Selective prebiotic formation of RNA pyrimidine and DNA purine nucleosides

Selective prebiotic formation of RNA pyrimidine and DNA purine nucleosides

Jianfeng Xu[1]†, Václav Chmela[1]†, Nicholas J. Green[1], David A. Russell[1], Mikołaj J. Janicki[2], Robert W. Góra[2], Rafał Szabla[3,4], Andrew D. Bond[5] and John D. Sutherland[1]*

[1]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, CB2 0QH, UK.

[2]Department of Physical and Quantum Chemistry, Wrocław University of Science and Technology, Faculty of Chemistry, Wybrzeże Wyspiańskiego 27, 50-370, Wrocław, Poland.

[3]EaStCHEM, School of Chemistry, University of Edinburgh, Joseph Black Building, David Brewster Road, Edinburgh, EH9 3FJ, UK.

[4]Institute of Physics, Polish Academy of Sciences, Al. Lotników 32/46, PL-02668 Warsaw, Poland.

[5]Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW, UK.

*Correspondence to: johns@mrc-lmb.cam.ac.uk

†These authors contributed equally to this work.

The nature of the first genetic polymer is the subject of major debate in the origin of life field[1]. Although the common 'RNA world' theory suggests RNA as the first replicable information carrier at the dawn of life, other evidence implies that life may have started with a heterogeneous nucleic acid genetic system including both RNA and DNA[2]. Such a theory streamlines the eventual 'genetic takeover' of homogeneous DNA from RNA as the principal information storage molecule in the central dogma, but requires a selective abiotic synthesis of both RNA and DNA building blocks in the same local primordial geochemical scenario. Herein, we demonstrate a high-yielding, completely stereo-, regio-, and furanosyl-selective prebiotic synthesis of the purine deoxyribonucleosides,

deoxyadenosine and deoxyinosine. Our synthesis utilizes key intermediates in the prebiotic synthesis of the canonical pyrimidine ribonucleosides, and we show that, once generated, the pyrimidines persist throughout the synthesis of the purine deoxyribonucleosides, ultimately leading to a mixture of deoxyadenosine, deoxyinosine, cytidine, and uridine. These results support the notion that purine deoxyribonucleosides and pyrimidine ribonucleosides may have coexisted before the emergence of life[3].

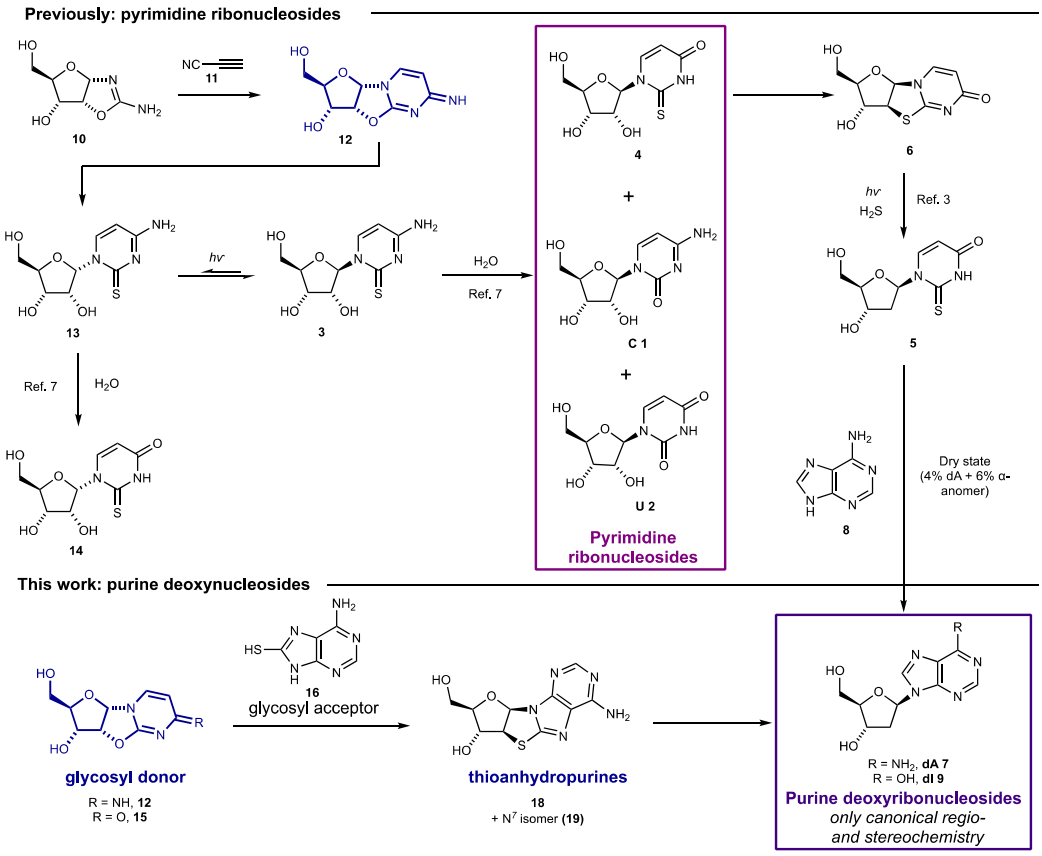
Introduction

The advent of life requires informational inheritance mediated by a suitable polymer that can undergo replication in the absence of enzymes. The RNA world hypothesis invokes RNA as this polymer[4, 5]. Considerable progress in the prebiotic synthesis of the pyrimidine ribonucleosides of RNA, cytidine (C) **1** and uridine (U) **2**, and their 2-thio derivatives, **3** and **4**[6, 7], together with recent advances in non-enzymatic RNA replication[8-10] have given credence to the RNA world theory. Progress towards the abiotic synthesis of purine nucleosides has been made, but only using routes that employ as starting materials chemically and enantiomerically pure sugars[11-15], which are not likely to be have been found on the primordial earth. Additionally, no prebiotically plausible route has been shown to provide a mixture containing a competent set of nucleosides for information storage at the polymeric level.

Extant biology, in contrast to the proposed RNA world, utilizes DNA as the central information-carrying molecule. This discrepancy between the RNA world and modern biology requires a 'genetic takeover' that invokes the power of primitive biosynthetic machinery and natural selection operating over millions of years, ultimately resulting in an ancestral biosynthetic route to DNA[16]. The superior

hydrolytic stability and replication fidelity[17] of DNA could have resulted in selection of primitive organisms capable of synthesizing DNA, and thus its rise to prominence in the central dogma, but the feasibility of this evolutionary process in a pre-DNA world is debated[1]. To circumvent this potentially problematic transition, an R/DNA world has been proposed, in which nascent biology had access to both RNA and DNA building blocks from the outset, without requiring elaborate biosynthesis[18-20]. In such a world, heterogeneous polymers would have initially been most common, but polymers with increased homogeneity, and hence properties closer to either that of RNA or DNA, would have been selected for over their mixed counterparts[2]. For the R/DNA world to be plausible, an efficient prebiotic synthesis of DNA building blocks is required, and one that provides building blocks for both RNA and DNA in the same localized geochemical scenario is preferable. We recently demonstrated proof of this principle by showing that 2'-deoxy-2-thiouridine **5** – a non-canonical deoxynucleoside – can be synthesized from thioanhydrouridine **6** – an RNA derivative – by way of a prebiotically plausible, hydrogen sulfide-mediated photoreduction[3]. Although this finding provides an important prebiotic link between RNA and DNA building blocks, the lability of **5** to hydrolysis may limit its phosphorylation and subsequent oligomerization[21, 22]. Additionally, the synthesis of canonical deoxyadenosine (dA) **7** from **5** and adenine **8** was low yielding (4%), and generated a more abundant undesired side product, the α-anomer of **7** (6%). Using guidance from a geochemical scenario[23], we now demonstrate a synthesis of purine deoxynucleosides that is based on prebiotically plausible reactions and substrates. We then evaluate our route at a systems level by enacting the synthesis on mixtures of materials likely to arise in a primordial environment, culminating in the demonstration of multiple reaction

sequences able to selectively furnish a mixture of U (**1**), C (**2**), dA (**7**) and deoxyinosine (dI, **9**).



