Multicomponent Assembly of Proposed DNA Precursors in Water

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Multicomponent Assembly of Proposed DNA Precursors in Water

Matthew W. Powner,*[,†,§] Shao-Liang Zheng,[‡] and Jack W. Szostak[§]

[†]Department of Chemistry, University College London, Christopher Ingold Laboratories, 20 Gordon Street, London, WC1H 0AJ, U.K.

[‡]Harvard X-ray Crystallography Centre, 12 Oxford Street, Cambridge, Massachusetts 02138, United States

[§]Howard Hughes Medical Institute and Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, United States

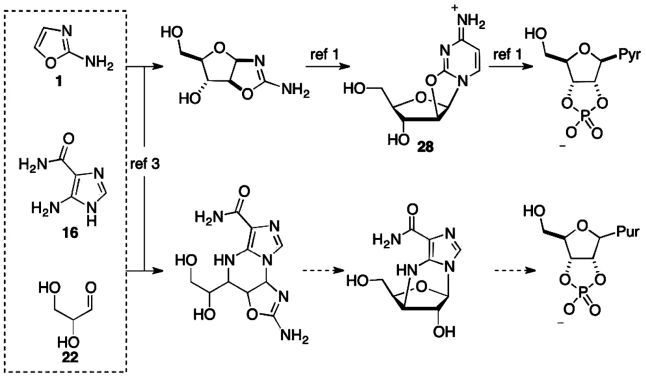
Supporting Information

ABSTRACT: We propose a novel pathway for the prebiotic synthesis of 2′-deoxynucleotides. Consideration of the constitutional chemical relationships between glycolaldehyde and β-mercapto-acetaldehyde, and the corresponding proteinogenic amino acids, serine and cysteine, led us to explore the consequences of the corresponding sulfur substitution for our previously proposed pathways leading to the canonical ribonucleotides. We demonstrate that just as 2-aminooxazole—an important prebiotic ribonucleotide precursor—is readily formed from glycolaldehyde and cyanamide, so is 2-aminothiazole formed from β-mercapto-acetaldehyde and cyanamide in water at neutral pH. Indeed, both the oxazole and the thiazole can be formed together in a one-pot reaction, and can be co-purified by crystallization or sublimation. We then show that 2-aminothiazole can take part in a 3-component carbon—carbon bond-forming reaction in water that leads to the diastereoselective synthesis of masked 2′-thiosugars regiospecifically tethered to purine precursors, which would lead to 2′-deoxynucleotides upon desulfurization. The possibility of an abiotic route to the 2′-deoxynucleotides provides a new perspective on the evolutionary origins of DNA. We also show that 2-aminothiazole is able to sequester, through reversible aminal formation, the important nucleotide precursors glycolaldehyde and glyceraldehyde in a stable, crystalline form.

INTRODUCTION

A plausible abiotic chemical route to the canonical nucleotides is a major goal in origins of life research.[1−3] Recently, we demonstrated the first chemical steps toward a divergent pyrimidine and purine ribonucleotide synthesis.[3] A one-pot multicomponent reaction was demonstrated to stereoselectively tether and consequently regiospecifically glycosylate purine precursors and masked pentose sugars, while concurrently furnishing known pyrimidine precursors (Scheme 1).

Scheme 1. Proposed Multicomponent Ribonucleotide Syntheses



Although RNA has often been considered as a candidate for the first biopolymer of life,[1,2] extant biology utilizes two chemically distinct, but related, nucleotide polymers, RNA and DNA. DNA is usually viewed as a late evolutionary adaptation of earlier RNA-based life.[4] However, it would not be possible to make DNA without deoxyribonucleotides, and yet in the absence of DNA there is no obvious reason for the evolution of the biochemical pathways for the synthesis of deoxyribonucleotides. Only after the emergence of DNA as an important cellular component would there have been a strong selective pressure favoring the emergence of biochemical pathways for the synthesis of deoxyribonucleotides from ribonucleotides. Although the advantages of DNA as a medium for information storage are clear, there has not until now been any reasonable hypothesis for how DNA could be 'invented' by primitive cells in the absence of pre-existing deoxyribonucleotides. In this paper, we propose a hypothesis for the prebiotic synthesis of 2′-deoxyribonucleotides.

Biochemically, 2′-deoxyribonucleotides are synthesized by chemo- and regio-specific reduction of ribonucleotides.[5] In the absence of the complex and energetically costly enzymatic resources required to regiospecifically deoxygenate ribonucleotides, abiological 2′-deoxygenation of nucleotides requires site-

Received: June 28, 2012
Published: July 27, 2012

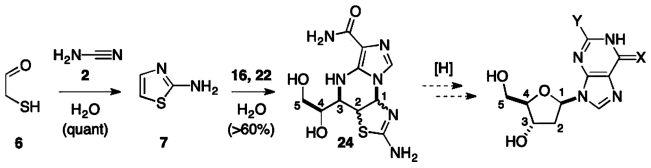
   dx.doi.org/10.1021/ja306176n | J. Am. Chem. Soc. 2012, 134, 13889−13895

specific *ribo*- or *arabino*-nucleoside deoxygenation. In light of the difficulty of regioselective nucleotide deoxygenation in the absence of sophisticated enzymatic control, it would seem that regiospecific positioning of a group/atom with the appropriate latent potential reactivity could form the basis of an alternative pathway to the 2′-deoxyribonucleotides. In principle, a sufficiently chemoselective reactivity difference could be exploited to generate both ribonucleotides and 2′-deoxyribonucleotides, simultaneously.

RESULTS AND DISCUSSION

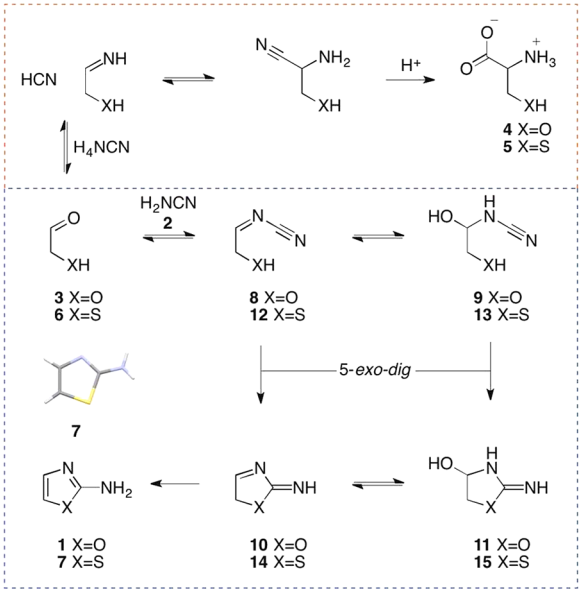
Constitutional analysis of DNA, with respect to RNA, in light of the low dissociation energy of C−S bonds,[6] suggests that regiospecific positioning of sulfur at C2′ could result in the chemical differentiation required for the divergent synthesis of DNA and RNA monomers from common precursors (Scheme 2).

