

# Prebiotic systems chemistry - Complexity overcoming clutter

Saidul Islam<sup>1</sup> and Matthew W. Powner<sup>1\*</sup>

<sup>1</sup>Department of Chemistry, University College London, 20 Gordon Street, London, WC1H 0AJ, UK

matthew.powner@ucl.ac.uk

## SUMMARY

Living organisms are the most complex chemical system known to exist, yet exploit only a small constellation of universally conserved metabolites to support indefinite evolution. The conserved chemical simplicity belying biological diversity strongly indicates a unified origin of life. Thus, the chemical relationship between metabolites suggests that a simple set of predisposed chemical reactions predicated the appearance of life on Earth. Conversely, if prebiotic chemistry produces highly complex mixtures, this then implies that the feasibility of elucidating life's origins is an insurmountable task. Prebiotic systems chemistry, however, has recently been exploiting the chemical links between different metabolites to provide unprecedented scope for exploration of the origins of life, and an exciting new perspective on a 4 billion-year-old problem. At the heart of the systems approach is an understanding that individual classes of metabolites cannot be considered in isolation. This review highlights several recent advances suggesting that the canonical nucleotides and proteinogenic amino acids are predisposed chemical structures.

## KEYWORDS

Origins of life, prebiotic chemistry, systems chemistry, predisposed chemistry, RNA, nucleotides, amino acids, sugars, metabolism, crystallization.

## BIGGER PICTURE

Advancing our understanding of the spontaneous emergence of life requires innovation across scientific disciplines as broad as astrophysics to phylogenetics, yet the primacy of chemistry cannot be overestimated. Cellular life is a chemical system of awe-inspiring complexity yet, perhaps surprisingly, life exploits only a

small constellation of universally conserved metabolites working in concert to support indefinite evolution. The conserved chemical simplicity that belies biodiversity is a strong indication that a simple set of predisposed reactions predicated the sudden appearance of life on Earth. The wonder of nature's greatest feat of invention—the self-assembly of living cells—positions the origins of life as one of the greatest challenges in chemistry. Building chemical systems that can self-assemble, process information, control the transport and accumulation of chemicals, orchestrate reaction pathways, and ultimately self-replicate will no doubt have a major impact on evolving technology, but nature has had a 4 billion-year head start in implementing controlled chemical evolution, and the lessons to be learnt from its prior art merely await discovery.

## **eTOC**

Prebiotic systems chemistry is providing unprecedented scope for exploring the origins of life and an exciting new perspective on a 4 billion-year-old problem. At the heart of this new systems approach is an understanding that individual classes of metabolites cannot be considered in isolation if the chemical origin of life on Earth is to be successfully elucidated. This review aims to highlight several recent advances that suggest the canonical nucleotides and proteinogenic amino acids are predisposed chemical structures.

## **INTRODUCTION**

Understanding the chemical processes that gave rise to primitive life on the early Earth is a fundamental problem that falls upon organic chemistry to resolve.<sup>1</sup> The small constellation of the core biological metabolites all share an atomic constitution of predominantly hydrogen, carbon, nitrogen, oxygen, phosphorus and sulfur, and these elements serve as the foundations of the triad of organic molecules—nucleic acids, lipids and proteins—that are thought to be immutably linked with sustained biological evolution.<sup>2-5</sup> There is an irrevocable link between the genetic code, the structure of the 20 proteinogenic amino acids, and the structure of the canonical nucleic acids (RNA and DNA) that is universal across all domains of life. The

persistent conservation of these core organic structures is highly indicative of a single origin of life on Earth.<sup>2-4,6-10</sup> However, distilling the complexity of modern cells to provide a reductionist explanation for life's chemical origins is one of the most challenging problems in science. Replication is clearly a key element of any biological system, and nucleotides appear ideally suited to this task.<sup>2,3,8,9</sup> However, biological nucleic acid replication is controlled by genetically encoded high-fidelity enzymes, and the self-assembly of such a complex and sophisticated system on the early Earth seems a remote possibility. The unexpected discovery of catalytic RNA molecules, or ribozymes, suggests a potential solution to this paradox. The capability of RNA to serve as a repository of genetic information *and* to perform catalytic functions provides a strong indication of the potential for an ancestral 'RNA world', a time period when self-replicating RNA molecules sustained biological information and also carried out the catalytic functions that are now mainly reserved for enzymes in extant life.<sup>2,7-9</sup> The critical role of RNA within the ribosome during genetically encoded peptide synthesis is the most recent and powerful piece of evidence of RNA's deeply ingrained origin in living systems, with crystallographic evidence illuminating the conspicuous absence of amino acids within as much as 18Å of the active site of peptidyl bond formation.<sup>11</sup> The predominance of RNA in the ribosome is seen as an indication that non-coding ribosomal RNA may be a vestigial remnant of an RNA-based metabolism, leading to the now oft-quoted phrase that the 'ribosome is a ribozyme' and perhaps a smoking gun for an 'RNA world'.<sup>12</sup>

Those who sympathize with elements of the RNA World attempt to recreate a scenario in which functional and replicating RNA may have self-assembled *de novo* on the early Earth, but demonstrating this is by no means a simple task. Synthetic organic chemistry overcomes the challenges imposed upon it through the development of exotic reaction methodologies to forge new chemical bonds, an array of orthogonal protecting group strategies and solvent changes to reduce deleterious side reactions, as well as increasingly sophisticated technologies to purify complex mixtures. These varied tools, technologies and a huge array of available chemical building blocks mean that the permutations of retrosynthetic analyses that can be applied to complex organic molecules are limited only by the imagination. Prebiotic chemistry, on the other hand, is not afforded such luxuries. The chemical process of abiogenesis had to occur under

stringent restrictions that a modern chemist undertaking routine synthesis in a laboratory would consider insufferable or archaic. The prebiotic chemical inventory would have been restricted to predominantly simple but highly reactive feedstock molecules such as nitriles, aldehydes, and acetylenes.<sup>13-15</sup> The concept of protecting groups in prebiotic chemistry is a relatively new and underutilized method of controlling site-selective modification, but this emerging strategy is being exploited to good effect and may have been an important way of controlling indiscriminate reactivity.<sup>16,17</sup> Physical processes such as crystallization, sublimation, thermophoresis and surface adsorption are examples of plausible means of purification of mixtures, and these may have played a critical role in the emergence of the canonical biological structures from potentially overwhelming mixtures of ‘prebiotic clutter’.<sup>8,9,18-20</sup>

A common perception (or misconception) of prebiotic chemistry is that the inherent reactivity of the simplest prebiotic feedstock molecules must tend towards uncontrollable reaction pathways that produce complex mixtures containing a multitude of undesirable products and intractable tars.<sup>8,9,18,21,22</sup> These deceptive perceptions have made it appear impossible to recapitulate the origins of life under primitive conditions, and accordingly led some to advocate structurally simpler pre-RNA or ‘XNA’ Worlds, where the basic functions of life may have been accomplished by non-biological or pre-biological systems (that is, before the existence of RNA).<sup>23-25</sup> Contemplating the likelihood of an ostensibly simpler transition from prebiotic chemistry to ‘XNA’ (XNA being the supposed transient non-biological informational forerunner to RNA) is understandable when taking into consideration the historical difficulties of prebiotic ribonucleotide assembly.<sup>8,9,26</sup> Broadening the scope of prebiotic chemistry beyond those components that exist in extant life reduces the specific constraints on synthesizing any of these individual components.<sup>25</sup> In other words, if one is allowed to retrospectively ‘move the goalposts’ and increase the number of targets by seeking (apparently) simpler alternatives to RNA, then it will surely also increase the chances of scoring a ‘goal’. This is certainly an intriguing and reasonable approach to take in hypothesizing about the first stages of life. If other non-biological systems are both prebiotically accessible and functionally capable then they are, of course, of fundamental interest. However, there are greater conceptual difficulties with any ‘XNA’ information transfer system that must subsequently invent RNA, rather than simply accessing RNA to begin

with. Just as RNA requires ribonucleotide building blocks, any XNA will also demand the self-assembly of its own monomers using prebiotic chemistry.<sup>26</sup> There is currently no evidence to suggest that the prebiotic chemistry leading to any specific classes of XNA would be afforded a simpler chemical pathway than RNA. However, if this chemical challenge is met, and a fully functioning and self-replicating XNA is produced *de novo*, one is then faced with the prospects of having to contend with a transition from a hypothetical XNA-based life to extant RNA-based life. This type of ‘genetic takeover’ requires a compatibility of XNA and RNA information transfer (through both high-fidelity base-pairing and maintained functional/catalytic prowess of XNA expressed phenotypes in progeny RNA systems) and, perhaps more importantly, the previously postponed synthesis of pools of RNA monomers is ultimately required to make this transition.<sup>26</sup> The feasibility of an XNA system inventing the “biosynthesis” of radically different monomers to support information transfer, under the constraints of competitive evolution, would seem highly improbable. Consequently, the demands of any pre-RNA scenario ostensibly create more problems than it resolves, and therefore it must be worthwhile exploring all the possible permutations of prebiotic RNA synthesis before yielding to the challenge and reconsidering RNA’s feasibility as the primordial ancestor. Notwithstanding its structural complexity, if RNA is a product of highly predisposed prebiotic chemistry then its chemical emergence and selection for a key role in nascent biological systems may have been by a direct route.

There is little doubt that the reductionist ‘RNA world’ hypothesis has been highly instructive in considering the origins of life. To some extent, however, understanding why RNA is the ideal candidate as the first information biopolymer of life has been hindered by approaches that have solely concentrated on realizing the prebiotic syntheses of a single class of metabolic macromolecule.<sup>4,5</sup> Replicating nucleic acids are not solitary actors within the biological systems. Cells require a number of chemical subsystems including peptides for functional support, lipids for compartmentalization, and numerous other small molecules that are part of the intricate network that forms the basis of metabolism. Cross-referencing the syntheses of these subsystems, rather than focusing on individual classes of metabolites, provides the real ‘smoking gun’ for a nucleic acid’s primacy. The implications of establishing a chemical relationship underlying biochemistry are manifold; overcoming the conceptual barriers that have prevented the

exploration of multiple fundamentally different classes of biomolecules arising within the same geochemical environment provides the best possible chance to observe emergent properties and the highest potential for sustained evolution.

This article serves as a focused analysis of recent findings to support the role of RNA at the origin of life, where systems chemistry is being employed to rediscover the mechanisms that underlie the earliest phases of evolution.<sup>4,5,10,27</sup> Recent results show that, despite employing ever more complicated mixed chemical systems, the canonical biological structures may be a product of a “natural selection” operating at the chemical level. Notably, these selection processes are not restricted solely to the emergence of RNA; credible generational links to other chemical subsystems, such as peptides and lipids, are now also being established. The work reviewed in the ensuing discussion also suggests that emergent physicochemical mechanisms may have guided chemistry through a labyrinth of possible routes to arrive at the vestiges of metabolism, proteinogenic amino acid, lipid, and RNA building blocks. These results collectively suggest that the primitive chemical pathway towards the canonical biological structures need not be the most obvious upon first inspection, but are nonetheless highly predisposed to yield the basic chemical constituents of cellular life.

## **THE TRADITIONAL DISCONNECTION OF RNA LEADS TO ‘IMPOSSIBLY DIFFICULT’ RIBONUCLEOTIDE ASSEMBLY**

### **The traditional disconnection of RNA**

The advent of any nucleotide-based biological system is contingent on the synthesis of the monomeric nucleotide building blocks in a primitive chemical environment, and this is no different for RNA.<sup>26</sup> The two likeliest candidate monomeric nucleotides are nucleoside-2',3'-cyclic phosphates **1** and activated nucleoside-5'-phosphates **2** (Figure 1). Nucleoside-2',3'-cyclic phosphates **1** are an attractive target because they are intrinsically activated towards oligomerization and yet possess a level of kinetic stability that prevents rapid hydrolysis in water to a mixture of nucleoside-2'- and 3'-phosphates. It is interesting that although nucleoside-5'-phosphates **2** are currently exploited by extant life, RNA polymers most readily hydrolyzes via

nucleoside-2',3'-cyclic phosphates **1** by virtue of intramolecular phosphodiester transfer, suggesting that simple chemical strategies for recycling RNA would most readily exploit **1** rather than **2**. Ribonucleotides are stereochemically and functionally complex organic molecules and to produce these thermodynamically unstable compounds in a completely achiral environment is a formidable challenge.

The difficulties associated with prebiotic syntheses of activated ribonucleotides—either **1** or **2**—have precluded RNA from gaining universal acceptance as the first biopolymer of life, but these synthetic problems are in large part a consequence of the chemical disconnection shown in Figure 1. Although the oligomerization of RNA involves iterative phosphodiester ligations that can be chemically coordinated to transfer information between a preexisting and nascent polymer, the assumed organic chemistry required to build the stereo- and regio-chemically complex ribonucleotide monomers is far more complicated. Monomer assembly appears at first to require an assortment of fundamentally different types of chemistries to build the constituent nucleobase (**3-6**) and enantiopure D-ribose (D-**7**; shown in the pentofuranose form) fragments, and then further distinct chemistries to permit their regio- and stereochemical fusion to form the furanosyl  $\beta$ -ribonucleoside **8** (Figure 1). Additional different chemical strategies are then required to regioselectively phosphorylate **8** in preparation for the ribonucleotide product (**1** or **2**) to generate an RNA oligomer by polycondensation. On top of all this, ribose **7** and the nucleobases **3-6** need to be built by the prebiotic chemistry of small feedstock molecules.<sup>13-15</sup> Prebiotic chemistries seeking to elucidate a mechanism to achieve this ‘traditional’ approach to assembling ribonucleotides is discussed in this first section of the tutorial review. This will put into context the challenges that one is faced with when trying to demonstrate the emergence of canonical RNA from complex prebiotic mixtures, and similar arguments can also be applied to the assembly of proteinogenic amino acids or any other class of essential metabolite under prebiotic constraints.

### **The ‘three pillars’ of prebiotic chemistry**

Many have become conditioned into accepting the ‘three pillars’ of prebiotic chemistry—sugar synthesis by the formose reaction, nucleobase synthesis by hydrogen cyanide **9** oligomerization, and the amino acid syntheses by the spark discharge experiment—as probable chemical routes to the building blocks

of core biological metabolites, such as the nucleic acids and polypeptides (Figure 2). These shall only be touched upon in this article as they have been comprehensively reviewed elsewhere.<sup>7</sup> The Miller-Urey spark discharge experiment was arguably the first truly prebiotic experiment that had significant implications for the theory of abiogenesis. This now infamous reaction yielded complex mixtures that contain some of the building blocks of proteins, amino acids, from simple gas mixtures supposed to simulate an early reducing atmosphere on Earth.<sup>28</sup>

Although the detailed mechanism for the formation of a milieu of products still remains sketchy, it is likely that aminonitriles **10** are presented by way of Strecker reactions of cyanide **9**, ammonia, and carbonyl compounds **11** formed *in situ* by a spark discharge passed through a mixture of methane, water, ammonia and hydrogen (Figure 2A). Advances in analytical techniques have allowed deeper analyses of the samples generated by spark discharge experiments. These continue to throw up new insights into their vast compositional complexity but cast doubt on their practical utility to resolve the problem of synthesizing proteinogenic peptides. In all cases, the formation of an inseparable mixture of proteinogenic and non-proteinogenic amino acids is observed alongside a huge range of other organics.<sup>29,30</sup> A similar situation arises with other exogenous amino acid samples. Meteorites, for example, have been observed to contain mixtures of proteinogenic and non-proteinogenic amino acids, but there is little evidence to suggest these samples inherently favour the former over the latter.<sup>28-31</sup> The similarity of the wide distribution of proteinogenic and non-proteinogenic amino acid compositions found in meteoritic samples and spark discharge experiments indicates that the mechanistic pathway to amino acid formation may bear resemblances. However, both lack the required precision to furnish proteinogenic amino acids.<sup>28-31</sup> Beginning from such a complex starting point, it is not clear how cellular life could have selected the few essential amino acids that are expressed in proteins.

The core structure of the canonical purines is constitutionally derived from an oligomer of hydrogen cyanide **9**, where the adenine **3** moiety of RNA is assumed to be a pentamer of **9** (Figure 2B). The sugar fragment of RNA, ribose **7**, is commonly assumed to be a product of oxygenous chemistry and a pentamer product of formaldehyde **12** (Figure 2B). Hydrogen cyanide **9** and formaldehyde **12** are presumed products



of the spark-discharge experiment that produce the aminonitriles **10** but they are also potential gateway feedstock molecules towards the nucleic acids. However, spark discharge conditions do not produce ribonucleotides *per se*, and so conditions for the chemical origins of amino acids and ribonucleotides are assumed to be fundamentally different. The disconnection of the ribonucleotides to nucleobases **3-6** and ribose **7** is an ostensibly simple one (Figure 1), but it immediately implies a need to separate the pH-dependent oligomerization of cyanide **9** and formaldehyde **12**. Simultaneous, but independent, oligomerizations of **9** and **12** within the same solution are prevented by glycolonitrile formation (Figure 2A, R=R<sub>1</sub>=H; **13**). Mixing stoichiometric cyanide **9** and formaldehyde **12** leads to rapid cyanohydrin formation.<sup>32,33</sup> Glycolonitrile **13** formation is highly favourable and though (in principle) reversible, the odds are stacked heavily against mutually compatible and independent oligomerizations of **9** and **12** to nucleobases and sugars within the same environment, with the equilibrium position overwhelmingly in favour of cyanohydrin formation ( $K_{eq} = 10^6 \text{ M}^{-1}$ ). Glycolonitrile **13** readily polymerizes, but not to give the sugars or nucleobases that are required to build ribonucleotides.<sup>34</sup> If there were some means of overcoming this vastly superior equilibrium position such that formaldehyde **12** and cyanide **9** oligomerization could proceed, one is then presented with the added complication of cyanohydrin self-destruction of C<sub>3</sub>, C<sub>4</sub>, C<sub>5</sub> and higher sugars under the high pH regimes required for nucleobase and sugar syntheses. Rapid imidolactone formation occurs by the intramolecular (*5-exo-dig*) hydroxyl addition to the nitrile carbon atom of the cyanohydrin that forms from the parent sugar, and this leads to its eventual hydrolysis to homologous acids rather than sugars.<sup>35,36</sup> The ramifications of this are the (apparent) need for segregation of nucleobases and sugars syntheses. Then, assuming this separation is geochemically possible, exclusive nitrogenous (cyanide **9**) or oxygenous (formaldehyde **12**) oligomerization does not allow effective formation of nucleobases or pentose sugars with the selectivity or the robustness to be considered as viable processes to yield either ribose **7** or nucleobases **3-6** (or any other non-biological nucleobase analogue) required for nucleotide assembly. The oligomerization of cyanide **9** rapidly deteriorates into an intractable mixture with most of the input cyanide **9** forming insoluble polymers (Figure 2B, blue). For example, aqueous solutions of ammonium cyanide at pH 9.2 lead only to minuscule quantities of adenine **3** amongst

many other species.<sup>18,37,38</sup> A representative sample of just a few of the products of hydrogen cyanide **9** oligomerizations is shown in Figure 2B. These processes are anything but robust, with adenine **3** usually being produced in less than 1% yield.