Results and Discussion

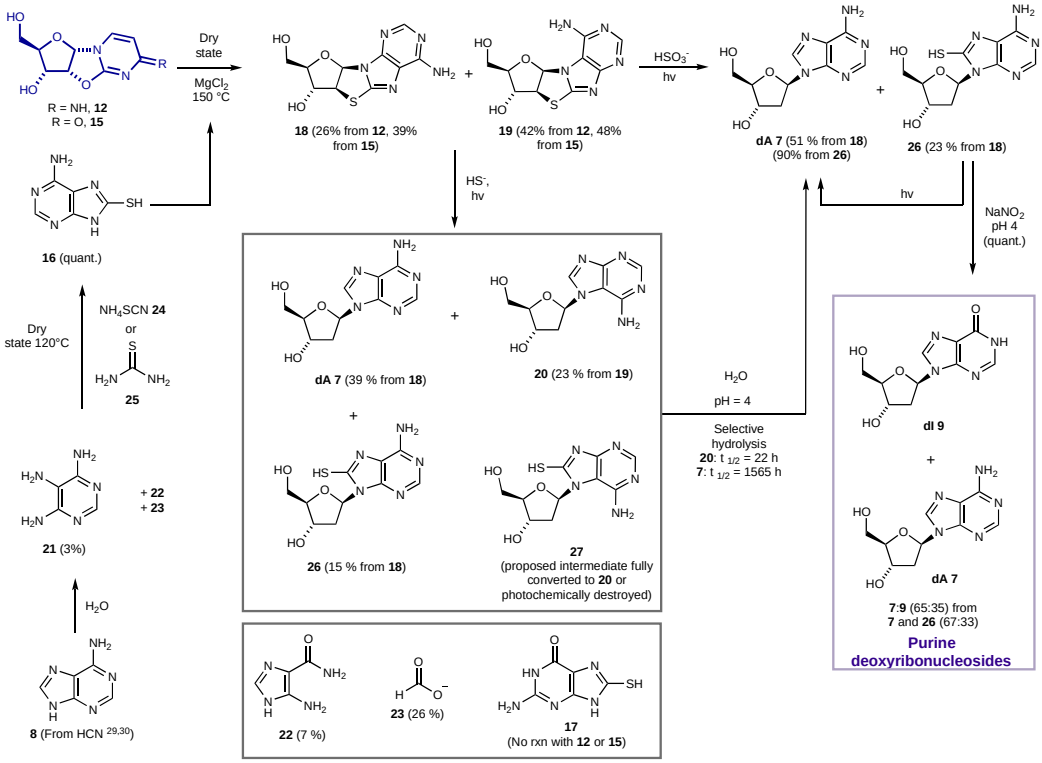
Prebiotically Guided Route to Purine Deoxyribonucleosides

A route to purine nucleosides that diverges from a prebiotic RNA synthesis is attractive because it implies that the constituents of a set of nucleosides capable of storing information – pyrimidines and purines – may have formed in the same location on a primordial Earth, rather than having been necessarily brought together by environmental processes after their separate formation. To develop such a route, we evaluated intermediates in the prebiotic RNA pyrimidine nucleoside synthesis[6, 7]

as ribosyl donors (Fig. 1). The RNA synthesis proceeds from RAO **10** which reacts with cyanoacetylene **11** to provide α-anhydrocytidine **12**. Thiolysis of **12** in formamide produces α-2-thiocytidine **13** which undergoes efficient UV-mediated photoanomerisation to 2-thiocytidine **3,** which hydrolyses to the canonical pyrimidines cytidine **1** and uridine **2**, and biologically important non-canonical pyrimidine **4.** Alternatively, in the dark, **13** is hydrolysed to α-2-thiouridine **14**[7]. Whilst **14** appeared initially only a by-product that would be produced in the dark on the early Earth, it is readily cyclised to anhydrouridine **15** at 80 °C (63% yield in water or 89% yield in formamide, Extended Data Fig. 2). We recognised α-anhydropyrimidines **12** and **15** as ideal glycosyl donors for 1',2'-cis tethered glycosylation[24]. Since the sugar of **12** and **15** is fixed in its furanosyl form, the formation of pyranosyl nucleosides – one of the critical downfalls of previous strategies – should be excluded. Additionally, the α-stereochemistries of C1' and C2' of **12** and **15** led us to expect transglycosylation to provide only β-anomers, the correct stereochemistry at C1' for all natural (deoxy)ribonucleosides. Finally, since **12** and **15** are ultimately derived from ribo-aminooxazoline (RAO) **10**, which crystallizes enantiopure from solutions of minimally enantioenriched carbohydrates or amino acids[25, 26], this route offered the so-far unmet potential to deliver enantio- and diastereomerically pure furanosyl-nucleosides by glycosylation.

Accordingly, we evaluated 8-mercaptoadenine **16** and 8-mercaptoguanine **17** as potential nucleophiles to participate in transglycosylation with **12** and **15** (Fig 2). Although **17** proved unreactive, **16** reacts with **12** and **15** at 150 °C in the dry state (Fig. 2), to provide two new β-configured nucleoside products in moderate yields (14% and 16% respectively from **15**, trace amounts from **12**). The minor product was

determined to be N^9-8, 2'-anhydro-thioadenosine **18** by X-ray crystallography and ^{1}H-NMR spiking experiments with a synthetic standard. The major product was inferred to be N^7-8,2'-anhydro-thioadenosine **19**, the regioisomer of **18**, by its subsequent conversion to 2'-deoxy-N^7-adenosine **20**. The presence of magnesium chloride in the reaction, presumably acting as a Lewis acid[27], dramatically improved the yield of **18** and **19** to 39% and 48% respectively from **15** (combined yield 87%) and 26% and 42% respectively from **12** (combined yield 68%). Thus, in a prebiotic environment where **12** or **15** and **16** are brought together, perhaps by converging streams that then undergo evaporation, **18** and **19** could be readily generated, especially in the presence of magnesium ions[28].



Any prebiotic synthesis requires a viable route to all reagents from plausible early-Earth feedstocks. We were drawn towards adenine **8** as a starting point for the

provision of 8-mercaptoadenine **16**, due to its widely accepted prebiotic plausibility as a relatively stable pentamer of hydrogen cyanide[29, 30]. Remarkably, despite the reactivity of related purines[31], adenine did not react with elemental sulfur at temperatures up to 300 °C. However, adenine does undergo (slow) hydrolysis in aqueous media. Miller et. al. reported a half-time for hydrolysis of adenine of about 1 year at 100 °C, and identified (but did not quantify) 4,5,6-triaminopyrimidine **21** (TAP) among the products of hydrolysis[32]. We reinvestigated this hydrolysis of adenine **8**, under conditions more suited to a laboratory time-scale (138 °C, phosphate buffer pH 8), and at partial conversion after 10–12 days confirmed the presence of TAP in yields of 2–3% (8–9% based on recovered adenine) (Fig. 2). Due to the differential solubilities of adenine and TAP, the supernatants of adenine hydrolysis reactions are enriched in TAP after cooling. A typical supernatant contains 5-aminoimidazole-4-carboxamide **22**, TAP **21**, and adenine **8** in a 4:2:1 ratio, and formate **23** as the only other major component (See Fig. S1–S5 for full details). We found that TAP (either commercially supplied or that in the crude adenine hydrolysate) is converted to 8-mercaptoadenine **16** by heating in the dry state with either ammonium thiocyanate **24** or thiourea **25**. **24** is an inevitable by-product of the photochemistry of hydrogen cyanide and hydrogen sulfide[33], two precursors likely to have been abundant on the primordial earth, and heavily implicated in the origin of life by our cyanosulfidic chemical network[23]. Thiourea **25** has also been widely invoked as a prebiotically plausible reagent[34]. Thus, we envision that a primordial environment supplied with adenine and water would continuously generate TAP, which can be enriched in aqueous solution by moving down a thermal gradient. Ammonium thiocyanate **24** can be mixed with the TAP at any stage, and eventual evaporation and dry state reaction leads to 8-mercaptoadenine **16**. This method of

accumulation of TAP also improves the plausibility of some aspects of other prebiotic syntheses[12].

With thioanhydropurine nucleosides **18** and **19** in hand, we moved on to evaluate their photoreduction chemistry to see if we might directly generate deoxyadenosine. Our previous synthesis of a deoxypyrimidine *via* a thioanhydropyrimidine **6** (Fig. 1) proceeded by the reduction of a C–S to a C–H bond mediated by a hydrated electron, generated by UV irradiation of hydrosulfide[3, 33]. **18** and **19** were separately subjected to UV irradiation at 254 nm in water with hydrogen sulfide (H_2S) as the reductant (Fig. 2). In the photoreduction of **18**, the natural regio-isomer N^9-deoxyadenosine **7** (dA) was detected in 39% yield, along with 15% of 8-mercapto-deoxyadenosine **26**. **26** was demonstrated to be a competent intermediate in the reaction by desulfurization to give **7** (dA) either by UV irradiation[35], or treatment with nitrous acid, which is produced from common atmospheric gases, nitrogen and carbon dioxide[36]. Nucleobase loss was also apparent (8-mecaptoadenine **16** in 10% yield and adenine **8** in 17% yield). The same reaction starting with **19** gave N^7-deoxyadenosine **20** in 23% yield with no other nucleoside products. Our proposed intermediate in this process, 8-mercapto-N^7-deoxyadenosine **27**, is either fully converted to **20** or photochemically destroyed. Photoreduction was also carried out on a mixture of **18** and **19** compatible with our synthesis by tethered transglycosylation. The ratio of N^9:N^7 regioisomers was increased from 38:62 of **18**:**19** in the starting mixture to 56:44 of **7**:**20** after photoreduction (31% yield for **7**, 17% yield for **20**), indicating an enhanced stability of intermediates or products bearing the natural N^9 glycosidic linkage, compared to N^7 isomers. Replacing hydrosulfide as the electron donor with bisulfite (HSO_3^-, pH 7)[37], which is readily formed by the dissolution of atmospheric SO_2 in water[38], improved both the yield and selectivity of

photoreduction. Photoreduction with bisulfite of **18** alone provided deoxyadenosine **7** (dA) in 51% yield and 8-mercapto-deoxyadenosine **26** in 23% yield, while a similar reaction with the N^7-regio-isomer **19** led only to its photochemical destruction. Photoreduction of a mixture of **18** and **19** with bisulfite led only to N^9-linked products, **7** and **26** in 44% and 18% yield respectively (Extended Data Fig. 3). Separate experiments probing the stability of starting materials and products under the reaction conditions indicated that the relative stabilities of intermediates are the cause of this selectivity. This strikingly selective destruction is highly suggestive of a potential mechanism by which primordial nucleosides were restricted to a near-canonical set[39, 40]. We found further evidence for such restriction in the hydrolysis rates of the N^9 and N^7 isomers of deoxyadenosine. In acetate buffer (pH 4, room temperature), the natural isomer **7** (dA) is more than 70 times more stable than **20** **(**half-lives of 1565 and 22 hours respectively), which is consistent with the reported difference in stabilities towards acid hydrolysis between the corresponding isomers of adenosine[41, 42].