Scheme 2. Proposed Multicomponent Deoxyribonucleotide Syntheses



Oxazole **1**—a key ribonucleotide precursor—is derived from cyanamide **2** and glycolaldehyde **3**.[1,3,7] The C2-carbon atom of aldehyde **3** is formally delivered to a ribonucleotide as the C2′ sugar carbon. It is of note that aldehyde **3** is also the aldehyde precursor of serine **4**, a proteinogenic amino acid, via Strecker synthesis with ammonium cyanide (Scheme 3).[8] The constitutional similarity of serine **4** and cysteine **5**, and their aldehyde precursors **3** and *β*-mercapto-acetaldehyde **6**, suggests that both **3** and **6** must be considered within our exploration of prebiotic azole synthesis.[9] Specifically, we propose that the divergent reactivity of glycolaldehyde **3** and *β*-mercapto-acetaldehyde **6** with cyanamide **2** could lead, via 2-aminooxazole **1** and 2-aminothiazole **7**, to ribonucleotides[1,3] and 2′-deoxyribonucleotides, by desulfurization of C2′ (Scheme 2).[10,11]

Scheme 3. Constitutional Analysis: Strecker-Type Synthesis of Amino Acids (Red Box) and Azole Synthesis in Water (Blue Box)



The importance of 2-aminothiazole **7**, and its derivatives, has been recognized in medicinal chemistry,[12] but though the synthesis of thiazole **7** in water has been reported,[13] the potential relevance of thiazole **7** to the origins of life remains unexplored. The formation of oxazole **1** from cyanamide **2** and glycolaldehyde **3** at neutral pH requires stoichiometric (or excess) cyanamide **2** and general acid−base catalysis.[1] Aldehyde **6**, upon imine formation with cyanamide **2**, has, like **3**, the potential to undergo 5-*exo*-*dig* cyclization to furnish imine **14** (or hemiaminal **15**) as shown in Scheme 3. Furthermore, we expected the reaction to proceed more rapidly at neutral pH and be less prone to stall at intermediates en route to thiazole **7** due to the increased nucleophilicity of sulfur and greater aromaticity of thiazoles with respect to oxazoles.

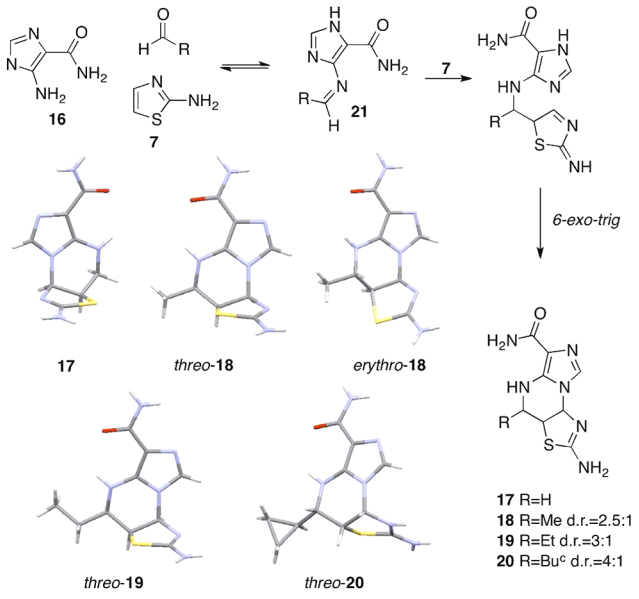
To test this supposition, *β*-mercapto-acetaldehyde **6** (8.5 mM) and **2** (15, 26, or 50 mM) were incubated in D_2O at pD 7 and 20 °C and a rapid (ca. 10 min) quantitative conversion of aldehyde **6** to hemiaminal **15** was observed (see the Supporting Information Figure S1). Over the course of several hours, **15** dehydrated to furnish thiazole **7** (see Supporting Information Figures S2 and S3).[14] It is of note that excess aldehyde **6** did not lead to further reaction of thiazole **7**. Moreover, at elevated concentration (>150 mM) 2-aminothiazole **7** was observed to directly crystallize from water, and a crystal structure of thiazole **7** is shown in Scheme 3.[15,16] Next we investigated the concomitant synthesis of oxazole **1** and thiazole **7**. In the presence of excess cyanamide **2** (2.2−5 equiv), rapid and stoichiometric conversion of glycolaldehyde **3** (1 equiv) and *β*-mercapto-acetaldehyde **6** (1 equiv) to hemiaminals **11** and **15**, respectively, was observed in water. At high pH/pD (pH/pD > 9) specific base catalysis or in 1 M phosphate solution (pH/pD 6.5−7.5) general acid−base catalysis furnish oxazole **1** and thiazole **7** over the course of 24−36 h at ambient temperature (Scheme 3; see Supporting Information Figure S8−S10).[13] At neutral pD, selective dehydration of **15** to liberate thiazole **7** was observed. Limiting cyanamide **2** (**2**/**3**/**6** ∼ 0.8:1:1) also resulted in the specific generation of thiazole **7**, even in 1 M phosphate solution (pH 7; See Supporting Information Figure S9). It is likely that the rapid 5-*exo*-*dig* cyclization of imine **12**/hemiaminal **13** leads to the selective sequestration of cyanamide **2** in thiazole **7**.

We did not observe significant homo- and hetero-aldol reaction during the reaction of glycolaldehyde **3** and mercaptoacetaldehyde **6** with either excess or limiting cyanamide **2** between ambient temperature and 60 °C in 1 M phosphate solution at pH 7. Interestingly, we also found that, like oxazole **1**, thiazole **7** is easily sublimed (see Supporting Information Figure S4), suggesting possible geochemical routes to the distribution, separation, and purification of thiazole **7** under abiotically plausible conditions. Sublimation from a mixture of oxazole **1** and thiazole **7** resulted in the isolation of both azoles as colorless crystalline solids. Partial sublimation of a mixture of oxazole **1** and thiazole **7** (1:1) resulted in the isolation 4:1−2:1 oxazole **1**/thiazole **7**, at 50 °C between 1 and 24 h at ambient pressure, indicating the more rapid but comparable sublimation rate of oxazole **1** relative to thiazole **7**. Complete sublimation was observed after 2 d at 60 °C to afford ∼1:1 mixture of crystalline azoles at ambient pressure.

