Constraints on the emergence of RNA through non-templated primer extension with mixtures of potentially prebiotic nucleotides

Xiwen Jia [1,2,3], Stephanie J. Zhang [1,2], Lijun Zhou [4,5] and Jack W. Szostak [3,*]

[1]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA
[2]Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114, USA
[3]Howard Hughes Medical Institute, Department of Chemistry, The University of Chicago, Chicago, IL 60637, USA
[4]Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
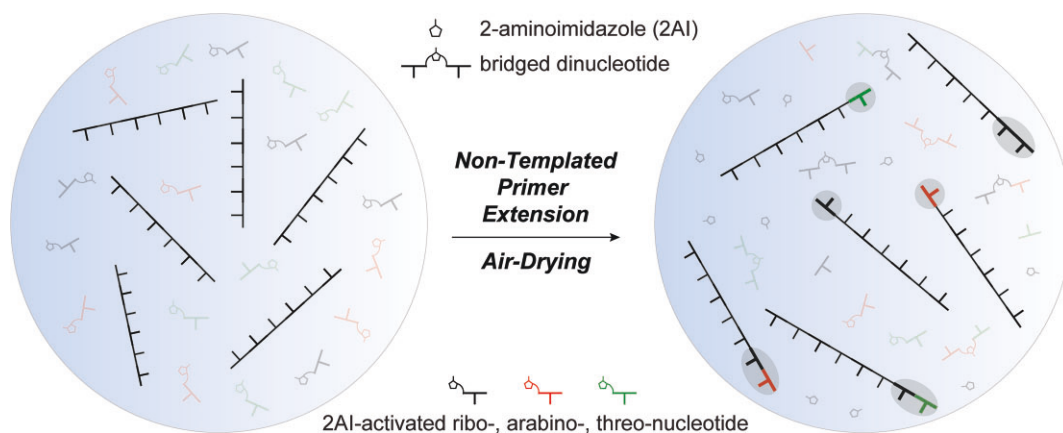[5]Penn Institute for RNA Innovation, University of Pennsylvania, Philadelphia, PA 19104, USA

[*]To whom correspondence should be addressed. Email: jwszostak@uchicago.edu
Present address: Stephanie J. Zhang, Department of Pathology, Brigham and Women's Hospital, 60 Fenwood Rd, Boston, MA 02115, USA.

Abstract

The emergence of RNA on the early Earth is likely to have been influenced by chemical and physical processes that acted to filter out various alternative nucleic acids. For example, UV photostability is thought to have favored the survival of the canonical nucleotides. In a recent proposal for the prebiotic synthesis of the building blocks of RNA, ribonucleotides share a common pathway with arabino- and threo-nucleotides. We have therefore investigated non-templated primer extension with 2-aminoimidazole-activated forms of these alternative nucleotides to see if the synthesis of the first oligonucleotides might have been biased in favor of RNA. We show that non-templated primer extension occurs predominantly through 5′-5′ imidazolium-bridged dinucleotides, echoing the mechanism of template-directed primer extension. Ribo- and arabino-nucleotides exhibited comparable rates and yields of non-templated primer extension, whereas threo-nucleotides showed lower reactivity. Competition experiments confirmed the bias against the incorporation of threo-nucleotides. The incorporation of an arabino-nucleotide at the end of the primer acts as a chain terminator and blocks subsequent extension. These biases, coupled with potentially selective prebiotic synthesis, and the templated copying that is known to favour the incorporation of ribonucleotides, provide a plausible model for the effective exclusion of arabino- and threo-nucleotides from primordial oligonucleotides.

Graphical abstract

Introduction

RNA is considered to be a promising candidate for the primordial genetic polymer, owing to its dual roles in encoding genetic information and catalyzing reactions ([1–3]). However, just how RNA might have emerged from prebiotic mixtures remains an open question. Potentially prebiotic synthetic pathways have been proposed in which ribo-, arabino- and threo-nucleotides would have shared common precursors ([4–6]) (Figure [1]). In these proposed pathways, cyanamide and glycolaldehyde react to yield 2-aminooxazole (2AO), which subsequently reacts with glyceraldehyde to form ribose aminooxazoline (RAO) and arabinose aminooxazoline (AAO), the precursors to the five-carbon sugar ribo- and arabino- pyrimidine nucleosides, respectively ([4]). While ribo- and arabino-
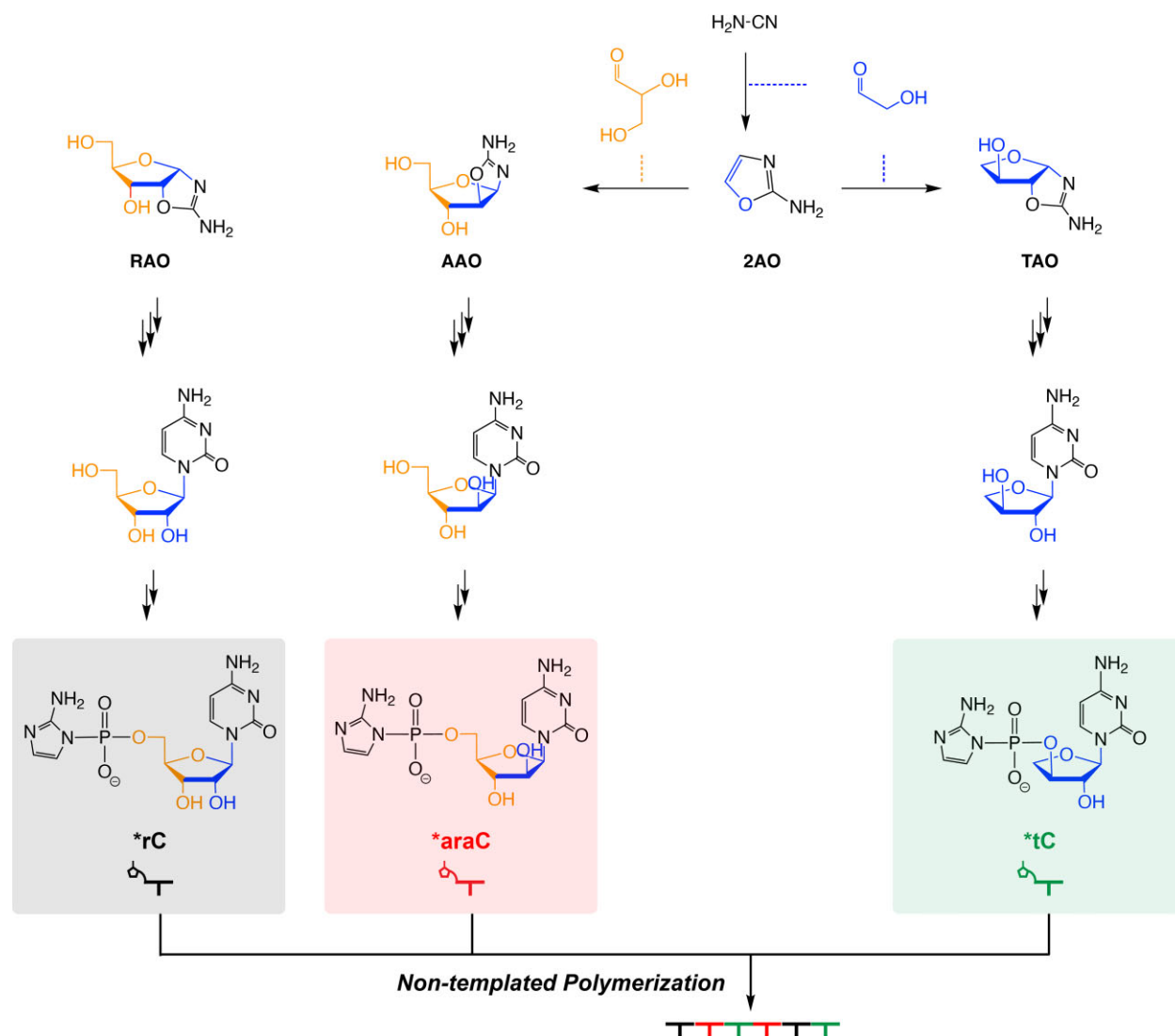
Figure 1. Potentially prebiotic synthetic pathway for activated ribo-, arabino-, and threo- cytidine mononucleotides. Phosphorylated and activated ribo-, arabino- and threo- cytidine mononucleotides (*rC, *araC and *tC) can undergo non-templated polymerization to form chimeric oligonucleotides.

nucleosides share common early steps in proposed synthetic pathways, the synthesis of arabino-nucleosides requires fewer steps, occurs over a broader pH range, and can be carried out in water (5). 2AO can react with a second glycolaldehyde to form threose aminooxazoline (TAO), the precursor to the four-carbon sugar threo- pyrimidine nucleosides (6). The resulting pyrimidine nucleosides can then undergo phosphory-lation to generate nucleotides (7–9), which can then be activated as imidazolides (10,11). These highly reactive mononucleotides can then participate in non-templated polymerization, generating oligonucleotides that could serve as primers and templates in subsequent nonenzymatic template-directed copying reactions.

The non-templated polymerization of activated mononucleotides has been demonstrated on mineral surfaces such as montmorillonite (12), and also in the ice eutectic phase (13). Montmorillonite immobilizes oligonucleotides and enables successive additions of activated mononucleotides, allowing incremental elongation of the primer in the absence of a template (12). The ice eutectic phase also concentrates the solutes, enabling non-templated polymeriza-

tion, while the low temperature slows down hydrolysis (13). Early studies of these processes employed imidazole-activated mononucleotides. Recent developments have shown that 2-aminoimidazole (2AI) activated nucleotides can enhance nonenzymatic template copying (14) because they spontaneously form highly reactive 5′-5′ imidazolium bridged dinucleotides (15) that are the substrates for the predominant pathway for template-directed primer extension (16). In addition, the activating group 2AI shares a common and potentially prebiotic synthetic pathway with the precursor of nucleotide synthesis, 2AO (17).

The effect of sugar heterogeneity in the primordial pool of activated nucleotides remains a major question to be addressed en route to the RNA World. Through kinetic and crystallographic studies, template-directed copying chemistry has been shown to favor RNA synthesis (18–20). The conformation of the 3′-terminal nucleotide of the primer and the identity of the incoming bridged dinucleotides are the major determinants of the kinetics of template-directed primer extension. However, sugar heterogeneity has rarely been studied in the non-templated reactions that must have occurred to pro-

duce the first primordial templates. Is there any selectivity in non-templated reactions, given that such a reaction would be free from constraints arising from template binding? Could the structural differences between ribo-, arabino- and threo-nucleotides influence their reactivity, shaping the composition of prebiotic oligomers? We have attempted to gain insight into the likely distribution of oligonucleotide products by simulating these conditions through competition experiments with mixtures of nucleotide substrates.

