LETTERS

Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions

Matthew W. Powner[1], Béatrice Gerland[1] & John D. Sutherland[1]

At some stage in the origin of life, an informational polymer must have arisen by purely chemical means. According to one version of the 'RNA world' hypothesis[1–3] this polymer was RNA, but attempts to provide experimental support for this have failed[4,5]. In particular, although there has been some success demonstrating that 'activated' ribonucleotides can polymerize to form RNA[6,7], it is far from obvious how such ribonucleotides could have formed from their constituent parts (ribose and nucleobases). Ribose is difficult to form selectively[8,9], and the addition of nucleobases to ribose is inefficient in the case of purines[10] and does not occur at all in the case of the canonical pyrimidines[11]. Here we show that activated pyrimidine ribonucleotides can be formed in a short sequence that bypasses free ribose and the nucleobases, and instead proceeds through arabinose amino-oxazoline and anhydronucleoside intermediates. The starting materials for the synthesis—cyanamide, cyanoacetylene, glycolaldehyde, glyceraldehyde and inorganic phosphate—are plausible prebiotic feedstock molecules[12–15], and the conditions of the synthesis are consistent with potential early-Earth geochemical models. Although inorganic phosphate is only incorporated into the nucleotides at a late stage of the sequence, its presence from the start is essential as it controls three reactions in the earlier stages by acting as a general acid/base catalyst, a nucleophilic catalyst, a pH buffer and a chemical buffer. For prebiotic reaction sequences, our results highlight the importance of working with mixed chemical systems in which reactants for a particular reaction step can also control other steps.

Because they comprise phosphate, ribose and nucleobases, it is tempting to assume that ribonucleotides must have prebiotically assembled from such building blocks. Thus, for example, it has previously been supposed that the activated ribonucleotide β-ribocytidine-2′,3′-cyclic phosphate **1** must have been produced by phosphorylation of the ribonucleoside **2**, with the latter deriving from the conjoining of the free pyrimidine nucleobase cytosine **3** and the furanose form of ribose **4** (Fig. 1, blue arrows). This mode of assembly is seemingly supported by the facts that cytosine **3** can be synthesized by condensation of cyanoacetaldehyde **5** and urea **6**[16] (the hydration products of cyanoacetylene **7**[17], and cyanamide **8**[18], respectively) and pentoses including ribose can be produced by aldol reaction of glyceraldehyde **9** and glycolaldehyde **10**[8,9]. The insuperable problem with this approach, however, is that one of the presumed steps, the condensation of ribose **4** and cytosine **3**, does not work[11]. The reasons for this are both kinetic (the N1 lone pair of **3** is unavailable owing to delocalization) and, in water, thermodynamic (the equilibrium constant is such that hydrolysis of **2** to **3** and **4** is favoured over condensation). The same is true for ribosylation of uracil, which has also not been demonstrated.

We have considered a large number of alternative ribonucleotide assembly modes, including those that extend back to the same small-molecule precursors as the traditionally assumed route described above[19]. By systematic experimental investigation of these options, we have discovered a short, highly efficient route to activated pyrimidine ribonucleotides from these same precursors that proceeds by way of alternative intermediates (Fig. 1, green arrows). By contrast with previously investigated routes to ribonucleotides, ours bypasses ribose and the free pyrimidine nucleobases. Mixed nitrogenous–oxygenous chemistry first results in the reaction of cyanamide **8** and glycolaldehyde **10**, giving 2-amino-oxazole **11**, and this heterocycle then adds to glyceraldehyde **9** to give the pentose amino-oxazolines including the arabinose derivative **12**. Reaction of **12** with cyanoacetylene **7** then gives the anhydroarabinonucleoside **13**, which subsequently undergoes phosphorylation with rearrangement to furnish β-ribocytidine-2′,3′-cyclic phosphate **1**. In a subsequent photochemical step, **1** is partly converted to the corresponding uracil derivative, and synthetic co-products are largely destroyed.