dx.doi.org/10.1021/ja306176n | J. Am. Chem. Soc. 2012, 134, 13889−13895

We then examined the multicomponent reaction of thiazole **7**, 4-amino-imidazole 5-carboxamide **16**,[17] and a series of aldehydes. Initially model aliphatic aldehydes were investigated to avoid issues of annulation control. Equimolar quantities of thiazole **7** and aliphatic aldehydes (including formaldehyde, acetaldehyde, propionaldehyde, and cyclobutaldehyde) with excess amino-imidazole **16** (1.2−3 equiv) furnished high yields (75−90%) of thiazolines **17**, **18**, **19**, and **20**, respectively, at pH 5. The multicomponent reactions were found to be robust at multi-gram scale and comparable diastereoselectivity was observed in the reaction of aliphatic aldehydes with oxazole **1** or thiazole **7** and amino-imidazole **16** (via imine **21** R = alkyl; d.r.= 2.5:1−4:1 *threo*/*erythro*).[3] Compounds **17**, *threo*-**18** and *erythro*-**18**, and the major isomers of **19** and **20** were isolated by fractional crystallization, and the crystal structures are shown in Scheme 4.[19]

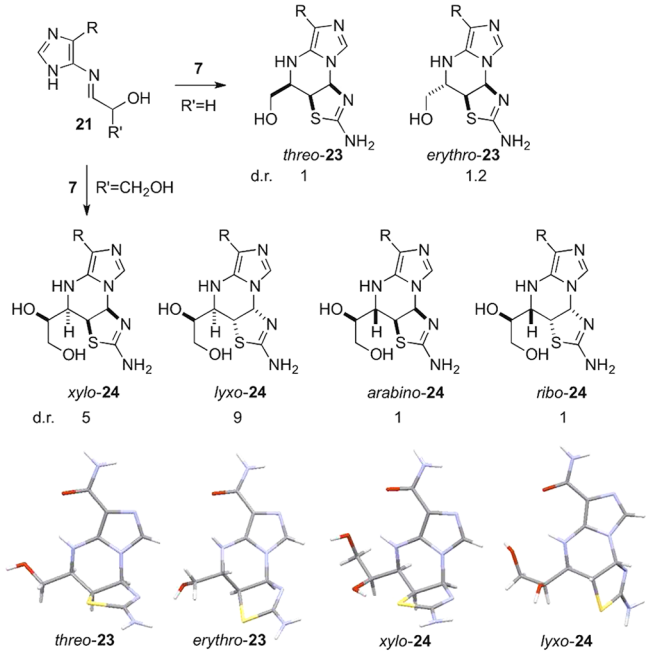
Scheme 4. Three-Component Reaction of 2-Aminothiazole 7, 5-Amino-imidazole-4-carboxamide 16, and Various Aliphatic Aldehydes



To continue our investigation into the potential of multicomponent assembly of nucleotides we studied the participation of thiazole **7** in carbon−carbon bond-forming reaction with α-hydroxyaldehydes and imines derived from α-hydroxyaldehydes in water.[20] Incubation of equimolar aldehyde **3**, 2-aminothiazole **7**, and amino-imidazole **16** for 5 d in water at pH 4−6 at 20 °C gave a high yield of C_4-thiazoline **23** (80% isolated yield) as 1.2:1 *erythro*/*threo* mixture of diastereomeric products. Surprisingly, this signified that the facial selectivity of thiazole **7** is significantly relaxed with respect to oxazole **1** and imine **21** (R = CH_2OH). However, both *erythro*-**23** and *threo*-**23** were formed with complete annulation control (>99%, see ^{1}H NMR spectrum shown in Supporting Information Figure S28; single crystal X-ray structures of *erythro*-**23** and *threo*-**23** are shown in Scheme 5).[22]

Upon investigation of the parallel MCR with the C_3-synthon glyceraldehyde **22**, comparable efficiency in the synthesis of C_5-thiazoline **24** (70% isolated yield) was observed. Again, exclusive *6-exo-trig* annulation is observed; however, in this case, **24** is furnished with significant diastereoselectivity (d.r.= 9:5:1.1:1, see Scheme 5 and Supporting Information Figure

Scheme 5. Three-Component Reaction of 2-Aminothiazole 7, 4-Aminoimidazole-5-carboxamide 16, and Glyceraldehyde 22



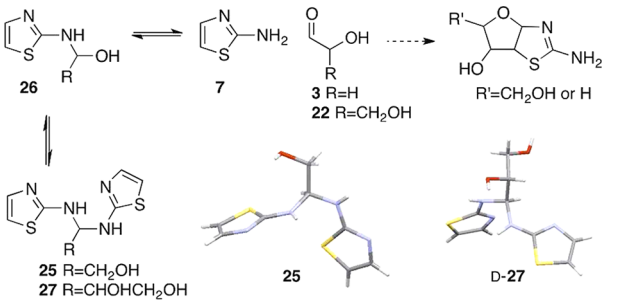
S35). Isolation of the two major isomers, crystallization, and X-ray diffraction proved their *xylo*- and *lyxo*-configuration.[22] The diastereoselectivity obtained, upon carbon−carbon bond formation is proposed to result from intramolecular hydrogen-bonding and concomitant steric impedance of thiazole addition to the *Si*-face of imine **21** (see Supporting Information Scheme S1).[3,23] Although the facial selectivity of the nucleophile **7** has been relaxed with respect to the comparable reaction of **1**, the facial selectivity of the glyceraldehyde-imine **21** (R = CH_2OH) is retained.[3] This is important with respect to purine nucleotide synthesis by 3′-purine tethering, which necessarily requires N3′-displacement and C3′-inversion. Furthermore, as well as regiospecific tethering of **16** and consequent delivery of N1 to the anomeric carbon atom, reaction of thiazole **7** with imine **21** regiospecifically delivers sulfur to C2′, which is of significant interest with regards to the potential synthesis of 2′-deoxynucleotides by C2′-desulfurization.