In this study, we first examined the mechanism of non-templated primer extension and identified imidazolium-bridged dinucleotides as the most reactive substrates. Subsequent kinetic analyses using 2AI-activated ribo-, arabino- and threo-nucleotides showed that ribonucleotides and arabino-nucleotides were incorporated into a growing primer at similar rates, while threo-nucleotides displayed lower reactivity. We then combined these nucleotides to mirror the heterogenous prebiotic environment, and used LC–MS to investigate the products of non-templated primer extension. Again, arabino-nucleotides and ribonucleotides were incorporated to similar extents, whereas threo-nucleotides were outcompeted. However, the incorporation of an arabino-nucleotide at the end of the primer prevents further extension. Our findings demonstrate an inherent bias against the incorporation of arabino- and threo-nucleotides during non-templated polymerization.

Material and methods

General Information

Materials
Reagents and solvents were obtained from Fischer Scientific, Sigma-Aldrich, Alfa Aesar, Acros Organics and were used without further purification unless otherwise noted. RNA and DNA oligonucleotides were purchased from Integrated DNA Technologies or synthesized in-house using the Expedite 8909 DNA/RNA synthesizer. Further information on oligonucleotide synthesis is included in the Supplementary data.

Nuclear magnetic resonance (NMR)
^{1}H and ^{31}P NMR spectra were acquired on a Varian Oxford AS-400 NMR spectrometer (400 MHz for ^{1}H, 162 MHz for ^{31}P) at 25°C. ^{1}H NMR spectra were referenced using sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) as internal standard (0 ppm at 25°C).

Low-resolution mass spectrometry (LRMS)
All samples were diluted to 200 μM in 50% (v/v) acetonitrile immediately prior to analysis. The spectra were obtained by direct injection on an Esquire 6000 mass spectrometer (Bruker Daltonics), which was operated in the alternating ion mode.

High-resolution liquid chromatography–mass spectrometry (HR LC–MS)
The eluted samples were separated and analyzed using an Agilent 1200 high-performance liquid chromatography (HPLC) coupled to an Agilent 6230 time-of-flight mass spectrometry (TOF MS) equipped with a diode array detector. The samples were separated by IP-RP-HPLC on a 100 mm × 1 mm (length × i.d.) Xbridge C18 column with a 3.5 μm particle size (Waters, Milford, MA). The samples were eluted between

2.5 and 15% methanol in 200 mM 1,1,1,3,3,3-hexafluoro-2-propanol with 1.25 mM triethylamine at pH 7.0 over 16 min with a flow rate of 0.1 ml/min at 50°C. The samples were analyzed in negative mode from 239 m/z to 3200 m/z with a scan rate of 1 spectrum/s.

Synthesis and characterization of 2-aminoimidazole activated mononucleotides and 5′-5′ imidazolium-bridged dinucleotides

Arabinose guanosine nucleoside (araG) phosphorylation
The Yoshikawa method was used to phosphorylate the 5′-hydroxyl of araG nucleoside (21). To a pre-chilled mixture of 9-(β-D-arabinofuranosyl)guanine (araG, 1 equiv.) in trimethyl phosphate (OP(OMe)$_3$, 0.1 M with respect to the nucleoside) was added phosphoryl chloride (POCl$_3$, 4 equiv.) and a trace amount of H$_2$O (7.5 μl) under vigorous stirring. The resultant mixture was allowed to stir at 0°C. After complete solubilization of the nucleoside, four portions of *N,N*-diisopropylethylamine (DIPEA, 0.5 equiv. each) were added dropwise at 20 min intervals. The reaction was monitored by LRMS. Once the starting material disappeared, the reaction was quenched using an aqueous solution of 1 M triethylamine-bicarbonate (TEAB, 5 volumes, pH 7.5). Any precipitate observed after the quench was re-solubilized using a minimal amount of acetonitrile. The products were purified by reverse phase flash chromatography with a 50 g C18Aq column. The desired product was separated from other compounds over 10 column volumes (CVs) of 0–15% acetonitrile in 2 mM aqueous TEAB buffer (pH 7.5) with a flow rate of 40 ml/min. Fractions containing the product were collected and lyophilized under high vacuum level at room temperature. The 9-(β-D-arabinofuranosyl)guanine monophosphate was used for activation without further purification.

Threose cytidine nucleoside (tC) phosphorylation and deprotection
The benzoyl protecting groups on the 2′-hydroxyl and amine enabled selective phosphorylation of the 3′-hydroxyl. The reaction is outlined in Supplementary Scheme S1. To azeotropically dried (toluene 3×, 2.5 ml each) tC (1 equiv.) and bis(2-cyanoethyl)-*N,N*-disopropylphosphoramidite in acetonitrile (0.1 ml, 1.25 equiv.) was added a solution of 5-(ethylthio)-1H-tetrazole (ETT, 2.5 equiv.) in acetonitrile dropwise. The mixture was stirred for 1 h. The reaction was monitored by thin layer chromatography (TLC). Meta-chloroperoxybenzoic acid (mCPBA, 3.0 equiv.) was added and the mixture was stirred for another 5 min, monitored by TLC. The reaction mixture was diluted with 15 ml DCM, transferred to a separatory funnel and 10 ml aqueous sodium bicarbonate (NaHCO$_3$) was added. The aqueous layer was extracted with dichloromethane (DCM) two times and the combined organic extracts were washed with brine. The extracts were dried with sodium sulfate and concentrated by rotary evaporator. The compound was dissolved in 0.5 ml water. Following the addition of 20 mg mCPBA, the reaction was well mixed. 8 ml of ammonium hydroxide (NH$_4$OH) was then added. The progress of the deprotection reaction was monitored by LRMS. A heat gun was used to evaporate the ammonia before concentrating the product in a rotary evaporator. The dried compound was dissolved in 5 ml 200 mM TEAB at pH 9 and purified by reverse phase flash chromatography with a 50 g C18Aq column. The desired product was separated from

other compounds over 10 CVs of 0–5% acetonitrile in 2 mM TEAB buffer (pH 7.5) with a flow rate of 40 ml/min. The collected fractions were measured by LRMS and lyophilized for further activation.

Synthesis and characterization of 2-aminoimidazole activated mononucleotides (*N)

The activation of NMPs followed a previously reported procedure ([14]). The activation of threose cytidine and guanosine monophosphate (tCMP and tGMP) was similar to the NMP activation detailed above except for the following steps: (i) tCMP or tGMP (1.0 equiv.), 2AI·HCl (10 equiv.) and TPP (0.5 equiv.) were suspended in 30 ml DMSO under Ar and (ii) the collected fractions were pH adjusted to 9 by 1 M NaOH before lyophilization. The detailed characterizations (NMR and HRMS) are included in the Supplementary data.

Synthesis and characterization of 5′-5′ 2-aminoimidazolium-bridged dinucleotides (N*N)

The synthesis of N*N followed a previously reported protocol ([22]). The detailed characterizations (NMR and HRMS) are included in the supplementary data.

Non-templated primer extension reactions in the aqueous phase and ^{31}P NMR hydrolysis experiments

Non-templated primer extension

Primers were designed to avoid the possibility of template-directed primer extension. Non-templated primer extension reactions were performed at 1 μM primer, 20 mM activated mononucleotides, 200 mM HEPES at pH 8.0 and 50 mM MgCl$_2$. At each time point, 1 μl of the reaction sample was added to 19 μl of quenching buffer containing 7 M urea, 1× TBE, 100 mM EDTA.

Non-templated primer extension with different terminal nucleotides

The primer was mixed with a 7.5-fold excess of a complementary oligonucleotide in an annealing buffer at 2.2 times the final concentration. Solutions containing 2.2 μM primer (XJ-FAM-12mer or XJ-15, Supplementary Table S1), 16.3 μM complementary oligomer (XJ-16, Supplementary Table S1), 22 mM HEPES at pH 8.0, 22 mM NaCl, and 0.44 mM EDTA at pH 8.0 were heated at 95ºC for 30 s and then slowly cooled to 25ºC at a rate of 0.1ºC/s in a thermal cycler machine. The annealed products were then diluted into the primer extension reaction buffer to give final concentrations of 1 μM primer, 7.5 μM complementary oligonucleotide, 200 mM HEPES at pH 8.0, and 50 mM MgCl$_2$.

Urea-PAGE analysis

A fluorophore-labeled primer was utilized to visualize primer extension by polyacrylamide gel electrophoresis (PAGE). Primers were either 12- or 6-nucleotides long. Primer extension products were resolved by 20% PAGE with 7 M urea, in 1× TBE gel running buffer. For extension reactions using a 6-mer primer, the reactions were desalted by ion pairing reverse phase (IP-RP) purification using C18 ZipTip pipette tips (Millipore, Billerica, MA) to avoid smearing of bands in PAGE. The tips were wetted with 100% (v/v) acetonitrile and equilibrated with 100 mM TEAA prior to sample binding. Extensive washing with 100 mM TEAA, then with LC–MS grade

water, was followed by elution in 50% (v/v) acetonitrile. The eluates were dried in a centrifugal vacuum concentrator to remove acetonitrile, followed by resuspension in 10 μl of 7 M urea and 1× TBE for sample loading. An increased acrylamide concentration (22%) was used for improved separation of smaller molecules.

The gels were scanned with an Amersham Typhoon RGB Biomolecular Imager (GE Healthcare Life Sciences) to visualize the fluorophore-labeled primer and extended primer bands. The relative band intensities were quantified using ImageQuant TL software. The rate of the extension was determined from the linear least-square fits of the data.

^{31}P NMR hydrolysis experiments

The hydrolysis experiments were performed at the respective concentrations of activated species (activated mononucleotides or bridged dinucleotides), 50 mM MgCl$_2$, 200 mM HEPES at pH 8.0, 10% (v/v) D$_2$O, incubation at room temperature. The resulting ^{31}P NMR spectra were processed using the MestReNova software (v14.2.0). Given the known total nucleotide concentration of 20 mM in solution, the peak integrations from the NMR spectra were converted to the concentrations of the individual species.

Spontaneous air-drying experiments, competition experiments and subsequent LC–MS analysis

Spontaneous air-drying experiments

PCR tubes containing 10 μl of non-templated primer extension reactions were prepared with 20 mM activated mononucleotides, 100 μM 6-mer primer, 200 mM HEPES at pH 8.0 and 50 mM MgCl$_2$. The reactions were allowed to dry spontaneously under ambient air, resulting in clear pastes (Supplementary Figure S1). The reaction pastes were then quenched at 24 h by adding 58 μl of quench buffer containing 100 mM EDTA at pH 8.0. Spontaneous air-drying significantly accelerated the non-templated primer extension, thereby generating enough primer +1 products for the following LC–MS experiments.