We had previously shown that in unbuffered aqueous solution, 2-amino-oxazole **11** adds to glyceraldehyde **9** to give the pentose amino-oxazolines including **12** in excellent overall yield[20]. Our starting point in the present work was therefore to find a prebiotically

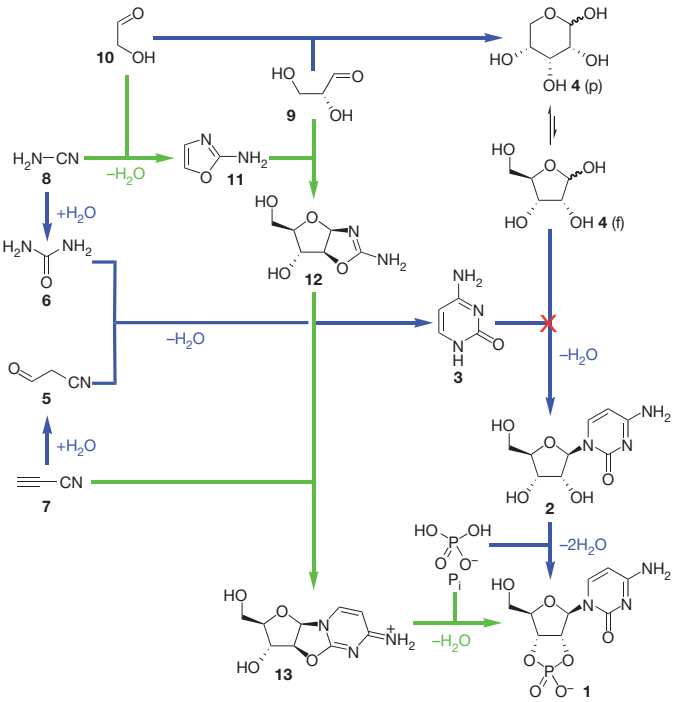

Figure 1 | Pyrimidine ribonucleotide assembly options. Previously assumed synthesis of β-ribocytidine-2′,3′-cyclic phosphate **1** (blue; note the failure of the step in which cytosine **3** and ribose **4** are proposed to condense together) and the successful new synthesis described here (green). p, pyranose; f, furanose.

[1]School of Chemistry, The University of Manchester, Oxford Road, Manchester M13 9PL, UK.

plausible synthesis of **11**. Constitutionally, **11** is the condensation product of **8** and **10**, and although there exists, in the conventional chemical literature, a procedure to bring about this condensation, it requires strongly alkaline conditions[21]. Because we wanted to generate **11**, and then allow it to react with **9**, which is unstable to alkali, under the same conditions, neutral-pH reaction conditions had to be found.

We initially investigated the reaction with **8** and **10** in a 1:1 ratio starting at neutral pH in unbuffered aqueous solution. Only a small amount of **11** was produced under these conditions, and ^{1}H NMR spectra were indicative of the formation of a variety of carbonyl addition adducts and other intermediates, for example **14**–**18** (Fig. 2a, b). The carbonyl addition adducts **14** were presumably formed reversibly, and so did not represent material irretrievably committed to other products, but rather intermediates stalled en route to **11**. At low concentrations of hydroxide, it appeared that two additional types of reaction needed to make **11** were very sluggish: intra-adduct attack of the glycolaldehyde-derived hydroxyl group on the cyanamide-derived nitrile carbon (for example **14** ($n = 0$) → **15**), and C–H deprotonation leading to aromatization (**17** → **11**).