The multicomponent tricyclic thiazoline products **17**−**24** were uncontaminated with homoaldolization and homo-Mannich byproducts. Additionally, oxazole **1** is known to react efficiently with α-hydroxyaldehydes in a pH-dependent bimolecular domino-reaction.[3] In striking contrast, no significant carbon−carbon bond formation was observed by ^{1}H NMR analysis upon incubation of thiazole **7** with either glycolaldehyde **3** or *rac*-glyceraldehdye **22** in D_2O (pD 3−9, 4−100 °C). Instead, over a period of 4 days at pD 7, crystal formation was observed from an aqueous solution containing **3** (100 mM) and thiazole **7** (100 mM) at 20 °C.[24] Single crystal X-ray diffraction of a crystal isolated directly from the reaction supernatant proved that the structure was aminal **25** (Scheme 6).[25]

Though NMR studies confirmed that *rac*-**22** formed hemiaminal **26** and aminal **27** with thiazole **7** in D_2O, no aminal crystallization was observed in solutions of *rac*-**22** and **7**, even at significantly elevated concentrations (500 mM to 2M) or with substoichiometric quantities of *rac*-**22**. To test whether

dx.doi.org/10.1021/ja306176n | J. Am. Chem. Soc. 2012, 134, 13889−13895

Scheme 6. Crystallization of Bis-(2-aminothiazole)-aminals of Glycolaldehyde 3 and D-Glyceraldehyde 22 from Water at pH 7



the change in solubility properties was due to the increased number of hydroxyl moieties or perhaps due to the differential solubility and packing properties of *rac*-**27** with respect to the achiral aminal **25**, we investigated the solubility properties of the aminal structures formed in a mixture of D-glyceraldehyde **22** and thiazole **7**. Intriguingly, we found that at 100 mM concentration, D-**27** now crystallized from water at 20 °C. NMR analysis showed a 1:2 ratio of **22**/**7** and recrystallization (10:1 H_2O/EtOH) allowed for crystallographic confirmation of the aminal structure (Scheme 6).[25]

SUMMARY

In summary, our results suggest that in a sulfur-rich prebiotic environment, oxazole **1** and thiazole **7** could be formed together under the same conditions, as long as both glycolaldehyde **3** and its thiol analogue, *β*-mercapto-acetaldehyde **6**, were present along with cyanamide **2**. In such mixed reactions, 2-aminothiazole **7** forms first, followed by 2-aminooxazole **1**. The absence of any interference between oxazole and thiazole synthesis shows that these potential precursors of deoxynucleotides and ribonucleotides could indeed be formed at the same time in the same conditions. We have also found that two abiotically plausible methods for the purification of oxazole **1**, namely, crystallization and sublimation, also apply to thiazole **7**. However, the chemical reactivities of oxazole **1** and thiazole **7** are quite different, suggesting that the subsequent pathways would diverge.

The most striking difference between the reactivity of oxazole **1** and thiazole **7** is the failure of the thiazole to react with aldehydes to form the thiazoline precursors of the (thio)-pyrimidine nucleotides. The most likely explanation for this difference is the increased aromaticity of the thiazole relative to the oxazole, and thus decreased nucleophilicity of C5. We are currently investigating an alternative synthesis of 2′-deoxyribo-pyrimidine nucleotides mediated by C2′-inversion of anhydronucleotide **28** (Scheme 1).[1a,11] However, during the course of efforts to observe two-component reactivity of thiazole **7** with aldehydes, we observed instead the crystallization of glycolaldehyde **3** and homochiral (but not racemic) glyceraldehyde **22** (two known ribonucleotide precursors)[1,3] as aminals directly from water. The facile reversibility of aminal formation suggests that the chemical sequestration and protection of C_2- and C_3-aldehyde nucleotide synthons by another important class of nucleotide synthon (azoles) may have provided a route to the concentration, purification, and stabilization of the necessary aldehyde precursors of nucleotides. Although these aldehydes can in principle be readily synthesized by a variety of routes,[26,27] their high reactivity has until now made it difficult to conceive of means for accumulating large reservoirs of these essential starting materials in a prebiotic context.

By exploring the multicomponent reactivity of thiazole **7**, together with aldehydes and the purine precursor amino-imidazole **16**, we have demonstrated a selective and high yielding 3-component carbon−carbon bond-forming reaction that chemospecifically furnishes masked-2′-thiosugars regiospecifically tethered to an amino-imidazole purine-precursor. The facial selectivity of thiazole **7**, upon addition to *α*-hydroxyaldehyde-imines, is quite relaxed (relative to oxazole **1**), resulting in low diastereoselectivity at C2′; however, this is inconsequential with regards to the proposed deoxynucleotide synthesis because C2′ would be rendered achiral following desulfurization. In contrast, we observed very high diastereoselectivity of nucleophilic addition of thiazole **7** to imine **21**, generating almost exclusively the *lyxo* and *xylo* isomers with controlled N9−C1′ annulation. The conversion of the resulting C_5-thiazolines to 2′-deoxyribo-purine nucleotides requires three additional steps: C3′-inversion to release the purine-precursor from its tethered attachment to C3′, completion of the purine heterocycle, and chemospecific desulfurization of C2′.[28] These steps are currently under investigation in our laboratories as part of our continuing studies aimed at the abiotic synthesis of 2′-deoxyribonucleotides.[29]

Finally, it is of note that if life emerged in an environment containing both ribo- and 2′-deoxyribonucleotides, it is likely that the first biopolymers were heterogeneous in composition, that is, composed of a mixture of ribo- and deoxyribonucleotides, and perhaps other compatible nucleotides. Recent work has shown that functional biopolymers, in the form of aptamers with highly specific molecular recognition properties, can be derived by in vitro evolution from libraries of polynucleotides composed of randomly interspersed ribo- and deoxyribonucleotides.[29] Thus, chimeric RNA/DNA polymers may have been sufficient for the emergence of life. The primordial biochemical exploitation of a mixed RNA and DNA genetic system could eliminate the requirement for a genetic takeover (of RNA by DNA), and would arguably result in a simplification of the transition from chemistry to biology. However, such mixed-composition polymers have neither the advantageous stability of DNA, nor the optimal functional characteristics of RNA. Thus, RNA and DNA may have emerged as early and contemporaneous specializations of a primitive mixed biopolymer.

EXPERIMENTAL SECTION

General Methods. Reagents and solvents were purchased from Sigma-Aldrich, TCI America, Frontier Scientific, or Cambridge Isotope Laboratories. Flash column chromatography was carried out on Merck 9385 silica gel 60 (230−400 mesh). NMR spectroscopy was carried out on a Varian NMR spectrometer (Oxford AS-400) or a Bruker NMR spectrometer (AvanceIII 600) operating at 25 °C probe temperature (unless otherwise specified). When possible, the chemical shift of the corresponding solvent was used as a reference. Chemical shift values are reported in parts per million (ppm) and *J*-couplings are recorded in hertz (Hz). Electrospray mass spectrometry was recorded on a Bruker Daltonics Esquire 6000 ESI-MS. High resolution mass spectrometry was carried out on a Waters Q-ToF micro LC/MS/MS system. Infrared spectra were recorded as manually pressed KBr discs on a PerkinElmer spectrum 100 series FT-IR spectrometer. Single crystal X-ray crystallography was carried out with a Bruker APEX II CCD diffractometer (Mo K*α* radiation, λ = 0.71073 Å), equipped with an Oxford Cryosystems nitrogen flow apparatus, at 100 K. Data

integration down to 0.76 Å resolution was carried out using SAINT V7.46 A (Bruker diffractometer, 2009) with reflection spot size optimization. Absorption corrections were made with the program SADABS (Bruker diffractometer, 2009). The structure was solved by the direct methods procedure and refined by least-squares methods against F^2 by using SHELXS-97 and SHELXL-97 (Sheldrick, 2008). Non-hydrogen atoms were refined anisotropically.

