**, 990–1001.

Rich, A. (1965). Evolutionary problems in the synthesis of proteins. In *Evolving Genes and Proteins*, ed. V. Bryson & H. J. Vogel, pp. 453–68. New York: Academic.

Richmond, M. H. & Smith, D. C. (1979). The cell as habitat: A discussion organised by M. H. Richmond & D. C. Smith, F.R.S. *Proceedings of the Royal Society B*, **204**, 113–286.

Rickard, D. T. (1969). The chemistry of iron sulphide formation at low temperatures. In *Stockholm Contributions in Geology*, vol. 20, ed. I. Hessland, pp. 67–95. Stockholm: The University of Stockholm.

Rimsaite, J. (1973). Genesis of chlorite, vermiculite, serpentine, talc and secondary oxides in ultrabasic rocks. In *Proceedings of the International Clay Conference 1972*, ed. J. M. Serratosa, pp. 291–302. Madrid: Division de Ciencias, CSIC.

Ring, D., Wolman, Y., Friedman, N. & Miller, S. L. (1972). Prebiotic synthesis of hydrophobic and protein amino acids. *Proceedings of the National Academy of Sciences USA*, **69**, 765–8.

Rohlfing, D. L. (1976). Thermal polyamino acids: synthesis at less than 100 °C. *Science*, **193**, 68–70.

Rosen, B. P. (1978). *Bacterial Transport*. New York: Dekker.

Ross, M., Takeda, H. & Wones, D. R. (1966). Mica polytypes: Systematic description and identification. *Science*, **151**, 191–3.

Rousseaux, J. M., Gomez Laverde, C., Nathan, Y. & Rouxhet, P. G. (1973). Correlation between the hydroxyl stretching bands and the chemical composition of trioctahedral micas. In *Proceedings of the International Clay Conference 1972*, ed. J. M. Serratosa, pp. 89–98. Madrid: Division de Ciencias, CSIC.

Rouxhet, P. G., Samudacheata, N., Jacobs, H. & Anton, O. (1977). Attribution of the OH stretching bands of kaolinite. *Clay Minerals*, **12**, 171–9.

Roy, D. M. & Roy, R. (1954). An experimental study of the formation and properties of synthetic serpentines and related layer silicate minerals. *American Mineralogist*, **39**, 957–75.

Rubey, W. W. (1951). Geologic history of seawater: An attempt to state the problem. *Bulletin of the Geological Society of America*, **62**, 1111–47.

Ruckert, H., Pfeil, E. & Scharf, G. (1965). Der sterische Verlauf der Zuckerbildung. *Chemische Berichte*, **98**, 2558–65.

Sagan, C. & Khare, B. N. (1971). Long wavelength ultraviolet photoproduction of amino acids on the primitive Earth. *Science*, **173**, 417–20.

Sanchez, R. A., Ferris, J. P. & Orgel, L. E. (1966a). Conditions for purine synthesis: Did prebiotic synthesis occur at low temperatures? *Science*, **153**, 72–3.

Sanchez, R. A., Ferris, J. P. & Orgel, L. E. (1966b). Cyanoacetylene in prebiotic synthesis. *Science*, **154**, 784–5.

Sanchez, R. A., Ferris, J. P. & Orgel, L. E. (1967). Synthesis of purine precursors and amino acids from aqueous hydrogen cyanide. *Journal of Molecular Biology*, **30**, 223–53.

Satir, P. (1974). How cilia move. *Scientific American*, **231**, no. 4, 44–52.

Saunders, P. T. & Ho, M. W. (1976). On the increase in complexity in evolution. *Journal of Theoretical Biology*, **63**, 375–84.

Schidlowski, M. (1978). Evolution of the Earth's atmosphere: current state and exploratory concepts. In *Origin of Life*, ed. H. Noda, pp. 3–20. Tokyo: Center for Academic Publications Japan.

Schmalhausen, I. I. (1968). *The Origin of Terrestrial Vertebrates*. New York: Academic.

Schopf, J. W. (1975). Precambrian paleobiology: Problems and perspectives. In *Annual Review of Earth and Planet Sciences*, vol. 3, ed. F. A. Donath, F. G. Stehli & G. W. Wetherill, pp. 213–49. Palo Alto, California: Annual Reviews Inc.

Schopf, J. W. (1976). Are the oldest 'fossils' fossils? *Origins of Life*, **7**, 19–36.

Schrauzer, G. N. & Guth, T. D. (1977). Photolysis of water and photoreduction of nitrogen on titanium dioxide. *Journal of the American Chemical Society*, **99**, 7189–93.

