Prebiotic Chemistry

DOI: 10.1002/ange.200601267

Direct Assembly of Nucleoside Precursors from Two- and Three-Carbon Units**

*Carole Anastasi, Michael A. Crowe, Matthew W. Powner, and John D. Sutherland**

The efficient synthesis of ribose, the sugar component of RNA, has long been a major goal of research aimed at investigating the origin of life. In strongly alkaline solution, glycolaldehyde and formaldehyde participate in the formose reaction—a series of aldol reactions and rearrangements—giving a range of sugars, including tetroses, pentoses, and hexoses.[1] However, aldose–ketose equilibration[2] and lack of stereocontrol in the formose reaction result in the yield of ribose being extremely low (< 1 %).[3] In an elegant series of experiments, Eschenmoser and co-workers showed that the alkaline reaction of glycolaldehyde phosphate and formaldehyde produced a much narrower range of products, with ribose-2,4-diphosphate predominating (10–20 %).[4] However, the high pH value required for this process in solution and the phosphorylation pattern of the products do not support its intermediacy in the production of RNA, since RNA is base-labile and contains 5′→3′ phosphate diester linkages. More recently, amino acids have been shown to catalyze the synthesis of sugars in water,[5] with Darbre and co-workers showing that the zinc–proline complex catalyzes the aldolization of glycolaldehyde and glyceraldehyde to give ribose (19 %),[5c] but as a component of a complex mixture containing tetroses, the other pentoses, and aldo- and ketohexoses.

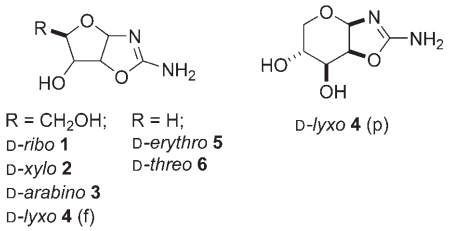
The instability of ribose ($t_{1/2} < 3$ h, pH ≈ 10, 55 °C) is another concern,[6,7] and one which prompted Benner and co-workers to consider complexation of the sugar with borate as a means of achieving stabilization.[8] It was shown that pentoses could be synthesized from glycolaldehyde and glyceraldehyde in the presence of borate, and that, once formed, the pentoses were significantly stabilized by complexation. However, this reaction again requires a high pH value and displays no selectivity for ribose over the other pentoses. Furthermore, although the complexation with borate is stabilizing, it is also likely to slow, or prevent,

[*] Dr. C. Anastasi, M. A. Crowe, M. W. Powner, Prof. Dr. J. D. Sutherland
School of Chemistry
The University of Manchester
Oxford Road, Manchester M13 9PL (UK)
Fax: (+44) 161-275-4939
E-mail: john.sutherland@manchester.ac.uk

[**] This work was carried out as part of the EU COST action D27 "Prebiotic Chemistry and Early Evolution", and was funded by the Engineering and Physical Sciences Research Council through the provision of a postdoctoral fellowship to C.A. and studentships to M.W.P. and M.A.C. We thank Dr. A. Parkes and Dr. M. Helliwell for preliminary experiments.

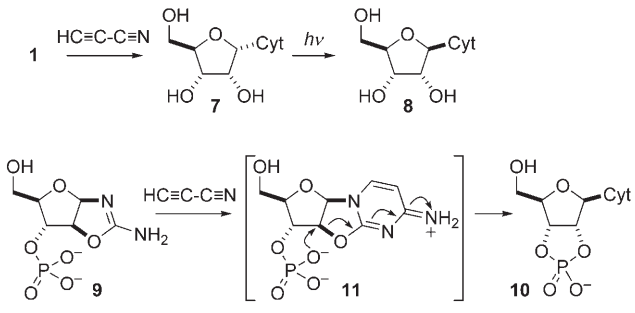
Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

elaboration of the ribose into ribonucleosides by addition of a nucleobase. Direct addition of a nucleobase to the free sugar is extremely inefficient with purines and is not possible with pyrimidines, but a stepwise assembly of pyrimidine nucleobases on a sugar scaffold has been demonstrated.[9–11] An ideal situation would therefore be one in which ribose could be stabilized as a derivative that could also function as an intermediate en route to a ribonucleoside, and in this regard the reaction of ribose with cyanamide has been investigated. Ribose, the other pentose sugars, and the tetroses all form aminooxazoline derivatives with cyanamide (Scheme 1 and



