



Ribonucleotides

John D. Sutherland

School of Chemistry, The University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom

Correspondence: john.sutherland@manchester.ac.uk

It has normally been assumed that ribonucleotides arose on the early Earth through a process in which ribose, the nucleobases, and phosphate became conjoined. However, under plausible prebiotic conditions, condensation of nucleobases with ribose to give β -ribonucleosides is fraught with difficulties. The reaction with purine nucleobases is low-yielding and the reaction with the canonical pyrimidine nucleobases does not work at all. The reasons for these difficulties are considered and an alternative high-yielding synthesis of pyrimidine nucleotides is discussed. Fitting the new synthesis to a plausible geochemical scenario is a remaining challenge but the prospects appear good. Discovery of an improved method of purine synthesis, and an efficient means of stringing activated nucleotides together, will provide underpinning support to those theories that posit a central role for RNA in the origins of life.

Whether RNA first functioned in isolation, or in the presence of other macromolecules and small molecules is still an open question, and a question that is addressable through chemistry (Borsenberger et al. 2004). If synergies are found between RNA assembly chemistry and that associated with the assembly of lipids and/or peptides, the purist RNA world concept (Woese 1967; Crick 1968; Orgel 1968) might have to be loosened to allow other such molecules a role in the origin of life. Metabolism, or the roots of metabolism, could also potentially have coevolved with RNA if organic chemistry happened to work in a particular way on a set of plausible prebiotic feedstock molecules in a dynamic geochemical setting. Such considerations point to the need for an open mind when considering the chemical

derivation of RNA. Notwithstanding these caveats, however, the self-assembly of polymeric RNA on the early Earth most likely involved activated monomers (Verlander et al. 1973; Ferris et al. 1996). These activated monomers could either have come together sequentially to make RNA one monomer at a time, or short oligoribonucleotides formed by such a process could have joined together by ligation in what would thus amount to a two-stage assembly of RNA polymers. Replication of RNA would then have involved template-directed versions of these or related chemistries (Orgel 2004).

The details of the polymerization processes that might plausibly have given rise to the first RNA molecules can only be investigated when there is some evidence as to the specific chemical nature of the activated monomers. Broadly

Editors: David Deamer and Jack W. Szostak
Additional Perspectives on The Origins of Life available at www.cshperspectives.org

Copyright © 2010 Cold Spring Harbor Laboratory Press; all rights reserved; doi: 10.1101/cshperspect.a005439
Cite this article as *Cold Spring Harb Perspect Biol* 2010;2:a005439

J.D. Sutherland

speaking, however, it is possible to differentiate different polymerization chemistries on the basis of the bonds formed in the polymerization step. P–O bond forming polymerization chemistry is reasonable to consider first because of the simplicity of P–O retrosynthetic disconnections of RNA (Corey 1988). In the case of the simplest P–O bond forming polymerization chemistry, the monomeric products of preceding prebiotic chemistry would either be activated ribonucleoside-5'-phosphates **1**—activated through having a leaving group attached to the phosphate—or ribonucleoside-2',3'-cyclic phosphates **2**, wherein the activation is intrinsic to the cyclic phosphate (Fig. 1). If all potential routes from prebiotic feedstock molecules to such monomers were to be investigated experimentally without success, then the potential for prebiotic self-assembly of monomers associated with more complicated polymerization chemistries, would additionally have to be investigated (this would include alternative P–O bond forming polymerization chemistry, as well as C–O and C–C bond forming chemistries). However, continuing with the simplest P–O bond forming polymerization chemistry, and the assumption that it seems reasonable to follow the simplest retrosynthetic disconnections first, **1** and **2** can then be conceptually reduced to ribose **3**, a nucleobase and phosphate (Joyce 2002; Joyce and Orgel 2006). Ribose **3** can then be disconnected to glycolaldehyde **4** and glyceraldehyde **5** through aldol chemistry, and the nucleobases disconnected to simpler carbon and nitrogen containing molecules—the pyrimidines to cyanoamide **6** and cyanoacetylene **7** (conventionally through the hydration products urea **8** and

cyoacetaldehyde **9**), and the purines to hydrogen cyanide and a C(IV) oxidation level molecule such as **6** or **8**. This retrosynthetic analysis ultimately breaks ribonucleotides down into molecules that are sufficiently simple so that they can be deemed prebiotically plausible feedstock molecules (Fig. 2) (Sanchez et al. 1966; Pasek and Lauretta 2005; Bryant and Kee 2006; Thaddeus 2006).

It is not just because the simplest retrosynthetic disconnections of ribonucleotides proceed by way of ribose, nucleobases, and phosphate that people have tried to synthesize them via these three building blocks under prebiotically plausible conditions for the last 40–50 years—in terms of their appearance to the human eye, ribonucleotides undoubtedly *consist* of these three building blocks. Experimentally, there have been several notably successful reactions that ostensibly support this nucleobase ribosylation approach: Orgel's and Miller's syntheses of cytosine **10** (Ferris et al. 1968; Robertson and Miller 1995); Benner's and Darbre's syntheses of ribose **3** by aldolization of glycolaldehyde **4** and glyceraldehyde **5** (Ricardo et al. 2004; Kofoed et al. 2005); Pasek's and Kee's demonstration of phosphate synthesis by disproportionation of meteoritic metal phosphides (Pasek and Lauretta 2005; Bryant and Kee 2006); and Orgel's urea-catalyzed phosphorylation of nucleosides (eg., **11** (B=C)→**2** (B=C)) (Lohrmann and Orgel 1971). Indeed, for many years, a prebiotically plausible synthesis of ribonucleotides from ribose **3**, the nucleobases, and phosphate has been tantalizingly close but for one step of the assumed synthesis—the joining of ribose to the nucleobases.

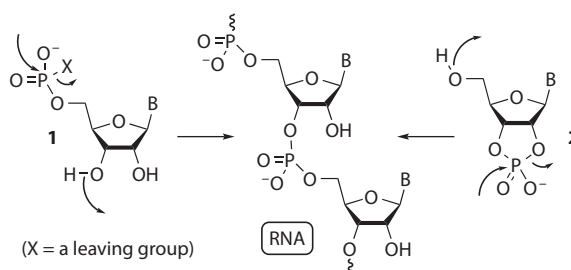


Figure 1. Activated ribonucleotides in the potentially prebiotic assembly of RNA. Potential P–O bond forming polymerization chemistry is indicated by the curved arrows.

This reaction works extremely poorly for the purines and not at all in the case of the pyrimidines (Fuller et al. 1972a, 1972b; Orgel 2004).