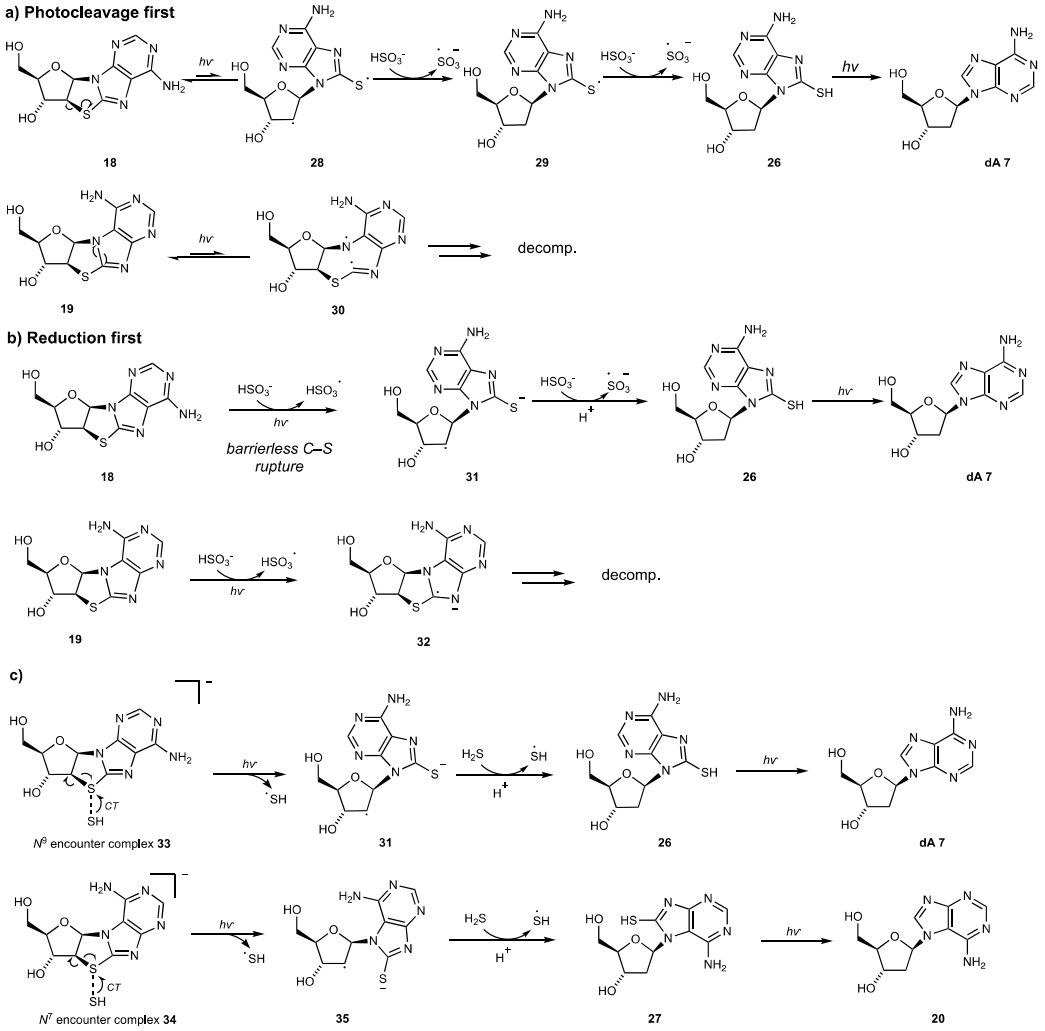
Photoreduction Mechanism

To provide mechanistic rationale for the observed photochemical selectivity, we performed quantum chemical calculations using density functional theory and algebraic diagrammatic construction to the second order [ADC(2)] methods[43, 44]. These calculations revealed, in the case of bisulfite, two possible competing mechanisms that explain the difference in reactivity of the two regioisomers. **18** and **19** can both undergo photoexcitation, but generate dissimilar biradical species (Fig. 3a). Photoexcitation of **18** leads to rupture of the C2'–S bond on the surface of the lowest excited singlet (S_1) state, generating biradical **28** (Fig 3a, N^9; Extended Data Fig. 4a). Reduction of this species by intermolecular hydrogen atom transfer (HAT)

205 or proton-coupled electron transfer (PCET) is likely to lead to C2'-reduced species **29**,
206 and ultimately, *via* a second HAT or PCET and subsequent photolysis of the C8–S
207 bond of **26**[35], deoxyadenosine **7** (dA) (Fig. 3a, N^9). In contrast, photoexcitation of **19**
208 leads to N7–C8 bond rupture through the S_1/S_0 state crossing (Fig. 3a, N^7; Extended
209 Data Fig. 4b), generating **30**, which is likely to undergo decomposition without C2'–S
210 reduction. Since bisulfite is well-known to provide a hydrated electron upon
211 irradiation[45], a second possibility is the reduction by hydrated electrons of **18** and **19**
212 in the ground state. Again, calculations suggest different fates of **18** and **19** upon
213 reduction. Reduction of **18** is predicted to proceed with concomitant barrierless C2'–S
214 bond rupture to give radical anion intermediate **31** (Fig. 3b, N^9; Extended Data Fig. 5)
215 whereas reduction of **19** is predicted to lead to formation of a C8, N9 radical anion **32**
216 which also is likely to undergo decomposition rather than C2' reduction (Fig. 3b N^7,
217 Extended Data Fig. 5). In the absence of any reducing agent, both **18** and **19** undergo
218 (equally) slow photochemical decomposition, presumably via the calculated biradical
219 structures **28** and **30**, but in the presence of bisulfite, reduction of the ground state or
220 photochemically generated intermediates results in remarkably different fates.

221 The successful reduction of **19** alongside **18** when using hydrosulfide as the
222 reducing agent is explained by a distinct mechanism. Calculations located stable
223 encounter complexes, **33** and **34**, between HS^- and thioanhydronucleosides **18** and **19**,
224 respectively (Fig. 3c, Extended Data Fig. 4c and 4d). This interaction is
225 predominantly stabilized by electrostatic and dispersion interactions and our
226 interaction energy decomposition demonstrates its stability in aqueous solution (see
227 the SI for detail). Similar S···S interactions were recently identified in intramolecular
228 complexes and were classified as chalcogen bonds[46]. Such an encounter complex
229 facilitates charge transfer (CT) from the hydrosulfide anion to the thioanhydropurine

fragment almost immediately after UV absorption by the complex to the S_1 state. Subsequent relaxation on the S_1 surface enables practically barrierless C2'–S bond breaking completed by a peaked S_1/S_0 state crossing for both intermediates **31** and **35**, thus facilitating C2'–S reduction of both **18** and **19** (Extended Data Fig. 4c and 4d). The products of this photochemical transformation, **26** and **27**, may further undergo photochemical sulfur cleavage through the mechanism described by Roberts *et al.*[35] (Fig. 3c). Thus, a HS^- thioanhydropurine encounter complex facilitates C–S bond cleavage and partially protects N^7 isomer **19** from the photodestruction observed in the presence of bisulfite. This finding not only explains the distinctive outcomes of photoreduction between the two reducing agents, but also points towards a potentially important stabilising role for hydrosulfide in prebiotic chemistry and photochemistry in general.



