The origin of RNA – a small part of the bigger picture

Frank R. Bowler (fbowler@mrc-lmb.cam.ac.uk)* and John D. Sutherland (johns@mrc-lmb.cam.ac.uk), MRC Laboratory of Molecular Biology, Cambridge, UK.

The astonishing diversity of life on Earth belies a fundamental uniformity in the chemistry that underpins it. These common molecular foundations have led many to conclude that the myriad species in existence today descend from a single ancestor: our "Last Universal Common Ancestor," or "LUCA". At a molecular level, LUCA would have shared many features with its modern-day progeny, including the central role of RNA in its biology, but whence LUCA? The ability of RNA to span the realms of both information storage (i.e., genotype) and catalytic function (i.e., phenotype) has spawned the theory of an earlier "RNA world," wherein RNA molecules catalysed their own replication. However, the recapitulation of an RNA-only system capable of self-replication and Darwinian evolution remains elusive. Furthermore, it is difficult to envision a mechanism whereby an RNA world could transition to incorporate the DNA and proteins of extant biochemistry.

It therefore seems prudent to consider alternatives to the purist RNA world model, which incorporate other macromolecules and small molecules. Akin to modern biochemistry, it is entirely possible (if not probable) that synergistic relationships will exist between these different molecular species, both regarding their assembly from simple (prebiotically available) feedstock chemicals and their ultimate role in a nascent living system. However, parsimony dictates that our search for the chemistry of life's origin should be informed by the molecules (and their probable antecedents) that we observe in today's living systems; otherwise, the chemical space that must be explored becomes boundless, and the problem insuperable.

In our laboratory, we take a "systems chemistry" approach to study the chemical origins of the (oligo)nucleotides, peptides, (vesicle-forming) amphiphiles and metabolites that are essential for life as we know it [1]. We utilise the power of modern analytical techniques (principally, but not exclusively, NMR spectroscopy) to probe "subsystems" of well-defined multi-component reactions, or one-pot multi-step sequences, constituting prebiotically available small molecules under conditions consistent with those on the early Earth. The ultimate assembly of these subsystems will yield an overall system that simultaneously and chemoselectively generates a range of biologically relevant molecules; dependent upon its emergent properties, the set or sequence of conditions required by this system will point towards likely geochemical scenarios for the origin of life.

Thus, a plausible chemical origin of the pyrimidine (i.e., uracil and cytosine) ribonucleotides was demonstrated through the application of systems chemistry [2]. Previously, it had been assumed that nucleotides – the monomeric building blocks of longer oligo- or polymeric RNA – must have arisen by phosphorylation of nucleosides, which were in turn generated by the reaction of pre-formed (nitrogenous) nucleobases and (oxygenous) ribose. However, the reaction to join ribose and a nucleobase works either extremely poorly (for the purine nucleobases, guanine and adenine) or not at all (for the pyrimidines). Now, it is known that simple prebiotically available small molecules can assemble in mixed nitrogenous-oxygenous systems to afford the pyrimidine ribonucleotides in good yields as their 2′,3′-cyclic phosphates.

Although the prebiotic synthesis of (pyrimidine) nucleoside-2′,3′-cyclic phosphates (N>Ps) can be viewed as predisposed, a route from these weakly activated monomers (and their purine counterparts) to the RNA polymers of extant biochemistry is still required. However, in aqueous solution, N>Ps do not react to give RNA – rather, they hydrolyze to their corresponding 2′- and 3′-monophosphates. That said, the condensation of N>Ps to short RNA oligomers (maximally 14 nucleotides long) has been shown to occur in the dry state following evaporation of aqueous solutions containing prebiotically plausible acid/base catalysts [3]. Some cyclic phosphate hydrolysis occurs, so that the resulting oligomers possess mixed 2′- and 3′-monophosphates at one end of the molecule. One route to longer RNA polymers might therefore be through the template-directed ligation of these oligomers in aqueous solutions following electrophilic activation of the terminal phosphate, but there is a problem: if the 2′- or 3′-phosphate of an oligonucleotide is activated in this way, it will only serve to regenerate a 2′,3′-cyclic phosphate, owing to the high effective molarity of the neighboring 3′- or 2′-hydroxyl group relative to the 5'-hydroxyl of the adjacent oligomer. These 2′,3′-cyclic phosphate-terminated oligomers can then only ligate very slowly, to give the ”wrong” linkage isomer: that is, a 2′,5′-internucleotide linkage, as opposed to the 3′,5′-internucleotide bonds that exclusively constitute RNA oligomers in contemporary biology.