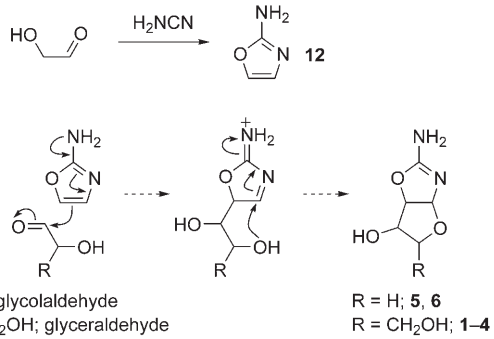
Scheme 1. Tetrose and pentose aminooxazolines. The D-tetroses, D-ribose, D-arabinose, and D-xylose all form furanose aminooxazolines.[7,9,12,13] D-Lyxose aminooxazoline displays anomalous behavior and exists in equilibrium between furanose (f) and pyranose (p) forms (**4**(p)/**4**(f), ca. 5:1).[13]

see the Supporting Information).[9,12,13] Springsteen and Joyce have shown that the D-ribose aminooxazoline derivative **1** degrades 70-fold more slowly than the free sugar ($t_{1/2}>$ 1 week, pH ≈ 10, 55 °C),[7] and therefore it seems reasonable to assume that **2**–**6** will also be more stable than their parent sugars. Additionally, these aminooxazolines react with cyanoacetylene to give cytidine nucleosides;[9,11–13] for example, **1** gives α-D-cytidine **7** (Scheme 2). However, RNA is composed of β-D-ribonucleotides, so stereochemical inversion at the anomeric center of **7** is subsequently required. Photoanomerization of **7** to β-D-cytidine **8** is a relatively inefficient reaction,[9] so we explored an alternative pathway involving **9**, the 3′-phosphate derivative of the arabinose aminooxazoline



Scheme 2. Conversion of aminooxazolines into cytidine nucleosides. Reaction of **1** with cyanoacetylene, followed by irradiation, gives β-D-cytidine **8**. Reaction of **9** (the 3′-phosphate of **3**) with cyanoacetylene gives the activated ribonucleotide **10** by intramolecular nucleophilic substitution of the intermediate **11**. These conversions establish **1** and **3** as being more advanced RNA precursors than free ribose. Cyt = cytosine.

3.[11] Reaction of **9** with cyanoacetylene gave the activated ribonucleotide **10** via the anhydronucleotide **11** (Scheme 2).[10] Their stability and subsequent conversion into nucleosides, therefore, suggest a central role for the aminooxazolines **1** and **3** in the prebiotic synthesis of RNA. Since **1** and **3** have only ever been synthesized from their parent pentose sugars, it has been necessary to invoke a somewhat unlikely assembly sequence in which the pentoses are exposed to cyanamide after their generation, but before their degradation.[7,14] To get round this problem, we wondered if it might be possible to bypass the free pentoses and assemble **1** and **3** directly from C_2 and C_3 units. It is known that reaction of cyanamide with glycolaldehyde gives 2-aminooxazole **12**,[16] and we recognized that if this compound contained a sufficiently nucleophilic C atom to react with an α-hydroxyaldehyde[17] the resultant intermediate might cyclize to an aminooxazoline (Scheme 3). Accordingly, we



Scheme 3. Synthesis of 2-aminooxazole **12** from prebiotic starting materials and the potential involvement of **12** in tetrose and pentose aminooxazoline formation.

investigated the reaction of **12** with α-hydroxyaldehydes, deciding, for the sake of analytical simplicity, to first study the potential $C_2 + C_2$ assembly of **5** and/or **6**. A neutral, aqueous solution of **12** and glycolaldehyde in a 1:1 stoichiometry was incubated at 40 °C for 12 h before being lyophilized. The resulting residue was dissolved in D_2O and analyzed by ^{1}H NMR spectroscopy (500 MHz, see the Supporting Information). This showed conversion of the starting materials into a mixture of *rac*-**5** and *rac*-**6** in approximately 90 % yield (*rac*-**5**/*rac*-**6** 46:43). The reaction was extremely clean and the high yield of C_4 products, uncontaminated by C_6 products, suggests that glycolaldehyde homoaldolization and subsequent addition of **12** to the resultant free tetroses do not occur. We next investigated the reaction of **12** with *rac*-glyceraldehyde hoping to uncover an efficient $C_2 + C_3$ route to *rac*-**1** and/or *rac*-**3**. On the basis of the reaction of **12** and glycolaldehyde, we also expected to see a slight selectivity for the *erythro* products in the reaction with glyceraldehyde, but it was not clear whether there would be any facial selectivity in the glyceraldehyde component. Under the same mild conditions, the reaction again proceeded extremely cleanly, with the products *rac*-**1** to *rac*-**4** obtained in approximately 95 % overall yield, as determined by ^{1}H NMR analysis, with *rac*-**1** and *rac*-**3** predominant (*rac*-**1**/*rac*-**3**/*rac*-**4**/*rac*-**2** 44:30:13:8)

(Figure 1 and see the Supporting Information).[19] The high yield of the *ribo* and *arabino* stereoisomers is remarkable given the outcome of previous $C_2 + C_3$ aldolization routes to pentoses (see the Supporting Information). We think that the stereochemical preference for *rac*-**1** and *rac*-**3** is due to control of facial selectivity of the *rac*-glyceraldehyde component through formation of a hydrogen-bonded chelate, along with the obligatory *Z* geometry and low facial selectivity of the nucleophilic C=C bond of **12** (see the Supporting Information).