Spontaneous air-drying experiments in the presence of complementary oligomers

The primer was mixed with a 7.5-fold excess of a complementary oligonucleotide in an annealing buffer at 2.2 times the final concentration. Solutions containing 217 μM primer (XJ-8, Supplementary Table S1), 1630 μM complementary oligomer (XJ-6mer-rc, Supplementary Table S1), 22 mM HEPES at pH 8.0, 22 mM NaCl and 0.44 mM EDTA at pH 8.0 were heated at 95ºC for 30s and then slowly cooled to 25ºC at a rate of 0.1ºC/s in a thermal cycler machine. The annealed products were then diluted into the primer extension reaction buffer to give final concentrations of 100 μM primer, 750 μM complementary oligonucleotide, 200 mM HEPES at pH 8.0, and 50 mM MgCl$_2$.

Competition experiments

In competition experiments, unlabeled primer was mixed with activated ribo-, arabino- and threo- nucleotides at the indicated ratio and left to air-dry spontaneously (reaction conditions as outlined in the spontaneous air-drying experiments). For an equimolar mixture (*rC:*araC:*tC = 1:1:1), the total of 20 mM activated mononucleotides consisted of 6.7 mM each of *rC, *araC and *tC. For the *rC:*araC:*tC = 10:1:1
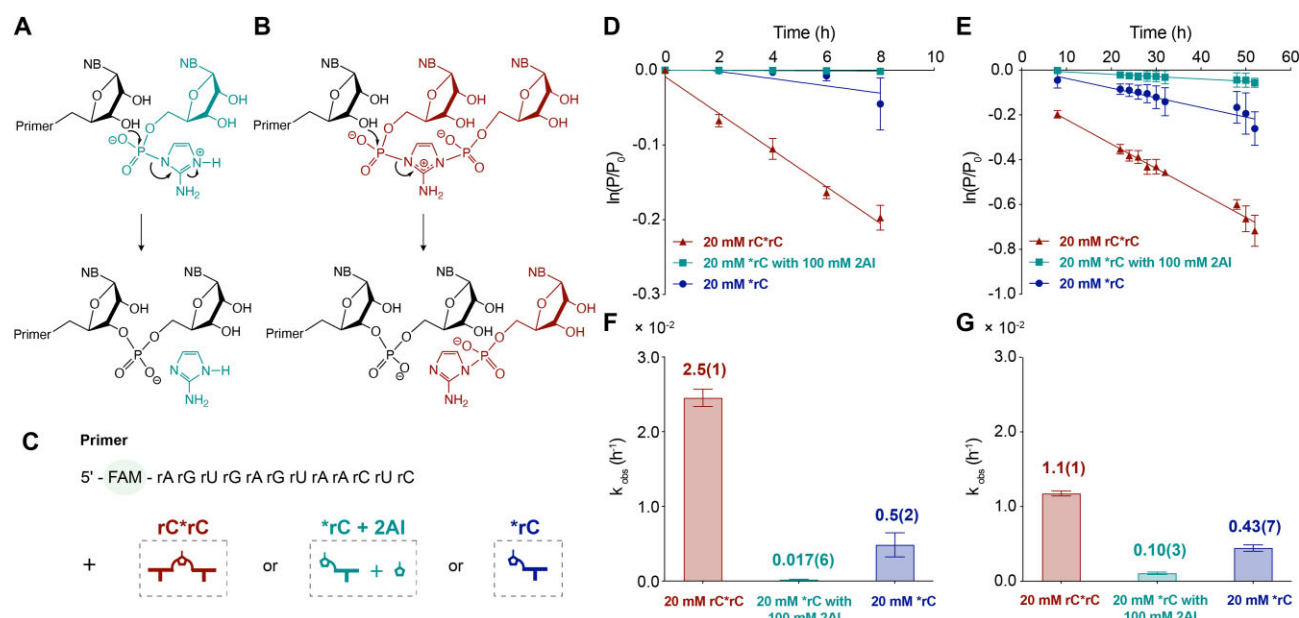
Figure 2. Mechanism of non-templated primer extension. Hypothesized mechanism of non-templated primer extension through (**A**) activated mononucleotides or (**B**) bridged dinucleotides. 'NB', nucleobase. Note that the scheme includes only the extension through 3′-OH moieties, although nucleotides can also add to 2′-OH groups. (**C**) Schematic representation of the non-templated primer extension using a FAM-labeled RNA primer with either bridged dinucleotides (rC*rC) or activated mononucleotides (*rC). 2AI was added to suppress the formation of bridged dinucleotides. (**D**) the initial and (**E**) later time course. Bar plots indicate the observed pseudo-first-order reaction rates (k_{obs}) calculated from (**F**) the initial 0–8 h and (**G**) the later 8–52 h time course. Reaction conditions: 1 μM FAM-labeled RNA primer (XJ-FAM-12mer, Supplementary Table S1), 200 mM HEPES at pH 8.0 and 50 mM MgCl$_2$ with activated species at 20 mM for rC*rC, 20 mM for *rC with 100 mM 2AI and 20 mM for *rC. Error bars represent standard deviations from the mean, n = 3 replicates. Note that non-templated primer extension with 20 mM *rC and 100 mM 2AI is below the detection limit of the gel electrophoresis assay at early time points and therefore the reported initial rate might not reflect its true rate. P/P$_0$ denotes the ratio of the unextended primer (P) to the initial total primer (P$_0$).

ratio, the total of 20 mM activated mononucleotides consisted of 16.7 mM *rC, 1.7 mM *araC and 1.7 mM *tC. For the experiment shown in Figure 4, the primer was XJ-5 (Supplementary Table S1) and for Figure 6 was XJ-10 (Supplementary Table S1).

Primer design for competition experiments

To improve the product separation in LC-MS, we used a 6-nt long 5′-OH primer instead of a fluorophore-labeled primer. In Ion Pairing Liquid Chromatography (IP-LC), separation is achieved through the difference in electrostatic affinities of charged molecules for a charged stationary phase (23). However, separating RNA oligomers of the same length can be challenging. To separate the +1 products ending in a ribo-, arabino-, or threo-nucleotide, we used shorter RNA oligomers with lower total charge. Furthermore, replacing a fluorophore-labeled primer with an unlabeled primer decreased the total charge and improved the separation of same-length RNA oligomers with different terminal nucleotides.

LC–MS sample preparation

Reaction mixtures to be subjected to LC–MS analysis were quenched with 100 mM EDTA at pH 8.0. The quenched reactions (60 μl) were aliquoted into 10 μl fractions for desalting by ion pairing reverse phase (IP-RP) purification on C18 ZipTip pipette tips (Millipore, Billerica, MA) as previously described. The eluates were combined into an HPLC vial insert and dried in a centrifugal vacuum concentrator. Samples were resuspended in LC–MS grade water prior to injection in HR LC–MS.

LC–MS analysis

Extracted compound chromatograms were generated using the Agilent MassHunter Qualitative Analysis software (B.07.00). For competition experiments, generated compound lists were matched with the calculated masses of all possible +1 non-templated primer extension products and their salt adducts. The observed and calculated masses of the relevant products are provided in Tables S2 and S3. For the purposes of this analysis, we assumed that the same-length oligomers ending in different terminal sugars have equivalent ionization efficiencies.

Results

Mechanism of non-templated primer extension

We hypothesized two distinct pathways for non-templated primer extension: one through the direct reaction of activated mononucleotides (Figure 2A) and the other through imidazolium bridged dinucleotides (Figure 2B). The activated mononucleotide pathway proceeds when the terminal hydroxyl group of the primer (3′-OH or 2′-OH) attacks the phosphorus atom of an activated mononucleotide, displacing 2-aminoimidazole (2AI) as the leaving group (Figure 2A). In contrast, the bridged dinucleotide pathway is characterized by the hydroxyl group attacking the phosphorus atom of an imidazolium bridged dinucleotide, displacing an activated mononucleotide as the leaving group (Figure 2B). These highly reactive dinucleotides form spontaneously by the reaction of two activated mononucleotides (24,25). While bridged dinucleotides are known to drive templated nonenzymatic copying

(16), their role in non-templated primer extension remains unclear due to the lack of stabilizing base pair interactions with the template.

To elucidate the mechanism of non-templated primer extension, we conducted comparative kinetic analyses of primer extension reactions using activated cytidine mononucleotides (*rC) or 5′-5′ imidazolium bridged dinucleotides (rC*rC) (Figure 2C). The reaction with 20 mM pre-formed purified bridged rC*rC started off fast (Figure 2D), but then slowed down (Figure 2E), presumably due to hydrolysis and an approach to a steady state equilibrium mixture of *rC and rC*rC. In contrast, the reaction with 20 mM *rC started off very slowly, but speeded up with time, again presumably due to the formation of an equilibrium mixture of *rC and rC*rC. Because *rC can spontaneously react to form bridged dinucleotides *in situ*, we added excess 2AI to reactions containing *rC to minimize the accumulation of bridged dinucleotides (15), which resulted in a sustained slow rate of primer extension (Figure 2D–G). To obtain a valid comparison of reaction rates with *rC and rC*rC, we first measured initial rates with 20 mM rC*rC ($k = 2.5(1) \times 10^{-2}$ h^{-1}) and with 20 mM *rC ($k = 5(2) \times 10^{-3}$ h^{-1}). To eliminate the effect of rC*rC formation in the reaction with *rC, we measured the reaction rate with 20 mM *rC in the presence of excess 2AI, and observed an even lower rate of extension ($<0.2 \times 10^{-3}$ h^{-1}) (Figure 2F). We observed a similar pattern of reactivity with activated guanosine nucleotides (Supplementary Figure S2). The primary role of bridged dinucleotides in non-templated primer extension is evident from our observation that the initial reaction rates are substantially elevated in reaction mixtures containing bridged dinucleotides (rC*rC and *rC alone) compared to those with minimal bridged dinucleotide formation (*rC + 2AI) (Figure 2D, F). The reactivity difference stems from the fact that bridged dinucleotides do not require protonation of the leaving group (Figure 2B), whereas activated mononucleotides require protonation (16) (Figure 2A). This is consistent with a comparison of their hydrolysis rates: the rate of hydrolysis of rC*rC is 0.105(7) h^{-1} while the hydrolysis of *rC is significantly lower at $1.96(6) \times 10^{-3}$ h^{-1} at pH 8.0 (Supplementary Figure S3). We conducted non-templated primer extension experiments as a function of pH to determine if increased protonation of the leaving group at lower pH would accelerate the reaction rate. Our results show that the rates for the reactions involving 5 mM *rG and 25 mM 2AI at pH 7, 8, and 9 are $1.9(3) \times 10^{-4}$, $5.3(6) \times 10^{-4}$ and $8.1(4) \times 10^{-4}$ h^{-1}, respectively (Supplementary Figure S4). The decreased reaction rate at lower pH is likely due to decrease in deprotonation of the 3′-OH of the primer.

