Spontaneous formation and base pairing of plausible prebiotic nucleotides in water

Brian J. Cafferty1,2, David M. Fialho1,2, Jaheda Khanam1,2, Ramanarayanan Krishnamurthy2,3 & Nicholas V. Hud1,2

The RNA World hypothesis presupposes that abiotic reactions originally produced nucleotides, the monomers of RNA and universal constituents of metabolism. However, compatible prebiotic reactions for the synthesis of complementary (that is, base pairing) nucleotides and mechanisms for their mutual selection within a complex chemical environment have not been reported. Here we show that two plausible prebiotic heterocycles, melamine and barbituric acid, form glycosidic linkages with ribose and ribose-5-phosphate in water to produce nucleosides and nucleotides in good yields. Even without purification, these nucleotides base pair in aqueous solution to create linear supramolecular assemblies containing thousands of ordered nucleotides. Nucleotide anomerization and supramolecular assemblies favour the biologically relevant β-anomer form of these ribonucleotides, revealing abiotic mechanisms by which nucleotide structure and configuration could have been originally favoured. These findings indicate that nucleotide formation and selection may have been robust processes on the prebiotic Earth, if other nucleobases preceded those of extant life.

1 School of Chemistry and Biochemistry, Georgia Institute of Technology, 901 Atlantic Drive, Atlanta, Georgia 30332, USA. 2 NSF-NASA Center for Chemical Evolution, Atlanta, Georgia 30332, USA. 3 Department of Chemistry, The Scripps Research Institute, La Jolla, California 92037, USA. Correspondence and requests for materials should be addressed to N.V.H. (email: hud@chemistry.gatech.edu).
The current role of mononucleotides and RNA polymers in numerous cellular functions gave rise to the long-standing hypothesis that these molecules were involved early in the emergence of life: the RNA World hypothesis. Supporting this hypothesis, model prebiotic reactions and analyses of carbonaceous meteorites provide evidence that the canonical nucleobases of RNA (adenine, guanine, cytosine, uracil) were likely present on the prebiotic Earth. In addition, progress has been made towards finding abiotic routes to ribose and related sugars from simple molecules (for example, formaldehyde, glyoxylate), as well as mechanisms for ribose phosphorylation and ribose selection from complex sugar mixtures. Nevertheless, despite decades of effort, the chemical origin of nucleosides and nucleotides (that is, nucleobases glycosylated with ribose and phosphorylated ribose) remains an unsolved problem. In the 1970s, Orgel and co-workers showed that adenosine (the nucleoside of adenine) can be formed in 1–5% yield if a solution of adenine and ribose is dried and heated, but no comparable reactions have been demonstrated for the other three canonical nucleosides. Frustrated by what became known as The Nucleoside Problem, Orgel proposed an alternative pathway for prebiotic pyrimidine nucleoside and nucleotide synthesis that bypassed glycosidic bond formation, a pathway in which the cytosine nucleoside is built stepwise on a sugar scaffold. This approach was furthered by Sutherland and co-workers, who developed a synthetic route to cytidine and uridine nucleotides starting from glycolaldehyde and glyceralddehyde. However, the necessity to temporally separate specific reagents and reaction steps caused some to question the relevance of this synthesis to the origin of RNA. Sutherland and co-workers have since proposed a spatially separated geochemical scenario that includes an ordered delivery of reagents on the prebiotic Earth that would coincide with the sequential steps of their pyrimidine nucleoside synthesis, a scenario that is initiated by meteorite impacts.

The persistent challenge of finding a simple, robust and plausible prebiotic route to the canonical nucleosides—juxtaposed with the exquisite functionality of RNA—have caused many researchers to consider RNA a product of chemical and/or biological evolution. Inspired by the possibility that RNA evolved from a proto-RNA with alternative nucleobases that more easily formed nucleosides, Miller and co-workers demonstrated that urazole (a triazole analog of uracil) is efficiently glycosylated by ribose in water. Subsequent demonstrations of nucleoside formation with different plausible prebiotic heterocycles suggest that other nucleosides may have been common on the prebiotic Earth. While encouraging, a model prebiotic reaction has yet to be reported that produces two extant nucleosides that form a Watson–Crick base pair or two noncanonical nucleosides. When combined, the nucleotides form supramolecular assemblies with Watson–Crick-like base pairs, even within the crude reaction mixtures. These assemblies are shown to preferentially incorporate and increase the fraction of the β-anomer of the melamine nucleotide over the α-anomer. These findings demonstrate prebiotically plausible mechanisms for the selection of nucleotides in both nucleobase and sugar structure.