Denied the opportunity of using hydroxide as a specific base catalyst to accelerate these slow steps, we sought a general base catalyst that could provide the same acceleration, but at neutral pH. Inorganic phosphate seemed to be ideal in this regard because its second pK_a value is close to neutrality. Furthermore, as phosphate is ultimately needed in some form to make activated nucleotides, we decided to include it from the start of the assembly sequence. We repeated the earlier reaction of cyanamide **8** and glycolaldehyde **10**, but in the presence of 1 M phosphate buffer at pH 7.0. ^{1}H NMR analysis revealed that 2-amino-oxazole **11** was produced in >80% yield (75% isolated yield) (Fig. 2c). With an excess of **8** over **10**, the synthesis of **11** still takes place in the presence of phosphate, but is followed by slower phosphate addition to residual **8** giving the intermediate adduct **19**, which partitions to urea **6** and cyanoguanidine **20** (Fig. 2d).

We then investigated whether the subsequent reaction of **11** with glyceraldehyde **9** would be tolerant to the residual presence of phosphate. In the absence of phosphate, the ribose and arabinose amino-oxazolines **21** and **12** are the major products, and the xylose derivative **22** is a minor product (Fig. 3a)[20]. The lyxose amino-oxazoline **23** is formed in intermediate amounts as an equilibrating mixture of pyranose and furanose isomers. All of the pentose amino-oxazolines have the potential to be converted reversibly into one or other of the 5-substituted 2-amino-oxazoles **24** and **25** by phosphate catalysis (by chemistry similar to that underlying the conversion of **16** to **11**), but to differing extents depending on their stability. After one day in the presence of phosphate, all of the amino-oxazolines showed some conversion to the corresponding 5-substituted 2-amino-oxazole (**24** or **25**), but the lyxose amino-oxazoline **23** proved the least stable and underwent the greatest conversion (Fig. 3b). We then took a crude sample of **11** that had just been prepared from cyanamide **8** and glycolaldehyde **10** in the presence of phosphate, and added glyceraldehyde **9** to it. After overnight incubation, ^{1}H NMR analysis revealed that although all four amino-oxazolines were still formed, the lyxose derivative **23** was selectively depleted and was now a minor product along with the xylose derivative **22** (Fig. 3c). With two of its stereoisomeric relatives now minor products, the path from the arabinose amino-oxazoline **12** to ribonucleotides looked clearer. Selective crystallization of ribose amino-oxazoline **21** offers a further means of enriching **12** such that it becomes the major product in solution[20,22].

We then proceeded to the second stage of pyrimidine nucleobase assembly. Although our focus was on the chemistry of the key arabinose amino-oxazoline **12**, the corresponding chemistry of the ribose amino-oxazoline **21** was also studied (Supplementary Information). It had earlier been shown that in unbuffered aqueous solution, **12** reacts with an excess of cyanoacetylene **7** giving β-arabinocytidine **26**, (Fig. 4a)[23]. The yield of **26** was relatively low, however, and we used ^{1}H NMR analysis to determine why. It transpires that the pH rises during the course of the reaction, resulting in hydrolysis of anhydronucleoside