We followed the distribution of activated monomers and bridged dinucleotides as a function of time under primer extension reaction conditions by ^{31}P NMR (Supplementary Figure S5). In a reaction mixture that initially contained 20 mM bridged dinucleotides, the concentration of bridged dinucleotides declined due to hydrolysis, with a half-life of 7 h (Supplementary Figure S3A). In contrast, a reaction initiated with 20 mM of *rC alone showed a rapid initial increase in the concentration of rC*rC to an expected equilibrium value of approximately 3 mM, followed by a long slow decline to 2.5 mM by 8 h. In contrast, the mixture of 20 mM *rC with 100 mM 2AI contained <0.25 mM of bridged dinucleotides at all times. In all cases the concentration of activated mononucleotides decreased very slowly with time, with

an estimated half life of ∼350 h (Supplementary Figure S3B). This concentration of bridged dinucleotides in these different reaction mixtures correlates well with the initial rates of primer extension in reactions initiated with either bridged dinucleotides or activated mononucleotides. After the initial equilibration period, the hydrolysis of bridged dinucleotides into activated mononucleotides and unactivated nucleotide monophosphates results in a slowly declining reaction rate (Figure 2E, G). Consequently, kinetic and NMR data support the conclusion that bridged dinucleotides are the primary substrates for non-templated primer extension.

Comparison of non-templated primer extension with activated ribo-, arabino- and threo- nucleotides

Given that ribo-, arabino- and threo- nucleotides might all be synthesized together in a prebiotic environment, we wanted to evaluate whether the intrinsic chemical and structural differences between these nucleotide analogs would affect their incorporation into oligonucleotides. To achieve this, we added activated ribo-, arabino- and threo- cytidine mononucleotides (*rC, *araC and *tC) separately to the primer (Figure 3A). We used activated mononucleotides because they naturally form bridged dinucleotides and their rate of reaction is only modestly lower than pure bridged dinucleotides in an extended time course ($k = 4.4(4) \times 10^{-3}$ h^{-1} versus $1.2(1) \times 10^{-2}$ h^{-1}). The non-templated primer extension rate using *rC was found to be slightly faster at $1.9(1) \times 10^{-3}$ h^{-1} than the extension rate observed when only *araC was used, which was $1.5(1) \times 10^{-3}$ h^{-1} (Figure 3B, C). However, *tC did not yield detectable extension. A similar trend was observed with activated ribo-, arabino-, and threo-guanosine mononucleotides (*rG, *araG and *tG) (Figure 3D–F). The *rG resulted in the fastest extension rate of $2.2(1) \times 10^{-3}$ h^{-1}, closely followed by *araG at $1.8(1) \times 10^{-3}$ h^{-1}. *tG had the slowest extension rate at $3.6(8) \times 10^{-4}$ h^{-1}, which was approximately 6-fold slower than that with *rG (Figure 3F).

Arabino-nucleotides are structurally similar to ribonucleotides, with the only difference being the stereochemistry of the sugar 2′-OH group. Kim *et al.* (18) showed that this difference does not impact the stability or formation of bridged dinucleotides under primer extension conditions. The rates of hydrolysis of bridged ribonucleotides (rA*rA) and arabino-nucleotides (araA*araA) are similar, at $k = 0.31$ h^{-1} and $k = 0.33$ h^{-1}, respectively. On the other hand, threo-nucleotides lack the 5′-methylene carbon found in ribo-nucleotides, resulting in a backbone repeat unit that is one atom shorter in TNA than in RNA. In addition, the rate of formation of bridged threo-nucleotides (tC*tC) is 8.8×10^{-4} h^{-1} mM^{-1} (20), while that of bridged ribonucleotides (rC*rC) is 4.5×10^{-3} h^{-1} mM^{-1} (15). The formation of 3′-3′ imidazolium-bridged threo-nucleotides is impeded by steric hindrance, in contrast to the facile formation of 5′-5′ imidazolium-bridged ribonucleotides (Supplementary Figure S6). This increased steric hindrance accounts for the observed 5-fold reduction in the rate of formation of bridged threo-nucleotides. In the non-templated primer extension reaction, the slower rate of formation of bridged threo-nucleotides and the resulting lower steady state level of bridged dinucleotides would be expected to result in a slower primer extension rate as non-templated primer extension primarily occurs via bridged dinucleotides.
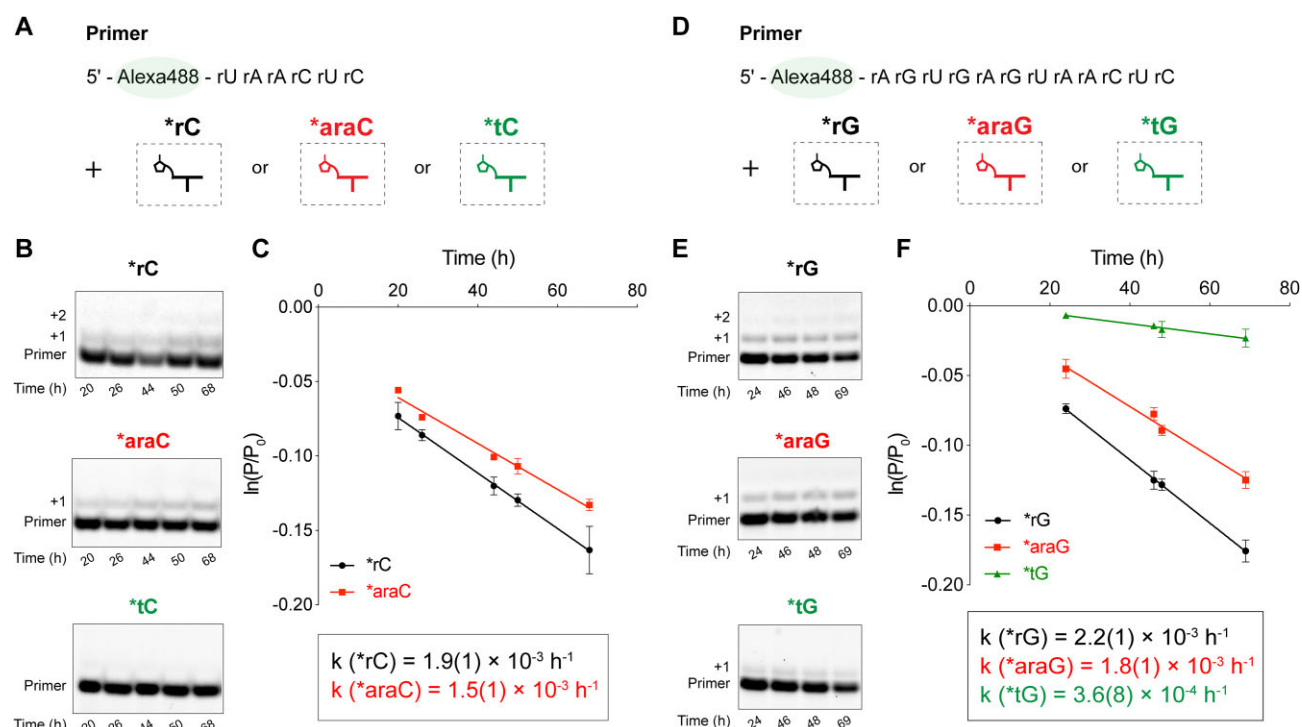
Figure 3. The intrinsic reactivities of activated ribo-, arabino- and threo-nucleotides in non-templated primer extension with (A–C) cytidine and (D–F) guanosine mononucleotides. (**A, D**) Schematic representation of non-templated primer extension reaction using an Alexa488-labeled RNA primer with the addition of activated ribo-, arabino- or threo-nucleotides, respectively. (**B, E**) Gel electrophoresis images. (**C, F**) Kinetic analysis with observed pseudo-first-order reaction rates (k_{obs}). Reaction conditions: 1 μM primer (XJ-Alexa-6mer or XJ-Alexa-12mer, Supplementary Table S1), 200 mM HEPES at pH 8.0 and 50 mM $MgCl_2$ with 20 mM of *rC or *rG, *araC or *araG, and *tC or *tG, respectively. Error bars represent standard deviations from the mean, $n = 2$ replicates. P/P_0 denotes the ratio of the unextended primer (P) to the initial total primer (P_0).

Non-templated primer extension with mixture of activated ribo-, arabino- and threo-nucleotides

To simulate the diverse chemical pool of activated nucleotides that could have co-existed on the early Earth, we proceeded to study the products of non-templated primer extension using a mixture of activated ribo-, arabino- and threo-nucleotides. To analyze the product distribution from these competition reactions, we used liquid chromatography-mass spectrometry (LC–MS). To distinguish the otherwise isobaric ribonucleotides and arabino-nucleotides, we used stable isotope-labeled (^{13}C, ^{15}N) ribonucleotides.

To accelerate the intrinsically slow non-templated reactions, we used a spontaneous air-drying method which reduces the time required to observe appreciable product yield. Spontaneous air-drying is a prebiotically plausible scenario driven by environmental factors such as fluctuations in temperature and humidity, precipitation and evaporation, day-night cycles and the presence of hydrothermal vents (26). Dry-down has been extensively used to model potentially prebiotic conditions in the field of prebiotic chemistry (7,27). We left the reaction mixture in an uncapped tube on the bench top to allow water to evaporate for 24 h (Supplementary Figure S1). We observed minimal variability within the same set of experiments by this method (Supplementary Figure S7). However, it is important to note that variability across different sets of experiments could be significant due to the variable flow rate and moisture level of the ambient air (Supplementary Table S4). The heterogeneity of the early Earth's environment would likely provide more variable conditions than found in a confined system or an aqueous environment (7). To maximize the amount of +1 products available for accurate LC–MS char-

acterization, we screened a range of oligonucleotide concentrations for air-drying experiments and chose 100 μM as the starting concentration for spontaneous air-drying and competition experiments (Supplementary Figure S8).