So, why does the ribosylation chemistry not work with free nucleobases and ribose **3** when, using the protecting and controlling groups of conventional synthetic chemistry, nucleobase ribosylation is possible? The reasons are predominantly kinetic and can be appreciated by consideration of the structure and reactivity of ribose **3** and representative nucleobases

(Fig. 3). Ribose **3** exists as an equilibrating mixture of different forms in aqueous solution (Fig. 3A) (Drew et al. 1998). The mixture is dominated by β - and α -pyranose isomers (**3** [β -p] and **3** [α -p]) with lesser amounts of β - and α -furanose isomers (**3** [β -f] and **3** [α -f]). The various hemiacetal ring forms equilibrate *via* the open chain aldehyde form **3** (a), which is a very minor component along with an open chain hydrate. The purine nucleobase adenine **12** also exists in various equilibrating forms in aqueous solution (Fig. 3B) (Fonseca Guerra et al. 2006). In this case, the isomers differ in the position of protonation, the major tautomer, **12** has N9 protonated, but other tautomers such as **13**—in which N1 is protonated—exist at extremely low concentration. To connect adenine **12** to ribose **3** to give a natural RNA ribonucleoside **11** (B=A) it is necessary for N9 of adenine to function as a nucleophile and C1 of **3** (α -f) to function as an electrophile. The latter is possible under acidic conditions when a small amount of **14**—a selectively protonated form of **3** (α -f)—is present at equilibrium. The protonation converts the anomeric hydroxyl group into a better leaving group and enhances the electrophilicity of C1. The major tautomer of adenine, **12**, is not nucleophilic at N9 because the lone pair on that atom is delocalized throughout the bicyclic ring structure. N9 of several minor tautomers such as **13** is nucleophilic because the nitrogen lone pair is localized, and so reaction with **14** is possible, though slow because of the low concentrations of the productively reactive species. To compound this sluggishness, the reaction is plagued with additional problems. First, the acid needed to activate **3** (α -f) also substantially protonates adenine, giving the cation **15**, which is not nucleophilic on N9 (Christensen et al. 1970; Zimmer and Biltonen 1972; Major et al. 2002). Second, the most nucleophilic nitrogen—the 6-amino group—of the major tautomer of adenine **12** reacts with **3** (a)—the most reactive form of **3** despite its scarcity—resulting in N6-ribosyl adducts as by far the major products (Fuller et al. 1972a, 1972b). Third, the other isomeric forms of ribose, **3** (β -p), **3** (α -p), and **3** (β -f), can

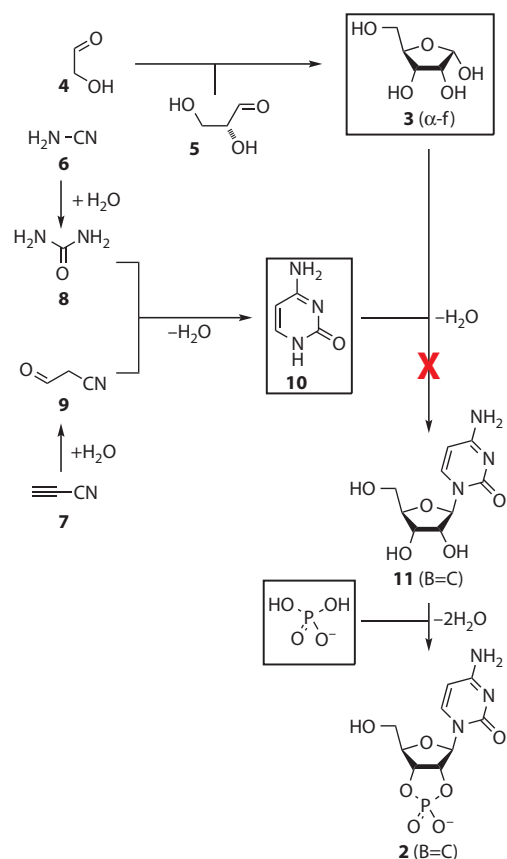


Figure 2. One of the synthetic routes to β -ribocytidine-2',3'-cyclic phosphate **2** (B=C) implied by the assumption that nucleosides can self-assemble by nucleobase ribosylation. The general synthetic approach has been supported by the experimental demonstration of most of its steps. However, prebiotically plausible conditions under which the key nucleobase ribosylation step works have not been found despite numerous attempts over several decades.

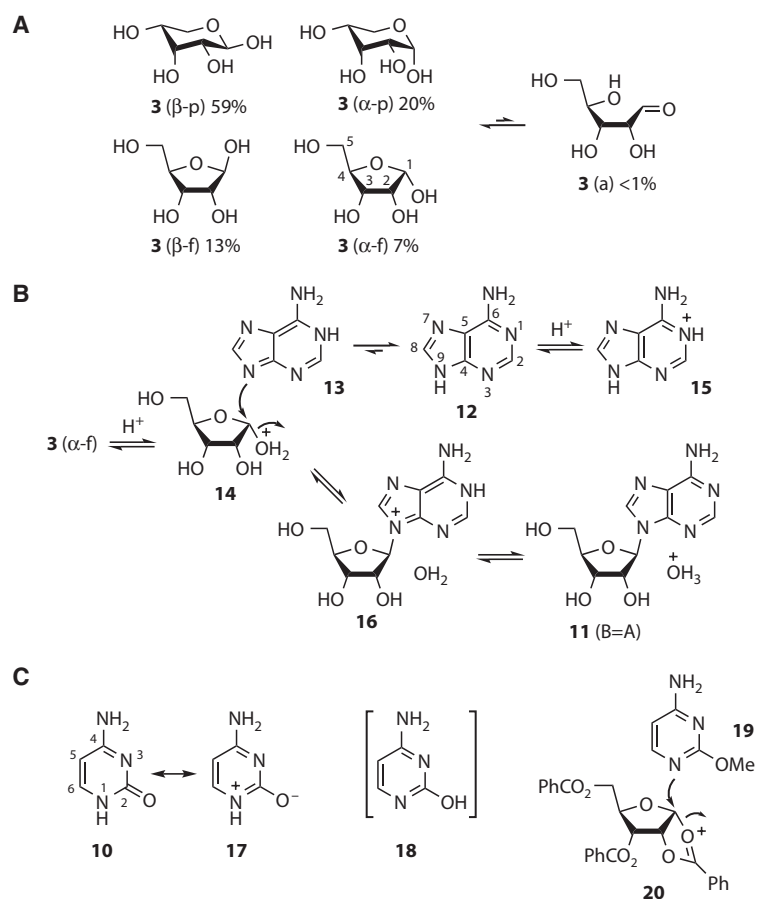


Figure 3. The difficulties of assembling β -ribonucleosides by nucleobase ribosylation. (A) The many different forms of ribose **3** adopted in aqueous solution. The pyranose (p) and furanose (f) forms interconvert via the open-chain aldehyde (a), which is also in equilibrium with an open-chain aldehyde hydrate (not shown). (B) Adenine tautomerism and the ribosylation step necessary to make the adenosine **11** (B=A) thought to be needed for RNA assembly. The low abundance of the reactive entities **13** and **14** is partly responsible for the low yield of **11** (B=A). (C) The reason for the lower nucleophilicity of N1 of the pyrimidines, and the conventional synthetic chemist's solution to the problems of ribosylation.

