Prebiotic synthesis of 3′-amino-TNA nucleotide triphosphates

Authors Daniel Whitaker and Matthew W. Powner*

Affiliations Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK. *e-mail: matthew.powner@ucl.ac.uk

Abstract: Nucleic acid replication is essential to the emergence of life. Unlike canonical ribonucleotides, aminonucleotides have shown great promise in non-enzymatic replication, but are assumed to be prebiotically irrelevant due to perceived difficulties with their selective formation on the early earth. Here we demonstrate that, contrary to expectation, 3′-amino-TNA is formed diastereoselectively and regiospecifically from prebiotic feedstocks in four high-yielding steps. Our results suggest that 3′-amino-TNA may have been present on the early earth. Formation of the amino-sugar moiety by 3-component coupling of glycolaldehyde **1**, 2-aminooxazole **3** and an aminonitrile **6** regiospecifically positions both the 3′-amine and glycosidic bond. Phosphate provides an unexpected resolution of the two diastereomers formed, leading to purification of the genetically relevant *threo*-isomer. Under phosphate catalysis, the inhibitory *erythro*-isomer rearranges to an unreactive guanidinium salt, whilst the Watson-Crick base paring *threo-is*omer co-crystalises with phosphate, leading to its spontaneous purification and accumulation. Nucleobase construction on the amino-sugar scaffold is observed upon reaction with cyanoacetylene **8**, and subsequent thiolysis and photochemical anomerisation sets up the ideal stereochemistry for Watson-Crick base pairing. The resulting 3′-amino-TNA can be phosphorylated directly in water, under mild conditions with cyclic trimetaphosphate $(PO_3Na)_3$, forming a nucleotide triphosphate (NTP) in a manner not feasible for canonical nucleosides. The ease with which these activated NTPs form, and the inherent selectivity for the Watson-Crick base pairing structure, warrants further study of 3′-amino-TNA as a precursor to the genetic material of life.

A significant obstacle to our understanding of life's emergence on earth is how replication, an essential step for Darwinian evolution,[1] could have begun. The cellular machinery currently used to enable replication of DNA is too complex to have emerged without invoking evolution, so must be the product of a simpler ancestral system.[2] The "RNA world" hypothesis posits that this ancestral system was composed of self-replicating RNA which both encodes genetic information and can exhibit catalytic function. While its capacity to act as both genotype and phenotype in extant biology makes RNA an attractive candidate for the first genetic material,

it is plausible that a different – perhaps constitutionally simpler – nucleotide may have preceded it,[3] or that the first genetic material was chimeric in nature, composed of several different classes of nucleotide.[4]

One particular stumbling block for an exclusively RNA-first scenario has been its sluggish non-enzymatic replication. Despite decades of excellent work,[5,1] RNA-templated primer extensions using all four nucleobases and chemically activated ribonucleotides has reached a copying limit of about 7 nucleotides.[6] Moreover, the conditions necessary to promote this type of oligomerisation (including relatively high concentrations of divalent metals) also promote strand hydrolysis. Experimentally, aminonucleotides have been employed to overcome these issues.[7,8,9,10,11,12] These substrates benefit from a nucleophilic amine in place of one of RNAs hydroxyls; the enhanced amine nucleophilicity enables far superior templated ligations. Indeed, recent work has shown that polymers of up to 25 aminonucleotides can be copied with reasonable efficiency even inside fatty acid vesicles.[13] As non-biological molecules aminonucleotides have so far been discussed explicitly as model substrates with no prebiotic relevance. However, the superior reactivity encouraged us to experimentally probe this untested view.

Almost 20 years ago, as part of a seminal study evaluating the potential of non-canonical nucleotides to have been the first genetic polymers at the origin of life, Eschenmoser evaluated RNA, threose nucleic acid (TNA)[14], 2′-amino-TNA and 3′-amino-TNA.[3] He considered the difference in "'constitutional' vs 'generational' complexity" of these nucleotides, comparing the structural complexity of the nucleotides with how readily they could have formed. Unusually for non-canonical nucleotides TNA and amino-TNA polymers are able to form very stable duplexes with both RNA and DNA as well as with themselves, setting the stage for a potential genetic takeover and for them to have played a direct role in the origins of extant life on earth.[3,15] TNA's four carbon sugar means it is "constitutionally" simpler than RNA, but Eschenmoser concluded that it was actually not "generationally" simpler since the most difficult steps for the emergence of polynucleotides seemed to be formation of the glycosidic bond, phosphorylation of the nucleosides and polymerisation of the resultant nucleotides. On the other hand, he considered polymers of amino-TNA to be "generationally" simpler than those of RNA due to the relative ease of polymerisation from their respective mononucleotides. However, the monomers of amino TNA appeared to be both generationally and constitutionally more complex than those of RNA due to the additional difficulty in regioselective incorporation

of an amine into a sugar. Indeed, when Eschenmoser considered the effect that nitrogenous compounds would have on the aldol chemistry that, at the time, seemed the most likely source of sugars on the early earth,[16] he concluded that "huge chemical complications" would arise due to Amadori and Mannich chemistry occurring alongside the already unselective aldol reactions. He nevertheless recognised that these complications may "harbour structural and transformational opportunities that may be relevant" to the origins of life. Recognising that if a facile synthesis of amino-TNA could be realised, then amino-TNA may be considered both 'constitutionally' and 'generationally' simpler than RNA, we therefore decided to re-examine this suggestion, in light of chemistry discovered since. Our re-evaluation indicates that one such opportunity is the regioselective incorporation of a free amine moiety at the 3' position of a sugar during aminooxazoline synthesis (Fig 1).

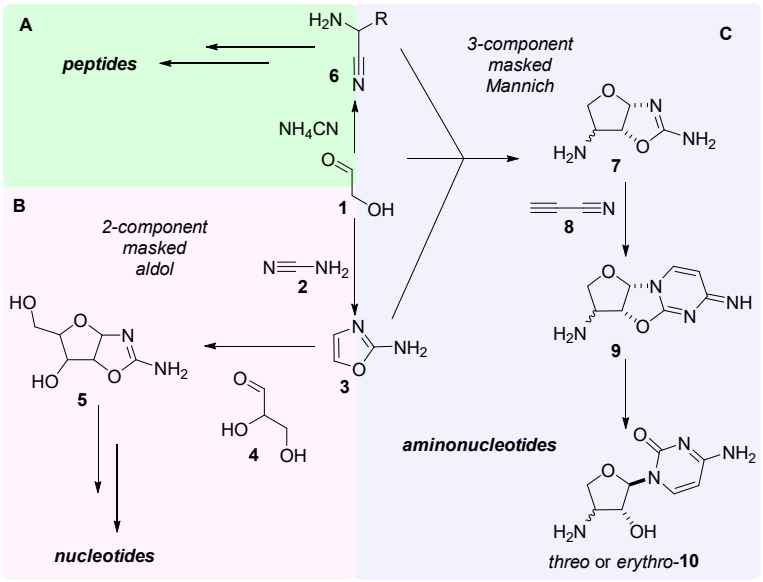


Figure 1 | Intersections of prebiotic peptide, sugar and nucleobase syntheses. Glycolaldehyde **1** is a prebiotic precursor to **A**) peptides via serine nitrile **6** (R=CH_2OH), **B**) nucleotides via 2-aminooxazole **3** and its masked-aldol reaction with glyceraldehyde **4**, and **C**) aminonucleotides **10** via 3-component Mannich reactivity.