Compounding the difficulties of selectively synthesizing nucleobases, the origin of the sugar adds another layer of complexity to the ribonucleotide question. Lingering doubts surrounding the provenance of ribose **7** have plagued prebiotic chemistry since its inception, with no apparently plausible solutions within reach.<sup>33</sup> The formose reaction is the direct oligomerization of formaldehyde **12** and has often been assumed to provide the ribose **7** moiety of RNA (Figure 2B, red). The exact details of how this reaction initiates are still not clear, but an initial homocoupling of formaldehyde **12** is thought to produce glycolaldehyde **14**. This process requires a polarity reversal ('umpolung') of formaldehyde **12** such that it can act as a nucleophile.<sup>4</sup> The reality of the formose reaction is that, much like cyanide **9** oligomerization, it descends into an inextricable mixture. The vast array of sugars produced is overwhelming and the intrinsic lack of selectivity for ribose **7** is its undoing. Ultimately, the formose reaction produces a disastrously complex mixture of linear and branched aldo- and ketosugars in the racemic forms.<sup>39</sup> Figure 2B (red) highlights this complexity; aldo- and keto- sugars are shown as the open chain forms only, but all of these exist in equilibrium with their hydrates, and C<sub>4</sub> and higher sugars cyclise to yield furanosyl or pyranosyl mixtures (Figure 2B, red). The consequences of such uncontrolled reactivity is that ribose **7** is formed in less than 1% yield amongst a plethora of isomers and homologues.<sup>33</sup> There have been significant efforts to optimize sugar synthesis, but none have ever given a convincing synthesis of ribose **7**, nor has the selectivity been streamlined to a point where the mixture is of synthetic utility.<sup>40-42</sup> The stability of ribose **7** is of even greater concern given that it is highly unstable under the alkaline conditions required for the formose reaction ( $t_{1/2} = 5$  h, pH = 12.5).<sup>21,43</sup> The instability of ribose **7** prevents its accumulation and requires it undergo extremely rapid onward conversion to ribonucleosides **8** before the free sugar is lost to rapid degradation.<sup>21,33,43</sup> This appears highly unlikely given the difficulties observed in combining nucleobase and sugar syntheses that have already been highlighted.<sup>32-35,38</sup> Furthermore, ribose **7** readily isomerizes to yield a mixture of pentose sugars even when it

is incubated under mild aqueous conditions (pH 7, 25 °C), where arabinose **15** (75%) actually predominates over ribose **7** (19%).<sup>33</sup>

The stabilization of sugars by borate complexation has been known for many years,<sup>44</sup> and this has been proposed as a way of guiding the synthesis and enhancing the stability of ribose **7** under the conditions of the formose reaction.<sup>45,46</sup> By exploiting the known propensity for ribose **7** to form borate esters, it is hypothesized that ribose **7** would persist under the harsh conditions of its prebiotic formation whilst other sugars would degrade. Ribose-borate complexes do indeed exhibit greater stability relative to other aldopentoses, but careful inspection of decomposition rates suggests that the stabilization offered to ribose **7** under formose conditions is modest at best ( $t_{1/2} = 45$  h, pH = 12.5).<sup>21,47</sup> Interestingly, ribulose **16**—the ketopentose sugar—is actually provided with a greater degree of enhanced stabilization by borate than ribose **7**.<sup>21</sup> Moreover, a major pathway of consumption of formaldehyde **12** in formose-type reactions is its disproportionation into formate and methanol and this is not ameliorated in borate-mediated reactions of formaldehyde **12**. Indeed, a greater excess of glycolaldehyde **14** over formaldehyde **12** is required to overcome borate inhibition, which still produces many sugar isomers with no intrinsic selectivity for ribose **7**. A complex pathway has been recently been invoked for prebiotic ribose **7** synthesis, but the need for a low pH (pH = 5.9) molybdic acid mediated isomerization of stalled branched intermediates produced by the high pH (pH = 12.5) borate-mediated aldol chemistry makes such a synthesis less convincing, and a closed reaction cycle has yet to be convincingly demonstrated.<sup>21</sup>

A phosphorylated variant of the aldose condensation has been shown by Eschenmoser and colleagues to be a potent method of suppressing the complexity of the formose reaction and the destructive effects that the high pH reaction conditions have on the sugar products.<sup>48</sup> Glycolaldehyde phosphate **14-P**, the product of  $\alpha$ -phosphorylation of glycolaldehyde **14** (by the action of amidotriphosphate **17** or diamidophosphate **18**),<sup>49,50</sup> can undergo homoaldol reactions to produce a mixture of tetrose-2,4-diphosphates and hexose-2,4,6-triphosphates **19**, with **19** being the overwhelmingly predominant compound (Figure 3). Glycolaldehyde phosphate **14-P** also reacts with formaldehyde **12** to give glyceraldehyde-2-phosphate **20-2P**, which underwent a diastereoselective crossed aldol condensation with glycolaldehyde

phosphate **14-P** to give predominantly an isomeric mixture of pentose-2,4-diphosphates **21**, in which *ribo-21* is the major product under kinetic control.<sup>48</sup> Whereas the formose reaction succumbs to the uncontrollable formation of tars, these sugar phosphates exhibit far greater stability than the non-phosphorylated parent sugars, the latter being eventually destroyed by the highly basic reaction conditions required for their initial formation.<sup>21,33,39,40</sup> Nevertheless, the diastereoselectivity is not absolute and the palette of sugar phosphate products is still complex due to the number of isomeric sugar phosphates produced (Figure 3).

Although the pentose sugar diphosphates **21** do share some constitutional similarity to the backbone of RNA, they do not possess the natural phosphorylation pattern or phosphorylation level to make a simple transition to the natural sugar phosphate backbone of RNA. Moreover, phosphorylation of the C4-hydroxyl moiety specifically prohibits the formation of a furanosyl ring, and the potential for nucleobase ribosylation with these sugar-phosphates remains unknown. This inherent divergence away from natural nucleotide structure is problematic if life's origins are viewed only through the lens of an attempted prebiotic nucleotide synthesis. Opening this remarkable example of prebiotic chemistry up to a systems chemical analysis however suggests that rather than serve as a feedstock to generate (non-natural) pentose sugar diphosphates **21**, glycolaldehyde phosphate **14-P** and glyceraldehyde-2-phosphate **20-2P** may have experienced an alternate fate and become implicated in a robust primitive triose glycolysis pathway,<sup>51</sup> a point which will be elaborated upon in the second part of this review.

Returning to sugar synthesis, the complexity of the formose reaction mixtures and the instability of ribose **7** under these conditions have led some to explore fundamentally different pathways to sugar syntheses, but a selective pathway to ribose **7** still remains elusive. Recent examples include the detection of pentose sugars upon photochemical and thermal action on model cosmic ice analogues.<sup>52</sup> These photochemically processed interstellar ice analogues—comprised of water, methanol and ammonia—were found to contain small amounts of various C<sub>2</sub> - C<sub>6</sub> aldoses (including 845 ppm aldopentoses of which 260 ppm was ribose **7**) along with larger quantities of ketose sugars (including 2010 ppm ribulose **16**) and various polyols and polyhydroxyacids. In another example, dihydroxyacetone **22** and ketose sugars, such as ribulose **16** and xylulose **23**, are formed in extremely high yield by reactions of lithium salts of

dihydroxyfumarate with glyoxylate and glyceraldehyde **20**.<sup>53</sup> These studies were based on the “glyoxylate scenario”, an interesting postulate in which the source of carbohydrates and other chemical components of potentially complex cyclic metabolic organization are proposed to be the products of HCN oligomerization.<sup>54</sup>

### Attempted syntheses of ribonucleosides from sugars and nucleobases

Assuming that some remarkable (but as yet unknown) prebiotic chemistry could generate ribose **7** and nucleobases with sufficient purity for onward reactivity, one then needs to contend with the prospects of uniting these to form  $\beta$ -ribonucleosides **8**. Consider first the laboratory-based chemical syntheses of ribonucleotides where most, if not all, require manipulation of sugars and nucleobases with protecting group strategies to overcome the thermodynamic and kinetic pitfalls that mitigate their fusion. Examples of highly sophisticated chemical interventions for the controlled formation of glycosidic bonds are legion,<sup>55</sup> but in the absence of such modern synthetic chemical ingenuity, it is not at all surprising that the direct reaction of ribose **7** and the canonical nucleobases (**3-6**) is an exceedingly poor reaction. The formation of a  $\beta$ -ribonucleoside **8** from the condensation of ribose **7** and the canonical purines **3** and **4** is low yielding, and generates an inseparable mixture of regioisomers in both the furanose and pyranose forms.<sup>9,10,56</sup>

*N*-Formamidopyrimidines were recently shown to act as masked purine nucleobases to generate purine ribonucleosides with absolute nucleobase regioselectivity by exploiting the fascinating symmetry of the parent heterocycle.<sup>57</sup> This approach is intriguing, but the benefits of *N*<sup>9</sup>-purination are offset by the formation of an inseparable mixture of  $\alpha/\beta$ -furanoside and  $\alpha/\beta$ -pyranoside nucleosides (even from a large excess of pure ribose **7** starting material) and poor selectivity for the natural  $\beta$ -furanosyl-ribonucleosides (**8**).<sup>58</sup> Furthermore, coupling formose-type sugar mixtures with *N*-formamidopyrimidines yield only analytically detectable amounts of the desired purine **8** (B = **3**, <0.2%) interspersed amongst a multitude of other isomers and homologues.<sup>57</sup> Whether a similar strategy can be applied to the formation of pyrimidine ribonucleotides is currently unknown. Ideally, both purines and pyrimidine ribonucleosides (**8**, B = **3-6**) would emerge from similar chemistries, but this is not the case as the canonical pyrimidines (cytosine **5** and

uracil **6**) are resistant to ribosylation, and their lack of reactivity can be ascribed to the lack of electron lone pair availability on the  $N^1$ -nitrogen atom.<sup>56</sup>

## Concluding remarks

In concluding the first part of this review it is worth highlighting that these preceding examples illustrate that, for many years, prebiotic chemistry has been geared towards the synthesis of the biologically relevant structures through reactions that are generally non-selective and produce complex mixtures. Further downstream processing of these product mixtures leads to significant problems because the already diminutive yields of each required constituents become progressively smaller after each step in a sequence towards the target. These synthetic challenges are exacerbated by an inordinate number of isomers and by-products that are generated by each iterative step, more so because plausible strategies to purify these mixtures have yet to be convincingly demonstrated. When considered as a complete synthesis, these observations especially underscore the insurmountable difficulties with the nucleobase + ribose **7** strategies (Figure 1), and this suggests that alternative chemical disconnections of RNA need to be considered.

## PREBIOTIC SYSTEMS CHEMISTRY AT THE ORIGIN OF LIFE

### Prebiotic systems chemistry

Prebiotic systems chemistry explores the possibilities of recreating the complex and emergent phenomena that are a hallmark of living systems. Cellular life operates through an integrated network of compounds and chemical pathways, and the possibility of mimicking this to create a self-replicating living system from the ‘bottom-up’ is a tantalising prospect. In essence, this constitutes a total chemical synthesis of a (minimal) cell,<sup>2</sup> and this arguably “*requires the greatest retrosynthetic analysis of all time*”.<sup>27</sup> The interactions of multiple chemical entities in concert, rather than in isolation, may actually curtail the combinatorial explosion of unwanted or undesirable compounds that has become synonymous with prebiotic chemistry. Seemingly fewer reaction pathways could be the product of suitable catalysts or ‘molecular chaperones’ that may be formed *in situ* and these may direct reactions towards a more specific reaction pathway. The possibility of self-purification may present itself as a result of a physical phenomenon inherent

to specific compounds and chemical environments, and combinations of such processes could, in principle, accumulate substrates of sufficient purity for onward reactivity to biologically relevant structures. The formation of complex mixtures does not necessarily mean a 'dead end', providing that there is a means of purifying or utilizing the required constituents in a prebiotically plausible manner. The following discussions provide some instructive examples of application of these strategies.

### **Synthesis of RNA, lipid, and protein precursors by cyanosulfidic protometabolism**

The origins of nucleic acids and peptides had long been considered as mutually exclusive due to the discordant reaction conditions that have been proposed for their respective formation, thus rendering their complementary syntheses incompatible.<sup>32-35</sup> However, this is now no longer the case and recent examples of prebiotic systems chemistry provide evidence to the contrary. In the first example, Sutherland and colleagues delineated a comprehensive reaction network that produces the building blocks of RNA, proteins and lipids. All of these emanate from hydrogen cyanide **9** in a remarkable set of processes that has been termed a 'cyanosulfidic protometabolism'.<sup>4,15,59,60</sup>

A mixed chemical system composed of bioessential elements (H, C, N, O, P and S) can simultaneously generate C<sub>2</sub> and C<sub>3</sub> sugars that are crucial for prebiotic ribonucleotide synthesis,<sup>61</sup> the hydrophilic moiety of glycerophospholipid membranes, and twelve proteinogenic amino acid precursors (Figure 4).<sup>15</sup> The driving force for this reaction network is a photochemical Kiliani-Fischer homologation of hydrogen cyanide **9**, with hydrogen sulfide as the reductant, and where some of these reactions are accelerated by Cu(I)-Cu(II) photoredox cycling. The network is initiated by a reduction of cyanide **9** to produce formaldehyde **12**, which is rapidly intercepted by cyanide **9** to produce the simplest cyanohydrin, glycolonitrile **13**. Whereas the formation of glycolonitrile **13** is a major pitfall in the traditional (formose/HCN-oligomerization) approach to prebiotic sugar and nucleobase synthesis,<sup>32-34</sup> Sutherland and colleagues use the incredible efficiency of cyanohydrin formation to their advantage. They exploit this reaction to fix carbon as a non-volatile and stable form, suppressing the detrimental disproportionation of formaldehyde **12**, which is a large contributor to its losses during the formose-type reactions.<sup>21,33,39</sup> Glycolonitrile **13** undergoes reduction to give glycolaldehyde **14** and thereby achieving an umpoled

15

formaldehyde-dimerization using cyanide as an umpolung formaldehyde-synthon. Homologation can then be reiterated to produce glyceraldehyde **20** by reduction of glyceronitrile **24** with high efficiency.<sup>59,60</sup> These C<sub>2</sub> and C<sub>3</sub> sugars serve as the only required sugars for the synthesis of pyrimidine ribonucleotides, a point that will be returned to later.<sup>61</sup> The formation and persistence of glyceraldehyde **20** is highly significant because, as previously mentioned, the corresponding cyanohydrin of glyceraldehyde **20** can undergo self-destruction by rapid imidolactone formation at elevated pH.<sup>35,36</sup> However, the cyanohydrin of glyceraldehyde **20** is stable under the mild (pH 7, phosphate buffered) reaction conditions employed in Sutherland's Kiliani-Fischer homologation.<sup>60</sup> Glyceraldehyde **20** synthesis by cyanide capture and photochemical reduction bears interesting similarities to the strategy used in the Calvin cycle, where enzyme-catalyzed CO<sub>2</sub> capture and reduction produces glyceraldehyde-3-phosphate **20-3P**.<sup>62</sup> Furthermore, exploiting hydrogen cyanide **9** as an environmental source of carbon has the additional advantage of nitrogen fixation as a valuable by-product of sugar synthesis.<sup>15,59,60</sup> Fixed nitrogen, in the form of ammonia, is released into solution during photochemical cyanide/cyanohydrin reduction to alleviate concerns over the prebiotic reduced nitrogen inventory.<sup>63</sup>

Sutherland recognised that cyanohydrins **13** and **24** could undergo a divergent fate and be converted to amino acid precursors as well as C<sub>2</sub> and C<sub>3</sub> sugars. Formaldehyde **12** and glycolaldehyde **14** are the Strecker aldehydes of glycine (**Gly**) and serine (**Ser**), and their cyanohydrins **13** and **24** are readily converted to the corresponding aminonitriles in the presence of ammonia.<sup>60</sup> Remarkably, glycolaldehyde **14** also provides the carbon framework of alanine (**Ala**) because glycolaldehyde **14** undergoes photochemical  $\alpha$ -deoxygenation to produce acetaldehyde **25**. The cyanohydrin of acetaldehyde **25** also undergoes reductive homologation to produce lactaldehyde, which serves as the building block for threonine (**Thr**) (Figure 4).<sup>60</sup> Dihydroxyacetone **22**, produced from phosphate-catalyzed interconversion with glyceraldehyde **20**, can undergo two sequential  $\alpha$ -deoxygenation to yield acetone **26**, and this provided the branched backbone of valine (**Val**) and leucine (**Leu**).<sup>15</sup> These deoxygenation reactions are particularly noteworthy due to the levels of sophistication required in achieving this chemical transformation even in the realms of synthetic



organic chemistry, where similar transformations generally require multistep manipulations of substrates with protecting and/or leaving groups prior to reduction.<sup>64</sup>

Further reaction pathways involving copper mediated cross-couplings of cyanide **9** and acetylene **27** generated the building blocks for asparagine (**Asn**), aspartate (**Asp**), glutamine (**Gln**), glutamate (**Glu**), proline (**Pro**) and arginine (**Arg**), completing a remarkable collection of twelve different proteinogenic amino acids (Figure 4).<sup>15</sup> To further corroborate this system's relationship to biochemistry, the formation of lipid precursors in the close proximity of ribonucleotide and amino acid precursors is striking. The reduction of dihydroxyacetone **22** to glycerol **28** and subsequent conversion to glycerol phosphates provides a direct connection to phospholipids.<sup>15</sup> Although the hydrophobic chains and the stereochemistry of phospholipids differ, the glycerol phosphate head group is conserved across amphiphiles in all domains of life.<sup>6</sup> What is immediately noticeable in Sutherland's reaction network is the relatively narrow product distribution: the runaway reactivity and the highly complex mixtures produced by the spark discharge experiments,<sup>28-30</sup> the formose reactions,<sup>33,39,40,52</sup> the borate-mediated aldol condensations,<sup>21,46</sup> and hydrogen cyanide **9** oligomerization<sup>18,37,38</sup> are simply not observed.

### **Bypassing ribose and nucleobases: Synthesis of activated ribonucleotides**

The recent syntheses of activated pyrimidine ribonucleotides  $\beta$ -cytidine-2',3'-cyclic phosphate **29** and  $\beta$ -uridine-2',3'-cyclic phosphate **30** under plausible prebiotic conditions are key working examples of systems chemistry implementation at the origin of life.<sup>61,65,66</sup> These syntheses are a complementary continuation to the production of the C<sub>2</sub> and C<sub>3</sub> sugars produced by the cyanosulfidic protometabolism.<sup>15,59,60</sup> The short reaction sequences that produces **29** and **30** overcome a number of major obstacles that had previously precluded their synthesis by invoking mixed chemical systems from the outset (Figures 5 and 6). Mixed oxygenous and nitrogenous systems allowed the construction of the sugar and nucleobase in a single hybrid scaffold, and a crucial component to the success of this strategy is the remarkable and multifaceted role of phosphate and its conspicuous presence from the very beginning of these syntheses. Phosphate participates in various capacities throughout the sequence by acting as a general acid-base catalyst, as a chemical and pH buffer, and then finally as an essential reagent.<sup>61,65-67</sup>

A pivotal disconnection resolved two critical problems in one single step.<sup>68</sup> Whereas the traditional retrosynthetic analysis of ribonucleotides places a demand on an endergonic C-N bond forming event between the elusive ribose sugar **7** and a preformed purine (adenine **3** or guanine **4**) or pyrimidine (cytosine **5** or uracil **6**) nucleobase, the formation of the intermediate pentose aminooxazolines **31** only requires the C<sub>2</sub> and C<sub>3</sub> sugars glycolaldehyde **14** and glyceraldehyde **20** (Figure 5A) and is a thermodynamically favourable process. The assembly of this crucial building block, which contains the pentofuranosyl backbone and a concealed glycosidic bond, is by virtue of a masked aldol reaction between 2-aminooxazole **32** and glyceraldehyde **20**.<sup>61,68</sup> 2-Aminooxazole **32** is formally the condensation product of glycolaldehyde **14** and cyanamide **33**, but incubation of **14** and **33** in water starting at pH 7 produces a multifarious reaction mixture, with only small amounts of 2-aminooxazole **32** actually present.<sup>61,69</sup> The addition of phosphate results in a dramatic change of circumstances—not only does phosphate buffer the reaction at pH 7 but it also acts a proficient general acid-base catalyst operating at its second pK<sub>a</sub>—resulting in the clean formation of 2-aminooxazole **32** in very high yield, with complex by-product formation completely suppressed.<sup>61</sup> 2-Aminooxazole **32** can, therefore, be considered as truly prebiotic—a product of the simplest aldose glycolaldehyde **14** and cyanamide **33** (Figure 5A).

