REVIEW

RNA: Prebiotic Product, or Biotic Invention?

by **Carole Anastasi**, **Fabien F. Buchet**, **Michael A. Crowe**, **Alastair L. Parkes**, **Matthew W. Powner**, **James M. Smith**, and **John D. Sutherland***

School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, UK (phone: (+44) 161 275 4614; fax: (+44) 161 275 4939; e-mail: john.sutherland@manchester.ac.uk)

Spectacular advances in structural and molecular biology have added support to the '*RNA world*' hypothesis, and provide a mandate for chemistry to explain how RNA might have been generated prebiotically on the early earth. Difficulties in achieving a prebiotically plausible synthesis of RNA, however, have led many to ponder the question posed in the title of this paper. Herein, we review recent experimental work on the assembly of potential RNA precursors, focusing on methods for stereoselective C–C bond construction by aldolisation and related processes. This chemistry is presented in the context of a broader picture of the potential constitutional self-assembly of RNA. Finally, the relative accessibility of RNA and alternative nucleic acids is considered.

Introduction. – A robust, prebiotically plausible synthesis of RNA, if achieved, will dramatically strengthen the case for the '*RNA world*' hypothesis [1][2]. Despite nearly half a century of effort, however, the prospects for such a synthesis have appeared somewhat remote. Difficulties in the generation and oligomerisation of activated nucleotides have led to suggestions that RNA might have been preceded by a 'simpler' informational macromolecule [1–3]. It has been suggested that a biology based on this simpler nucleic acid might have then 'invented' RNA. According to this scheme, functional superiority of RNA would have subsequently driven the transition to a biology based on RNA, and the RNA world would have been born (*Fig. 1*) [1][2].

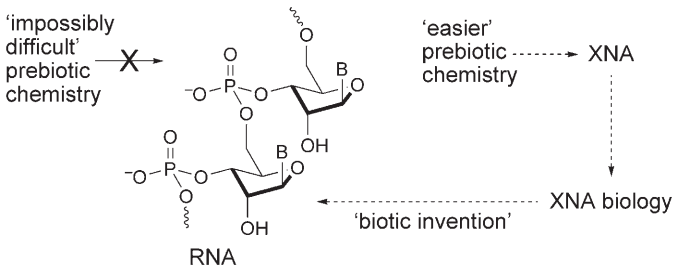

Fig. 1. *RNA as a biotic invention.* According to this scenario, potentially prebiotic chemistry leading to RNA has been tried (plain arrow), and found not to work (cross). A 'simpler' nucleic acid XNA is assumed to arise more easily, develop a biology, and 'invent' RNA (dashed arrows).

In this paper, we emphasise the large number of potential chemical routes to RNA from prebiotic feedstock molecules, and suggest that, until all these routes have been experimentally explored, it is premature to conclude that constitutional self-assembly of RNA is unlikely. Results of our preliminary exploration of other routes to RNA are then reviewed, and a comparison is made between the (potential) assembly chemistry of RNA and some other nucleic acids that have been suggested as RNA precursors such as threose nucleic acid (TNA) [4][5] (although *Albert Eschenmoser* has pointed out to *J. D. S.* that TNA did not result from a search for a simpler nucleic acid alternative/precursor to RNA) or glycol nucleic acid (GNA) [6].

Constitutional Analysis of RNA Reveals Many Potential Assembly Routes. – RNA has traditionally been viewed as a polymer of activated nucleotides **1** or **2**, and these nucleotides have been assumed to derive from phosphate, ribose, and nucleobase building blocks (*Fig. 2*) [1][2]. In terms of constitution, ribose is a pentamer of formaldehyde, and some sort of oxygenous chemistry of this prebiotic feedstock molecule has usually been envisaged as the source of ribose in RNA [7]. The nucleobases can be retrosynthetically broken down into small nitrogenous prebiotic feedstock molecules such as hydrogen cyanide, cyanoacetylene, and cyanamide. Various syntheses of the purine [8][9] and pyrimidine [10] nucleobases from these feedstocks, or their hydrated derivatives, have been described, perhaps the most

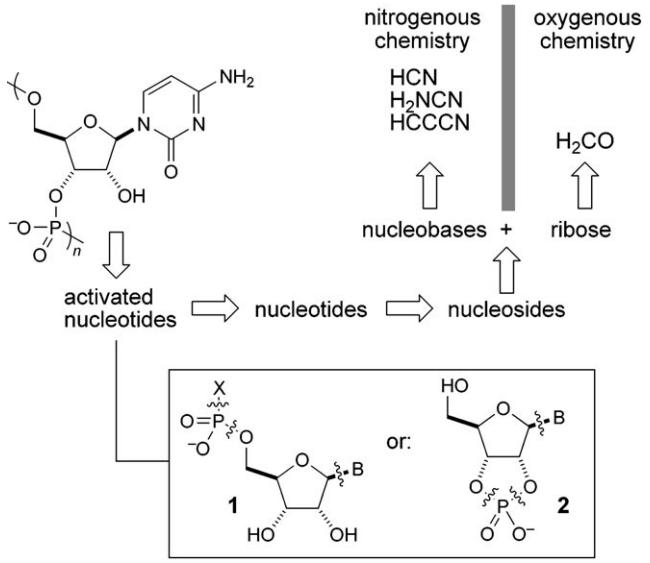

Fig. 2. *Traditional retrosynthetic disconnection of RNA.* This superficially simple retrosynthesis has been widely adopted presumably because of the apparent obviousness of the P–O and C(1′)–nucleobase bond cleavages, and the fact that the synthesis implied is reminiscent of the way RNA is biosynthesised in contemporary biology. Failure to realise the implied synthesis using prebiotically plausible chemistry has led to a widespread feeling that RNA must have been preceded by a 'simpler' nucleic acid. An alternative interpretation of the negative experimental results is that the traditional dogma should be rejected, and other disconnections should be explored.

notable being *Oro*'s synthesis of adenine by pentamerisation of HCN [8]. These simplistic constitutional relationships, and the apparent obviousness of the retrosynthetic disconnections that reduce a nucleotide to phosphate, ribose, and nucleobase building blocks, have contributed to a process by which the traditional disconnection of RNA has become dogma. Additionally, there has been a consensus of opinion that the oxygenous chemistry giving rise to ribose must have been separated (spatially and/or temporally) from the nitrogenous chemistry giving rise to the nucleobases [4]. For the purposes of this paper, it is convenient to depict routes to RNA from prebiotic feedstock molecules in a graphical manner that shows rough constitutional relationships but omits chemical details. The route to RNA implied by the traditional disconnection is shown in this manner in *Fig. 3*. As mentioned in the *Introduction*, there are difficulties with this potential route to RNA. Although it has proved possible to synthesise the nucleobases from nitrogenous feedstocks, the yields are low [8–10]. Likewise, although mixtures of sugars can be elaborated from simple precursors, yields of ribose are low [7], it is unstable [11][12], and no prebiotically realistic means of purifying it has been described. Furthermore, attachment of preformed pyrimidine nucleobases to ribose has not proved possible for kinetic and thermodynamic reasons,

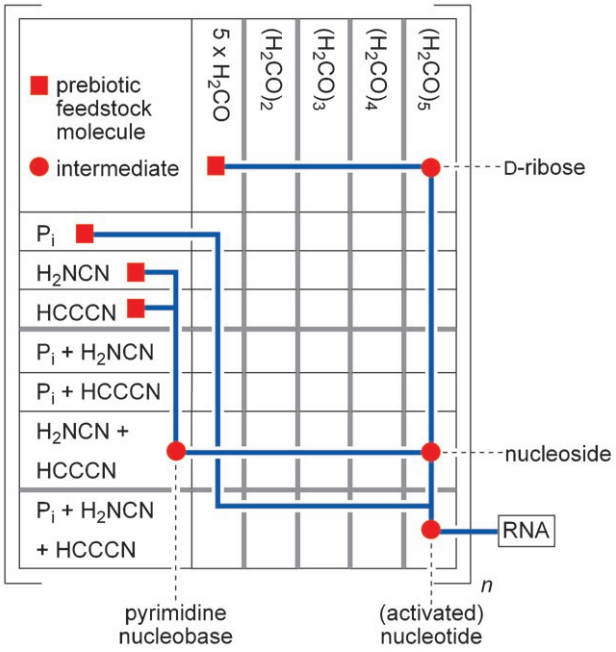

Fig. 3. *The traditionally assumed prebiotic synthesis of RNA is just one of many potential syntheses.* This constitutional self-assembly grid makes clear the many potential routes to RNA from its constituent feedstock molecules. The traditionally assumed synthesis (blue lines) involves oxygenous chemistry from formaldehyde to ribose, and nitrogenous chemistry leading to nucleobases. Attachment of the nucleobases to ribose is followed by phosphorylation, activation, and polymerisation. For simplicity, the grid only shows pyrimidine RNA assembly options, but similar grids can be constructed for assembly of purine RNA in which the nitrogenous constituents are $5 \times$ HCN (adenine), or $(4 \times \text{HCN}) + \text{H}_2\text{NCN}$ (guanine, diaminopurine).