How the first nucleotides formed on earth is still an open question,[17] but there is a strong growing consensus that the earliest life would have emerged from a system containing peptides and lipids as well as nucleotides.[18] Despite this, most proposed mechanisms by which nucleotides could have formed on the early earth[19] are still based on two assumptions – first, that the sugar and base moieties formed separately and then combined at a late stage,[20] and second, that nucleotide synthesis and amino acid/peptide synthesis occurred separately. We

have previously challenged the first assumption, discovering highly diastereoselective routes to ribonucleotides[21,22] in which sugar and base are constructed in the same environment on the same scaffold. This sequence represents the only diastereoselective prebiotically plausible formation of ribonucleotides known so far, and proceeds via the coupling of cyanamide **1** with glycolaldehyde **2** to form 2-aminooxazole **3**, which then reacts with glyceraldehyde **4** forming the key intermediate aminooxazoline **5** (Fig. 1B).

The advantages delivered through unification of nitrile and sugar chemistry in our synthesis have led us to consider the relationship between peptides and nucleotides, and how their syntheses would intersect.[23] Aminonitriles **6** have long been considered to be precursors of amino acids[24,25] since their formation via the Strecker reaction is both facile and predisposed to form biogenic α-aminonitriles. However, the intrinsic reactivity of these molecules has been overlooked, with a major focus on how to improve their relatively sluggish hydrolysis to amino amides and then acids. Unlike amino acids, aminonitriles are predisposed towards reactivity in water due to the kinetically inert but high energy nitrile moiety and the suppressed pK_a of the amine rendering them nucleophilic at neutral pH. We recently demonstrated that this dual reactivity could be exploited to form peptide bonds in water, with high selectivity for biologically relevant α-peptides over unnatural isomers.[26,27] But how then does this chemistry interact with nucleotide synthesis? And can it be used to overcome the perceived complexity of aldol, Mannich and Amadori reactions (Fig. 2A) expected in nitrogenous aldol chemistry?

Results and Discussions

Amino-sugar synthesis

We have previously reported that at pH 5–6 in the presence of **AICA** the reaction between sugars and 2-aminooxazole **3** is out-competed by a three component reaction. Imine formation between **AICA** and the sugar (e.g. glycolaldehyde **1**) preceded reaction with **3**, forming an aminooxazoline product (**11**) regiospecifically functionalised by an amine at C3' (Fig. 2B).[28] In a related reaction between amino acids, glyceraldehyde **4** and 2-aminooxazole **3**, Blackmond observed a mixture of both 2- and 3-component coupling products (Fig. 2D).[29] The 3-component reaction was diastereoselective with enantiopure proline which, due to the low chemoselectivity, led to enantioenrichment of glyceraldehyde **4** and therefore consequently

enantioenrichment of the 2-component aminooxazoline products **5**. We recognised that if a variant of this reaction could be rendered selective for 3-component coupling with an -NH_2 source, a simple route to 2,3'-diaminooxazolines **7** would be uncovered. We envisaged that aminonitriles **6**, due to their low amine pK_a, would be efficiently incorporated and that, due to the reversibility of the Strecker reaction, the free amine could then be liberated after coupling, and therefore deliver a selective route to 3'-amino-nucleosides.

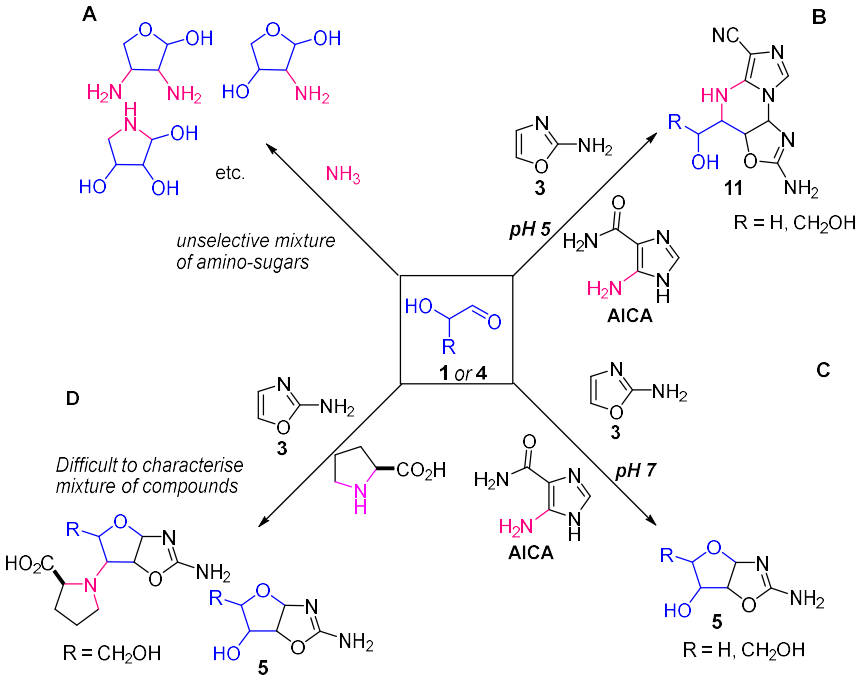
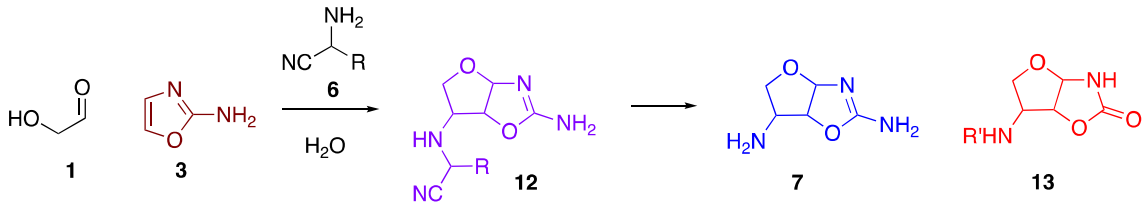


Figure 2 | Formation of amino-sugar derivatives from C2 and C3 sugars. **A**) Uncontrolled aldol and Amadori reactions between ammonia and sugars;[16] **B**) Selective 3-component coupling between 2-aminooxazole **3**, glycolaldehyde **1** and AICA at pH 5 to form tricyclic product **11**;[28] **C**) Selective 2-component coupling between 2-aminooxazole **3** and glycolaldehyde **1** in the presence of AICA to form aminooxazolines **5** at pH 7;[29] **D**) Unselective incorporation of amino acids into aminooxazolines via coupling of glyceraldehyde **4** with 2-aminooxazole **3** and proline.[40]