The first chiral aldose in the sugar series, glyceraldehyde **20**, is then added to the crude mixture leading to the formation of pentose aminooxazoline **31** in nearly quantitative yield. The reaction exhibits pronounced diastereoselectivity for *ribo*-**31** and *arabino*-**31**, and these make up 74% of the total composition of **31**.<sup>61</sup> The significance of this reaction cannot be overestimated: the free pentose sugars are avoided completely, and yet the five-carbon pentose moiety of RNA is synthesized with complete furanosyl selectivity along with the concomitant installation of a carbon-nitrogen bond that is predestined to become the crucial glycosidic bond. This is all achieved at neutral pH and in one single, high yielding reaction (Figure 5A).<sup>61,68</sup> The differential solubility of the pentose aminooxazolines **31** results in purification by fortuitous crystallization of pure *ribo*-**31** from a mixture of the four other pentose isomers, leaving *arabino*-**31** as the major aminooxazoline in solution and pure *ribo*-**31** in the solid state.<sup>68</sup> Furthermore, the crystallization of pure homochiral D-*ribo*-**31** occurred spontaneously when scalemic glyceraldehyde **20** with

an enantiomeric excess (*ee*) of 60% was used in the reaction with 2-aminooxazole **32**.<sup>68</sup> This remarkable observation has significant implications for the origins of homochirality of ribonucleotides. With the natural biological chirality of nucleic acids now installed into *D-ribo-31*, any further manipulation of this building block means that stereochemistry should remain conserved. This important point will be returned to later on during a discussion on the chiral resolution of ribonucleotides. Phosphate also interconverts *D-ribo-31* and *D-arabino-31* through general acid-base catalyzed C2'-epimerization, allowing material transfer between *D-arabino-31* and *D-ribo-31*,<sup>67</sup> both of which are precursors to activated pyrimidine ribonucleotides **29** and **30**.<sup>61,66</sup> The remaining pentose aminooxazolines, *xylo-31* and *lyxo-31*, are only minor products. Furthermore, *xylo-31* and *lyxo-31* are not stereochemically related to *ribo-31* by a simple interconversion mechanism like *arabino-31*, and therefore can be readily and permanently removed from isomeric mixtures of **31** by crystallization of *ribo-31*. Thus *xylo-31* and *lyxo-31* are not considered to be substrates for conversion to **29** and **30**.

Ribose aminooxazoline, *ribo-31*, reacts with cyanoacetylene **34** in water to produce  $\alpha$ -*ribo*-cytidine  $\alpha$ -**35**, but the yields of this reaction are invariably low (Figure 6A).<sup>70</sup> A significant excess of cyanoacetylene **34** is needed for complete conversion of *ribo-31* to  $\alpha$ -**35**, and it emerged that the rising pH observed during the course of this reaction causes the intermediate *ribo-36* to hydrolyze.<sup>61</sup> This pH rise also initiates high levels of over-cyanovinylation to generate a highly complex and intractable mixture of cytidine-like products.<sup>61,70</sup> Phosphate resolves all of these problems by acting as a chemical buffer to react with excess cyanoacetylene **34**, and as a pH buffer to mitigate hydrolysis. The reward for the introduction of phosphate to tame the reactivity of **34** is the production of anhydronucleosides **36** in quantitative yield without its decomposition during the reaction.

The clean and robust formation of *ribo-36* and *arabino-36*, which up to this point had previously been inaccessible by prebiotic chemistry, was highly significant and proved to be an essential missing element required to realizing two very distinct pathways to activated ribonucleotides, but both of which eventually converge to one product. Remarkably, both *arabino-36* and *ribo-36* can be manipulated to provide high yields of canonical pyrimidine nucleotides **29** and **30**, and these two pathways together are a

testament to the robustness of this strategy for ribonucleotide synthesis, demonstrating a predisposed generational degeneracy in these RNA building blocks.<sup>61,66</sup>

The conversion of *arabino-36* to **29** by chemoselective C3'-OH phosphorylation is as much noteworthy as it is counterintuitive because empirical predictions based on simple consideration of steric hindrance would imply that the primary C5'-OH is the most reactive alcohol (Figure 5B). Thankfully, such simple predictions proved to be misguided and crystallographic evidence and computational studies revealed that regioselective 3'-OH phosphorylation is predisposed by the structure of *arabino-36*.<sup>61,71</sup> Regioselective phosphorylation of *arabino-36* is governed by an intramolecular stereoelectronic effect that reduces reactivity of the C5'-OH by an n- $\pi^*$  interaction with the antibonding orbital on C2 carbon atom of the anhydronucleobase.<sup>71</sup> Urea ((NH<sub>2</sub>)<sub>2</sub>CO), the product of phosphate-catalyzed hydration of cyanamide **33**, manifests itself as a nucleophilic catalyst at elevated temperatures to activate inorganic phosphate within a viscous melt (upon evaporation of water), and mediate phosphoryl transfer from inorganic phosphate to *arabino-36* producing **29** as the main nucleotide product (Figure 5B).<sup>61</sup>

In the final step, UV irradiation was found to provide a plausible mechanism to destroy the remaining by-products from the synthesis of **29** and partially convert **29** to **30**. The specific structural and conformational attributes of **29** bestow it with a protective mechanism not shared by the other minor isomeric pyrimidine nucleoside and nucleotide by-products leftover over from the synthesis of **29**. The conformation adopted by **29** permits the C5'-OH to sit in close proximity to the C5 carbon atom of the cytosine nucleobase and promotes an intramolecular cyclization to form a 5,5'-bicyclic structure. This allows the nucleoside-2',3'-cyclic phosphates **29** and **30** to be impervious to numerous mechanisms of UV-degradation that rapidly destroy closely related isomers. Therefore, irradiation of **29** over three days led to partial conversion to **30**, with all other by-products converted to non-nucleotide species. This leaves **29** and **30** free to undergo oligomerization without interference or inhibition by other isomeric species, and this final photochemical sanitization step concludes the first short and robust pathway to pyrimidine ribonucleotides **29** and **30**.<sup>61</sup>

A second route to **29** by  $\alpha \rightarrow \beta$  photoanomerization was very recently established, providing further validation of the importance of anhydronucleosides **36** in the synthesis of the canonical ribonucleotides. Incubation of *ribo*-**36** above pH 7 leads to clean hydrolysis to produce  $\alpha$ -*ribo*-cytidine  $\alpha$ -**35** in near quantitative yield (Figure 6A). Whilst  $\alpha$ -**35** is tantalisingly only one stereochemical inversion away from natural pyrimidine ribonucleoside  $\beta$ -**35**, attempts to photochemically epimerize the glycosidic bond of  $\alpha$ -**35** result in low yields of the photoanomerization product upon irradiation with UV-light.<sup>70</sup> The underlying cause of this low yield is a combination of nucleobase loss and oxazolidinone **37** formation outcompeting stereochemical inversion of  $\alpha$ -**35** (Figure 6A).<sup>72</sup>

The formation of oxazolidinone **37** is the result of intramolecular reaction of the C2'-OH moiety of  $\alpha$ -**35** with its photoexcited nucleobase. This was corroborated by subsequent investigation of  $\alpha$ -cytidine-2-phosphate  $\alpha$ -**35**-2P, in which the detrimental effect of C2'-OH was blocked by phosphorylation.<sup>73</sup> Photoanomerization was improved 10-fold upon C2' phosphorylation, but a selective prebiotic pathway to  $\alpha$ -cytidine-2-phosphate  $\alpha$ -**35**-2P has not yet been demonstrated, with known pathways yielding mixtures of 2' and 3' phosphates (Figure 6A).<sup>73</sup> Sutherland and colleagues, however, recently found a solution to the inefficient photoanomerization of  $\alpha$ -**35** without having to find a means to block the detrimental effect of the C2'-OH. Sutherland's second route proceeds via *ribo*-**36**, thus accruing an even greater amount of pyrimidine ribonucleotides **29** and **30** by allowing exploitation of both *ribo*-**31** and *arabino*-**31** en route to nucleotides.<sup>61,66</sup> The reaction of *ribo*-**36** with hydrosulfide (HS<sup>-</sup>) yields the corresponding  $\alpha$ -thioribocytidine  $\alpha$ -**38** (Figure 6B). Whereas  $\alpha$ -ribocytidine  $\alpha$ -**35** underwent predominantly UV-promoted degradation,<sup>70,72</sup>  $\alpha$ -thioribocytidine  $\alpha$ -**38** undergoes an extraordinarily efficient  $\alpha \rightarrow \beta$  photoanomerization to produce the required C1'- $\beta$  stereochemistry, thus removing the demand for non-natural C2'-OH phosphorylation (as was the case with  $\alpha$ -**35**-2P) to control pyrimidine photoanomerization.<sup>73</sup> Phosphorylation of  $\beta$ -**38** under urea-mediated conditions produced the activated ribonucleotide and converted the nucleobase thiocarbonyl (C=S) to a carbonyl (C=O) to produce **29** in a single step.<sup>66</sup>

With two of the four canonical ribonucleotides synthesized, the compatible synthesis of the two remaining purine ribonucleotides currently remains elusive. However, the prospects of the purines yielding to continued efforts are promising. A congruent pathway to the analogous purines that runs in parallel to pyrimidine ribonucleotide synthesis would strongly validate both as prebiotically plausible routes to all four canonical ribonucleotides, and preliminary studies have provided potential leads.<sup>74</sup> Purine ribonucleotide precursors **39** have been efficiently assembled in a multicomponent reaction that simultaneously yields pyrimidine precursor **31** (Figure 7). This multicomponent reaction exploits prebiotically plausible purine precursors 5-aminoimidazole-4-carboxamide (AICA, **40**) and 4-amino-5-cyanoimidazole (AICN, **41**),<sup>75,76</sup> which undergo high yielding Mannich-type reactivity with 2-aminooxazole **32** and glyceraldehyde **20**, resulting in *N*<sup>9</sup> glycosylation at the C1' carbon of the newly formed five-carbon backbone with absolute regioselectivity and high diastereoselectivity.<sup>74</sup> What is particularly interesting about this mixed chemical system is that it exhibits pH dependent-divergent behaviour. This provides a tremendous opportunity to produce both purine and pyrimidine ribonucleotide precursors within the same environment through mixed reactivity of aldehydes and imines. For example, pentose aminooxazolines **31** are predominant products at pH 7, whilst three-component chemistry to produce **39** is dominant between pH 4-5, and a mixture is observed between pH 5-6.5.<sup>74</sup>

HCN-oligomers are an especially attractive alternative to fully formed nucleobases (Figure 2B) for regio- and stereocontrolled purine glycosylation. The synthesis of HCN-tetramers, such as AICA **40**, AICN **41** and diaminomaleonitrile **42** (Figure 2B, blue), is significantly more robust than the directed oligomerization of **9** to purine nucleobases themselves under prebiotic conditions.<sup>75,76</sup> HCN-oligomers (Figure 2B, blue) likely served a critical role in purine nucleotide synthesis, and accordingly a controlled pathway to these HCN-oligomers that avoids the adverse complexity and low yields of the direct high-pH hydrogen cyanide **9**-oligomerization is needed.<sup>18,37,38</sup> A comparison can be drawn with formose prebiotic sugar syntheses. Attempts to yield prebiotic sugars had (until recently) relied universally upon uncontrolled high-pH formaldehyde **12**-oligomerization, however, this strategy proved to be highly unsatisfactory and does not exhibit the control required to yield the elusive ribose **7** that is required for ribonucleotide

assembly. These acute difficulties provided a strong mandate to explore other avenues, which eventually led to Sutherland's elegant Kiliani-Fischer chemistry as a solution to the prebiotic sugar synthesis. These dramatic improvements in prebiotic sugar synthesis were made by exploration of chemical spaces that does not fall under the 'traditional' strategies of prebiotic chemistry.<sup>59,60</sup> Furthermore, Sutherland's C<sub>2</sub> and C<sub>3</sub> sugar syntheses led to the discovery of a methodology to circumvent the inefficient synthesis of amino acids via Miller-Urey spark discharge chemistry.<sup>15,28,60</sup> Similarly, new methods are now needed for controlling HCN-oligomerization. Strategies for HCN-oligomerization that avoids the irreparable complexity and low yields observed in purine nucleobase synthesis by (high pH) HCN-oligomerization are required to complete the transition away from uncontrolled high pH oligomerization chemistry.

### **Crystallization-driven assembly of ribonucleotides from complex mixtures**

The convergent syntheses of the pyrimidine ribonucleotides **29** and **30**, summarized in Figures **5** and **6**, currently stand as the most complete prebiotic model for ribonucleotide assembly.<sup>61,66</sup> There are, however, some caveats to these models and valid questions have persisted with regards to the reliance on pure glycolaldehyde **14** and glyceraldehyde **20**, and the order of assembly of the key pentose aminooxazoline **31** intermediate.<sup>9,22,25,77</sup> The plausibility of pentose aminooxazoline **31** syntheses is reliant on the sequential addition of cyanamide **33** to pure glycolaldehyde **14** to produce 2-aminooxazole **32**, followed by its delivery to pure glyceraldehyde **20** in a second step. This sequence of addition must be strictly adhered to because glycolaldehyde **14** and glyceraldehyde **20** lack the intrinsic selectivity to react with cyanamide **33**, leading to the formation of a mixture of their corresponding oxazoles **32** and **43** (Figure 8).<sup>69,74</sup> The situation is further exacerbated by the presence of other prebiotic sugars, which also competitively react with cyanamide **33** to form their respective aminooxazolines.<sup>78</sup> One potential way to overcome this problem is by exploiting the volatility of 2-aminooxazole **32**. 2-Aminooxazole **32** has a high vapour pressure and readily sublimates on warming, and accordingly may have been 'rained in' to a separate pool containing glyceraldehyde **20** at a different location.<sup>79</sup> There is, however, a conceptual flaw in this geochemical scenario; there are currently no known prebiotic syntheses that exclusively produce **14** or **20**. All known prebiotic syntheses of sugars generate mixtures of glycolaldehyde **14** and glyceraldehyde **20** alongside other aldehydes and ketones,

making the prospects of exclusive syntheses of either C<sub>2</sub> or C<sub>3</sub> sugar overwhelmingly small. Sutherland's Kiliani-Fischer homologation of cyanide **9**, outlined already in the preceding discussion, is the most streamlined and highest yielding pathway to glycolaldehyde **14** and glyceraldehyde **20**, yet even this produces both as a mixture.<sup>60</sup>

To complicate matters further, glyceraldehyde **20** isomerizes to the thermodynamically more stable ketose isomer dihydroxyacetone **22** under various conditions, such as specific base, metal ion, and general acid-base catalysis.<sup>80</sup> Phosphate is, of course, one such catalyst, but it is also crucial that its presence is assured from the outset of the prebiotic synthetic sequence of ribonucleotide assembly because phosphate catalyzes the formation of 2-aminooxazole **32** and curtails the unwanted side reactions of cyanoacetylene **34**.<sup>61</sup> Worryingly, a dichotomy presents itself as a result of introducing phosphate into the mix early on. Phosphate is beneficial to the syntheses of 2-aminooxazole **32** and anhydronucleosides **36**, but it appears to be detrimental to glyceraldehyde **20** that is required for pentose aminooxazoline **31** formation. Glyceraldehyde **20** is rapidly (but cleanly) converted to dihydroxyacetone **22** by phosphate catalysis in less than 1 hour under the conditions of 2-aminooxazole **32** synthesis. To demonstrate the consequences of this, the amount of pentose aminooxazoline **31** formation is so little when pre-equilibrated mixtures of glycolaldehyde **14** and triose sugars **20** and **22** are exposed to 2-aminooxazole **32** that even the privileged crystallinity of *ribo*-**31** can no longer resolve this mixture.<sup>36</sup> So while the reaction of glyceraldehyde **20** and 2-aminooxazole **32** is exceedingly clean, the prospects of this reaction furnishing sufficiently pure ribonucleotide precursors is thwarted by an increasing amount of 'prebiotic clutter'. Specifically, there is no resolution between the reactivity of C<sub>2</sub> and C<sub>3</sub> sugars. Moreover, there is a lack of aldose-ketose resolution for the C<sub>3</sub> sugars glyceraldehyde **20** and dihydroxyacetone **22**, and though dihydroxyacetone **22** is thermodynamically favoured it is glyceraldehyde **20** that is required for nucleotide assembly. The inability to form pentose aminooxazolines **31** with sufficient purity from complex prebiotic mixtures is a serious problem for the efficacy of prebiotic nucleotide synthesis. However, a solution to all of these problems has recently been found by way of 2-aminothiazole **44** mediated physicochemical processes that result in the high yielding and chemoselective formation of pentose aminooxazolines **31** from remarkably complex



mixtures.<sup>36</sup> This is achieved by selective sequestration of glycolaldehyde **14** and glyceraldehyde **20** from highly complex sugar mixtures (Figure 9).

These C<sub>2</sub> and C<sub>3</sub> sugars crystallized as the corresponding amino acids **45** and **46** in a time-resolved sequence required for ribonucleotide synthesis, and were cleanly converted to pentose aminooxazolines **31**. These results reinforce the utility of the C<sub>2</sub> and C<sub>3</sub> sugars **14** and **20** (rather than ribose **7**), the pentose aminooxazolines **31**, and the critical role of phosphate in ribonucleotide assembly. They also highlight that 2-aminothiazole **44** is essential in orchestrating these processes. The role of **44** is particularly noteworthy because it plays a purely catalytic role in ribonucleotide assembly as **44** is not ultimately expressed in the constitution of these products (Figure 9).

The crystallization-driven assembly of pentose aminooxazolines **31** from complex mixtures was conceived through a systems chemical analysis that recognized the relationship between amino acids and ribonucleotides (Figure 10).<sup>79</sup> Glycolaldehyde **14** provides the carbon framework for serine (**Ser**) and 2-aminoxazole **32**, and accordingly occupies a generational node between nucleic acid and amino acid syntheses. Prebiotic nucleic and amino acid syntheses require that **14** undergoes divergent fates—either through the reaction with cyanamide **33** to produce ribonucleotides, or with ammonium cyanide (NH<sub>4</sub>CN) to produce serine (**Ser**) (Figure 10). Extrapolation of this relationship to other essential amino acids then leads to β-mercaptoacetaldehyde **47** by application of the analogous retrosynthetic analysis to cysteine (**Cys**).