Incorporation of ribo-, arabino- and threo-nucleotides in competition experiments

We examined non-templated primer extension in the presence of different ratios of ribo-, arabino-, and threo-nucleotides and analyzed the products of primer extension by LC–MS (Figure 4A). In a mixture of activated mononucleotides in a 1:1:1 ratio (*rC:*araC:*tC = 1:1:1), we observed both the primer and the + 1 products in the total compound chromatogram (TCC) (Figure 4B). To distinguish between the + 1 extended products for each mononucleotide, we used extracted compound chromatograms (ECC) for the expected extended products and overlaid them with the +1 product region of the TCC (Figure 4C). We confirmed that each peak in the ECC corresponded to the expected +1 products (Supplementary Table S2). We observed the expected mass-to-charge ratios (m/z) of the different +1 products and their salt adducts (Supplementary Figure S9), demonstrating that each species could be easily distinguished. We calculated the percentage of each nucleotide in the observed +1 products (Figure 4D, Supplementary Table S4A), revealing that arabino-nucleotide incorporation ($44.7 \pm 3.5\%$) was the highest, followed by ribonucleotide ($30.9 \pm 1.1\%$) and threo-nucleotide incorporation ($24.4 \pm 2.5\%$). The slightly higher incorporation profile of arabino-nucleotide is unlikely to be related to the formation of an RNA dimer (prCprC), as only a trace amount (1% of the initial input concentra-
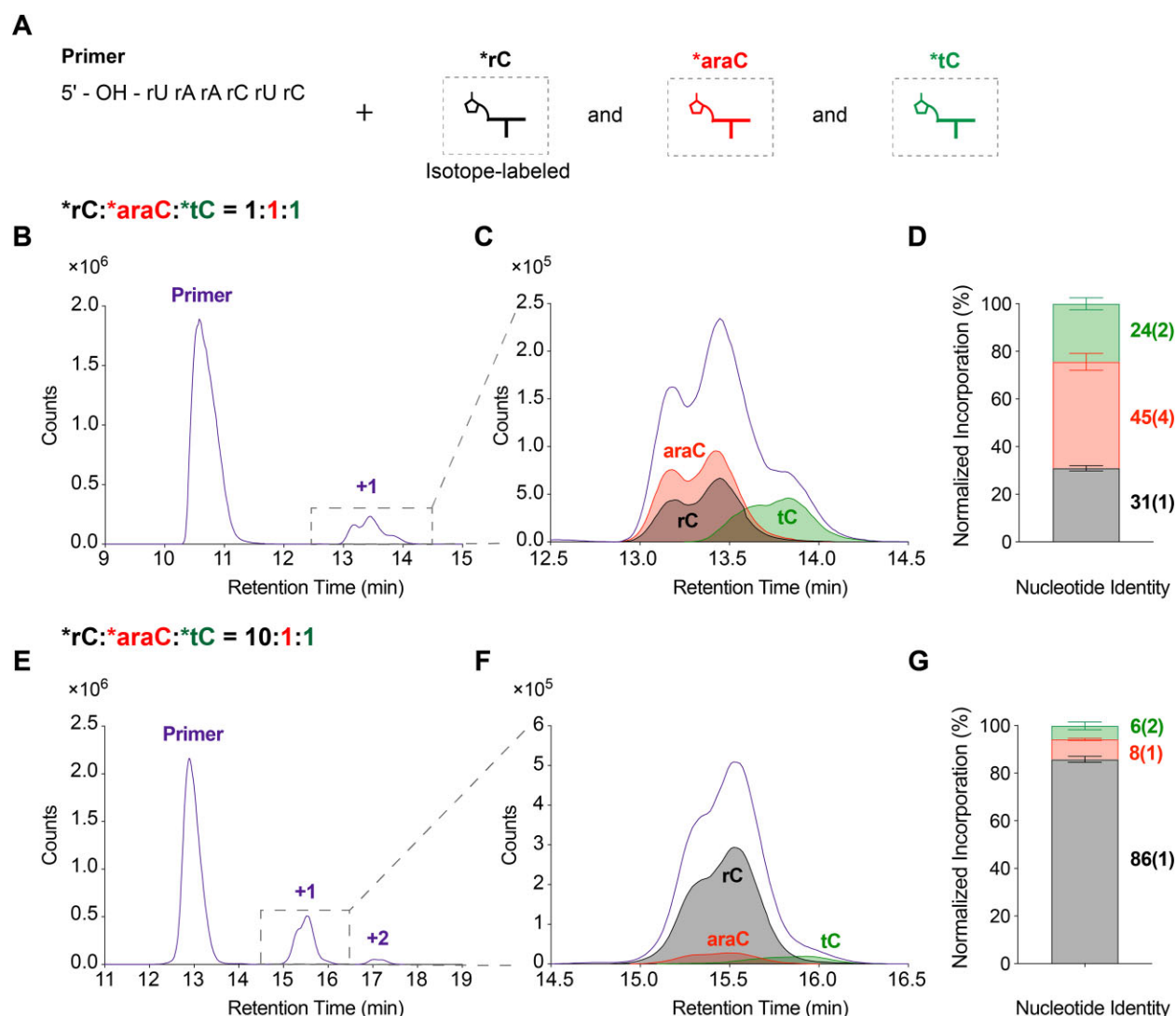
Figure 4. Incorporation of ribo-, arabino- and threo-nucleotides in competition experiments. (**A**) Schematic representation. (**B–D**) *rC:*araC:*tC = 1:1:1 and (E-G) 10:1:1 competition experiments with (**B, E**) total compound chromatogram (TCC) of the + 1 extended products (purple); (**C, F**) overlay of extracted compound chromatograms (ECC) of the +1 extended products ending in rC (black), araC (red) and tC (green); (**D, G**) stacked barplot representing the normalized incorporation of different nucleotides. Error bars represent standard deviations from the mean, $n = 6$ replicates.

tion) was detectable by ^{31}P NMR after 72 h of reaction (Supplementary Figure S10).

We performed an additional competition experiment with *rC, *araC and *tC at a ratio of 10:1:1 to simulate the effect of enrichment for ribonucleotides during prebiotic synthesis (Figure 1). It has been observed that ribose aminooxazoline (RAO), the precursor to ribonucleotides, can crystallize out, leaving other aminooxazolines in solution (28). Nucleotides derived from crystallized RAO would be predominantly ribonucleotides. Furthermore, homochiral RAO crystals can form through crystallization on magnetite (Fe_3O_4) surfaces, which induce chiral symmetry-breaking (29). The crystallization of enantiopure RAO, in turn, triggers the avalanche magnetization of magnetite, forming a closed feedback loop between chiral molecules and magnetite surfaces (30). Alternatively, enantiopure RAO can be obtained through physical and chemical amplification processes using enantioenriched amino acids (31). Taken together, these phenomena suggest that prebiotic mixtures of nucleotides might have been strongly enriched in ribonucleotides. We therefore sought to

investigate how changes in the stoichiometry of prebiotic mixtures, particularly the enrichment of ribonucleotides, would influence the composition of the products in downstream nontemplated polymerization.

In competition reactions initiated with *rC:*araC:*tC at a 10:1:1 ratio, we observed primer, +1 and +2 products (Figure 4E). By overlaying the TCC of the +1 products with the ECC of the expected products (Figure 4F), we found that ribonucleotide incorporation was the highest (85.8 ± 1.3%), followed by arabino-nucleotide (8.4 ± 0.4%), and then threo-nucleotide incorporation (5.8 ± 1.6%) (Figure 4G, Supplementary Table S4A). The ratio of incorporated ribonucleotide to arabino-nucleotide was ~10 to 1 in the + 1 products, while threo-nucleotide incorporation was slightly disfavored. Analysis of the +2 products revealed that the predominant product is the primer extended by two rC nucleotides (83.9 ± 1.1%), followed by extension with a rC and then araC (11.5 ± 1.2%), and then by extension with a rC and a tC (4.5 ± 0.1%) (Supplementary Table S5). These results suggest that the composition of non-templated extension prod-

ucts will approximately mirror the input ratios of activated ribo-, arabino- and threo-mononucleotides, with a moderate selection bias against threo-nucleotides.

Non-templated reaction at both internal and terminal hydroxyls

The chromatograms shown in Figure 4 exhibited a double peak pattern for +1 products in both the TCC and the ECC. To trace the origins of the two peaks in the chromatograms, we considered all possible reaction sites in an RNA oligomer. Nucleotides contain two major nucleophilic groups, of which the 2′-hydroxyl group is more reactive than the 3′-hydroxyl (32,33). During non-templated primer extension with activated species, oligonucleotides with mixed 2′-5′ and 3′-5′ phosphodiester linkages can form. We first examined the possibility of the double peak pattern arising from a mixture of terminal 3′-5′ and 2′-5′ phosphodiester linkages by co-injecting oligonucleotides synthesized with terminal 3′-5′ and 2′-5′ linkages from the corresponding phosphoramidites at 1:1 and 1:10 ratios. These terminal linkage regioisomers could not be resolved by LC-MS (Supplementary Figure S11A, B), showing that terminal regioisomers were not responsible for the observed double peak pattern.

To narrow down the search for the cause of the double peak pattern, we performed a series of non-templated reactions with various primer constructs. We first examined the possibility of reaction at the 5′-OH group of the primer by carrying out a dry-down reaction with a 5′-hexynyl DNA primer and a 5′-OH DNA primer, each with a terminal dideoxy nucleotide. In both cases, no extension products were observed by LC–MS (Supplementary Figure S11C, D). This result suggested that at least one of the two HPLC peaks might have derived from reaction at internal 2′-hydroxyls.

Further tests with modified RNA primers confirmed that the internal 2′-OHs in RNA oligomers can react with activated nucleotides (Figure 5A-D, Supplementary Figure S12), causing the double peak pattern. Comparing the TCC and ECC of the non-templated products of a 5′-hexynyl RNA primer ending with dideoxycytosine (Supplementary Figure S12B) and an all RNA 5′-hexynyl primer (Supplementary Figure S12C) shows that both internal and terminal hydroxyls can react under spontaneous air-drying conditions. The experimentally observed ratio of internal reaction to terminal extension was ~1:1.6 (Figure 5D). Since there are five internal hydroxyls, but only two terminal hydroxyls (Supplementary Figure S12A), on average the internal hydroxyls must have only ~25% of the reactivity of the terminal hydroxyls (Supplementary Table S4B, Figure 5D). In the presence of complementary short oligomers, total internal hydroxyl reactivity was reduced to 21(1)% (Figure 5E–H), corresponding to an average ratio of internal to terminal hydroxyl reactivity of 10(1)% (Supplementary Table S4B, Figure 5H). This observation suggests that internal 2′-OH modification would be reduced in the presence of oligonucleotides.