also react with **13** when they are protonated at their anomeric hydroxyl groups. Fourth, N9 of the minor adenine tautomer **13** is not the only nucleophilic ring nitrogen of adenine; N1, N3, and N7 of the major tautomer **12** are also nucleophilic. These latter two points mean that the small amount of protonated adenosine **16**, that is formed when N9 of **13** reacts with **14**, is accompanied by a multitude of isomeric products. The final problem with the synthesis is reversibility. Any adenosine **11** (B=A) that is produced is formed in acid at equilibrium, and the equilibrium in aqueous solution lies

in favor of **3** and **15**. The only way round this is to carry out the reaction in the dry-state in the presence of acidic catalysts. The best that has been achieved—a 4% yield of **11** (B=A)—involves such a dry-state reaction with an excess of ribose, followed by heating with concentrated ammonium hydroxide solution to hydrolyse N6-ribosyl adducts (Fuller et al. 1972a).

The situation with prebiotic pyrimidine ribosylation is even worse (Fig. 3c). Thus, for example, N1 of cytosine **10** is not nucleophilic because the lone pair is delocalized round the ring and into the carbonyl group (as indicated

by the resonance canonical structure **17**). Experimentally, cytosine **10** cannot be ribosylated on N1 even using the conditions established for unselective ribosylation of adenine **12** (Fuller et al. 1972a). If there is any tautomeric form such as **18** with a localized N1 lone pair present at equilibrium, it must be in such a low concentration as to be effectively unreactive. The same considerations hold true for uracil.

In conventional synthetic chemistry, the aforementioned difficulties in nucleobase ribosylation can be overcome with directing, blocking, and activating groups on the nucleobase and ribose (Ueda and Nishino 1968). Thus, the cytosine derivative **19** is directed to function as a nucleophile at N1 by alkylation of O2. The ribose-derived intermediate **20** is constrained to the furanose form by benzylation of the C5-hydroxyl group, and neighboring group participation from an O2-benzoyl group directs β -glycosylation. These molecular interventions are synthetically ingenious, but serve to emphasize the enormous difficulties that must be overcome if ribonucleosides are to be efficiently produced by nucleobase ribosylation under prebiotically plausible conditions. This impasse has led most people to abandon the idea that RNA might have assembled abiotically, and has prompted a search for potential pre-RNA informational molecules (Joyce et al. 1987; Eschenmoser 1999; Schöning et al. 2000; Zhang et al. 2005; Sutherland 2007). However, we realized that there were other possible synthetic approaches that although less obvious, still had the potential to make the ribonucleotides **1** and **2** (Anastasi et al. 2007). Furthermore, and as pointed out earlier, there are also alternative bond forming polymerization chemistries imaginable. Our plan was to work through these possibilities by systematic experimentation before deciding whether the abiogenesis of RNA is possible or not.

RECENT RESULTS

As building blocks for pyrimidine ribonucleotides, we have investigated the chemistry of glycolaldehyde **4**, glyceraldehyde **5**, cyanamide **6**, cyanoacetylene **7**, and phosphate—the same

feedstock molecules ultimately invoked in the nucleobase ribosylation approach (Fig. 2) (in ongoing work on the prebiotic synthesis of purine nucleotides, the systems chemistry of mixtures containing hydrogen cyanide, is additionally being investigated). However, we have not presumed a particular order of assembly, and have systematically investigated many options over the last decade or so (Sutherland and Whitfield 1997; Ingar et al. 2003; Smith et al. 2004; Smith and Sutherland 2005; Anastasi et al. 2007, 2008). Furthermore, aware of the potential for certain molecules to act as catalysts of reactions in which other molecules are joined, an important aspect of our approach has been the study of the chemistry of multi-component chemical systems, including mixed oxygenous and nitrogenous chemistry (Szostak 2009). It has long been held in prebiotic chemistry that the aldol chemistry of aldehydes potentially required to make sugars should not be mixed with, for example, the oligomerization of hydrogen cyanide and cyanamide potentially required to make purine nucleobases (Shapiro 1988). This is because the two chemistries have been assumed to interfere with each other, suggesting the potential for a “combinatorial explosion” of minor by-products. However, given the impasse of preformed nucleobase ribosylation, and the conceptual difficulties with which a transition from a pre-RNA world to the RNA world is fraught, taking a few synthetic risks seemed justified.

Cutting straight to the chase, we found that mixed oxygenous-nitrogenous chemistry of glycolaldehyde **4**-cyanamide **6** mixtures can be tamed by the inclusion of phosphate such that the heterocycle 2-aminooxazole **21** is obtained in >80% yield (Fig. 4) (Powner et al. 2009). The phosphate functions both as a pH buffer and as a general acid-base catalyst. The heterocycle **21** readily sublimates, potentially offering a prebiotically plausible purification through sublimation and precipitation or rain-in. In a smooth C–C bond forming reaction in water at neutral pH, **21** then adds to glyceraldehyde **5**, giving the four pentose aminooxazolines, but with high stereoselectivity for the *ribo*- and *arabino*-configured materials (Anastasi et al.

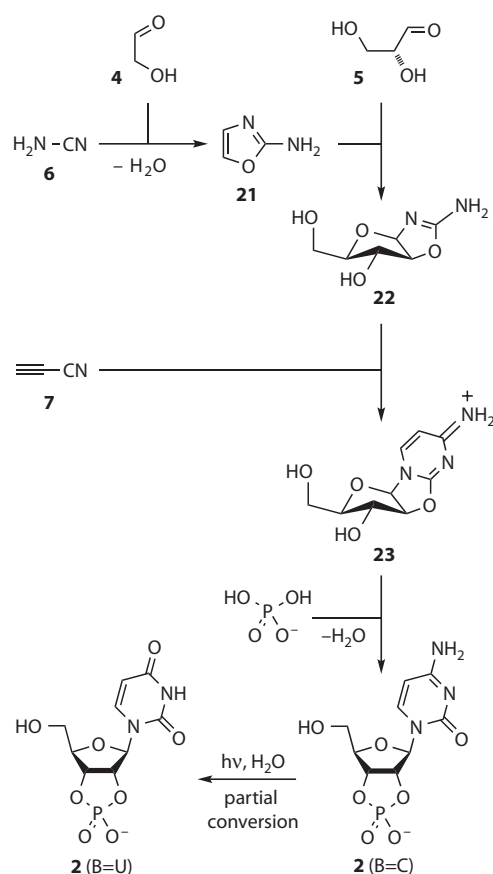


Figure 4. The recently uncovered route to activated pyrimidine nucleotides 2. The nucleobase ribosylation problem is circumvented by the assembly proceeding through 2-aminooxazole 21, which can be thought of as the chimera of half a pentose sugar and half a nucleobase. The second half of the pentose—glyceraldehyde 5—and the second half of the nucleobase—cyanoacetylene 7—are then added sequentially to give the anhydronucleoside 23. Phosphorylation and rearrangement of 23 then furnishes 2 (B=C), and UV irradiation effects the partial conversion of 2 (B=C) to 2 (B=U).

