



Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates

Hyo-Joong Kim^{a,1} and Steven A. Benner^{a,b}

^aFirebird Biomolecular Sciences LLC, Alachua, FL 32615; and ^bFoundation for Applied Molecular Evolution, Alachua, FL 32615

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved September 18, 2017 (received for review June 14, 2017)

According to a current “RNA first” model for the origin of life, RNA emerged in some form on early Earth to become the first biopolymer to support Darwinism here. Threose nucleic acid (TNA) and other polyelectrolytes are also considered as the possible first Darwinian biopolymer(s). This model is being developed by research pursuing a “Discontinuous Synthesis Model” (DSM) for the formation of RNA and/or TNA from precursor molecules that might have been available on early Earth from prebiotic reactions, with the goal of making the model less discontinuous. In general, this is done by examining the reactivity of isolated products from proposed steps that generate those products, with increasing complexity of the reaction mixtures in the proposed mineralogical environments. Here, we report that adenine, diaminopurine, and hypoxanthine nucleoside phosphates and a noncanonical pyrimidine nucleoside (zebularine) phosphate can be formed from the direct coupling reaction of cyclic carbohydrate phosphates with the free nucleobases. The reaction is stereoselective, giving only the β -anomer of the nucleotides within detectable limits. For purines, the coupling is also regioselective, giving the *N*-9 nucleotide for adenine as a major product. In the DSM, phosphorylated carbohydrates are presumed to have been available via reactions explored previously [Krishnamurthy R, Guntha S, Eschenmoser A (2000) *Angew Chem Int Ed* 39:2281–2285], while nucleobases are presumed to have been available from hydrogen cyanide and other nitrogenous species formed in Earth’s primitive atmosphere.

prebiotic synthesis | nucleotide | phosphorylated carbohydrate

Life on Earth is thought to have begun with the emergence of an informational molecule that could be replicated, with errors, where those errors are themselves replicable. These are believed to be necessary and perhaps sufficient features to support Darwinian evolution, which in turn is believed to be the only mechanism by which organic matter can spontaneously self-assemble to give properties that we value in life (1, 2). Based on an analysis of its role in modern biology and its presumed increased role in ancient biology, RNA is one of the most prominently sought first informational molecules, although threose nucleic acid (TNA) (3) and peptide nucleic acid (4) have both been proposed as alternative candidates. However, considering their watery environment, the first genetic polymers likely had repeating charges in their backbones (5). TNA and RNA both have these, and are polyelectrolytes.

Therefore, most current efforts in prebiotic chemistry have concentrated on seeking pathways to make RNA (or TNA) from materials that were formed without life on early Earth (6). Nucleosides are subunits of these biopolymers, and multiple prebiotic routes to these have been proposed.

For example, the direct condensation of ribose itself with purine nucleobases (adenine and hypoxanthine) is known to provide the corresponding ribonucleosides (7). However, such procedures, as previously reported, suffer from low yields for β -furanonucleosides and the formation of mixtures as a result of multiple nucleophilic centers on the heterocycle, multiple ring

sizes (furanose and pyranose) of the ribose, and reactions that lack stereo- and/or regioselectivity.

Accordingly, a second approach considers condensation of fragments of the nucleobase and/or the carbohydrate to form a composite, with the nucleoside “finished” after the glycosyl bond is formed. For example, Carell and coworkers (8) reported the reaction of ribose (and ribose-borate) with formylated aminopyrimidine nucleobase fragments that, after coupling, further react to produce purine nucleosides. This reaction is regioselective with respect to the nucleobase, giving *N*-9 purine nucleosides in high yield. With respect to the carbohydrate, both furanoses and pyranoses were formed.

Pyrimidine nucleosides have always been more difficult to obtain under prebiotic conditions. However, Sanchez and Orgel (9) reported many years ago the formation of cytidine derivative by the reaction of a ribose derivative with cyanamide to give an aminooxazoline; the synthesis of the heterocycle was finished by reaction with cyanoacetylene and phosphate-assisted ring opening to make cytidine nucleotide (10). More recently, Powner et al. (11) produced pyrimidine nucleoside phosphates by a process where both the carbohydrate and the nucleobase were introduced as fragments, and finished after the two precursor fragments were coupled.