Prebiotic Route to A Purine/Pyrimidine Genetic System

Whilst our attempts to glycosylate 8-mercaptoguanine **17** to provide thioanhydroguanosine (and ultimately deoxyguanosine) failed, the triple selectivity and high yield of our route to deoxyadenosine combined with recent results from the Szostak group[47] suggest a possible alternative genetic alphabet that does not include (deoxy)guanosine. Guanosine is yet to succumb to a plausible prebiotic synthesis, but Szostak *et al.* have recently shown that inosine (I), which is capable of base-pairing with cytosine, can replace guanosine in non-enzymatic RNA replication systems with no loss of rate or fidelity. (Deoxy)adenosine **7** (dA) is readily converted to

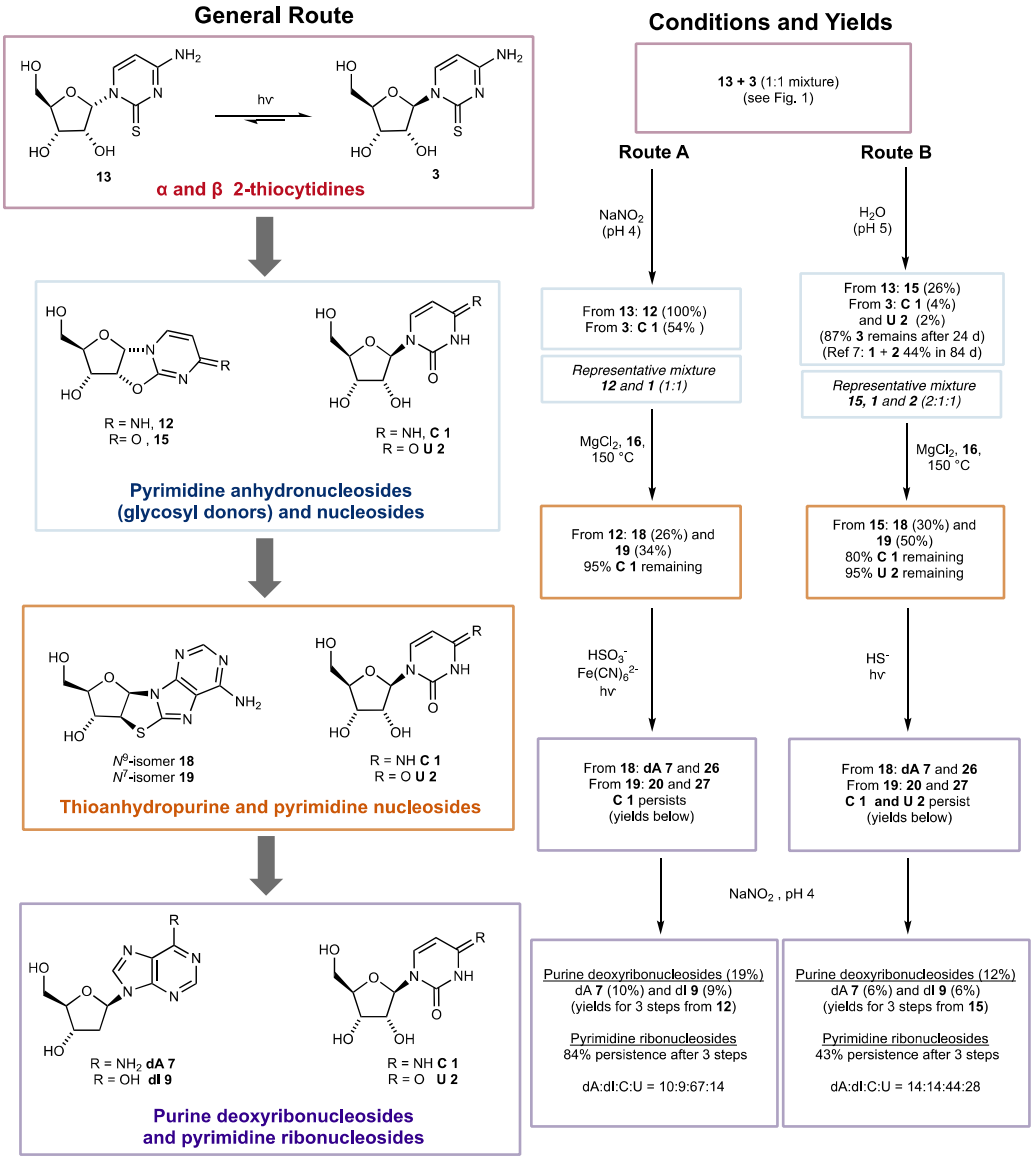
(deoxy)inosine **9** (dI) (Fig. 2) by deaminative hydrolysis, which spontaneously occurs very slowly in nucleic acid polymers[48], and is greatly accelerated by the presence of nitrous acid[49]. To demonstrate that this conversion can occur under mild conditions consistent with our primordial geochemical scenario[50], we treated deoxyadenosine **7** (dA) with nitrous acid at pH 4 (the same conditions by which we could effect desulfurization of **26**). After four days at room temperature, approximately 40% of **7** (dA) had been converted to **9** (dI), providing a 60:40 mixture of **7** (dA) and **9** (dI) (Fig 2). A control experiment monitoring the decomposition of deoxyadenosine **7** (dA) at pH 4, without nitrous acid, showed only a trace of depurination ($t_{1/2}$ = 1600 h). When a 67:33 mixture of **7** and **26**, representative of the outcome of photoreduction, was submitted to the reaction conditions, **26** underwent relatively rapid desulfurization first, with deoxyadenosine **7** (dA) undergoing slower deaminative hydrolysis to ultimately provide a 65:35 mixture of **7** (dA) and **9** (dI). Thus, mixtures of deoxyadenosine **7** (dA) and deoxyinosine **7** (dA) are readily obtainable from partial deaminative hydrolysis of deoxyadenosine **7** (dA) or its precursor **26**, thereby supplying half of a potential primordial alphabet. Despite the potential for a mismatch in reactivity between deoxypurines and pyrimidines, a 47:53 mixture of deoxyadenosine **7** (dA) and cytidine **1** (C) underwent nitrous acid-promoted deamination to provide all four (deoxy)nucleosides deoxyadenosine **7** (dA), deoxyinosine **9** (dI), cytidine **1** (C), and uridine **2** (U) (30:17:42:11 ratio) (Extended Data Fig. 6). A similar primordial mixture may have been a starting point for genetic information storage. Furthermore, in the absence of significant geochemically plausible sources of pyrimidine deoxynucleotides and purine ribonucleotides, heteropolymers made from a mixture of purine deoxyribonucleotides and pyrimidine ribonucleotides should possess heritable backbone heterogeneity and thus a 1:1

phenotype to genotype correspondence, which is potentially advantageous in the evolution of catalytic activity[18].

Systems Level Prebiotic Plausibility

Having demonstrated the potential of a divergent route to yield a local mixture of **7** (dA), **9** (dI), **1** (C) and 2 (U), we sought to evaluate the key question of whether all four nucleosides could persist after divergence in the sequence. We chose a 1:1 mixture of α– and β–2-thiouridines **13** and **3** as our starting point, which could be obtained from the partial photoanomerisation of **13**, and evaluated two particular combinations of reactions as representative permutations of a primordial geochemical process (Fig. 4, Route A and B). In route A, exposure of the mixture to nitrous acid (pH 4) generates a mixture of **12** and **1** (100% yield for **12** from **13**, 54% yield for **1** from **3**). **12** is formed from **13** by intramolecular addition of the C2' hydroxyl to C2 of an S-nitrosyl intermediate, and subsequent elimination of SNO^-. Dry state glycosylation of **16** and a 1:1 mixture of **12** and **1** (C), in the presence of $MgCl_2$, leads to a mixture of **18** and **19** as described in our route development above, however, critically, 95% of **1** persists in this mixture. Subsequent photoreduction in the presence of ferrocyanide and bisulfite generates the expected mixture of purine nucleosides **7** (dA), **26**, **20** and **27** alongside **1** (C). Finally, a second exposure to nitrous acid converts this mixture into the components of a competent genetic system, **7** (dA), **9** (dI) (10% and 9% yield respectively from **12** for 3 steps), **1** (C) and **2** (U) (84% combined persistence after 3 steps) with no significant nucleoside impurities. Products derived from **19** – with the wrong N^7 regiochemistry – are hydrolysed in the last step. It is noteworthy that this route is only viable from a systems level approach – for instance, the pyrimidines are fairly rapidly destroyed in the photoreduction step

in the absence of the thioanhydropurines (Extended Data Fig. 7). Route B presents an alternative in which initial hydrolysis of the mixture of **13** and **3** generates glycosyl donor **15** (26% yield) alongside pyrimidine nucleosides (4% of **1** (C), 2% of **2** (U), 92% **3** remaining). **3** has previously been shown to hydrolyse to **1** (C) and **2** (U) in greater yields (44%) over longer periods[7]. A representative mixture of **15**, **1** (C) and **2** (U) (2:1:1) was then subjected to tethered glycosylation, resulting in **18** and **19** as above (30% and 50% yield respectively) with 80% and 95% persistence of **1** (C) and **2** (U). Photoreduction of the mixture, this time with hydrogen sulfide, provides purine products **7**, **26**, **20** and **27** alongside the pyrimidines **1** (C) and **2** (U). Finally, nitrosation furnished the key mixture of **7** (dA) and **9** (dI) (6% for each from **15** for three steps) alongside pyrimidine nucleosides (43% persistence over 3 steps, final ratio of dA:dI:C:U in the mixture is 14:13:45:28, Extended Data Fig. 8). Thus, sequences comprised of various orders of operations and various photoreduction conditions, which might plausibly emulate a terrestrial geochemical scenario, generate the components of a mixed genetic system alongside one another. The exact ratio of **1** (C) and **2** (U) (ribosylpyrimidines) to **7** (dA) and **9** (dI) (deoxyribosylpurines) in the final mixture will depend on the ratio of α–(anhydro)pyrimidines (**13**, **12**, and **15**) to β–(thio)pyrimidines (**1**, **2** and **3**) earlier in the sequence, which will vary based on environmental conditions.