and attachment of the purines lacks stereo- and regiocontrol and is very low-yielding [13]. In short, the traditional retrosynthetic disconnection of RNA equates with chemistry that just does not work! This is why alternate, 'simpler' precursor nucleic acids are, at least ostensibly, conceptually attractive [1–3]. However, it is also possible that the dogma of the traditional disconnection is wrong, and that there might be an extremely plausible prebiotic synthesis of RNA, as yet undiscovered. Indeed, it is clear from *Fig. 3* that there are many other *potential* routes from the prebiotic feedstock molecules to RNA. We set out to investigate this uncharted chemical space several years ago on the basis that it seemed premature to write off RNA as the ancestral genetic and functional macromolecule until all of this chemical space was comprehensively explored. This exploration is by no means complete, and has not yet revealed a complete predisposed route to RNA, but it has uncovered some interesting chemistry that may be of relevance. As an example, building on early work of the groups of *Orgel* [14] and *Nagyvary* [15], we discovered a route to activated pyrimidine nucleotides in which the nucleobase is assembled in stages on a sugar phosphate template (*Fig. 4*) [16]. Treatment of D-arabinose-3-phosphate with cyanamide in water at neutral pH gave the amino-oxazoline **3** in excellent yield. When cyanoacetylene was then added to an aqueous solution of **3**, β-D-cytidine-2′,3′-cyclic phosphate **2** (B = C) was formed [16]. This process is etiologically noteworthy for a number of reasons: the selection of the furanose ring system from a starting material which exists predominantly in pyranose forms, the establishment of the β-configuration at C(1′), the inversion of configuration

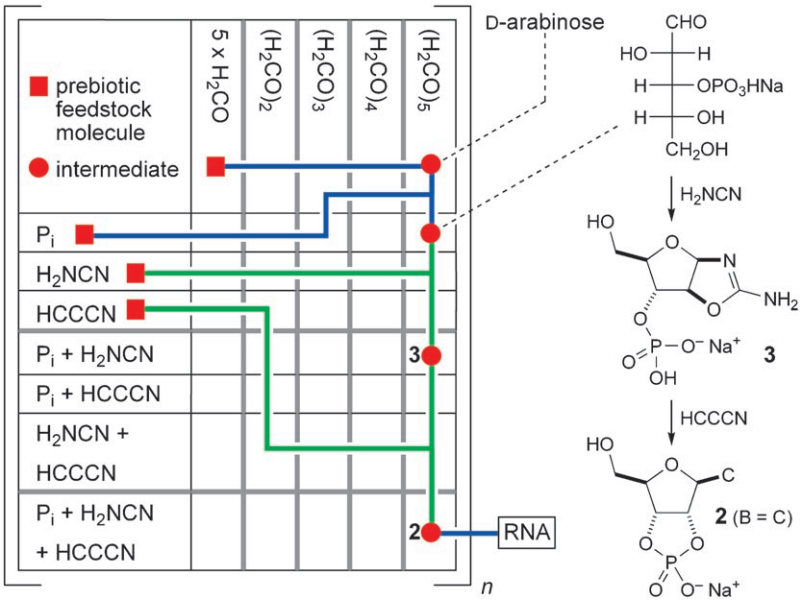

Fig. 4. *An alternate route to RNA involving stepwise nucleobase assembly on a sugar phosphate template.* The conversion of D-arabinose-3-phosphate to β-D-cytidine-2′,3′-cyclic phosphate **2** has been experimentally demonstrated (green lines). The potential synthesis of the sugar phosphate has been investigated and appears unlikely [17], and the polymerisation of **2** remains to be fully explored (blue lines).

at C(2′), and the increase in activation of the phosphate group from a monoester to a diester. The big problem[1]) with this approach, however, is that the particular pentose phosphate needed, arabinose-3-phosphate, is not one that can be accessed using prebiotically plausible phosphorylation chemistry (or simple aldolisation of phosphorylated C_2 or C_3 units) [17].

Thinking Outside the Box – A Potential Aldolisation Route to the p-RNA Backbone. – *Eschenmoser* made one of the first suggestions for a synthesis of RNA different to that implied by the traditional disconnection. In his seminal work on pyranosyl-RNA (p-RNA) [18–20], it was pointed out that the backbone of this potentially prebiotic nucleic acid can be formally derived from bis(glycolaldehyde) phosphate **4** and formaldehyde by aldolisation (*Fig. 5*) [20]. After stereoselective nucleobase addition to this aldol backbone giving p-RNA, a deep-seated isomerisation to RNA could be imagined.

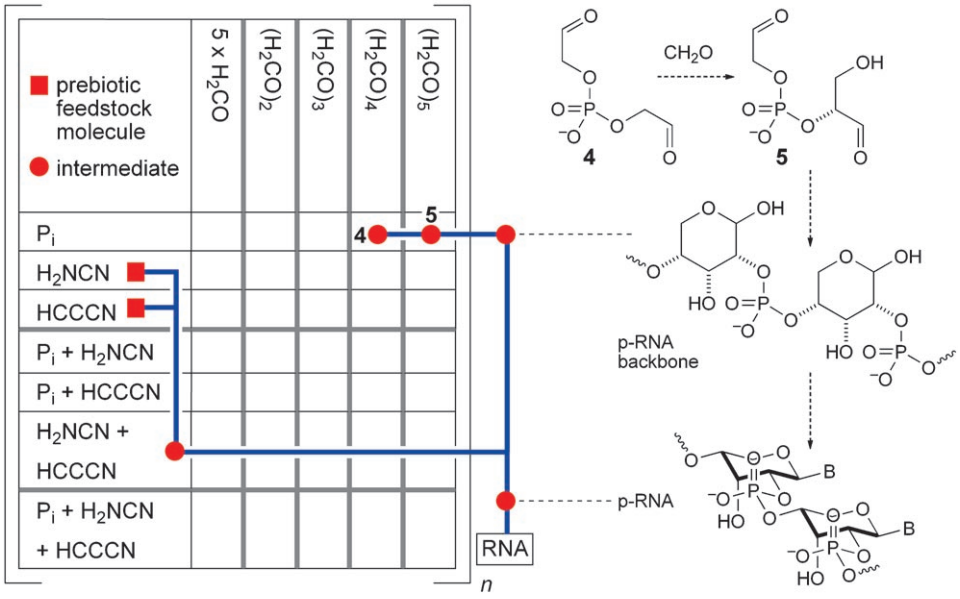

Fig. 5. *A potentially alternate route to RNA involving aldol oligomerisation.* This radically different RNA assembly option proposed by *Eschenmoser* was experimentally unexplored at the outset of the work reviewed herein.