It turns out that the reaction of β-mercaptoacetaldehyde **47** and nucleotide precursor cyanamide **33** is one of the most chemoselective and robust reactions observed in prebiotic chemistry, so much so that it can be considered as a prebiotic variant of the venerable class of ‘click’ reactions. Cyanamide **33** sequesters β-mercaptoacetaldehyde **47** in nearly quantitative yield to produce 2-aminothiazole **44** even in stoichiometric competition with 26 other aldehydes, ketones and sugars (which include formaldehyde **12**, glycolaldehyde **14**, glyceraldehyde **20** and ribose **7**).<sup>36</sup> The reaction also tolerates the presence of other thiols and sulfides. Furthermore, the reduction of prebiotically plausible disulfides (observed in UV-light mediated amino acid synthesis in H<sub>2</sub>S-enriched gaseous mixtures)<sup>81</sup> also produces 2-aminothiazole **44** upon reduction by

ammonium cyanide (NH<sub>4</sub>CN). These latter conditions are notable for their relationship to the Strecker amino acid synthesis. The driving force for the chemoselective synthesis of 2-aminothiazole **44** in such complex mixtures is twofold. The first element is a kinetic factor, and this is dictated by the high nucleophilicity of sulfur resulting in the rapid and preferential trapping of  $\beta$ -mercaptoacetaldehyde **47** by cyanamide **33**, despite the concentration of  $\alpha$ -hydroxyaldehydes exceeding that of  $\beta$ -mercaptoacetaldehyde **47** by an order of magnitude.<sup>36</sup> The second is a thermodynamic driving force because  $\beta$ -mercaptoacetaldehyde **47** is privileged to readily form an aromatic heterocycle upon condensation with cyanamide **33**. The multiple and highly predisposed pathways to 2-aminothiazole **44** suggest that it would have been present in cyanosulfidic environments, and this is corroborated by its constitutional relationship to cysteine (Cys) and pyrimidine ribonucleotides. The role of 2-aminothiazole **44**, however, turns out to be a rather different to the previously outlined role of 2-aminooxazole **32**.

2-Aminothiazole **44** exhibits a greater degree of aromaticity than 2-aminooxazole **32** and this may explain why conditions for its masked aldol reaction with glycolaldehyde **14** or glyceraldehyde **20** have yet to be shown.<sup>79</sup> However, the lack of such reactivity turns out to be rather fortunate for ribonucleotide assembly. Instead, 2-aminothiazole **44** induced the unexpected precipitation of glycolaldehyde **14** and glyceraldehyde **20** as crystalline amins when incubated at neutral pH in water,<sup>79</sup> and the incubation of glycolaldehyde **14** and glyceraldehyde **20** with 2-aminothiazole **44** in phosphate buffer at pH 7 results in rapid and near quantitative formation of the corresponding amins **45** and **46**.<sup>36</sup> These amination forming reactions are robust, underpinned by their tolerance of a wide range of potentially prebiotic compounds such as amino acids, nucleobases, ammonia, aminonitriles, borate, and iron cyanide complexes. Although the potential of these amins to serve as a stable reservoir of C<sub>2</sub> and C<sub>3</sub> aldoses was immediately recognised, the contemporaneous precipitation of the amins **45** and **46** from water did not immediately provide a satisfactory solution to separation of glycolaldehyde **14** and glyceraldehyde **20**.<sup>36</sup>

In a direct competition, glycolaldehyde **14** and glyceraldehyde **20** initially formed a mixture of C<sub>2</sub>-amination **45** and C<sub>3</sub>-amination **46**, but intriguingly only the C<sub>2</sub>-amination **45** remained after 12 hours. Amination formation is reversible, and during the course of the reaction glyceraldehyde **20** had in fact undergone

equilibration with dihydroxyacetone **22**. Further experiments confirmed that glycolaldehyde **14** and glyceraldehyde **20** are resolved under thermodynamic control, driven by phosphate-catalyzed isomerization of glyceraldehyde **20** to the more stable ketose isomer dihydroxyacetone **22**. This allowed the exclusive accumulation of pure C<sub>2</sub>-aminal **45** from a mixture of C<sub>2</sub> and C<sub>3</sub> sugars (Figure 9A). So whilst it was possible to separate glycolaldehyde **14** from triose sugars **20** and **22**, there was now the need to find a way to convert dihydroxyacetone **22** back to glyceraldehyde **20**. Phosphate has already been shown to play a crucial role in numerous ways in recent prebiotic chemistry,<sup>15,59-61,66,67</sup> and it was once again found to have a central role in resolving the problem of triose equilibration. Prolonged incubation of dihydroxyacetone **22** with 2-aminothiazole **44** in phosphate buffer at pH 7 led to the deposition of a precipitate that transpired to be pure C<sub>3</sub>-aminal **46**, rather than the corresponding aminal of dihydroxyacetone **22**.<sup>36</sup>

The rate of C<sub>3</sub>-aminal **46** accumulation from dihydroxyacetone **22** (9%, 24 h → 62%, 312 h) is much slower than direct precipitation from pure glyceraldehyde **20**, which is complete within a matter of a few hours. This leads to a prebiotically plausible mechanism for C<sub>2</sub> and C<sub>3</sub> sugar resolution. Dihydroxyacetone **22** predominates over glyceraldehyde **20** at equilibrium and therefore presents an ideal opportunity to resolve glycolaldehyde **14** and glyceraldehyde **20** as crystalline compounds in the order required for the synthesis of pentose aminooxazolines **31** under plausible geochemical conditions. This crystallization process overcomes the long-standing conundrum of the thermodynamic preference for dihydroxyacetone **22** over glyceraldehyde **20** to achieve selective accumulation and resolution of C<sub>2</sub> and C<sub>3</sub> sugars as aldose derivatives. Moreover, the privileged crystallinity of C<sub>2</sub>- and C<sub>3</sub>-aminals **45** and **46** provides a direct physical mechanism to cleanly separate glycolaldehyde **14** and glyceraldehyde **20** from what would have previously been considered to be wholly intractable mixtures (Figure 9A). The first precipitate, C<sub>2</sub>-aminal **45**, crystallizes out of complex sugar mixtures within 2 hours. Aminal **45** subsequently undergoes quantitative conversion to 2-aminooxazole **32** upon exposure to cyanamide **33**, resulting in the first stage of 2-aminothiazole **44** regeneration. This crude 1:2 mixture of **32** and **44** is then exposed to the C<sub>3</sub>-aminal **46** that had crystallized (much later) from the same complex sugar mixture, to result in the formation of the key pentose aminooxazoline intermediates **31**. The 2-aminothiazole **44** mediated multistep synthesis of **31** is

remarkably as clean as the reaction of pure glyceraldehyde **20** and 2-aminooxazole **32**.<sup>61,68</sup> The large excess of 2-aminothiazole **44** liberated in the latter stages of the synthesis is merely present as a spectator and does not interfere with the formation of 2-aminooxazole **32** or the synthesis of pentose aminooxazolines **31**, effectively rendering it a catalytic chemical chaperone that directs multistep prebiotic synthesis. Once liberated from the aminor reservoirs it is free to reiterate the crystallization process again. Interestingly, 2-aminothiazole **44**, like 2-aminooxazole **32**, is volatile and readily sublimates.<sup>79</sup> This provides a simple physical mechanism by which **44** could be re-delivered, for example, to the locations where the formation of glycolaldehyde **14** and glyceraldehyde **20** is taking place.<sup>15,59,60</sup> Phosphate and 2-aminothiazole **44** did not impede the crystallization of pure *ribo*-**31** crystals from the crude pentose aminooxazoline **31** mixture obtained from aminor-mediated synthesis, and X-ray crystallographic analysis of these crystals unexpectedly revealed the formation of novel conglomerates, where the individual crystals are purely homochiral in a bulk mixture that is wholly racemic. The reasons for this surprising observation are not known at present, but previously *rac-ribo*-**31** had only been observed to form enantiomorphously twinned crystals in which individual crystals are racemic but may contain homochiral domains.<sup>78</sup> However, the importance of *D-ribo*-**31** in prebiotic ribonucleotide synthesis suggests that this observation could be crucial to understanding the origins of biological homochirality.<sup>20</sup>

The benefits of crystallization in directing multistep prebiotic ribonucleotide assembly are indisputable. The crystallinity of aminorals **45** and **46**, and *ribo*-**31** provides them with a protective mechanism that simply cannot be offered by strategies that propose the formation of ribonucleotides *via* ribose **7**. The stability of *ribo*-**31** in solution is already 70 times higher than ribose **7** under comparative conditions, and it also exhibits greater stability under both acidic and basic conditions. However, the crystallinity of *ribo*-**31** adds an even greater element of stability, rendering it “*essentially inert once it forms crystals*”.<sup>78</sup> The additional effect of concentrating large quantities of key building blocks as precipitates during prebiotic assembly chemistry, some of which are considered somewhat unstable (including for example cyanoacetylene **34** as a copper salt),<sup>15</sup> is a highly appealing mechanism to placate the specific concerns of

product selection, reagent concentration and to tame highly reactive substrates and indiscriminate reactivity thereof.<sup>22,25,77,82</sup>

### Chiral resolution of ribonucleotides

Glyceraldehyde **20** is the first chiral sugar from which other sugars and nucleotides are derived, and therefore serve as a building block for downstream synthesis of more complex chiral building blocks. Significantly, glyceraldehyde **20** possesses a single stereogenic centre that is readily racemised, making it a highly appealing target for dynamic kinetic resolution. Breaking the symmetry at glyceraldehyde **20** would set in motion a domino effect such that all subsequent sugars and nucleotides could be built from the chiral pool of homochiral D-**20**. Furthermore, the issue of homochirality only needs to be resolved once for sugars and nucleotides (at the level of glyceraldehyde **20**), whereas the case for the amino acids is not so straightforward. There are 20 different proteinogenic amino acids and each (excluding Gly) requires a related or separate solution to resolve their chirality. The crystallization of pure homochiral D-*ribo*-**31** from the reaction of scalemic glyceraldehyde **20** is an important observation in this context, providing a simple mechanism to enrich chirality en route to ribonucleotides (Figures 5 and 6).<sup>68</sup> Accordingly, breaking symmetry at the C<sub>3</sub> sugar seems more logical than tackling this problem at a more advanced stage in ribonucleotide synthesis.

Presently, known routes to non-racemic glyceraldehyde **20** require homochiral amino acids to impart chirality (Figure 11), and accordingly highlight an important potential ‘chicken and egg dilemma’ or the requirement for a chiral symmetry-breaking event prior to amino acid synthesis to furnish homochiral amino acids.<sup>20</sup> The first and most direct route to non-racemic glyceraldehyde **20** is the amino acid-catalyzed hydroxymethylation of glycolaldehyde **14** by formaldehyde **12**, producing enantiomeric excesses in up to 44% *ee*.<sup>20,83,84</sup> The physical properties of *rac*-glyceraldehyde **20** and homochiral L- and D-glyceraldehyde **20** could in principle be further exploited to enhance this rather impressive enantiomeric excess. Racemic glyceraldehyde **20** adopts a stable symmetric dimeric structure **49** which facilitates its packing into a crystal lattice, whereas homochiral D-glyceraldehyde **20** forms dimer **50** and syrups upon concentration (Figure 11B). The differential solubilities of racemic and homochiral glyceraldehyde **20** allow for a (low) initial

enantiomeric excess of pure glyceraldehyde **20** to be enriched by sequestration of racemic glyceraldehyde **20** in the solid state (as the dimer **49**) from highly concentrated scalemic mixtures that have undergone very rapid equilibration in water. This would then leave behind an aqueous solution with further enantioenriched glyceraldehyde **20**.<sup>84</sup> Problematically for this model, *ee*-enrichment by glyceraldehyde crystallization requires highly pure glyceraldehyde **20** and the amino acid-catalyzed condensation of glycolaldehyde **14** and formaldehyde **12** to produce glyceraldehyde **20** is not a straightforward reaction. Low conversion of **14** (2.4% after 4 days) and appreciable amounts of dihydroxyacetone **22** formed alongside glyceraldehyde **20** in these amino acid-catalyzed reactions.<sup>83</sup> A compromise must be struck in these reactions between selectivity for glyceraldehyde **20** with low yield, or a higher yield alongside the production of much greater product complexity. It is worth noting that glycolaldehyde **14** conversion can be increased under more forcing conditions, but this is at the expense of a multitude of other products beginning to form with no indications to suggest that there is an improvement in glyceraldehyde **20** yield.<sup>85</sup>

The second route to D-glyceraldehyde D-**20** is actually a resolution of *rac*-glyceraldehyde **20** rather than a chiral synthesis *per se*, and this is achieved by trapping L-glyceraldehyde (L-**20**) as a three-component reaction by-product (Figure 11A).<sup>86</sup> L-Amino acids, and particularly L-proline (L-**Pro**), kinetically resolve glyceraldehyde **20** by preferentially scavenging L-**20** during the synthesis of oxazoline **51**, whilst the majority of D-**20** is then free to undergo a bimolecular reaction with 2-aminooxazole **32** to yield enantioenriched **31**.<sup>86</sup> The crystallinity of *ribo*-**31** then ensures that homochiral D-*ribo*-**31** separates from the bulk solution containing a mixture of enantiomerically enriched **51** and **31**, if pentose aminooxazolines **31** with >20% *ee* is synthesized.<sup>68,86</sup> However, up to half of the glyceraldehyde **20** and 2-aminooxazole **32** would be lost as part of the three component products **51** and **52** along with the natural L-amino acid required for peptide synthesis.<sup>86</sup> Ideally, a dynamic resolution would allow for the total conversion of glyceraldehyde **20** without destroying the natural L-amino acids. Nevertheless this multicomponent reaction provides another insight into the potential benefits that can be amassed through system chemistry; by exploring the interrelated reactions of amino acids and nucleotide precursors, an intriguing link between D-nucleotide synthesis and L-amino acids was uncovered.

The most efficient syntheses of C<sub>3</sub> sugars are observed under conditions that result in their eventual isomerization to dihydroxyacetone **22**.<sup>53,60</sup> The implications of this is that a chirogenic reaction generating scalemic or homochiral glyceraldehyde **20** requires its onward reactivity before it undergoes racemization, and this would occur unless scalemic or chiral glyceraldehyde **20** is removed before the erosion of enantiomeric excesses sets in. It is particularly fascinating that the phosphate-catalyzed C<sub>3</sub>-aminal **46** formation from dihydroxyacetone **22** both (re)generates the chiral centre and is coupled to crystal lattice formation.<sup>36</sup> A dynamic kinetic resolution of C<sub>3</sub> sugars, *via* aminal crystallization, is particularly attractive. Dynamic chiral resolution of triose sugars (**20** and **22**) during aminal-crystallization could in principle eventually convert all C<sub>3</sub> sugars into a stable, homochiral source of glyceraldehyde **20** whilst removing this precursor from the effector of racemization. It seems highly probable that other chiral catalysts (such as amino acids, peptides or heterogeneous mineral surfaces) will facilitate enantioenriched glyceraldehyde synthesis and that other crystalline derivatives of glyceraldehyde **20** are yet to be discovered, some of which may exhibit homochiral conglomerate crystallization behaviour that could facilitate dynamic resolution of homochiral glyceraldehyde **20**. A systematic search for these is currently underway.

### **Crystallization-driven assembly of proteinogenic amino acids from complex mixtures**

As outlined in the preceding discussion, the prebiotic synthesis of various amino acids has been known for more than 60 years. However, it has remained a mystery for all that time why the puzzling chemical complexity of the mixtures formed under abiotic constraints is not reflected by the relatively simple repertoire of the proteinogenic amino acids assigned to the genetic code.<sup>15,28-31</sup> Considerable amounts of  $\alpha$ -methyl amino acids, the Strecker products of ketones, are generated in the spark discharge experiments and found in meteoritic samples.  $\alpha$ -Amino isobutyric acid and isovaline are examples of two highly abundant non-proteinogenic amino acids that have been detected.<sup>28-31</sup> Additionally, Sutherland's chemical pathway to twelve of the proteinogenic amino acids (described previously), also produces ketones as essential intermediates en route to several amino acids and lipids.<sup>15</sup> These ketones are just as susceptible as aldehydes to highly efficient aminonitrile formation, thus yielding  $\alpha,\alpha$ -disubstituted amino acids that are not assigned to the universal genetic code.<sup>36</sup>

2-Aminothiazole **44** mediated crystallization was found once again to provide a solution to this long-standing prebiotic problem (Figure 9B).<sup>36</sup> The hints of such a possibility were found in the observation that dihydroxyacetone **22** had conspicuously failed to precipitate as the corresponding aminal. 2-Aminothiazole **44** selectively crystallizes Strecker aldehydes (**11**, blue) in high yields, whereas ketones (**11**, red) have so far never been observed to crystallize as their respective aminals. Separation of Strecker aldehydes from a complex mixture of prebiotic aldehydes (**11**, blue) and closely related ketones (**11**, red) in the presence of 2-aminothiazole **44** is readily observed. These ketones, including precursors of  $\alpha$ -methyl amino acids frequently found in prebiotic amino acid syntheses and abiotic exogenous samples,<sup>15,28-31,60</sup> can then be readily separated from Strecker precursors of proteinogenic amino acids.<sup>36</sup> Intriguingly, quantitative crystallization of a mixture of aminals **48**, derived only from aldehydes (**11**, blue), was observed thereby allowing all the ketones (**11**, red) to be washed away. These Strecker-aldehyde aminals **48** were directly converted to the aminonitriles **10** upon exposure to cyanide **9** and ammonia, providing the first mechanism to exclude non-natural  $\alpha,\alpha$ -disubstituted amino acids from prebiotic amino acid syntheses.<sup>36</sup> The propensity for 2-aminothiazole **44** to directly sequester aldehydes from mixtures that also contains ketones now provides a simple and robust physical method to drive the chemical selection of proteinogenic amino acid precursors. More profound, however, is the existence of a common physicochemical mechanism that selects for the proteinogenic amino acid formation *and* canonical ribonucleotide assembly from prebiotically plausible mixtures. This duality of function in the assembly of the monomeric units of two essential biopolymers (peptide and nucleic acids) further corroborates the role of 2-aminothiazole **44** in choreographing the syntheses of prebiotic metabolites.<sup>36</sup>

### **Systems chemistry approach to prebiotic triose glycolysis**

It is tempting to draw comparisons between the enzymatic mechanism of aldose-ketose equilibration of triose sugar phosphates during glycolysis and the physicochemical conversion of dihydroxyacetone **22** into glyceraldehyde **20** during phosphate catalysed aminal-crystallization. Triosephosphate isomerase (TIM) is a near kinetically perfect enzyme that has gone through extensive evolutionary optimization to achieve diffusion-limited isomerization of dihydroxyacetone phosphate **22-P** and glyceraldehyde-3-phosphate **20-**



3P.<sup>87</sup> The equilibrium position of this reaction lies heavily in favour of dihydroxyacetone phosphate (**22-P**:**20-3P**; 20:1), but is overcome by consumption of glyceraldehyde-3-phosphate **20-3P** in the subsequent steps of glycolysis. TIM has such a highly optimized active site that the decomposition of the common enediol intermediate is stereoelectronically disfavoured,<sup>88</sup> and the decomposition of triose phosphates **22-P** and **20-3P** by E1cB (Elimination Unimolecular conjugate Base) elimination to methylglyoxal **53** and phosphate occurs one hundred times faster in the absence of the protective mechanism offered within the confines of the enzyme's active site (Figure 12).

The triose sugar phosphates **22-P** and **20-3P** are highly vulnerable without such a sophisticated stabilizing effect imparted by the enzymatic cavity, leaving the prospects of exploiting dihydroxyacetone phosphate **22-P** and/or glyceraldehyde-3-phosphate **20-3P** improbable in a prebiotic setting unless a favourable onward reaction has a faster rate than their decomposition. In contrast to the triose sugar phosphates **22-P** and **20-3P**, dihydroxyacetone **22** and glyceraldehyde **20** exhibit far greater stability than their phosphorylated variants under comparative conditions,<sup>80,89</sup> but prior to the discovery of phosphate-catalyzed C<sub>3</sub>-aminal **46** crystallization from dihydroxyacetone **22** there were no known mechanisms of overturning the unfavourable triose aldose-ketose equilibration that had plagued prebiotic chemistry for decades. This elicits speculation that phosphate may have served as a simple primordial isomerase (with the assistance of aldehyde-selective aminal sequestration) before the invention and evolution of enzymes that could be optimized to handle unstable triose phosphates **22-P** and **20-3P**.