Results

Spontaneous formation of nucleosides and nucleotides. Given previous reports that phosphorylated sugars can be produced in model prebiotic reactions, we were motivated to explore the potential for BA and melamine to be glycosylated by R5P, which could represent a model prebiotic route to nucleotides that can form base pairs that are similar to those formed by the canonical nucleobases (Fig. 1a). Nucleotides spontaneously form when BA is mixed in water with one equivalent of R5P at 20 °C (without the need for drying). The reaction is surprisingly efficient with BA + R5P conjugates exceeding 80% in the unpurified (crude) reaction mixture after 24 h and with significant yields for reactions performed over the range of pH 3–11 (Fig. 1b and Supplementary Fig. 1). Reactions between melamine and R5P are also productive, forming melamine + R5P conjugates at yields ranging from 33 to 55% when the reaction was carried out for 24 h at 65 °C (Fig. 1c). Glycosylation of melamine with R5P was observed from pH 3 to 9 and at 20 °C (Supplementary Figs 2 and 3). Glycosylation of BA and melamine by (unphosphorylated) ribose was also found to occur spontaneously in water, producing four nucleoside isomers for each reaction (Supplementary Figs 4 and 5). The robustness (for example, wide pH and temperature range) and good yields for both nucleoside and nucleotide formation with BA and melamine are noteworthy among model prebiotic reactions, especially considering that none of the four canonical nucleobases form nucleosides in detectable yields when heated with R5P in water (Supplementary Fig. 6). On the contrary, the canonical nucleosides and nucleotides of RNA are thermodynamically disfavoured (but kinetically stable) in water.

To characterize the nucleotides formed from the BA reactions with R5P, the products formed in water after 24 h at 20 °C were isolated by column chromatography for further analysis. Two-dimensional (2D) NMR spectroscopy confirmed that the BA + R5P conjugates are C-nucleotides, with a C-C glycosidic bond between ribose and BA (5-ribosuranosyl-C-barbiturate-5’-monophosphate, C-BMP; Fig. 2b and Supplementary Fig. 7). One-dimensional (1D) rotating frame NOE (ROE) spectroscopy confirmed that the β-anomer is preferentially formed in a 67:33 ratio over the α-anomer (Fig. 2c). Similar to our previous report that reactions between 2,4,6-triaminopyrimidine and ribose yield the C-linked β-ribosuranoside as the major product, BA reactions with R5P again demonstrate that the biologically significant β-anomeric sugar form can be preferentially selected through nucleosylation with an alternative nucleobase. Intermediate products suggest that the BA nucleotide-formation reaction proceeds through a Knoevenagel condensation (Supplementary Fig. 8a). We note that β-C-BMP shares a close structural relationship with the C-nucleotide pseudouridine, the most common posttranscriptional modification of RNA in biology.

We next isolated and characterized the melamine + R5P conjugates formed from a reaction between melamine and R5P.
that was performed at 65 °C over 24 h. 2D NMR analysis of the melamine + R5P conjugates confirmed that glycosylation of melamine occurs at an exocyclic amine (Fig. 3b and Supplementary Fig. 9). Glycosylation proceeds through a reversible Schiff base intermediate (Supplementary Fig. 8b), which is partially stabilized by ring closure, to form two N-nucleotides (N-ribofuranosyl-melamine-5′-monophosphate (MMP)). 1D ROE analysis confirms that the α- and β-anomers of MMP equilibrate to approximately equal amounts in aqueous solution (Fig. 3c). The stability of the glycosidic bond of exocyclic N-linked triaminotriazine nucleosides in water had previously been shown to be quite poor, with total hydrolysis occurring on the order of minutes. The hydrolytic stability of a 5 mM solution of MMP was evaluated to determine the rate of hydrolysis at pH 5 and 5.5°C. Under these conditions, the hydrolysis rate was determined to be 3 × 10⁻⁶ min⁻¹, with a half-life of 6 months (Supplementary Fig. 10). Remarkably, the dissociation constant between melamine and R5P was found to be 3.7 mM. For comparison, the work of Miller and co-workers revealed that the Kₐ for the ribonucleoside of urazole is 700 mM at 5°C, whereas the Kₐ for the ribonucleoside of uracil is estimated to be around 700 M. Thus, the equilibrium of nucleoside formation between melamine and ribose could be several orders of magnitude more favourable than that of the extant nucleobases.