Effect of an arabino-nucleotide at the 3′-end of the primer

Given the comparable levels of incorporation of ribo- and arabino-nucleotides in +1 extended products of an RNA primer in competition experiments, we asked whether the incorporation of an arabino-nucleotide would affect subsequent extension. We therefore measured the rate of non-templated primer extension of a FAM-labeled RNA primer ending in either rC or araC. We carried out these experiments in the pres-

ence of a complementary RNA oligomer to minimize the formation of 2′-branched products (Figure 6A, Supplementary Table S1). The rate of the rC-terminated primer extension was $4.1(2) \times 10^{-3}$ h^{-1} while extension of the araC-terminated primer was barely detectable (Figure 6B). Thus the incorporation of an arabino-nucleotide into the growing chain strongly impedes further primer extension, similar to its chain-terminating effect in templated primer extension (18). Based on this finding, we can infer the order of incorporation for multiple-nucleotide extension products containing both nucleotides: rC is incorporated first, followed by araC in the extended +2 product profile (Supplementary Table S5).

Competition experiments with a 5′-hexynyl DNA primer with a 3′-terminal ribonucleotide

To investigate the non-templated addition of ribo-, arabino- and threo-nucleotides specifically to the primer terminus, we carried out dry-down reactions using a 5′-hexynyl DNA primer with a 3′-terminal ribonucleotide, so that only the terminal diol is available for reaction (Figure 6C). We then analyzed the observed masses of + 1 products (Supplementary Table S3). The m/z profiles of all +1 products and their salt adducts were well separated, allowing us to distinguish the different species (Supplementary Figure S13). ECCs generated from these LC–MS datasets showed no peak doublet pattern in any of the three +1 products corresponding to ribo-, arabino-, and threo-nucleotide incorporation, at input ratios of *rC:*araC:*tC of 1:1:1 (Figure 6D, E) or 10:1:1 (Figure 6F, G). Consistent with the trends observed in the previous competition experiments, at a 1:1:1 input ratio, the modified primer results also showed that ribonucleotides and arabino-nucleotides exhibited comparable levels of incorporation ($39.0 \pm 1.6\%$ and $45.0 \pm 2.3\%$, respectively). As before, threo-nucleotide incorporation was significantly reduced ($16.1 \pm 2.3\%$) (Figure 6E, Supplementary Table S4C). To simulate the enrichment of ribonucleotides in prebiotic mixtures, we again changed the input ratio of *rC:*araC:*tC to 10:1:1. Upon the analysis of the resulting +1 products, ribonucleotide incorporation was predominant ($89.2 \pm 0.2\%$), followed by arabino-nucleotide ($8.3 \pm 0.5\%$), and then by threo-nucleotide incorporation ($2.6 \pm 0.4\%$) (Figure 6G, Supplementary Table S4C).

Discussion

Potentially prebiotic synthetic routes suggest that ribo-, arabino- and threo-nucleotides may have been synthesized together on the early Earth. Once activated, these nucleotides could have co-polymerized to form oligomers, resulting in a heterogenous mixture of strands capable of taking part in template-directed nonenzymatic copying reactions and potentially cycles of replication (Figure 7). Despite the significance of non-templated polymerization in understanding the emergence of RNA from prebiotic mixtures, this step has been poorly studied. We have investigated the mechanism of non-templated primer extension and its behavior with activated ribo-, arabino- and threo-nucleotides. In addition, we performed one-pot competition experiments with mixtures of these activated nucleotides at varying input ratios to simulate the content of prebiotic mixtures and to monitor the composition of the resulting non-templated primer extension products.

Consistent with nonenzymatic template-directed copying, non-templated primer extension primarily occurs through
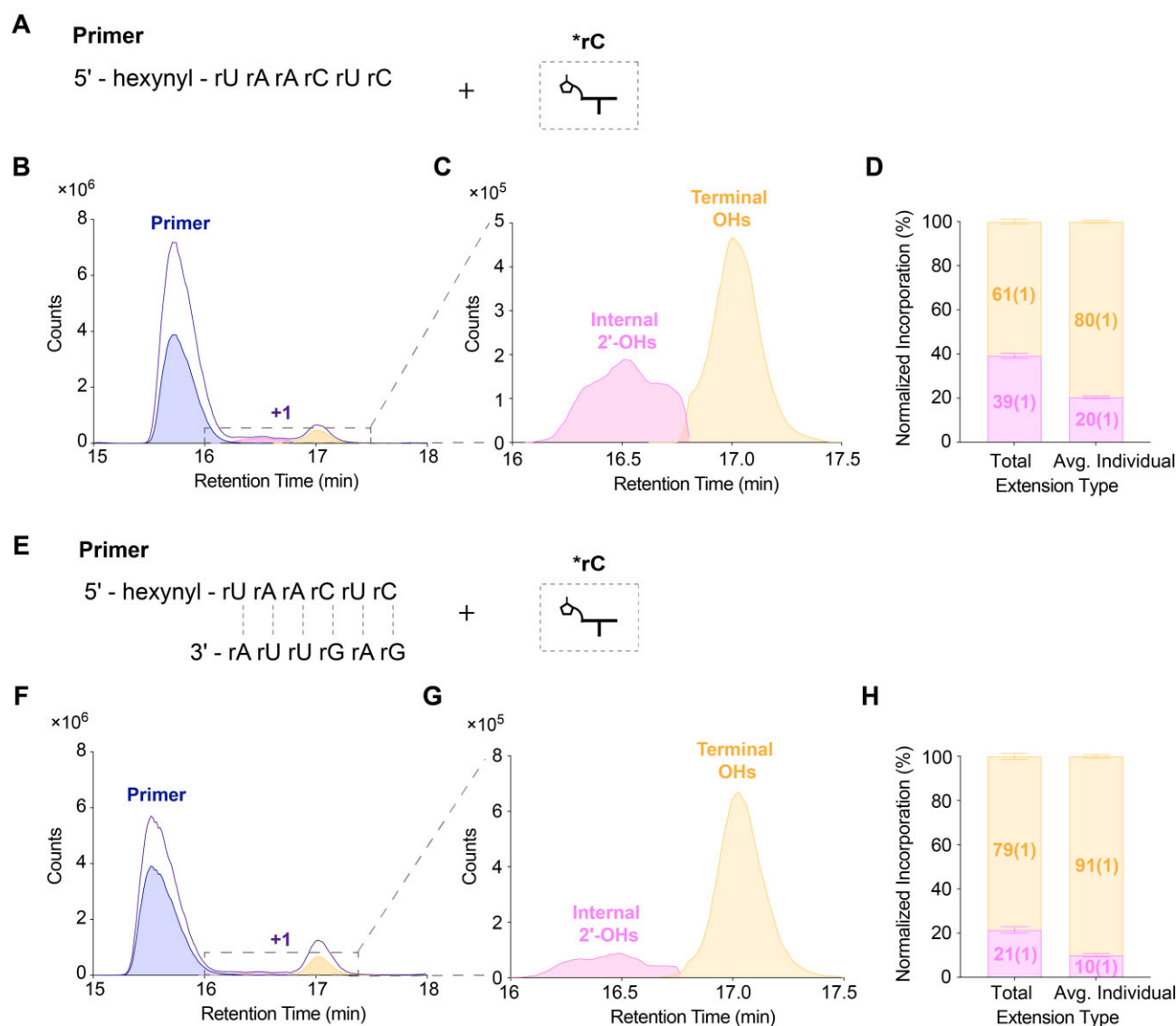
Figure 5. Complementary oligomers reduce the reactivity of internal 2′-OHs. Non-templated primer extension of a 5′-hexynyl RNA primer in the absence (A-D) or presence (E-H) of a complementary oligomer. (**A, E**) Schematic representation; (**B, F**) overlay of TCC (purple) and ECC (blue: primer; pink: internal 2′-OHs; yellow: terminal OHs); (**C, G**) overlay of ECC corresponding to reaction of internal 2′-OHs or terminal OHs; (**D, H**) stacked barplots representing the normalized incorporation at internal 2′-OHs or terminal-OHs in total versus average individual reaction. Error bars represent standard deviations from the mean, $n = 6$ replicates.

imidazolium-bridged dinucleotides. We observed a much faster rate of non-templated primer extension with bridged dinucleotides than with activated mononucleotides (Figure 2F, G). The superior leaving group of the bridged dinucleotides makes them the primary substrates in both non-templated and templated primer extension. The actual reaction rates varied with time (Figure 2D, E), because of the time required for formation of bridged dinucleotides from monomers and the more rapid hydrolysis of bridged dinucleotides than activated monomers (Supplementary Figure S3). In a prebiotic scenario, the extent of non-templated oligomerization would depend strongly on the effectiveness of the ambient activation chemistry, as well as on environmental factors such as pH and temperature that would affect both activation and hydrolysis.

Contrary to individual reactivity assays where threo-nucleotide incorporation is disfavored by a factor of 6 relative to ribonucleotides (Figure 3F), competition experiments show a reduced bias against threo-nucleotides (Figure 4, 6). A possible explanation for this difference is that in a mixture, hetero-bridged dinucleotides such as rN*tN and araN*tN (Supplementary Table S6) would form more readily than tN*tN, due to the less hindered attack by the 2AI moiety of a *tN on the phosphate groups of *rN or *araN nucleotides. By mitigating the sterically crowded nature of threo-nucleotides, the formation of hetero-bridged dinucleotides may facilitate the incorporation of threo-nucleotides in non-templated reactions. This mechanism is consistent with the observation that activated downstream oligoribonucleotides accelerate template-directed primer extension with threo-nucleotides (20).

