Pergamon

Tetrahedron, Vol. 53, No. 34, pp. 11493-11527, 1997

0040-4020/97 $17.00 + 0.00

PII: S0040-4020(97)00438-9

TETRAHEDRON REPORT NUMBER 425

Prebiotic Chemistry: A Bioorganic Perspective

John D Sutherland* and J Nicole Whitfield

The Dyson Perrins Laboratory, South Parks Road, Oxford OX1 3QY, UK.
*E-mail: john.sutherland@dpl.ox.ac.uk Tel/FAX: +44 1865 275644

Contents

1 Scope

The complexity and diversity of contemporary life forms appears at first sight to provide few clues towards the possible origins of life. Closer inspection however reveals extraordinary similarities in all organisms at the most basic of biochemical levels. Such properties include the universal reliance on the nucleic acids RNA and DNA for the propagation of genetic information, a single genetic code for protein synthesis and the presence of the same 20 amino acids in all proteins. These findings have led to the widely accepted notion that all extant life forms evolved from a single organism termed "the last common ancestor".[1] The search for the nature of such an organism has occupied many years of research and has seen approaches from many angles including chemistry, biology, geology and space research. In this report we survey results from bioorganic experiments aimed at shedding light on possible chemical mechanisms of early evolution and suggest how theories might be modified to accomodate recent experimental results.

2 The Primordial Earth

The formation of the Earth from a diffuse cloud of cosmic gas and dust occurred some 4.6 x 10^9 years ago (**Fig. 1**) but the frequency of impacts from meteorites and the high temperature of the molten surface during the subsequent 0.6 x 10^9 years, would have provided unsuitable conditions for stable life-forms to emerge.[2]

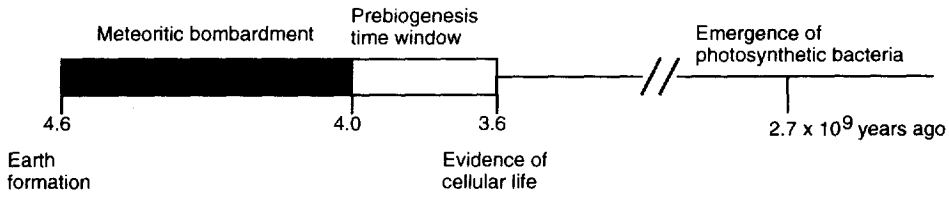


Fig. 1 Timescale for the emergence of the first primitive life forms

Once the level of meteoric bombardment had dropped and the rapid cooling process to form the Earth's crust was complete (*ca.* 4.0 x 10^9 years ago) it is proposed that bodies of water were formed and organic chemistry became established. The oldest known fossils date back to *ca.* 3.6 x 10^9 years ago[3] and show resemblance to modern blue-green algae. The complex translational apparatus found in these organisms suggests that by this time cellular life must have already been well established. Biogenesis from organic chemistry to a primitive cell must therefore have occurred in the relatively small time (at least in the geological sense) of 0.4 x 10^9 years.

There is little agreement on the nature of the primordial atmosphere with opinions ranging from it being strongly reducing to neutral. It is however generally accepted that there was no free oxygen until the advent of photosynthetic bacteria *ca.* 2.7 x 10^9 years ago.[4] Though probably not as reducing as first suggested by Oparin and Haldane in the 1930s,[5] this gives scope for efficient organic synthesis. In this setting, energy required for chemical synthesis would be available from the sun in the form of ultra violet radiation,[6] blocked today by high level ozone.

3 Prebiotic Feedstock Molecules

Prebiotic chemistry has a vital role to play in indicating the possible nature of an early chemical environment. Its aim is to establish by experiment: chemical structures, reactions and pathways which may have been involved in a process of self-organisation to form the last common ancestor. A select few small molecules are considered to comprise the building blocks for natural self-assembly,[7] *inter alia:* water, ammonia, hydrogen cyanide, formaldehyde, acetonitrile, acrylonitrile, cyanogen and cyanoacetylene. It is possible that a first phase of organic chemistry involved the aforementioned compounds bar water and formaldehyde. Subsequent melting of ice provides water and atmospheric chemistry, formaldehyde and a second phase of organic chemistry ensued. This scenario is suggested by the requirement for spatially or temporally separate cyanide and formaldehyde chemistries given the high association constant for the formation of glyconitrile when the two species are brought into contact in water. The identification of prebiotic feedstock molecules derives from many sources including the study of interstellar dust clouds and carbonaceous chondrites such as the Murchison meteorite[8] along with laboratory experiments designed to simulate the primordial atmosphere. Cyano compounds have been shown to be abundant in outer space[9] and have also been produced in many prebiotic experiments. For example, hydrogen cyanide is formed in good yield from gaseous mixtures of nitrogen, hydrogen and carbon monoxide in spark discharge experiments[10] or by the action of u.v. radiation on mixtures of methane and ammonia.[11] A spark discharge passed through methane and nitrogen or through HCN produces cyanoacetylene[12] and cyanogen[9] respectively. Similar experiments have demonstrated the formation of formaldehyde.[13] Not only do the facile formation and wide distribution of these compounds suggest their involvement in prebiogenesis on the primitive Earth but they open up a plethora of possible subsequent reaction pathways, driven by the high chemical potential of the multiple bonds they contain.

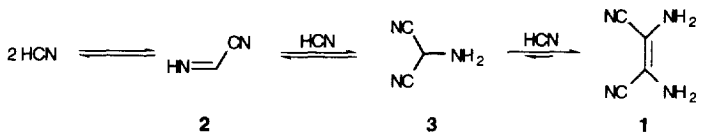
A diverse range of experiments on potentially prebiotic chemistry has been carried out in the laboratory to date and some of the more significant findings will now be reviewed.

4 Synthesis of Amino Acids

Some of the first experiments in prebiotic chemistry were carried out by Miller and Urey in the early 1950s and involved the production of proteinogenic amino acids in a spark discharge apparatus.[14] Miller passed an electric discharge through a gaseous mixture of methane, ammonia, hydrogen and water vapour and was able to isolate many organic compounds including the amino acids glycine, alanine and aspartic acid. He proposed that they were formed by a Strecker synthesis[15] involving reaction of aldehydes and HCN with ammonia. Other similar experiments have since been carried out and many other amino acids have been isolated.[16] Interestingly these have been found in the Murchison meteorite in approximately the same proportions.[8]

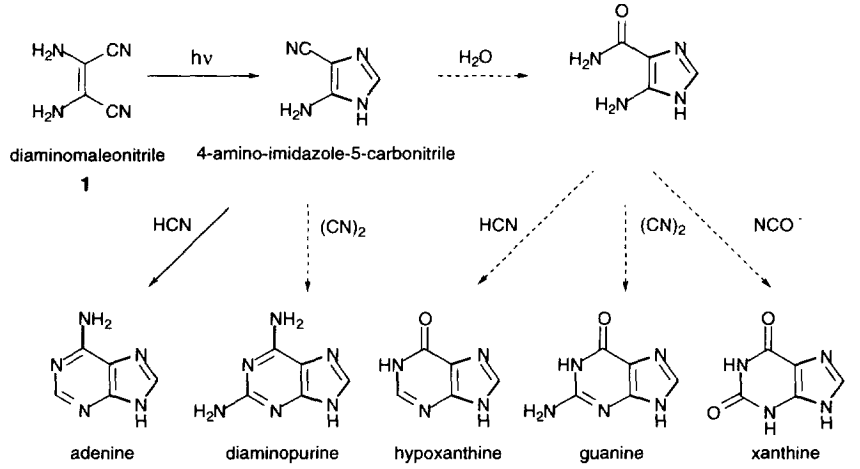
5 Synthesis of Purines and Chemistry of HCN

Another early series of important experiments was performed by Oró in 1960[17] who formed adenine in 0.5% yield by heating solutions of ammonium cyanide (>1.0 M) at 70°C for several days, then hydrolysing with 6N HCl. Lowe and co-workers confirmed this finding and also identified other purines in the same mixture.[18] A later study of the self-condensation reactions of HCN by Ferris and Orgel[19] revealed a stable tetramer, diaminomaleonitrile (DAMN) **1** as the only easily isolable oligomer formed irreversibly in aqueous solutions of 0.1-1.0 M HCN at room temperature and optimally at pH 9.2 (HCN has a pK_a of 9.2). Oligomerisation is thought to proceed *via* the less stable dimer (iminoacetonitrile) **2** and trimer (aminomalononitrile) **3**, **Scheme 1**.



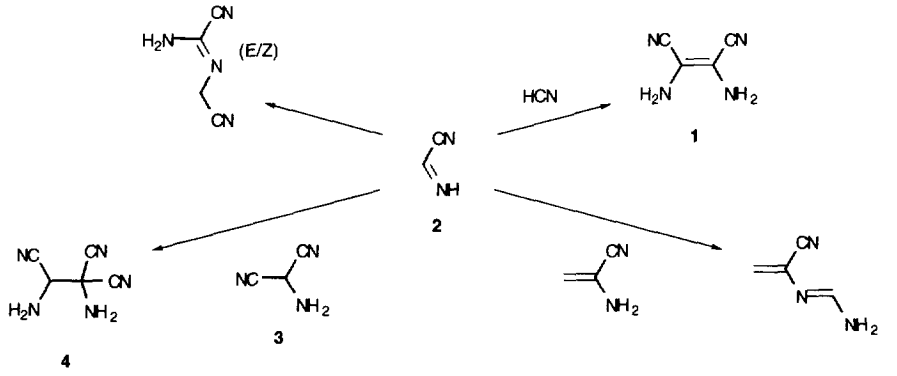
Scheme 1

During their subsequent studies of the chemistry of DAMN **1**, Ferris and Orgel elucidated the mechanism of the Oró experiment and developed a photochemical alternative which appeared to satisfy proposed prebiotic conditions more convincingly.[20] This mechanism is outlined in **Scheme 2**, which also illustrates the potential for production of many other purine bases under these conditions.



Scheme 2

More recently Eschenmoser's group have succeeded in the low temperature preparation and characterisation of the HCN dimer, *N*-(methylimino)acetonitrile **2**.[21] Some of the chemistry of this highly reactive compound is shown in **Scheme 3**.

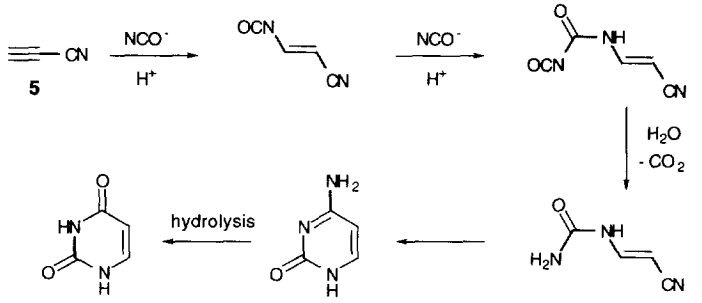


Scheme 3

Of particular interest is the formation of the new HCN pentamer **4** (an isomer of adenine) on treatment of **2** with the HCN trimer **3**. On reaction of **2** with HCN in the presence of base (triethylamine in the laboratory), DAMN **1** is formed indicating that the chemistry of its formation is predisposed under a variety of conditions.

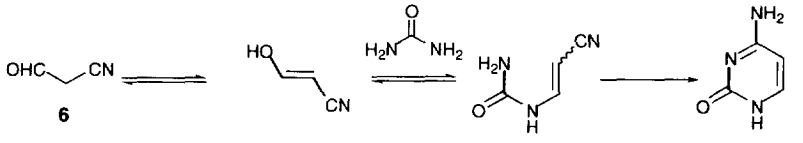
6 Synthesis of Pyrimidines

A role for cyanoacetylene **5** has been suggested in the formation of the pyrimidine bases. The finding of pyrimidines in the Murchison meteorite at one fifth of the concentration of purines, suggests that an efficient synthesis should be possible under potentially prebiotic conditions. It is recognised that cytosine is readily hydrolysed to uracil ($t_{1/2}$ = 200 years at 30°C in neutral solution)[22] and so experiments have focused on attempts to synthesise the former. The first experiments involved mixing 1.0 M cyanate (formed from the hydrolysis of cyanogen) with 0.1 M cyanoacetylene to produce cytosine in *ca.* 5% yield, after heating at 100°C for 24 h, **Scheme 4**.[12]



Scheme 4

Other routes to pyrimidines have been recently reported by Robertson and Miller.[23] The reaction of cyanoacetaldehyde **6** (the hydration product of cyanoacetylene) with urea (a likely prebiotic compound derived from cyanate)[24] was shown to produce cytosine and the measured yields of 30-50% were found to be directly proportional to the concentration of urea, **Scheme 5**.