In conclusion, a highly efficient synthesis of both deoxyadenosine **7** (dA) and deoxyinosine **9** (dI), requiring only prebiotically plausible reagents and conditions, is reported. In contrast to all previous attempts to synthesize purine nucleosides, our synthesis is both prebiotically plausible and strictly stereo-, regio-, and furanosyl-selective for the only isomer of the deoxypurine nucleosides used in modern biology. The pathway proceeds mostly via simple hydrolysis or dry state processes, with a key reduction step promoted by UV irradiation supported by distinct mechanisms. The (photo)chemical selection exhibited by this route hints at an explanation for Nature's

choice of one isomer of nucleic acid from the many that are conceivable. Critically, we have demonstrated that sequences leading selectively to both RNA pyrimidine and DNA purine nucleosides can occur together simultaneously, providing mixtures which could conceivably complete a genetic alphabet. The fact that DNA building blocks can be co-produced with the RNA pyrimidine nucleosides is consistent with and perhaps evidence for the coexistence of RNA and DNA building blocks at the dawn of life.

Data and materials availability: Supplementary Information is available containing all procedures, characterization data, NMR spectra, HPLC traces, X-Ray data and CCDC numbers, and theoretical methods and data. Any additional data are available from the corresponding author upon reasonable request.

Code availability: All custom code used to generate the data in this study is available upon reasonable request.

References:

1. Samanta, B & Joyce, G. F. A reverse transcriptase ribozyme. *Elife* **6**, e31153 (2017).

2. 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).

3. Xu, J., Green, N. J., Gibard, C., Krishnamurthy, R. & Sutherland, J. D. Prebiotic phosphorylation of 2-thiouridine provides either nucleotides or DNA building blocks via photoreduction. *Nat. Chem.* **11**, 457-462 (2019).

4. Gilbert, W. Origin of life: The RNA world. *Nature* **319**, 618 (1986).

5. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418**, 214-221 (2002).

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

7. Xu, J. *et al.* A prebiotically plausible synthesis of pyrimidine beta-ribonucleosides and their phosphate derivatives involving photoanomerization. *Nat. Chem.* **9**, 303-309 (2017).

8. 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).

9. Walton, T. & Szostak, J. W. A highly reactive imidazolium-bridged dinucleotide intermediate in nonenzymatic RNA primer extension. *J. Am. Chem. Soc.* **138**, 11996-12002 (2015).

10. Li, L. *et al.* Enhanced nonenzymatic RNA copying with 2-aminoimidazole activated nucleotides. *J. Am. Chem. Soc.* **139**, 1810-1813 (2017).

11. Fuller, W. D., Orgel, L. E. & Sanchez, R. A. Studies in Prebiotic Synthesis: VI. Solid-State Synthesis of Purine Nucleosides. *J. Mol. Evol.* **1**, 249-257 (1972).

12. Becker S. et al. A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352**, 833-836 (2016).
13. Kim, H. & Benner, S. A. Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotides from nucleobases and phosphorylated carbohydrates. *Proc. Nat. Acad. Sci. USA* **114**, 11315-11320 (2017).
14. Becker S. et al. Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science* **366**, 76-82 (2019).
15. Teichert, J. S., Kruse, F. M. & Trapp, O. Direct prebiotic pathway to DNA nucleosides. *Angew. Chem. Int. Ed.* **55**, 9944-9947 (2019).
16. Reichard, P. From RNA to DNA, why so many ribonucleotide reductases? *Science* **260**, 1773-1777 (1993).
17. Leu, K., Obermayer, B., Rajamani, S., Gerland, U. & Chen, I. A. The prebiotic evolutionary advantage of transferring genetic information from RNA to DNA. *Nucleic Acids Res.* **39**, 8135-8147 (2011).
18. Sutherland, J. D. & Whitfield, J. N. Prebiotic chemistry: a bioorganic perspective. *Tetrahedron* **53**, 11493-11527 (1997).
19. Trevino, S. G., Zhang, N., Elenko, M. P., Lupták, A. & Szostak, J. W. Evolution of functional nucleic acids in the presence of nonheritable backbone heterogeneity. *Proc. Nat. Acad. Sci. USA* **108**, 13492-13497 (2011).
20. Gavette, J. V., Stoop, M., Hud, N. V. & Krishnamurthy, R. RNA–DNA chimeras in the context of an RNA world transition to an RNA/DNA world. *Angew. Chem. Int. Ed.* **55**, 13204-13209 (2016).
21. Schoffstall, A. M. Prebiotic phosphorylation of nucleosides in formamide. *Orig. Life* **7**, 399-412 (1976).

22. Lohrmann, R. & Orgel, L. E. Urea-Inorganic Phosphate Mixtures as Prebiotic Phosphorylating Agents. *Science* **171**, 490-494 (1971).

23. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **7**, 301-307 (2015).

24. Ishiwata, A., Lee, Y. J. & Ito, Y. Recent advances in stereoselective glycosylation through intramolecular aglycon delivery. *Org. Biomol. Chem.* **8**, 3596-3608 (2010).

25. Springsteen, G. & Joyce, G. F. Selective derivatization and sequestration of ribose from a prebiotic mix. *J. Am. Chem. Soc.* **126**, 9578-9583 (2004).

26. Anastasi, C., Crowe, M. A., Powner, M. W. & Sutherland, J. D. Direct Assembly of Nucleoside Precursors from Two- and Three-Carbon Units. *Angew. Chem. Int. Ed.* **45**, 6176-6179 (2006).

27. Vorbrüggen, H. & Ruh-Pohlenz, C. *Handbook of nucleoside synthesis*. (Wiley, 2001).

28. Holm, N. G., Oze, C., Mousis, O., Waite, J. H. & Guilbert-Lepoutre, A. Serpentinization and the formation of H_2 and CH_4 on celestial bodies (planets, moons, comets). *Astrobiology* **15**, 587-600 (2015).

29. Sanchez, R. A., Ferris, J. P. & Orgel, L. E. Studies in prebiotic synthesis II: Synthesis of purine precursors and amino acids from aqueous hydrogen cyanide. *J. Mol. Biol.* **80**, 223-253 (1967).

30. Hudson, J. S. *et al*. A unified mechanism for abiotic adenine and purine synthesis in formamide. *Angew. Chem. Int. Ed.* **51**, 5134-5137 (2012).

31. Giner-Sorolla, A., Thom, E. & Bendich, A. Studies on the Thiation of Purines. *J. Org. Chem.* **29**, 3209-3212 (1964).
32. Levy, M. & Miller, S. L. The stability of the RNA bases: implications for the origin of life. *Proc. Natl. Acad. Sci. USA* **95**, 7933-7938 (1998).
33. Ritson, D. J. & Sutherland, J. D. Synthesis of aldehydic ribonucleotide and amino acid precursors by photoredox chemistry. *Angew. Chem. Int. Ed.* **52**, 5845-5847 (2013).
34. Robertson, M. P., Levy, M. & Miller, S. L. Prebiotic synthesis of diaminopyrimidine and thiocytosine. *J. Mol. Evol.* **43**, 543-550 (1996).
35. Roberts, S. J. *et al*. Selective prebiotic conversion of pyrimidine and purine anhydronucleosides into Watson-Crick base-pairing arabino-furanosyl nucleosides in water. *Nat. Commun.* **9**, 4073-4082 (2018).
36. Ranjan, S., Todd, Z. R., Rimmer, P. B., Sasselov, D. D. & Babbin, A. R. Nitrogen oxide concentrations in natural waters on early Earth. *Geochem. Geophy. Geosy*. **20**, 2021-2039 (2019).
37. Xu, J. *et al*. Photochemical reductive homologation of hydrogen cyanide using sulfite and ferrocyanide. *Chem. Commun.* **54**, 5566-5569 (2018).
38. Marion, G. M., Kargel, J. S., Crowley, J. K. & Catling, D. C. Sulfite–sulfide–sulfate–carbonate equilibria with applications to Mars. *Icarus*, **225**, 342–351 (2013).
39. Rios, A. C. & Tor, Y. On the origin of the canonical nucleobases: an assessment of selection pressures across chemical and early biological evolution. *Isr. J. Chem.* **53**, 469 – 483 (2013).