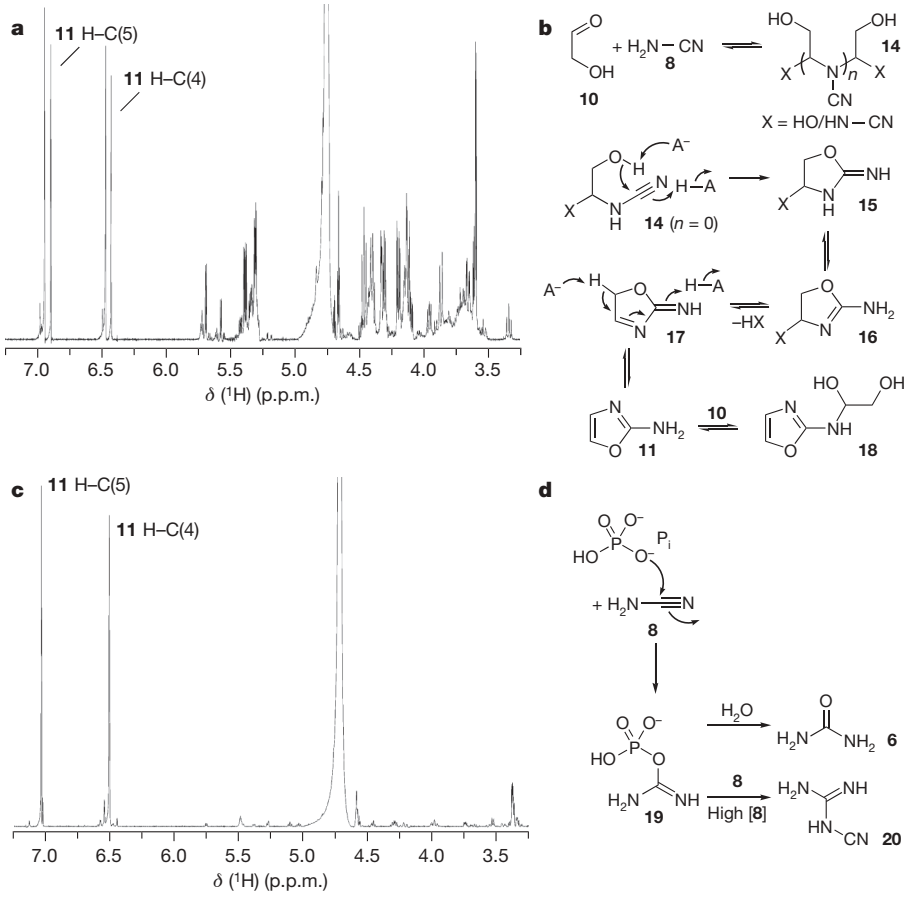

Figure 2 | Development of the synthesis of 2-amino-oxazole 11. a, ^{1}H NMR spectrum of the products of reaction of cyanamide **8** and glycolaldehyde **10** in the absence of phosphate. δ, chemical shift. **b**, Mechanism for the condensation of **8** and **10**. Curved arrows depict base-catalysed steps thought to be rate limiting in the absence of phosphate. **c**, ^{1}H NMR spectrum of the products of reaction of **8** and **10** in the presence of phosphate. **d**, Slower side reaction between **8** and phosphate. P_i, inorganic phosphate; H–A, general acid; A^-, general base.

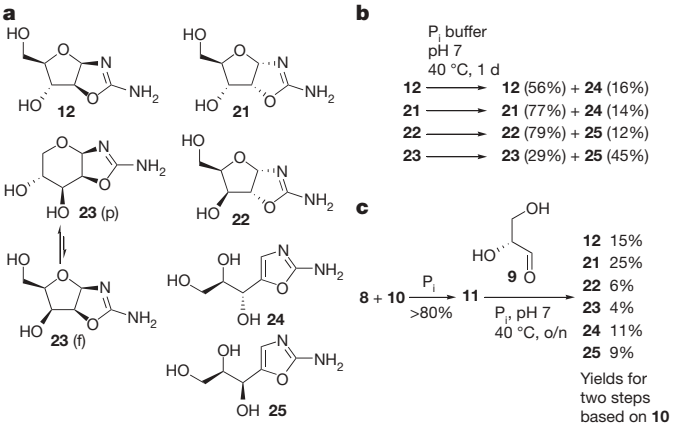

Figure 3 | Pentose amino-oxazoline stability, and assembly chemistry. **a**, Structures of the arabinose (**12**), ribose (**21**), xylose (**22**) and lyxose (**23**) amino-oxazolines and their elimination products **24** and **25**. **b**, Relative stabilities of the amino-oxazolines in the presence of phosphate. **c**, Formation of amino-oxazolines by addition of glyceraldehyde **9** to a solution of 2-amino-oxazole **11**, with the latter freshly formed *in situ* from cyanamide **8** and glycolaldehyde **10**. P_i, inorganic phosphate; o/n, overnight.

intermediates and causing hydroxyl groups to undergo reaction with cyanoacetylene **7** (Supplementary Information). To prevent the rise in pH during the reaction, inorganic phosphate was added as a buffer. When the buffering pH was 6.5, the reactions were extremely clean, with little evidence for anhydronucleoside hydrolysis. Furthermore, excess cyanoacetylene **7** that did not evaporate underwent reaction with phosphate at this pH, giving cyanovinyl phosphate **27** instead of cyanovinylating hydroxyl groups. Using phosphate as a dual-function pH and chemical buffer in this way, the arabinose anhydronucleoside **13** could be produced in extremely high yield from **12**.