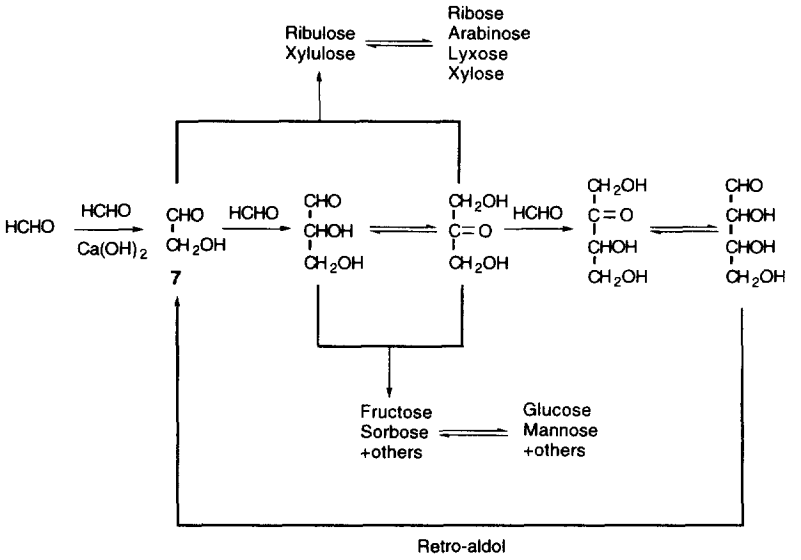
Scheme 5

Miller suggests that the high solubility of urea in water allows the required high concentrations to be reached and invokes the possibility of evaporating lagoons to enhance this concentration effect. The requirements for this model would be involatile, soluble reactants; criteria which are met in this instance. There is much room for conjecture in this theory but it cannot be dismissed. In an adaptation of an earlier experiment Robertson and Miller also mixed cyanoacetaldehyde **6** with guanidine and obtained the important intermediate diaminopyrimidine in

yields of up to 50% compared with the 2% previously reported[25] but again a requirement for high concentration was seen. Hydrolysis of diaminopyrimidine potentially provides a direct route to uracil.

7 Synthesis of Sugars

The prebiotic importance of formaldehyde was first demonstrated in 1861 when Butlerow performed what has become known as the "formose reaction".[26] The reaction involves the synthesis of a (sweet-tasting!) complex mixture of sugars from formaldehyde *via* an array of aldol condensations. It is initiated by an alkaline earth catalyst such as calcium hydroxide which promotes the formation of glycoaldehyde **7**. The cascade of reactions which follows produces, in high yield, a mixture of triose, tetrose, pentose, hexose and heptose sugars in aldose and ketose form, branched and straight chain in structure, **Scheme 6**.

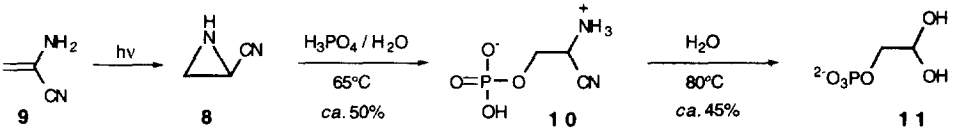


Scheme 6

8 Synthesis of Sugar Phosphates

The work of Eschenmoser's group at the ETH in Zürich has suggested that the study of phosphorylated sugars is of central importance, since the aldolisation chemistry of such compounds is inherently simplified by the lack of aldose/ketose equilibria. The fact that nature now uses sugars almost exclusively in their phosphorylated form adds credence to this suggestion although it is noted that phosphorylation also serves to capture sugars by rendering them membrane-insoluble.

A study of the potentially prebiotic compound aziridine-2-carbonitrile **8** (previously demonstrated to be a photochemical product of 2-aminopropenonitrile **9**) by the ETH group was extremely illuminating.[27] Reaction of **8** with inorganic phosphate produces 3-phosphoserinenitrile **10** which undergoes a retro-Strecker reaction to produce the phosphorylated sugar, glycoaldehyde phosphate **11**, **Scheme 7**.



9 8 10 11

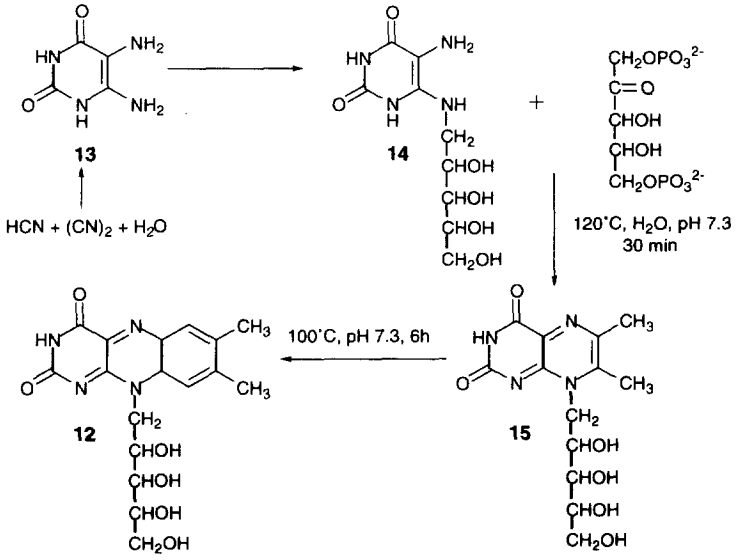
Scheme 7

On carrying out the analogous formose experiment Eschenmoser and co-workers discovered that the product mixture was indeed simpler.[28] When formaldehyde was included in the reaction mixture the products, formed in up to 50% yield, were mostly pentose-2,4-diphosphates and when formaldehyde was excluded an 80% yield of mostly hexose-2,4,6-triphosphates was obtained. The reactions showed a remarkably high degree of diastereoselectivity with the major isolable products being ribose-2,4-diphosphate and allose-2,4,6-triphosphate respectively. In both these sugars the chiral centres of the carbon chain show the same relative configuration, which is also that chosen by nature and that which derives from a minimisation of steric interactions in the transition state during formation. Hence, in this example the stereochemical course of the reaction is inherent in the reacting system itself, a point to which we will return later.

Individual sugar phosphates thus formed are more stable to hydrolysis than free sugars and could be envisaged as easier to concentrate from complex mixtures, their charged nature giving scope for selective adsorption onto mineral surfaces as a means of separation and concentration.

9 Synthesis of Riboflavin

There has been much discussion on the role of coenzymes in evolution which has led to a suggestion that they may represent "enzyme-fossils" from an earlier life-form.[29] The fact that many cofactors contain ribose is often taken as additional evidence for an 'RNA world'. Riboflavin **12** is the functionally important part of the coenzyme flavin adenine dinucleotide, FAD and has features common to many coenzymes including a heterocyclic core related to guanine and a ribose-sub-unit, usually phosphorylated at C5', **Scheme 8**.



Scheme 8

The heterocyclic moiety can be seen to derive from 5,6-diaminouracil **13** which itself can be made from HCN and cyanogen.[30] The subsequent coupling of this heterocyclic base to ribose can readily be envisaged but it is not immediately obvious how the carbocyclic portion of riboflavin could be synthesised. The biosynthesis requires several enzyme-catalysed steps but has been shown to proceed *in vitro* in the absence of enzymes by chemistry that can properly be described as predisposed.[7, 31] Heating the ribityl-diaminopyrimidine derivative **14** with a five carbon sugar derivative in aqueous solution for 30 min at pH 7 and 120°C in the absence of air produces the bicyclic lumazine derivative **15** which when heated for 6 hours at 100°C remarkably (and beautifully) produces riboflavin **12**. Isotopic labelling studies have revealed that the latter process follows the same subtle mechanism as the enzyme-catalysed process. This is an important point; a predisposed reaction need not proceed by an overtly simple mechanism. Indeed, in the sense that predisposition relies upon subtle reactivity features of functional groups that are contextually specific, we should anticipate certain reaction pathways to be substantially favoured in situations where a large number of outcomes are potentially possible. If we adhere to Eschenmoser's concept that the products of predisposed chemistry should be viewed as *elementary* structures then we should perhaps ascribe the term *complex* to those structures for which formation by predisposed reactions cannot be demonstrated (or imagined). This contrasts somewhat with conventional chemical reasoning in which *complex* is applied to those structures which, by virtue of functional group density, size, stereochemistry, concatenation *etc.*, are so beloved as targets by synthetic chemists. It follows from simple evolutionary reasoning that most natural products likely predate the enzymes that now catalyse their biosynthesis. It is also likely that the biosynthetic steps mirror prior uncatalysed but predisposed chemistry. Given the small number of enzymatic reaction types then a firm understanding of predisposition might provide inherently and operationally simple synthetic strategies to a variety of natural products. In a sense this has already been exploited in the realm of biomimetic chemistry but if we follow Eschenmoser's lead we might view the biosynthetic process as being *chemomimetic* rather than the chemical process as being *biomimetic*.[7]

10 Theories

Theories on the origin of life fall broadly into two categories.

(i) The *heterotrophic* origin of life has been crudely described as the 'soup theory' and relies on the notion that the primitive atmosphere was reducing[5] allowing a build up of organic molecules in oceans. It is supposed that once formed, the first organisms would have a continuous supply of the materials required for replication.

(ii) The *autotrophic* origin of life derives from the CO_2-rich model of the primitive Earth[32] and supposes that the potential for organic synthesis would be negligible under these less reducing conditions. The constitution of the first evolving systems would therefore be inorganic and a mechanism for the biosynthesis of organic material is expected to evolve at a later stage.

The various theories deriving from these two categories will now be discussed.

11 The RNA World

When faced with the question of how the modern interdependent system of functional proteins and information carrying nucleic acids arose, one is immediately faced with a paradox. Proteins are synthesised according to a precise sequence of nucleotides and are assembled at the RNA-containing ribosome from individual aminoacyl-transfer RNA molecules. Nucleic acids however are themselves reproduced during transcription by the action of highly optimised enzymes. It is difficult to envisage one system without the other but it is considered unlikely by

most workers in the field that such complex molecules and their intricate interdependence could have emerged simultaneously.[1] A theory gaining credence as research progresses is that there was a time before the origin of protein synthesis when life was comprised entirely of replicating RNA species, a scenario which has been termed the 'RNA world'.[33] On closer inspection RNA can be seen to possess many properties deemed essential to satisfy any acceptable definition of the term 'living'. Basic requirements for life include:

(i) the ability to carry hereditary information,

(ii) the ability to propagate this information by efficient replication,

(iii) a phenotype which is sequence-dependent.

RNA is commonly found in modern biology in what are considered to be some of the oldest biochemical processes. It has an essential role in protein synthesis where native DNA is transcribed into a strand of m-RNA and t-RNA molecules ensure that the correct amino acids are delivered to the ribosome (itself rich in RNA) where the peptide bonds are formed. RNA is used as a primer template during the initiation of DNA synthesis. The biosynthesis of deoxyribonucleotides involves the enzymatic reduction of ribonucleotides (a point which we will return to later), whereas the latter are synthesised directly by an established biosynthetic pathway. Several plant pathogenic viruses have a genome composed solely of RNA, eg. the tobacco mosaic virus.[34]

The inherent template properties of RNA greatly facilitate a potential mechanism for replication. The formation of specific base pairs allows total sequence control in a nascent strand of RNA which will itself produce a direct copy of the parent strand during the next round of replication. Orgel[35] and Ferris[36] have investigated the template properties of RNA and have shown that certain sequences of RNA can direct the synthesis of a complementary strand without the aid of enzymes. This crucial work will be reviewed in Section 17.

Recent findings in the area of catalytic RNA have added great support to the RNA world theory. Ribozymes were discovered simultaneously by Altman[37] and Cech[38] and have since then been explored by many researchers. Most, but increasingly not all natural ribozymes are limited to sequence specific phosphodiester bond cleavage or ligation. An example is the self-splicing activity of *Tetrahymena*. In eukaryotic genes the functional portion of the gene is initially formed in sections (the exons) separated by stretches of non-functional sequence (the introns or intervening sequences (IVS)). The introns of the gene must be removed by cleavage/ligation reactions or 'splicing' before the gene can function. The *Tetrahymena* RNA undergoes 'self-splicing' in the presence of Mg^{2+} or Mn^{2+} and guanosine, **Fig. 2**. Nucleophilic attack at phosphorus by the 3'-OH group of the external guanosine causes cleavage of the RNA strand. Excision of the intervening sequence is brought about by a further nucleophilic displacement to reform the RNA strand in a shorter but functional form. The excised portion of the RNA genome is subsequently modified by the organism and the resultant, slightly shorter RNA molecule was found by Cech to be a catalyst for many reactions including nucleotidyl-[39] and phosphoryl-transfer,[40] hydrolysis of an aminoacyl ester[41] and even the synthesis of a complementary RNA strand.[42] This latter property is potentially very important in the context of prebiotic replication and was shown by Szostak to be more effective using the shorter *sunY* IVS.[43] Cech showed that the ribose 2'-OH functionality is important for catalytic activity in this ribozyme and proposed that this could be due to several factors. The 2'-OH is both a hydrogen bond donor and acceptor which allows the ribozyme to achieve a degree of tertiary structure, to provide a substrate binding site despite the negatively charged phosphate groups. The 2'-OH itself is involved in RNA hydrolysis by direct attack at phosphorus attached to the 3'-OH to produce the cyclic phosphate. The reactivity of the 3'-OH in

terminal ribonucleosides has enhanced nucleophilicity due to its occasional deprotonation by the neighbouring 2'-OH. This is reflected in the pK_a of ribonucleosides, ~12.5 compared with deoxyribonucleosides, ~16.