2006). Remarkably, the *ribo*-isomer then selectively crystallizes out of solution, leaving the *arabino*-isomer 22 the most abundant in solution. Reaction of 22 with cyanoacetylene 7 furnishes the anhydronucleoside 23 if the pH of the reaction is maintained at pH=6.5 through the use of phosphate as a buffer (Powner et al. 2009). Heating 23, inorganic phosphate, and urea 8 in the dry-state or in formamide

solution then gives the activated ribonucleotide 2 (B=C). Finally, irradiation of 2 (B=C) with UV light in solution, followed by heating (Powner and Sutherland 2008), converts it to a mixture of 2 (B=C) and 2 (B=U), the two activated pyrimidine nucleotides needed for RNA synthesis.

The presumed mechanisms of the various reactions that make up the self-assembly sequence (Fig. 5) reveal some remarkable aspects of selectivity, and show the synthetic power of systems chemistry (Anastasi et al. 2006; Powner et al. 2009). In the first step, we found that the reaction of glycolaldehyde 4 and cyanamide 6 generated a series of dead end adducts and a small amount of 2-aminooxazole 21 in the absence of phosphate, but in its presence 21 was produced cleanly in extremely high yield. This improvement in the yield of 21 is attributed to general acid–base catalysis of two steps—that in which the nitrile carbon undergoes attack by a hydroxyl group nucleophile, and the C–H deprotonation step that generates the aromatic ring of 21 (Fig. 5A). Phosphate is an ideal general acid–base catalyst in this context as the reaction is ideally conducted at near neutral pH values close to the second pK_a of phosphoric acid. Phosphate is also known to catalyze both the dimerization and the hydrolysis of cyanamide 6 (Lohrmann and Orgel 1968), but these reactions are slower than the formation of 21. However, if an excess of 6 is allowed to react with glycolaldehyde 4 in the presence of phosphate, the formation of 21 is followed by the formation of cyanoguanidine and urea 8.

In the second step of the self-assembly process, 2-aminooxazole 21 adds to glyceraldehyde 5, giving all four pentose aminooxazolines but, as mentioned previously, with high stereoselectivity for the *ribo*-isomer, and the key *arabino*-isomer 22 (Fig. 5B). The reaction initially follows the course of an electrophilic aromatic substitution reaction with the formation of a cationic intermediate. However, rather than rearomatizing by C–H deprotonation, this intermediate instead undergoes intramolecular addition of a hydroxyl group (Anastasi et al. 2006). Although the C5'-hydroxyl group

is intrinsically the most nucleophilic, it is the C4'-hydroxyl group that adds kinetically because its addition proceeds via a five-membered (cf., six-membered) transition state. The chemistry is partly reversible, however, especially in the presence of phosphate, and the product of attack by the C5'-hydroxyl group—a pyranose aminooxazoline—can be accessed, as can the electrophilic aromatic substitution product—an open sugar chain aminooxazole isomer—when the key C–H deprotonation step is general base catalysed by phosphate (Fig. 5B, red arrows). In the absence of phosphate, the *ribo-arabino-* and *xylo-*aminooxazolines exist

in the furanose form, and it is only the *lyxo-*aminooxazoline that exists partly in the pyranose form, presumably because of steric encumbrance in the furanose form (Saewan et al. 2005). Under partial thermodynamic control in the presence of phosphate, the formation of aminooxazoline versus substituted aminooxazole is finely balanced and depends on the stability of the different aminooxazoline stereoisomers. Intriguingly, the *ribo-*, *arabino-*, and *xylo-*products preferentially exist as aminooxazolines but the *lyxo-*product preferentially exists as the substituted aminooxazole. Phosphate thus makes the reaction of 2-aminooxazole **21**