We envisaged the reaction of glycolaldehyde **1**, 2-aminooxazole **3** and an aminonitrile **6** would form tetrose 3'-amino-aminooxazolines **7**. The nucleobase could then be constructed on this compound by reaction with cyanoacetylene **8**, and anomerisation of the base would furnish the base-pairing α-*threo*-nucleoside *threo*-**10** (Fig. 1C). We therefore began our investigation by looking at the three-component reaction between glycolaldehyde **1**, 2-aminooxazole **3** and glycine nitrile **6a** (R=H) in water at neutral pH. Although the reaction was not selective, **12a** was clearly visible in the ^{1}H NMR spectrum. Employing a slight excess of **6a** improved the

chemoselectivity, with the yield increasing to an appreciable 40% (Fig. 3, Table entry 1). Based on our previous work with AICA we expected that three-component coupling would be favoured at low pH, and indeed at pH 5 we observed very efficient three-component coupling, forming **12a** in 94% yield after 16 h (Fig. 3, Table entry 2). The yield reached a maximum at pH 5. At lower pH lower yields were observed – at pH 4 the reaction stalled after four hours and 60% conversion, but all three starting materials were accounted for (Table 1, entry 3). When initiated at pH 4, the solution was observed to drop to pH 2 as the reaction proceeded, and at pH 2 coupling is observed to proceed only very slowly (Fig. S1). Interestingly, **6a** (pK_{aH} = 5.2) can act as the ideal buffer, as well as a coupling partner, keeping the pH close to the optimum value of pH 5. This buffering effect is highly apparent in reactions initiated at pD 4.5, where with 1 equiv. of **6a** the reaction stops at 55% conversion, but is then restarted if the solution is adjusted back from pD 2.5 to pD 4.5 reaching 75% conversion. However, when the reaction was initiated at pD 4.5 with 3 equiv. of **6a** it did not stall, the pH was maintained and **12a** was furnished in 90% yield. At pH 5 3-component coupling was seen to dominate even at concentrations as low as 10 mM, albeit more slowly, with **12a** formed in 70% yield after 10 days (Fig. 3, Table entry 4).

If glycyl aminooxazoline **12a** is heated at 60 °C a slow retro-Strecker reaction occurs, liberating a free amino group to form **7** (Fig. 3, Table entry 5). However, hydrolysis of the aminooxazoline moiety of **12a** to oxazolidinone **13** occurred concomitantly with glycyl-retro-Strecker upon heating. Similar hydrolysis of *pentose*-aminooxazolines (**5**) to their corresponding oxazolidinones has previously been reported to occur on heating in phosphate solution.[30] This C2-hydrolysis can be supressed with more sterically encumbered aminonitriles which undergo a more rapid retro-Strecker reaction. For example, α-alkylated aminonitriles **12b-e** (entries 5–9) all undergo retro-Strecker reaction at room temperature, proceeding efficiently over several days (Fig. 3). The cyanohydrin formed in this process is not consumed and would be recycled into aminonitriles in the presence of a suitable source of ammonia.[31] Indeed, in the case of the hydrophobic R groups (e.g. iPr and iBu) cyanohydrin crystallisation was seen to occur spontaneously from the reaction mixture.



Entry	R	pH	t / days	12 / %	7 / %	13 / %	*threo:erythro*
1	H	7	1	40	0	0	1:3
2	H	5	1	94	0	0	1:3
3	H	4	1	60	0	0	1:3
4[a]	H	5	10	70	0	0	1:3
5	H	5	15[b]	19	37	38	1:3
6	Me	5	5	43	37	12	1.2:1
7	Me	4.5	5	22	52	18	1:1
8	iPr	4.5	5	4	73	16	1.3:1
9	iBu	4.5	5	8	57	21	1:1
10	sec-Bu	4.5	5	4	79	17	1.3:1

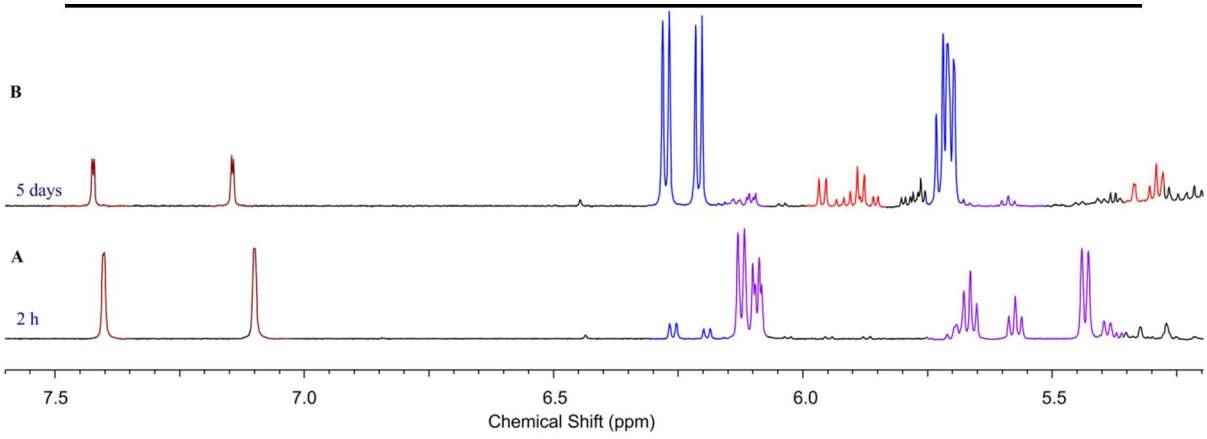


Figure 3 | Three-component coupling of aldehyde 1, oxazole 3 and aminonitrile 6 to yield amino-nucleotide precursor 7 in water. Table. Optimisation of 3-component coupling of glycolaldehyde (**1**; 0.2 M) and 2-aminooxazole (**3**; 0.2 M) with aminonitrile (**6**; 0.6 M) at room temperature. [b] 10 mM **1**, 10 mM **3** and 30 mM **6**. [c] room temperature for 1 day, then 60 °C for 14 days. R' = H or CHRCN **^{1}H NMR spectra** [400 MHz, H_2O/D_2O (9:1), 7.6 – 5.2 ppm] **A)** Formation of oxazoline **12e** (R = sec-Bu) after 2 h at room temperature and pH 4.5 **B)** Retro-Strecker of **12e** at room temperature and pH 4.5 to form **7** after 5 days.

Although aminonitriles are known to be unstable compared to cyanohydrins and ammonium at low concentrations of free ammonia[32] they are reasonably kinetically stable at neutral or acidic pH. Commeyras, however, has shown that α,α-disubstituted aminonitriles undergo retro-Strecker reactions easily even at acidic pH at room temperature, and that N-alkylation further increases the rate of retro-Strecker.[33] On the other hand, Commeyras reported that Ala-CN (**6b**) was roughly 20 times more stable than Gly-CN (**6a**),[34] but our data suggests that α-alkylation

promotes more rapid retro-Strecker (**12b** > **12a**) and that Gly-CN (**6a**) and its derivatives are kinetically more stable to retro-Strecker reactions than the more substituted aminonitriles.

Aminonitriles **6** are more reactive coupling partners than either amino acids or ammonia under the optimal conditions for 3-component coupling. For example, when 200 mM **1**, **3**, **6a** and glycine were mixed at pH 5, **12a** (72%) was seen to form as the predominant product after 24 h. Furthermore, if **1** and **3** were mixed in the presence of 600 mM **6a** in 1 M NH_4Cl at pH 5, **12a** was seen to form in 95% yield after 24 h.

When the reactions of α-substituted aminonitriles initiated at pH 5 were monitored for several weeks the concentration of **7** was seen to fall slightly over time, without concomitant increases in either **12** or oxazolidinone **13**. The concentration of *threo*-**7** seemed to fall less than *erythro*-**7**. To explore this intriguing behaviour further we isolated **7** as a mixture of diastereomers.

threo-Selective crystallisation

When **7** (3:2 *threo*:*erythro;* Fig 4A) was mixed with 200 mM phosphate at near-neutral pH a colourless precipitate was seen to form rapidly. Proton NMR spectra indicated that the supernatant was significantly enriched in *erythro*-**7**, and *threo*-**7** had precipitated (Fig. 4B). Over time *erythro*-**7** was then seen to transform into a new compound, assigned as the bridged bicyclic compound **14** on the basis of 2D NMR and mass spectral data (Fig. 4C; see Fig. S42-S44).[3] **14** was found to be unreactive in the subsequent synthetic steps. The high pK_{aH} of the guanidine moiety (pK_a of guanidinium – 13.7) means that **14** is fully protonated, and so its nucleophilicity is quenched at neutral pH. The same rearrangement would also be expected to occur in 3′-amino-*ribo*-aminooxazolines, since *ribo* and *erythro* aminooxazolines have the same stereochemical orientation at their 2′ and 3′ carbons, which would preclude the formation of 3′-amino-RNA.