Our finding that the reaction of the amino-oxazoline **12** with cyanoacetylene **7** could be controlled, by the pH and chemical buffering action of phosphate, to produce the arabinose anhydronucleoside **13** in excellent yield opened up the possibility of a combined phosphorylation–rearrangement[24,25] reaction to convert **13** to the activated ribonucleotide **1**. Furthermore, the formation of cyanovinyl phosphate **27**, as a co-product in the nucleobase assembly process, extended the range of potential phosphorylating agents for such a process because, in aqueous solution, **27** undergoes reaction with inorganic phosphate to give pyrophosphate[17]. Accordingly, we investigated the phosphorylation of the anhydronucleoside **13** using both inorganic phosphate and pyrophosphate.

Prebiotic phosphorylation of nucleosides has been demonstrated by heating either in the dry state with urea[26] or in formamide solution[27]. We were particularly attracted by the possibility of using urea **6** in the phosphorylation of **13** because it is a co-product of the chemical system in which 2-amino-oxazole **11** is produced from glycolaldehyde **10** and cyanamide **8**, if the latter is initially present in excess (Fig. 2d). After preliminary experiments, and through consideration of the phosphorylation mechanism (Supplementary Information), we found that when **13** was heated with 0.5 equiv. of pyrophosphate in urea containing ammonium salts, **1** was formed as the major product in addition to **28** and **29** (the 5′-phosphate derivatives of **13** and **1**, respectively) and small amounts of the hydrolysis product β-arabinocytidine **26** and its 5′-phosphate derivative **30** (Fig. 4b, procedure A; Supplementary Information). Alternatively, **1** was formed in very good yield—along with **29**, the hydrolysis products **26** and **30**, and the nucleobases cytosine **3** and diaminopyrimidine **31**—by heating **13** with inorganic phosphate and urea in formamide solution (Fig. 4b, procedure B; Supplementary Information).

The conversion to **1** in both cases is thought to involve phosphorylation of the 3′-hydroxyl group of **13** to give the 3′-phosphate **32**, which can undergo rearrangement, through intramolecular nucleophilic substitution (Fig. 4c)—a reaction not possible in the *ribo*-analogue because of the *cis* relationship of the 2′- and 3′-oxygens. The efficiency of the conversion of **13** to **1** is thought to be due to a high selectivity for 3′-phosphorylation over 5′-phosphorylation. Such selectivity is particularly noteworthy because of the increased steric hindrance normally associated with a secondary alcohol in comparison with a primary alcohol. To investigate this, we determined the X-ray crystal structure of **13**, and found that its sugar moiety has the C(4′)-*endo* pucker with a C(4′)–C(5′) +*sc* conformation (Fig. 4d). This conformation has the effect of making the 5′-hydroxyl group of **13** abnormally hindered for a nucleoside derivative relative to the 3′-hydroxyl group. Assuming that the solid-state conformation of **13** is also the predominant conformation in urea and in formamide solution, the relative ease with which the 3′- and