2-Aminothiazole (7). Mercaptoacetaldehyde **6** and cyanamide **2** (0.5−1.5 equiv) were dissolved in H_2O, D_2O or phosphate buffer (1 mL) at pH/pD 7.0. Reaction progress was monitored by the ^{1}H NMR spectroscopy (see the Supporting Information, Figure S1). An initial concentration of 100 mM led to direct crystallization of **7** (60% isolated yield of crystals). Alternatively, upon complete conversion (by NMR spectroscopy), the solution was concentrated and **7** recovered by sublimation. Compound **7** (1−5 g) was placed in a covered beaker or Erlenmeyer flask and warmed from beneath to 40 °C. Sublimate was collected on flask/beaker sides and cover at ambient temperature (see Figure S4) or on a water-cooled coldfinger (5 °C). Crystalline **7** was isolated. IR (solid, cm^{-1}): 3407, 3285, 3120, 3086, 1622, 1518, 1488. ^{1}H NMR (400 MHz; DMSO): δ_H 6.90 (d, J = 3.7, 1H, H−(C4)), 6.84 (s, 2H, NH_2), 6.51 (d, J = 3.7, 1H, H−(C5)). ^{13}C NMR (101 MHz; DMSO): δ_C 169.5, 139.3, 107.1. MS ESI (pos.) 101 (100%, $[M + H]^+$). Cambridge Crystallographic Data Centre deposition number 864225.

One-Pot Synthesis of 2-Aminothiazole (7) and 2-Amino-oxazole (1). Glycolaldehyde **3** (1 equiv), mercaptoacetaldehyde **6** (1 equiv), and cyanamide **2** (0.5−5 equiv) were dissolved in H_2O/D_2O (1 mL) at pH/pD 7, 10 or 12, or 1 M phosphate buffer (1 mL) at pH/pD 6.5−7.0. Reaction progress was monitored by the ^{1}H NMR spectroscopy (see the Supporting Information, Figure S8−10). An initial concentration of 200 mM of **2**, **3** and **6** led to direct crystallization of **15** (5−10% isolated yield of crystals). Alternatively, upon complete conversion (by NMR spectroscopy), the solution was concentrated and **7** and **1** were recovered by sublimation. Compounds **7** and **1** (1:1; 100 mg) as fine powder or film (evaporated water) were deposited in a covered glass tube (10 cm × 1 cm). A temperature gradient (40, 50, or 60 °C to ambient temperature) was set up across the tube and a mixture of crystalline thiazole **7** and oxazole **1** were isolated at ambient temperature as a deposit upon the glass wall of the tube.

General Procedure for Isolation of Bis-thiazole Aminal Crystals. Aldehyde (0.49−1.0 equiv) was added to a solution of **7** in H_2O at pH 7 ± 0.2. The solution was then incubated at room temperature for 5−10 d; the resultant solid was then isolated by filtration or trituration and washed twice with ethanol and once with ice cold water. D-Glyceraldehyde **22**/thiazole **7** crystals were then recrystallized from aqueous ethanol (10:1 H_2O/EtOH).

*Glycolaldehyde Bis-thiazole Aminal **17***. IR (solid, cm^{-1}): 3213 (br), 2966, 2934, 2869, 1589, 1539, 1497. Cambridge Crystallographic Data Centre deposition number 864235.

*Glyceraldehyde Bis-thiazole Aminal **19***. IR (solid, cm^{-1}): 3326 (br), 3244, 3113, 2948, 2933, 2901, 2882, 1536, 1489. Cambridge Crystallographic Data Centre deposition number 864236.

General Procedure for Multicomponent Reaction. Aldehyde (50−250 mM, 1 equiv) was added to a solution of **7** (1.1 equiv) and 5-aminoimidazole-4-carboxamide **16** (1.5−3.0 equiv) in H_2O at pH 5.0 ± 0.2. The reaction was stirred under an argon atmosphere until no residual aldehyde was detectable following lyophilization and analysis by ^{1}H NMR. The reaction was lyophilized and purified by silica gel column chromatography (CH_2Cl_2/MeOH, 10:0 to 6:4). The major product(s) were concentrated and recrystallized from ethanol/water or dichloromethane/methanol mixtures.

17: Yield 90% R_f = 0.25 (CH_2Cl_2/MeOH 7:3). IR (solid, cm^{-1}): 3333, 3179, 2977, 2922, 2860, 1636, 1607, 1580, 1558. ^{1}H NMR (400 MHz; D_2O): δ_H 7.22 (s, 1H, Ar), 5.85 (d, J = 6.0, 1H, H−(C1′)), 4.18 (ddd, J = 7.8, 6.0, 4.1, 1H, H−(C2′)), 3.39 (ABX, J = 12.8, 4.1, 1H, H−(C3′)), 3.20 (ABX, J = 7.8, 1H, H−(C3′)). ^{1}H NMR (400 MHz; DMSO): δ_H 7.26 (d, J = 1.1, 1H, Ar), 5.91 (d, J = 6.1, 1H, H−(C1′)), 4.29−4.25 (ddd, J = 7.0, 6.1, 3.8, 1H, H−(C2′)), 3.41 (ABX, J = 13.0, 3.8, 1H, H−(C3′)), 3.18 (ABX, J = 13.0, 7.0, 1H, H−(C3′)). ^{13}C NMR (101 MHz; DMSO): δ_C 166.6, 161.4, 141.6, 129.4, 111.4, 79.5, 48.0, 42.2. HRMS ESI [M + Na$^+$] calc. for $C_8H_{10}N_6OSNa$ 261.0535; obs. 261.0526. Cambridge Crystallographic Data Centre deposition number 864226.