In the absence of the catalytic prowess of enzymes to process labile substrates such as glyceraldehyde-3-phosphate **20-3P**, prebiotic chemistry must find alternative routes to access pathways that may have harvested chemical energy. Modern triose glycolysis is a metabolic pathway that harnesses energy from the multistep degradation of sugars to pyruvate **54**. The intermediates of triose glycolysis are some of the most interconnected nodes in central metabolism essential for amino acids and lipid syntheses, and access to the citric acid cycle (Figure 12). Glyceraldehyde-3-phosphate **20-3P** and phosphoenolpyruvate **55** are crucial intermediates in a sequence of enzymatically controlled reactions that convert glucose into pyruvate **54** through multiple discrete reactions that include phosphorylation, elimination, isomerization, and redox

processes. Triose glycolysis is a highly conserved metabolic pathway present across biological kingdom, and considered to be a relic of primitive metabolism.<sup>90</sup>

Recently it was shown that the ribonucleotide building blocks, glycolaldehyde **14** and glyceraldehyde **20**, can also be converted into nature's highest energy phosphorylating agent—phosphoenolpyruvate **55**—in a simple non-enzymatic reaction sequence.<sup>51</sup> This pathway provides evidence to support the idea that the complex pentose (C<sub>5</sub>) or hexose (C<sub>6</sub>) sugars, or the corresponding sugar phosphates, are not required to access primitive metabolic pathways. It is significant that access to prebiotic triose glycolysis is through  $\alpha$ -phosphorylation rather than extant life's modern route of terminally phosphorylated triose sugars.<sup>51,90</sup> Glyceraldehyde-2-phosphate **20-2P**, the predisposed product of  $\alpha$ -aldose phosphorylation (Figure 3), is significantly more stable than glyceraldehyde-3-phosphate **20-3P**.<sup>48</sup> Incubation of glycolaldehyde phosphate **14-P** and formaldehyde **12** in phosphate buffer at pH 7 results in a rapid tandem hydroxymethylation–dehydration to give phosphoenol pyruvaldehyde **56** in high yield and under remarkably mild conditions (Figure 12). The rapid formation and dehydration of glyceraldehyde-2-phosphate **20-2P** under these mild conditions is noteworthy; previously high pH conditions were required to allow the aqueous synthesis of glyceraldehyde-2-phosphate **20-2P** (Figure 3).<sup>48</sup> Phosphoenolpyruvaldehyde **56** then undergoes facile oxidation under a number of potentially primitive oxidative conditions to give access to phosphoenolpyruvate **55** in excellent yield. Glyceraldehyde-2-phosphate **20-2P** mitigates the deleterious E1cB elimination of phosphate that glyceraldehyde-3-phosphate **20-3P** is susceptible to in the absence of enzymatic protection.<sup>89</sup> Moreover, though glyceraldehyde-2-phosphate **20-2P** is more stable than glyceraldehyde-3-phosphate **20-3P** it is also intrinsically activated towards an E1cB-type elimination of water to access the valuable high-energy phosphoenol moiety of **55** and **56**. The reaction network was further expounded to give access to all the intermediates of the triose glycolysis pathway, and even phosphoserine **57**, through subtle variations of simple reaction conditions. Therefore,  $\alpha$ -phosphorylation of the simplest aldose sugars form the basis of a reconstituted primitive triose glycolysis pathway, providing yet another example of the divergence of prebiotic chemistry into essential metabolites.<sup>51</sup> The exploitation of glyceraldehyde-2-phosphate **20-2P** to build a primitive high yielding triose glycolysis pathway stands in

stark contrast to the elevated pH-dependent aldolization chemistry of glyceraldehyde-2-phosphate **20-2P** that was previously described by Eschenmoser and colleagues as a potential source of sugar phosphates for primordial nucleic acid assembly (Figure 3).<sup>1,48</sup> Rather than serve as precursors for non-canonical pentose-2,4-diphosphates, glycolaldehyde phosphate **14-P** and glyceraldehyde-2-phosphate **20-2P** can be readily directed into alternative prebiotic reaction pathways to undergo transformations to yield nature's highest energy phosphate at the origin of life.<sup>51</sup>

## CONCLUSIONS AND OUTLOOK

Significant progress is being made towards demonstrating the abiotic generation of the natural biological structures through the implementation of systems chemistry. Prebiotic chemistry must continue to embrace this approach to have the best possible chance of uncovering the roots of biological organization. The predisposed reaction pathways that were essential to the origins of life need not be the most obvious routes to the canonical biological structures, but they must be carefully considered in a broader context of the whole system under investigation. The lessons learnt during the discovery of selective syntheses of pyrimidine ribonucleotides and proteogenic amino acids are illustrative of this approach to prebiotic chemistry. For example, the isomerization of glyceraldehyde **20** to dihydroxyacetone **22** was near-universally perceived to be detrimental to the selective synthesis of ribonucleotides and amino acids, but what had at first appeared to be an undesirable affliction turned out to be a tremendous opportunity following a system chemical analysis.<sup>5,15,36</sup>

Physicochemical processes clearly need to receive far greater attention and will undoubtedly play a key role in understanding the chemical origins of life. Coupling chemical reactions to physical processes (such as separation and transportation) provides simple, yet plausible, mechanisms to accrue relatively pure materials required to specifically build the canonical components of extant life.<sup>19,20,91</sup> Indeed, reflecting on the relationship between the chemical subsystems required for primitive life may provide not only the most compelling evidence for the deep-seated chemical origin of life, but also the most facile mechanism through which to make new discoveries within the field of prebiotic chemistry.<sup>4,5</sup>

A deeper understanding is required of how a purely chemical system acquired the traits of one that is living. The transition from the products of simple chemical pathways and prebiotic building blocks to a proto-life remains one of the most fascinating mysteries in science, and yet this problem can only be addressed by building an incremental understanding of the generational and functional relationships between metabolite and macromolecules within the broader context of an all-inclusive system. Recent studies exploring the dynamic behaviour and the synergistic effects of nucleic acids, peptides, lipids and other prebiotic small molecules demonstrate these interactions can give rise to remarkable emergent properties that would not have been observed in isolation.<sup>42,91-100</sup> Under the auspices of prebiotic systems chemistry, it is now time for organic chemistry to rise to the challenge of retracing nascent life's lost chemical roots.

## **AUTHOR CONTRIBUTIONS**

S.I and M. W. P wrote this paper.

## **ACKNOWLEDGEMENTS**

This work was supported by the Simons Foundation (318881), the Engineering and Physical Sciences Research Council (EP/K004980/1), and the Leverhulme Trust (RGP-2013-189).

## **REFERENCES AND NOTES**

1. Eschenmoser, A. (2011). Etiology of potentially primordial biomolecular structures: from vitamin B<sub>12</sub> to the nucleic acids and an inquiry into the chemistry of life's origin: a retrospective. *Angew. Chem. Int. Ed. Engl.* *50*, 12412-12472
2. Szostak, J. W., Bartel, D. P., and Luisi, P. L. (2001). Synthesizing life. *Nature* *409*, 387-390
3. Pace, N. R. (2001). The universal nature of biochemistry. *Proc. Natl. Acad. Sci. USA* *98*, 805-808
4. Sutherland, J. D. (2016). The origin of life—out of the blue. *Angew. Chem. Int. Ed. Engl.* *55*, 104-121
5. Powner, M. W., and Sutherland, J. D. (2011). Prebiotic chemistry: a new modus operandi. *Philos. Trans. R. Soc. B.* *366*, 2870-2877
6. Lombard, J., López-García, P., and Moreira, D. (2012). The early evolution of lipid membranes and the three domains of life. *Nature Rev. Microbiol.* *10*, 507-515

7. Ruiz-Mirazo, K., Briones, C., and de la Escosura, A. (2014). Prebiotic systems chemistry: new Perspectives for the origins of life. *Chem. Rev.* *114*, 285-366
8. Joyce, G. F. (2002). The antiquity of RNA-based evolution. *Nature* *418*, 214-221
9. Robertson, M. P., and Joyce, G. F. (2012). The origins of the RNA world. *Cold Spring Harb. Perspect. Biol.* *4*, a003608
10. Powner, M. W., Sutherland, J. D., and Szostak, J. W. (2011). The origins of nucleotides. *Synlett* *14*, 1956-1964
11. Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science* *289*, 920-930
12. Steitz, T. A., and Moore, P. B. (2003). RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem. Sci.* *28*, 411-418
13. Öberg, K. I. (2016). Photochemistry and astrochemistry: photochemical pathways to interstellar complex organic molecules. *Chem. Rev.* *116*, 9631-9663
14. Raulin, F., Brasse, C., Poch, O., and Coll, P. (2012). Prebiotic-like chemistry on Titan. *Chem. Soc. Rev.* *41*, 5380-5393
15. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D., and Sutherland, J. D. (2015). Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* *7*, 301-307
16. Bowler, F. R., Chan, C. K. W., Duffy, C. D., Gerland, B., Islam, S., Powner, M. W., Sutherland, J. D., and Xu, J. F. (2013). Prebiotically plausible oligoribonucleotide ligation facilitated by chemoselective acetylation. *Nat. Chem.* *5*, 383-389
17. Fernández-García, C., and Powner, M. W. Selective acylation of nucleosides, nucleotides, and glycerol-3-phosphocholine in Water. *Synlett* *28*, 78-83
18. Schwartz, A. W. (2007). Intractable mixtures and the origin of life. *Chem. Bio. Div.* *4*, 656-664
19. Budin, I., and Szostak, J. W. (2010). Expanding roles for diverse physical phenomena during the origin of life. *Annu. Rev. Biophys.* *39*, 245-263

20. Hein, J. E., and Blackmond, D. G. (2012). On the origin of single chirality of amino acids and sugars in biogenesis. *Acc. Chem. Res.* *45*, 2045-2054
21. Kim, H.-J., Ricardo, A., Illangkoon, H. I., Kim, M. J., Carrigan, M. A., Frye, F., and Benner, S. A. (2011). Synthesis of carbohydrates in mineral-guided prebiotic cycles. *J. Am. Chem. Soc.* *133*, 9457-9468
22. Benner, S. A., Kim, H.-J., and Yang, Z. (2012). Setting the stage: the history, chemistry, and geobiology behind RNA. *Cold Spring Harb. Perspect. Biol.* *4*, a003541
23. Joyce, G. F., Schwartz, A. W., Miller, S. L., and Orgel, L. E. (1987). The case for an ancestral genetic system involving simple analogs of the nucleotides. *Proc. Natl. Acad. Sci. USA* *84*, 4398-4402
24. Orgel, L. (2000). Origin of life - A simpler nucleic acid. *Science* *290*, 1306-1307
25. Hud, Nicholas V., Cafferty, Brian J., Krishnamurthy, R., and Williams, Loren D. (2013). The origin of RNA and “my grandfather’s axe”. *Chem. Biol.* *20*, 466-474
26. Anastasi, C., Buchet, F. F., Crowe, M. A., Parkes, A. L., Powner, M. W., Smith, J. M., and Sutherland, J. D. (2007). RNA: prebiotic product, or biotic invention? *Chem. Bio. Div.* *4*, 721-739
27. Wagner, A. J., and Blackmond, D. G. (2016). The future of prebiotic chemistry. *ACS Cent. Sci.* *2*, 775-777
28. Miller, S. L. (1953). A production of amino acids under possible primitive earth conditions. *Science* *117*, 528-529
29. Johnson, A. P., Cleaves, H. J., Dworkin, J. P., Glavin, D. P., Lazcano, A., and Bada, J. L. (2008). The Miller volcanic spark discharge experiment. *Science* *322*, 404-404
30. Wolman, Y., Haverland, W. J., and Miller, S. L. (1972). Nonprotein amino acids from spark discharges and their comparison with the Murchison meteorite amino acids. *Proc. Natl. Acad. Sci. USA* *69*, 809-811

31. Matrajt, G., Pizzarello, S., Taylor, S., and Brownlee, D. (2004). Concentration and variability of the AIB amino acid in polar micrometeorites: implications for the exogenous delivery of amino acids to the primitive earth. *Meteorit. Planet. Sci.* *39*, 1849-1858
32. Schlesinger, G., and Miller, S. L. (1973). Equilibrium and kinetics of glyconitrile formation in aqueous solution. *J. Am. Chem. Soc.* *95*, 3729-3735
33. Shapiro, R. (1988). Prebiotic ribose synthesis - a critical analysis. *Orig. Life Evol. Biosph.* *18*, 71-85
34. Arrhenius, T., Arrhenius, G., and Paplawsky, W. (1994). Archean geochemistry of formaldehyde and cyanide and the oligomerization of cyanohydrin. *Orig. Life Evol. Biosph.* *24*, 1-17
35. Serianni, A. S., Nunez, H. A., and Barker, R. (1980). Cyanohydrin synthesis: studies with [<sup>13</sup>C] cyanide. *J. Org. Chem.* *45*, 3329-3341
36. Islam, S., Bučar, D.-K., and Powner, M. W. (2016). Prebiotic selection and assembly of proteinogenic amino acids and natural nucleotides from complex mixtures. *Nat. Chem.* DOI: 10.1038/nchem.2703
37. Oró, J. (1960). Synthesis of adenine from ammonium cyanide. *Biochem. Biophys. Res. Commun.* *2*, 407-412
38. Shapiro, R. (1995). The prebiotic role of adenine - a critical analysis. *Orig. Life Evol. Biosph.* *25*, 83-98
39. Mizuno, T., and Weiss, A. H. (1974). Synthesis and utilization of formose sugars. In *Advances in Carbohydrate Chemistry and Biochemistry*. R. S. Tipson and H. Derek, eds. (Academic Press) Vol. 29 pp. 173-227
40. Zubay, G. (1998). Studies on the lead-catalyzed synthesis of aldopentoses. *Orig. Life Evol. Biosph.* *28*, 13-26
41. Kofoed, J., Reymond, J. L., and Darbre, T. (2005). Prebiotic carbohydrate synthesis: zinc-proline catalyzes direct aqueous aldol reactions of  $\alpha$ -hydroxy aldehydes and ketones. *Org. Bio. Chem.* *3*, 1850-1855

42. Gardner, P. M., Winzer, K., and Davis, B. G. (2009). Sugar synthesis in a protocellular model leads to a cell signalling response in bacteria. *Nat. Chem.* *1*, 377-383
43. Larralde, R., Robertson, M. P., and Miller, S. L. (1995). Rates of decomposition of ribose and other sugars - implications for chemical evolution. *Proc. Natl. Acad. Sci. USA* *92*, 8158-8160
44. Chapelle, S., and Verchere, J.-F. (1988). A <sup>11</sup>B and <sup>13</sup>C NMR determination of the structures of borate complexes of pentoses and related sugars. *Tetrahedron* *44*, 4469-4482
45. Prieur, B. E. (2001). Étude de l'activité prébiotique potentielle de l'acide borique. *C. R. Acad. Sci. Ser.IIc* *4*, 667-670
46. Ricardo, A., Carrigan, M. A., Olcott, A. N., and Benner, S. A. (2004). Borate minerals stabilize ribose. *Science* *303*, 196-196
47. Furukawa, Y., Horiuchi, M., and Kakegawa, T. (2013). Selective stabilization of ribose by borate. *Orig. Life Evol. Biosph.* *43*, 353-361
48. Müller, D., Pitsch, S., Kittaka, A., Wagner, E., Wintner, C. E., and Eschenmoser, A. (1990). Chemie von  $\alpha$ -aminonitrilen. Aldomerisierung von glycolaldehyd-phosphat zu racemischen hexose-2,4,6-triphosphaten und (in gegenwart von formaldehyd) racemischen pentose-2,4-diphosphaten: rac-allose-2,4,6-triphosphat und rac-ribose-2,4-diphosphat sind die reaktionshauptprodukte. *Helv. Chim. Acta* *73*, 1410-1468
49. Krishnamurthy, R., Arrhenius, G., and Eschenmoser, A. (1999). Formation of glycolaldehyde phosphate from glycolaldehyde in aqueous solution. *Orig. Life Evol. Biosph.* *29*, 333-354
50. Krishnamurthy, R., Guntha, S., and Eschenmoser, A. (2000). Regioselective  $\alpha$ -phosphorylation of aldoses in aqueous solution. *Angew. Chem. Int. Ed. Engl.* **39**, 2281-2285
51. Coggins, A. J., and Powner, M. W. (2016). Prebiotic synthesis of phosphoenol pyruvate by  $\alpha$ -phosphorylation-controlled triose glycolysis. *Nat. Chem.* 10.1038/nchem.2624
52. Meinert, C., Myrgorodska, I., Marcellus, P. d., Buhse, T., Nahon, L., Hoffmann, S. V., d'Hendecourt, L. L. S., and Meierhenrich, U. J. (2016). Ribose and related sugars from ultraviolet irradiation of interstellar ice analogs. *Science* *352*, 208-212



53. Sagi, V. N., Punna, V., Hu, F., Meher, G., and Krishnamurthy, R. (2012). Exploratory experiments on the chemistry of the “glyoxylate scenario”: Formation of ketosugars from dihydroxyfumarate. *J. Am. Chem. Soc.* *134*, 3577-3589
54. Eschenmoser, A. (2007). On a hypothetical generational relationship between HCN and constituents of the reductive citric acid cycle. *Chem. Bio. Div.* *4*, 554-573
55. Vorbrüggen, H., and Ruh-Pohlenz, C. (2000). Synthesis Of Nucleosides. In *Organic Reactions*, L. A. Paquette, ed. (John Wiley & Sons, Inc.) Vol 55, 1-630
56. Sutherland, J. D. (2010). Ribonucleotides. *Cold Spring Harb. Perspect. Biol.* *2*, a005439
57. Becker, S., Thoma, I., Deutsch, A., Gehrke, T., Mayer, P., Zipse, H., and Carell, T. (2016). A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* *352*, 833-836
58. Fiore, M., and Strazewski, P. (2016). Bringing prebiotic nucleosides and nucleotides down to earth. *Angew. Chem. Int. Ed. Engl.* *55*, 13930-13933
59. Ritson, D., and Sutherland, J. D. (2012). Prebiotic synthesis of simple sugars by photoredox systems chemistry. *Nat. Chem.* *4*, 895-899
60. Ritson, D. J., and Sutherland, J. D. (2013). Synthesis of aldehydic ribonucleotide and amino acid precursors by photoredox chemistry. *Angew. Chem. Int. Ed. Engl.* *52*, 5845-5847
61. Powner, M. W., Gerland, B., and Sutherland, J. D. (2009). Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* *459*, 239-242
62. Ducat, D. C., and Silver, P. A. (2012). Improving carbon fixation pathways. *Curr. Opin. Chem. Biol.* *16*, 337-344
63. Ferris, J. P., and Nicodem, D. E. (1972). Ammonia photolysis and the role of ammonia in chemical evolution. *Nature* *238*, 268-269
64. McCombie, S. W., Motherwell, W. B. and Tozer, M. J. (2012). The Barton-McCombie Reaction. In *Organic Reactions*, S.E. Denmark, ed. (John Wiley & Sons, Inc.) Vol 77, 161-432.
65. Szostak, J. W. (2009). Origins of life: systems chemistry on early earth. *Nature* *459*, 171-172