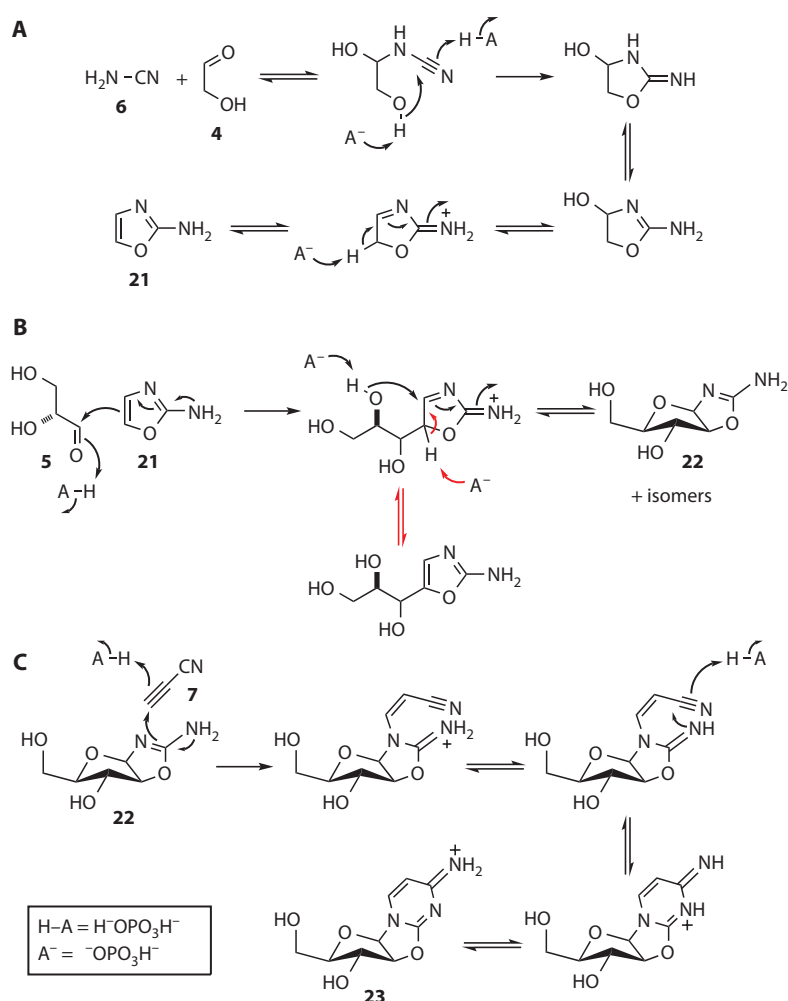
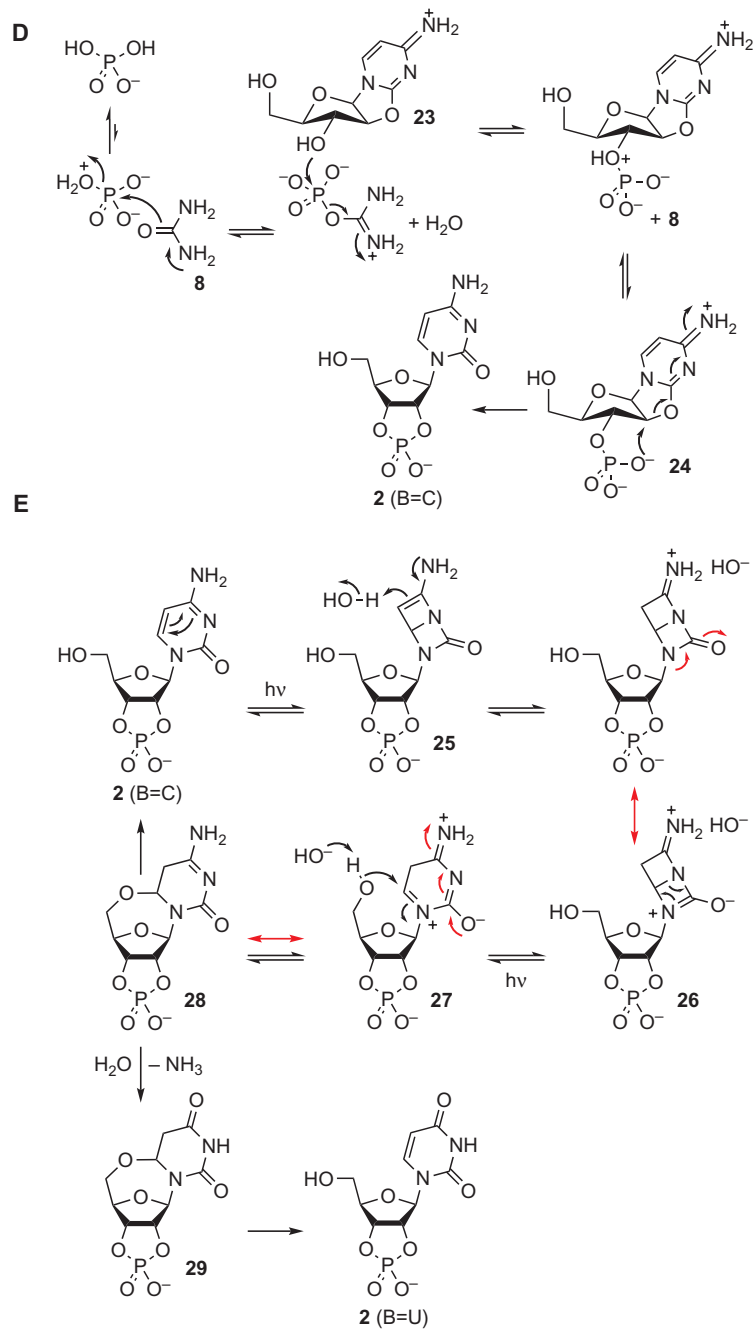


Figure 5. Mechanistic details of each of the five steps—(A) through (E)—of the recently uncovered route to activated pyrimidine nucleotides **2**. In (E), resonance arrows are shown in red and reaction arrows in black.



and glyceraldehyde **5** more selective for the formation of the desired *arabino*-aminooxazoline **22** (Powner et al. 2009).

In addition to its role as a pH buffer in the addition of **22** to cyanoacetylene **7** (at pH values higher than 6.5, or in the absence of phosphate, the anhydronucleoside **23** undergoes hydrolysis of the C2–O2' bond giving β -arabinocytidine [Sanchez and Orgel 1970]), phosphate almost certainly functions as a general acid catalyst for the first addition step, thereby allowing the reaction to proceed by a path that does not require the intermediacy of a high-energy cyanovinyl anion (Fig. 5C). Furthermore, reaction of excess cyanoacetylene with the hydroxyl groups of the anhydronucleoside product **23** is prevented by phosphate since HPO_4^{2-} is more nucleophilic than these hydroxyl groups. In the course of acting as a chemical buffer in this way, the phosphate is partly transformed into cyanovinyl phosphate, and this latter compound, over time, reacts with more phosphate to give pyrophosphate (Ferris et al. 1970).

The urea **8** produced as a by-product in the reaction of glycolaldehyde **4** with cyanamide **6**, if the latter is in excess, then acts as a catalyst for the incorporation of phosphate in the fourth step of the self-assembly sequence (Fig. 5D). At high temperatures in the dry-state (Lohrmann and Orgel 1971), or in formamide solution (Schoffstall 1976), the weakly nucleophilic urea oxygen atom attacks a minor tautomer of H_2PO_4^- , displacing water via a dissociative, monomeric metaphosphate-like transition state to give an imidoyl phosphate. Although this reaction is reversible in principle, water is lost at high temperature, driving the formation of the high-energy intermediate, which can then phosphorylate the hydroxyl groups of nonvolatile compounds such as the anhydronucleoside **23**. Again, this reaction is reversible in principle, though it is unlikely to be fully so in urea melts because the viscosity of the reaction medium likely prevents free diffusion. It transpires that **23** adopts a conformation that potentially makes the C5'-hydroxyl group more hindered than the C3'-hydroxyl group; thus, phosphorylation of the latter is apparently preferred kinetically (Powner et al. 2009). The dianionic form of

the phosphate group of the O3'-phosphorylated anhydronucleoside **24** can then attack C2' in an intramolecular $\text{S}_{\text{N}}2$ reaction, thereby inverting the configuration of that carbon, and giving the ribocytidine-2',3'-cyclic phosphate **2** (B=C) (Tapiero and Nagyvary 1971). Alternatively, if kinetic selectivity is not fully operative, the O5'- and O3'-phosphorylation products of **23** could both be sampled at equilibrium, and the equilibrium could then be displaced to the detriment of the former through irreversible conversion of the latter to **2** (B=C). Pyrophosphate—the product of reaction of the cyanovinyl phosphate by-product in the preceding step with additional phosphate (Ferris et al. 1970)—can also function as a phosphorylating agent in urea melts. In this case, the imidoyl phosphate is generated by nucleophilic displacement of phosphate (cf., water) via a dissociative, monomeric metaphosphate-like transition state. Pyrophosphate is not soluble in formamide, but there is another twist to phosphorylation of **23** with inorganic phosphate with urea in formamide solution, and that is that formamide is also weakly nucleophilic on its oxygen atom, and so can also react to give imidoyl phosphate intermediates—indeed, phosphorylation can be accomplished in formamide in the absence of urea (Schoffstall 1976). Formamide can be considered to be prebiotically available because it is the hydration product of hydrogen cyanide. The latter is inevitably also produced when cyanoacetylene **7** is made by high-energy atom recombination chemistry (Sanchez et al. 1970). Given that the purine nucleobases are constitutionally derivable from hydrogen cyanide, the case for the involvement of this compound and its hydration product in the prebiotic synthesis of pyrimidine and purine nucleotides seems extremely strong (Sanchez et al. 1967).