Analysis of the precipitate demonstrated that the phosphate salt of *threo*-**7** had indeed undergone a highly diastereoselective crystallisation (Fig. 4, inset). This provides a natural route for separation and concentration of the genetically useful diastereomer, *threo*-**7**, while removing the potentially inhibitory *erythro*-isomer. This *threo*-selective precipitation is observed at concentrations of *threo*-**7** as low as 10 mM in 100 mM phosphate. In contrast, precipitation of *ribo*-aminooxazoline **5**, thought to provide a selection mechanism for RNA,[23,29]

occurs efficiently only above 100 mM and does not co-accumulate **5** with phosphate (see Fig. S37-S40).

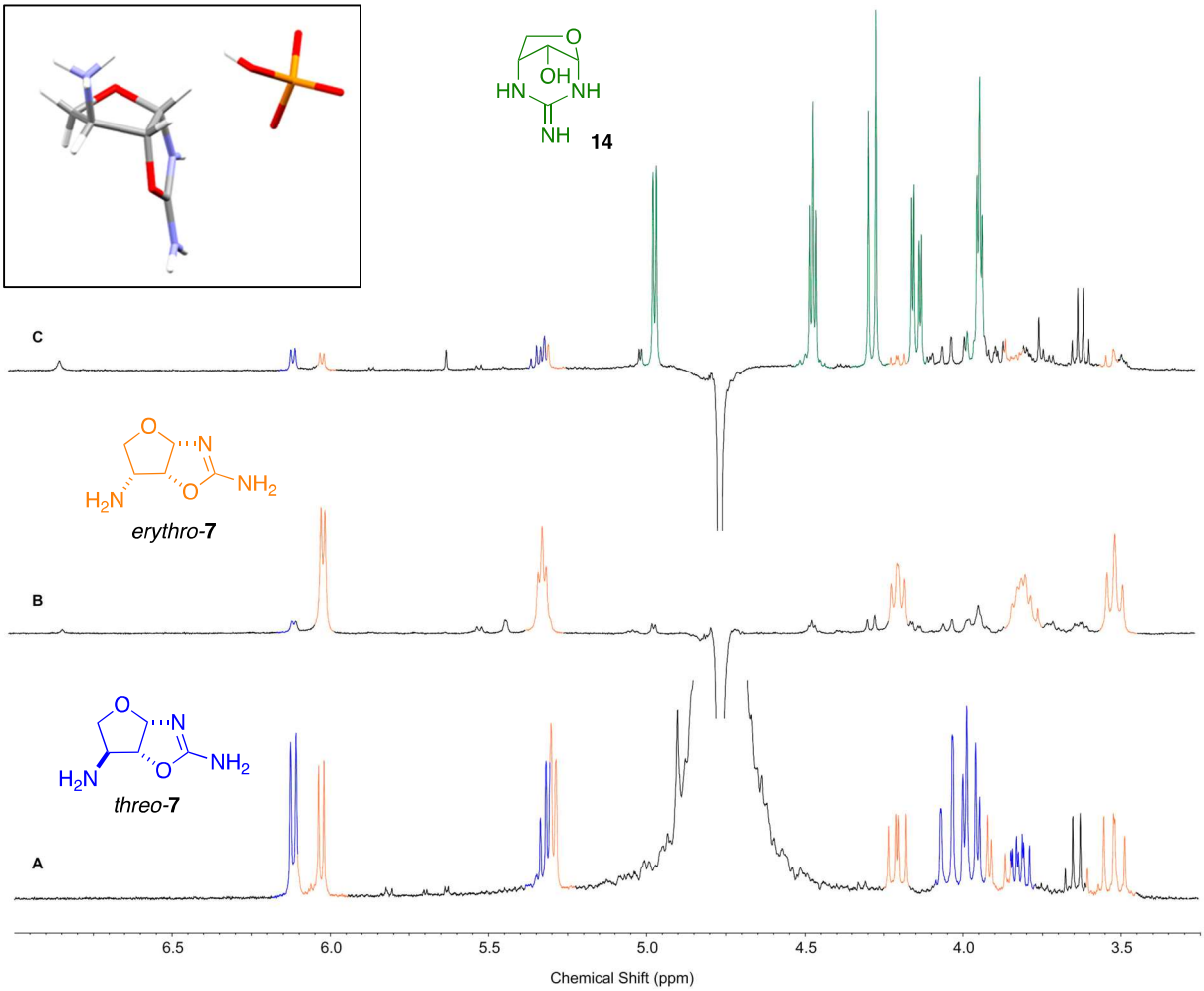


Figure 4 | ^{1}H NMR spectra to show the precipitation of *threo*-7 and rearrangement of *erythro*-7 to 14 in phosphate buffer. **A)** [400 MHz, D_2O, zg30, 7.0 – 3.3 ppm] mixture of *threo*-**7** (blue) and *erythro*-**7** (orange) (200 mM, 3:2) in D_2O at pD 6.5 without addition of phosphate; **B)** [400 MHz, H_2O/D_2O 9:1, noesygppr1d, 7.0 – 3.3 ppm] mixture of *threo*-**7** (blue) and *erythro*-**7** (orange) (200 mM, 3:2) in 200 mM phosphate at pH 6.5, soluble fraction after 1 h; **C)** [400 MHz, H_2O/D_2O 9:1, noesygppr1d, 7.0 – 3.3 ppm] soluble fraction after 20 h showing formation of **14** (green). **Inset** Crystal structure of *threo*-**7**·H_3PO_4 ; grey = carbon, white = hydrogen, red = oxygen, blue = nitrogen and orange = phosphorus. CCDC 2087673.

The diastereoselectivity imparted in **7** by phosphate is therefore twofold: phosphate drives the co-precipitation, purification and concentration of the *threo*-isomer, whilst catalysing the conversion of the *erythro*-isomer into an unreactive by-product. That the chemistry is wholly predisposed to favour the genetically useful isomer is an interesting and unprecedented result.

Nucleobase elaboration

Phosphate-buffered addition of cyanoacetylene **8** to pentose aminooxazolines **5** in near-neutral water quantitatively converts them to 2,2′-anhydrocytidines,[14,21] building the nucleobase on the sugar scaffold. When *threo*-**7** was mixed with **8** at pH 7 competitive addition to both nucleophilic sites on the molecule was observed, forming *threo*-**9** and *threo*-**15** as the predominant products in 74% combined yield after 24 h (Fig. 5A). The NMR spectrum in D_2O very clearly shows that only two species derived from *threo*-**7** are present after 24 h, *threo*-$\mathbf{9_d}$ and *threo*-$\mathbf{15_d}$ (Fig. S54). However, addition of **8** to the 3′-amine does not significantly inhibit addition to the aminooxazoline,[35] and once excess cyanoacetylene **8** has been consumed *threo*-**15** slowly converts to *threo*-**9**. If no adjustment to the reaction conditions are made, hydrolysis of *threo*-**15** to *threo*-**9** (33%) is complete after 28 days, but this sluggish hydrolysis is then accompanied by hydrolysis to *threo*-**17**, which forms in 22% yield after 28 days. Some precipitation of nucleoside-derived products is seen to occur over this time alongside significant precipitation of cyanoacetaldehyde oligomer products which leads to partial enrichment of nucleosides in the supernatant over time. However, the sluggish hydrolysis of *threo*-**15** to *threo*-**9** can be greatly accelerated, for example by acid catalysis, and at pH 2 the conversion is rapid (< 10 min) and quantitative.