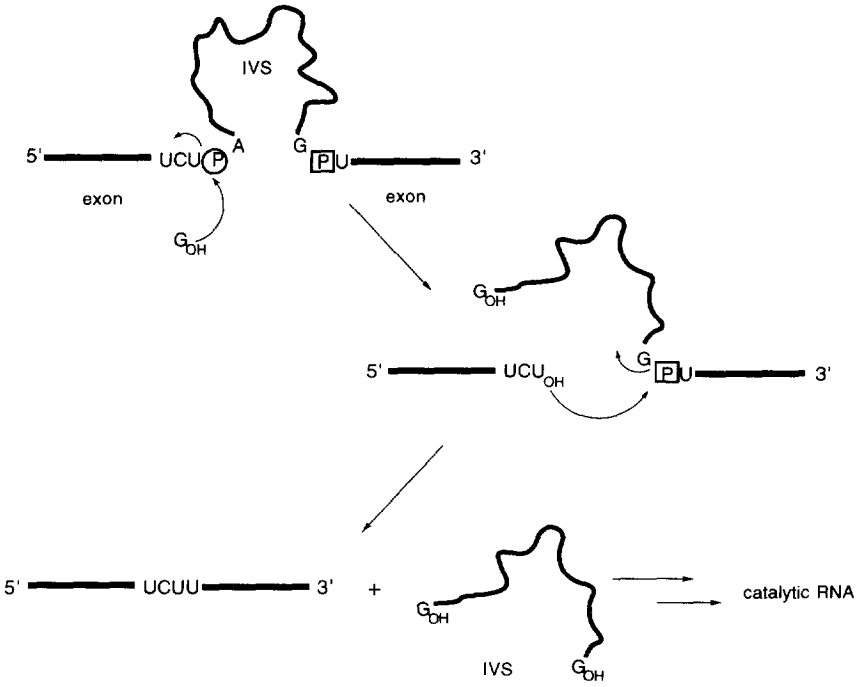


Fig. 2 Mechanism of the self-splicing intervening sequence (IVS) in *Tetrahymena*

Other examples of RNA catalysis include the t-RNA processing enzyme ribonuclease P[37] and perhaps one of the most striking findings is that the ribosome, the site of modern protein synthesis, retains some peptidyl-transfer activity when all the protein it contains is removed.[44] This latter observation suggests a role for RNA in the formation of proteins and adds great support to the theory that RNA preceded proteins in evolution.

12 *In Vitro* Evolution

During any replication process mistakes will inevitably be made by mutational errors or random recombination events. In fact such errors serve an evolutionary purpose in introducing variation into the population. If the phenotype is sequence dependent then Darwinian selection can take place at a molecular level and an environment for 'survival of the fittest' is created. This has been investigated by so-called '*in vitro* evolution' experiments.[45] By first creating a large pool of random RNA molecules, workers have simulated evolution *in vitro* by selecting for molecules with certain catalytic properties, amplifying them using reverse transcription and PCR and then introducing mutations in the new population. Another round of selection identifies molecules with enhanced catalytic properties over the first generation and these are again amplified before being subjected to further cycles of this process. In this way efficient ribozymes have been developed for a wide range of catalytic processes. For example the activity of the *Tetrahymena* IVS discussed above has a strict dependence on Mg^{2+} ions but has been 'evolved' to utilise Ca^{2+} ions which actually inhibit the native ribozyme.

Other experiments have identified new ribozymes which are able to catalyse aminoacylation,[46] hydrolyse peptide bonds,[47] self-alkylate at *N*-7 on a single internal guanosine residue[48] and isomerise a bridged biphenyl ring system.[49] Very recently, Ekland and Bartel isolated an RNA molecule which is able to catalyse the extension of an RNA primer by the addition of nucleoside triphosphates in the same way that polymerase enzymes do today and with very high fidelity.[50] This work provides evidence that primitive RNA may have had the capacity to catalyse its own replication, although the results are somewhat limited in efficiency producing at best an RNA strand of only six nucleotides in length.

Whilst these experiments do not suggest a role for any precise catalytic properties of primitive RNA molecules, they do illustrate the catalytic potential in a random population of replicating RNA molecules and hint at a larger catalytic repertoire which may ultimately include an efficient catalysis of replication and reactions comprising a primitive metabolism. This would indeed increase support for the RNA world theory.

13 Pyranosyl RNA and Hexose Nucleic Acids

During his study of the chemistry of glycoaldehyde phosphate, Eschenmoser was struck by the facile and diastereoselective formation of hexose-2,4,6-triphosphates in the analogous formose reaction (Section 8). He considered the reasons for Nature's selection of pentose sugars as the building blocks for life when hexose sugars would undoubtedly have been present at the same time and suggested that the reason must be one of functionality.[51] In a breathtaking series of experiments provoked by the question 'Why Pentose- and not Hexose-Nucleic Acids?' the ETH group synthesised polynucleotides containing hexopyranose sugars in place of the usual pentofuranose sugars. Initial work focused on 'homo-DNA'[52,53,54] a 2',3'-dideoxy-β-D-glucopyranosyl (6'-4') oligonucleotide. His group undertook conformational analysis, chemical synthesis and characterisation of these compounds with interesting results which suggested a linear polymer similar to A-DNA in structure but with different base pairing properties to DNA, **Table 1**.[55]

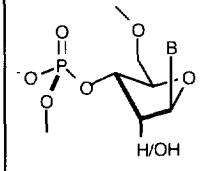
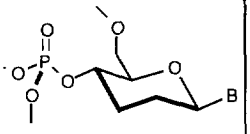
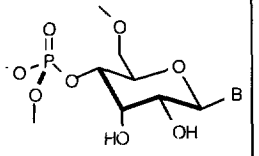
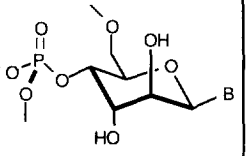
DNA (RNA)	Homo-DNA	Allopyranosyl-NA	Altropyranosyl-NA
			
G > A C > T	G > A ~ G > A C > A G > T	A ~ G > G A G > C	A ..... A
Selective pH-independent semi-independent of sequence	Selective but less so than RNA (DNA) pH-independent semi-independent of sequence	Much weaker than homo-DNA pH-dependent Strongly sequence-dependent	Stronger than allopyranosyl-NA weaker than homo-DNA

Table 1 Base-pairing properties of hexopyranosyl-nucleic acids

It was found that homo-DNA oligonucleotides form antiparallel purine-pyrimidine duplexes which are more stable than the corresponding DNA duplexes by virtue of having a less negative entropy of association. The homo-DNA structure is not *elementary* however and is therefore presumed not to be of prebiotic significance.

When hexopyranosyl-(6'-4')-analogues of RNA derived from fully hydroxylated hexoses were synthesised it was found that base-pairing was greatly inferior to that in RNA. The reason was found to be intrastrand steric hindrance in those conformations capable of base-pairing or to put it in Eschenmoser's own words: "too many atoms". Assuming that predisposed synthetic pathways for these latter hexopyranosyl oligonucleotides existed, Eschenmoser has gone so far as to say that they "stood no chance in the evolution of nucleic acids in competition with RNA, and this for functional reasons."[56] What is of more importance however, is pyranosyl-RNA (p-RNA), a β-D-allopyranosyl (4'-2') oligonucleotide and a structural isomer of RNA, **Fig. 3**.[57]

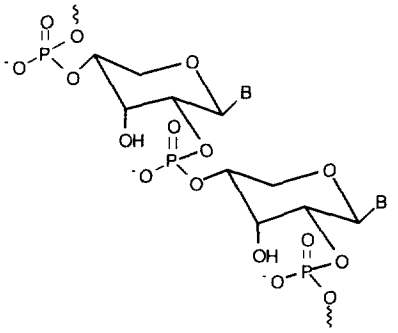


Fig. 3 The structure of pyranosyl RNA

A detailed study of p-RNA, largely by nmr spectroscopy, revealed an almost linear structure showing strict Watson-Crick base pairing properties, due to larger conformational rigidity than in RNA itself. Eschenmoser has suggested that the strong base pairing in the p-RNA duplex gives it a higher potential for constitutional self-assembly than RNA. He also suggests that the notable lack of *intra*-strand base-stacking interactions between individual purine bases in p-RNA would facilitate template directed synthesis when using purine-rich templates, a problem which has arisen in the study of the assembly of RNA.[58] Recently experiments on the capacity of p-RNA to replicate have been reported.[59] C-rich tetramer-2'-phosphates can be ligated by a water-soluble carbodiimide in the presence of G-rich octamer templates. Perhaps more important are experiments in which p-RNA oligonucleotide-2',3'-cyclophosphates are ligated on p-RNA templates, thus, for example a 45% yield of an octamer 4'-GCCCGCCC-2',3'-cp (-2',(3')-p) was obtained on mixing the tetramer 4'-GCCC-2',3'-cp with the template octamer 2'-CGGGCGGG-4'. Hydrolysis has however been shown to compete with ligation. The absence of ligation in the absence of templates strongly suggests that such ligations will be strongly sequence-selective. These experiments bode well for an activated nucleotide oligomerisation route to p-RNA. Assembly of the p-RNA backbone by aldol polymerisation of glyceraldehyde-2-phosphoglycoaldehyde appears unlikely given the tendency of this phosphodiester to cyclise to an, as yet, uncharacterised product.[60]

Eschenmoser also proposes that p-RNA could be a potential precursor to RNA and envisages a transition to RNA *via* an acid catalysed intramolecular rearrangement of p-RNA although the mechanism of this *potential* rearrangement has not been specified.

14 Peptide Nucleic Acid

Peptide nucleic acid (PNA) is a DNA analogue first investigated by Nielsen and co-workers[61] who researched its structure, synthesis and properties in terms of its potential as an antisense agent, **Fig. 4**. They showed that its backbone of repeating achiral *N*-(2-aminomethyl)glycine units forms a double helix as in DNA and that PNA/DNA and PNA/RNA duplexes were held together more tightly than the strands of duplex DNA or RNA themselves.[62, 63] This discovery has led to a suggestion for a role for PNA in the origin of life, its properties illustrating a potential means of transition between different genetic systems.[64]

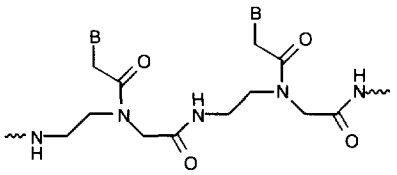


Fig. 4 The structure of peptide nucleic acid

Template directed synthesis using various combinations of PNA, DNA and RNA templates and monomers has met with limited success. The difficulties experienced include short polymer chain lengths and a requirement for specific phosphate leaving groups in the activated nucleotides, problems seen in earlier experiments with RNA polymer synthesis.

15 Iron Pyrites

The most detailed theory for an autotrophic origin of life has been developed by Wächterhäuser[65] and centres around the thermodynamically favourable reaction:

$$FeS + H_2S \rightleftharpoons FeS_2 + H_2$$

Scheme 9

The large decrease in Gibbs function accompanying this reaction (ΔG = -9.23 kcal mol^{-1}) is proposed to serve as the first energy source for life which would, according to this theory have evolved on the surfaces of FeS and FeS_2. Today's enzyme/nucleic acid system is seen as a secondary development brought about by the FeS/H_2S mediated reduction firstly of carbon containing gases, presumably CO_2 to form organic compounds and their subsequent conversion to biologically useful molecules. Experiments to date have not demonstrated the first and potentially most vital of these processes but several reductions of organic molecules have been achieved.[66, 67]

16 Clay Crystals

Cairns-Smith has proposed a theory which cites clay minerals as containing prebiotic genetic information in the pattern of ions in the clay crystal lattice.[68] He suggests that any imperfections in a growing crystal lattice will necessarily be replicated as the crystal grows and describes this process as 'rudimentary biological evolution'. The clay mineral is seen as a surface on which a primitive metabolism emerged and a subsequent process termed 'genetic take-over' derived the present biological system. The mechanism of this take-over has not been specified.

17 Prebiogenesis of the Natural Nucleic Acids

In this section we survey previous approaches to potential prebiotic syntheses of the natural nucleic acids, RNA and DNA. This is a challenge so far unconquered but one that is central to a substantiation of any theory concerning the chemical origins of life.

Previous attempts to demonstrate a potentially prebiotic synthesis of RNA have centered on an apparently obvious disconnection of the polymer structure, perhaps in part dictated by the current mode of enzyme catalysed nucleic acid biosynthesis, **Fig. 5**. In this model the monomers for RNA synthesis are considered to be activated nucleotides **16** (X = leaving group). Despite many decades of research, these ideas have failed to come to fruition and some of those working in the field have begun to doubt the viability of an RNA world.

The first problems pertain to the origin of the activated monomers **16**. Again the apparently most obvious disconnection has been considered and the monomers have been assumed to derive from ribose, a heterocyclic base, phosphate and an activating group. The current theory of prebiotic ribose formation invokes the formose reaction but as discussed earlier, this cascade of aldol reactions produces a vast array of sugars which are likely to have very similar physical and chemical properties. The yield of racemic-ribose in this mixture is estimated at <1%.[69]

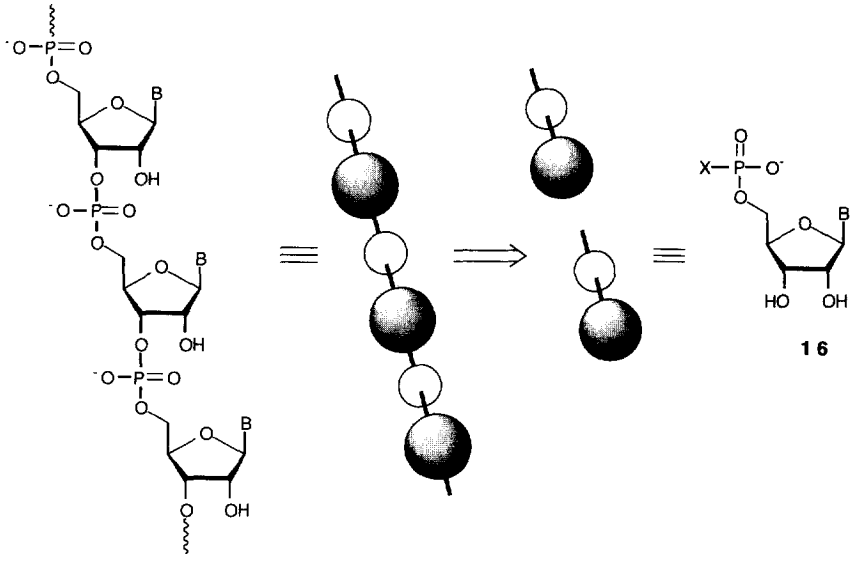


Fig. 5 Schematic representation of RNA disconnection to activated nucleotide precursors

The presence of other sugars would have served only to inhibit RNA synthesis without an efficient method of selection for ribose and more specifically D-ribose, the sugar component in homochiral RNA. Suggestions have been made of a role for minerals in providing a surface for adsorption, with binding facilitated by the 2,3-diol unit of ribose.[70] It does however seem rather difficult to imagine that the required degree of selectivity could be achieved.