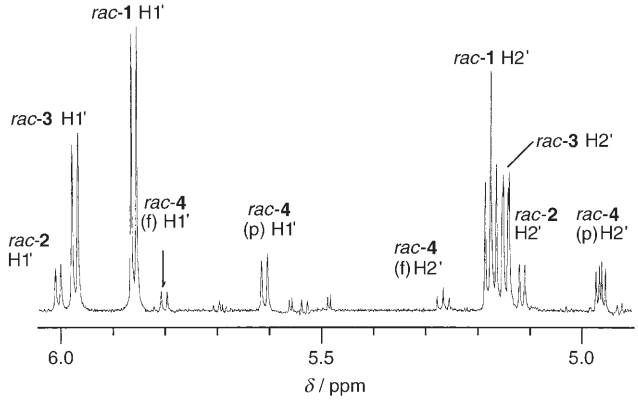


Figure 1. Stereoselectivity in the reaction of **12** with *rac*-glyceraldehyde. Partial ^{1}H NMR spectrum (500 MHz, D_2O) of the crude products from the reaction of **12** with *rac*-glyceraldehyde showing the predominance of *rac*-**1** and *rac*-**3**.

It was noted in their initial synthesis that compound **3** is more soluble than **1** in water ,[9] and it was found that **1** is the least soluble of all the pentose aminooxazolines **1**–**4**, thus permitting its isolation from mixtures by direct crystallization.[7] Since *rac*-**1** and *rac*-**3** are both produced in high yield in our reaction, and since it is not clear which is the more important ribonucleoside precursor, we sought a means of separating them that is prebiotically plausible. We realized that crystallization of *rac*-**1** from the crude products of the reaction of **12** and *rac*-glyceraldehyde would not only furnish a pure sample of *rac*-**1**, but would also change the composition of the supernatant such that *rac*-**3** might now be the predominant solute. To test this idea, we carried out a reaction between **12** and *rac*-glyceraldehyde using the same 1:1 stoichiometry as before, but at higher concentrations. After incubation of the reaction at 40 °C for 12 h, it was cooled to 4 °C whereupon crystals formed. The crystals were separated from the supernatant and both were examined by ^{1}H NMR spectroscopy. As expected, the crystals proved to be pure *rac*-**1**, and from weighing these crystals, it was determined that approximately half the total amount of *rac*-**1** expected from the reaction had crystallized. Thus, the amount of *rac*-**1** in the supernatant was correspondingly reduced and the major product in the supernatant was now *rac*-**3** (*rac*-**3**/*rac*-**1**/*rac*-**4**/*rac*-**2** 28:25:11:9). In this way, starting from a solution containing only **12** and *rac*-glyceraldehyde, *rac*-**1** can be produced in pure form as a solid, and *rac*-**3** as the predominant species in solution, simply by the warming and cooling that might have occurred during a day/night cycle on early Earth.

Ideally, crystalline enantiopure **1**, rather than *rac*-**1**, should be produced by this approach, and on the face of it, this would require the starting glyceraldehyde to be enantiopure. However, the incomplete crystallization of *rac*-**1** from solution suggested that it might be possible to obtain enantiopure **1** from scalemic (sca-) glyceraldehyde if enantiopure and racemic **1** had different solubilities. The space group of both the enantiopure and racemic **1** is $P2_12_12_1$, and the racemic material has been shown to form enantiomorphously twinned crystals.[7,12] Hexahelicene crystals belong to the same space group and racemic (and near-racemic) material also forms enantiomorphously twinned crystals, but above a threshold *ee* value, enantiopure crystals may be formed.[20] We considered that **1** might show similar behavior, in which case the formation of crystalline enantiopure **1** from sca-glyceraldehyde might be possible. To test this supposition, we prepared solutions of D-, L- and sca-**1** with a range of *ee* values, and allowed partial crystallization to take place. Crystals from each sample were then dissolved in a phosphate buffer at pH 4.5, and the optical rotation of the resultant solutions determined. To our delight, enantiopure **1** crystallizes from solutions of *rac*-**1** with $\geq 60\,\%$ *ee* (Figure 2). We