Schwartz, A. W. (1971). Phosphate solubilisation and activation on the primitive Earth. In *Chemical Evolution and the Origin of Life*, ed. R. Buvet & C. Ponnamperuma, pp. 207–15. Amsterdam: North-Holland.

Schwartz, A. W. (1981). Chemical evolution – the genesis of the first organic compounds. In *Marine Organic Chemistry*, ed. E. K. Duursma & R. Dawson, pp. 7–30. Amsterdam: Elsevier.

Schwartz, A. W. & Chittenden, G. J. F. (1977). Synthesis of uracil and thymine under simulated prebiotic conditions. *Biosystems*, **9**, 87–92.

Schwertmann, U. (1979). Non-crystalline and accessory minerals. In *International Clay Conference 1978*, ed. M. M. Mortland & V. C. Farmer, pp. 491–9. Amsterdam: Elsevier.

Schwertmann, U. & Thalmann, H. (1976). The influence of [Fe(II)], [Si] and pH on the formation of lepidocrocite and ferrihydrite during oxidation of aqueous $FeCl_2$ solutions. *Clay Minerals*, **11**, 189–200.



Serratosa, J. M. (1979). Surface properties of fibrous clay minerals (paygorskite and sepiolite). In *International Clay Conference 1978*, ed. M. M. Mortland & V. C. Farmer, pp. 99–109. Amsterdam: Elsevier.

Sheppard, R. A. & Gude, A. J. (1973). *Zeolites and Associated Authigenic Silicate Minerals in Tuffaceous Rocks of the Big Sandy Formation, Mohave County, Arizona*. U.S. Geological Survey Professional Paper, No. 830.

Sherrington, C. S. (1940). *Man on his Nature*, p. 88. Cambridge: Cambridge University Press.

Sherwood, E., Nooner, D. W., Eichberg, J., Epps, D. E. & Oró, J. (1978). Prebiotic condensation reactions using cyanamide. In *Origin of Life*, ed. H. Noda, pp. 105–11. Tokyo: Center for Academic Publications Japan.

Shigemasa, Y., Sakazawa, C., Nakashima, R. & Matsuura, T. (1978). Selectivities in the formose reaction. In *Origin of Life*, ed. H. Noda, pp. 211–16. Tokyo: Center for Academic Publications Japan.

Shirozu, H. & Bailey, S. W. (1966). Crystal structure of a two-layer Mg-vermiculite. *American Mineralogist*, **51**, 1124–43.

Sieskind, O. & Ourisson, G. (1971). Interactions argile–matière organique: formation de complexes entre la montmorillonite et les acides stéarique et béhénique. *Comptes Rendus de L'Académie des Sciences Paris*, C272, 1885–8.

Siffert, B. (1979). Genesis and synthesis of clays and clay minerals: recent developments and future prospects. In *International Clay Conference 1978*, ed. M. M. Mortland & V. C. Farmer, pp. 337–47. Amsterdam: Elsevier.

Siffert, B. & Wey, R. (1973). Contribution à la connaissance de la synthèse des kaolins. In *Proceedings of the International Clay Conference 1972*, ed. J. M. Serratosa, pp. 159–72. Madrid: Division de Ciencias, CSIC.

Sillen, L. G. (1965). Oxidation state of the Earth's ocean and atmosphere. I. A model calculation on earlier states. The myth of the 'probiotic soup'. *Arkiv för Kemi*, **24**, 431–56.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, **106**, 467–82.

Sipling, P. J. & Yund, R. A. (1974). Kinetics of Al/Si disordering in alkali feldspars. In *Geological Transport and Kinetics*, ed. A. W. Hoffmann, B. J. Giletti, H. S. Yoder & R. A. Yund, pp. 185–94. Washington: Carnegie Institute Publication 634.

Smith, A., Folsome, C. & Bahadur, K. (1981). Carbon dioxide reduction and nitrogenase activity in organo-molybdenum microstructures, *Experientia*, **37**, 357–9.

Smith, J. V. & Yoder, H. S. (1956). Experimental and theoretical studies of mica polymorphs. *Mineral Magazine*, **31**, 209–35.

Smith, R. W. & Hem, J. D. (1972). Effect of ageing on aluminium hydroxide complexes in dilute aqueous solutions. U.S. Geological Survey Water Supply Paper 1827-D.

Solomon, D. H., Loft, B. C. & Swift, J. D. (1968a). Reactions catalysed by minerals. IV. The mechanism of the benzidine blue reaction on silicate minerals. *Clay Minerals*, **7**, 389–98.