18: Yield >85% d.r.= 2.5:1 *erythro*-**18** (minor): R_f = 0.25 (CH_2Cl_2/MeOH 7:3). IR (solid, cm^{-1}): 3437, 3387, 3292, 3123, 2975, 2856, 1638, 1613, 1571. ^{1}H NMR (400 MHz; DMSO): δ_H 7.05 (s, 1H, Ar), 6.97 (s, 2H, NH_2), 6.68 (br d, J = 19.9, 2H, NH_2), 5.95 (d, J = 6.4, 1H, H−(C1′)), 5.89 (s, 1H, HN), 4.54 (ddd, J = 6.4, 2.5, 1.7, 1H, H−(C2′)), 3.78 (qd, J = 6.4, 2.6, 1H, H−(C3′)), 1.25 (d, J = 6.4, 3H, H−(C4′)). ^{13}C NMR (101 MHz; DMSO): δ_c 166.7, 165.2, 141.4, 128.4, 111.6, 79.4, 58.6, 46.7, 22.7. HRMS ESI [M + Na$^+$] calc. for $C_9H_{12}N_6OSNa$ 275.0691; obs. 275.0685. Cambridge Crystallographic Data Centre deposition number 864227.

threo-**18** (major): R_f = 0.25 (CH_2Cl_2/MeOH 7:3). IR (solid, cm^{-1}): 3391, 3302, 3103, 2972, 2865, 1651, 1621, 1582. ^{1}H−NMR (400 MHz; DMSO): δ_H 7.33 (s, 1H), 6.96 (s, 1H), 6.76 (d, J = 27.2, 1H), 6.20 (s, 1H), 6.00 (d, J = 6.1, 1H), 3.85 (dd, J = 9.5, 6.3, 1H), 3.24−3.17 (m, 1H), 1.31 (d, J = 6.2, 1H). ^{13}C NMR (101 MHz; DMSO): δ_c 166.7, 159.2, 141.3, 129.7, 111.3, 79.1, 54.6, 48.9, 20.4. HRMS ESI [M + Na$^+$] calc. for $C_9H_{12}N_6OSNa$ 275.0691; obs. 275.0686. Cambridge Crystallographic Data Centre deposition number 864228.

19: Yield >85% d.r.= 3:1 R_f = 0.25 (CH_2Cl_2/MeOH 7:3). *threo*-**19** (major): IR (solid, cm^{-1}): 3432, 3371, 3314, 3105, 2966, 2930, 1645, 1604, 1575. ^{1}H NMR (400 MHz; DMSO): δ_H 7.04 (s, 1H, Ar), 6.97 (s, 2H, NH_2), 6.67 (br.s, 2H, NH_2), 5.95 (d, J = 6.4, 1H, H−(C1′)), 5.85 (s, 1H, NH), 4.60 (d, J = 6.3, 1H, H−(C2′)), 3.60 (t, J = 6.9, 1H, H−(C3′)), 1.74−1.63 (m, 1H, H−(C4′H)), 1.46−1.35 (m, 1H, H−(C4′H)), 0.94 (t, J = 7.3, 3H, Me). ^{13}C NMR (101 MHz; DMSO): δ_c 166.8, 165.3, 141.3, 128.3, 111.6, 79.6, 56.5, 52.5, 29.8, 10.6. HRMS ESI [M + Na$^+$] calc. for $C_{10}H_{14}N_6OSNa$ 289.0848; obs. 289.0851. Cambridge Crystallographic Data Centre deposition number 864229.

20: Yield >85% d.r.= 4:1 R_f = 0.2 (CH_2Cl_2/MeOH 7:3). *threo*-**20** (major): IR (solid, cm^{-1}): 3404, 3109, 2923, 2525, 2293, 1605, 1581, 1548, 1512. ^{1}H NMR (400 MHz; DMSO): δ_H 6.70 (s, 1H, Ar), 5.58 (d, J = 6.4, 1H, H−(C1′)), 4.28 (dd, J = 6.3, 2.1, 1H, NH)), 2.57 (dd, J = 9.0, 2.2, 1H, H−(C2′)), 0.39 (qt, J = 8.1, 4.2, 1H, H−(C3′)), 0.18 (quintet, J = 8.7, 2H, H_2−(C4′)), 0.00 (d, J = 3.9, 2H, H_2−(C4″)). ^{13}C NMR (101 MHz; DMSO): δ_c 164.4, 163.1, 138.8, 126.1, 109.2, 77.2, 54.9, 53.6, 15.1, 2.0, 0.0 HRMS ESI [M+Na$^+$] calc. for $C_{11}H_{14}N_6OSNa$ 301.0846; obs. 301.0842. Cambridge Crystallographic Data Centre deposition number 864230.

23: Yield >75% d.r.= 1:1. *erythro*-**23** (major): R_f = 0.25 (CH_2Cl_2/MeOH 7:3). IR (solid, cm^{-1}): 3363, 3213, 3192, 2920, 2844, 2375, 1626, 1592, 1547, 1519. ^{1}H NMR (400 MHz; D_2O): δ_H 7.23 (s, 1H, Ar), 5.81 (d, J = 7.5, 1H, H−(C1′)), 4.99 (dd, J = 7.5, 2.5, 1H, H−(C2′)), 3.75 (ddd, J = 2.5, 5.9, 5.2 1H, H−(C3′)), 3.53 (ABX, J = 12.5, 5.2, 1H, H−(C4′)), 3.49 (ABX, J = 12.1, 5.9, 1H, H−(C4′)). ^{1}H NMR (400 MHz; DMSO-d_6): δ_H 7.05 (s, 1H, Ar), 6.97 (s, 2H, $CONH_2$), 6.66 (bs, 2H, NH_2), 6.05 (s, 1H, NH), 6.02 (d, J = 6.3, 1H, H−(C1′)), 5.29 (t, J = 0.5, 1H, OH), 4.48 (d, J = 5.9, 1H, H−(C2′)), 3.78 (m, 1H, H−(C3′)), 3.60 (m, 1H, H−(C4′)), 3.27 (m, 1H, H−(C4′) partially obscured by DMSO residual solvent signal). ^{13}C NMR (101 MHz; DMSO): δ_C 166.8, 165.9, 141.2, 128.3, 111.8, 80.2, 65.3, 53.6, 53.0. HRMS ESI [M + Na$^+$] calc. for $C_9H_{12}N_6O_2SNa$ 291.0640; obs. 291.0640. Cambridge Crystallographic Data Centre deposition number 864231.

threo-**23** (minor): R_f = 0.20 (CH_2Cl_2/MeOH 7:3). IR (solid, cm^{-1}): 3330, 3212, 3111, 2950, 2926, 2853, 2394, 1631, 1568. ^{1}H NMR (400 MHz; D_2O): δ_H 7.35 (s, 1H, Ar), 5.98 (d, J = 5.9, 1H, H−(C1′)), 4.09 (dd, J = 7.9, 6.5, 1H, H−(C2′)), 3.80 (ABX, J = 12.0, 2.4, 1H, H−(C4′)), 3.62 (ABX, J = 11.9, 5.7, 1H, H−(C4′)), 3.45−3.41 (m, 1H, H−(C3′)). ^{1}H NMR (400 MHz; DMSO): δ_H 7.31 (s, 1H, Ar), 6.99 (s, 2H, NH_2), 6.76 (bd, J = 10.1, 2H, C(O)NH_2), 6.35 (s, 1H, NH), 6.02 (d, J = 6.2, 1H, H−(C1′)), 5.21 (dd, J = 6.5, 4.6, 1H, HO), 3.95 (dd, J = 9.7, 6.2, 1H, H−(C2′)), 3.73 (dt, J = 11.1, 3.8, 1H, H−(C4′)), 3.46−3.41 (m, inc. J = 11.1, 1H, H−(C4′) partially obscured by H_2O signal), 3.18 (m, inc. J = 7.2, 1H, H−(C3′)). ^{13}C NMR (101 MHz; DMSO): δ_C 166.5, 159.2, 140.9, 129.3, 111.0, 79.0, 63.2, 54.4, 49.0.