We return here to a search for pathways to nucleos(t)ides that used completed heterocycles as starting materials (Fig. 1). This would avoid the requirement for prebiotic availability of the reactive cyanamide and cyanoacetylene. Nucleobases as metastable units may have been prebiotically available by either polymerization of hydrogen cyanide (HCN) (12, 13) or by thermal condensation of formamide in the presence of borate minerals (14).

Significance

Much recent research into the origins of life focuses on the hypothesis that RNA emerged on early Earth by an abiotic process, and gave Earth its first access to Darwinian evolution. This article provides a key step in this process. Here, we show that the phosphorylated ribonucleoside building blocks for RNA can be made stereoselectively under a prebiotic plausible condition with canonical and noncanonical purines, and with one noncanonical pyrimidine. It also shows that threose nucleoside phosphates can be synthesized in a similar way. This result is significant in terms of numbers of steps, high stereo- and regiochemistry, scope, involvement of minerals, and likelihood of the prebiotic availability of its starting materials.

Author contributions: H.-J.K. designed research; H.-J.K. performed research; H.-J.K. and S.A.B. analyzed data; and H.-J.K. and S.A.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

¹To whom correspondence should be addressed. Email: hkim@firebirdbio.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1710778114/-DCSupplemental.

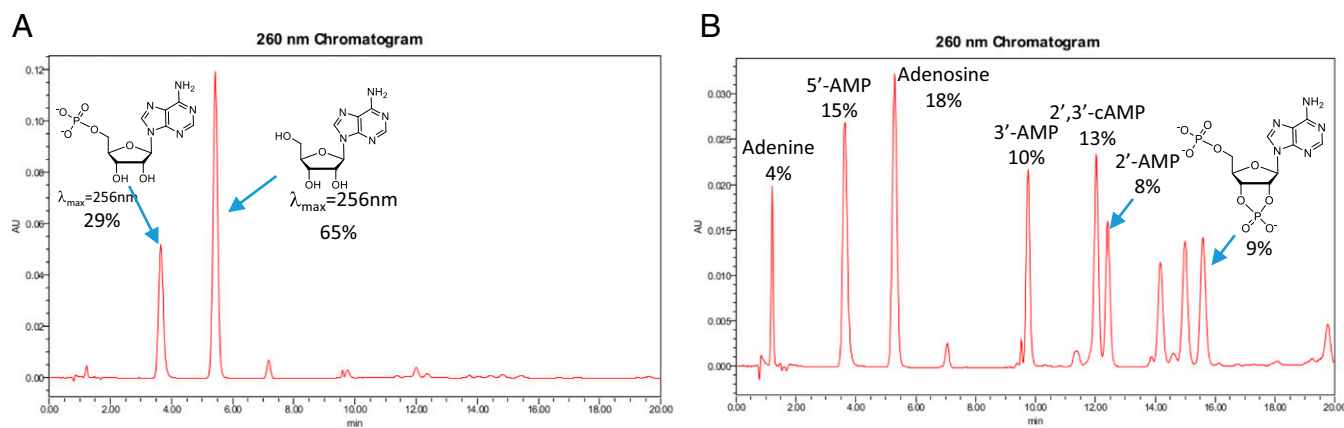
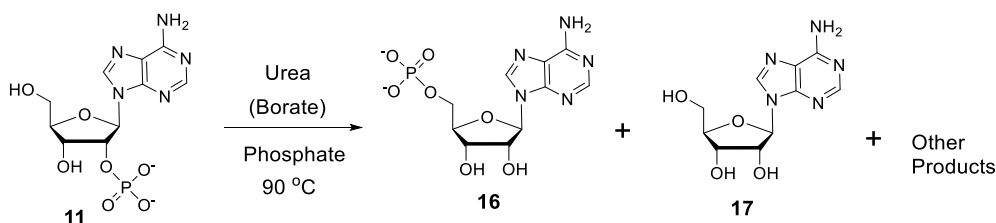


Fig. 3. Reversed-phase HPLC profile of the incubation of adenosine-2'-phosphate (11). (A) Incubation of adenosine-2'-phosphate in the presence of urea, borate, and sodium phosphate. (B) Incubation of adenosine-2'-phosphate in the presence of urea, sodium phosphate, and no borate.

of hydroxyl groups by coordinating to *cis*-diol of ribonucleotide and forces phosphorylation on the 5'-OH (28).