We decided to investigate this potential route to RNA by synthesising **4** and studying its aldolisation behaviour in the presence of formaldehyde to see if the p-RNA backbone could be made in this way [21][22]. It was already known at the time that cyclisation of **4** in dilute alkali is rapid [23] potentially making hydroxymethylation to **5** difficult, and suggesting that, even if **5** could be produced, it too would be more likely

[1]) A smaller problem is that the reaction of **3** with cyanoacetylene also results in the production of β-D-arabinocytidine-3′-phosphate. An anhydronucleotide intermediate partitions to this by-product and **2** (B = C).

to cyclise than oligomerise. These *caveats* notwithstanding, the idea of an aldolisation route to nucleic acid backbones struck us (as it struck the idea's originator [20]) as an excellent way of avoiding the pitfalls inherent in the traditionally assumed P–O bond forming oligomerisations to make RNA in water[2]), and so we proceeded to look into it. To start with, we considered the stereochemistry of the potential competing ring-closure of **5** to pentose-2,4-cyclic phosphates **6**–**9** [22]. Steric hindrance in all idealised transition states leading to the *arabino*- and *lyxo*-products, **6** and **7**, respectively, initially gave us hope as it suggested there would be a barrier to the formation of these products that might lead to oligomerisation of **5** being favoured over cyclisation (*Fig. 6*). However, such hope evaporated when we realised that certain idealised transition states leading to the *ribo*- and *xylo*-products, **8** and **9**, respectively, suffered no such steric hindrance (*Fig. 7*). In the event, we found that it was possible to hydroxymethylate **4** with formaldehyde at pH 9.5, albeit in competition with cyclisation of **4** to erythrose- and threose-2,4-cyclic phosphates **10** and **11**, respectively [21]. The intermediate **5** did not oligomerise, however, and, as we had begun to suspect, cyclisation to **8** and **9** was observed instead (*Fig. 8*). In D_2O solution, **9** exists exclusively in the hydrated aldehyde form, and **8** is mainly hydrated but also exists in pyranose form [22]. This structural complexity, and the presence of unidentified materials amongst the products of the reaction of **4** and formaldehyde, made accurate quantitation difficult, but the relative abundance of identifiable tetrose and pentose derivatives was $\mathbf{11} \gg \mathbf{10} \approx \mathbf{8} > \mathbf{9}$, and the absolute combined yield of these four products was of the order of 70–80%. The lack of oligomerisation of the intermediate **5** to the p-RNA backbone militates against an aldolisation approach to p-RNA and thence RNA[3]). Nevertheless, the stereoselective synthesis of the pentose phosphates **8** and **9** is of potential etiological significance.

Amino-oxazolines Revisited, Can Ribose Be Bypassed? – Even if a way round the difficulties of producing the p-RNA backbone by aldolisation could be found, the route to p-RNA and RNA just discussed suffers from the additional problem of nucleobase addition to the ribose residues of the backbone. Direct nucleobase addition is likely to be no easier than it is with free ribose, and stepwise assembly of pyrimidine bases similar to that shown in *Fig. 4* is not possible because of the phosphorylation of the 2-position. This phosphorylation prevents formation of 1′,2′-fused amino-oxazolines, and we found that pentose-2-phosphate derivatives were unreactive towards cyanamide [22]. Considerations along these lines, along with our inability to find a prebiotically plausible route to D-arabinose-3-phosphate [17], made us reconsider RNA assembly options proceeding through unphosphorylated amino-oxazolines. *Orgel* and co-workers had first explored this in the late 1960s, and found that when D-ribose reacts with cyanamide an amino-oxazoline **12** is formed in good yield (*Fig. 9*) [14].

---

[2]) Perhaps the major pitfall is competing hydrolysis of the activated phosphate group.

[3]) Competing cyclisation is also a potential problem with other monomer oligomerisation chemistries. In the case of the monomers for RNA formation by P–O bond-forming oligomerisation, the geometry of **2** precludes cyclisation to the nucleoside-3′,5′-cyclic phosphate, and, although formation of the latter from **1** is potentially kinetically feasible, it is thermodynamically unfavourable because of strain in the six-membered cyclic nucleotide [24].

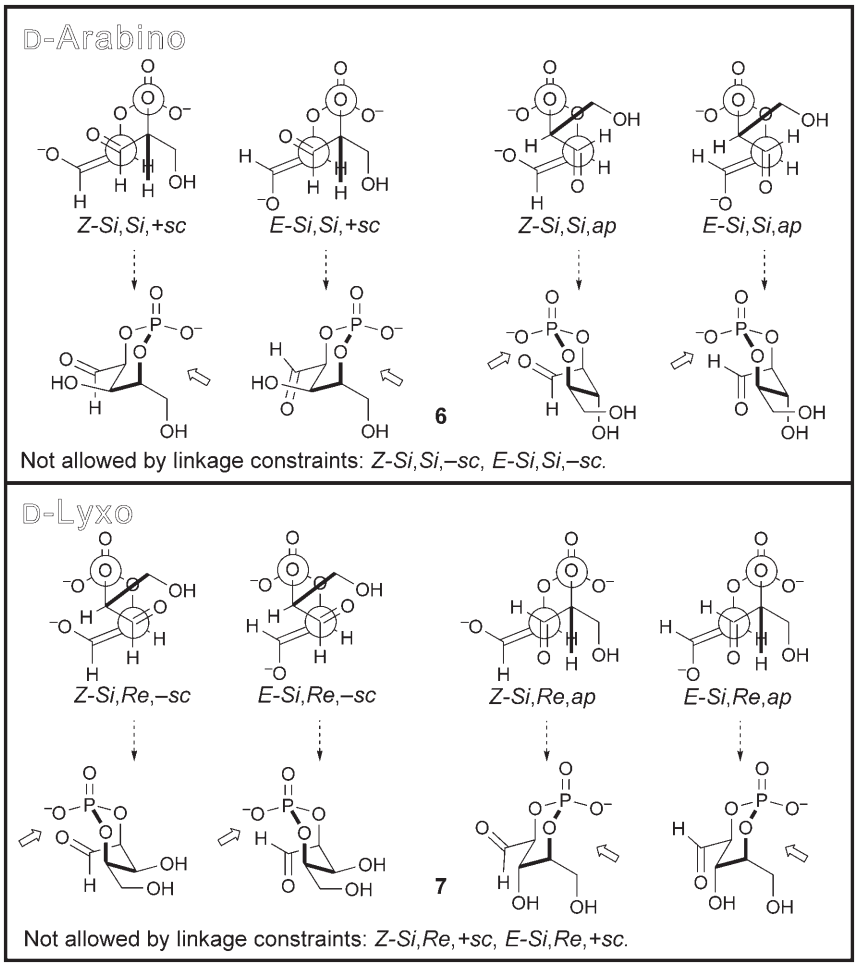

Fig. 6. *Predicted steric constraints in the kinetic aldolisation of* **5** *to the* D*-arabinose- and* D*-lyxose-2,4-cyclic phosphates* **6** *and* **7**. Approach geometries and nascent product conformers (D-configured only) for the cyclisation of **5** (*via* aldol reaction of the glycolaldehyde enolate with the glyceraldehyde C=O group). The four stereochemical descriptors relate to the: enolate geometry, enolate face, carbonyl face, and the rotamer of the nascent C(2)–C(3) bond (OPO_3^{2-} with respect to C=O).

Our group [25] and *Joyce*'s group [12] re-investigated this reaction recently and determined the structure of **12** by X-ray crystallography. *Sanchez* and *Orgel* had further shown that **12** reacts with cyanoacetylene to give α-D-cytidine **13**, again in good yield, and had then gone on to demonstrate that **13** could be converted to β-D-cytidine **14** by a photo-anomerisation process [14]. The photoanomerisation was very low yielding, however, and this, coupled to the notorious difficulties in making ribose in a prebiotically plausible manner, made the overall scheme less attractive as part of a route to RNA.