These results demonstrate that the skeletal structure of 3'-amino-*threo*-nucleosides can be regioselectively established in only two steps from prebiotically plausible two- and three-carbon units. Formation of aminonucleotides would then require anomerisation of the base and phosphorylation. We set out to consider how these could have occurred on the early earth. Anhydronucleoside hydrolysis seems to be the disposed pathway for anhydronucleosides in water (Fig 5B). It is very sluggish, taking weeks to complete at neutral pH, but is accelerated at higher pH occurring within 6 hours at pH 8 or within 10 mins at pH 11 to furnish *threo*-**17** (quant.). However, *threo*-**17** has not been shown to be an effective component of a proto-genetic material, and by comparison with the weaker duplex formation in α-DNA,[36] would be expected to form less stable duplexes than polymers of *threo*-**10**. Therefore, if *threo*-**17** were to be a precursor to a genetic polymer on the early earth anomerisation of the 1′ stereocenter would likely have been highly beneficial.

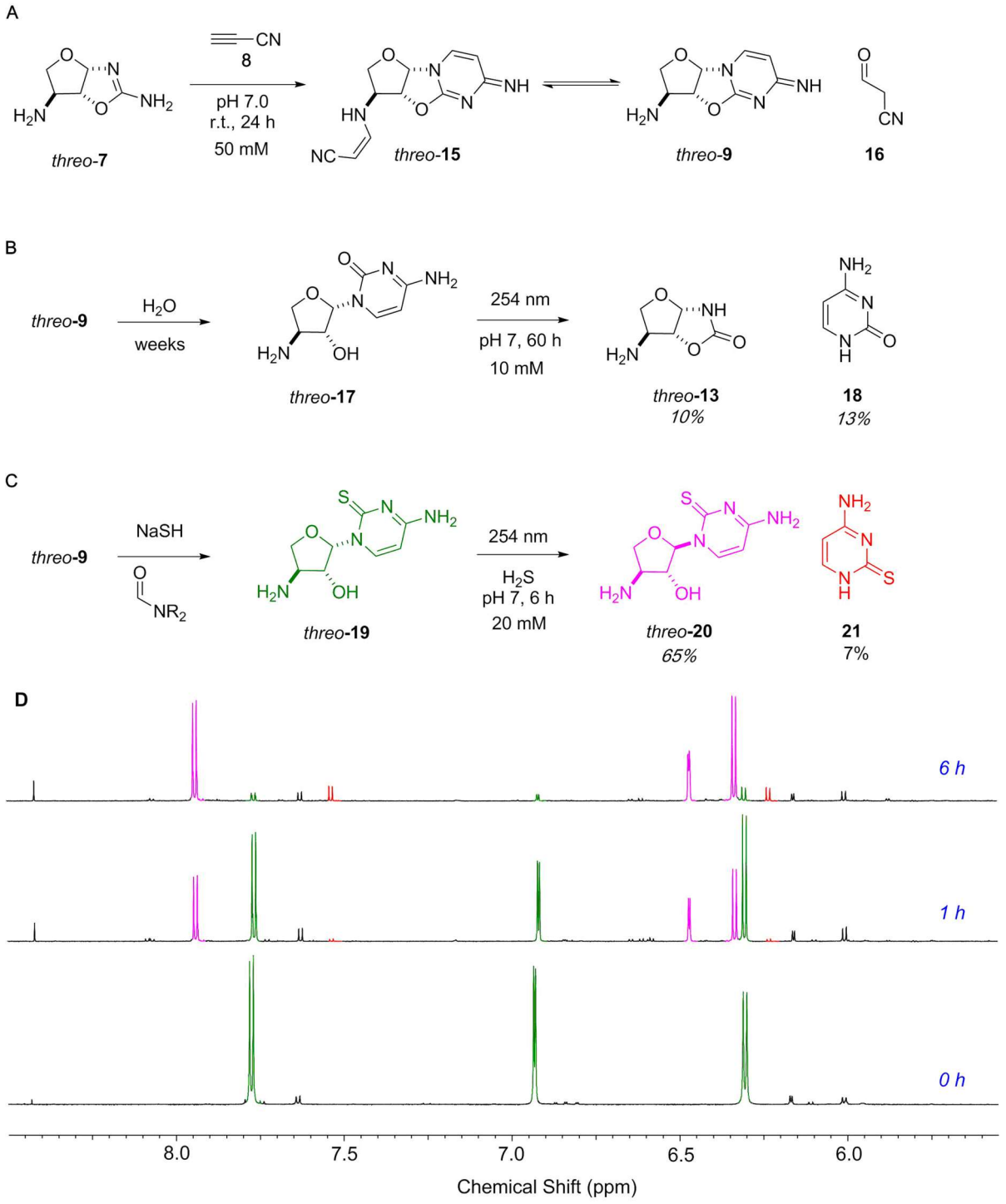


Figure 5 | Formation of anhydrocytidine *threo*-9 and its photochemical products. A) Reaction of *threo*-**7** with cyanoacetylene **8** to yield anhydronucleotide *threo*-**9**. **B**) Hydrolysis of *threo*-**9** to *threo*-**17** and subsequent photochemical formation of *threo*-**13**; **C**) Thiolysis of *threo*-**9** to *threo*-**19** and subsequent photochemical anomerisation of *threo*-**19** to *threo*-**20.** R=Me or H; **D**) ^{1}H NMR [700 MHz, H_2O/D_2O 9:1, noesygppr1d, 5.5-8.5 ppm] spectra showing the photoanomerisation of *threo*-**19** to yield *threo*-**19** and **21** on irradiation at 245 nm in water.

Irradiation at 254 nm is the most promising mechanism uncovered so far for anomerisation of pyrimidine nucleosides. However, for canonical nucleobases this process is still inefficient, with their nucleotides undergoing condensation to form oxazolidinones as the major product upon irradiation, with anomerisation only occurring in 4-6% yield.[37] In contrast, 2-thiocytosine ribonucleosides undergo extremely clean anomerisation to preferentially form a *trans*-disposition between the 1' and 2' substituents.[38] We recently demonstrated that the same anomerisation can occur on a *threo* skeleton,[14] and therefore decided to investigate aminonucleoside photoanomerisation. As expected, when *threo*-**17** was irradiated at 254 nm in neutral water it slowly formed oxazolidinone **13** along with base release to give cytosine **18** (Fig. 5B). We therefore considered how sulfur would have been incorporated into *threo*-**9** to give *threo*-**19.**

Thiolysis of α-anhydrocytidines occur in the presence of sodium hydrosulfide in formamide.[38,14] Although these conditions are not trivial to rationalise as prebiotically plausible, all the chemical inputs are individually plausible, and the demonstrated efficiency and value of this chemical transformation outweighs speculative plausibility at this stage – it seems reasonable that there may be a different, and more plausible route to *threo*-**19** and it would be counterproductive to discount uniquely effective chemistry on unverifiable grounds of plausibility. We therefore tested the reaction of *threo*-**14** and hydrosulfide in DMF and observed quantitative conversion to *threo*-**19** in the crude NMR. Similar, although less clean, thiolysis was observed in formamide.