TNA is an alternative potential prebiotic informational polymer that might have appeared on Earth before the RNA world (3). Accordingly, we examined the direct condensation of threose and adenine under the same dry conditions. The reaction provided two products having together as much as a 70% yield (*SI Appendix, section 2.9*). Those products showed the same mass spectra as synthetic threofuranosyl adenine (40 in *SI Appendix*). However, their HPLC mobilities did not match those of 40. This suggested that those products have nucleosidic bonds on exocyclic NH_2 group of adenine (*SI Appendix, section 1.9*). This suggestion was confirmed by UV spectroscopy.

However, reaction of threose-1,2-cyclic phosphate (7) and purine and noncanonical pyrimidine nucleobases gave threose nucleoside-2'-phosphates (Fig. 4 and *SI Appendix, section 5.3*). These structures were proven by comparison with authentic synthetic compounds (*SI Appendix, sections 1 and 2*). Here, the glycosidic bond is formed to a ring nitrogen, not to an exocyclic amino group, as judged by UV spectroscopy.

Threose-1,2-cyclic phosphate proved to be under these conditions a slightly more receptive carbohydrate than ribose. Thus, whereas ribose-1,2-cyclic phosphate (5) does not react with uracil (20) to any detectable extent, threose-1,2-cyclic phosphate (7) reacted with uracil (20) to produce the corresponding nucleotide in 0.4% yield as its 2'-phosphorylated derivative. Here, of course, the phosphate is at a site involved in internucleotides in TNA. Thus, TNA remains an interesting alternative to RNA, if only as the sequential process that eventually generates RNA on a prebiotic Earth (29).

Discussion

These results show that the reaction of ribose-1,2-cyclic phosphate (5) and threose-1,2-cyclic phosphate (7) with a range of purine nucleobases gives ribonucleosides or threosynucleosides first as their 2'-phosphates. The condensation reaction is stereoselective, giving β -nucleotides only, and regioselective; *N*-9 nucleotides predominate

in the case of adenine, and 2,6-diaminopurine and *N*-3 nucleotides are major products for 2-pyrimidinone. Thus, they suggest that the direct coupling of preformed carbohydrates and preformed heterocycles need not be considered a prebiotic dead-end.

The stereoselectivity of the cyclic carbohydrate phosphates 5 and 7. Since these reactions are not catalyzed by any enzyme-like macromolecules, the reactive intermediate is not thought to be involvement of an oxocarbenium ion, but rather to proceed by concerted mechanism (30). Under this rationalization, reaction from the α -face is blocked by the cyclic phosphate, which activates the 1-position of the carbohydrates. This reactivity and stereoselectivity resemble the modern organic synthetic method that uses Vorbrüggen reaction conditions, which also has a cyclic intermediate as a reactive species (31).

The coupling reaction proceeds under dry state at elevated temperature ($\geq 70\text{ }^\circ\text{C}$). Further, the reaction requires divalent metal ions, magnesium or calcium. In some cases, the presence of ammonium formate increases the reaction yield (*Methods and SI Appendix, section 5*).

Finally, although borate need not be present for the coupling reaction, it does control subsequent reactivity of the 2'-phosphorylated nucleoside derivatives. In the presence of borate, these 2'-phosphorylated nucleoside derivatives undergo rearrangement in urea in the presence of inorganic phosphate to give the 5'-phosphorylated nucleoside derivatives. Absent borate, complex mixtures are seen. Borate is also useful in making the precursor carbohydrates without substantial decomposition.

Whether the coupling of carbohydrate-1,2-cyclic phosphate and nucleobases is a feasible prebiotic reaction pathway to nucleoside phosphates therefore rests on the availability of various precursor components. These include cyclic trimetaphosphate, ammonia, and ribose (or threose), which generate the cyclic phosphate precursors. These in turn rely on the availability of glyceraldehyde and glycolaldehyde, or glycolaldehyde with formaldehyde in the presence of borate. For the nucleobases, this presumes a hydrogen

linear gradient started from 100% A to 70% A at 15 min with flow rate of 10 mL/min. Peak detection was conducted using the 260-nm absorbance.