The poor regioselectivity of non-templated polymerization results in significant formation of 2′-5′ linkages (34). In addition, internal 2′ hydroxyl groups in the primer can react with activated nucleotides forming branched structures (Figure 5).
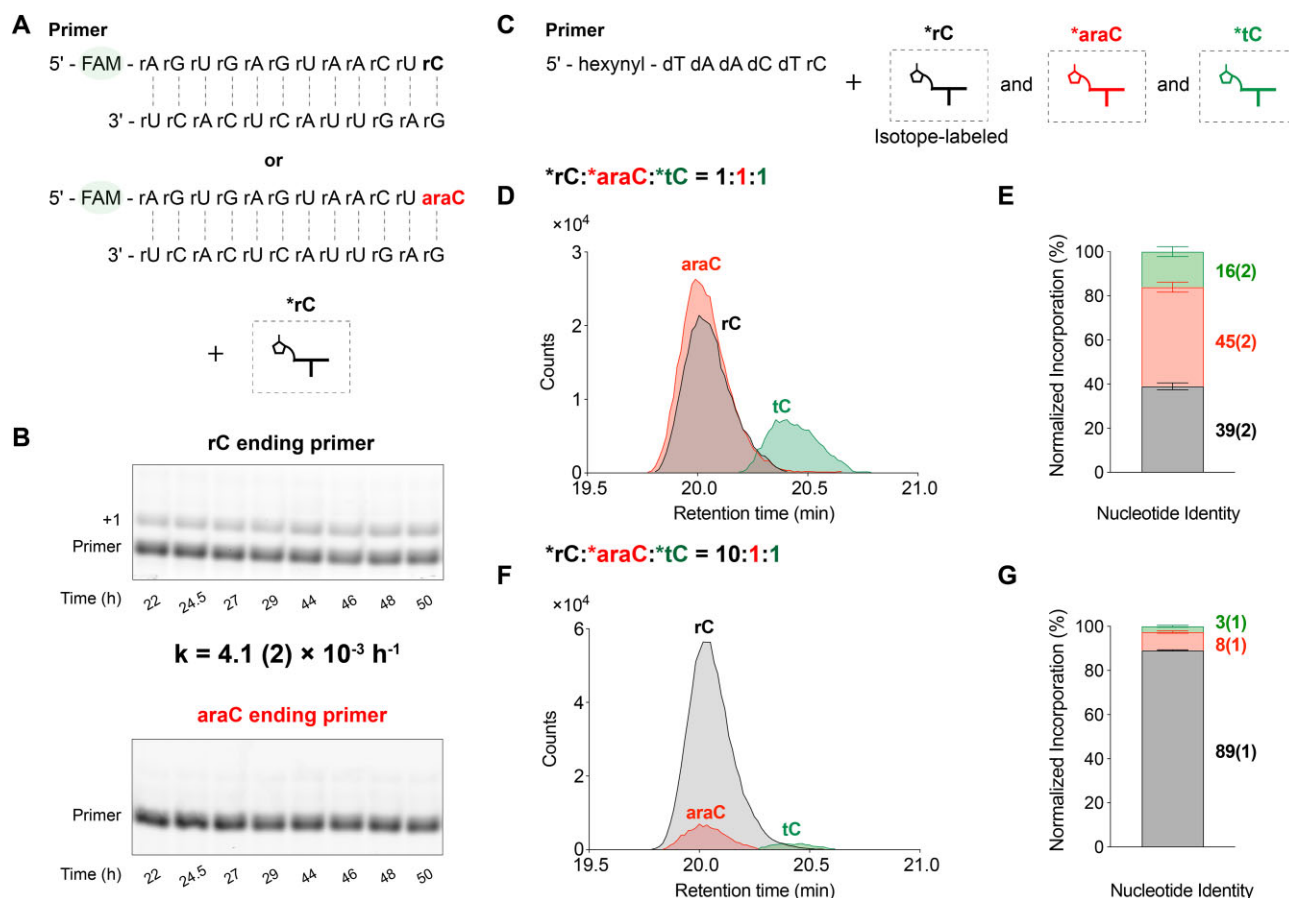
Figure 6. Nucleotide selectivity of non-templated primer extension. (**A**) Schematic representation of nonenzymatic primer extension with a primer that contains either a terminal ribonucleotide or a terminal arabino-nucleotide. The presence of a complementary oligonucleotide minimizes the formation of the 2′-branched products. (**B**) Gel electrophoresis images and rate for primer extension. Error represents standard deviation from the mean, $n = 3$ replicates. (**C**) Schematic representation of the competition experiments with (D, E) *rC:*araC:*tC = 1:1:1 and (F, G) 10:1:1. The 5′-hexynyl DNA primer with a 3′-terminal ribonucleotide was used to avoid 5′ and 2′ modifications. (**D, F**) Overlay of the ECC of the +1 extended products ending in rC (black), araC (red), and tC (green); (**E, G**) stacked barplots representing the normalized incorporation percentage of different nucleotides. Error bars represent standard deviations from the mean, $n = 6$ replicates.

Our competition experiments suggest that sugar heterogeneity will be enhanced by the copolymerization of mixed nucleotides (Figure 4, 6). However, incorporating an arabinonucleotide at the 3′-end of the primer caps the strand and blocks further elongation (Figure 6 A, B), potentially inhibiting the formation of oligonucleotides long enough to exhibit catalytic functions. Moreover, this initial heterogeneity is likely to be reduced during subsequent template-directed copying steps. For example, a template containing a 2′-5′ linkage is copied at a reduced rate but no detectable 2′-5′ linkages are formed in the product (35). Similarly, arabino- and threonucleotides are outcompeted by ribonucleotides in template directed copying experiments (18,20). Moreover, the inability of ANA to form stable duplexes with complementary ANA and the weaker pairing with complementary RNA sequences may introduce a further bias against ANA (36–38). On the other hand, a degree of backbone heterogeneity is compatible with RNA folding into functional structures and might also be helpful for strand separation by lowering the melting temperature of product duplexes, thus facilitating subsequent cycles of primer extension (39,40). 2′-5′ branched RNA structures (41) form biologically as the lariat products of RNA splicing (42), suggesting that branched RNAs could also have prebiotic

roles. In the presence of complementary oligomers, the probability of reactions at internal 2′-OHs decreases markedly (Figure 5H, Supplementary Table S4B). While branched products may not be ideal for propagating genetic information, they might still serve as splints to assist template-directed copying (43,44) and RNA replication under the Virtual Circular Genome (VCG) Model (45).

In summary, our study strengthens our understanding of the physico-chemical selection steps that led to the emergence of RNA from prebiotic mixtures. The copolymerization of mixtures of different types of nucleotides results in extension products that roughly mirror the input ratio of nucleotides. This process results in a modest bias against threo-nucleotide incorporation, although the presence of ribo- and arabinonucleotides actually mitigates the intrinsic bias against threonucleotides. The incorporation of arabino-nucleotides into the growing chain inhibits further primer extension. As a result, bias in favor of ribonucleotides during synthesis would be retained during non-templated polymerization, and then more strongly enhanced during subsequent template-directed reactions (18,20). The combined selection in favor of ribonucleotides across sequential physical and chemical steps likely resulted in RNA-enriched oligomers after successive rounds
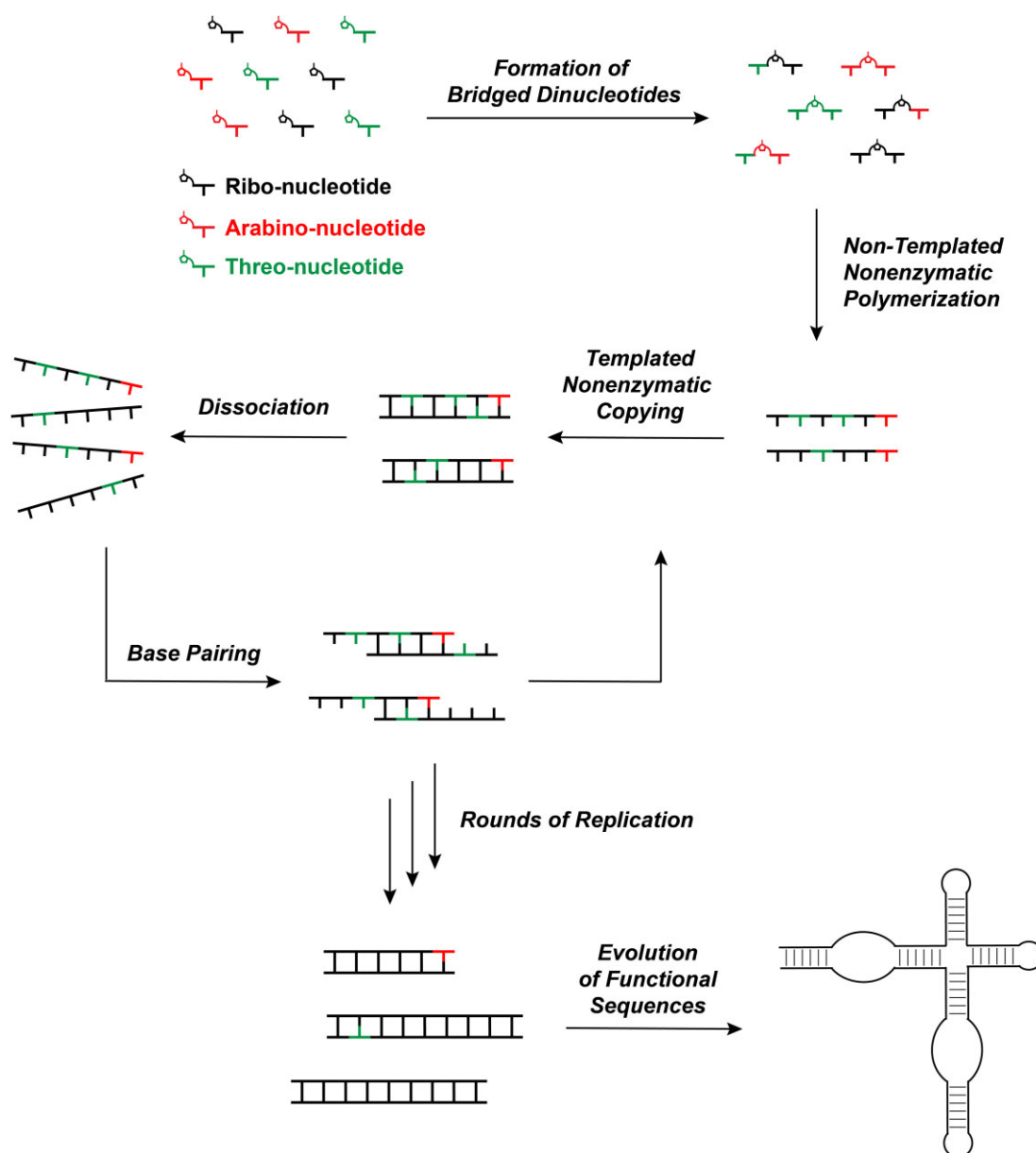
Figure 7. Schematic diagram illustrating pathways for the emergence of RNA from prebiotic mixtures and evolution of functional sequences.

of replication, setting the stage for the evolution of complex, functional RNA sequences (Figure 7).

Data availability

The data underlying this article are available in the article and in its online Supplementary material.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

The authors thank Professor John C. Chaput for providing the threo-nucleosides. We thank Dr Victor Lelyveld, Dr Harry Aitken and Dr Seohyun Chris Kim for helpful discussions and technical assistance. We thank Dr Victor Lelyveld, Dr Saurja DasGupta and Dr Filip Bošković for comments on the manuscript.

Conflict of interest statement

None declared.