Solomon, D. H., Loft, B. C. & Swift, J. D. (1968b). Reactions catalysed by minerals. V. The reaction of leuco dyes and unsaturated organic compounds with clay minerals. *Clay Minerals*, **7**, 399–408.

Sonneborn, T. M. (1965). Degeneracy of the genetic code: extent, nature and genetic implications. In *Evolving Genes and Proteins*, ed. V. Bryson & H. J. Vogel, pp. 377–97. New York: Academic.

Spiegelman, S. (1970). Extracellular evolution of replicating molecules. In *The Neurosciences: Second Study Program*, ed. F. O. Schmitt, pp. 927–45. New York: Rockefeller University Press.

Spiegelman, S., Mills, D. R. & Kramer, F. R. (1975). The extracellular evolution of structure in replicating RNA molecules. In *Stability and Origin of Biological Information*, ed. I. R. Miller, pp. 123–72. New York: John Wiley & Sons.

Spirin, A. S. (1978). Energetics of the ribisome. *Progress in Nucleic Acid Research and Molecular Biology*, **21**, 39–62.

Spiro, T. G., Allerton, S. E., Renner, J., Terzis, A., Bils, R. & Saltman, P. (1966). The hydrolytic polymerisation of iron (III). *Journal of the American Chemical Society*, **88**, 2721–6.

Steinfink, H. (1958a). The crystal structure of chlorite: I. A monoclinic polymorph. *Acta Crystallographica*, **11**, 191–5.

Steinfink, H. (1958b). The crystal structure of chlorite: II. A triclinic polymorph. *Acta Crystallographica*, **11**, 195–8.

Stillinger, F. H. & Wasserman, Z. (1978). Molecular recognition and self-organisation in fluorinated hydrocarbons. *Journal of Physical Chemistry*, **82**, 929–40.

Stoeckenius, W. (1976). The purple membrane of salt-loving bacteria. *Scientific American*, **234**, no. 6, 38–46.

Stroud, R. M. (1974). A family of protein-cutting proteins. *Scientific American*, **231**, no. 1, 74–88.

Suquet, H., Iiyama, J. T., Kodama, H. & Pezerat, H. (1977). Synthesis and swelling properties of saponite with increasing layer charge. *Clays and Clay Minerals*, **25**, 231–42.

Szent-Györgyi, A. (1972). *The Living State*. New York: Academic.

Takeda, H. & Ross, M. (1975). Mica polytypism: Dissimilarities in the crystal structure of coexisting 1M and $2M_1$ biotite. *American Mineralogist*, **60**, 1030–40.

Taylor, P. (1980). The stereochemistry of iron sulphides – a structural rationale for crystallisation of some metastable phases from aqueous solution. *American Mineralogist*, **65**, 1026–30.

Taylor, W. H. (1933). The structure of sanidine and other felspars. *Zeitschrift für Kristallographie*, **85**, 425–42.

Taylor, W. H., Darbyshire, J. A. & Strunz, H. (1934). An X-ray investigation of the felspars. *Zeitschrift für Kristallographie*, **87**, 464–98.

Tazaki, K. (1979a). Scanning electron microscope study of imogolite formation from plagioclase. *Clays and Clay Minerals*, **27**, 209–12.

Tazaki, K. (1979b). Micromorphology of halloysite produced by weathering of plagioclase in volcanic ash. In *International Clay Conference 1978*, ed. M. M. Mortland & V. C. Farmer, pp. 415–22. Amsterdam: Elsevier.



Theng, B. K. G. (1974). *The Chemistry of Clay-Organic Reactions*. Bristol: Hilger.

Thiemann, W. (1974). Disproportionation of enantiomers by precipitation. *Journal of Molecular Evolution*, **4**, 85–97.

Thiemann, W. & Darge, W. (1974). Experimental attempts for the study of the origin of optical activity on Earth. *Origins of Life*, **5**, 263–83.

Thilo, E. (1962). Condensed phosphates and arsenates. *Advances in Inorganic Chemistry and Radiochemistry*, **4**, 1–75.

Thilo, E. & Wieker, W. (1957). Die anionenhydrolyse kondensierter phosphate in verdünnter wässriger lösung. *Zeitschrift für anorganische und allgemeine Chemie*, **291**, 164–85.

Thompson, T. D. & Brindley, G. W. (1969). Absorption of pyrimidines, purines and nucleosides by Na-, Mg- and Cu(II)-illite (clay-organic studies XVI). *American Mineralogist*, **54**, 858–68.

Topal, M. D. & Fresco, J. R. (1976). Complementary base pairing and the origin of substitution mutations. *Nature*, **263**, 285–9.