HPLC Analysis of the Reaction Products of Ribose-1,2-Cyclic Phosphate (5) and Nucleobases. HPLC analysis was done with a C-18 reversed-phase narrow-bore column (3 mm i.d., 150 mm length, 5 μ m; SunFire; Waters) on a Waters 2695 separation module equipped with 996 photodiode array detector. The column was eluted with a gradient of (A) aqueous 20 mM KH_2PO_4 with 5 mM tetrabutylammonium bromide (pH 3.3, adjusted by phosphoric acid) and (B) 100% acetonitrile. The elution program created a linear gradient that started from 99% A to 2.5 min, 97% at 5.5 min, 89.5% at 17.5 min, and 65.0% at 23.5 min with total flow rate of 0.8 mL/min. Peak detection and integration were conducted with the signal at 260 nm for adenine, hypoxanthine, 2,6-diaminopurine and 300 nm for pyrimidin-2-one. Full UV spectra (230 ~ 400 nm) were also obtained.

The yield of the coupling reaction was determined by the peak integration of the products compared with the integration of nucleosides having the same nucleobases (*SI Appendix, section 5.2*).

Reaction of Threose-1,2-Cyclic Phosphate (7) and Nucleobases. Threose-1,2-cyclic phosphate (7) was prepared following the published method (15). The reaction was conducted in an Eppendorf tube containing nucleobase [adenine (8), hypoxanthine (9), 2-hydroxypyrimidine hydrochloride (14), each 5 μ L of 3.75 mM], threose-1,2-cyclic phosphate (7) (5 μ L of 15 mM), either MgCl_2 (5 μ L of 3.75 mM) or CaCl_2 (5 μ L of 3.75 mM) with/without ammonium formate (5 μ L of 3.75 mM). The tube was placed in an oven at 70 $^\circ\text{C}$ for 18 h with the lid open. It was resuspended in water (0.3 mL) and analyzed by reversed-phase HPLC with 20 μ L of injection. The yield of the reactions is summarized in *SI Appendix, section 5.3*.

Reaction of Threose-1,2-Cyclic Phosphate (7) and Uracil. The reaction was conducted in an Eppendorf tube containing uracil (5 μ L of 3.75 mM), threose-1,2-cyclic phosphate (7) (5 μ L of 15 mM), MgCl_2 (5 μ L of 3.75 mM), and sodium hydroxide (5 μ L of 100 mM). The tube was placed in an oven at 70 $^\circ\text{C}$ for 18 h with the lid open. It was resuspended in 1 M TEAA buffer (0.3 mL) and analyzed by reversed-phase HPLC with 20 μ L of injection.

HPLC Analysis of the Reaction Products of Threose-1,2-Cyclic Phosphate (7) and Nucleobases. HPLC analysis was done with a C-18 reversed-phase narrow-bore column (3 mm i.d., 150 mm length, 5 μ m; SunFire; Waters) on a Waters 2695 separation module equipped with 996 photodiode array detector. The column was eluted with a gradient of (A) aqueous 25 mM triethylammonium acetate and (B) 100% acetonitrile. The elution program created a linear gradient started from 100% (by volume) A to 85% A at 10 min with flow rate of 0.5 mL/min. Peak detection and integration were conducted with the signal at 260 nm for adenine, hypoxanthine, uracil and 300 nm for 2-hydroxypyrimidine. Full UV spectra (230 ~ 400 nm) were also obtained.

The yield of the coupling reaction was determined by the peak integration of the products compared with the integration of nucleosides having the same nucleobases (*SI Appendix, section 5.4*).

ACKNOWLEDGMENTS. We thank Professor Andrew Ellington and one unnamed referee for calling our attention to specific literature. This publication was made possible through the support of John Templeton Foundation Grant 54466. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