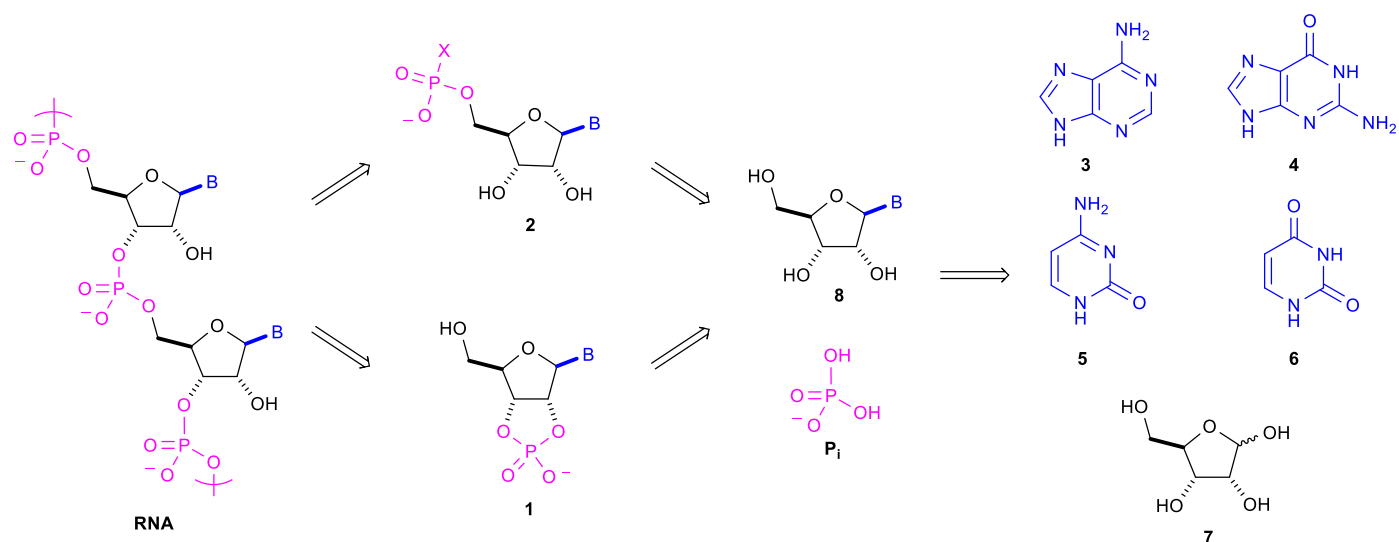
66. Xu, J., Tsanakopoulou, M., Magnani, C. J., Szabla, R., Šponer, J. E., Šponer, J., Góra, R. W., and Sutherland, J. D. (2016). A prebiotically plausible synthesis of pyrimidine  $\beta$ -ribonucleosides and their phosphate derivatives involving photoanomerization. *Nat. Chem.* doi:10.1038/nchem.2664
67. Powner, M. W., and Sutherland, J. D. (2010). Phosphate-mediated interconversion of ribo- and arabino-configured prebiotic nucleotide intermediates. *Angew. Chem. Int. Ed. Engl.* *49*, 4641-4643
68. Anastasi, C., Crowe, M. A., Powner, M. W., and Sutherland, J. D. (2006). Direct assembly of nucleoside precursors from two- and three-carbon units. *Angew. Chem. Int. Ed. Engl.* *45*, 6176-6179
69. Islam, S., Aguilar, J. A., Powner, M. W., Nilsson, M., Morris, G. A., and Sutherland, J. D. (2013). Detection of potential TNA and RNA nucleoside precursors in a prebiotic mixture by pure shift diffusion-ordered NMR spectroscopy. *Chem.- Eur. J.* *19*, 4586-4595
70. Sanchez, R. A., and Orgel, L. E. (1970). Studies in prebiotic synthesis. V. Synthesis and photoanomerization of pyrimidine nucleosides. *J. Mol. Evol.* *47*, 531-543
71. Choudhary, A., Kamer, K. J., Powner, M. W., Sutherland, J. D., and Raines, R. T. (2010). A Stereoelectronic effect in prebiotic nucleotide dyntesis. *ACS Chem. Biol.* *5*, 655-657
72. Powner, M. W., Anastasi, C., Crowe, M. A., Parkes, A. L., Raftery, J., and Sutherland, J. D. (2007). On the prebiotic synthesis of ribonucleotides: photoanomerisation of cytosine nucleosides and nucleotides revisited. *ChemBioChem* *8*, 1170-1179
73. Powner, M. W., and Sutherland, J. D. (2008). Potentially prebiotic synthesis of pyrimidine  $\beta$ -D-ribonucleotides by photoanomerization/hydrolysis of  $\alpha$ -D-cytidine-2'-phosphate. *ChemBioChem* *9*, 2386-2387
74. Powner, M. W., Sutherland, J. D., and Szostak, J. W. (2010). Chemoselective multicomponent one-pot assembly of purine precursors in water. *J. Am. Chem. Soc.* *132*, 16677-16688
75. Ferris, J. P., and Orgel, L. E. (1966). An unusual photochemical rearrangement in the synthesis of adenine from hydrogen cyanide. *J. Am. Chem. Soc.* *88*, 1074-1074
76. Sanchez, R. A., Ferris, J. P., and Orgel, L. E. (1968). Studies in prebiotic synthesis: IV. Conversion of 4-aminoimidazole-5-carbonitrile derivatives to purines. *J. Mol. Evol.* *38*, 121-128

77. Schwartz, A. W. (2013). Evaluating the plausibility of prebiotic multistage syntheses. *Astrobiology* 13, 784-789
78. Springsteen, G., and Joyce, G. F. (2004). Selective derivatization and sequestration of ribose from a prebiotic mix. *J. Am. Chem. Soc.* 126, 9578-9583
79. Powner, M. W., Zheng, S. L., and Szostak, J. W. (2012). Multicomponent assembly of proposed DNA precursors in water. *J. Am. Chem. Soc.* 134, 13889-13895
80. Nagorski, R. W., and Richard, J. P. (2001). Mechanistic imperatives for aldose–ketose isomerization in water: specific, general base- and metal ion-catalyzed isomerization of glyceraldehyde with proton and hydride transfer. *J. Am. Chem. Soc.* 123, 794-802
81. Sagan, C., and Khare, B. N. (1971). Long-wavelength ultraviolet photoproduction of amino Acids on the primitive earth. *Science* 173, 417-420
82. Shapiro, R. (2007). A simpler origin for life. *Sci. Am.* 296, 46-53
83. Breslow, R., Ramalingam, V., and Appayee, C. (2013). Catalysis of glyceraldehyde synthesis by primary or secondary amino acids under prebiotic conditions as a function of pH. *Orig. Life Evol. Biosph.* 43, 323-329
84. Breslow, R., and Cheng, Z. L. (2010). L-amino acids catalyze the formation of an excess of D-glyceraldehyde, and thus of other D sugars, under credible prebiotic conditions. *Proc. Natl. Acad. Sci. USA* 107, 5723-5725
85. Weber, A. L. (2001). The sugar model: catalysis by amines and amino acid products. *Orig. Life Evol. Biosph.* 31, 71-86
86. Hein, J. E., Tse, E., and Blackmond, D. G. (2011). A route to enantiopure RNA precursors from nearly racemic starting materials. *Nat. Chem.* 3, 704-706
87. Knowles, J. R., and Albery, W. J. (1977). Perfection in enzyme catalysis: the energetics of triosephosphate isomerase. *Acc. Chem. Res.* 10, 105-111
88. Knowles, J. R. (1991). Enzyme catalysis: not different, just better. *Nature* 350, 121-124

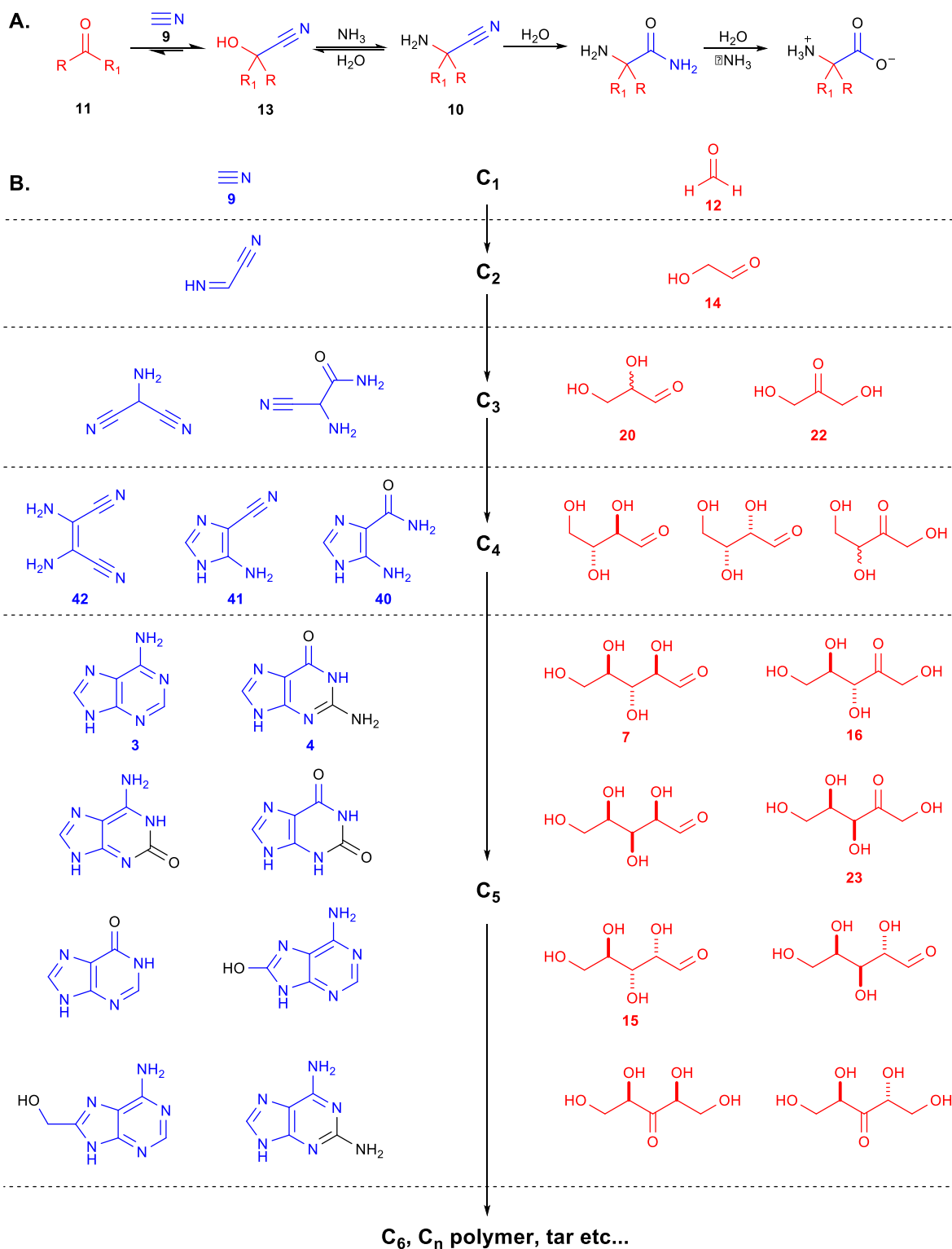
89. Richard, J. P. (1984). Acid-base catalysis of the elimination and isomerization reactions of triose phosphates. *J. Am. Chem. Soc.* *106*, 4926-4936
90. Potter, S., and Fothergill-Gilmore, L. A. (1993). Molecular evolution: The origin of glycolysis. *Biochem. Mol. Biol. Educ.* *21*, 45-48
91. Kreysing, M., Keil, L., Lanzmich, S., and Braun, D. (2015). Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length. *Nat. Chem.* *7*, 203-208
92. Strulson, C. A., Molden, R. C., Keating, C. D., and Bevilacqua, P. C. (2012). RNA catalysis through compartmentalization. *Nat. Chem.* *4*, 941-946
93. Adamala, K., and Szostak, J. W. (2013). Competition between model protocells driven by an encapsulated catalyst. *Nat. Chem.* *5*, 495-501
94. Jia, T. Z., Fahrenbach, A. C., Kamat, N. P., Adamala, K. P., and Szostak, J. W. (2016). Oligoarginine peptides slow strand annealing and assist non-enzymatic RNA replication. *Nat. Chem.* *8*, 915-921
95. Grochmal, A., Prout, L., Makin-Taylor, R., Prohens, R., and Tomas, S. (2015). Modulation of reactivity in the cavity of liposomes promotes the formation of peptide bonds. *J. Am. Chem. Soc.* *137*, 12269-12275
96. Murillo-Sánchez, S., Beaufils, D., González Mañas, J. M., Pascal, R., and Ruiz-Mirazo, K. (2016). Fatty acids' double role in the prebiotic formation of a hydrophobic dipeptide. *Chem. Sci.* *7*, 3406-3413
97. Adamala, K. P., Engelhart, A. E., and Szostak, J. W. (2016). Collaboration between primitive cell membranes and soluble catalysts. *Nat. Commun.* *7*, 11041
98. Tamura, K., and Schimmel, P. R. (2006). Chiral-selective aminoacylation of an RNA minihelix: Mechanistic features and chiral suppression. *Proc. Natl. Acad. Sci. USA* *103*, 13750-13752
99. Izgu, E. C., Björkbom, A., Kamat, N. P., Lelyveld, V. S., Zhang, W., Jia, T. Z., and Szostak, J. W. (2016). N-Carboxyanhydride-mediated fatty acylation of amino acids and peptides for functionalization of protocell membranes. *J. Am. Chem. Soc.* *138*, 16669-16676

100. Adamala, K., and Szostak, J. W. (2013). Nonenzymatic template-directed RNA synthesis inside model protocells. *Science* 342, 1098-1100

### FIGURE TITLE AND LEGENDS

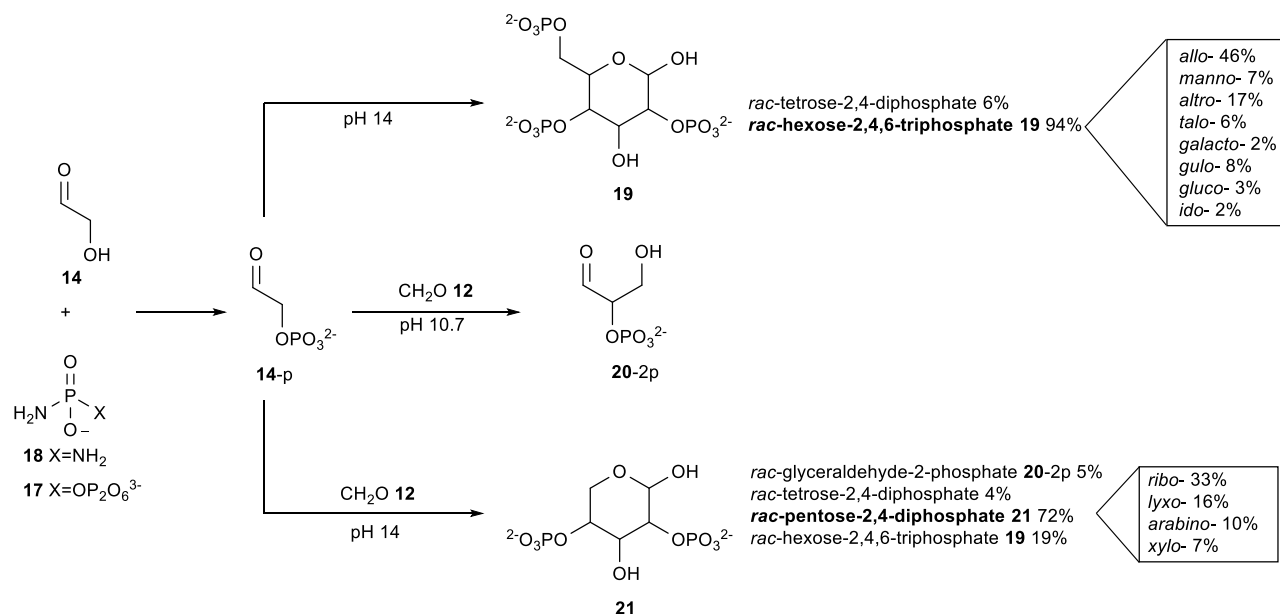


**Figure 1. The traditional retrosynthetic analyses disconnect RNA to ribofuranosyl sugar, inorganic phosphate and canonical RNA nucleobases.** The traditional disconnection of RNA is modular. Phosphodiester cleavage at the 5' or 3' hydroxyl moiety of RNA suggests that its prebiotic synthesis may proceed via the polymerization of activated ribonucleotide monomers **1** or **2** (X = nucleofuge). These nucleotides are then further disconnected to inorganic phosphate ( $P_i$ ) and  $\beta$ -ribose (**7**). These nucleosides are then further disconnection to inorganic phosphate ( $P_i$ ) and  $\beta$ -ribose (**7**). It is then assumed that nucleosides **8** are formed from the stereo- and regioselective  $N^9$ - or  $N^1$ -ribosylation of purine and pyrimidine bases by D-ribose **7** (shown in the minor equilibrium pentofuranosyl form). These analyses have failed to resolve prebiotic nucleotide synthesis for numerous reasons. Reactions of D-**7** with adenine **3** or guanine **4** produce low yields of the natural  $\beta$ -ribose nucleoside (**8**, B = **3** or **4**) as a wide mixture of regioisomers,  $\alpha$ - and  $\beta$ -anomers, and furanosyl and pyranosyl isomers. The ribosylation of cytosine **5** and uracil **6** is not permitted. Attempts to synthesize nucleobases **3-6** and ribose sugar D-**7** have also produced intractably complex mixtures (see Figure 2).