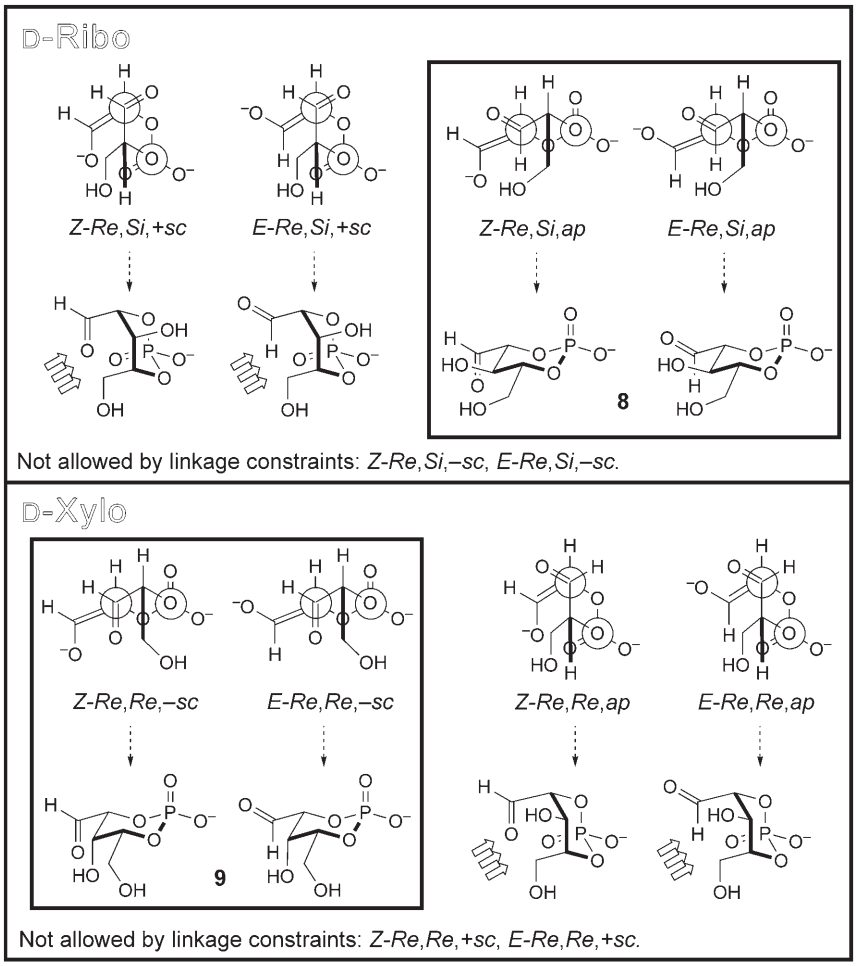

Fig. 7. *Predicted stereochemical preferences in the kinetic aldolisation of* **5** *to the* D*-ribose- and* D*-xylose-2,4-cyclic phosphates* **8** *and* **9**

We have investigated in some detail the mechanism of formation of amino-oxazolines, such as **12**, from the free sugars and cyanamide [25][26]. In the course of this study, we found that the behaviour of lyxose was somewhat different to that of the other pentoses. All of the pentoses formed furanosylamino-oxazolines, but lyxose also formed a pyranosylamino-oxazoline **15** (p) [26]. Furthermore, the pyranosyl- and furanosyl-D-lyxose amino-oxazolines were shown to be in equilibrium (**15** (p)/**15** (f) *ca.* 5 : 1) presumably *via* the open-chain iminium ion **16** (*Fig. 10*). Thinking about this mechanism, we realised that **16** and other stereoisomeric iminium ions might be accessible by reaction of 2-aminooxazole (**17**) with glyceraldehyde **18**, if **17** was a sufficiently good C-nucleophile to react with an α-hydroxy aldehyde [27]. If these iminium ions could be formed by reaction of **17** and **18**, then conversion to the amino-

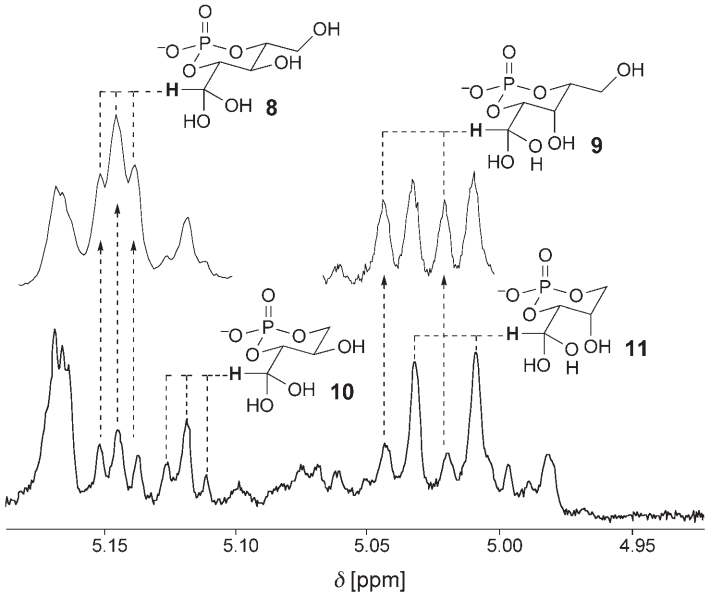

Fig. 8. *^{1}H-NMR Analysis of the aldolisation of* **4** *and 1 equiv. of formaldehyde.* NMR Spectroscopy, especially in conjunction with sample spiking with authentic standards, is the mainstay of our analytical repertoire. This figure is an expansion of the region of the ^{1}H-NMR spectrum of the reaction products of **4** and formaldehyde, showing signals due to aldehyde hydrates with spectra from spiked samples inset [21].

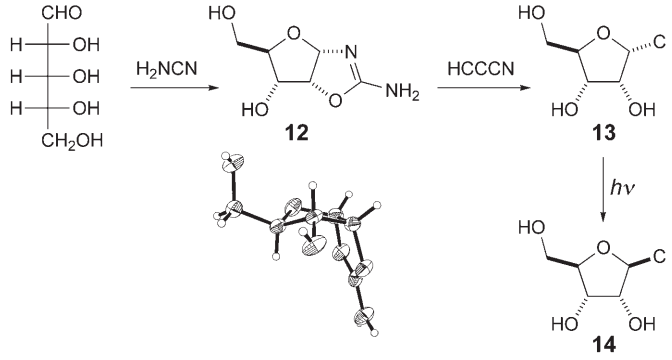

Fig. 9. *Synthesis of cytidine nucleosides* via *the amino-oxazoline* **12**. The conversion of D-ribose to α-D-cytidine **13** is prebiotically very plausible, but synthesis of D-ribose, and photoanomerisation of **13** to **14** are not (see text).

oxazolines **12**, **15**, **19**, and **20** seemed their most likely fate. To our delight, these expectations were realised, and it turned out that the reaction of **17** with **18** proceeded smoothly in essentially quantitative yield [28]. The reaction was also highly stereoselective with the *ribo*- and *arabino*-products **12** and **19**, respectively, predominating

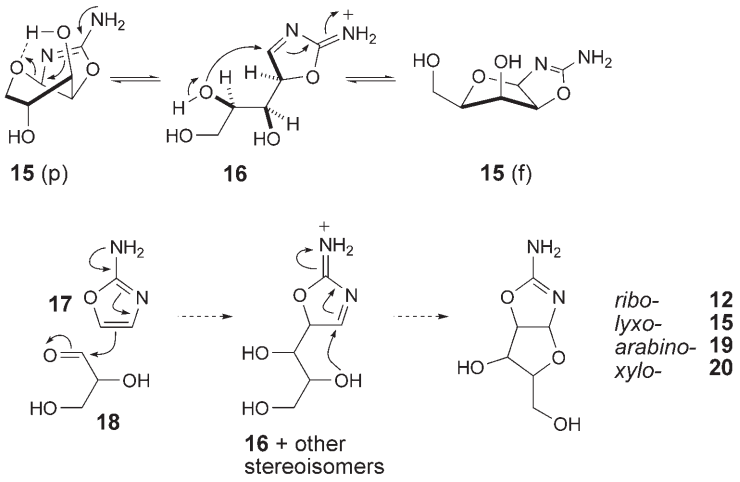

Fig. 10. *Interconversion of pyranosyl- and furanosyl-*D*-lyxose amino-oxazolines, and possible generation of pentose amino-oxazolines involving C–C bond formation.* By thinking about a possible mechanism for the interconversion of **15** (p) and **15** (f), we recognised the potential for generation of all the pentose amino-oxazolines from **17** and **18**.