The stability of sugars on a geological time scale is another point of concern. Miller has recently gone so far as to state that sugars could not have been a part of the backbones of the first genetic material after he discovered the short half life for the decomposition of ribose at pH 7.0 (73 min at 100°C and 44 years at 0°C).[71] Ribose is also known to isomerise rapidly to give a mixture of arabinose (75%), ribulose (6%) and ribose (19%) in just a few weeks at 25°C, pH 7.0.[72] The work of Eschenmoser on glycoaldehydephosphate **11** and its conversion to

significantly more stable ribose-2,4-diphosphate in good yield has been encouraging, but a facile rearrangement to ribose 3,5-diphosphate has yet to be demonstrated.

Next we must consider prebiotic synthesis of the RNA bases. As mentioned previously this is relatively facile for purines but the pyrimidine experiments are less convincing, relying on high concentrations of reactants which themselves seem susceptible to hydrolysis. The products are also hydrolysed, cytosine being ultimately converted to uracil.

In attempting to join the base and sugar components difficulties are again encountered. As pointed out by Shapiro[69] the conditions required for prebiotic base synthesis are not compatible with those for sugar synthesis. Amines, amino acids and ureas may all react with formaldehyde and the intermediates in the formose reaction. Miller suggests that the two components must have been made in different places[73] but the instability of the sugars and the lack of explanation of how they were brought into the vicinity of the bases makes this sound unfeasible (but the separation might have been temporal, *vide supra*). Experiments to demonstrate condensation of base and sugar have been carried out with mixed success. When ribose and purines are heated in the presence of inorganic salts a mixture of α- and β-nucleosides is formed in 2-10% yield.[74] Analogous reactions with pyrimidines however give < 0.1% yield of products.

Once the nucleoside is formed it must, according to this line of reasoning, be phosphorylated. Experiments have shown that inorganic polyphosphate can be produced under potentially prebiotic conditions by heating aqueous suspensions of hydroxylapatite, $Ca_{10}(PO_4)_6(OH)_2$ at 100°C for 24 hours.[75] When nucleosides are added to this reaction mixture, 2'-, 3'- and 5'- monophosphates, 2',5'- and 3',5'-diphosphates and 2',3'-cyclophosphates are formed in an overall yield of 25%.[76]

The final step invoked in this monomer synthesis is activation of the phosphate group which is a requirement for the template directed reactions. It was found that 5' activation was the most useful since the other isomeric activated phosphates were found to undergo facile cyclisation. In carrying out the activation however another selectivity problem is encountered.

Using a synthetic oligonucleotide template, Orgel[77,78] was able to demonstrate the first example of the use of activated nucleotides in the directed synthesis of a complementary oligomeric chain. The polymerisation is not without its limitations however and these supplement the evidence against the prebiotic formation of RNA in this way. The nature of the phosphate activating group has been shown to be of marked importance in the outcome of such polymerisations[79] with the 5'-phospho-2-methylimidazolides proving the most successful, **Fig. 6**.

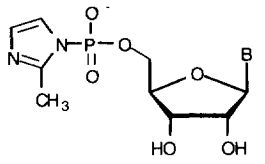


Fig. 6 Structure of 5'-phospho-2-methylimidazolides used in prebiotic RNA synthesis experiments

The chain lengths in the resultant oligonucleotides from Orgel's experiments have not exceeded 20-mers whereas a viable genetic system would require 30-60-mers at least.[80] Competing formation of 2'-5' linkages[81] appears to account in part for the low yields of higher oligomers but a major limiting factor is thought to be the phenomenon of 'enantiomeric cross inhibition.' [70] It transpires that the D-anti and L-syn forms of the monomer form close structural homologues when they become bound to a complementary template, **Fig. 7**.

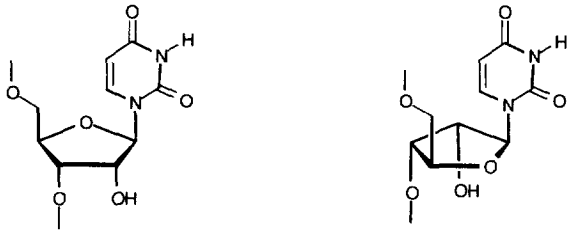


Fig. 7 Structural homology between ribose mononucleotides bound to a complementary template

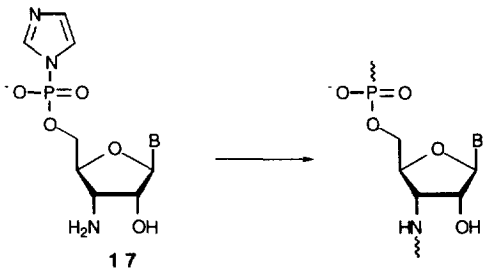
In a growing homochiral chain, incorporation of the 'wrong' enantiomer is found to cause inhibition of polymerisation and chain termination occurs. This is because the 'wrong' enantiomer causes a slight distortion in the structure of the polymer structure when it adds to the growing 3' end of the chain, sufficient to prevent attachment of subsequent residues by distorting base-base interactions. The same phenomenon is seen with α-D-syn and β-D-anti forms of the monomer and given that the mixture of monomers afforded by the proposed prebiotic synthesis would include α-D-syn, α-L-anti, β-D-anti and β-D-syn stereochemistries, polymerisation would appear to be an extremely unlikely event.

Ferris suggested that the process was open to catalysis and showed that the formation of phosphodiester bonds could be facilitated using a derivative of DAMN **1** as a condensing agent.[82] Ferris has since shown that the process of template directed synthesis can be catalysed by minerals, specifically montmorillonite. Most recently in collaboration with Orgel he has created up to 55-mers by adopting a 'feeding' strategy.[83] A template absorbed onto Na^+ montmorillonite was repeatedly incubated with low concentrations of activated nucleotides. The process was described as being akin to primitive Earth rocks being washed repeatedly with the monomers in solution. The ready hydrolysis of these activated monomers in water casts doubt on the feasibility of such a hypothesis.

A final point to consider is that, if this mechanism is to serve as a model for prebiotic replication, the newly formed oligonucleotide must be shown to dissociate from its original template and become a template itself for the next round of replication. Synthesis of its complementary strand will then provide an exact replica of the original oligonucleotide. This second generation template directed synthesis has yet to be demonstrated.

18 Alternative Backbone Nucleic Acids

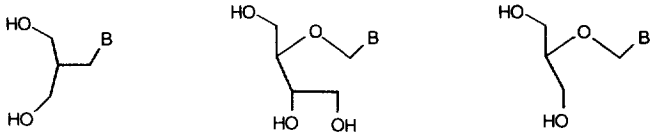
The evidence against the involvement of activated nucleotides in the prebiotic synthesis of RNA has led to suggestions that the first genetic material was a nucleic acid with a modified backbone. Such materials are envisaged as having a transitional role leading ultimately to the formation of RNA, which must be seen to confer an obvious advantage to an evolving population, such as novel chemical properties. An example of such a system acting as a template was reported by Orgel[84] who used 3'-amino-nucleoside-5'-phosphoimidazolides **17** and carbodiimide condensing agents to produce RNA-like 3',5'-linked phosphoramidites, **Scheme 10**.



17

Scheme 10

Many acyclic monomers have been suggested, **Fig. 8**.[85, 86]



18

Fig. 8 Suggested acyclic monomers for prebiotic nucleic acid synthesis

18 is proposed to have its prebiotic origin in glycerol which is more stable than ribose and formaldehyde. Due to an *apparent* simplification of stereochemistry, hypothetical condensation with the bases produces a relatively small number of compounds which in turn demands less of the prebiotic environment during polymerisation. In the work of Joyce *et al.* monomers were chosen as 2',5'-diphosphates to avoid side products due to cyclisation and because pyrophosphate linkages are easier to form than phosphomonoester linkages in this context.[85] Problems of enantiomeric cross inhibition cannot occur with an achiral monomer but once the monomer is incorporated in the polymer it becomes chiral. The L-syn and D-anti forms in this series are found to have an even greater structural homology but the added flexibility in the chain prevents distortion and no inhibition is expected.[87]

In the light of all the evidence against the activated nucleotide disconnection of the RNA chain most workers in the field agree that this is not the way in which the first RNA molecules came into being. Indeed many have now resigned themselves to the notion that RNA played no role in the origin of life despite the extensive, albeit circumstantial evidence cited earlier. It appeared, however, that there was a disconnection of RNA which had not been considered and which seemed worthy of investigation.

19 An Alternative RNA Disconnection

In Oxford we viewed the foregoing analysis of the prebiogenesis of RNA in a different way. Rather than assuming that the evidence pointed to an early alternative to RNA we reasoned that the retrosynthetic analysis was inappropriate and that a radically different disconnection of RNA was demanded. A carbon-carbon bond disconnection was proposed involving formation of the ribose moiety rather than the phosphodiester linkage as the crucial step, **Fig. 9**.[88] The proposal reduces the RNA polymer to the achiral monomer **19**.

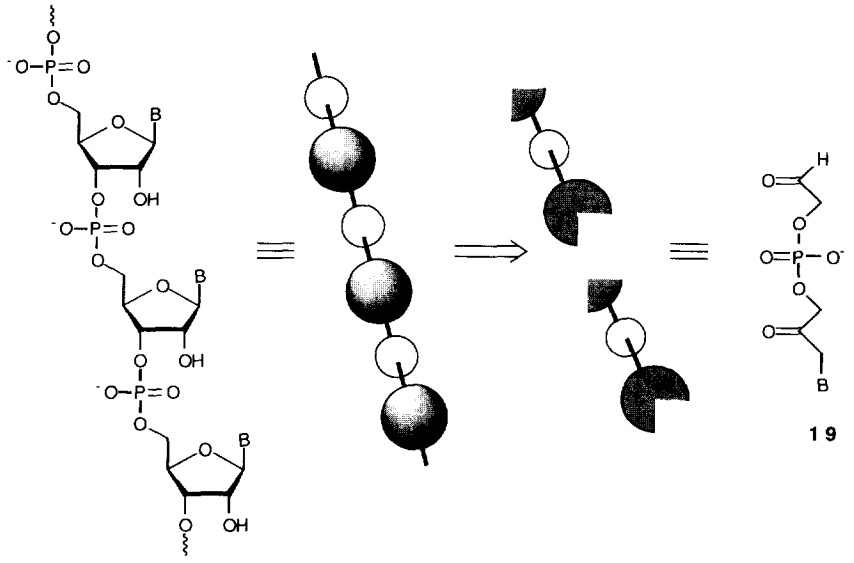
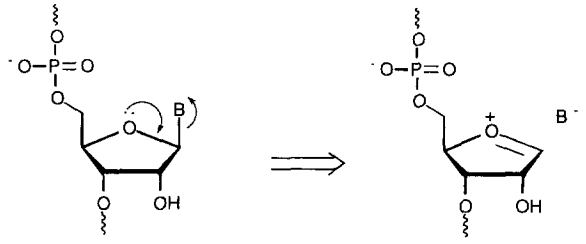


Fig. 9 Schematic representation of an alternative RNA disconnection proposed in Oxford

Many of the difficulties in previous approaches originate from the inherent chirality in the chosen activated nucleotide monomer **16**. It seemed sensible therefore to suggest an achiral precursor to RNA and to require that the origin of homochirality in RNA (and in all ensuing biology) is intrinsic to the polymerising system itself. If the stereochemistry of a newly formed ribose centre is able to influence that formed in the next residue in the chain, then individually homochiral strands of RNA could result. This is expected to be manifested through minimisation of unfavourable interactions in the transition states involved in the reaction. Other workers have shown that the polymerisation of methylmethacrylate can be contrived to produce a quasi-racemic mixture of individually homochiral chains using a chiral template.[89]

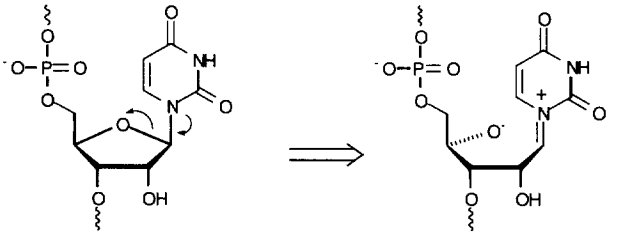
The suggested glycosyl disconnection requires intramolecular redox transfer in order to convert C1' into a reducible centre and two possible mechanisms by which this could occur were considered.

(i) *Oxonium ion disconnection,* **Scheme 11** - This approach would require Lewis acid catalysis which would be incompatible with the presence of the bases in the required deprotonated form and hence this option was disfavoured. In addition it was felt that some physicochemical properties of the bases were likely to be important in the assembly of the monomers for polymerisation and that the base should accordingly be introduced at an earlier stage in the sequence.



Scheme 11

(ii) *Iminium ion disconnection,* **Scheme 12** - This approach invokes the involvement of the lone pair of the component base.