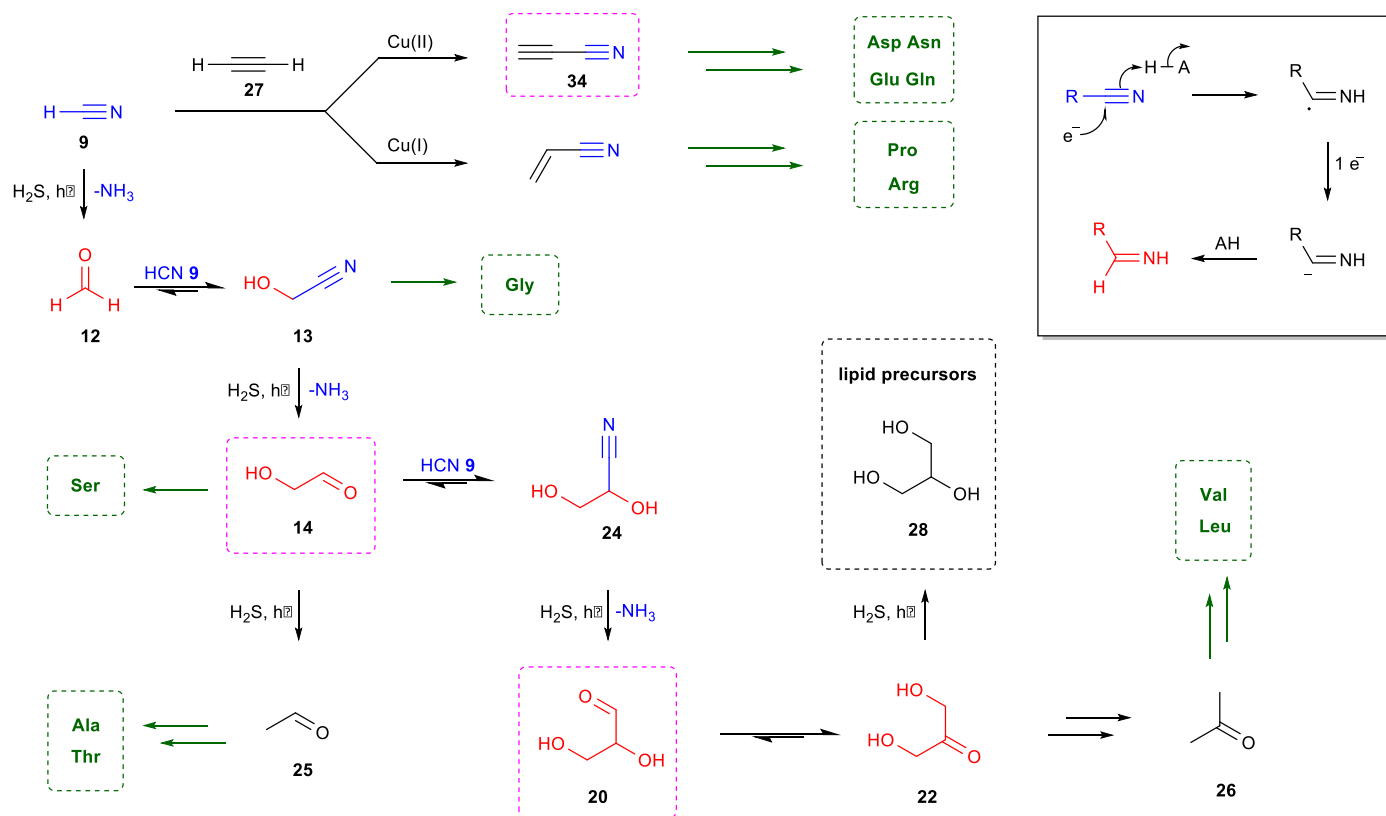
**Figure 2. The ‘three pillars’ of prebiotic chemistry.** The spark discharge aminonitrile synthesis, nucleobase synthesis by HCN oligomerization, and sugar synthesis by the formose reaction. **(A)** The Miller-Urey experiment is one of the earliest examples of prebiotic chemistry, producing molecules of biological relevance by passing a spark discharge through simple gases mixtures to simulate a potential early Earth atmosphere. The Miller-Urey synthesis results in a myriad of organic molecules, including a mixture of

proteinogenic and non-proteinogenic aminonitriles **10**, presumably upon *in situ* formation of aldehydes and ketones **11** and hydrogen cyanide **9**. The subsequent acid hydrolysis of Miller-Urey spark discharge products yields amino acids. **(B)** Oligomers of HCN **9** (nitrogenous chemistry; blue) and formaldehyde **12** (oxygenous chemistry; red). The oligomerization of **9** and **12** require spatiotemporal separation to produce the nucleobases and sugars that are ostensibly required for RNA synthesis to avoid the formation of cyanohydrin **13** from **9** and **12**, which is an overwhelmingly favourable process (shown in panel A). However, irrespective of this (hypothetical) spatiotemporal separation, high pH HCN-oligomerization produces intractable mixtures, which includes large amounts of insoluble polymer and very little of purines, such as adenine **3**, required for RNA. A selection of HCN-oligomers is shown in blue. High pH formaldehyde **12**-oligomerization is catastrophically complex, forming very little ribose **7** that is ostensibly required for RNA synthesis. The number of diastereoisomers and constitutional isomers produced increases rapidly upon homologation.

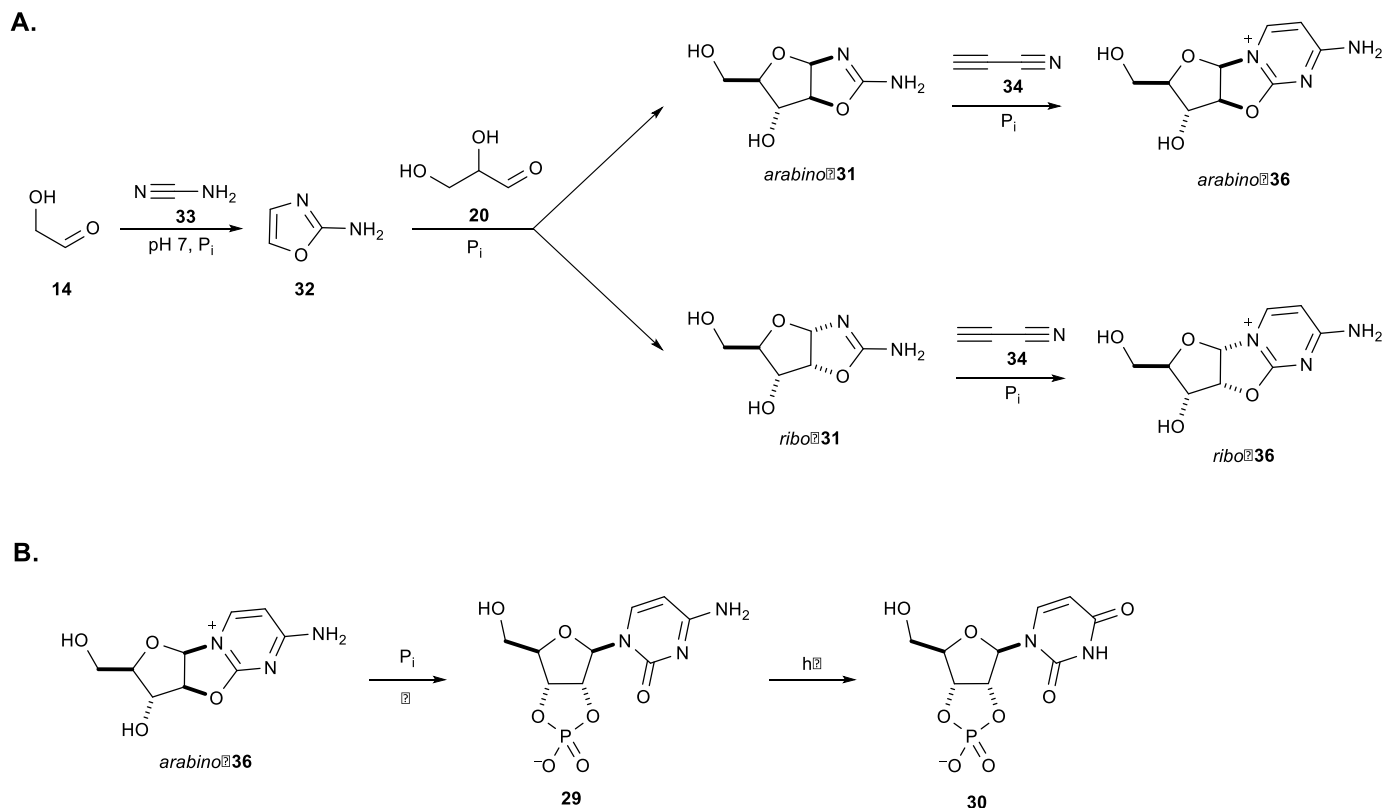


**Figure 3. Selective  $\alpha$ -phosphorylation of glycolaldehyde and aldol reactions of glycolaldehyde phosphate.** Glycolaldehyde **14** undergoes highly efficient and selective prebiotic  $\alpha$ -phosphorylation by amidophosphates **17** or **18**. Glycolaldehyde phosphate **14-P** undergoes efficient homoaldol condensation to produce a mixture of *rac*-hexose-2,4,6-triphosphates **19**. Hydromethylation of glycolaldehyde phosphate **14-P** with formaldehyde **12** produces glyceraldehyde-2-phosphate **20-2P** at pH 10.7. Under conditions conducive towards aldol reactions (pH 14), the **20-2P** synthesized undergoes a crossed aldol reaction with **14-P**, affording a mixture of sugar phosphates. In comparison to formose chemistry, these reactions are remarkable for their ‘erythroid’ selectivity, where *allo*-**19** and *ribo*-**21** predominate under kinetic control.

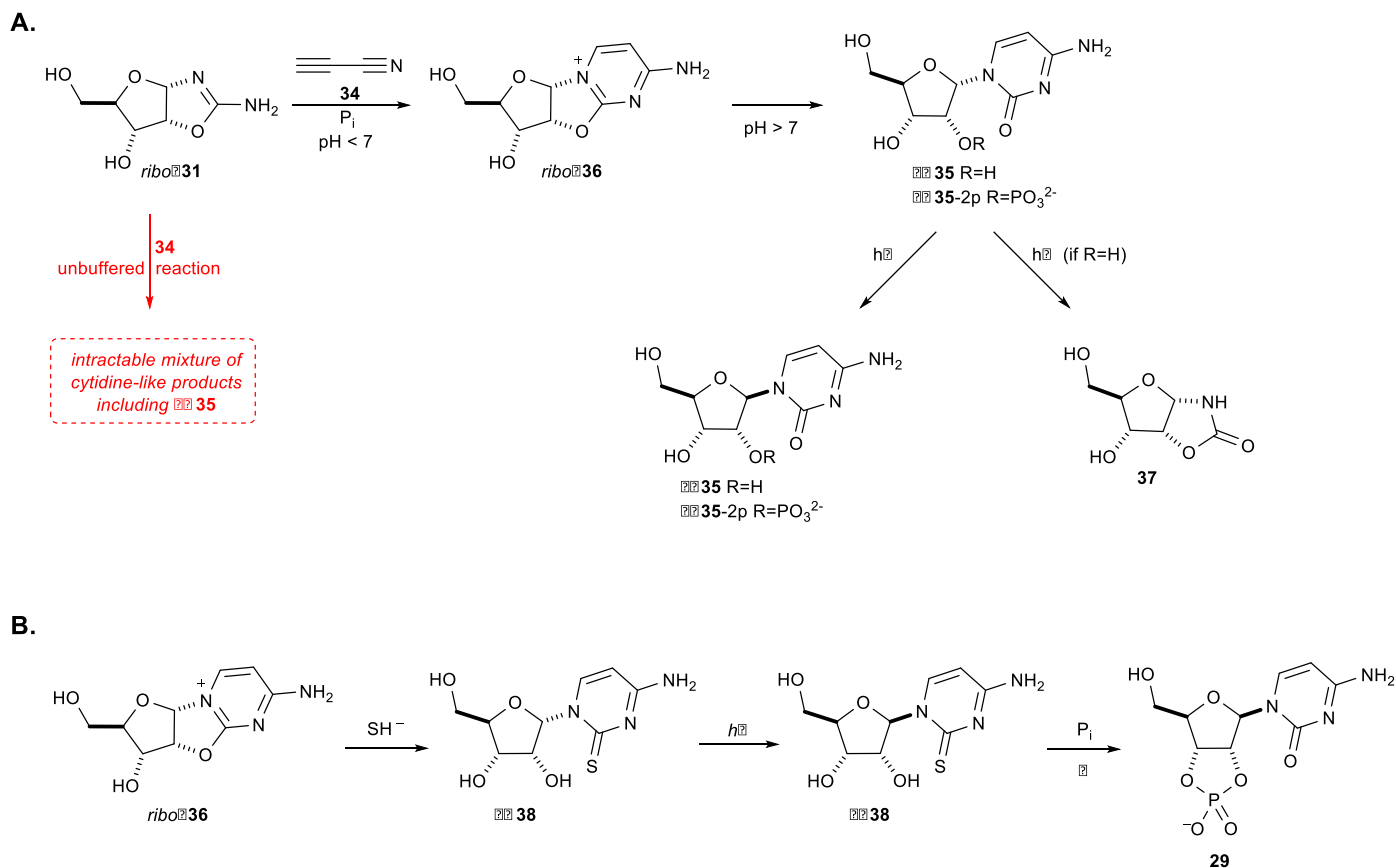




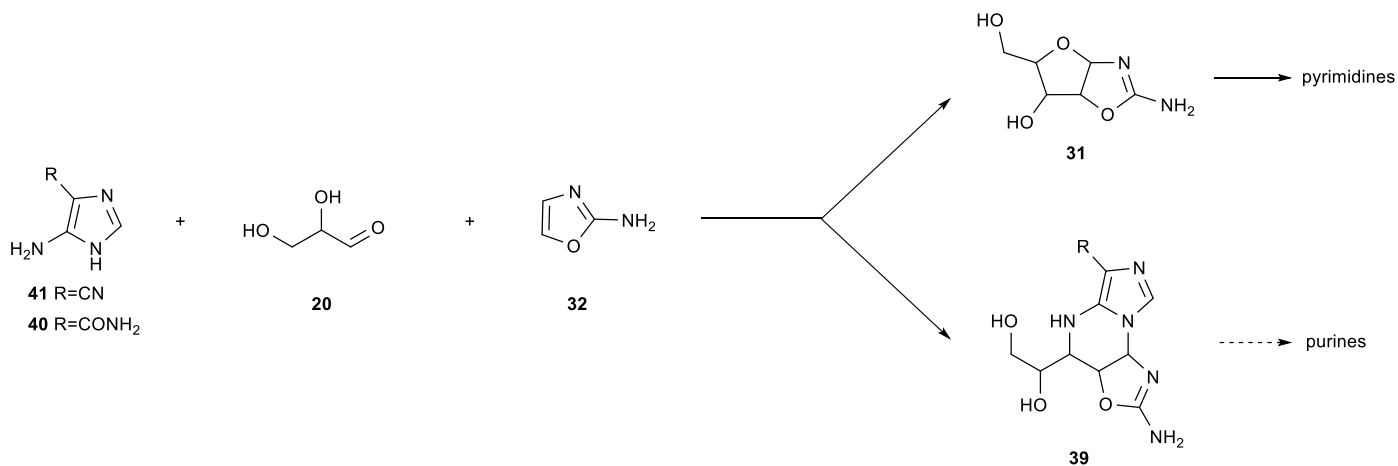
**Figure 4. Summary of Sutherland's cyanosulfidic protometabolism.** The building blocks of the nucleic acids, peptides and lipid membranes all emanate from a single carbon source, hydrogen cyanide **9**. The sequence starts by reduction of hydrogen cyanide **9** (inset, R = H). Iterative formaldehyde-homologation occurs by cyanohydrin **13** formation and reduction, producing glycolaldehyde **14** and glyceraldehyde **20** that are required for prebiotic ribonucleotide assembly. Glycolaldehyde **14** and glyceraldehyde **20** also serve as nodes to various amino acids and lipid precursor glycerol **28**. Copper-catalyzed cross-couplings of **9** with acetylene **27** allow access to numerous other amino acids, increasing the total number of proteinogenic amino acids produced by this proposed cyanosulfidic scenario to twelve. **AH** = general acid; **Gly** = glycine; **Ser** = serine; **Ala** = alanine; **Thr** = threonine; **Val** = valine; **Leu** = leucine; **Asn** = asparagine; **Asp** = aspartate; **Glu** = glutamate; **Gln** = glutamine; **Pro** = proline; **Arg** = arginine.



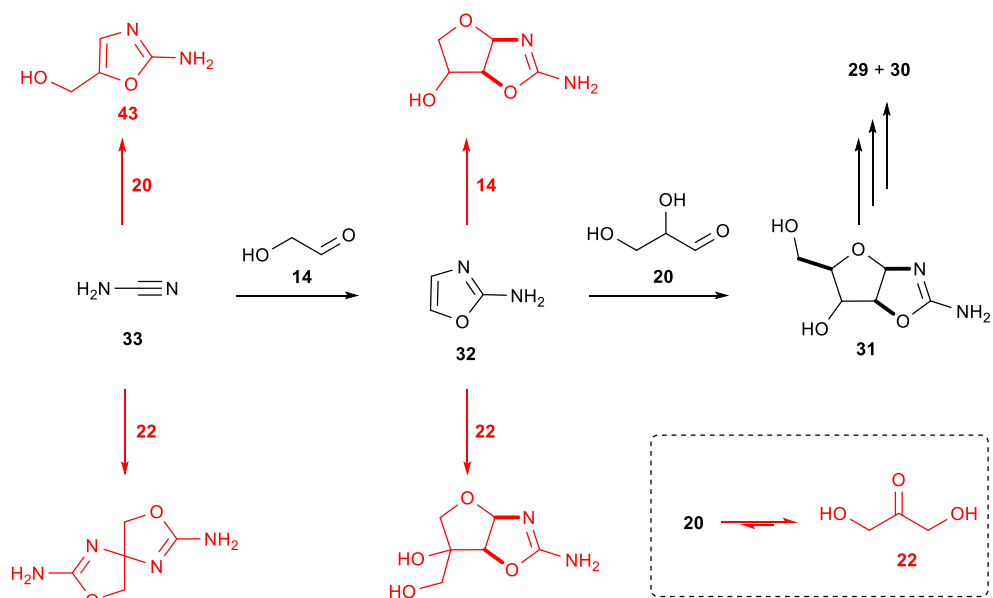
**Figure 5. Summary of the prebiotic syntheses of the activated pyrimidine ribonucleotides.** The long-standing problems with the synthesis and fusion of free nucleobases and D-ribose **7** were finally overcome by completely bypassing these with a mild and high yielding synthesis of pyrimidine ribonucleotides using a systems chemistry approach. **(A)** Controlled synthesis of anhydronucleosides, *ribo-36* and *arabino-36*, proceeds by stepwise assembly using C<sub>2</sub> and C<sub>3</sub> sugars only. The combination of mixed nitrogenous and oxygenous chemistry produces 2-aminooxazole **32**, which then reacts with glyceraldehyde **20** to produce pentose aminooxazolines **31** with the highest *ribo*- and *arabino*-selectivity observed in prebiotic chemistry. Elaboration of **31** to anhydronucleosides **36** by cyanoacetylene **34** is essentially quantitative in the presence of phosphate. **(B)** The *arabino-36* isomer is converted to the activated ribonucleotides by phosphorylation. Regioselective C3'-OH phosphorylation of *arabino-36* results in the thermodynamically-driven isomerization of the anhydronucleoside, which unravels to the ribonucleotide β-cytidine-2',3'-cyclic phosphate **29**. Photochemical sanitization destroys any unwanted by-products, but also produces β-uridine-2',3'-cyclic phosphate **30**. The *ribo-36* isomer can also be converted to ribonucleotides **29** and **30** via photochemical anomerization and prebiotic phosphorylation (see Figure 6).