When **19** was irradiated at 254 nm in the presence of H_2S in neutral water clean anomerisation was found to occur, providing *threo*-**20** (65%) along with 7% 2-thiocytosine (Fig. 5C + 5D). 2-Thiopyrimidines are intriguing molecules from the point of view of Watson-crick base pairing – 2-thiouridine and 2-thiocytidine are found in tRNA[39,40] and the enhanced selectivity for U:A base pairing over the U:G wobble pair when uridine is replaced by 2-thiouridine[41] has been exploited to improve the fidelity of non-enzymatic primer extension[42] and ribozyme-catalysed copying of RNA.[43] Further to this, 5-methyl-2-thiocytidine has been shown to selectively bind with inosine,[44] which has been considered as a likely precursor to guanosine in early genetic materials due to its easier synthesis and greatly increased solubility.[45] 2-thioU:A and 2-thioC:I Watson-Crick pairing deserves further investigation from the point of

view of duplex stability and copying efficiency, particularly in light of these compounds potentially being synthetic precursors to the oxygenic variants.

How the cytosine-nucleoside *threo*-**10** could have formed from its 2-thiocytosine presents an interesting problem here due to the nucleophilicity of the 3'-amine. Activation of the sulfur, either by alkylation or oxidation, would render it a better leaving group and so greatly accelerate substitution at this position,[14] but in the case of *threo*-**20** such activation leads to rapid cyclisation to **24** (Fig. 6), even at pH 5 where the amine would be highly protonated (we have not observed either cyclisation or hydrolysis of *threo*-**20** in the absence of an activating agent in rigorously deoxygenated water, even over several weeks, although we cannot rule out that substitution occurs extremely slowly).

Phosphorylation

Considering the issue of cyclisation, we next reflected on our target structure and realised it was the phosphorylated amino-TNA nucleo*tide*, not the unphosphorylated nucleo*side*. Accordingly, we recognised that this cyclisation problem may have arisen by assuming that the nucleoside should be an intermediate in the synthesis. The nucleophilicity of the amine would be suppressed by phosphorylation, and therefore the solution to monomer cyclisation may lie in incorporation of the intermediate into a nucleotide – where we expected cyclisation to be suppressed in favour of hydrolysis – before converting the pyrimidine to the canonical nucleobase.

Nucleoside phosphorylation is extremely challenging in water and has remained a long standing problem, with reasonably efficient phosphorylation only being seen for vicinal diols in water.[46] Furthermore, mechanisms by which nucleotide triphosphates, nature's chosen monomers and universal energy currency, could have formed on the early earth are even more poorly understood than the formation of nucleotide monophosphates – the most promising methods all employ nucleotide monophosphates as a starting point,[46,47] but such processes compete with formation of inorganic polyphosphates,[48] and leave the question of how nucleotide monophosphates could form in the first place an open question. Aminonucleotides provide a potential way around this problem due to the increased nucleophilicity of the amine moiety relative to the solvent water.[49] We reasoned that if an aminonucleotide were exposed to a condensed phosphate species nucleotide polyphosphates could form directly and selectively

in water. Regiospecific addition would be expected to occur, which would help with directionality in subsequent oligonucleotide formation.

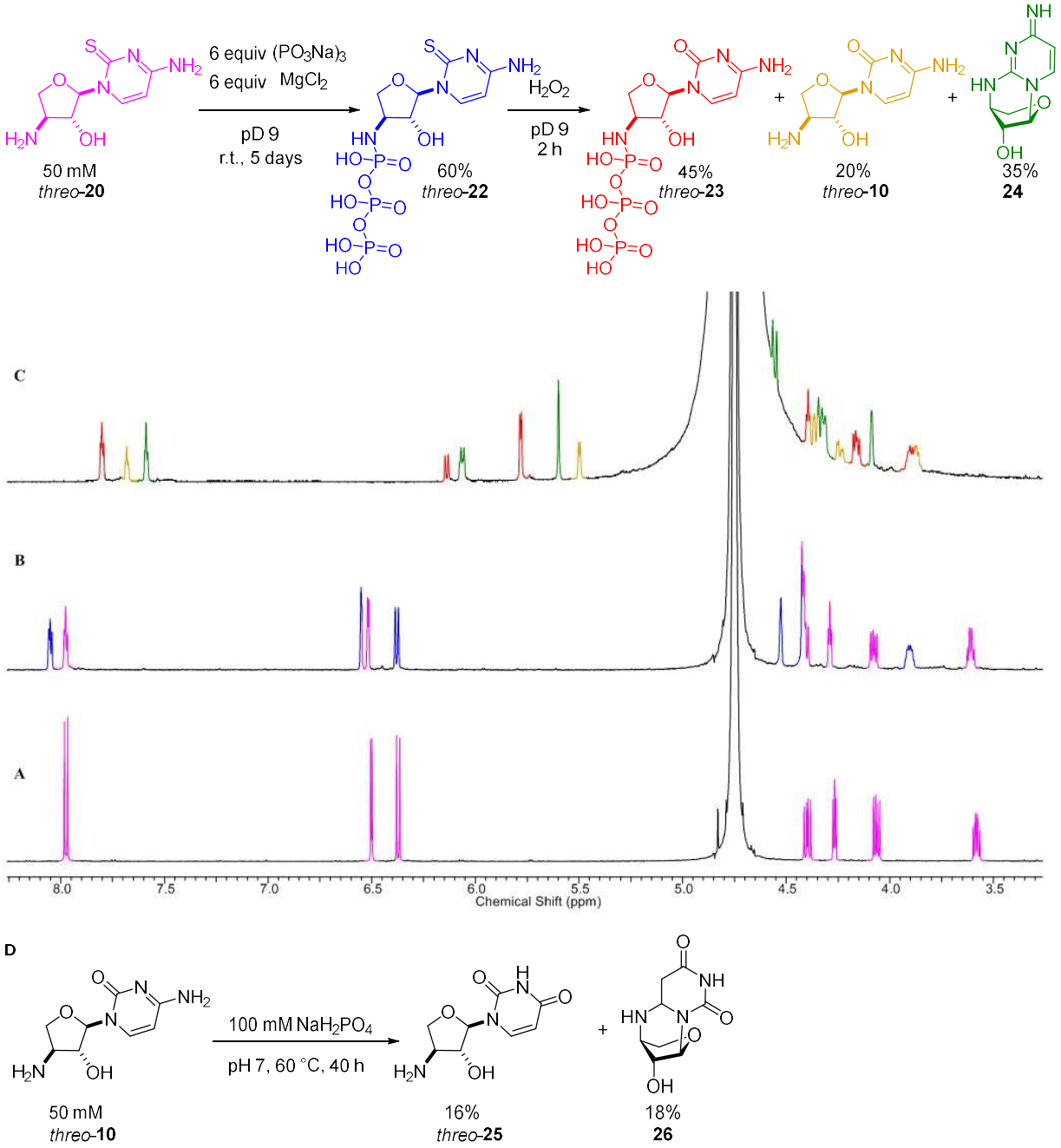


Figure 6 | Aqueous phosphorylation of amino-nucleoside 20 with cyclic trimetaphosphate $(PO_3Na)_3$ and hydrolysis to canonical nucleobases. ^{1}H NMR [700 MHz, D_2O, 3.25-8.25 ppm] spectra showing the reaction of 50 mM *threo*-**20**, 6 equiv. $(PO_3Na)_3$, 6 equiv. $MgCl_2$, pD 9 after **A)** 1 h; **B)** 5 days and **C)** subsequent reaction of solution shown in **B** with 3 equiv. H_2O_2 resulting in the formation of *threo*-**23**, *threo*-**10** and **24**. Note that the apparent change in multiplicity at C6-H is due to partial deuteration at C5-H. **D)** Phosphate catalysed hydrolysis of cyditine *threo*-**24** to uridine *threo*-**25**.