(*Fig. 11*). The *ribo*-isomer **12** is the least soluble pentose amino-oxazoline [12][14], and we found that if the products of the reaction of **17** and **18** were simply cooled, then partial crystallisation of **12** took place. It transpires that *rac*-**12** is more soluble than enantiomerically pure **12**, and that both forms give crystals with the $P2_12_12_1$ space group. When solutions of scalemic (*sca*-) **12** with ee > 60% were cooled, the crystals of **12** that formed were composed of enantiomerically pure material. Extending this result, we found that if **17** was allowed to react with *sca*-**18** with a D-ee of 60%, and the products were cooled, then enantiomerically pure crystals of D-**12** were again formed. We thus have a very powerful reaction that not only leads to a nucleoside precursor in high yield, but also purifies it by a crystallisation process that effectively amplifies any ee values in the input glyceraldehyde. Since **17** is a condensation product of glycolaldehyde and cyanamide [29], this route to **12**, and thence **13**, can be seen to involve mixed oxygenous and nitrogenous chemistry (*Fig. 12*). Despite the simplicity of this route to enantiomerically pure **12** and **13**, several additional features are needed before it can be viewed as part of a prebiotically plausible route to nucleotide precursors of RNA. Firstly, glyceraldehyde **18** is not a known prebiotic feedstock molecule, and a route to it from an available compound such as dihydroxyacetone (DHA) [30] must be found. This conversion must additionally generate **18** with ee > 60% if **12** and **13** are subsequently to be produced in enantiomerically pure form. Secondly, the photo-anomerisation of **13** to **14** is inefficient [14], and must either be optimised, or an alternative must be found. Thirdly, phosphate has to be introduced at some point so that nucleotides, rather than nucleosides, are the end products.

We have only just started to investigate these additional features, but found some interesting chemistry early on. We recognised that if **17** underwent reaction with D-glyceraldehyde-3-phosphate **21**, D-pentose amino-oxazoline-5′-phosphates **22**–**25**

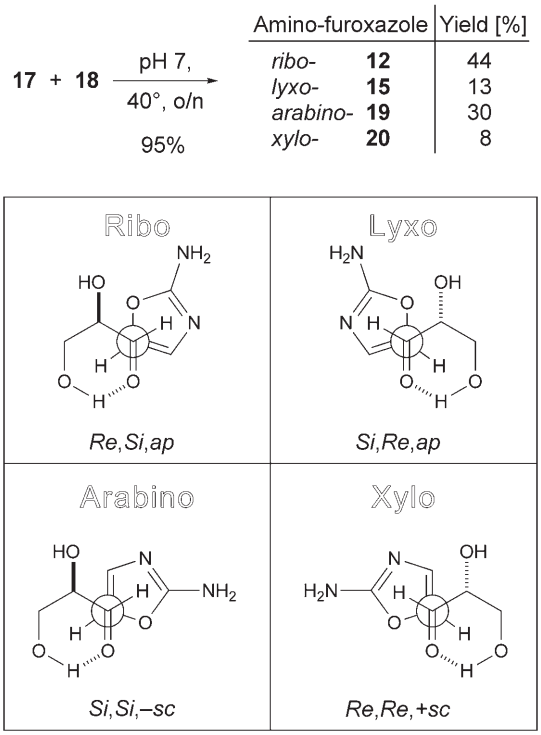

Fig. 11. *Results and rationalisation of the addition of* **17** *to* **18**. The figure shows approach trajectories associated with idealized staggered transition structures leading to the various products. The stereochemical descriptors refer, in order, to the face of **17** and the face of the C=O group of **18** involved in bond formation, and the relative disposition of the C=O group to the CH–O bond of **17**. A significant preference for addition of **17** to the *Si* face of **18** (left two structures), and a slight preference for the *erythro*-arrangement of substituents about the newly formed C(2′)–C(3′) bond (upper two structures) are observed.

would most likely be formed (*Fig. 13*). However, we expected such a reaction to be difficult to realise experimentally because of the known propensity of **21** to undergo elimination of phosphate [31]. Somewhat to our surprise, the addition chemistry predominated over the destructive elimination of phosphate from **21**[4]), and the amino-oxazoline-5′-phosphates **22**–**25** were formed in good yield [32]. This result is testimony to the mildness of the conditions under which **17** reacts with α-hydroxy aldehydes. When we first examined the product mixture by ^{1}H-NMR spectroscopy, it was evident that one stereoisomer was greatly favoured, but we could not assign it with certainty. The D-*ribo*-configured stereoisomer **22** can be alternatively prepared from D-ribose-5-phosphate and cyanamide [14][32], so we used a sample synthesised in this way to spike our reaction products. In this way, we showed that the favoured isomer is **22**. To further prove this, and to assign and quantify the other isomers, we enzymatically

[4]) We used the D-enantiomer simply because it was commercially available and *rac*-**21** was not.

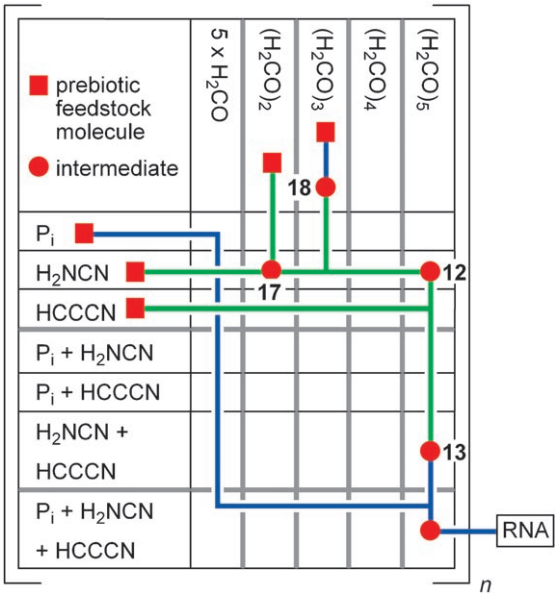

Fig. 12. *The chemistry of* Fig. 11 *in the context of a complete RNA synthesis.* One way of incorporating the remarkable conversion of **17** and **18** to **12** into a complete RNA synthesis is shown. Synthesis of **17** from glycolaldehyde and cyanamide [29], and conversion of **12** to **13** [14] extend the experimentally demonstrated chemistry (green lines). Needed are a conversion of dihydroxyacetone [30] to **18**, a means of phosphorylating and anomerising **13**, and activated nucleotide polymerisation (blue lines).

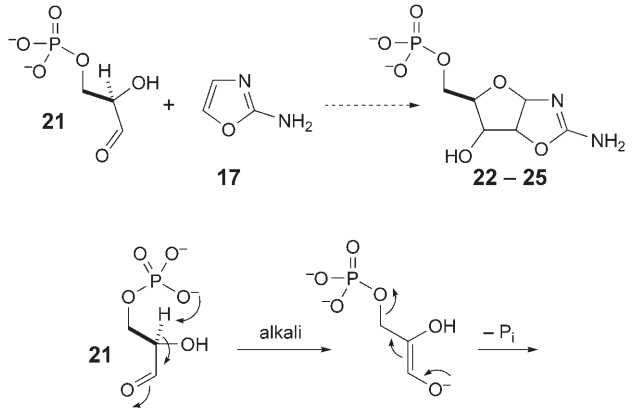

Fig. 13. *Potential generation of pentose amino-oxazoline-5′-phosphates* **22**–**25**. The hypothetical addition of **17** to **21** (dashed arrow) was initially considered unlikely given the known tendency of **21** to undergo (auto-induced) *E*1*cb* irreversible elimination of inorganic phosphate.

dephosphorylated a sample of the reaction products and separately spiked aliquots with samples of the free amino-oxazolines D-**12**, D-**15**, D-**19**, and D-**20** [26]. The overall yield of products is lower than that of the reaction of **17** with **18**, but in the phosphorylated

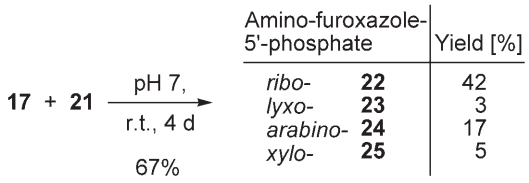

Fig. 14. *Stereoselectivity in the addition of* **17** *to* **21**

series the relative predominance of the *ribo*-configured product is higher (*Fig. 14*). It is not yet clear whether this stereoselectivity is kinetic, thermodynamic, or some combination of both[5]).

The reaction of **17** with **21** produces the ribose amino-oxazoline-5′-phosphate **22** in good yield, and reaction of **22** with cyanoacetylene generates 5′-nucleotides neatly getting round the requirement for the regioselective phosphorylation of **13** or **14** for example. However, the requirement for **21** raises the question of the provenance of such an unstable intermediate, and (most simplistically) invokes the regioselective phosphorylation of glyceraldehyde **18**. Regioselective and prebiotically plausible phosphorylation of D-**18** has been demonstrated, but the 2-phosphate is produced and not the 3-phosphate **21** [33]. More work is, therefore, needed to try and find a way of converting ribose amino-oxazoline **12** or its derivatives into the activated nucleotides **1** or **2**.