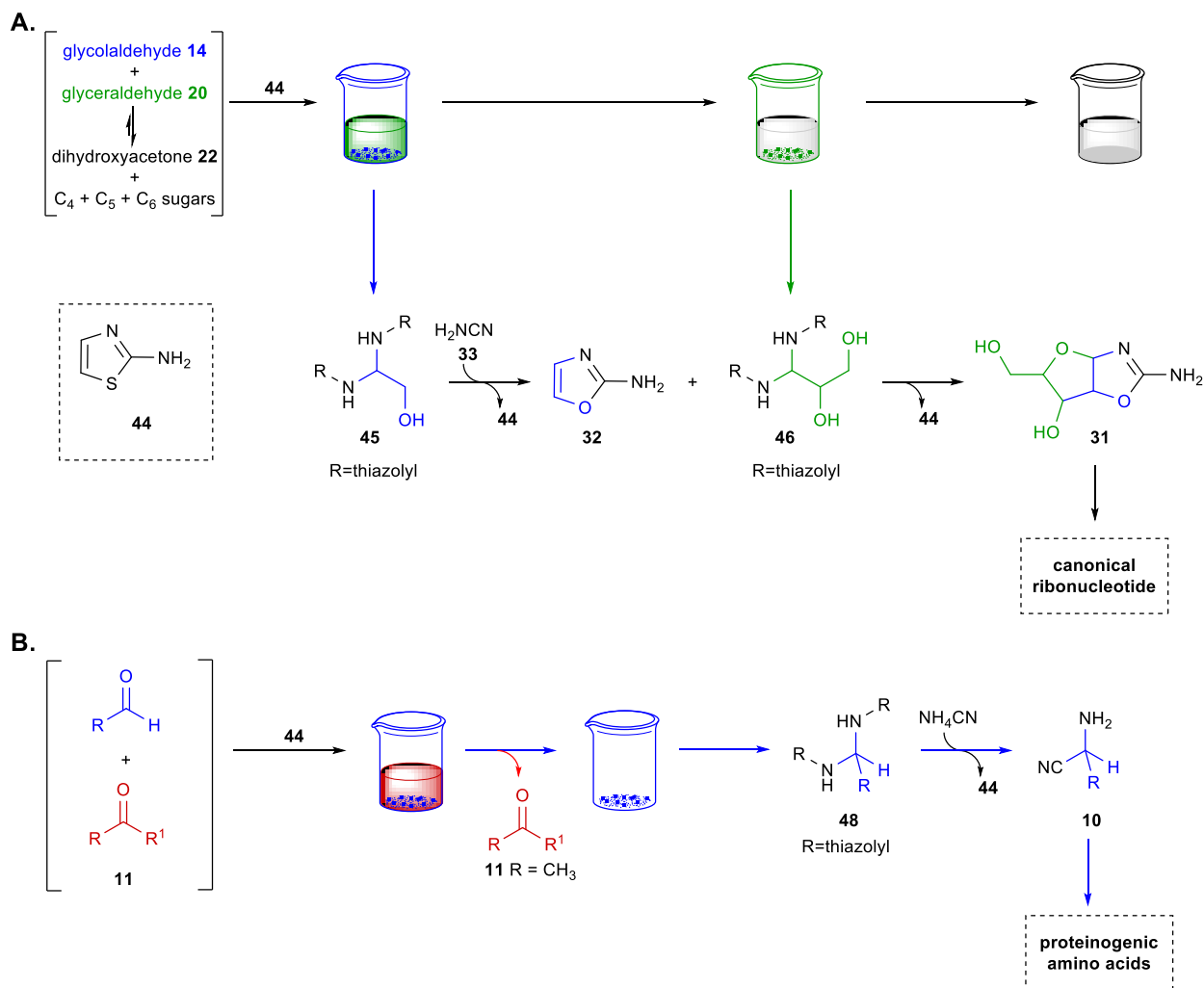
**Figure 6. Conversion of ribose amino oxazoline to activated pyrimidine ribonucleotides.** The obligate *cis*-1',2'-relationship of ribose amino oxazoline (*ribo*-31) dictates the stereospecific formation of  $\alpha$ -(thio)ribocytidines  $\alpha$ -35 or  $\alpha$ -38 that then require a photochemical  $\alpha \rightarrow \beta$  anomerization to yield canonical ribonucleotide 29. **(A)** The controlled synthesis of *ribo*-36 requires phosphate buffering to prevent the formation of a complex mixture of cytidine-like products (red). The anhydronucleoside *ribo*-36 undergoes hydrolysis to  $\alpha$ -ribocytidine  $\alpha$ -35 but attempts to photoanomerize the glycosidic bond are low yielding and occur alongside nucleobase loss and oxazolidinone 37 formation as major competing pathways. Irradiation of  $\alpha$ -cytidine-2'-phosphate  $\alpha$ -35-2P prevents these major degradative pathways and substantially improves photoanomerization ( $\alpha$ -35-2P  $\rightarrow$   $\beta$ -35-2P), but a chemospecific prebiotic route to  $\alpha$ -cytidine-2'-phosphate  $\alpha$ -35-2P is currently not known. **(B)** Thiolysis of *ribo*-36 yields  $\alpha$ -thioribocytidine  $\alpha$ -38. In stark contrast to  $\alpha$ -35,  $\alpha$ -38 underwent extraordinarily efficient C1'  $\alpha \rightarrow \beta$  epimerization to  $\beta$ -38. A one-step phosphorylation produces the activated ribonucleotide and converted the nucleobase to the canonical cytosine nucleobase. The enforced east-west conformation of nucleotide-2',3'-cyclic phosphates facilitates hydrolysis of the C2-thiocarbonyl to yield 29.



**Figure 7. Simultaneous pH-controlled multicomponent assembly of purine and pyrimidine nucleotide precursors.** HCN-tetramers AICA **40** and AICN **41** participate in a high yielding pH-dependent three-component reaction with glyceraldehyde **20** and 2-aminooxazole **32**. This produces potential purine ribonucleotide precursors **39**. The Mannich-type reactivity results in *N*<sup>9</sup>-purination with absolute regioselectivity. At pH 6–6.5, both purine **39** and pyrimidine **31** ribonucleotide precursors are observed, suggesting that a divergent synthesis of purine and pyrimidine ribonucleotides from within one pool of reagents is an enticing prospect.

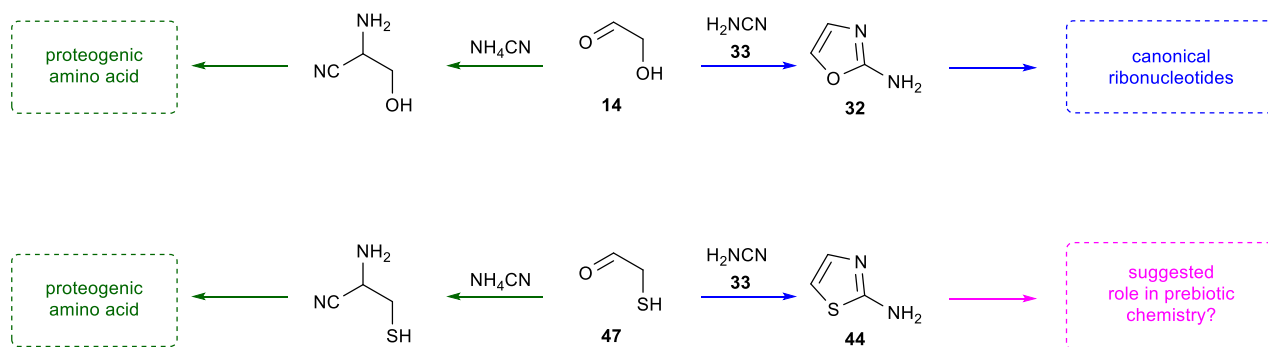


**Figure 8. The synthesis of activated ribonucleotides needs to overcome the formation of prebiotic ‘clutter’.** The synthesis of activated pyrimidine ribonucleotides **29** and **30** is dependent on the controlled formation of pentose aminooxazolines **31** (black), but the synthesis of **31** is wholly reliant on the ordered introduction of pure glycolaldehyde **14** (to cyanamide **33**) and glyceraldehyde **20** (to 2-aminooxazole **32**) to prevent the formation of numerous deleterious by-products (red). Ribonucleotide synthesis fails without the adherence to this order of synthetic steps. Glyceraldehyde **20** is highly susceptible to equilibration with dihydroxyacetone **22**, especially in phosphate buffer, which results in diminishing amounts of pentose aminooxazolines **31** being formed (inset).



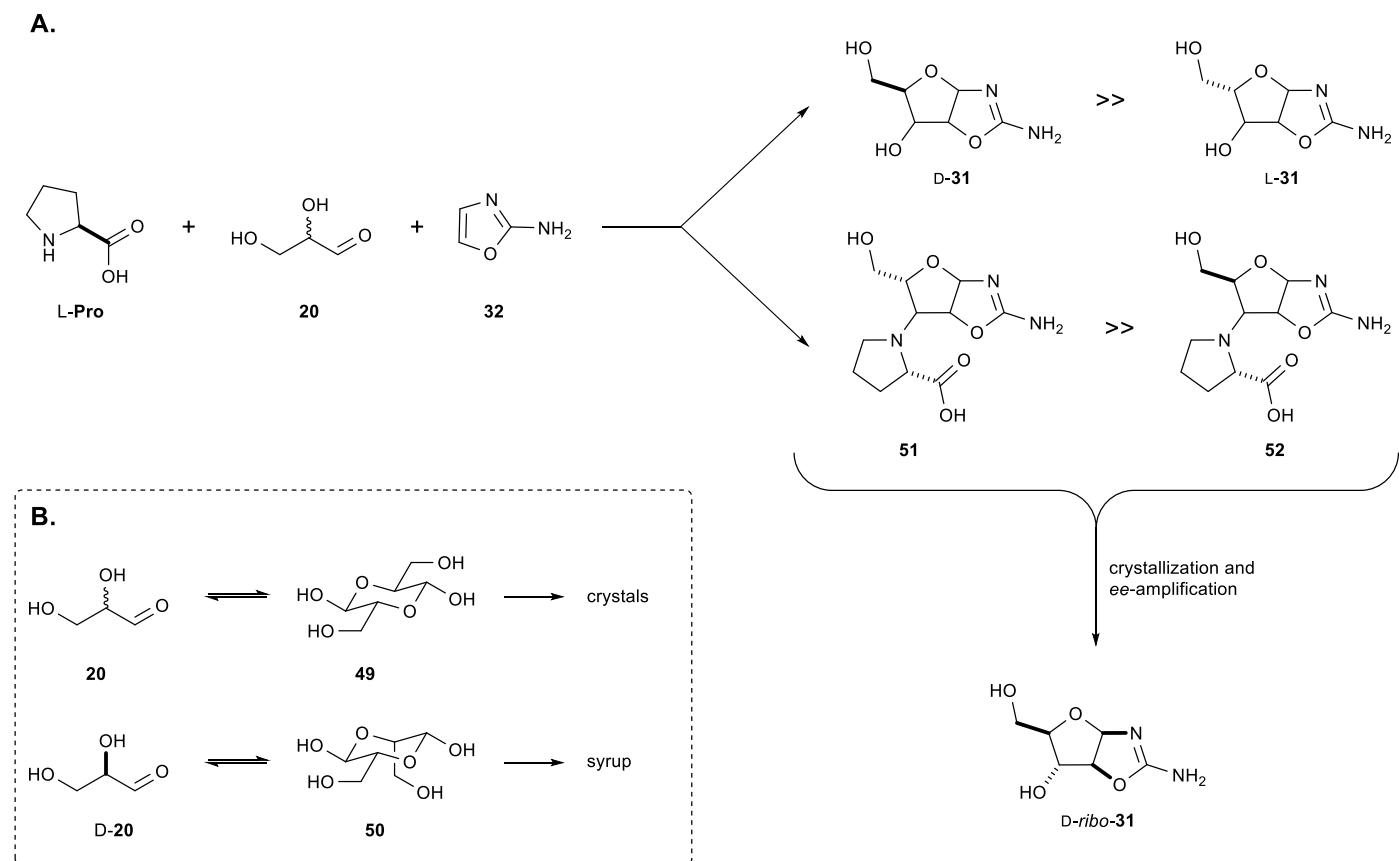
**Figure 9. Prebiotic selection of proteinogenic amino acids and natural nucleotides from complex mixtures through 2-aminothiazole-controlled aldehyde-sequestration.** The predisposed formation of 2-aminothiazole **44** from nucleotide and amino acid precursors, even within highly complex mixtures, provided a strong indication that the role of 2-aminothiazole **44** in the prebiotic chemical space had to be investigated in the context of both amino acid and nucleotide abiogenesis. This led to the discovery that 2-aminothiazole **44** is highly adept in controlling the selection and assembly of natural nucleic acid and proteinogenic amino acids. **(A)** Prebiotic selection of natural nucleotide precursors from a highly complex C<sub>2</sub>, C<sub>3</sub>, C<sub>4</sub>, C<sub>5</sub> and C<sub>6</sub> aldose and ketose sugar mixture. The crystallization-controlled synthesis of pentose aminooxazolines **31** is achieved by the initial sequestration of glycolaldehyde **14** as the stable C<sub>2</sub>-aminal **45** (blue), leaving behind a complex sugar mixture completely devoid of glycolaldehyde **14**. The slower sequestration of glyceraldehyde **20**, by trapping it as the C<sub>3</sub>-aminal **46** (green), overcomes the thermodynamic bias of C<sub>3</sub> sugars in favour of the ketose isomer dihydroxyacetone **22**. This apparently

‘unfavourable’ preference to yield dihydroxyacetone **22** at equilibrium turns out to be critical for ribonucleotide assembly; aldose-ketose isomerism retards C<sub>3</sub>-aminal **46** sequestration and therefore provides the required resolution of C<sub>2</sub> and C<sub>3</sub> sugars that is crucial to pentose aminooxazolines **31** synthesis from complex mixtures. Sequestration of amins **45** and **46** leaves behind a mixture of the residual C<sub>4</sub>, C<sub>5</sub> and C<sub>6</sub> sugars (black/grey), none of which crystallized upon prolonged incubation with **44**. The C<sub>2</sub>-aminal **45** reacts with cyanamide **33** to yield 2-aminooxazole **32**, and this mixture is then exposed to the C<sub>3</sub>-aminal **46** to produce pentose aminooxazolines **31**. Ribose aminooxazoline, *ribo-31*, precipitates from this mixture as conglomerate crystals. The 2-aminothiazole **44** regenerated can iterate the crystallization process indefinitely, rendering its unprecedented role in prebiotic chemistry as a highly effective traceless and catalytic ‘chaperone’ for multistep synthesis. **(B)** Chemoselective and high-yielding 2-aminothiazole **44**-driven crystallization of Strecker aldehydes (blue) from a complex mixture containing closely related prebiotic ketones (red). Crystallization of amins **48** (formed only from Strecker aldehydes (blue)) is coupled to a facile conversion to  $\alpha$ -aminonitriles **10**, providing a direct mechanism to exclude non-natural  $\alpha,\alpha$ -disubstituted amino acids (which are found extensively in carbonaceous chondrites and spark discharge experiments but not the proteome) from prebiotic amino acid synthesis.

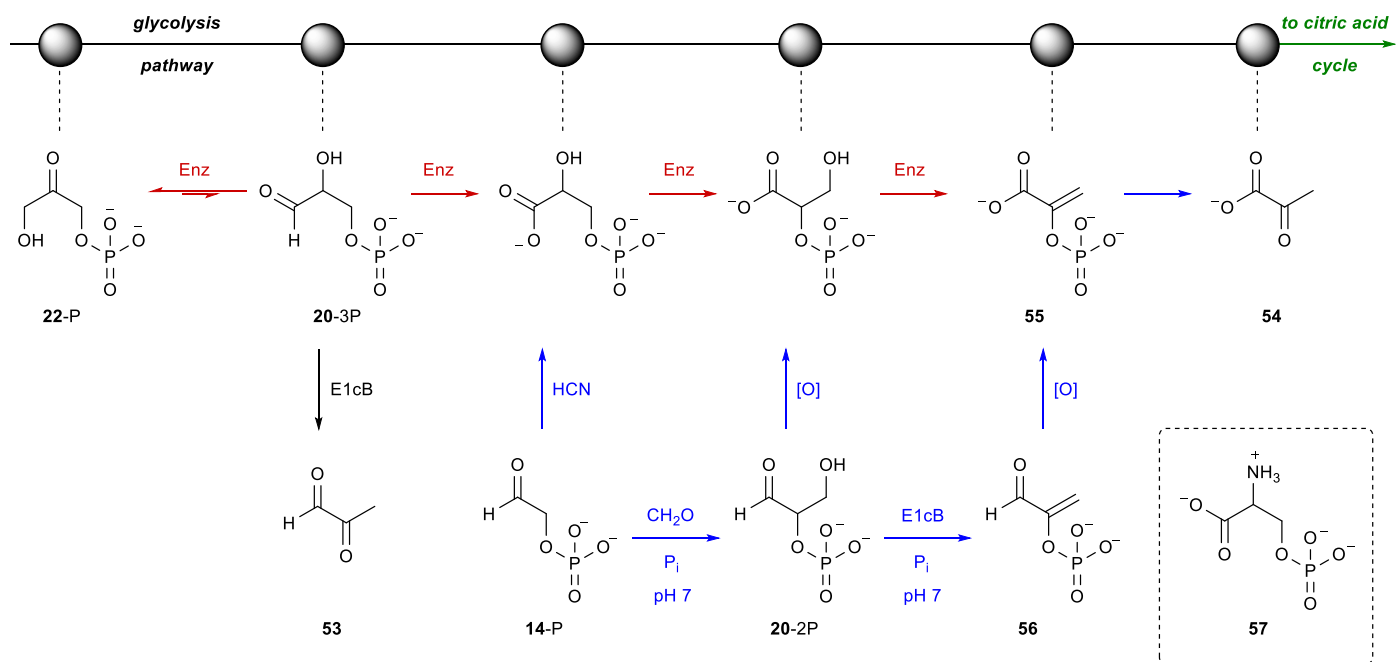


**Figure 10. Systems chemical analysis of amino acid and nucleotide syntheses.** Analysis of the prebiotic amino acid and nucleotide syntheses reveal that glycolaldehyde **14**—a serine and ribonucleotide precursor—lies at a generational node between these two metabolite classes. The same analysis applied to cysteine suggested that  $\beta$ -mercaptoacetaldehyde **47** would be of comparable importance to glycolaldehyde **14** and that the reactivity of 2-aminothiazole **44** may have key implications for the concomitant prebiotic synthesis of amino acid and nucleotides (see Figure 9).





**Figure 11. Strategies towards enantioenriched glycerinaldehyde and ribonucleotide precursors.** The synthesis of homochiral ribonucleotides appears to require prebiotically enantioenriched glycerinaldehyde **20** synthesis. **(A)** Resolution of glycerinaldehyde **20** can be achieved with enantiomerically pure L-proline (L-**Pro**) by trapping L-**20** as a three component product **51** during the reaction of *rac*-**20** with 2-aminooxazole **32**. Excess D-**20** can then undergo preferential two component reaction with 2-aminooxazole **32** to give enantioenriched D-pentose aminooxazolines D-**31**. **(B)** Racemic glycerinaldehyde *rac*-**20** can exist as dimer **49**. Dimer **49** accommodates all functional groups in an equatorial position around the dioxane ring, and so **49** readily forms a stable crystalline solid with a melting point of 145°C. Homochiral D-**20** requires one axial hydroxymethyl group around the 1,4-dioxane ring, resulting in a less favourable crystal packing. Accordingly, homochiral D (or L)-**20** exist as syrups at room temperature. Amplification of glycerinaldehyde **20** enantiomeric excesses can exploit these solubility differences.



**Figure 12. Prebiotic reconstruction of the triose glycolysis pathway by selective  $\alpha$ -phosphorylation of the simplest sugars.** Phosphoenolpyruvate **55**, nature's highest energy phosphate, is produced in a short sequence of mild reactions that only requires simple C<sub>1</sub> and C<sub>2</sub> building blocks. This process begins with formaldehyde **12** and glycolaldehyde phosphate **14-P**, the product of  $\alpha$ -phosphorylation of glycolaldehyde **14** and amidophosphates **17** or **18** (Figure 3). This primitive glycolysis pathway is notable for its robust and mild steps and the simple distribution of biologically important products. Triose phosphates **20-3P** and **22-P** undergo irreparable E1cB (Elimination Unimolecular conjugate Base) elimination of phosphate to produce methylglyoxal **53** in the absence of enzymatic control (red), thus limiting their potential for prebiotic utility. Conversely, glyceraldehyde-2-phosphate **20-2P** is predisposed to robustly yield the high-energy phosphoenol moiety that is central to the glycolysis.