HRMS ESI [M+H$^+$] calc. for $C_6H_{13}N_6O_2S$ 269.0821; obs. 269.0832. Cambridge Crystallographic Data Centre deposition number 864232.

24: Yield >70% *lyxo*-**24**: R_f = 0.20 (CH_2Cl_2/MeOH 7:3). IR (solid, cm^{-1}): 3344, 3106, 2917, 2833, 1627, 1553. ^{1}H NMR (400 MHz; DMSO): δ_H 7.14 (s, 1H, Ar), 6.90 (s, 2H, NH_2), 6.59 (bs, 1H, NH_2), 6.35 (s, 1H, NH), 5.93 (d, J = 6.1, 1H, H−(C1′)), 4.98 (d, J = 6.1, 1H, HO), 4.93 (t, J = 5.4, 1H, HO), 4.19 (dd, J = 7.6, 6.2, 1H, H−(C2′)), 3.59−3.54 (m, 1H, H−(C4′)), 3.49 (ABX, J = 10.9, 5.0, 1H, H−(C5′)), 3.46−3.41 (m, inc. J = 10.9, 1H, H−(C5′)), 3.36−3.34 (m, inc. J = 7.6 Hz, 1H, H−(C3′)). ^{13}C NMR (100 MHz; DMSO): δ_C 166.6, 160.5, 148.7, 141.2, 128.6, 110.8, 78.38, 78.35, 71.2, 63.2, 53.3, 50.5. HRMS ESI [M + Na$^+$] calc. for $C_{10}H_{14}N_6O_3SNa$ 321.0746; obs. 321.0742. Cambridge Crystallographic Data Centre deposition number 864233.

xylo-**24**: R_f = 0.20 (CH_2Cl_2/MeOH 7:3). IR (solid, cm^{-1}): 3309, 3125, 2936, 2835, 1632, 1565. ^{1}H NMR (400 MHz; D_2O): δ_H 7.26 (s, 1H, Ar), 5.94 (d, J = 5.8, 1H, H−(C1′)), 4.21 (t, J = 6.7, 1H, H−(C2′)), 3.73 (m, 1H, H−(C4′)), 3.69 (ABX, J = 11.8, 4.5, 1H, H−(C5′)), 3.64 (ABX, J = 11.8, 4.5, 1H, H−(C5′)), 3.52 (dd, J = 7.5, 1.9, 1H, H−(C3′)). ^{1}H NMR (400 MHz; DMSO): δ_H 7.05 (s, 1H, Ar), 6.99 (s, 1H, NH_2), 6.67 (bs, 1H, NH_2), 6.18 (s, 1H, NH), 6.01 (d, J = 6.3, 1H, H−(C1′)), 5.44 (d, J = 6.0, 1H, HO), 4.85 (t, J = 5.5, 1H, HO), 4.55 (d, J = 6.1, 1H, H−(C2′)), 3.69 (d, J = 8.3, 1H, H−(C4′)), 3.54 (ABX(OH), J = 11.3, 5.5, 1H, H−(C5′)), 3.47 (ABX(OH), J = 11.3, 5.5, 1H, H−(C5′)), 3.20 (m, 1H, H−(C3′)). ^{13}C NMR (101 MHz; DMSO): δ_C 166.5, 165.9, 141.0, 127.9, 111.5, 100.0, 79.7, 75.1, 63.5, 53.5. HRMS ESI [M + Na$^+$] calc. for $C_{10}H_{14}N_6O_3SNa$ 321.0746; obs. 321.0750. Cambridge Crystallographic Data Centre deposition number 864234.

■ ASSOCIATED CONTENT

Supporting Information

NMR and IR spectra, photographic images of sublimation experiments, crystallographic information (CIFs) and a stereochemical model for the addition of thiazole **7** to imine **21**. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

matthew.powner@ucl.ac.uk

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

M.W.P. was supported by a fellowship from the Harvard Origins Initiative, the Howard Hughes Medical Institute and University College London. This research was funded in part by Grant CHE-0809413 from the NSF. J.W.S. is an Investigator of the Howard Hughes Medical Institute. The authors thank J.C. Blain, Dr. R.B. Moreno, Dr. A. Ricardo, and Prof. J.D. Sutherland for helpful discussions and The Mass Spectrometry Facility at the FAS Center for Systems Biology for mass spectrometry support.

■ REFERENCES

(1) (a) Powner, M. W.; Gerland, B.; Sutherland, J. D. *Nature* **2009**, *495*, 239. (b) Powner, M. W.; Sutherland, J. D. *Angew. Chem.* **2010**, *122*, 4745; *Angew. Chem., Int. Ed.* **2010**, *49*, 4641. (c) Powner, M. W.; Sutherland, J. D.; Szostak, J. W. *Synlett* **2011**, *14*, 1956. (d) Hein, J. E.; Tse, E.; Blackmond, D. *Nat. Chem.* **2011**, *3*, 704.

(2) (a) Eschenmoser, A. *Science* **1999**, *284*, 2118. (b) *The RNA World*, 2nd ed.; Gesteland, R. F., Cech, T. R., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: New York, 1999.

(3) Powner, M. W.; Sutherland, J. D.; Szostak, J. W. *J. Am. Chem. Soc.* **2010**, *132*, 16677. Erratum. *J. Am. Chem. Soc.* **2011**, *133*, 4149.

(4) (a) Benner, S. A.; Ellington, A. D.; Tauer, A *Proc. Nati. Acad. Sci. U.S.A.* **1989**, *86*, 7054. (b) Robertson, M. P.; Joyce, G. F. *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, a003608.

(5) (a) Reichard, P. *Annu. Rev. Biochem.* **1988**, *57*, 349. (b) Stubbe, J. *J. Biol. Chem.* **1990**, *256*, 5329. (c) Stubbe, J. *Curr. Opin. Struct. Biol.* **2000**, *10*, 731.