Toupance, G., Raulin, F. & Buvet, R. (1971). Primary transformation processes under the influence of energy for models of primordial atmospheres. In *Chemical Evolution and the Origin of Life*, ed. R. Buvet & C. Ponnamperuma, pp. 83–95. Amsterdam: North-Holland.

Troland, L. T. (1914). The chemical origin and regulation of life. *The Monist*, **24**, 92–133.

Tseng, S-S. & Chang, S. (1975). Photochemical synthesis of simple organic free radicals on simulated planetary surfaces – an ESR study. *Origins of Life*, **6**, 61–73.

Turner, R. C. & Ross, G. J. (1970). Conditions in solution during the formation of gibbsite in dilute Al salt solutions. 4. Effect of Cl concentration and temperature and a proposed mechanism for gibbsite formation. *Canadian Journal of Chemistry*, **48**, 723–9.

Urey, H. C. (1952). *The Planets*. New Haven, Connecticut: Yale University Press.

Van Olphen, H. (1966). Maya blue: A clay-organic pigment? *Science*, **154**, 645–6.

Van Olphen, H. (1977). *An Introduction to Clay Colloid Chemistry*, 2nd edn. New York: John Wiley & Sons.

Van Oosterwyck-Gastuche, M. C. & La Iglesia, A. (1978). Kaolinite synthesis. II: A review and discussion of the factors influencing the rate process. *Clays and Clay Minerals*, **26**, 409–17.

Van Wazer, J. R. (1958). *Phosphorus and its Compounds*, vol. I. New York: Interscience.

Vedder, W. (1964). Correlations between infrared spectrum and chemical composition of mica. *American Mineralogist*, **49**, 736–68.

Veitch, L. G. & Radoslovich, E. W. (1963). The cell dimensions and symmetry of layer-lattice silicates: III. Octahedral ordering. *American Mineralogist*, **48**, 62–75.

Velde, B. (1965). Experimental determination of muscovite polymorph stabilities. *American Mineralogist*, **50**, 436–49.

Velde, B. (1977). *Clays and Clay Minerals in Natural and Synthetic Systems.* Amsterdam: Elsevier.

Verwey, E. J. W. & Heilmann, E. L. (1947). Physical properties and cation arrangement of oxides with spinel structures. I. Cation arrangement in spinels. *Journal of Chemical Physics*, **15**, 174–80.

Wada, K. (1979). Structural formulas for allophanes. In *International Clay Conference 1978*, ed. M. M. Mortland & V. C. Farmer, pp. 537–45. Amsterdam: Elsevier.

Wada, S. I. & Wada, K. (1977). Density and structure of allophane. *Clay Minerals*, **12**, 289–98.

Waddington, C. H. (1957). *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology.* London: Allen & Unwin.

Waddington, C. H. (1968). The basic ideas of biology. In *Towards a Theoretical Biology, 1, Prolegomena*, ed. C. H. Waddington, p. 3. Edinburgh: Edinburgh University Press.

Wagener, K. (1974). Amplification processes for small primary differences in the properties of enantiomers. *Journal of Molecular Evolution*, **4**, 77–84.

Wakamatsu, H., Yamada, Y., Saito, T., Kumashiro, I. & Takenishi, T. (1966). Synthesis of adenine by oligomerisation of hydrogen cyanide. *Journal of Organic Chemistry*, **31**, 2035–6.

Wald, G. (1957). The origin of optical activity. *Annals of the New York Academy of Sciences*, **69**, 352–68.

Walker, G. L. (1971). *On the Formation of Sugars from Formaldehyde.* Ph.D. Thesis, University of Glasgow.

Walker, J. C. G. (1976a). Implications for atmospheric evolution of the inhomogeneous accretion model of the origin of the Earth. In *The Early History of the Earth*, ed. B. F. Windley, pp. 537–46. London: John Wiley & Sons.

Walker, J. C. G. (1976b). *Evolution of the Atmosphere.* New York: Hafner.

Watson, D. M. S. (1946). The evolution of the proboscidea. *Biological Reviews*, **21**, 15–29.

Watson, J. D. (1976). *Molecular Biology of the Gene*, 3rd edn. Menlo Park, California: Benjamin.

Watson, J. D. & Crick, F. H. C. (1953). Genetical implications of the structure of deoxyribonucleic acid. *Nature*, **171**, 964–7.

Weir, A. H., Nixon, H. L. & Woods, R. D. (1962). Measurement of thickness of dispersed clay flakes with the electron microscope. *Clays and Clay Minerals*, **9**, 419–23.

Weiss, A. (1963). Mica-type layer silicates with alkylammonium ions. *Clays and Clay Minerals*, **10**, 191–224.