**MMP and C-BMP form supramolecular assemblies.** We next tested the ability of MMP and C-BMP to exhibit base pairing in aqueous solution. Base pairing is not exhibited by the canonical mononucleotides, but such a property would have been advantageous for the prebiotic mutual selection and colocalization of base-pairing nucleotides, particularly if protobiopolymers emerged from a complex 'prebiotic soup'. The crude products of melamine and BA reactions with R5P were mixed and the mixtures were analysed by circular dichroism (CD) spectroscopy as an initial test of whether melamine and BA nucleotides can form chiral assemblies even without purification. To maximize assembly of the two heterocycles, solution pH was adjusted to between pH 4 and 5. This pH range was chosen due to pKₐ considerations of the parent heterocycles (BA pKₐ = 4; melamine pKₐ = 5), which would be expected to minimize the difference in relative ionization between the bases. This mixture of nucleotides, side products and unreacted starting materials from the two reactions (50 mM each in total BA and melamine) exhibits a substantial CD signal, whereas separately the crude products of each reaction exhibit either no CD signal (the melamine-R5P reaction) or a much lower CD signal and with a different wavelength profile (the BA-R5P reaction; Fig. 4a). The loss of the CD signal upon heating the mixture of the crude reactions above 30 °C, and return of signal upon cooling to 5 °C (Supplementary Fig. 11a) indicates the reversible formation of non-covalent assemblies.

The assemblies formed upon combining the crude reaction mixtures were imaged by atomic force microscopy (AFM), which revealed linear supramolecular polymers with diameters of ca 2 nm (Fig. 4b). These structures are fully consistent with the presence of stacked hydrogen-bonded hexads with paired BA and melamine bases (Fig. 4f), assemblies that have been observed previously with analogous molecules that also form hexads. The length of these supramolecular polymers (typically >1 μm) indicates that tens of thousands of heterocycles are paired within...
The absence of H5-C5 correlations in the HSQC supports the C-nucleoside multiple bond correlation (HMBC) spectra showing 1H-13C couplings for proton–proton magnetization transfer (ROE) as shown in Figure 2 containing a 1:2 ratio of C-BMP nucleotides. 

Figure 2 | NMR characterization of C-BMP nucleotides. (a) Chemical structure of α-C-BMP and β-C-BMP with arrows indicating through-space proton–proton magnetization transfer (ROE) as shown in c. (b) Heteronuclear single-quantum correlation (HSQ) and heteronuclear multiple bond correlation (HMBC) spectra showing 1H-13C couplings for α-C-BMP and β-C-BMP. 1H-13C correlation observed between H1' and C5 of β-C-BMP and H1'-C4/C6 of α-C-BMP and β-C-BMP in the HMBC as well as the absence of H5-C5 correlations in the HSQ spectrum support the C-nucleoside assignment. α-Anomer cross peaks are shown in red, β-anomer cross peaks in blue and overlapping cross peaks in purple. 1H Correlation spectroscopy (COSY) spectrum of a mixture of C-BMP nucleotides is provided in the Supplementary Information. (c) 1H NMR and 1D ROE spectra of a solution containing a 1:2 ratio of α-C-BMP to β-C-BMP. (Top) 1H NMR spectrum with resonance assignments as indicated in a. (Middle) Irradiation of the H1' of α-C-BMP results in through space magnetization transfer to the H2' and H3' of α-C-BMP. (Bottom) Irradiation of the H1' of β-C-BMP results in through space magnetization transfer to the H2' and H4' of β-C-BMP. * Indicates TOCSY transfer from β-H1' to β-H2'.