In the last step of the self-assembly sequence, **2** (B=C) is partly converted to **2** (B=U) by UV irradiation. Remarkably, the irradiation conditions destroy other cytidine nucleotides, and since some such compounds will likely arise from similar assembly chemistry proceeding from aminooxazolines other than **22**, the irradiation might also serve a sanitizing function such

J.D. Sutherland

that **2** could be produced in the absence of stereoisomeric contaminants. The proposed mechanism of the photochemistry (Fig. 5E) is speculative, but supported by a significant body of fact. It has long been known that irradiation of other cytidine nucleosides and nucleotides leads to C5-photohydrates that eliminate water to regenerate the original cytosine derivative (Miller and Cerutti 1968). During prolonged irradiation, such cytosine derivatives go through many cycles of photoexcitation and thermal relaxation. Deleterious photochemistry—nucleobase loss, C1'- and C2'-epimerization, hydrolysis to uracil derivatives, etc.—increases with increased irradiation, presumably because high energy intermediates are thus accessed for longer (Powner et al. 2007; Powner and Sutherland 2008). The molecular details of these processes have been elusive, but recent work has suggested plausible mechanisms that can be further tested. It is thought that irradiation of the cytosine ring leads to the formation of a Dewar pyrimidine **25** through a disrotatory 4π electrocyclic process (Shaw and Shetlar 1990). Steric inhibition of resonance is then expected to make C5-protonation facile, giving an amidinium ion that can undergo an electrocyclic ring-opening—most easily visualized, as shown, for the resonance canonical form **26** (Fig. 5E). The consequence of this second pericyclic process is the formation of an N1,C6-iminium ion **27**. In most molecular contexts, an N1,C6-iminium ion would be expected to undergo attack by hydroxide anion, giving a cytidine photohydrate, but in the irradiation of **2** (B=C), the 2',3'-cyclic phosphate allows ready access to otherwise inaccessible western sugar ring conformers (Saenger 1984) in which the C5'-hydroxyl group is proximal to C6 and can thus add to it instead. Thus, the specific nature of the cyclic phosphate of **2** (B=C) controls the cytosine photochemistry through enforced intramolecular nucleophilic addition. The resultant C5,O5'-anhydronucleotide **28** is thought to undergo thermal elimination back to **2** (B=C) more slowly than other cytidine photohydrates eliminate water because of conformational restriction about the C5–C6

bond. This has the effect that **2** (B=C) undergoes fewer photochemical excitation and thermal relaxation cycles than other cytidine nucleosides and nucleotides. The consequences of this are twofold. Firstly, deleterious chemistry from high-energy (charged) intermediates is reduced because these intermediates are accessed less often. Secondly, the persistence of the C5,O5'-anhydronucleotide allows for greater hydrolysis to the corresponding uracil derivative **29**. Relaxation of this uracil derivative is then responsible for the high yield of **2** (B=U) relative to the yield of the corresponding uridine nucleosides and nucleotides in the irradiation of other cytidine nucleosides and nucleotides. Indeed, most other pyrimidine nucleosides and nucleotides are destroyed by the prolonged irradiation that partly converts **2** (B=C) to **2** (B=U). This has the beneficial effect that **2** (B=C) and **2** (B=U) can be effectively sanitized by destruction of nucleoside and nucleotide by-products of the assembly process described herein.

TOWARD A GEOCHEMICALLY PLAUSIBLE SYNTHESIS

The contextually specific chemoselectivity, and the systems chemistry aspects of this newly uncovered self-assembly route to activated pyrimidine nucleotides suggest that the chemistry should be viewed as predisposed. However, the route as operated thus far in the laboratory is associated with several steps, and the conditions for these steps are different. Furthermore, purification in between certain steps was carried out to make analysis of the chemistry easier. Clearly, these issues need to be addressed before the synthesis can be seen as geochemically plausible.

First, the sequence of different conditions needs considering. Miller's iconic experiment (Miller 1953) has apparently conditioned many in the field to think that a prebiotic synthesis needs to occur under one set of conditions. But, think of chemistry occurring now on the Earth. There is no doubt that in many locations, conditions vary with time—there are light periods and dark periods, hot periods and cold periods, dry periods and wet periods.

If the same were true on the primordial Earth, then why should we expect prebiotic synthesis just to occur in one particular period? Furthermore, the frequent impacts of asteroids, meteorites, and comets on the early Earth would surely have exacerbated the changeability of geochemical conditions. Given that there are so many geochemically plausible conditions, and sequences of conditions, it is difficult to try and predict solely from geochemistry the actual conditions that pertained for the prebiotic synthesis of any particular compound or class of compounds. Surely, it makes more sense to first find predisposed chemical routes to molecules of interest and then ask whether the sequence of conditions is geochemically plausible. If it then turns out that this sequence of conditions additionally supports the synthesis of other important molecules, then evidence might accumulate that the sequence of conditions actually did take place during the origin of life on Earth. Furthermore, the search for further relevant prebiotic synthesis would then be expedited. From what is now known about pyrimidine nucleotide self-assembly, one geochemically plausible sequence of conditions—coincidentally redolent of Darwin's suggestion (Darwin 1871)—involves a warm pond that evaporates and dries out, stays dry and is heated for a time, is then filled again by rain, and subsequently bathed in sunshine.

Secondly, one has to wonder if any purification steps could reasonably be invoked in such a scenario, and ask if the other chemistry would work in the absence of purification. 2-Aminooxazole **21** sublimates readily on being warmed at atmospheric pressure, and resolidifies on cooling surfaces. This property suggests that **21** could plausibly be purified prebiotically—if necessary—and be relocated by precipitation. If delivered to glyceraldehyde **5** and phosphate, then reaction of **21** and **5** giving the pentose aminooxazolines, including the key *arabino*-isomer **22**, certainly seems plausible. Subsequent delivery of cyanoacetylene **7**—produced by atmospheric chemistry—by rain in also seems plausible, though there is no doubt that it would be accompanied by hydrogen cyanide. Will hydrogen cyanide interfere

with the reaction of **22** and **7**, giving the anhydronucleoside **23**? The answer awaits experimental evaluation. If the hydrogen cyanide does not interfere, but slowly hydrolyses to formamide, then when the water in the pond evaporates, a formamide solution would remain. If this solution gets heated, the phosphorylation and rearrangement of **23** to **2** ($B=C$) will take place whether the pond additionally happened to contain urea or not. The issue seems not to be whether the chemistry could take place or not, but the yield and purity of products. Because of the other aminooxazolines, there would doubtless be other stereoisomeric products at this point, but this would not matter as subsequent dilution with water and irradiation would not just partly convert **2** ($B=C$) to **2** ($B=U$), but also destroy these other products. So, there is more work to be done, and although it is premature to conclude that the newly discovered self-assembly route to activated pyrimidine nucleotides (Powner et al. 2009) is geochemically plausible, the signs are looking good! What is now needed on top of further experimental work to assess the geochemical plausibility of the synthesis, is a thorough investigation of the synthesis of activated purine nucleotides.