Scheme 12

The low availability of the nitrogen lone pair is acknowledged but it is known that *N*-1-alkyl-pyrimidines can undergo electrophilic substitution reactions at C-5[90] for which the *N*-1 lone pair must be involved. It was noted that the iminium ion could also be stabilised if the base was deprotonated to produce a mesomeric heterocyclic betaine. The deprotonation of guanine and uracil is possible at pH values which do not result in RNA hydrolysis and adenine and cytosine could then be deprotonated *via* base pairing between minor tautomers[91] as illustrated for adenine in **Fig. 10**.

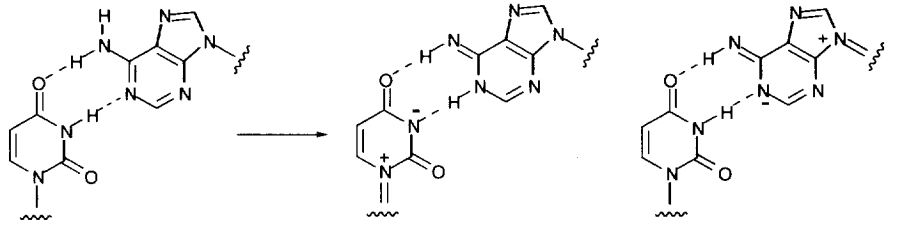


Fig. 10 Deprotonation of adenine *via* base pairing of a minor tautomer to uracil

Such zwitterionic compounds are known, for example the natural product herbipolin,[92] **Fig. 11** which has recently been shown to form self-complementary base pairs.[93]

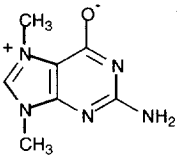
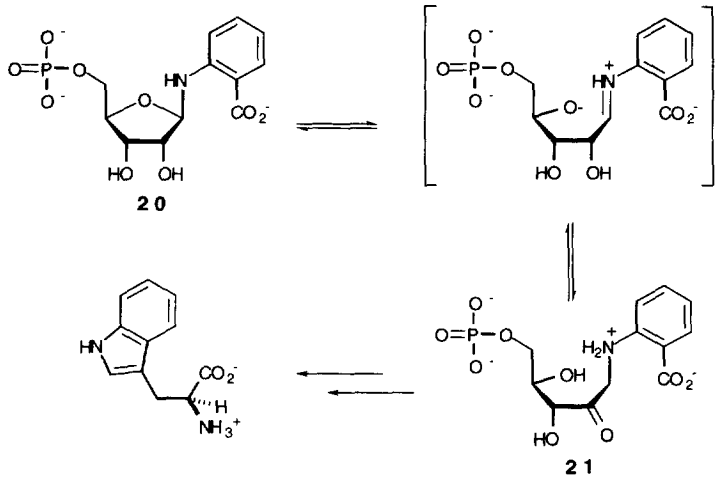


Fig. 11 Structure of naturally occurring mesomeric betaine, herbipolin

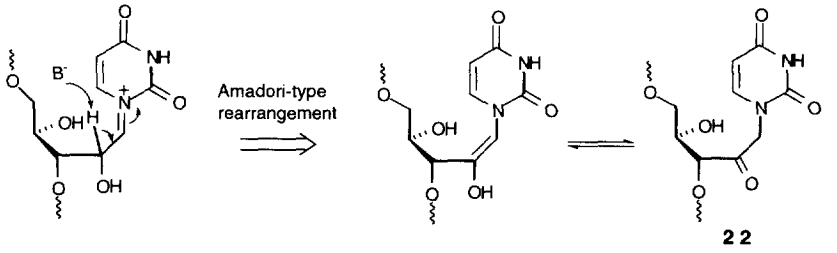
The intramolecular C-1' to C-2' redox transfer necessary for the aldol disconnection to be applied, could conceivably occur by two mechanisms.

(i) *Retro-Amadori rearrangement* - The Amadori rearrangement[94] involves azaenolisation followed by ketonisation as seen in the contemporary biosynthesis of tryptophan where it is mediated by the enzyme phosphoribosylanthranilate isomerase, **Scheme 13**.[95, 96]



Scheme 13

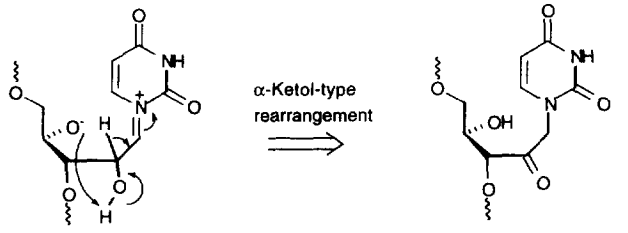
The proposed *retro*-Amadori rearrangement in our system is illustrated in **Scheme 14** and the similarity of the two cases is apparent.



Scheme 14

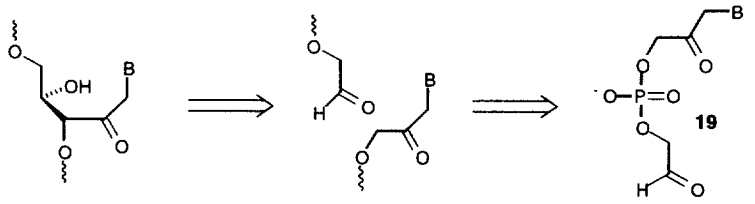
The major difference is the substitution of the anthranilate secondary amino group in **21** for the aromatic nitrogen in **22** and this is likely to be very important in determining the position of equilibrium in these systems. In the former, *N*-protonation of the product is influential in displacing the equilibrium to the right since in the starting material **20** a stereoelectronic n-$\sigma^*_{C\text{-}O}$ interaction disfavours protonation. The equilibrium in our system is difficult to predict but should not be affected to the same extent.

(ii) *1,2-Hydride shift* - This is suggested as similar to a base catalysed α-ketol rearrangement and in this case is assisted (in a retrosynthetic sense) by intramolecular general base catalysis from the C-4' alkoxide, **Scheme 15**.



Scheme 15

From here the retrosynthetic approach requires one final disconnection which involves (in the forwards sense) an aldol condensation. Iteration of the overall disconnection reduces RNA to the monomers **19**, **Scheme 16**.



Scheme 16

Support for the involvement of aldol condensations in such a context was found in the work of Sugiyama *et al.* [97] These authors reported that photochemical degradation of DNA by radical oxidation at C2' resulted in an erythrose-containing component which was subject to alkaline dependent strand scission by a retro-aldol mechanism.

A number of model compounds (**Fig. 12**) have been studied in this laboratory in order to enable prediction of the behaviour of the monomer **19** and to investigate our proposal for a potentially prebiotic RNA synthesis.

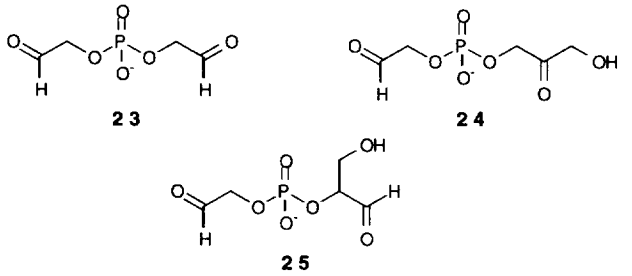
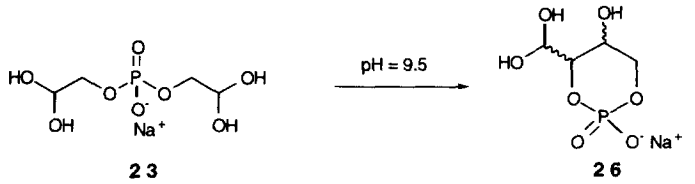


Fig. 12 Structures of model compounds studied in Oxford

Synthesis and analysis of the solution behaviour of compounds **23**, **24** and **25**[98, 99] showed that **23** and **25** exist as fully hydrated species whereas **24** exists as a 1:1 mixture of ketone and hydrated ketone with the aldehydic portion of **24** being fully hydrated in both species. This observation is of great importance when considering the propensity for enolisation of the two carbonyl groups of **24**. Kinetic control in this process is assumed, based on the evidence of Eschenmoser's group for the reaction of glycoaldehyde phosphate and formaldehyde.[28] The rate of enolisation is then dependent on two factors, given by rate = k[C=O]. It was found that hydration slows the rate of enolisation since the effective concentration of C=O is reduced. Exchange experiments showed that after 1.5 h at pD 9.5 in D_2O the ketone moiety of **24** was predominantly monodeuterated, supporting the ability of the ketone to enolise and demonstrating that it does so faster than the aldehyde moiety. This supports our proposal which requires the phosphoketonic enolate of **19** to be the preferred aldol *donor* during polymerisation. Of the two carbonyl functionalities the aldehyde is expected to be the preferred aldol *acceptor* during subsequent addition since, although similar factors apply, the rate constant for attack at ketone is expected to be much less than that for attack at aldehyde due to steric hindrance. This hydration control thus potentially provides us with the desired 3'-5'-linked polymer.

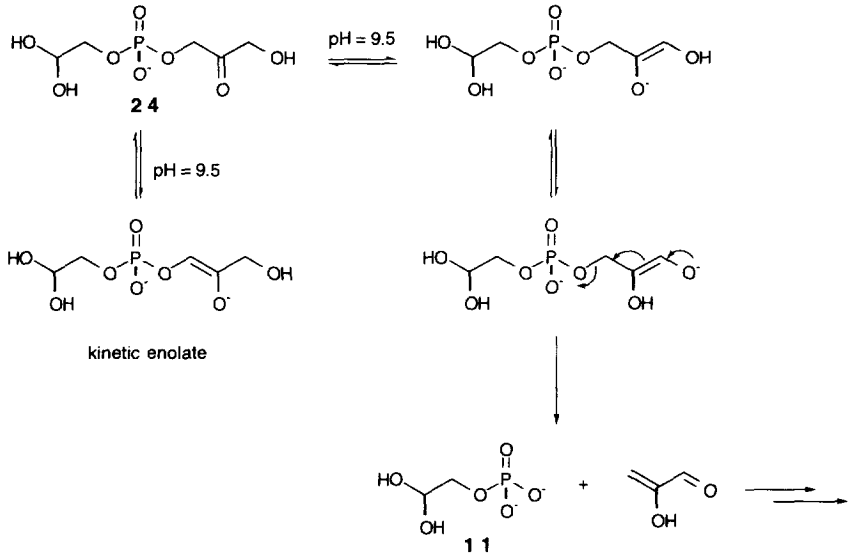
A study of the aldolisation behaviour of these model compounds has revealed some encouraging observations. When subjected to alkaline conditions, specifically pH 9.5, **23** was found to cyclise to **26**, **Scheme 17**.[100]



Scheme 17

This confirms that the aldehydic portion of **23** can act as an aldol acceptor but experiments in D_2O reveal that enolate protonation/deuteration is competitive with cyclohydroxyalkylation. No evidence for polymerisation was seen by nmr or electrospray mass spectrometry.

The behaviour of **24** over extended periods at this pH is, however very different. Cyclisation does not occur, as evidenced by retention of the characteristic glycoaldehyde signals in ^1H nmr spectra. Instead it was found that an elimination mechanism was operating, **Scheme 18**.



Scheme 18

These results suggest that the added steric bulk over *bis*-glycoaldehydephosphate makes cyclisation unfavourable in this instance and at the chosen concentration the observed slow elimination reaction is then able to proceed. In the case of the actual monomer **19** this steric hindrance is accentuated but elimination should not occur in this way due to *restricted* nitrogen lone pair availability.

In light of the foregoing results with model compounds it was decided that the synthesis of the four variants of monomer **19** was justified; this work is described in the following paper in this issue. All four monomers display similar behaviour in neutral aqueous solution being fully hydrated in their aldehydic moieties and approximately 30% hydrated in their ketonic moieties. This hydration behaviour parallels that of the model phosphodiester **24**; on standing in neutral D_2O, the phosphoketonic methylene protons of the monomer are gradually exchanged indicating that formation of the phosphoketonic enolate is kinetically preferred (and very accessible) as was the

case with **24**. In recent preliminary experiments we have shown that at pH 9.5 the adenine and uracil variants of **19** do indeed polymerise to give short oligomers. The exact nature of the oligomeric products remains to be established but certain important conclusions can already be drawn:

i) the polymerisation involves aldol reactions

ii) the monomers appear to cyclise reversibly but not to the extent that polymerisation is precluded (*cf.* the case with **23**)

iii) RNA is not produced; no signals for the glycosyl protons of RNA are observed in ^{1}H nmr experiments

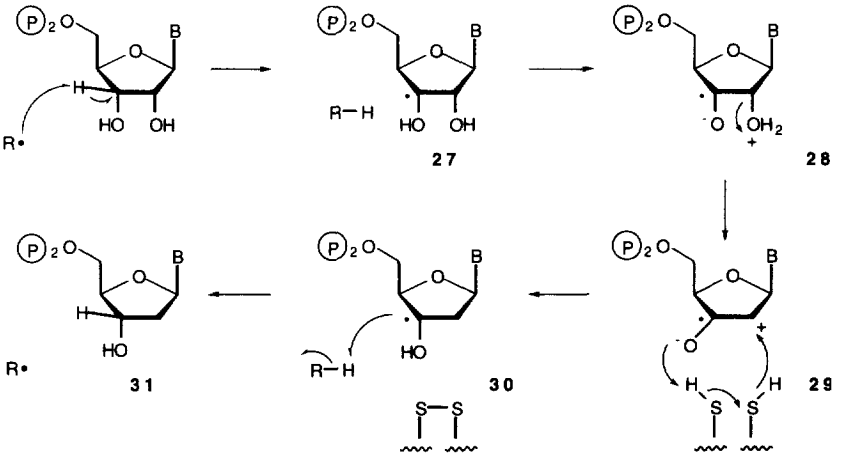
iv) RNA itself is only partially stable to the prolonged incubation conditions at alkaline pH.