(6) Average bond enthalpies: C−S (255 kJ mol^{-1})[6a] < C−O (355−380 kJ mol^{-1}).[6b,c] (a) Grelbig, T.; Pötter, B.; Seppelt, K. *Chem. Ber.* **1987**, *120*, 815. (b) Lovering, E. G.; Laider, K. J. *Can. J. Chem.* **1960**, *38*, 2367. (c) Levi, G. I.; Baladin, A. A. *Bull. Acad. Sci. USSR, Div. Chem. Sci.* **1960**, 149.

(7) Schimpl, A.; Lemmon, R. M.; Calvin, M. *Science* **1965**, *147*, 149.

(8) (a) Strecker, A. *Liebigs Ann. Chem.* **1850**, *75*, 27. (b) Miller, S. L.; Urey, H. C. *Science* **1959**, *130*, 245.

(9) (a) Sagan, C.; Khare, B. S. *Science* **1971**, *173*, 417. (b) Parker, E. T.; Cleaves, H. J.; Dworkin, J. P.; Glavin, D. P.; Callahan, M.; Aubrey, A.; Lazcanoe, A.; Bada, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 5526.

(10) Examples of desulfurization are legion in synthetic chemistry. (a) Belen'kii, L. I. *Chemistry of Organosulfur Compounds*: Ellis Horwood, Chichester, 1990, p 193. (b) Larock, R. C. *Comprehensive Organic Transformations*, 2nd ed.; Whiley-VCH: New York, 1999, pp 53−60 and references therein. (c) Anderson, C. D.; Goodman, L.; Baker, B. R. *J. Am. Chem. Soc.* **1959**, *81*, 3967. (d) Uesugi, S.; Shida, T.; Ikehara, M. *Chem. Pharm. Bull.* **1980**, *28*, 3621.

(11) C2′-inversion of ancitabine **28** (Scheme 1) by thiophosphates[11a,b] and the subsequent reduction of 2-thiocytidine[11c] have been reported by Nagyvary and co-workers. However, Divakar et al. questioned the validity of these results based upon the synthesis of similar compounds.[11d] (a) Bradbury, E.; Nagyvary, J. *Nucleic Acids Res.* **1976**, *3*, 2437. (b) Patel, A. D.; Schrier, W. H.; Nagyvary, J. *J. Org. Chem.* **1980**, *45*, 4830. (c) Patel, A. D.; Schrier, W. H.; Hrncir, M. A.; Nagyvary, J. *J. Origins Life* **1981**, 201. (d) Divakar, K. J.; Mottoh, A.; Reese, C. B.; Sanghvi, Y. S. *J. Chem. Soc., Perkin Trans. 1* **1990**, 969.

(12) (a) Barone, R.; Chanon, M.; Gallo, R. *Aminothiazoles and Their Derivatives in The Chemistry of Heterocyclic Compounds, 2;*. Vol. *34*;Metzger, J. V., Ed.; Wiley: New York, 1979.

(13) Currie, R. B.; Brutel, R.; Beutel, R. H. 1957, US PAT. 2,850,504.

(14) Significant H−(C4) exchange was observed in D_2O.

(15) Caraoni, P. C.; Reboul, J. P. *Acta Crystallogr.* **1982**, *B38*, 1255.

(16) Cambridge Crystallographic Data Centre deposition number for **7** is 864225.

(17) **16** is a product of HCN-oligomerization and known nucleobase precursor.[18] Notably, HCN also is required for Strecker synthesis.[8]

(18) (a) Oro, J. *Biochem. Biophys. Res. Commun.* **1960**, *2*, 407. (b) Lowe, C. U.; Rees, M. W.; Markham, R. *Nature* **1963**, *199*, 219.

(19) Cambridge Crystallographic Data Centre deposition numbers for **17**, *erythro*-**18**, *threo*-**18**, *threo*-**19**, and *threo*-**20** are 864226, 864227, 864228, 864229, and 864230, respectively.

(20) Proposed prebiotic nucleotide syntheses must be evaluated for their potential to yield alternative structures capable of base-pairing. Constitutionally, **3** (C_2) and **22** (C_3) can react with azoles (C_2 sugar-synthon) to furnish threose nucleic acid (TNA; C_4-sugar)[21] and DNA/RNA (C_5-sugar) precursors, respectively.[1]

(21) Schöning, K.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. *Science* **2000**, *290*, 1347.

(22) Cambridge Crystallographic Data Centre deposition numbers for *erythro*-**23**, *threo*-**23**, *lyxo*-**24**, and *xylo*-**24** are 864231, 864232, 864233, and 864234, respectively.

(23) (a) Anastasi, C.; Crowe, M. A.; Powner, M. W.; Sutherland, J. D. *Angew. Chem.* **2006**, *118*, 6322; *Angew. Chem., Int. Ed.* **2006**, *45*, 6176. (b) Sagi, V. S.; Karri, P.; Hu, F; Krishnamurthy, R. *Angew. Chem.* **2011**, *123*, 8277; *Angew. Chem., Int. Ed.* **2011**, *50*, 8127.

(24) The formation of aminal **25** was observed during the formation of 2-aminothiazole **7** in presence of stoichiomtric glycolaldehyde **3** and mercaptoacetaldehyde **6** and substiochiometric cyanamide **2**.

(25) Cambridge Crystallographic Data Centre deposition numbers for **25** and D-**27** are 864235 and 864236, respectively.

(26) (a) Butlerov, A. M. *Acad. Sci.* **1861**, *53*, 145. (b) Speck, J. C., Jr. *Adv. Carbohydr. Chem. Biochem.* **1958**, *13*, 63. (c) Breslow, R. *Tetrahedron Lett.* **1959**, *21*, 22. (d) Mizuno, T.; Weiss, A. H. *Adv. Carbohydr. Chem. Biochem.* **1974**, *29*, 173. (e) Ahmed, S. N.; McKee, M. L.; Shevlin, P. B. *J. Am. Chem. Soc.* **1985**, *107*, 1320. (f) Flanagan, G.; Ahmed, S. N.; Shevlin, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 3892. (g) Breslow, R.; Cheng, Z.-L. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 5723. (h) Kim, H.-J.; Ricardo, A.; Illangkoon, H. I.; Kim, M. J.; Carrigan, M. A.; Frye, F.; Benner, S. A. *J. Am. Chem. Soc.* **2011**, *133*, 9457.

(27) (a) Hollis, J. M.; Lovas, F. J.; Jewell, P. R. *Astrophys. J. Lett.* **2000**, *540*, L107. (b) Halfen, D. T.; Apponi, A. J.; Woolf, N.; Polt, R.; Ziurys, L. M. *Astrophys. J.* **2006**, *639*, 237.

(28) N.B. C_4-thiazoline desulfurization would generate monovalent nucleosides incapable of forming polymeric phosphodiesters.

(29) Trevino, S. G.; Zhang, N.; Elenko, M. P.; Luptak, A.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 13492.