We elected to study the interaction between *threo*-**20** and trimetaphosphate, $(PO_3Na)_3$.[50] In neutral water at room temperature no reaction was seen to occur over 5 days, so we investigated the effect of pH. If the phosphorylation is initiated in alkaline solution (pD 12) with 6 equiv. $(PO_3Na)_3$ 50% nucleotide triphosphate *threo*-**22** is observed to form over 25 days. Over this period the solution also fell to pD 9, so we next initiated the phosphorylation at pD 9 in the presence of a divalent metal ion (Mg^{2+}) to accelerate the reaction and observed 60% *threo*-**22** form after only 5 days (Fig. 6B). Intriguingly, were *threo*-**22** to undergo templated polymerisation it would do so in the same direction as biological DNA and RNA polymerisation. The ease with which this triphosphate forms relative to canonical nucleotides suggests that the concomitant phosphorylation and polymerisation of aminonucleotides should be revisited investigating monomers in which the activated phosphate is positioned on the amine rather than the hydroxyl.

We next exposed *threo*-**22** to sulfur-activating agents (such as H_2O_2) and found that hydrolysis to *threo*-**23** and *threo*-**10** occurs, demonstrating that phosphorylation of the 3′-NH_2 enables selective hydrolysis to occur in place of the cyclisation that is observed without phosphorylation. Further hydrolysis of the cytosine base to uracil can also occur to give both canonical pyrimidine bases, a process which is accelerated by heating in phosphate buffer (Fig. 6D). *threo*-**10** undergoes hydrolysis (34% after 40 h) to furnish *threo*-**25** at 40 °C in 100 mM phosphate at neutral pH. Uridine *threo*-**25** is observed to also undergo reversible addition of its 3′-amine to the C6-carbon of the uracil base when heated to yield **26**.

We have found that regio- and stereo-selective incorporation of an amine into a nucleotide sugar moiety can be achieved with prebiotically plausible reactions. 3’-amino-TNA NTPs are formed readily from two and three carbon building blocks. The phosphate-mediated diastereoselectivity of these reactions, for the genetically useful *threo*-nucleotides, are unparalleled in prebiotically plausible nucleotide syntheses. Notably, the 3'-amino-isomer that is exclusively formed has been shown to form far more stable duplexes with both itself and with RNA than the 2'-amino-isomer[15] - the network appears to be predisposed to generate the superior genetic molecule. The ease and selectivity for the formation of 3′-amino-TNA suggests that it would have been present on the early earth. The ability of these nucleosides to react with activated phosphates in water has been demonstrated to be far superior to that of canonical nucleotides which, along with the known performance of amino-nucleotides in non-enzymatic replication studies and interspecies Watson-Crick base pairing with D/RNA,

suggests that these molecules may have formed replicating polymers more easily than RNA. The propensity to 3'-triphosphorylate might also have primed biochemistry to adopt this oligomerisation register, which was then imparted to D/RNA. The implications of this scenario for polymerisation and information transfer will be the object of further studies.

Acknowledgements: We thank K. Karu, M. Puchnarewicz (UCL Mass Spectrometry Facility), M. Corpinot and K. Bucar (Crystallography) and A. E. Aliev (NMR spectroscopy).

Funding: The Leverhulme Trust (RPG-2019-214) and the Simons Foundation (318881FY19), provided financial support.

Data statement: All data (experimental procedures and characterization data) supporting the findings of this study are available within the article and its Supplementary Information. X-ray crystallographic data has also been deposited at the Cambridge Crystallographic Data Centre (CCDC) under the following CCDC deposition number: 2087673. These can be obtained free of charge from CCDC via www.ccdc.cam.ac.uk/data_request/cif.

Author contributions: M.W.P. conceived the research. M.W.P. and D.W. designed and analysed the experiments and wrote the paper; D.W. conducted the experiments.

---

[1] Szostak, J. W. The Narrow Road to the Deep Past: In Search of the Chemistry of the Origin of Life. *Angew. Chem. Int. Ed.* **56**, 11037–11043 (2017).
[2] Robertson, M. P. & Joyce, G. F. The origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* **4**, 1 (2012).
[3] Schöning, K.-U. *et al.* The -L-Threofuranosyl-(3'→2')-oligonucleotide System ('TNA'): Synthesis and Pairing Properties. *Helv. Chim. Acta* **85**, 4111–4153 (2002).
[4] Bhowmik, S. & Krishnamurthy, R. The role of sugar-backbone heterogeneity and chimeras in the simultaneous emergence of RNA and DNA. *Nat. Chem.* **11**, 1009–1018 (2019).
[5] Weimann, B. J., Lohrmann, R., Orgel, L. E., Schneider-Bernloehr, H. & Sulston, J. E. Template-Directed Synthesis with Adenosine-5'-phosphorimidazolide. *Science* **161**, 387–387 (1968).
[6] Li, L. *et al.* Enhanced nonenzymatic RNA copying with 2-aminoimidazole activated nucleotides. *J. Am. Chem. Soc.* **139**, 1810–1813 (2017).
[7] Zielinski, W. S. & Orgel, L. E. Oligomerization of activated derivatives of 3'-amino-3'-deoxyguanosine on poly(C) and poly(dC) templates. *Nucleic Acids Res.* **13**, 2469–2484 (1985).
[8] Röthlingshöfer, M. *et al.* Chemical primer extension in seconds. *Angew. Chem. Int. Ed.* **47**, 6065–6068 (2008).
[9] Chen, J. J., Cai, X. & Szostak, J. W. N2'→P3' Phosphoramidate Glycerol Nucleic Acid as a Potential Alternative Genetic System. *J. Am. Chem. Soc.* **131**, 2119–2121 (2009).
[10] Blain, J. C., Ricardo, A. & Szostak, J. W. Synthesis and Nonenzymatic Template-Directed Polymerization of 2'-Amino-2'-deoxythreose Nucleotides. *J. Am. Chem. Soc.* **136**, 2033–2039 (2014).