These conclusions have fascinating potential ramifications which will be covered after a brief digression on a subject which, in the context of these recent results now has added significance.

20 On the Interconnection between RNA and DNA

Our own involvement in this area was stimulated by an attempt to analyse contemporary biochemical processes according to the manner in which they might have evolved. Specifically we sought to highlight how the two pillars of organic chemistry, structure and mechanism have parallels in biochemical evolutionary theory. In this latter area we may view a *structure* as being an operating sub-system of an overall metabolism while *mechanism* pertains to the development of a new structure or the transition between two structures. Like their organic reaction counterparts, biochemical evolutionary mechanisms have rules. Most importantly, a biochemical evolutionary mechanism must proceed by a number of incremental steps where the change in each step can be associated with only a few mutations at the genotypic level and where an advantage accrues to the organism for each step. We chose to illustrate this concept by considering the current dogma that RNA preceeded DNA in evolution and that DNA only arose after the production of a ribonucleotide reductase system.

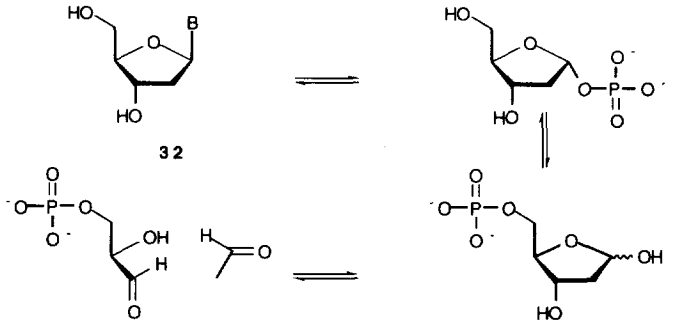
The analysis starts by considering the mechanism of ribonucleotide reductase, **Scheme 19**.



Scheme 19

Radical abstraction of the β-hydrogen from C-3' produces a carbon-centred radical **27** which can participate in a prototropic equilibrium between the 3'-hydroxyl group and the 2'-hydroxyl group. Loss of water from the intermediate **28** produces a resonance stabilised species **29** (formally an enoloxy radical) which is then reduced by a dithiol to give the 2'-deoxy carbon-centred radical **30**. The reaction cycle is completed by hydrogen atom transfer to regenerate the initiating radical and give the 2'-deoxynucleotide **31**. This process intrigued us as it is one of the few central metabolic reactions which is inconceivable in the absence of an enzyme catalyst. In this sense it is different to the vast majority of reactions in biochemistry which, as indicated before, could have occurred before the advent of catalysis. In the case of reactions which proceed to some extent in the absence of a catalyst and give products which have a use, then the evolution of a catalyst can offer immediate advantage in that it allows for more rapid (regulated) synthesis of the product. In some cases the enzymic product will be produced from a different starting material in which case the underlying enzyme chemistry can either be predisposed or not. If, however, a compound is first produced as the result of enzymic catalysis (of non-predisposed chemistry) then the product (or the consequent reduction of starting material) must offer some immediate advantage to the organism. In the case of ribonucleotide reductase two contrasting scenarios can be envisaged and both must be considered as relatively late evolutionary events for an advanced catalytic repertoire must have been available to allow for the evolution of a catalyst for such a demanding reaction. An implicit assumption is that the advantage of a combined DNA and RNA world was the evolutionary driving force in both scenarios. In the first scenario we imagine that deoxyribonucleotides were *previously available* by some *different route* and the advantages of a combined DNA and RNA world had already been experienced. The chance evolution of ribonucleotide reductase could then offer a major advantage to the organism if the *previous route* to deoxynucleotides was the *less efficient.*

The second scenario, which has been assumed by others, is that deoxyribonucleotides were *first produced* by the action of ribonucleotide reductase.[101] In order for advantage to accrue to the organism then DNA polymerase and DNA-dependent RNA polymerase would have had to coevolve along with the ribonucleotide reductase. This simultaneous evolution of an enzyme catalysing difficult chemistry *and* two new polymerases with subtly differing specificities appeared to us to be a contravention of biochemical evolutionary mechanism. We were thus forced to consider the ramifications of the first scenario in more detail. Our thoughts focussed on the possible nature of the previous, different route to deoxyribonucleotides and we realised that an alternate route in contemporary biochemistry exists in the chemistry catalysed by the enzymes encoded by the *deo* operon, **Scheme 20**.[102]



Scheme 20

In the contemporary organism, *Escherichia coli*, the *deo* operon is induced by high levels of deoxyribonucleosides **32** which are consequently degraded by the encoded enzymes to glyceraldehyde-3-phosphate and acetaldehyde. Importantly, however, the equilibrium constants for the three enzyme catalysed processes are close to unity so the system *could* serve as a biosynthetic route to deoxyribonucleosides. The chemical reactions involved in this scheme are undoubtedly less demanding than the reaction catalysed by ribonucleotide reductase. Our tentative conclusion was that this system might have served as the biosynthetic route to deoxyribonucleotides *prior* to the evolution of ribonucleotide reductase. Our initial research goal was to investigate whether the apparently simple chemical interconnection between acetaldehyde, glyceraldehyde-3-phosphate and deoxynucleosides could be realised. For a variety of reasons this research was not pursued but it still remains a worthwhile objective. As far as evolutionary theory is concerned though, after ribonucleotide reductase had become the major route to deoxyribonucleotides, alteration of the regulation of the system would allow its use only in a catabolic sense.

If deoxyribonucleosides were biosynthesised along with their ribonucleoside counterparts by an early organism then, it seems reasonable to suppose that activation to deoxyribonucleoside triphosphates would have occurred since the difference between the two sugars is small as far as recognition by primitive enzymes is concerned. In the same way it seems highly unlikely that a primitive polymerase could have distinguished effectively between ribo- and deoxyribonucleoside triphosphates. The conclusion to our line of reasoning was that there could have been a period when the nucleic acid was neither RNA nor DNA but a chimaeric heteropolymer which we term R/DNA. This heteropolymer could solve many of the difficulties associated with a genetic/metabolic system based on either of the two homopolymers alone. RNA has a strong tendency to adopt secondary folded structures and this is one feature which allows it act catalytically. It is to be expected however that replication of folded structures would be more difficult than replication of linear polymers such as DNA. If a statistically representative population of R/DNA were available then those molecules with the highest ribose content would naturally be the most folded and the best catalysts, molecules of the same sequence but with a higher deoxyribose content would be less structured and most easily replicated. In terms of progression to the RNA and DNA world, we could imagine that genetic duplication of the R/DNA polymerase would allow divergent evolution to produce two new polymerases one more specific for R/DNA and the other for R/DNA. The benefit of this divergence would be the production of a population more biased in favour of the best catalysts (R/DNA) and the most easily replicated species (R/DNA) rather than the statistically favoured compromise molecules (R/DNA). A viable biochemical evolutionary mechanism for the production of separate RNA and DNA polymerases can thus be drawn, **Figure 13**.

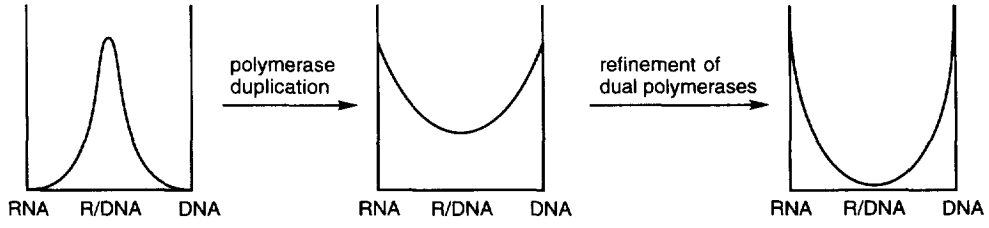
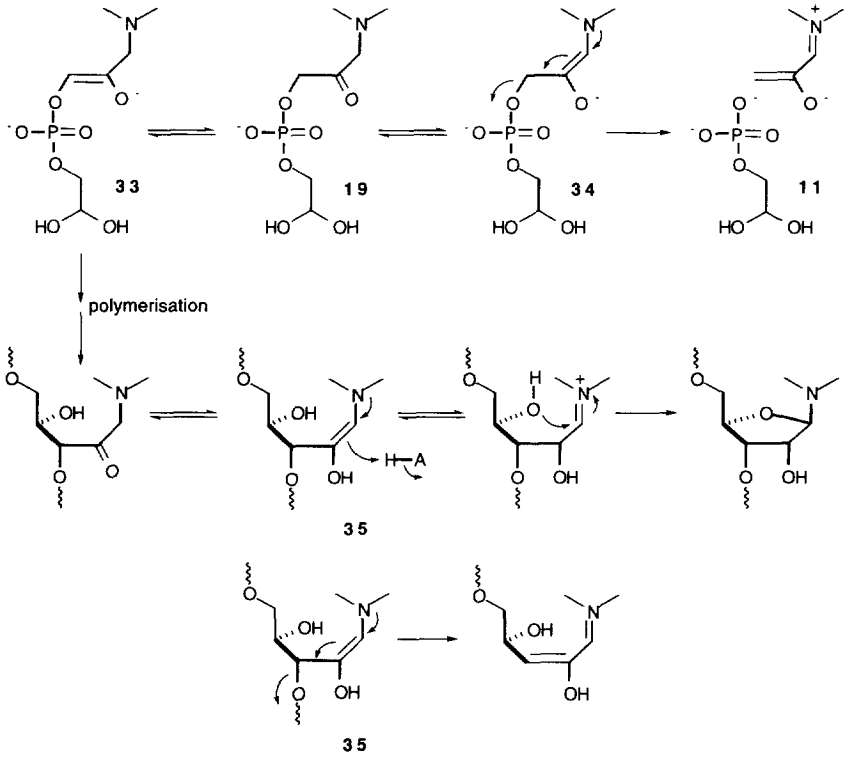


Figure 13

From what has been said in this section it is clear that the search for predisposed chemical routes to polymeric nucleic acid should not necessarily be restricted to RNA, a predisposed route to R/DNA could potentially be even more significant.

21 Refinements to Theory and Future Directions

Results from our work in Oxford suggest that certain key elements of the novel disconnection of RNA described in Section 19 are valid. The major problem with the approach as it stands though is the degree to which the nitrogen lone pair in monomers **19** is available. The weight of evidence suggests that the lone pair is insufficiently available for the proposed redox transfer to take place. Modifications to the monomers that might render the lone pair more available were sought with the *caveat* that over-availability would allow the elimination of glycoaldehydephosphate from kinetically less accessible enolates, **Scheme 21**.



Scheme 21

More specifically the modification of the monomer should:

i) increase the kinetic preference for the enolate **33** to the extent that polymerisation occurs before the alternative enolate **34** is sampled and elimination of glycoaldehydephosphate is possible,

ii) slightly increase the bulk of the system to favour polymerisation over cyclisation further,

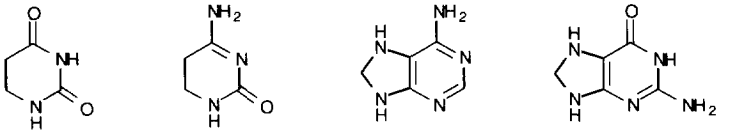
iii) increase the availabilty of the nitrogen lone pair to participate in redox transfer once polymerisation has occurred,

iv) allow the polymer to adopt a conformation in which protonation of enamine **35** is facile but elimination from **35** is precluded (such conformational control is evident when contrasting the enzymes triosephosphate isomerase and methylglyoxal synthase),

v) still allow the production of RNA (or, given what has been said in Section 20, R/DNA).

A consideration of these reasons pointed to the low availability of the adjacent nitrogen lone pair in the component base as the major factor. This realisation prompted the development of a modification to the original proposal, which appears to satisfy some of the more elusive issues in prebiotic chemistry.

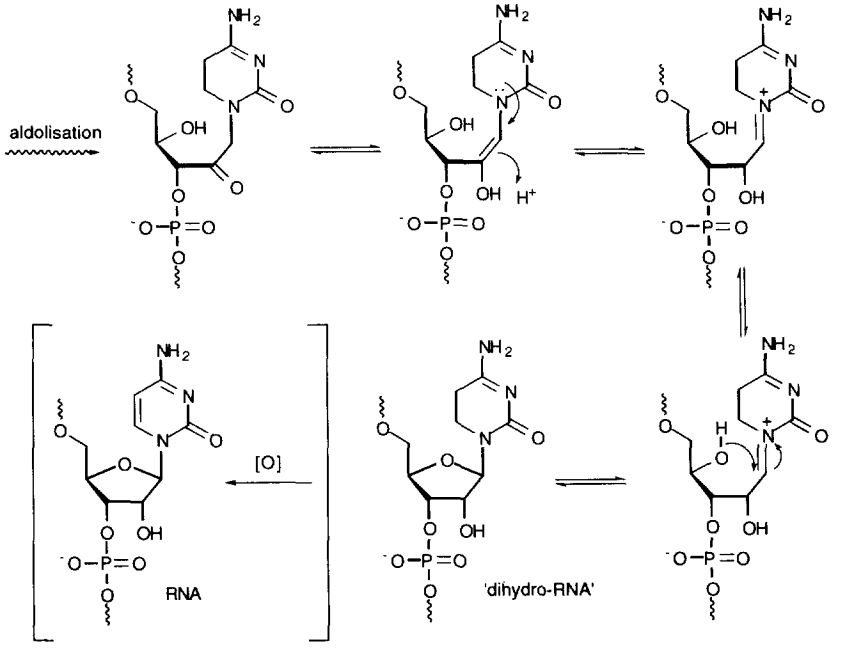
It is now proposed that the original base components of monomer **19** were the dihydro forms of the present-day bases, **36-39**, **Scheme 22** (although pyrimidines reduced at C-2 should also be considered).