40. Rios, A. C., Yu, H. T. & Tor, Y. Hydrolytic fitness of *N*-glycosyl bonds: comparing the deglycosylation kinetics of modified, alternative, and native nucleosides. *J. Phys. Org. Chem.* **28,** 173-180 (2014).
41. Panzica, R. P., Rousseau, R. J., Robins, R. K., & Townsend, L. B. Relative stability and a quantitative approach to the reaction mechanism of the acid-catalyzed hydrolysis of certain 7-and 9-β-D-ribofuranosylpurines. *J. Am. Chem. Soc.* **94**, 4708-4714 (1972).
42. Lindahl, T. & Nyberg, B. Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **11**, 3610-3618 (1972).
43. Hättig, C. Structure Optimizations for Excited States with Correlated Second-Order Methods: CC2 and ADC(2). *Adv. Quantum Chem.* **50**, 37–60 (2005).
44. Dreuw, A. & Wormit, M. The algebraic diagrammatic construction scheme for the polarization propagator for the calculation of excited states. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **5**, 82–95 (2015).
45. Sauer, M. C., Crowell, R. A. & Shkrob, I. A. Electron Photodetachment from Aqueous Anions. 1. Quantum Yields for Generation of Hydrated Electron by 193 and 248 nm Laser Photoexcitation of Miscellaneous Inorganic Anions. *The Journal of Physical Chemistry A* **108**, 5490-5502 (2004).
46. Pascoe, D. J., Ling, K. B. & Cockroft, S. L. The origin of chalcogen-bonding interactions. *J. Am. Chem. Soc.* **139**, 15160-15167 (2017).
47. 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. USA* **115**, 13318-13323 (2018).

48. Karran, P. & Lindahl, T. Hypoxanthine in deoxyribonucleic acid: generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus. *Biochemistry* **19**, 6005-6011 (1980).

49. Shapiro, R., & Pohl, S. H. Reaction of ribonucleosides with nitrous acid. Side products and kinetics. *Biochemistry* **7**, 448-455 (1968).

50. Mariani, A. D., Russell, A., Javelle, T. & Sutherland, J. D. A light-releasable potentially prebiotic nucleotide activating agent. *J. Am. Chem. Soc.* **140**, 8657−8661 (2018).

Fig. 1. Previous synthesis of RNA pyrimidine nucleosides 1 (C), 2 (U) and a deoxypyrimidine nucleoside 5, and the present work. RAO **10** is a starting point in the network since it crystallises in enantiopure form from minimally enantio-enriched solutions. It can be elaborated via **12** and **3** to the pyrimidine nucleosides. Although we had developed a low-yielding route to deoxyadenosine **7** (dA) from **6** via **5**, we recognized that **12** and **15** are ideal candidates for tethered glycosylation with **16**. The products, thioanhydropurines **18** and **19**, are reduced photochemically in a similar way to **6**, providing an efficient route to deoxynucleosides. Critically, once produced, pyrimidines **1** (C) and **2** (U) survive the sequence that produces purines **7** (dA) and **9** (dI), and we show that the four nucleosides **1** (C), **2** (U), **7** (dA) and **9** (dI) can be produced alongside one another.

Fig. 2. Prebiotic route to purine deoxyribonucleosides, 7 (dA) and 9 (dI). The route starts with α-anhydropyrimidines **12** and **15**, which are intermediates in the

RNA pyrimidine synthesis, and 8-mercaptoadenine **16**, which is available from adenine **8** via hydrolysis and reaction with ammonium thiocyanate or thiourea. Dry state tethered glycosylation of **16** and **12** or **15** provides thioanhydropurines **18** and **19**, which can be photochemically reduced by two routes. If bisulfite is the reductant, only N^9-configured products **7** (dA) and **26** are formed. **26** can be converted to **7** by further irradiation, or by nitrosation. If hydrosulfide is used as the reductant, both N^9-configured **7** (dA) and **26** as well as N^7-configured **20** is formed. **20** has a half-time of hydrolysis nearly two orders of magnitude lower than **7** (dA) and so is selectively degraded. To generate deoxyinosine **9** (dI) alongside deoxyadenosine **7** (dA), the products of either photoreduction are treated with nitrous acid at pH 4.

Fig. 3 Proposed mechanism of photoreduction of N^7-8,2'-anhydro-thioadenosine 18 and N^9-8,2'-anhydro-thioadenosine 19 nucleosides. a) Potential mechanism involving bisulfite proceeding with initial photoexcitation of the thioanhydronucleosides to **28**, followed by reduction of C2', sulfur, and C8. Photoexcitation of the N^7 isomer **19** to **30** leads to decomposition. b) Potential mechanism involving bisulfite proceeding via intial reduction of ground state thioanhydronucleosides, followed by desulfurisation of **26**. Reduction of **19** gives **32** which leads to decomposition. c) Distinct mechanism involving reduction of thioanhydronucleoside–hydrosulfide encounter complexes, **33** and **34**, which both undergo charge transfer and concomitant C–S bond cleavage to produce **31** and **35**. **31** and **35** undergo reduction at C2' and desulfurisation to furnish **7** (dA) and **20**.

Fig. 4. A systems-level approach to a potential primordial genetic alphabet composed of 1 (C), 2 (U), 7 (dA) and 9 (dI). A mixture of the α– and β–epimers of

2-thiocytidine **13** and **3**, which interconvert in UV light, can generate a mixture containing **1** (C), **2** (U), **7** (dA) and **9** (dI). A general route is shown at left. The thiopyrimidines are initially converted into the canonical pyrimidines (cytidine **1** and uridine **2**) and the α-anhydropyrimidines **12** and **15**. The latter undergo tethered glycosylation and then photoreduction to selectively provide purine deoxyribonucleosides **7** (dA) and **9** (dI) as depicted in Fig. 2. The pyrimidines **1** (C) and **2** (U) persist through each step of this sequence, ultimately generating a mixture of all four nucleosides. Specific conditions and yields for two possible particular routes (Routes A and B) are shown at right.

Acknowledgments: The authors thank all JDS group members for fruitful discussions. This research was supported by the Medical Research Council (MC_UP_A024_1009), the Simons Foundation (290362 to JDS, 494188 to RS), and a grant from the National Science Centre Poland (2016/23/B/ST4/01048 to RWG). MJJ acknowledges support within the ‘‘Diamond Grant’’ (0144/DIA/2017/46) from the Polish Ministry of Science and Higher Education and a computational grant from Wrocław Centre of Networking and Supercomputing (WCSS).

Author contributions: Experimental: J. X., V. C., N. J. G., D. A. R., A. D. B.; Theoretical: M. J. J., R. W. G., R. S.; Crystallography: A. D. B.; Supervision: J. D. S.; all authors co-wrote the manuscript.

Competing interests: The authors declare no competing financial interests.

Supplementary Materials is available for this paper.

Correspondence and requests for materials should be addressed to J. D. S.

Reprints and permission information is available at

http://www.nature.com/reprints.

Extended Data Fig. 1. A summary of the main findings of the work. Previously, a prebiotically plausible synthesis of beta-ribopyrimdines C and U has been identified using α–-thiocytidine. Herein, we demonstrate that the same intermediate can undergo a distinct prebiotically plausible process that could have happened in a similar, or the same, environment. The new process furnishes β–*D*-N^9-deoxyribopurine nucleosides, dA and dI, alongside the pyrimidines. Remarkable selectivity enforced by UV irradiation and hydrolysis operates throughout the reported ribosylpyrimidine synthesis and the newly discovered deoxyribosylpurine synthesis, resulting in a set of nucleosides with only the canonical regio- and stereochemistry. The coexistence in one location of a set of nucleosides similar to this is thought by many to be a precondition for the spontaneous emergence of life on Earth.

Extended Data Fig. 2. ^{1}H NMR spectra of conversion of α-anhydrouridine 15 from α-thiouridine 14. a) ^{1}H NMR spectrum of α-anhydrouridine **15**; b) ^{1}H NMR spectrum of the reaction mixture after heating α-thiouridine **14** in H_2O; c) ^{1}H NMR spectrum of the reaction mixture after heating α-thiouridine **14** in formamide.

Extended Data Fig. 3. ^{1}H NMR spectra of photoreduction of N^7-8,2'-anhydro-thioadenosine 18 and N^9-8,2'-anhydro-thioadenosine 19 mixture with bisulfite. a)

^{1}H NMR spectrum of the crude mixture before irradiation (the ratio of N^7 : N^9 isomer was 4 : 5); b) ^{1}H NMR spectrum of the mixture after irradiation for 7 hrs (the N^9 isomers dA **7** and **26** are the only detectable products).