References

1. Szostak,J.W., Bartel,D.P. and Luisi,P.L. (2001) Synthesizing life. *Nature*, **409**, 387–390.

2. Orgel,L.E. (1968) Evolution of the genetic apparatus. *J. Mol. Biol.*, **38**, 381–393.

3. Gilbert,W. (1986) Origin of life: the RNA world. *Nature*, **319**, 618–618.

4. Powner,M.W., Gerland,B. and Sutherland,J.D. (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, **459**, 239–242.

5. Roberts,S.J., Szabla,R., Todd,Z.R., Stairs,S., Bučar,D.-K., Šponer,J., Sasselov,D.D. and Powner,M.W. (2018) Selective prebiotic conversion of pyrimidine and purine anhydronucleosides into Watson-Crick base-pairing arabino-furanosyl nucleosides in water. *Nat. Commun.*, **9**, 4073.

6. Colville,B.W.F. and Powner,M.W. (2021) Selective prebiotic synthesis of α-threofuranosyl cytidine by photochemical anomerization. *Angew. Chem. Int. Ed.*, **60**, 10526–10530.

7. Gibard,C., Bhowmik,S., Karki,M., Kim,E.-K. and Krishnamurthy,R. (2018) Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nat. Chem.*, **10**, 212–217.

8. Jiménez,E.I., Gibard,C. and Krishnamurthy,R. (2021) Prebiotic phosphorylation and concomitant oligomerization of deoxynucleosides to form DNA. *Angew. Chem. Int. Ed.*, **60**, 10775–10783.

9. Fernández-García,C., Grefenstette,N.M. and Powner,M.W. (2017) Prebiotic synthesis of aminooxazoline-5′-phosphates in water by oxidative phosphorylation. *Chem. Commun.*, **53**, 4919–4921.

10. Zhang,S.J., Duzdevich,D. and Szostak,J.W. (2020) Potentially prebiotic activation chemistry compatible with nonenzymatic RNA copying. *J. Am. Chem. Soc.*, **142**, 14810–14813.

11. Zhang,S.J., Duzdevich,D., Ding,D. and Szostak,J.W. (2022) Freeze-thaw cycles enable a prebiotically plausible and continuous pathway from nucleotide activation to nonenzymatic RNA copying. *Proc. Natl. Acad. Sci. U.S.A.*, **119**, e2116429119.

12. Ferris,J.P., Hill,A.R., Liu,R. and Orgel,L.E. (1996) Synthesis of long prebiotic oligomers on mineral surfaces. *Nature*, **381**, 59–61.

13. Kanavarioti,A., Monnard,P.-A. and Deamer,D.W. (2001) Eutectic phases in ice facilitate nonenzymatic nucleic acid synthesis. *Astrobiology*, **1**, 271–281.

14. Li,L., Prywes,N., Tam,C.P., O'Flaherty,D.K., Lelyveld,V.S., Izgu,E.C., Pal,A. and Szostak,J.W. (2017) Enhanced nonenzymatic RNA copying with 2-aminoimidazole activated nucleotides. *J. Am. Chem. Soc.*, **139**, 1810–1813.

15. Walton,T. and Szostak,J.W. (2017) A kinetic model of nonenzymatic RNA polymerization by cytidine-5′-phosphoro-2-aminoimidazolide. *Biochemistry*, **56**, 5739–5747.

16. Walton,T., Zhang,W., Li,L., Tam,C.P. and Szostak,J.W. (2019) The mechanism of nonenzymatic template copying with imidazole-activated nucleotides. *Angew. Chem. Int. Ed.*, **58**, 10812–10819.

17. Fahrenbach,A.C., Giurgiu,C., Tam,C.P., Li,L., Hongo,Y., Aono,M. and Szostak,J.W. (2017) Common and potentially prebiotic origin for precursors of nucleotide synthesis and activation. *J. Am. Chem. Soc.*, **139**, 8780–8783.

18. Kim,S.C., Zhou,L., Zhang,W., O'Flaherty,D.K., Rondo-Brovetto,V. and Szostak,J.W. (2020) A model for the emergence of RNA from a prebiotically plausible mixture of ribonucleotides, arabinonucleotides, and 2′-deoxynucleotides. *J. Am. Chem. Soc.*, **142**, 2317–2326.

19. Kim,S.C., O'Flaherty,D.K., Giurgiu,C., Zhou,L. and Szostak,J.W. (2021) The emergence of RNA from the heterogeneous products of prebiotic nucleotide synthesis. *J. Am. Chem. Soc.*, **143**, 3267–3279.

20. Zhang,W., Kim,S.C., Tam,C.P., Lelyveld,V.S., Bala,S., Chaput,J.C. and Szostak,J.W. (2021) Structural interpretation of the effects of threo-nucleotides on nonenzymatic template-directed polymerization. *Nucleic Acids Res.*, **49**, 646–656.

21. Yoshikawa,M., Kato,T. and Takenishi,T. (1967) A novel method for phosphorylation of nucleosides to 5′-nucleotides. *Tetrahedron Lett.*, **8**, 5065–5068.

22. Ding,D., Zhou,L., Giurgiu,C. and Szostak,J.W. (2022) Kinetic explanations for the sequence biases observed in the nonenzymatic copying of RNA templates. *Nucleic Acids Res.*, **50**, 35–45.

23. Welch,C.J., Foley,J.P., Biba,M. and Mao,B. (2014) Liquid chromatography methods for the separation of short RNA oligonucleotides. *LCGC Europe*, **27**, 632–639.

24. Walton,T. and Szostak,J.W. (2016) A highly reactive imidazolium-bridged dinucleotide intermediate in nonenzymatic RNA primer extension. *J. Am. Chem. Soc.*, **138**, 11996–12002.

25. Kervio,E., Sosson,M. and Richert,C. (2016) The effect of leaving groups on binding and reactivity in enzyme-free copying of DNA and RNA. *Nucleic Acids Res.*, **44**, 5504–5514.

26. Fares,H.M., Marras,A.E., Ting,J.M., Tirrell,M.V. and Keating,C.D. (2020) Impact of wet-dry cycling on the phase behavior and compartmentalization properties of complex coacervates. *Nat. Commun.*, **11**, 5423.

27. Ritson,D.J., Battilocchio,C., Ley,S.V. and Sutherland,J.D. (2018) Mimicking the surface and prebiotic chemistry of early Earth using flow chemistry. *Nat. Commun.*, **9**, 1821.

28. Anastasi,C., Crowe,M.A., Powner,M.W. and Sutherland,J.D. (2006) Direct assembly of nucleoside precursors from two- and three-carbon units. *Angew. Chem. Int. Ed.*, **45**, 6176–6179.

29. Ozturk,S.F., Liu,Z., Sutherland,J.D. and Sasselov,D.D. (2023) Origin of biological homochirality by crystallization of an RNA precursor on a magnetic surface. *Sci. Adv.*, **9**, eadg8274.

30. Ozturk,S.F., Bhowmick,D.K., Kapon,Y., Sang,Y., Kumar,A., Paltiel,Y., Naaman,R. and Sasselov,D.D. (2023) Chirality-induced avalanche magnetization of magnetite by an RNA precursor. *Nat. Commun.*, **14**, 6351.

31. Hein,J.E., Tse,E. and Blackmond,D.G. (2011) A route to enantiopure RNA precursors from nearly racemic starting materials. *Nature Chem*, **3**, 704–706.

32. Lohrmann,R. and Orgel,L.E. (1978) Preferential formation of (2'–5')-linked internucleotide bonds in non-enzymatic reactions. *Tetrahedron*, **34**, 853–855.

33. Kanavarioti,A., Lee,L.F. and Gangopadhyay,S. (1999) Relative reactivity of ribosyl 2′-OH vs. 3′-OH in concentrated aqueous solutions of phosphoimidazolide activated nucleotides. *Orig Life Evol Biosph*, **29**, 473–487.

34. Sulston,J., Lohrmann,R., Orgel,L.E. and Miles,H.T. (1968) Nonenzymatic synthesis of oligoadenylates on a polyuridylic acid template. *Proc. Natl. Acad. Sci. U.S.A.*, **59**, 726–733.

35. Giurgiu,C., Li,L., O'Flaherty,D.K., Tam,C.P. and Szostak,J.W. (2017) A mechanistic explanation for the regioselectivity of nonenzymatic RNA primer extension. *J. Am. Chem. Soc.*, **139**, 16741–16747.

36. Martín-Pintado,N., Yahyaee-Anzahaee,M., Campos-Olivas,R., Noronha,A.M., Wilds,C.J., Damha,M.J. and González,C. (2012) The solution structure of double helical arabino nucleic acids (ANA and 2′F-ANA): effect of arabinoses in duplex-hairpin interconversion. *Nucleic Acids Res.*, **40**, 9329–9339.

37. Noronha,A.M., Wilds,C.J., Lok,C.-N., Viazovkina,K., Arion,D., Parniak,M.A. and Damha,M.J. (2000) Synthesis and biophysical properties of arabinonucleic acids (ANA): circular dichroic spectra, melting temperatures, and ribonuclease H susceptibility of ANA·RNA hybrid duplexes. *Biochemistry*, **39**, 7050–7062.

38. Watts,J.K., Martín-Pintado,N., Gómez-Pinto,I., Schwartzentruber,J., Portella,G., Orozco,M., González,C. and Damha,M.J. (2010) Differential stability of 2′F-ANA•RNA and ANA•RNA hybrid duplexes: roles of structure, pseudohydrogen bonding, hydration, ion uptake and flexibility. *Nucleic Acids Res.*, **38**, 2498–2511.

39. Engelhart,A.E., Powner,M.W. and Szostak,J.W. (2013) Functional RNAs exhibit tolerance for non-heritable 2′–5′ versus 3′–5′ backbone heterogeneity. *Nature Chem*, **5**, 390–394.

40. Sheng,J., Li,L., Engelhart,A.E., Gan,J., Wang,J. and Szostak,J.W. (2014) Structural insights into the effects of 2′-5′ linkages on the RNA duplex. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 3050–3055.

41. Smail,B.A., Clifton,B.E., Mizuuchi,R. and Lehman,N. (2019) Spontaneous advent of genetic diversity in RNA populations through multiple recombination mechanisms. *RNA*, **25**, 453–464.

42. Wang,Y. and Silverman,S.K. (2005) Efficient one-step synthesis of biologically related lariat RNAs by a deoxyribozyme. *Angew. Chem. Int. Ed.*, **44**, 5863–5866.

43. Zhou,L., O'Flaherty,D.K. and Szostak,J.W. (2020) Template-directed copying of RNA by non-enzymatic ligation. *Angew. Chem.*, **132**, 15812–15817.

44. Zhou,L., O'Flaherty,D.K. and Szostak,J.W. (2020) Assembly of a ribozyme ligase from short oligomers by nonenzymatic ligation. *J. Am. Chem. Soc.*, **142**, 15961–15965.

45. Zhou,L., Ding,D. and Szostak,J.W. (2020) The virtual circular genome model for primordial RNA replication. *RNA*, **27**, 1–11.