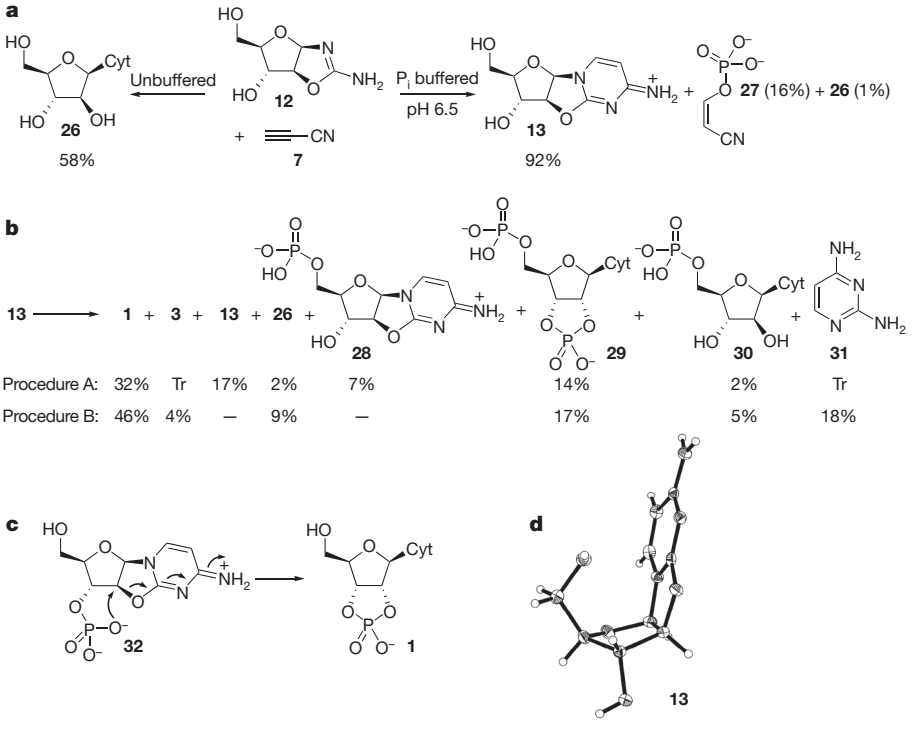

Figure 4 | Formation and phosphorylation of the arabinose anhydronucleoside 13. **a**, Major products isolated from the reaction of arabinose amino-oxazoline **12** and cyanoacetylene **7** in the presence and absence of phosphate. **b**, Products of phosphorylation of **13** using pyrophosphate and urea in the dry state (procedure A) or inorganic phosphate and urea in formamide solution (procedure B) (see main text and Supplementary Information). **c**, Rearrangement of **32**, the 3′-phosphate of **13**, to **1** by intramolecular nucleophilic substitution. **d**, X-ray crystal structure of **13**. Cyt, N1-linked cytosine; Tr, trace.

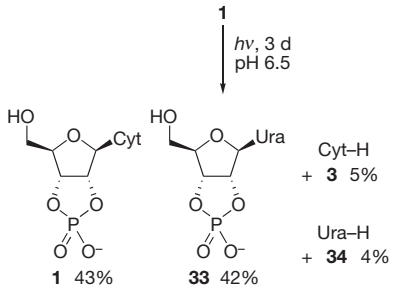

Figure 5 | Photochemistry of β-ribocytidine-2′,3′-cyclic phosphate 1. Under conditions of irradiation that destroy most other pyrimidine nucleosides and nucleotides (Supplementary Information), 1 undergoes partial hydrolysis and slight nucleobase loss. Ura, N1-linked uracil; Cyt–H, cytosine; Ura–H, uracil.

5′-hydroxyl groups are phosphorylated can be understood on the basis of these conformationally controlled steric effects.

If the products of the phosphorylation reaction in urea were subsequently to dissolve in aqueous medium at neutral pH and incubate for any significant length of time, then any residual anhydronucleoside/anhydronucleotide would undergo hydrolysis. Assuming such a rehydration, after phosphorylation in urea or urea–formamide mixtures, the major nucleosides/nucleotides that would accompany **1** would be **26**, **29** and **30** (in addition to any products ultimately deriving from the ribose amino-oxazoline **21** that was a by-product in the synthesis of arabinose amino-oxazoline **12**; see Supplementary Information). It is apparent that although **1** would be one of the major products, these co-products might interfere with any subsequent incorporation of **1** into RNA. Accordingly, we sought a means of selectively destroying these co-products. Furthermore, we also hoped to find a way of converting **1** partly to the corresponding activated uracil nucleotide, β-ribouridine-2′,3′-cyclic phosphate **33**. It transpires that irradiation achieves both of these goals.