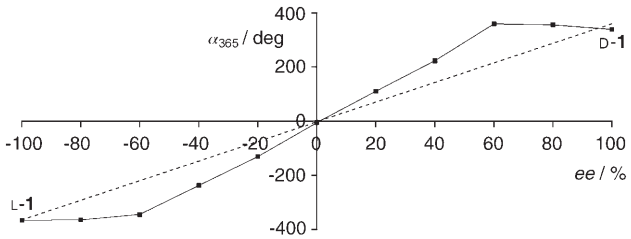


Figure 2. Asymmetric amplification in the crystallization of **1** from solutions of scalemic material. Crystals obtained from solutions of **1** with varying *ee* values (abscissa) were re-dissolved in a pH 4.5 buffer, and optical rotations (α) were measured (ordinate). The dotted line indicates the rotation that would have been expected if the crystals had the same *ee* value as the starting solutions.

then carried out a reaction between 2-aminooxazole **12** and D-glyceraldehyde with 60 % *ee* and obtained crystals of optically pure D-**1**. This amplification of optical asymmetry during the formation of a nucleoside precursor has potential relevance to the origin of biomolecular homochirality.

Since **12** is also formed from glycolaldehyde and cyanamide under mild conditions, we are now investigating the chemistry of glycolaldehyde, glyceraldehyde, and cyanamide mixtures with a view to establishing conditions for the direct synthesis of *rac*-**1** to *rac*-**6**. We are also investigating the stereochemical consequences of carrying out reactions of **12** with *rac*-glyceraldehyde in the presence of chiral additives.

Received: March 31, 2006
Revised: May 31, 2006
Published online: August 17, 2006

Keywords: carbohydrates · nucleobases · nucleosides · prebiotic chemistry · ribose

[1] T. Mizuno, A. H. Weiss, *Adv. Carbohydr. Chem. Biochem.* **1974**, *29*, 173.
[2] J. C. Speck, Jr., *Adv. Carbohydr. Chem. Biochem.* **1958**, *13*, 63.
[3] R. Shapiro, *Origins Life Evol. Biospheres* **1988**, *18*, 71.
[4] D. Müller, S. Pitsch, A. Kittaka, E. Wagner, C. E. Wintner, A. Eschenmoser, *Helv. Chim. Acta* **1990**, *73*, 1410.
[5] a) S. Pizzarello, A. L. Weber, *Science* **2004**, *303*, 1151; b) A. Córdova, I. Ibrahem, J. Casas, H. Sundén, M. Engqvist, E. Reyes, *Chem. Eur. J.* **2005**, *11*, 4772; c) J. Kofoed, J.-L. Reymond, T. Darbre, *Org. Biomol. Chem.* **2005**, *3*, 1850.
[6] R. Larralde, M. P. Robertson, S. L. Miller, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 8158.
[7] G. Springsteen, G. F. Joyce, *J. Am. Chem. Soc.* **2004**, *126*, 9578.
[8] A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner, *Science* **2004**, *303*, 196.
[9] R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1970**, *47*, 531.
[10] C. M. Tapiero, J. Nagyvary, *Nature* **1971**, *231*, 42.
[11] A.-A. Ingar, R. W. A. Luke, B. R. Hayter, J. D. Sutherland, *ChemBioChem* **2003**, *4*, 504.
[12] V. Borsenberger, M. A. Crowe, J. Lehbauer, J. Raftery, M. Helliwell, K. Bhutia, T. Cox, J. D. Sutherland, *Chem. Biodiversity* **2004**, *1*, 203.
[13] N. Saewan, M. A. Crowe, M. Helliwell, J. Raftery, K. Chantrapromma, J. D. Sutherland, *Chem. Biodiversity* **2005**, *2*, 66.
[14] The temporal or spatial separation of the oxygenous chemistry (required for sugar synthesis) from the nitrogenous chemistry (required for nucleobase assembly) has previously been assumed, because of the "drastically enhanced chemical complications" expected for a scenario without that separation.[15]
[15] K.-U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, A. Eschenmoser, *Science* **2000**, *290*, 1347.
[16] 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.
[17] By way of partial precedent, 2-amino-4-methyloxazole has been shown to react as a C-based nucleophile with 4-methoxybenzaldehyde in hot toluene.[18] However, in this case the intermediate lacks a suitably positioned hydroxy group and so cannot cyclize to an aminooxazoline. Instead, the intermediate re-aromatizes giving 2-amino-5-[1-hydroxy-1-(4-methoxyphenyl)methyl]-4-methyloxazole.
[18] H. R. Khan, G. Crank, *Tetrahedron Lett.* **1987**, *28*, 3381.
[19] The identity of the products was confirmed by spiking with authentic samples of **1**–**4** (for example, see the Supporting Information).
[20] B. S. Green, M. Knossow, *Science* **1981**, *214*, 795.