A Prebiotically Plausible Synthesis of RNA Still Looks Difficult, But Will the Synthesis of Ostensibly Simpler Nucleic Acids Be Any Easier? – Despite the discovery of new routes to ribose amino-oxazoline **12** and the 5′-phosphate **22**, a predisposed synthesis of RNA still looks a long way off. In particular, the polymerisation of the activated nucleotides **1** [34] and/or **2** [35] has, thus far, been relatively unsuccessful, and must be considered further. Our initial exploration of the previously uncharted areas of RNA constitutional self-assembly space has uncovered a rich chemistry, however, and further exploration is surely warranted. It is too early to consider what might be there to discover, but mixed nitrogenous/oxygenous chemistry clearly has great synthetic potential if it can be controlled. Although our exploratory work is incomplete, some points concerning the relative simplicity of RNA and other 'simpler' nucleic acids have already begun to emerge. The ensuing discussion will only deal with TNA and GNA in comparison to RNA, but the analysis can easily be extended to include other nucleic acids.

As a starting point, the three nucleic acids may be compared structurally (*Fig. 15*). At first glance, it is tempting to rank these polymers according to structural criteria such as the number of atoms and stereogenic centres per repeating unit. According to these simplistic criteria, TNA appears simpler than RNA [36], and GNA appears simpler still. Furthermore, the furanose rings of RNA and TNA also make them seem structurally more complex than GNA with its acyclic repeating unit. A slightly different

---

[5]) We note, however, that if there is a kinetic element, the transition state model of the reaction in the non-phosphorylated series (*Fig. 11*) will have to be modified.

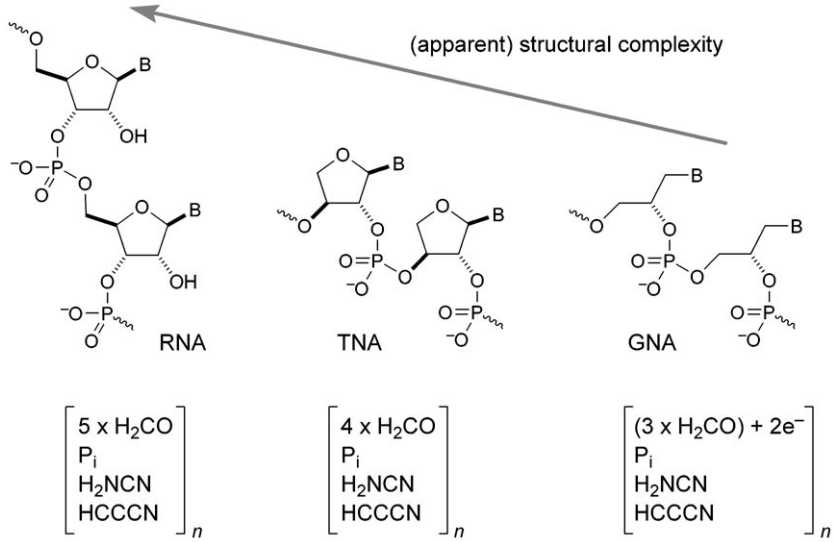

Fig. 15. *Structural and constitutional comparison of RNA and two of its proposed precursors.* Whilst structural complexity is somewhat in the eye of the beholder, the trend shown by the arrow is difficult to deny. Constitutional complexity with respect to an assumed (and experimentally supported) basis set of feedstock molecules is less open to debate, and GNA can be seen to be more reduced than RNA or TNA.

picture emerges when the three nucleic acids are compared at the level of constitution. The repeating units of RNA and TNA are based on formaldehyde oligomers without any redox adjustment, but the repeating unit of GNA is based on a reduced formaldehyde trimer. The more reduced constitution of GNA is expected to be manifested in a requirement for reduction at some point in any assembly sequence starting from formaldehyde, or its dimer or trimers. In addition, the redox alteration being at the position of nucleobase linkage implies fundamentally different nucleobase attachment, or stepwise synthesis chemistries relative to RNA and TNA. Thus, although the repeating unit of GNA has fewer atoms, stereogenic centres, and rings than those of RNA and TNA, a strong case can be made that GNA is constitutionally the most complex of the three nucleic acids. As regards a constitutional comparison between RNA and TNA, there seems little doubt that TNA is simpler than RNA. There is a special constitutional simplicity to the repeating unit of TNA in that it is formally comprised of two glycolaldehyde units [4]. This is in contrast to the situation with RNA (glycolaldehyde plus glyceraldehyde), and $C_2+C_2=C_4$ seems a lot simpler than $C_2+C_3=C_5$. The results of the comparative constitutional analysis of the three nucleic acids may be, therefore, summarised in terms of relative simplicity: TNA > RNA > GNA.

In addition to constitutional considerations, it is evident that the potential generational chemistry of a nucleic acid must be considered in depth before any assessment of likelihood of prebiotic existence can be made. As this review has made clear, the experimental exploration of the potential generational chemistry of RNA is far from over. A similar exploration of the chemistry of TNA and GNA has not yet

begun, and will be required before concrete etiological statements can be made. This exploration might profitably start with an assessment of the simplest options of monomers for oligomerisation by P–O bond formation. We have deferred our discussion of the polymerisation of the RNA precursors **1** and **2** up to now, so that the potential for P–O bond forming oligomerisation can be dealt with in a comparative sense (*Fig. 16*).

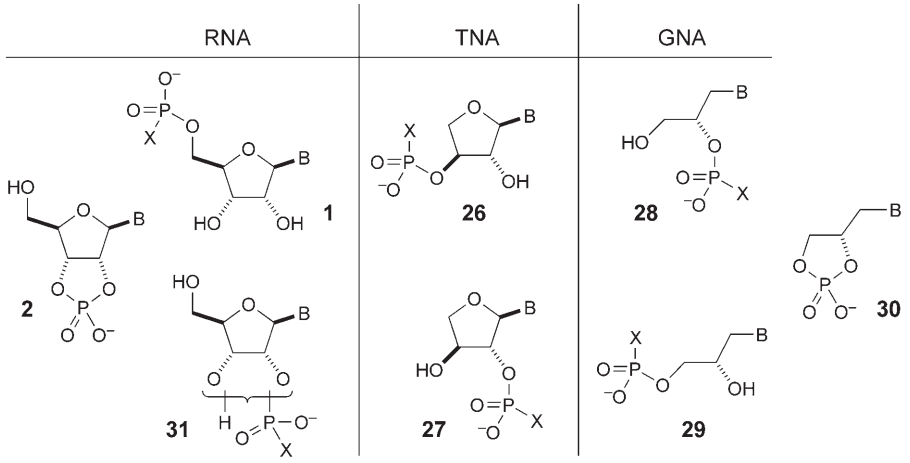

Fig. 16. *Candidate monomers for P–O bond forming polymerisations to RNA, TNA, and GNA*

On the face of it, the simplest such TNA monomers are the 3′- and 2′-activated nucleotides **26** and **27**, but there are major problems associated with both. Polymerisation of **26** or **27** would necessarily involve a secondary alcohol attacking an activated phosphate, and secondary alcohols are much less nucleophilic than primary alcohols or H_2O. Such a polymerisation in H_2O, therefore, looks doomed without selective catalysis. In the case of GNA, the 2′- and 3′-activated phosphates **28** and **29** are the first monomers that spring to mind, and **28** looks better than **29** because its oligomerisation would involve a primary rather than a secondary alcohol. A moment's reflection, however, suggests that cyclisation of either **28** or **29** to **30** is likely to preclude their oligomerisation[6]). In the RNA series, the 2′/3′-activated nucleotides **31** also undergo cyclisation, rather than oligomerisation, in this case to nucleoside-2′,3′-cyclic phosphates **2**. However, the big difference between **2** and **30** is the fact that **2** is still oligomerisation competent (because of the 5′-OH group), whereas **30** is not. Importantly, hydrolytic conversion of **2** to 2′/3′- nucleotides can be (continuously) repaired by reaction with activating agents such as cyanoacetylene [37]. This repair of competing hydrolysis, and the reversibility of 2′,3′-cyclic phosphate oligomerisation, opens up the prospect for a dynamic combinatorial synthesis of RNA oligomers [37]. The exciting thing about such a synthesis is that 3′→5′-linked homochiral duplexes might be the preferred products at (quasi)equilibrium for thermodynamic reasons [38].