[11] Zhou, L., O'Flaherty, D. K. & Szostak, J. W. Template-Directed Copying of RNA by Non-enzymatic Ligation. *Angew. Chem. Int. Ed.* **59**, 15682–15687 (2020).
[12] Hänle, E. & Richert, C. Enzyme-Free Replication with Two or Four Bases. *Angew. Chem. Int. Ed.* **57**, 8911–8915 (2018).
[13] O'Flaherty, D. K., Zhou, L. & Szostak, J. W. Nonenzymatic Template-Directed Synthesis of Mixed-Sequence 3'-NP-DNA up to 25 Nucleotides Long Inside Model Protocells. *J. Am. Chem. Soc.* **141**, 10481–10488 (2019).
[14] Colville, B. W. F. & Powner, M. W. Selective prebiotic synthesis of α-threofuranosyl cytidine by photochemical anomerization. *Angew. Chemie Int. Ed.* **17**, anie.202101376 (2021).
[15] Wu, X., Guntha, S., Ferencic, M., Krishnamurthy, R. & Eschenmoser, A. Base-Pairing Systems Related to TNA: α-Threofuranosyl Oligonucleotides Containing Phosphoramidate Linkages. *Org. Lett.* **4**, 1279–1282 (2002
[16] Eschenmoser, A. The TNA-Family of Nucleic Acid Systems: Properties and Prospects. *Orig. Life Evol. Biosph.* **34**, 277–306 (2004).
[17] Whitaker, D. & Powner, M. W. Prebiotic nucleic acids need space to grow. *Nat. Commun.* **9**, 5172 (2018).
[18] Islam, S. & Powner, M. W. Prebiotic Systems Chemistry: Complexity Overcoming Clutter. *Chem* **2**, 470–501 (2017).
[19] Yadav, M., Kumar, R. & Krishnamurthy, R. Chemistry of Abiotic Nucleotide Synthesis. *Chem. Rev.* **120**, 4766–4805 (2020).
[20] Becker, S. *et al.* Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science* **366**, 76–82 (2019).
[21] Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).
[22] Stairs, S. *et al.* Divergent prebiotic synthesis of pyrimidine and 8-oxo-purine ribonucleotides. *Nat. Commun.* **8**, 15270 (2017).
[23] Islam, S., Bučar, D.-K. & Powner, M. W. Prebiotic selection and assembly of proteinogenic amino acids and natural nucleotides from complex mixtures. *Nat. Chem.* **9**, 584–589 (2017).
[24] Ponnamperuma, C. & Woeller, F. α-Aminonitriles formed by an electric discharge through a mixture of anhydrous methane and ammonia. *Currents in Moden Biology 1*, 156–158 (1967).
[25] Patel, B. H., Percivalle, C., Ritson, D. J., D., D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat Chem* **7**, 301–307 (2015).
[26] Canavelli, P., Islam, S. & Powner, M. W. Peptide ligation by chemoselective aminonitrile coupling in water. *Nature* **571**, 546–549 (2019).
[27] Foden, C. S. *et al.* Prebiotic synthesis of cysteine peptides that catalyze peptide ligation in neutral water. *Science.* **370**, 865–869 (2020).
[28] Powner, M. W., Sutherland, J. D. & Szostak, J. W. Chemoselective Multicomponent One-Pot Assembly of Purine Precursors in Water. *J. Am. Chem. Soc.* **132**, 16677–16688 (2010).
[29] Hein, J. E., Tse, E. & Blackmond, D. G. A route to enantiopure RNA precursors from nearly racemic starting materials. *Nat. Chem.* **3**, 704–706 (2011).
[30] Powner, M. W. & Sutherland, J. D. Phosphate-Mediated Interconversion of Ribo- and Arabino-Configured Prebiotic Nucleotide Intermediates. *Angew. Chem. Int. Ed.* **49**, 4641–4643 (2010).
[31] Ashe, K. *et al.* Selective prebiotic synthesis of phosphoroaminonitriles and aminothioamides in neutral water. *Commun. Chem.* **2**, 23 (2019).
[32] Moutou, G. *et al.* Equilibrium of α-aminoacetonitrile formation from formaldehyde, hydrogen cyanide and ammonia in aqueous solution: Industrial and prebiotic significance. *J. Phys. Org. Chem.* **8**, 721–730 (1995).
[33] Taillades, J. & Commeyras, A. Systemes de strecker et apparentes—I Etude de la decomposition en solution aqueuse des α-alcoyl-aminonitriles tertiaires. Mécanisme d'élimination du groupement nitrile. *Tetrahedron* **30**, 127–132 (1974).
[34] Taillades, J. *et al.* N-carbamoyl-α-amino acids rather than free α-amino acids formation in the primitive hydrosphere: A novel proposal for the emergence of prebiotic peptides. *Orig. Life Evol. Biosph.* **28**, 61–77 (1998).

[35] Ferris, J. P., Sanchez, R. a & Orgel, L. E. Studies in prebiotic synthesis: III. Synthesis of pyrimidines from cyanoacetylene and cyanate. *J. Mol. Biol.* **33**, 693–704 (1968).
[36] Ni, G. *et al.* Review of α-nucleosides: From discovery, synthesis to properties and potential applications. *RSC Adv.* **9**, 14302–14320 (2019).
[37] Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis V. Synthesis and photoanomerization of pyrimidine nucleosides. *J. Mol. Biol.* **47**, 531–543 (1970).
[38] Xu, J. *et al.* A prebiotically plausible synthesis of pyrimidine β-ribonucleosides and their phosphate derivatives involving photoanomerization. *Nat. Chem.* **9**, 303–309 (2017).
[39] Grosjean, H., de Crécy-Lagard, V. & Marck, C. Deciphering synonymous codons in the three domains of life: Co-evolution with specific tRNA modification enzymes. *FEBS Lett.* **584**, 252–264 (2010).
[40] Ajitkumar, P. & Cherayil, J. D. Thionucleosides in transfer ribonucleic acid: diversity, structure, biosynthesis, and function. *Microbiol. Rev.* **52**, 103–113 (1988).
[41] Testa, S. M., Disney, M. D., Turner, D. H. & Kierzek, R. Thermodynamics of RNA-RNA duplexes with 2- or 4-thiouridines: Implications for antisense design and targeting a group I intron. *Biochemistry* **38**, 16655–16662 (1999).
[42] Heuberger, B. D., Pal, A., Del Frate, F., Topkar, V. V. & Szostak, J. W. Replacing uridine with 2-thiouridine enhances the rate and fidelity of nonenzymatic RNA primer extension. *J. Am. Chem. Soc.* **137**, 2769–2775 (2015).
[43] Prywes, N., Michaels, Y. S., Pal, A., Oh, S. S. & Szostak, J. W. Thiolated uridine substrates and templates improve the rate and fidelity of ribozyme-catalyzed RNA copying. *Chem. Commun.* **52**, 6529–6532 (2016).
[44] Ohkubo, A. *et al.* Formation of new base pairs between inosine and 5-methyl-2-thiocytidine derivatives. *Org. Biomol. Chem.* **10**, 2008–2010 (2012).
[45] Kim, S. C., O'Flaherty, D. K., Zhou, L., Lelyveld, V. S. & Szostak, J. W. Inosine, but none of the 8-oxo-purines, is a plausible component of a primordial version of RNA. *Proc. Natl. Acad. Sci.* **115**, 13318–13323 (2018).
[46] Gibard, C., Bhowmik, S., Karki, M., Kim, E. K. & Krishnamurthy, R. Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nat. Chem.* **10**, 212–217 (2018).
[47] Lohrmann, R. Formation of nucleoside 5'-polyphosphates from nucleotides and trimetaphosphate. *J. Mol. Evol.* **6**, 237–252 (1975).
[48] Yamagata, Y. Prebiotic formation of ADP and ATP from AMP, calcium phosphates and cyanate in aqueous solution. *Orig. Life Evol. Biosph.* **29**, 511–520 (1999).
[49] Feldmann, W. & Thilo, E. Zur Chemie der kondensierten Phosphate und Arsenate. XXXVIII. Amidotriphosphat. *Zeitschrift für Anorg. und Allg. Chemie* **328**, 113–126 (1964).
[50] Yamagata, Y., Watanabe, H., Saitoh, M. & Namba, T. Volcanic production of polyphosphates and its relevance to prebiotic evolution. *Nature* **352**, 516–519 (1991).