FUTURE CHALLENGES

If further experimentation supports nucleoside-2',3'-cyclic phosphates **2** as the activated monomers for prebiotic RNA synthesis, then an obvious next question is how such monomers oligomerise. Pioneering work from the 1970s partly addresses this and suggests that general acid-base catalysis is the key (Verlander et al. 1973). In the dry-state, mixtures of aliphatic amines and the corresponding ammonium salts are effective oligomerization catalysts. It is thought that the amines deprotonate the C5'-hydroxyl group of one nucleoside-2',3'-cyclic phosphate **2**, enabling it to attack the phosphate group of another. The ammonium salts serve to protonate the oxygen leaving group generating a C2'- or C3'-hydroxyl group. Short oligomers are easily produced, but the chemistry generates both 3',5'- and

2',5'-internucleotide links, and further work is needed to address this selectivity issue.

The oligomerization of racemic mixtures of nucleoside-2',3'-cyclic phosphates **2** clearly has the potential to generate extremely complex mixtures of diastereoisomers, so a major goal has to be the generation of these monomers in enantiopure form. This issue has not yet been solved, but it is apparent from the self-assembly chemistry that if glyceraldehyde **5** can be obtained in enantiopure form then the resultant pyrimidine nucleotides will be similarly enantiopure. How glyceraldehyde **5** might be obtained in enantiopure is a question that can only be answered when a prebiotically plausible synthesis of **5** is shown. The conditions potentially required to generate glyceraldehyde **5** by the aldolization of glycolaldehyde **4** and formaldehyde are the same conditions that favour the conversion of **5** to the more stable dihydroxyacetone. A synthesis of **5**—and indeed **4**—from formaldehyde involving umpolung would therefore be preferable, but prebiotically plausible carbonyl umpolung has not yet been shown.

Finally, the potential for simultaneous self-assembly of RNA and other molecules such as lipids and peptides must be investigated. Compartmentalization of RNA leading to protocells is seen as a crucial step towards systems that can undergo Darwinian evolution (Szostak et al. 2001), so can lipids be produced alongside nucleotides? Peptides, even dipeptides, offer additional scope for catalysis, and their prebiotic synthesis and use could help drive the evolution of the ribosome and the genetic code, so can peptides also be produced under the same conditions? Experimental investigation of such questions will rely heavily on analytical chemistry because of the complexity of the multicomponent systems that will have to be investigated. There is cause for optimism, however, as prebiotic systems chemistry, though still in its infancy, is already suggesting new solutions to old problems.

REFERENCES

- Anastasi C, Buchet FE, Crowe MA, Helliwell M, Raftery J, Sutherland JD. 2008. The search for a potentially

prebiotic synthesis of nucleotides via arabinose-3-phosphate and its cyanamide derivative. *Chem Eur J* **14**: 2375–2388.

Anastasi C, Buchet FE, Crowe MA, Parkes AL, Powner MW, Smith JM, Sutherland JD. 2007. RNA: Prebiotic product or biotic invention? *Chem Biodiversity* **4**: 721–739.

Anastasi C, Crowe MA, Powner MW, Sutherland JD. 2006. Direct assembly of nucleoside precursors from two- and three-carbon units. *Angew Chem Int Ed* **45**: 6176–6179.

Borsenberger V, Crowe MA, Lehbauer J, Raftery J, Helliwell M, Bhutia K, Cox T, Sutherland JD. 2004. Exploratory studies to investigate a linked prebiotic origin of RNA and coded peptides. *Chem Biodiversity* **1**: 203–246.

Bryant DE, Kee TP. 2006. Direct evidence for the availability of reactive, water soluble phosphorus on the early Earth. H-phosphinic acid from the Nantan meteorite. *Chem Commun* **14**: 2344–2346.

Christensen JJ, Rytting JH, Izatt RM. 1970. Thermodynamic pK , ΔH° , ΔS° , and ΔC_p° values for proton dissociation from several purines and their nucleosides in aqueous solution. *Biochemistry* **9**: 4907–4913.

Corey EJ. 1988. Retrosynthetic thinking - essentials and examples. *Chem Soc Rev* **17**: 111–133.

Crick FHC. 1968. The origin of the genetic code. *J Mol Biol* **38**: 367–379.

Darwin C. 1871. Letter to JD Hooker of 1st February. Cambridge University Library MS. DAR 94: 188–189 (*Calendar* 7471).

Drew KN, Zajicek J, Bondo G, Bose B, Serianni AS. 1998. ¹³C-labeled aldopentoses: detection and quantitation of cyclic and acyclic forms by heteronuclear 1D and 2D NMR spectroscopy. *Carbohydrate Res* **307**: 199–209.

Eschenmoser A. 1999. Chemical etiology of nucleic acid structure. *Science* **284**: 2118–2124.

Ferris JP, Goldstein G, Beaulieu DJ. 1970. Chemical evolution, IV. An evaluation of cyanovinyl phosphate as a prebiotic phosphorylating agent. *J Am Chem Soc* **92**: 6598–6603.

Ferris JP, Sanchez RA, Orgel LE. 1968. Studies in prebiotic synthesis: III. Synthesis of pyrimidines from cyanoacetylene and cyanate. *J Mol Biol* **33**: 693–704.

Ferris JP, Hill AR Jr, Liu R, Orgel LE. 1996. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381**: 59–61.

Fonseca Guerra C, Bickelhaupt FM, Saha S, Wang F. 2006. Adenine tautomers: relative stabilities, ionization energies, and mismatch with cytosine. *J Phys Chem A* **110**: 4012–4020.

Fuller WD, Sanchez RA, Orgel LE. 1972a. Studies in prebiotic synthesis VI. Synthesis of purine nucleosides. *J Mol Biol* **67**: 25–33.

Fuller WD, Sanchez RA, Orgel LE. 1972b. Studies in prebiotic synthesis. VII Solid-state synthesis of purine nucleosides. *J Mol Evol* **1**: 249–257.

Ingar A-A, Luke RWA, Hayter BR, Sutherland JD. 2003. Synthesis of cytidine ribonucleotides by stepwise assembly of the heterocycle on a sugar phosphate. *ChemBioChem* **6**: 504–507.