36 **37** **38** **39**

Scheme 22

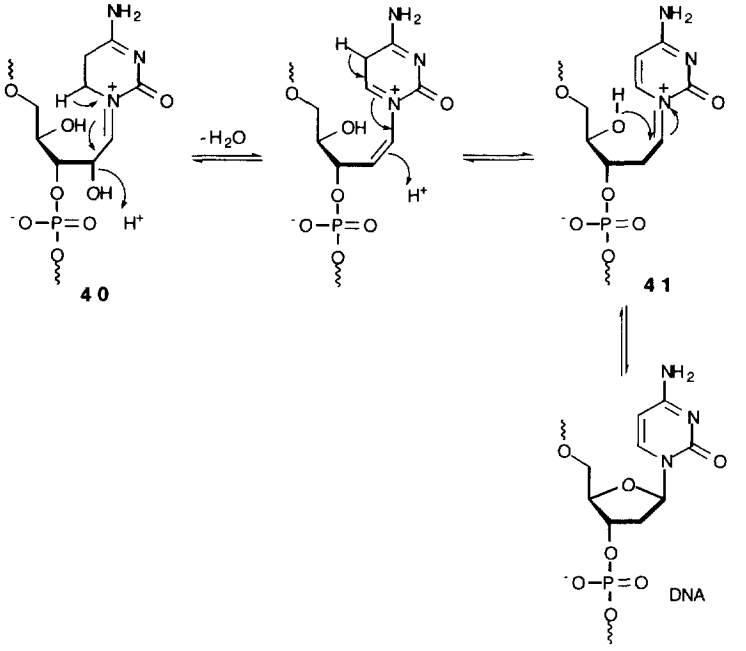
The *N*-1 or *N*-9 lone pair is significantly less delocalised in these heterocycles than in the natural, aromatic bases and is therefore potentially more available for donation to C-1' during the required azaenolisation step, as proposed in the original polymerisation route (Section 19). If the analogous sequence of steps occurs in this system the result would be an RNA-type polymer containing dihydrobases, **Scheme 23**.



Scheme 23

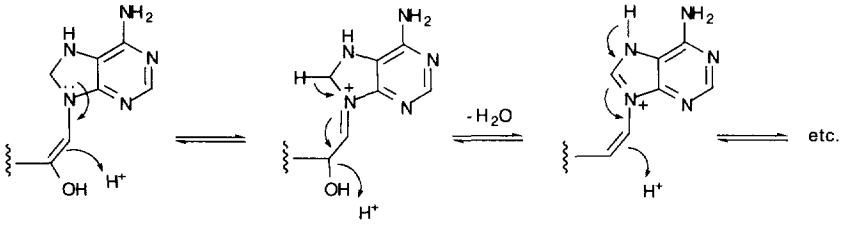
Subsequent oxidation of the dihydrobases to afford natural RNA should be a relatively facile process, particularly for the purines where aromatisation provides a strong thermodynamic driving force.

In these systems an alternative reaction course is also conceivable, **Scheme 24**.



Scheme 24

Dehydration of intermediate **40**, *via* initial loss of the acidic C-6 proton (the conjugate base is a 1,3-dipole) is followed by loss of the now very acidic C-5 proton and subsequent protonation at C-2'. The system in **41** is then set up for ring closure, the result being the formation of *DNA* containing the natural unsaturated bases. The dihydropurines can be envisaged as acting in the same way, **Scheme 25**.

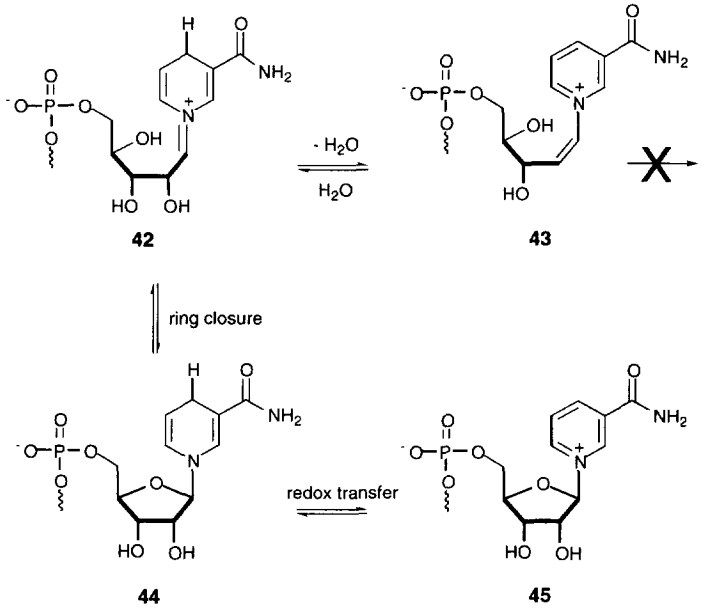


Scheme 25

Aromatisation of the heterocycle (or the extension of conjugation in the pyrimidine series) is the obvious driving force for this sequence of steps. Combination of dihydro-RNA oxidation and dehydration would then produce R/DNA which, as discussed in Section 20, is a potentially more realistic (and useful) primitive nucleic acid.

This proposal represents a novel view of the chemistry associated with the origin of life, and challenges the widely accepted notion of the RNA world, Section 11. On reflection, however, the properties of RNA which have led to this notion, such as its ability to act as a template, its information content and the potential for genetic variation, also apply to DNA. The discovery that RNA molecules are able to effect catalysis has lent great credence to the theory of the RNA world, but work by Szostak has illustrated that DNA can also function as a catalyst.[103] This new theory therefore appears to circumvent the problem of a transition from the proposed RNA world to the interdependent DNA-RNA-protein world of the present day (Section 20), by providing a mechanism whereby all can co-evolve. In this scenario the dominant *genetic* material after time would be expected to be DNA, its formation being driven by heterocyclic aromatisation and its greater stability rendering the formation/prolonged existence of longer oligomers more probable.

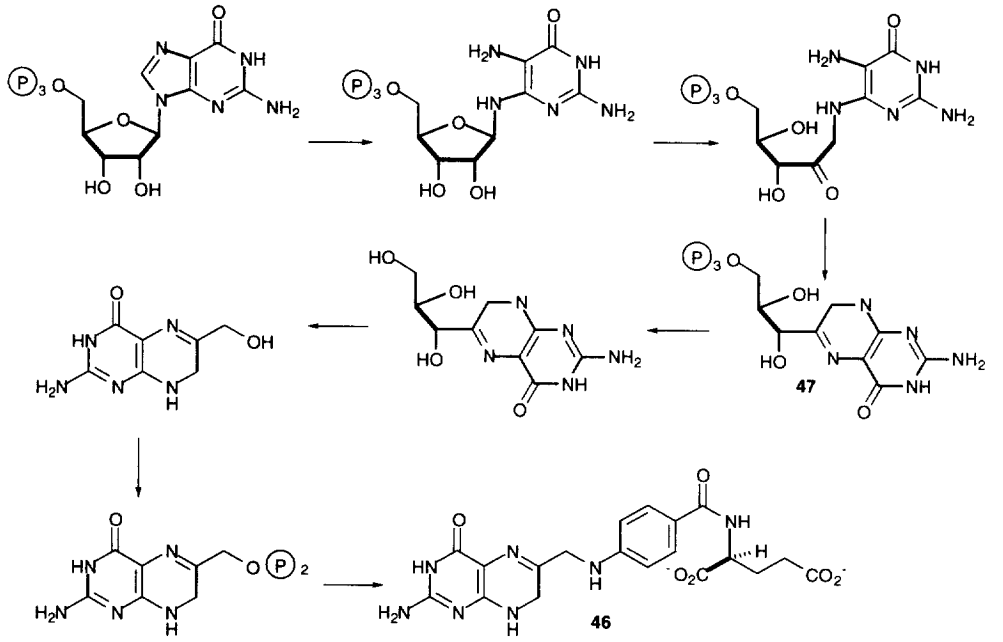
The new theory has other tantalising consequences. The origin of coenzymes has long been a debated issue and many suggest that their ribonucleotide-like structures point to their derivation from RNA and cite this as evidence for the RNA world. In a strikingly plausible manner, however, incorporation of the relevant heterocycles into monomer **19** leads to a mechanism for the synthesis of a large number of these coenzymes. By way of example, consider nicotinamide adenine dinucleotide, **Scheme 26**.



Scheme 26

In this case, loss of the C-4 proton from intermediate **42** affords full aromatisation of the heterocycle in **43** and the system cannot lose a second proton to allow ring closure to a deoxyribonucleotide. The resultant equilibrating mixture is thus likely to ring close directly to ribose, leading to the formation of an NADH-like material **44**. The redox potential of the dihydropyridine ring system of **44** is such that hydride can be (reversibly) transferred giving **45**. It seems logical to propose therefore, that coenzymes containing ribose were perhaps a direct result of the incorporation of heterocycles into monomer **19**, giving compounds which after aldol reaction were unable to dehydrate to the deoxyribose derivatives. Hydrolysis of the 3'-phosphate and modification of the 5'-end of these

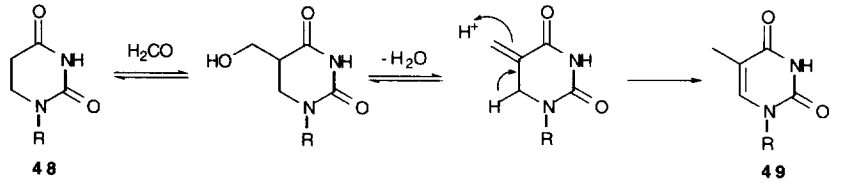
nucleosides could lead to an array of catalysts which were later subject to natural selection. Support for this assertion is found in the contemporary biosynthesis of folic acid **46** *via* the intermediacy of neopterin triphosphate **47**, **Scheme 27**.



Scheme 27

The similarity of this biosynthesis to the novel theory we propose for prebiogenesis is apparent. That the first three steps are effected by the same enzyme, GTP cyclohydrolase suggests that the underlying chemistry is indeed predisposed.

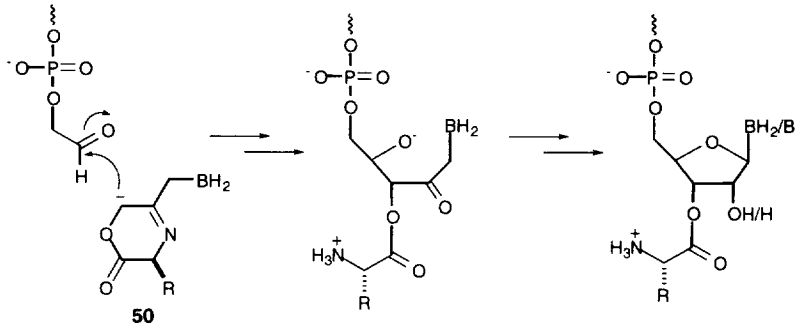
Modern day transfer RNA contains many unusual residues which interestingly in the context of this proposal, include dihydrouridine **48** and ribothymidine **49**, the former would be a direct product of polymerisation according to our modified theory while the latter could easily be derived by subsequent reaction with formaldehyde **Scheme 28**.



Scheme 28

A final point pertains to the evolution of proteins and the genetic code. The progressive evolution of the genetic code requires a set of several aminoacyl t-RNAs (tR/DNAs?) which can assemble proteins having a variety of catalytic properties. If these proteins can confer an evolutionary advantage over their catalytic predecessors this will then ensure that their encoding genes are selected in an evolutionary sense. It is difficult to see how this set of

aminoacyl t-RNAs could be acquired without some selective advantage for each of them, before their eventual combined function became apparent. The question arises, on what basis could the intermediates have been selected? This dilemma is avoided if one assumes that aminoacyl t-RNAs were necessarily formed prebiotically as a by-product of some other process. The suggestion is now, that this novel theory allows for this to occur. The dihydrooxazine-2-one **50** is constitutionally related to the monomer **19** if glycoaldehydephosphate is replaced by an amino acid. Participation of **50** during the polymerisation of monomer **19** can be envisaged and the formation of aminoacyl t-R/DNA is theoretically possible, **Scheme 29**.



50

Scheme 29

Hydroxyalkylation of the dihydrooxazine-2-one is expected to occur anti-facially to the R group, this provides a potential correlation between the relative stereochemistries of amino acids and sugars now found in biochemistry. We have recently synthesised two variants of **50** in which the base is aromatic (R = ^{i}Pr, BH_2 = A,U)[104] and have almost completed the synthesis of a dihydro-variant (R = ^{i}Pr, BH_2 = *2H*-U).

In summary we propose that an environment containing the monomers **50** and **19** (B = *2H*-A, C, G, U and other heterocycles) has the potential to produce, simultaneously, R/DNA, cofactors, aminoacyl-R/DNAs (and proteins). This proposal differs markedly from the 'RNA-world' hypothesis and invokes a high degree of (theoretically) predisposed chemistry. We air these conjectures as they clearly indicate the need for a great deal of investigative, experimental chemistry; given the potential results, however, we are of the firm opinion that this experimental work is justified.

22 Acknowledgements

Research in this laboratory has been supported by the EPSRC and BBSRC. We gratefully acknowledge helpful discussions with Dr. George Weaver, Dr. Anthony Lawrence and Steve Cook of this department and Dr. Martin Wright (Scripps Research Institute, USA). Professor A. Eschenmoser and Dr. S. Pitsch (ETH, Zürich) have consistently and generously provided wise advice to the authors who thank them for this with respect.