C-BMP and MNP nucleotides also form water-soluble supramolecular assemblies when mixed with free melamine and free BA, respectively. These C-BMP-melamine and MNP-BA assemblies were visualized by AFM (Fig. 4d,e and Supplementary Fig. 11c,d), revealing 2 nm fibres that are again indicative of a single assembly. When solutions containing purified MNP and C-BMP were combined and analysed by AFM, supramolecular polymers with a diameter of 2 nm were also observed (Fig. 4c and Supplementary Fig. 11b). As can be seen in Fig. 4b,c, the assemblies formed from purified MNP and BNP are noticeably shorter than those formed upon combining the crude reaction mixtures. This observation is consistent with length being limited by greater peripheral charge, an effect that is expected to be greater for supramolecular polymers formed from the purified nucleotides than it is for assemblies containing both nucleotides and the parent heterocycles (that is, those present in the crude reaction mixtures).
stacked hexad assemblies. Although free BA and free melamine form insoluble precipitates when mixed in aqueous solution (Supplementary Fig. 12), the steric bulk and charge provided by conjugation with R5P on one nucleobase favours the formation of the water soluble, linear assemblies of stacked hexads.

Supramolecular assemblies preferentially incorporate β-MMP. 1H-NMR spectroscopy was used to further characterize C-BMP-melamine and MMP-BA assemblies (Supplementary Figs 13 and 14). Nucleotides incorporated into these supramolecular assemblies exhibit extreme 1H NMR line broadening (to baseline), which render them invisible to solution state NMR spectroscopy. In contrast, free nucleotides that exist in equilibrium with the assemblies exhibit virtually no change in 1H resonance chemical shift or line width. Quantitative analysis of 1H resonance intensity can be used to determine the fraction of free nucleotides, and subtraction of these values from the known total concentration of nucleotides in a sample reveals the...
concentration of assembled nucleotides. Analysis of $^1$H spectra of solutions containing MMP + BA, 50 mM in nucleotide and heterocycle, showed temperature-dependent assembly of nucleotides from 5 to 40 °C (Fig. 5a,b and Supplementary Fig. 14). Unexpectedly, the preferential incorporation for $\beta$-MMP over $\alpha$-MMP was observed at all temperatures where assemblies are present. In particular, a twofold preference for $\beta$-MMP incorporation is observed at 5 °C, and only incorporation of $\beta$-MMP at 20 °C (Fig. 5b).

As the $\beta$-anomer is selectively incorporated into the assemblies over the $\alpha$-anomer, and anomerization of MMP occurs in the solution (see the Methods for details), we next explored if supramolecular assembly will affect the anomeric ratio of MMP. As noted above, at equilibrium, a solution of MMP contains a mixture of $\alpha$- and $\beta$-anomers (45% $\alpha$ and 55% $\beta$), however, when 50 mM MMP was incubated with 50 mM BA at 5 °C, we observe the conversion of $\alpha$-MMP to $\beta$-MMP, a conversion that reached a maximum of 63% $\beta$-MMP after 8 days (Fig. 5c). This conversion was not observed when BA was omitted or when MMP was below the minimal concentration required for MMP-BA assembly. Therefore, we have found that the assemblies formed with BA preferentially select/stabilize the $\beta$-anomer of MMP, and as a result, with anomerization, the assemblies enrich $\beta$-MMP (Fig. 5d). These observations indicate that solutions containing MMP and BA assemblies will change overtime in both molecular (that is, enrichment of $\beta$-MMP) and supramolecular composition (that is, more MMP and BA assembled).

Discussion

The data presented here demonstrate the efficient single-step syntheses of complementary nucleosides and nucleotides, starting with the plausible proto-nucleobases melamine and BA and ribose or R5P. Although R5P forms an exocyclic N-glycosidic bond with melamine (MMP) and R5P forms a C-nucleotide with BA (C-BMP), both of these nucleotides favour their $\beta$ anomers. Specifically, for the free nucleotides in solution, the $\beta$ and $\alpha$ anomers of C-BMP exist at equilibrium in a 67:33 ratio, and for MMP the $\beta$-anomer is favoured in a 55:44 ratio over the $\alpha$-anomer. Perhaps coincidentally, extant life uses the $\beta$-anomeric form of ribonucleotides, and our observations indicate that this form may have been enriched on the early Earth for $\beta$-ribonucleotides for which glycosidic bond anomerization is under equilibrium control. This nucleotide structural preference is apparently not limited to the nucleotides of melamine and BA, as previous studies have also revealed that nucleosides formed by drying and heating ribose with two other pyrimidines, 2-pyrimidinone$^{21,37}$ and 2,4,6-triaminopyrimidine$^{23}$, also favour the $\beta$-anomer.