Joyce GF. 2002. The antiquity of RNA-based evolution. *Nature* **418**: 214–221.



- Joyce GF, Orgel LE. 2006. In *The RNA world*, 3rd ed. (ed. Gesteland R.F., Cech T.R., Atkins J.E.), pp. 23–56. Cold Spring Harbor Laboratory Press.
- Joyce GF, Schwartz AW, Miller SL, Orgel LE. 1987. The case for an ancestral genetic system involving simple analogues of the nucleotides. *Proc Natl Acad Sci* **84**: 4398–4402.
- Kofoed J, Reymond J-L, Darbre T. 2005. Prebiotic carbohydrate synthesis: Zinc-proline catalyses direct aqueous aldol reactions of α -hydroxy aldehydes and ketones. *Org Biomol Chem* **3**: 1850–1855.
- Lohrmann R, Orgel LE. 1968. Prebiotic synthesis: Phosphorylation in aqueous solution. *Science* **161**: 64–66.
- Lohrmann R, Orgel LE. 1971. Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents. *Science* **171**: 490–494.
- Major DT, Laxer A, Fischer B. 2002. Protonation studies of modified adenine and adenine nucleotides by theoretical calculations and ^{15}N NMR. *J Org Chem* **67**: 790–802.
- Miller SL. 1953. A production of amino acids under possible primitive earth conditions. *Science* **117**: 528–529.
- Miller N, Cerutti P. 1968. Structure of the photohydration products of cytidine and uridine. *Proc Natl Acad Sci* **59**: 34–38.
- Orgel LE. 1968. Evolution of the genetic apparatus. *J Mol Biol* **38**: 381–393.
- Orgel LE. 2004. Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol* **39**: 99–123.
- Pasek MA, Lauretta DS. 2005. Aqueous corrosion of phosphide minerals from iron meteorites: A highly reactive source of prebiotic phosphorus on the surface of the early Earth. *Astrobiology* **5**: 515–535.
- Powner MW, Sutherland JD. 2008. Potentially prebiotic synthesis of pyrimidine β -D-ribonucleotides by photoanomerization/hydrolysis of β -D-cytidine-2'-phosphate. *ChemBioChem* **9**: 2386–2387.
- Powner MW, Gerland B, Sutherland JD. 2009. Synthesis of activated pyrimidine nucleotides in prebiotically plausible conditions. *Nature* **459**: 239–242.
- Powner MW, Anastasi C, Crowe MA, Parkes AL, Raftery J, Sutherland JD. 2007. On the prebiotic synthesis of ribonucleotides: Photoanomerisation of cytosine nucleosides and nucleotides revisited. *ChemBioChem* **8**: 1170–1179.
- Ricardo A, Carrigan MA, Olcott AN, Benner SA. 2004. Borate minerals stabilize ribose. *Science* **303**: 196.
- Robertson MP, Miller SL. 1995. An efficient prebiotic synthesis of cytosine and uracil. *Nature* **375**: 772–774.
- Saenger W. 1984. *Principles of nucleic acid structure*. Springer-Verlag, New York.
- Saewan N, Crowe MA, Helliwell M, Raftery J, Chantrapromma K, Sutherland JD. 2005. Exploratory studies to investigate a linked prebiotic origin of RNA and coded peptides. 4th Communication: Further observations concerning pyrimidine nucleoside synthesis by stepwise nucleobase assembly. *Chem Biodiversity* **2**: 66–83.
- Sanchez RA, Orgel LE. 1970. Studies in prebiotic synthesis V. Synthesis and photoanomerization of pyrimidine nucleosides. *J Mol Biol* **47**: 531–543.
- Sanchez RA, Ferris JP, Orgel LE. 1966. Cyanoacetylene in prebiotic synthesis. *Science* **154**: 784–785.
- Sanchez RA, Ferris JP, Orgel LE. 1967. Studies in prebiotic synthesis II. Synthesis of purine precursors and amino acids from aqueous hydrogen cyanide. *J Mol Biol* **30**: 223–253.
- Schoffstall AM. 1976. Prebiotic phosphorylation of nucleosides in formamide. *Origins Life* **7**: 399–412.
- Schöning K-U, Scholz P, Guntha S, Wu X, Krishnamurthy R, Eschenmoser A. 2000. Chemical etiology of nucleic acid structure: the α -threofuranosyl-(3' \rightarrow 2') oligonucleotide system. *Science* **290**: 1347–1351.
- Shapiro R. 1988. Prebiotic ribose synthesis: A critical analysis. *Origins Life Evol Biosphere* **18**: 71–85.
- Shaw AA, Shetlar MD. 1990. 3-Ureidoacrylonitriles: Novel products from the photoisomerization of cytosine, 5-methylcytosine, and related compounds. *J Am Chem Soc* **112**: 7736–7742.
- Smith JM, Sutherland JD. 2005. Aldolisation of bis(glycolaldehyde) phosphate and formaldehyde. *Chem Bio Chem* **6**: 1980–1982.
- Smith JM, Borsenberger V, Raftery J, Sutherland JD. 2004. Exploratory studies to investigate a linked prebiotic origin of RNA and coded peptides. 2nd Communication: Derivation and reactivity of xylose phosphates. *Chem Biodiversity* **1**: 1418–1451.
- Sutherland JD. 2007. Looking beyond the RNA structural neighborhood for potentially primordial genetic systems. *Angew Chem Int Ed* **46**: 2354–2356.
- Sutherland JD, Whitfield JN. 1997. Studies on a potentially prebiotic synthesis of RNA. *Tetrahedron* **53**: 11595–11626.
- Szostak JW. 2009. Systems chemistry on early Earth. *Nature* **459**: 171–172.
- Szostak JW, Bartel DP, Luisi PL. 2001. Synthesizing life. *Nature* **409**: 387–390.
- Tapiero CM, Nagyvary J. 1971. Prebiotic formation of cytidine nucleotides. *Nature* **231**: 42–43.
- Thaddeus P. 2006. The prebiotic molecules observed in the interstellar gas. *Phil Trans R Soc B* **361**: 1681–1687.
- Ueda T, Nishino H. 1968. On the Hilbert-Johnson procedure for pyrimidine nucleoside synthesis. *J Am Chem Soc* **90**: 1678–1679.
- Verlander MS, Lohrmann R, Orgel LE. 1973. Catalysts for the self-polymerization of adenosine cyclic 2',3'-phosphate. *J Mol Evol* **2**: 303–316.
- Woese C. 1967. *The genetic code*. Harper & Row, New York.
- Zhang L, Peritz A, Meggers E. 2005. A simple glycol nucleic acid. *J Am Chem Soc* **127**: 4174–4175.
- Zimmer S, Biltonen R. 1972. Thermodynamics of proton dissociation of adenine. *J Solution Chem* **1**: 291–298.