---

[6]) In the same way that cyclisation of **5** to **8** and **9** prevented formation of the RNA backbone by aldolisation. The lack of strain in **30** suggests that the cyclisation of **28** and **29** will be facile.

There is a large body of experimental work relating to the oligomerisation of 5′-activated nucleotides **1** which will not be reviewed here [34]. Suffice it to say that the enhanced nucleophilicity of the 2′/3′ diol(ate) allows (template-directed) oligomerisation if the activating group is carefully chosen, but 2′→5′- and 3′→5′-linkages are formed in addition to other by-products. So, although there are problems that remain to be solved, there are options for P–O bond forming oligomerisation with RNA that do not exist with TNA or GNA, especially the prospect for thermodynamic, rather than kinetic selection in the oligomerisation of nucleoside-2′,3′-cyclic phosphates **2** [37].

In addition to the poor prospects for their oligomerisation, there are also problems in the potential generation of **26** and/or **27**, and **28** and/or **29** by a process of constitutional self-assembly. These problems are comparable, and, in some cases, worse than the problems faced in the RNA series. If the threose moiety of either **26** or **27** is to derive from two glycolaldehyde units by some sort of aldolisation process then a 3′-OH group would be produced. The 3′-phosphate group of **26** would, therefore, have to be added regioselectively after sugar assembly, and our experience in the pentose series suggests that this will be difficult [17]. Tetrose-2-phosphates can be produced by direct phosphorylation [33], but this does not necessarily facilitate a synthesis of **27** as the 2-phosphate blocks the stepwise nucleobase assembly options [22]. Alternatively, the sugar phosphate moiety of **27** could be produced directly by crossed aldolisation of glycolaldehyde phosphate and glycolaldehyde. Such an aldolisation might be demonstrable, but the problem of assembling the base on a 2-phosphate still remains. This brief analysis, therefore, suggests that the best route to either **26** or **27** would involve phosphorylation after nucleoside assembly, but there is nothing to suggest that this process could be achieved regioselectively[7]), and attempted polymerisation of a mixture of both isomers is likely to be unproductive.

As regards GNA, even if conditions could be found that favoured oligomerisation of **28** or **29** over cyclisation, there are major potential problems with the generational chemistry of either monomer because of the reduced state of C(1′). This potentially implies nucleobase (precursor) attachment by S_N2 reaction rather than by the addition/elimination chemistry allowed when C(1′) is at the higher, aldehyde oxidation level.

At this point, some tentative conclusions can be drawn. On the basis of its constitution, and the probable ease with which the monomers **28** and/or **29** will cyclise to **30**, GNA is likely to be less accessible by prebiotically plausible chemistry than either RNA or TNA. Of these latter two nucleic acids, TNA is constitutionally the simpler, but generational considerations are more important. Similar generational difficulties can be envisaged for the TNA monomers **26** and/or **27**, and the RNA monomers **1** and/or **2**, but the oligomerisation of the latter (especially **2**) offers more promise than oligomerisation of the former. The analysis is thus slightly in favour of RNA as the first informational macromolecule, but a precursor TNA world cannot yet be ruled out. The situation, therefore, has to be analysed further, and the possibilities for a genetic transition from TNA to RNA need to be explored. If such a transition is likely to be difficult, then further search for a prebiotically plausible synthesis of RNA might be

---

[7]) In contrast, phosphorylation of β-D-ribonucleosides, such as **14**, shows significant selectivity for formation of the 2′,3′-cyclic phosphate **2** (and/or the hydrolysis products of **2**) [39].

more easily justified than an experimental investigation of the possibility of a precursor TNA world.

What About a Transition Between Different Nucleic Acid-Based Biologies? – The likelihood of such a transition can only be speculated upon until it is experimentally investigated by synthetic biology. However, different evolutionary scenarios can be distinguished depending on the chemical properties of the incoming and outgoing nucleic acids. The key chemical properties would appear to be cross-pairing ability, and (sequence-dependent) structural relatedness. These properties are clearly related in that the more different the structure of two nucleic acids, the less likely they are to be able to cross-pair. To keep the arguments simple, let us 'cut to the chase' and consider the potential TNA → RNA transition. TNA and RNA can cross-pair and have related structures, in particular with respect to the intranucleotide P···P distance – one of the structural parameters that determines cross-pairing capability [40]. It is (remotely) possible that any particular sequence might have a similar tertiary structure in both TNA and RNA forms, but it seems unlikely that this will be true for all sequences. Since most function depends on structure, it, therefore, appears unlikely that much function could be retained in a one-step transition from a TNA system to an RNA system. Such a transition would, therefore, be a retrograde evolutionary step. In fact, the problem is more serious as a one-step transition is unlikely because *Darwinian* evolution occurs in a series of small steps. Either TNA could have been gradually converted to RNA *via* a heteropolymer T/RNA, or TNA could have gradually produced an RNA copy of itself, before being outdone by this copy. The gradual introduction of RNA residues into a TNA sequence would change structure, and unless, the ribo-substitution could be copied strand to strand, change structure in a random and unselectable way violating the principle of continuity. Furthermore, the incorporation of the RNA residues would have to involve a mechanism different to that of the incorporation of TNA residues because the monomers are not structurally analogous[8]). So, for a TNA → RNA transition to have occurred, an RNA copy of the TNA genome must have been gradually produced. This would have required the TNA biology to evolve catalysts for the synthesis of RNA monomers, and catalysts to oligomerise these monomers copying the TNA genome to an RNA genome. Such an evolutionary scenario is only barely imaginable even with some sort of progressive selection for each step, but when the chances are that most TNA-encoded function would be lost in the process, the scenario loses credibility.

In conclusion, TNA does not appear to be significantly simpler than RNA from a generational point of view, indeed there are options for P–O bond forming oligomerisation of RNA monomers that do not exist for TNA monomers. In addition, the prospects for a TNA → RNA transition appear remote, and so we prefer to continue to focus our experimental effort on exploration of the potential constitutional self-assembly of RNA (*Fig. 17*).

---

[8]) A situation that can be contrasted with the occasional misincorporation of RNA residues in DNA in contemporary biology where the monomers ((deoxy)nucleoside triphosphates) are structurally analogous.

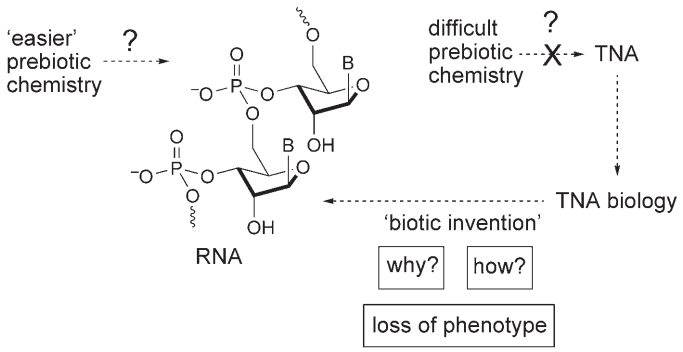

Fig. 17. *RNA as a prebiotic product.* An alternative scenario to that presented in *Fig. 1*, and one which cannot be ruled out until the complete constitutional self-assembly chemical space of RNA has been exhaustively explored.