23 References

1 L. E. Orgel, *Scientific American,* 1994, **10**, 53.
2 W. G. Ernst, *Earth's Earliest Biosphere (ed J. W. Schopf)*; Princeton University Press: 1983, pp 41-52.
3 R. Buick, J. S. R. Dunlop and D. I. Groves, *Alcheringa,* 1981, **5**, 161.
4 J. M. Hayes *Earth's Earliest Biosphere (ed J. W. Schopf)*; Princeton University Press: 1983, pp 291-301.
5 A. I. Oparin, *The Origin of Life*; Macmillan: New York, 1938.
6 S. L. Miller, *Cold Spring Harbour Symp. Quant. Biol.,* 1987, **52**, 17.
7 A. Eschenmoser and E. Loewenthal, *Chem. Soc. Rev.,* 1992, 1.
8 J. R. Cronin, *Advances in Space Research,* 1989, **9**, 54.
9 J. P. Ferris and J. W. J. Hagan, *Tetrahedron,* 1984, **40**, 1093.
10 P. H. Abelson, *Proc. Natl. Acad. Sci., USA,* 1966, **55**, 1365.
11 J. P. Ferris and C. T. Chen, *Nature,* 1975, **258**, 587.
12 R. A. Sanchez, J. P. Ferris and L. E. Orgel, *Science,* 1966, **154**, 784.
13 R. Stribling and S. L. Miller, *Origins of Life,* 1987, **17**, 261.
14 S. L. Miller, *Science,* 1953, **117**, 528.
15 A. Strecker, *Liebigs Ann. Chem.,* 1850, **75**, 27.
16 S. L. Miller, *J. Am. Chem. Soc.,* 1955, **77**, 2351.
17 J. Oró, *Biochem. Biophys. Res. Commun.,* 1960, **2**, 407.
18 C. U. Lowe, M. W. Rees and R. Markham, *Nature,* 1963, **199**, 219.
19 R. A. Sanchez, J. P. Ferris and L. E. Orgel, *J. Mol. Biol.,* 1967, **30**, 223.
20 J. P. Ferris and L. E. Orgel, *J. Am. Chem. Soc.,* 1966, **88**, 1074.
21 Y-B. Xiang, S. Drenkard, K. Baumann, D. Hickey and A. Eschenmoser, *Helv. Chim. Acta,* 1994, **77**, 2209.
22 R. Shapiro and R. S. Klein, *Biochemistry,* 1966, **5**, 2358.
23 M. P. Robertson and S. L. Miller, *Nature,* 1995, **375**, 772.
24 F. Wöhler, *Poggendorfs Ann. Phys. Chem.,* 1828, **12**, 253.
25 J. P. Ferris, O. S. Zamek, A. M. Altbuch and H. J. Freiman, *J. Mol. Evol.,* 1974, **3**, 301.
26 A. Butlerow, *Liebigs Ann. Chem.,* 1861, **120**, 295.
27 S. Drenkard, J. Ferris and A. Eschenmoser, *Helv. Chim. Acta,* 1990, **73**, 1373.
28 D. Müller, S. Pitsch, A. Kittaka, E. Wagner, C. E. Wintner and A. Eschenmoser, *Helv. Chem. Acta,* 1990, **73**, 1410.
29 A. Eschenmoser, *Angew. Chem. Intl. Ed. Engl.,* 1988, **27**, 5.
30 U. Trinks, Thesis No. 8368, ETH Zürich: 1987.
31 T. Paterson and H. C. S. Wood, *J. Chem. Soc., Perkin 1,* 1972, 1041.
32 J. F. Kasting, *Science,* 1993, **259**, 920.
33 W. Gilbert, *Nature,* 1986, **319**, 618.
34 R. H. Symons, *Annu. Rev. Biochem.,* 1992, **61**, 641.
35 T. Inoue, G. F. Joyce, K. Grzeskowiak and L. E. Orgel, *J. Mol. Biol.,* 1984, **178**, 669.
36 J. P. Ferris and G. Ertem, *Science,* 1992, **257**, 1387.
37 C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace and S. Altman, *Cell,* 1983, **35**, 849.

38 K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling and T. R. Cech, *Cell,* 1982, **31**, 147.

39 A. J. Zaug and T. R. Cech, *Science,* 1986, **231**, 470.

40 A. J. Zaug, M. D. Been and T. R. Cech, *Nature,* 1986, **324**, 429.

41 J. A. Piccirilli, T. S. McConnell, A. J. Zaug, H. F. Noller and T. R. Cech, *Science,* 1992, **256**, 1420.

42 J. A. Doudna and J. W. Szostak, *Nature,* 1989, **339**, 519.

43 J. A. Doudna, S. Couture and J. W. Szostak, *Science,* 1991, **251**, 1065.

44 H. F. Noller, V. Hoffarth and L. Zimniak, *Science,* 1992, **256**, 1416.

45 G. F. Joyce, *Curr. Opin. Struct. Biol.,* 1994, **4**, 331.

46 M. Illangesekare, G. Sanchez, T. Nickles and M. Yarus, *Science,* 1995, **267**, 643.

47 X. Dai, A. D. Mesmaeker and G. F. Joyce, *Science,* 1995, **267**, 237.

48 C. Wilson and J. W. Szostak, *Nature,* 1995, **374**, 777.

49 J. R. Prudent, T. Uno and P. G. Schultz, *Science,* 1994, **264**, 1924.

50 E. H. Ekland, J. W. Szostak and D. P. Bartel, *Science,* 1995, **269**, 364.

51 A. Eschenmoser and M. Dobler, *Helv. Chim. Acta,* 1992, **75**, 218.

52 M. Böhringer, H.-J. Roth, J. Hunziker, M. Göbel, R. Krishnan, A. Giger, B. Schweizer, J. Schreiber, C. Leumann and A. Eschenmoser, *Helv. Chim. Acta,* 1992, **75**, 1416.

53 J. Hunziker, H.-J. Roth, M. Böhringer, A. Giger, U. Diederichson, M. Göbel, R. Krishnan, B. Juan, C. Leumann and A. Eschenmoser, *Helv. Chim. Acta,* 1993, **76**, 259.

54 G. Otting, M. Billeter, K. Wüthrich, H-J. Roth, C. Leumann and A. Eschenmoser, *Helv. Chim. Acta,* 1993, **76**, 2701.

55 A. Eschenmoser, *Pure Appl. Chem.,* 1993, **65**, 1179.

56 A. Eschenmoser and V. Kisakürek, *Helv. Chim. Acta,* 1996, **79**, 1249.

57 S. Pitsch, S. Wendeborn, B. Jaun and A. Eschenmoser, *Helv. Chim. Acta,* 1993, **76**, 2161.

58 L. E. Orgel, *Nature,* 1992, **358**, 203.

59 S. Pitsch, R. Krishnamurthy, M. Bolli, S. Wendeborn, A. Holzner, M. Minton, C. Lesueur, I. Sclövogt, B. Jaun and A. Eschenmoser, *Helv. Chim. Acta,* 1995, **78**, 1621.

60 S. D. Cook, P. S. Whitney and J. D. Sutherland, unpublished results.

61 M. Egholm, O. Buchardt, P. E. Nielsen and R. H. Berg, *J. Am. Chem. Soc.,* 1992, **114**, 1895.

62 M. Egholm, P. E. Nielsen, O. Buchardt and R. H. Berg, *J. Am. Chem. Soc.,* 1992, **114**, 9677.

63 P. Wittung, P. E. Nielsen, O. Buchardt, M. Egholm and B. Nordén, *Nature,* 1994, **368**, 561.

64 C. Böhler, P. E. Nielsen and L. E. Orgel, *Nature,* 1995, **376**, 578.

65 E. Drobner, H. Huber, G. Wächterhäuser, D. Rose and K. O. Stetter, *Nature,* 1990, **346**, 742.

66 D. Hafenbradl, M. Keller, G. Wächterhäuser and K. O. Stetter, *Tetrahedron Lett.,* 1995, **36**, 5179.

67 M. Keller, E. Blöchl, G. Wächtershäuser and K. O. Stetter, *Nature,* 1994, **368**, 836.

68 A. G. Cairns-Smith, *Genetic Takeover and the Mineral Origins of Life*; Cambridge University Press: Cambridge, 1982.

69 R. Shapiro, *Prebiotic Ribose Synthesis : A Critical Analysis*; D. Reidel Publishing Co.: 1988; Vol. 18.

70 G. F. Joyce, G. M. Visser, C. A. A. v. Boeckel, J. H. v. Boom, L. E. Orgel and J. v. Westrenen, *Nature,* 1984, **310**, 602.

71 R. Larralde, M. P. Robertson and S. L. Miller, *Proc. Natl. Acad. Sci. USA,* 1995, **92**, 8158.

72 R. B. Tewari and R. N. Goldberg, *Biophysical Chemistry,* 1985, **22**, 197.
73 G. Schlesinger and S. L. Miller, *J. Am. Chem. Soc.,* 1973, **95**, 3729.
74 W. D. Fuller, R. A. Sanchez and L. E. Orgel, *J. Mol. Biol.,* 1972, **67**, 25.
75 R. Osterberg, L. E. Orgel and R. Lohrmann, *J. Mol. Evol.,* 1973, **2**, 231.
76 R. H. Lohrmann and L. E. Orgel, *Science,* 1971, **171**, 490.
77 T. Inoue and L. E. Orgel, *J. Mol. Biol.,* 1982, **162**, 201.
78 T. Inoue and L. E. Orgel, *J. Am. Chem. Soc.,* 1981, **103**, 7666.
79 B. J. Weimann, R. H. Lohrmann, L. E. Orgel, H. Schneider-Bernloehr and J. E. Sulston, *Science,* 1968, **161**, 387.
80 J. W. Szostak and A. D. Ellington, *The RNA World*; Cold Spring Harbour Laboratory Press: Cold Spring Harbour, New York, 1993, pp 511-533.
81 R. H. Lohrmann and L. E. Orgel, *Tetrahedron,* 1978, **34**, 853.
82 J. P. Ferris, H. Yanagawa, P. A. Dudgeon, J. W. J. Hagan and T. E. Mallare, *Origins of Life,* 1984, **15**, 29.
83 J. P. Ferris, A. R. H. Jr, R. Liu and L. E. Orgel, *Nature,* 1996, **381**, 59.
84 W. S. Zielinski and L. E. Orgel, *Nature,* 1987, **327**, 346.
85 G. F. Joyce, A. W. Schwartz, S. L. Miller and L. E. Orgel, *Proc. Natl. Acad. Sci., USA,* 1987, **84**, 4398.
86 A. W. Schwartz, J. Visscher, R. Van der Woerd and C. G. Bakker *Cold Spring Harbour Symposia on Quantitative Biology*, 1987, **52**, 37.
87 G. Spach, *Origins of Life,* 1984, **14**, 433.
88 J. D. Sutherland and G. W. Weaver, *Tetrahedron Lett.,* 1994, **35**, 9105.
89 G. Challa and Y. Y. Tan, *Pure Appl. Chem.,* 1981, **53**, 627.
90 M. P. Robertson and S. L. Miller, *Science,* 1995, **268**, 702.
91 W. Saenger, *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984, pp21.
92 H. Bredereck, O. Christmann and W. Kosser, *Chem. Ber.,* 1960, **93**, 1206.
93 S. Metzger and B. Lippert, *Angew. Chem. Intl. Ed. Engl.,* 1996, **35**, 1228.
94 J. E. Hodge, *Adv. Carbohyd. Chem.,* 1955, **10**, 169.
95 L. Stryer, *Biochemistry*; W. H. Freeman: New York, 1981, pp 496.
96 U. Hommel, A. Lustig and K. Kirschner, *Eur. J. Biochem.,* 1989, **180**, 33.
97 H. Sugiyama, Y. Tsutsmi, K. Fujimoto and I. Saito, *J. Am. Chem. Soc.,* 1993, **115**, 4443.
98 J. D. Sutherland and G. W. Weaver, *Tetrahedron Lett.,* 1994, **35**, 9109.
99 J. D. Sutherland and G. W. Weaver, *Tetrahedron Lett.,* 1994, **35**, 9113.
100 S. D. Cook and J. D. Sutherland, *Tetrahedron Lett.,* 1996, **37**, 5779.
101 P. Reichard, *Science,* 1993, **260**, 1773.
102 E. C. C. Lin, *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology (Ed. in Chief, F. C. Neidhardt)* American Society for Microbiology: Washington DC, 1987, pp. 244-284.
103 B. Cuenoud and J. W. Szostak, *Nature,* 1995, **375**, 611.
104 S. D. Cook and J. D. Sutherland, unpublished results.

(*Received* 16 *April* 1997)

Biographical Sketch

John D. Sutherland

J. Nicole Whitfield

John Sutherland was born in 1962 and received his BA (first class) in 1984 from Oxford University. He has remained in Oxford since, taking his D. Phil. in 1988 and as a Junior Research Fellow before taking up the post of University Lecturer in 1990. His research interests focus on the application of chemical and biochemical techniques to the study of evolution. He is also involved in combinatorial chemistry and screening methodology and maintains an active interest in the biosynthesis of β-lactam antibiotics.

Nicole Whitfield, born in 1971 received her BA (first class) in 1993 from Oxford University. She remained in Oxford to read for her D. Phil. under the supervision of John Sutherland, investigating the synthesis and polymerisation behaviour of potentially prebiotic monomers of RNA. Having successfully completed her D. Phil. in 1996 she moved to Glaxo Wellcome Research and Development to take up a position as a Medicinal Chemist.