Weiss, A. (1981). Replication and evolution in inorganic systems. *Angewandte Chemie, International Edition*, **20**, 850–60.

Weiss, A. & Roloff, G. (1966). *Proceedings of the International Clay Conference (Jerusalem)*, vol. 1, 263–75.

Weiss, A. H., Trigerman, S., Dunnells, G., Likholobov, V. A. & Biron, E. (1979). Ethylene glycol from formaldehyde. *Industrial and Engineering Process Design and Development*, **18**, 522–7.

Wells, A. F. (1975). *Structural Inorganic Chemistry.* Oxford: Clarendon.

Wells, N., Childs, C. W. & Downes, C. J. (1977). Silica Springs, Tongariro National Park, New Zealand – analyses of the spring water and characterisation of the alumino-silicate deposit. *Geochimica et Cosmochimica Acta*, **41**, 1497–1506.

Wey, R. & Siffert, B. (1962). Réactions de la silice monomoléculaire en solution avec les ions $Al^{3+}$ et $Mg^{2+}$. In *Genèse et Synthèse des Argiles*, ed. M. R. Hocart, pp. 11–23. Paris: CNRS.

White, H. B. (1976). Coenzymes as fossils of an earlier metabolic state. *Journal of Molecular Evolution*, **7**, 101–4.

Wickner, W. (1980). Assembly of proteins into membranes. *Science*, **210**, 861–8.

Wilkins, R. W. T. (1967). The hydroxyl-stretching region of the biotite mica spectrum. *Mineral Magazine*, **36**, 325–33.

Williams, D. G. & Garey, C. L. (1974). Crystal imperfections with regard to direction in kaolinite minerals. *Clays and Clay Minerals*, **22**, 117–25.

Wilson, D. B. (1978). Cellular transport mechanisms. *Annual Review of Biochemistry*, **47**, 933–65.

Wilson, M. D. & Pittman, E. D. (1977). Authigenic clays in sandstones: Recognition and influence on reservoir properties and paleoenvironmental analysis. *Journal of Sedimentary Petrology*, **47**, 3–31.

Winterburn, P. J. (1974). Polysaccharide structure and function. In *Companion to Biochemistry*, ed. A. T. Bull, J. R. Lagnado, J. O. Thomas & K. F. Tipton, pp. 307–41. London: Longman.

Woese, C. R. (1967). *The Genetic Code.* New York: Harper & Row.

Woese, C. R., Magrum, L. J. & Fox, G. E. (1978). Archaebacteria. *Journal of Molecular Evolution*, **11**, 245–52.

Wolfe, R. W. & Giese, R. F. (1973). A quantitative study of one-layer polytypism in kaolin minerals. In *Proceedings of the International Clay Conference 1972*, ed. J. M. Serratosa, pp. 27–33. Madrid: Division de Ciencias, CSIC.

Wolman, Y., Haverland, W. J. & Miller, S. L. (1972). Nonprotein amino acids from spark discharges and their comparison with the Murchison meteorite amino acids. *Proceedings of the National Academy of Sciences USA*, **69**, 809–11.

Wolman, Y., Miller, S. L., Ibanez, J. & Oró, J. (1971). Formaldehyde and amonnia as precursors to prebiotic amino acids. *Science*, **174**, 1039.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the 6th International Congress of Genetics*, **1**, 356–66.

Yamagata, Y. (1966). A hypothesis for the asymmetric appearance of biomolecules on Earth. *Journal of Theoretical Biology*, **11**, 495–8.

Yershova, Z. P., Nikitina, A. P., Perfil'ev, Yu. D. & Babeshkin, A. M. (1976). Study of chamosites by gamma-resonance (Mössbauer) spectroscopy. In *Proceedings of the International Clay Conference 1975*, ed. S. W. Bailey, pp. 211–19. Wilmette, Illinois: Applied Publishing Ltd.

Yoder, H. S. & Eugster, H. P. (1955). Synthetic and natural muscovites. *Geochimica et Cosmochimica Acta*, **8**, 225–80.

Youatt, J. B. & Brown, R. D. (1981). Origins of chirality in Nature: A reassessment of the postulated role of bentonite. *Science*, **212**, 1145–6.

Zeeman, E. C. (1976). Catastrophe theory. *Scientific American*, **234**, no. 4, 65–83.

Zuckerkandl, E. (1975). The appearance of new structures and functions in proteins during evolution. *Journal of Molecular Evolution*, **7**, 1–57.

# Index

After a page entry, *d* indicates that a definition is to be found on that page, *e* indicates an electron micrograph and *s* a structure.