Work in our laboratory has been generously supported by the *Engineering and Physical Sciences Research Council* through the provision of postdoctoral fellowships to *C. A.* and *A. L. P.*, and studentships to *F. F. B.*, *M. A. C.*, *M. W. P.*, and *J. M. S.*

REFERENCES

[1] G. F. Joyce, L. E. Orgel, 'Prospects for Understanding the Origin of the RNA World', in 'The RNA World', 2nd edn., Eds. R. F. Gesteland, T. R. Cech, J. F. Atkins, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1999.
[2] G. F. Joyce, *Nature* **2002**, *418*, 214.
[3] G. F. Joyce, A. W. Schwartz, S. L. Miller, L. E. Orgel, *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4398.
[4] K.-U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, A. Eschenmoser, *Science* **2000**, *290*, 1347.
[5] K.-U. Schöning, P. Scholz, X. Wu, S. Guntha, G. Delgado, R. Krishnamurthy, A. Eschenmoser, *Helv. Chim. Acta* **2002**, *85*, 4111; J. C. Chaput, J. K. Ichida, J. W. Szostak, *J. Am. Chem. Soc.* **2003**, *125*, 856; J. C. Chaput, J. W. Szostak, *J. Am. Chem. Soc.* **2003**, *125*, 9274.
[6] L. Zhang, A. Peritz, E. Meggers, *J. Am. Chem. Soc.* **2005**, *127*, 4174.
[7] R. Shapiro, *Orig. Life Evol. Biosphere* **1988**, *18*, 71; T. Mizuno, A. H. Weiss, *Adv. Carbohydr. Chem. Biochem.* **1974**, *29*, 173; S. Pizzarello, A. L. Weber, *Science* **2004**, *303*, 1151; A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner, *Science* **2004**, *303*, 196; J. Kofoed, J.-L. Reymond, T. Darbre, *Org. Biomol. Chem.* **2005**, *3*, 1850.
[8] J. Oro, A. P. Kimball, *Arch. Biochem. Biophys.* **1962**, *96*, 293.
[9] J. P. Ferris, W. J. Hagan Jr., *Tetrahedron* **1984**, *40*, 1093; R. A. Sanchez, J. P. Ferris, L. E. Orgel, *J. Mol. Biol.* **1967**, *30*, 223; R. F. Shuman, W. E. Shearin, R. J. Tull, *J. Org. Chem.* **1979**, *44*, 4532; R. A. Sanchez, J. P. Ferris, L. E. Orgel, *J. Mol. Biol.* **1968**, *38*, 121.
[10] J. P. Ferris, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1968**, *33*, 693; M. P. Robertson, S. L. Miller, *Nature* **1995**, *375*, 772, and correction, *377*, 257.
[11] R. Larralde, M. P. Robertson, S. L. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8158.
[12] G. Springsteen, G. F. Joyce, *J. Am. Chem. Soc.* **2004**, *126*, 9578.
[13] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1972**, *67*, 25; W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Evol.* **1972**, *1*, 249.
[14] R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1970**, *47*, 531.
[15] C. M. Tapiero, J. Nagyvary, *Nature* **1971**, *231*, 42.
[16] A.-A. Ingar, R. W. A. Luke, B. R. Hayter, J. D. Sutherland, *ChemBioChem* **2003**, *4*, 504.

[17] C. Anastasi, F. F. Buchet, M. A. Crowe, J. D. Sutherland, in preparation.
[18] S. Pitsch, S. Wendeborn, B. Jaun, A. Eschenmoser, *Helv. Chim. Acta* **1993**, *76*, 2161.
[19] S. Pitsch, S. Wendeborn, R. Krishnamurthy, A. Holzner, M. Minton, M. Bolli, C. Miculca, N. Windhab, R. Micura, M. Stanek, B. Jaun, A. Eschenmoser, *Helv. Chim. Acta* **2003**, *86*, 4270.
[20] S. Pitsch, E. Pombo-Villar, A. Eschenmoser, *Helv. Chim. Acta* **1994**, *77*, 2251.
[21] J. M. Smith, J. D. Sutherland, *ChemBioChem* **2005**, *6*, 1980.
[22] J. M. Smith, V. Borsenberger, J. Raftery, J. D. Sutherland, *Chem. Biodiv.* **2004**, *1*, 1418.
[23] D. Müller, S. Pitsch, A. Kittaka, E. Wagner, C. E. Wintner, A. Eschenmoser, *Helv. Chim. Acta* **1990**, *73*, 1410; S. D. Cook, J. D. Sutherland, *Tetrahedron Lett.* **1996**, *37*, 5779.
[24] J. A. Gerlt, N. I. Gutterson, P. Datta, B. Belleau, C. L. Penney, *J. Am. Chem. Soc.* **1980**, *102*, 1655; F. J. Marsh, P. Weiner, J. E. Douglas, P. A. Kollman, G. L. Kenyon, J. A. Gerlt, *J. Am. Chem. Soc.* **1980**, *102*, 1660.
[25] V. Borsenberger, M. A. Crowe, J. Lehbauer, J. Raftery, M. Helliwell, K. Bhutia, T. Cox, J. D. Sutherland, *Chem. Biodiv.* **2004**, *1*, 203.
[26] N. Saewan, M. A. Crowe, M. Helliwell, J. Raftery, K. Chantrapromma, J. D. Sutherland, *Chem. Biodiv.* **2005**, *2*, 66.
[27] H. R. Khan, G. Crank, *Tetrahedron Lett.* **1987**, *28*, 3381.
[28] C. Anastasi, M. A. Crowe, M. W. Powner, J. D. Sutherland, *Angew. Chem., Int. Ed.* **2006**, *45*, 6176.
[29] A. F. Cockerill, A. Deacon, R. G. Harrison, D. J. Osborne, D. M. Prime, W. J. Ross, A. Todd, J. P. Verge, *Synthesis* **1976**, 591.
[30] S. L. Widicus-Weaver, *Astrophys. J.* **2005**, *624*, L33; G. Cooper, N. Kimmich, W. Belisle, J. Sarinana, K. Brabham, L. Garrel, *Nature* **2001**, *414*, 879.
[31] A. Hall, J. R. Knowles, *Biochemistry* **1975**, *14*, 4348; J. P. Richard, *J. Am. Chem. Soc.* **1984**, *106*, 4926.
[32] C. Anastasi, M. A. Crowe, J. D. Sutherland, *J. Am. Chem. Soc.* **2007**, *129*, 24.
[33] R. Krishnamurthy, S. Guntha, A. Eschenmoser, *Angew. Chem., Int. Ed.* **2000**, *39*, 2281.
[34] W. Huang, J. P. Ferris, *J. Am. Chem. Soc.* **2006**, *128*, 8914; A. Kanavarioti, E. E. Baird, T. B. Hurley, J. A. Carruthers, S. Gangopadhyay, *J. Org. Chem.* **1999**, *64*, 8323; I. A. Kozlov, L. E. Orgel, *Helv. Chim. Acta* **1999**, *82*, 1799; A. Kanavarioti, C. F. Bernasconi, E. E. Baird, *J. Am. Chem. Soc.* **1998**, *120*, 8575; J. P. Ferris, A. R. Hill Jr., R. Liu, L. E. Orgel, *Nature* **1996**, *381*, 59; G. F. Joyce, G. M. Visser, C. A. A. van Boeckel, J. H. van Boom, L. E. Orgel, J. van Westrenen, *Nature* **1984**, *310*, 602.
[35] M. S. Verlander, R. Lohrmann, L. E. Orgel, *J. Mol. Evol.* **1973**, *2*, 303; M. S. Verlander, L. E. Orgel, *J. Mol. Evol.* **1974**, *3*, 115; M. Renz, R. Lohrmann, L. E. Orgel, *Biochim. Biophys. Acta* **1971**, *240*, 463; D. A. Usher, A. H. McHale, *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 1149; M. Bolli, R. Micura, S. Pitsch, A. Eschenmoser, *Helv. Chim. Acta* **1997**, *80*, 1901.
[36] L. E. Orgel, *Science*, **2000**, *290*, 1306.
[37] M. A. Crowe, J. D. Sutherland, *ChemBioChem* **2006**, *7*, 951.
[38] P. A. Giannaris, M. J. Damha, *Nucleic Acids Res.* **1993**, *21*, 4742; M. Wasner, D. Arion, G. Borkow, A. Noronha, A. H. Uddin, M. A. Parniak, M. J. Damha, *Biochemistry* **1998**, *37*, 7478; H. Urata, H. Shimizu, H. Hiroaki, D. Kohda, M. Akagi, *Biochim. Biophys. Res. Commun.* **2003**, *309*, 79.
[39] R. Lohrmann, L. E. Orgel, *Science* **1971**, *171*, 490; R. Saffhill, *J. Org. Chem.* **1970**, *35*, 2881.
[40] P. S. Pallan, C. J. Wilds, Z. Wawrzak, R. Krishnamurthy, A. Eschenmoser, M. Egli, *Angew. Chem., Int. Ed.* **2003**, *42*, 5893.

Received August 21, 2006