Limited irradiation of aqueous solutions of cytosine nucleosides with ultraviolet light having an emission maximum at 254 nm results in the reversible formation of photohydrates and partial hydrolysis to the corresponding uracil nucleosides[28]. Prolonged irradiation causes additional chemistry to take place[29], and results in the destruction of most pyrimidine nucleosides and nucleotides (for example **26**, **30** and the major nucleoside/nucleotide products deriving from ribose amino-oxazoline **21**; see Supplementary Information). By contrast, however, we found that prolonged irradiation of β-ribocytidine-2′,3′-cyclic phosphate **1** causes significant hydrolysis to β-ribouridine-2′,3′-cyclic phosphate **33**, with very little destructive photochemistry other than slight nucleobase loss; cytosine **3** and uracil **34** were both detected (Fig. 5). This finding suggests that there must be some protective mechanism functioning with **1** and **33** that does not operate with other pyrimidine nucleosides and nucleotides. Whatever the mechanism (Supplementary Information), the protection against the destructive effects of irradiation provides a means whereby **1** and **33**, the two activated pyrimidine ribonucleotides needed for RNA synthesis, can be enriched relative to other end products of the assembly process we have discovered.

Our findings suggest that the prebiotic synthesis of activated pyrimidine nucleotides should be viewed as predisposed[30]. This predisposition would have allowed the synthesis to operate on the early Earth under geochemical conditions suitable for the assembly sequence. Although the issue of temporally separated supplies of glycolaldehyde and glyceraldehyde remains a problem, a number of situations could have arisen that would result in the conditions of heating and progressive dehydration followed by cooling, rehydration and ultraviolet irradiation. Comparative assessment of these models is beyond the scope of this work, but it is hoped that the chemistry described here will contribute to such an assessment.

Received 11 December 2008; accepted 24 March 2009.