Unlike the mononucleotides of extant RNA, we observe that MMP and C-BMP will pair as monomers in aqueous solution, with each other and with their unmodified pairing partners (that is, MMP with C-BMP, MMP with BA, and C-BMP with melamine), producing, in all cases, supramolecular structures that indicate the highly efficient stacking of H-bonded hexads that are themselves composed of Watson–Crick-like base pairs.
This pairing and stacking is sufficiently robust to drive assembly in the presence of the side products and unreacted starting materials of the crude nucleotide reactions. Based on these observations, and because C-BMP and MMP structurally resemble two nucleotides found in life today (UMP and AMP, respectively) and have been reported to pair with extent, complementary nucleobases, it is tempting to speculate that these heterocycles could represent ancestral nucleotides of the contemporary genetic polymers. In particular, the ability for C-BMP and MMP to form noncovalent supramolecular assemblies could have facilitated the prebiotic localization, organization and subsequent linking of these (or similar) nucleotides into covalent polymers that were then capable of storing and transferring information (for example, by templating the formation of sequence-specific assemblies for the polymerization of additional monomers). The ability of C-BMP and MMP to form supramolecular assemblies might have also facilitated the emergence of early RNA-like polymers by selecting nucleotides with sugars (or earlier trifunctional linkers) that were structurally compatible with the assemblies and their subsequences and coupling into covalent polymers. In the present study, we have, for practical reasons, used D-ribose and D-R5P for our nucleoside and nucleotide reactions with melanin and BA, but L-ribose and L-R5P would exhibit equivalent reactivity with these two heterocycles. Nevertheless, it has been often postulated that a racemic mixture of nucleotides would have inhibited the prebiotic synthesis of RNA polymers, and so the question of how the present system might address this challenge deserves some discussion. Although we have not shown chiral nucleotide selection, in the current study we have demonstrated that the β-anomer of MMP is enriched in supramolecular assemblies over the α-anomer of MMP, and this selection leads to a detectable increase in the ratio of the β-anomer over the α-anomer of MMP in the entire solution (presumably due to anomerization and selective stabilization by the assembly). As a recent example of the ability of supramolecular polymers to promote local chiral resolution, Aida and co-workers demonstrated that racemic solutions of chiral macrocycles self-sort into homochiral supramolecular polymers. It is therefore possible that supramolecular assemblies, formed by nucleotides with different sugars, including different anomers and enantiomers, could have been selectively enriched in individual supramolecular assemblies before polymerization. Current investigations of this possibility are actively being pursued in our laboratory.

Methods

Materials. Melamine and BA were purchased from Acros Organic, β-R5P disodium salt and α-ribose were purchased from Sigma-Aldrich. All chemicals were used as received.

Synthesis of C-BMP. BA (2.5 mmol) and R5P (2.5 mmol) were dissolved in 5 ml of H2O and the pH was adjusted to 9 with NaOH (unless otherwise noted). The solution was stirred for 24 h at 20 °C, at which time a clear, pale yellow solution was present. This solution is referred to as the crude BA-R5P reaction mixture. BA (2.5 mmol) and R5P (2.5 mmol) were dissolved in 5 ml of H2O and the pH was adjusted to 9 with NaOH (unless otherwise noted). The solution was stirred for 24 h at 20 °C, at which time a clear, pale yellow solution was present. This solution is referred to as the crude BA-R5P reaction mixture. BA (2.5 mmol) and R5P (2.5 mmol) were dissolved in 5 ml of H2O and the pH was adjusted to 9 with NaOH (unless otherwise noted). The solution was stirred for 24 h at 20 °C, at which time a clear, pale yellow solution was present. This solution is referred to as the crude BA-R5P reaction mixture. BA (2.5 mmol) and R5P (2.5 mmol) were dissolved in 5 ml of H2O and the pH was adjusted to 9 with NaOH (unless otherwise noted). The solution was stirred for 24 h at 20 °C, at which time a clear, pale yellow solution was present. This solution is referred to as the crude BA-R5P reaction mixture. BA (2.5 mmol) and R5P (2.5 mmol) were dissolved in 5 ml of H2O and the pH was adjusted to 9 with NaOH (unless otherwise noted). The solution was stirred for 24 h at 20 °C, at which time a clear, pale yellow solution was present. This solution is referred to as the crude BA-R5P reaction mixture.