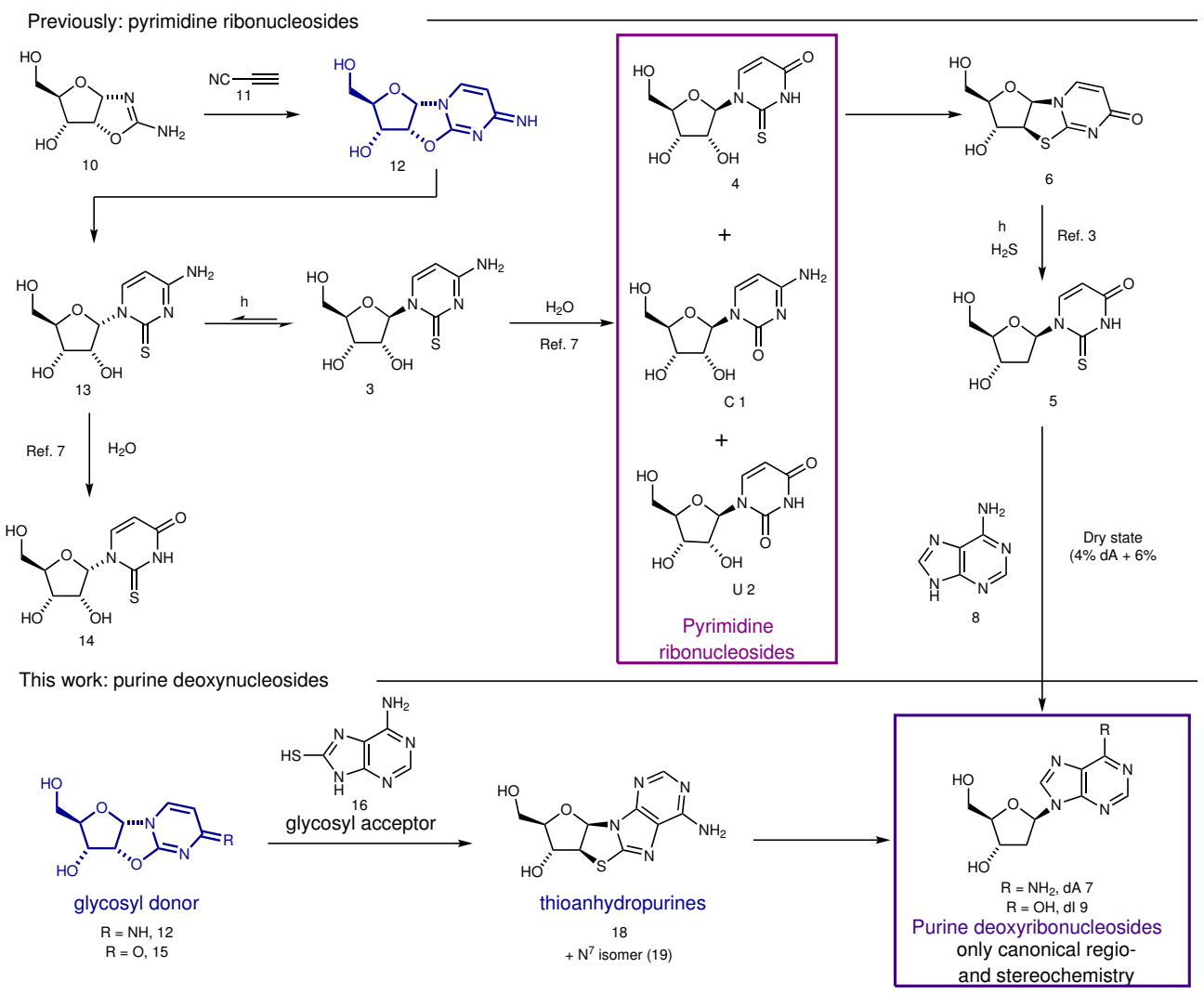
Extended Data Fig. 4. Potential energy surfaces and S_1/S_0 state crossings of the key photochemical steps in deoxyadenosine synthesis calculated at the ADC(2)/ma-def2-TZVP level (see the SI for more details). a) C-S bond opening may spontaneously occur in **18** leading to a peaked S_1/S_0 state crossing, however, a reducing agent is necessary to maintain that geometry after reaching the S_0 state; b) N7-C8 bond rupture is the lowest energy photochemical process in **19** and results in destruction of the purine ring; c) and d) encounter complexes of **18** and **19** with HS^-, which readily undergo photochemical C–S bond rupture induced by charge transfer from HS^- to chromophore.

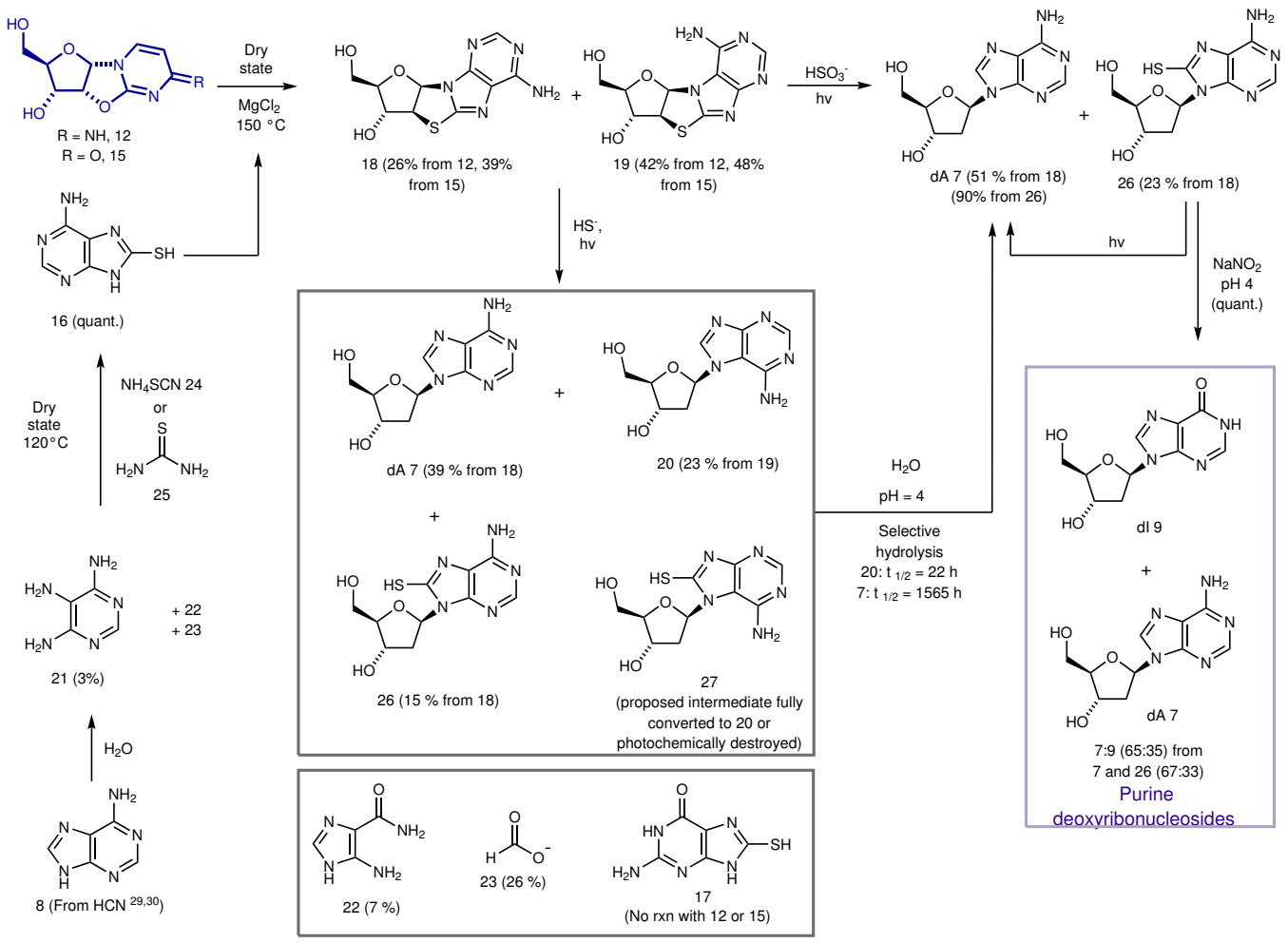
Extended Data Fig. 5. Equilibrium geometries of C2, S8 radical anion 31 and C8, N9 radical anion 32 radical anions which may be formed after accepting a hydrated electron from the environment and the adiabatic electron affinities calculated at the ωB97X-D/IEFPCM/ma-def2-TZVP.