1. Woese, C. *The Genetic Code* 179–195 (Harper & Row, 1967).
2. Crick, F. H. C. The origin of the genetic code. *J. Mol. Biol.* **38,** 367–379 (1968).
3. Orgel, L. E. Evolution of the genetic apparatus. *J. Mol. Biol.* **38,** 381–393 (1968).
4. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418,** 214–221 (2002).
5. Joyce, G. F. & Orgel, L. E. in *The RNA World* (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 23–56 (Cold Spring Harbor Laboratory Press, 2006).
6. Ferris, J. P., Hill, A. R. Jr, Liu, R. & Orgel, L. E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381,** 59–61 (1996).
7. Verlander, M. S., Lohrmann, R. & Orgel, L. E. Catalysts for the self-polymerization of adenosine cyclic 2′,3′-phosphate. *J. Mol. Evol.* **2,** 303–316 (1973).
8. Kofoed, J., Reymond, J.-L. & Darbre, T. Prebiotic carbohydrate synthesis: zinc-proline catalyzes direct aqueous aldol reactions of α-hydroxy aldehydes and ketones. *Org. Biomol. Chem.* **3,** 1850–1855 (2005).
9. Ricardo, A., Carrigan, M. A., Olcott, A. N. & Benner, S. A. Borate minerals stabilize ribose. *Science* **303,** 196 (2004).
10. Fuller, W. D., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis VI. Synthesis of purine nucleosides. *J. Mol. Biol.* **67,** 25–33 (1972).
11. Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **39,** 99–123 (2004).
12. Thaddeus, P. The prebiotic molecules observed in the interstellar gas. *Phil. Trans. R. Soc. B* **361,** 1681–1687 (2006).
13. Sanchez, R. A., Ferris, J. P. & Orgel, L. E. Cyanoacetylene in prebiotic synthesis. *Science* **154,** 784–785 (1966).
14. Pasek, M. A. & Lauretta, D. S. Aqueous corrosion of phosphide minerals from iron meteorites: a highly reactive source of prebiotic phosphorus on the surface of the early Earth. *Astrobiology* **5,** 515–535 (2005).
15. Bryant, D. E. & Kee, T. P. Direct evidence for the availability of reactive, water soluble phosphorus on the early Earth. H-phosphinic acid from the Nantan meteorite. *Chem. Commun.* 2344–2346 (2006).
16. Robertson, M. P. & Miller, S. L. An efficient prebiotic synthesis of cytosine and uracil. *Nature* **373,** 772–774 (1995).
17. Ferris, J. P., Goldstein, G. & Beaulieu, D. J. Chemical evolution, IV. An evaluation of cyanovinyl phosphate as a prebiotic phosphorylating agent. *J. Am. Chem. Soc.* **92,** 6598–6603 (1970).
18. Lohrmann, R. & Orgel, L. E. Prebiotic synthesis: phosphorylation in aqueous solution. *Science* **161,** 64–66 (1968).
19. Anastasi, C. *et al.* RNA: prebiotic product, or biotic invention? *Chem. Biodivers.* **4,** 721–739 (2007).
20. Anastasi, C., Crowe, M. A., Powner, M. W. & Sutherland, J. D. Direct assembly of nucleoside precursors from two- and three-carbon units. *Angew. Chem. Int. Edn Engl.* **45,** 6176–6179 (2006).
21. Cockerill, A. F. *et al.* An improved synthesis of 2-amino-1,3-oxazoles under basic catalysis. *Synthesis* 591–593 (1976).
22. Springsteen, G. & Joyce, G. F. Selective derivatization and sequestration of ribose from a prebiotic mix. *J. Am. Chem. Soc.* **126,** 9578–9583 (2004).
23. Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis V. Synthesis and photoanomerization of pyrimidine nucleosides. *J. Mol. Biol.* **47,** 531–543 (1970).
24. Tapiero, C. M. & Nagyvary, J. Prebiotic formation of cytidine nucleotides. *Nature* **231,** 42–43 (1971).
25. Shannahoff, D. H. & Sanchez, R. A. 2,2′-Anhydronucleosides. Novel syntheses and reactions. *J. Org. Chem.* **38,** 593–598 (1973).
26. Lohrmann, R., &. Orgel, L. E. Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents. *Science* **171,** 490–494 (1971).
27. Schoffstall, A. M. Prebiotic phosphorylation of nucleosides in formamide. *Orig. Life* **7,** 399–412 (1976).
28. Miller, N. & Cerutti, P. Structure of the photohydration products of cytidine and uridine. *Proc. Natl Acad. Sci. USA* **59,** 34–38 (1968).
29. Powner, M. W. *et al.* On the prebiotic synthesis of ribonucleotides: photoanomerisation of cytosine nucleosides and nucleotides revisited. *ChemBioChem* **8,** 1170–1179 (2007).
30. Eschenmoser, A. & Loewenthal, E. Chemistry of potentially prebiological natural products. *Chem. Soc. Rev.* **21,** 1–16 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work has been funded by the UK Engineering and Physical Sciences Research Council through the provision of a postdoctoral fellowship (B.G.) and a PhD studentship (M.W.P.). We thank J. Raftery and M. Helliwell for X-ray crystallography.

Author Information X-ray crystallographic data (excluding structure factors) for **13** have been deposited at the Cambridge Crystallographic Data Centre, UK, under deposition number CCDC 701055. These data can be obtained free of charge from the Cambridge Crystallographic Data Centre (http://www.ccdc.cam.ac.uk/data_request/cif). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.D.S. (john.sutherland@manchester.ac.uk).