1. Woese C (1967) *The Genetic Code* (Harper & Row, New York), pp 179–195.
2. Crick FHC (1968) The origin of the genetic code. *J Mol Biol* 38:367–379.
3. Schöning K, et al. (2000) Chemical etiology of nucleic acid structure: The alpha-thiofuranosyl-(3'→2') oligonucleotide system. *Science* 290:1347–1351.
4. Nielsen PE, Egholm M, Berg RH, Buchardt O (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* 254:1497–1500.
5. Westheimer FH (1987) Why nature chose phosphates. *Science* 235:1173–1178.
6. Orgel LE (2004) Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol* 39:99–123.
7. Fuller WD, Sanchez RA, Orgel LE (1972) Studies in prebiotic synthesis: VII. Solid-state synthesis of purine nucleosides. *J Mol Biol* 1:249–257.
8. Becker S, et al. (2016) A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* 352:833–836.
9. Sanchez RA, Orgel LE (1970) Studies in prebiotic synthesis. V. Synthesis and photoanomerization of pyrimidine nucleosides. *J Mol Biol* 47:531–543.
10. Tapiero CM, Nagyvary J (1971) Prebiotic formation of cytidine nucleotides. *Nature* 231:42–43.
11. Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459:239–242.
12. Oro J (1961) Mechanism of synthesis of adenine from hydrogen cyanide under possible primitive earth conditions. *Nature* 191:1193–1194.
13. Orgel LE (2004) Prebiotic adenine revisited: Eutectics and photochemistry. *Orig Life Evol Biosph* 34:361–369.
14. Saladino R, Barontini M, Cossetti C, Di Mauro E, Crestini C (2011) The effects of borate minerals on the synthesis of nucleic acid bases, amino acids and biogenic carboxylic acids from formamide. *Orig Life Evol Biosph* 41:317–330.
15. Krishnamurthy R, Guntha S, Eschenmoser A (2000) Regioselective α -phosphorylation of aldose in aqueous solution. *Angew Chem Int Ed* 39:2281–2285.
16. Feldmann W, Thilo E (1964) Zur chemie der kondensierten phosphat und arsenate. XXXVIII. Amidotriphosphat. *Z Anorg Allg Chem* 328:113–126.
17. Pasek MA, Kee TP, Bryant DE, Pavlov AA, Lunine JI (2008) Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry. *Angew Chem Int Ed Engl* 47:7918–7920.
18. Benner SA, Kim HJ, Carrigan MA (2012) Asphalt, water, and the prebiotic synthesis of ribose, ribonucleosides, and RNA. *Acc Chem Res* 45:2025–2034.
19. Ricardo A, Carrigan MA, Olcott AN, Benner SA (2004) Borate minerals stabilize ribose. *Science* 303:196.
20. Neveu M, Kim HJ, Benner SA (2013) The “strong” RNA world hypothesis: Fifty years old. *Astrobiology* 13:391–403.
21. Kim HJ, et al. (2011) Synthesis of carbohydrates in mineral-guided prebiotic cycles. *J Am Chem Soc* 133:9457–9468.
22. Weber AL, Pizzarello S (2006) The peptide-catalyzed stereospecific synthesis of tetroses: A possible model for prebiotic molecular evolution. *Proc Natl Acad Sci USA* 103:12713–12717.
23. Löb W (1913) Über das verhalten des formamids unter der wirkung der stillen entladung ein beitrage zur frage der stickstoff-assimilation. *Ber Dtsch Chem Ges* 46:684–697.
24. Ritson DJ, Sutherland JD (2013) Synthesis of aldehydic ribonucleotide and amino acid precursors by photoredox chemistry. *Angew Chem Int Ed Engl* 52:5845–5847.
25. Jørgensen JK, et al. (2012) Detection of the simplest sugar, glycolaldehyde, in a solar-type protostar with ALMA. *ApJL* 757:L4.
26. Bean HD, et al. (2007) Formation of a β -pyrimidine nucleoside by a free pyrimidine base and ribose in a plausible prebiotic reaction. *J Am Chem Soc* 129:9556–9557.
27. Walton T, Szostak JW (2016) A highly reactive imidazolium-bridged dinucleotide intermediate in nonenzymatic RNA primer extension. *J Am Chem Soc* 138:11996–12002.
28. Kim HJ, et al. (2016) Evaporite borate-containing mineral ensembles make phosphate available and regioselectively phosphorylate ribonucleosides: Borate as a multifaceted problem solver in prebiotic chemistry. *Angew Chem Int Ed Engl* 55:15816–15820.
29. Yu H, Zhang S, Chaput JC (2012) Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor. *Nat Chem* 4:183–187.
30. Unrau PJ, Bartel DP (2003) An oxocarbenium-ion intermediate of a ribozyme reaction indicated by kinetic isotope effects. *Proc Natl Acad Sci USA* 100:15393–15397.
31. Vorbrüggen H, Ruh-Pohlentz C (2000) Synthesis of nucleosides. *Org React* 55:1.
32. Fathi R, Jordan F (1986) α -D-Ribofuranosyl 1,2-cyclic monophosphate. Isolation, NMR spectroscopic properties, and rates and mechanism of acid and alkaline hydrolysis. *J Org Chem* 51:4143–4146.