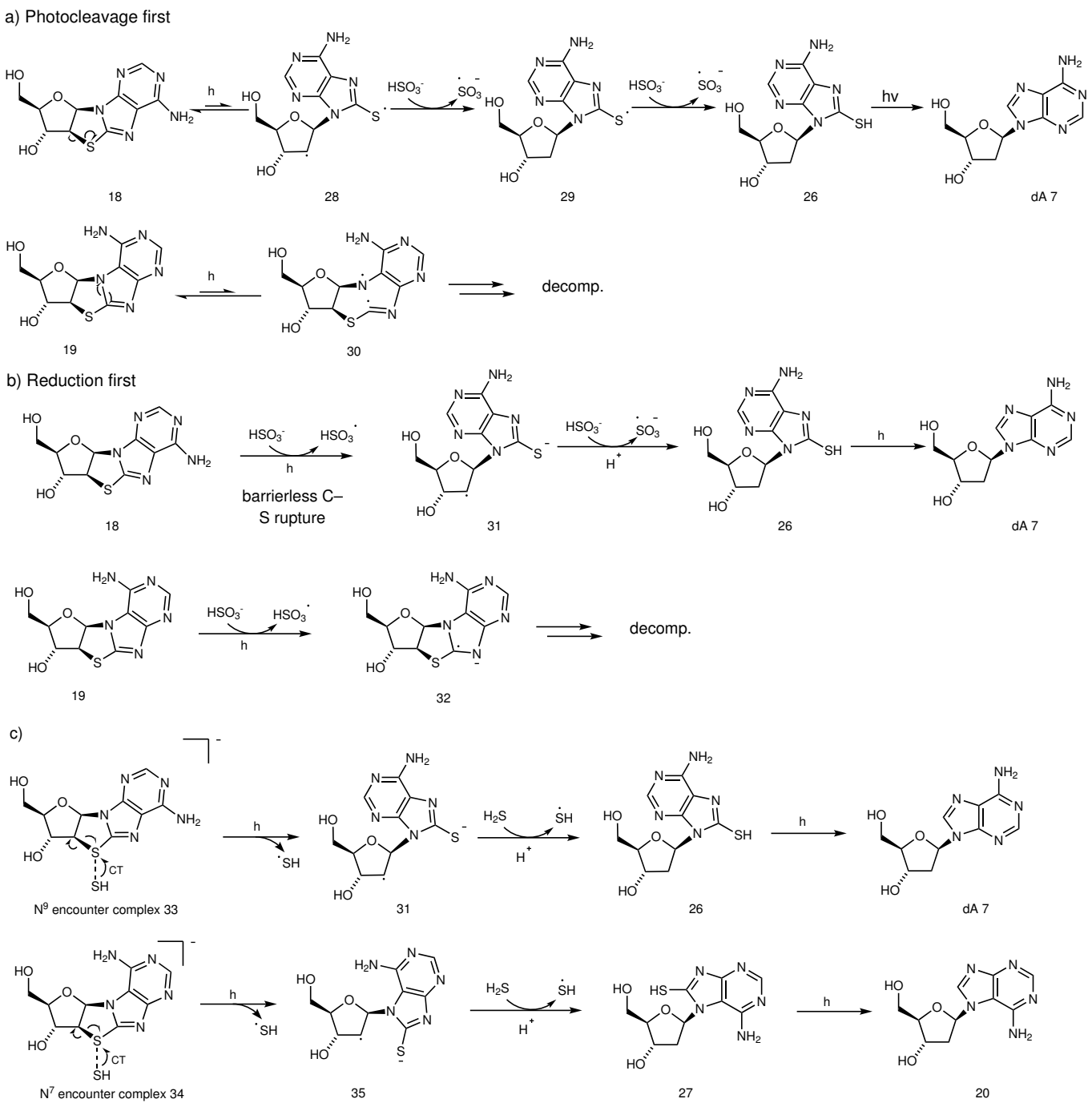
Extended Data Fig. 6. ^{1}H NMR spectra for the reactions of deoxyadenosine 7 and cytidine 1 with nitrous acid. a) ^{1}H NMR spectrum of the mixture of deoxyadenosine **7** and cytidine **1**; b) ^{1}H NMR spectrum of the reaction mixture after 4 days, showing the ratio of all four (deoxy)nucleosides deoxyadenosine **7**, deoxyinosine **9**, cytidine **1**, and uridine **2** is 30:17:42:11.

Extended Data Fig. 7. ^{1}H NMR spectra for stability study of cytidine 1 and uridine 2 at 254 nm irradiation with bisulfite. a) ^{1}H NMR spectrum of the mixture of cytidine **1**, bisulfite and $K_4Fe(CN)_6$ in the dark; b) as a), ^{1}H NMR spectrum after 10 hours of irradiation; c) ^{1}H NMR spectrum of the mixture of uridine **2**, bisulfite and $K_4Fe(CN)_6$ in the dark; d) as c), ^{1}H NMR spectrum after 10 hours of irradiation; e) ^{1}H NMR spectrum of the mixture of cytidine **1**, uridine **2**, N^9-thioanhydroadenosine **18**, bisulfite and $K_4Fe(CN)_6$ in the dark; f) as e), ^{1}H NMR spectrum after 10 hours of irradiation.

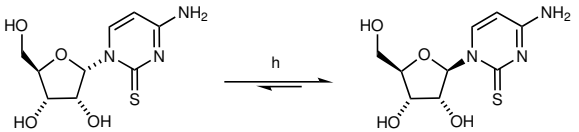
Extended Data Fig. 8. ^{1}H NMR spectra for sequential reactions with the mixture of α-anhydrouridine 15, cytidine 1 and uridine 2. a) ^{1}H NMR spectrum of the mixture after heating with 8-mercaptoadenine **16** and magnesium chloride at 150 °C for 1.5 days; b) ^{1}H NMR spectrum of the same mixture after irradiation with hydrogen sulfide at 254 nm; c) ^{1}H NMR spectrum of the same mixture after reacting with nitrous acid for 2 days (dA **7**:dI **9**:C **1**:U **2**= 14:14:44:28).


Previously: pyrimidine ribonucleosides
10
11
12
13
3
h
14
Ref. 7
H2O
4
C 1
U 2
Pyrimidine ribonucleosides
6
h
H2S
Ref. 3
5
8
Dry state
(4% dA + 6%
This work: purine deoxynucleosides
glycosyl donor
R = NH, 12
R = O, 15
16
glycosyl acceptor
thioanhydropurines
18
+ N7 isomer (19)
R = NH2, dA 7
R = OH, dI 9
Purine deoxyribonucleosides
only canonical regio-
and stereochemistry

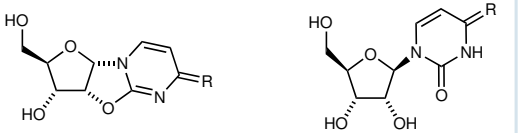

R = NH, 12
R = O, 15
Dry state
MgCl2
150 °C
18 (26% from 12, 39% from 15)
19 (42% from 12, 48% from 15)
HSO3-
hv
dA 7 (51 % from 18)
(90% from 26)
26 (23 % from 18)
16 (quant.)
HS-,
hv
hv
NaNO2
pH 4
(quant.)
dA 7 (39 % from 18)
20 (23 % from 19)
dI 9
+
dA 7
26 (15 % from 18)
27
(proposed intermediate fully converted to 20 or photochemically destroyed)
H2O
pH = 4
Selective hydrolysis
20: t 1/2 = 22 h
7: t 1/2 = 1565 h
7:9 (65:35) from 7 and 26 (67:33)
Purine deoxyribonucleosides
Dry state 120°C
NH4SCN 24
or
25
21 (3%)
+ 22
+ 23
H2O
8 (From HCN 29,30)
22 (7 %)
23 (26 %)
17
(No rxn with 12 or 15)


a) Photocleavage first
h
HSO3-
SO3-
hv
18
28
29
26
dA 7
19
30
decomp.
b) Reduction first
barrierless C–S rupture
31
32
c)
CT
N9 encounter complex 33
N7 encounter complex 34
35
27
20

General Route



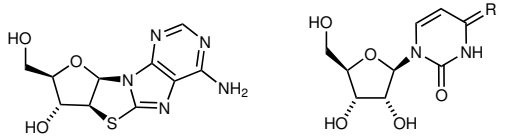
2-thiocytidines



R = NH, 12
R= O , 15

R = NH, C 1
R = O U 2

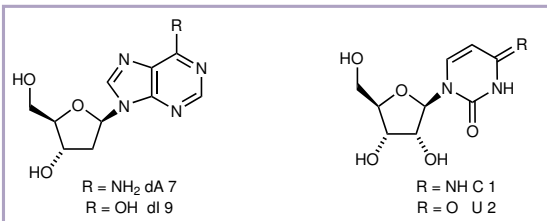
Pyrimidine anhydronucleosides
(glycosyl donors) and nucleosides



N^9-isomer 18
N^7-isomer 19

R = NH C 1
R = O U 2

Thioanhydropurine and pyrimidine nucleosides



R = NH_2 dA 7
R = OH dI 9

R = NH C 1
R = O U 2

Purine deoxyribonucleosides and
pyrimidine ribonucleosides

Conditions and Yields

13 + 3 (1:1 mixture)
(see Fig. 1)

Route A

$NaNO_2$
(pH 4)

From 13: 12 (100%)
From 3: C 1 (54%)

Representative mixture
12 and 1 (1:1)

$MgCl_2$, 16,
150 °C

From 12: 18 (26%)
and 19 (34%)
95% C 1 remaining

HSO_3^-
$Fe(CN)_6^{2-}$
h

From 18: dA 7 and 26
From 19: 20 and 27
C 1 persists
(yields below)

$NaNO_2$, pH 4

Purine deoxyribonucleosides (19%)
dA 7 (10%) and dI 9 (9%)
(yields for 3 steps from 12)

Pyrimidine ribonucleosides
84% persistence after 3 steps

dA:dI:C:U = 10:9:67:14

Route B

H_2O
(pH 5)

From 13: 15 (26%)
From 3: C 1 (4%)
and U 2 (2%)
(87% 3 remains after 24 d)
(Ref 7: t ½ 2.44% in 84 d)

Representative mixture
15, 1 and 2 (2:1:1)

$MgCl_2$, 16,
150 °C

From 15: 18 (30%) and
19 (50%)
80% C 1 remaining
95% U 2 remaining

HS^-
h

From 18: dA 7 and 26
From 19: 20 and 27
C 1 and U 2 persist
(yields below)

Purine deoxyribonucleosides (12%)
dA 7 (6%) and dI 9 (6%)
(yields for 3 steps from 15)

Pyrimidine ribonucleosides
43% persistence after 3 steps

dA:dI:C:U = 14:14:44:28