In recent work, we uncovered a prebiotically plausible chemical system that facilitates the rapid ligation of these 2′- and 3′-phosphate-terminated oligomers (Figure 1) [4]. Specifically, we found that (oxidatively or electrophilically) activated acetyl groups can regio- and chemoselectively protect the terminal 2′-hydroxyl group of an (oligo)nucleotide possessing a 3′-phosphate. This then permits rapid template-directed ligation with a neighbouring RNA oligomer. Remarkably, protection of the 3′-hydroxyl group of a 2′-phosphorylated (oligo)nucleotide is less efficient, which favors the generation of natural 3′,5′-linkages from ligation reactions containing mixed 2′/3′-phosphate-terminal oligomers. Whilst some linkage heterogeneity might have been tolerated or even beneficial for early (non-enzymatic) RNA replication, the transition to homogeneously 3′,5′-linked extant RNA is likely to have been easier if the starting point was RNA already enriched in these linkages. Intriguingly, this acetylation-facilitated ligation chemistry hints at a link between two previously distinct models for the origin of life: the RNA world scenario, and the metabolism-first ”iron-sulfur world” capable of generating activated acetyl groups on the surface of iron-nickel sulfide minerals.

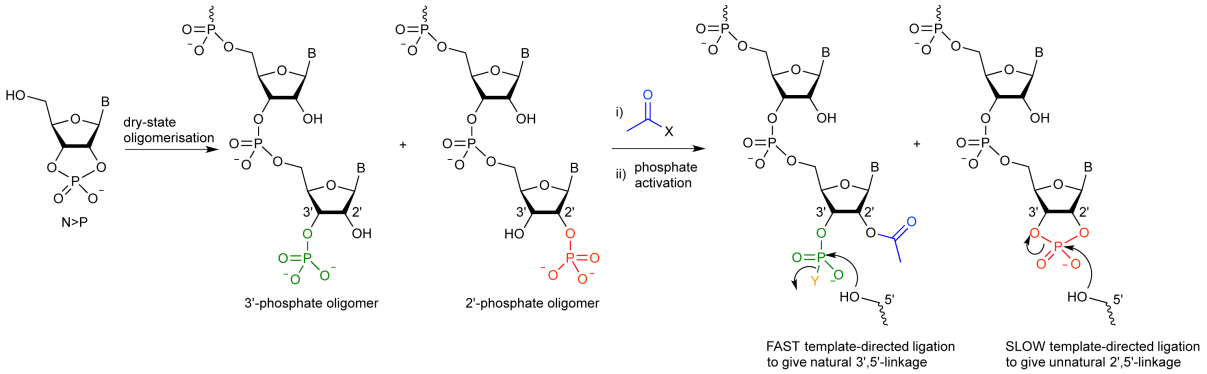

Figure 1. The dry-state oligomerisation of nucleoside 2′,3′-cyclic phosphates (N>P) yields short oligomers with mixed 3′- and 2′-phosphate termini. Chemoselective acetylation permits the rapid, template-directed ligation of those oligomers with a 3′-phosphate. (Note that mixed internucleotide linkages – specifically a 2:1 ratio of 3′,5′:2′,5′-linkages – result from the original dry-state condensation, although 3′,5′-linkages (shown) will be favored in subsequent template-directed ligation reactions by virtue of their greater hydrolytic stability in the

context of an RNA:RNA (A-form) duplex, and the reduced thermodynamic stability of duplexes containing 2',5'-linkages.) X, Y = leaving groups; B = nucleobase.

In a further development of prebiotic systems chemistry in our laboratory, it has been found that the simple two- and three carbon sugars glycolaldehyde and glyceraldehyde emerge from the photoredox cycling of copper cyanide complexes [5]. These sugars are requisite building blocks for the synthesis of the pyrimidine ribonucleotides, and are synthesised using hydrogen cyanide as the sole source of carbon. In the original system, glycolaldehyde and glyceraldehyde were observed as their cyclic cyanate adducts (i.e., oxazolidinones); however, the introduction of hydrogen sulfide into the system as the ultimate reductant yielded the (hydrated) free aldehydes [6]. Through iterative cycles of cyanohydrin formation, reduction and imine hydrolysis, together with an observed deoxygenation of glycolaldehyde, the aldehydic (Strecker) precursors to the amino acids glycine, serine, alanine and (*allo*-)threonine were observed. Thus, at least four of the proteinogenic amino acids in extant biochemistry may have emerged from a simple chemical system that also generated the sugar precursors necessary for ribonucleotide synthesis.

Life is chemistry – but even in its most basic form, life depends upon the intricate interplay of many complex chemical subsystems. It is therefore through the study of prebiotically plausible chemical subsystems, and the phenomena that emerge therefrom, that we can best approach a fundamental understanding of life's chemical origins. Recent work in our laboratory has further highlighted the need for an open mind when considering the possibility that the chemical conditions necessary for RNA abiogenesis may have simultaneously given rise to other biologically relevant molecules and metabolic forerunners, on which RNA replication and function may have depended in a nascent living system.

References:

1. Powner, M. W. & Sutherland, J. D. Prebiotic chemistry: a new *modus operandi*. *Phil. Trans. R. Soc. B* **366,** 2870-2877 (2011).
2. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459,** 239-242 (2009).
3. Verlander, M. S., Lohrmann, R. & Orgel, L. E. Catalysts for the self-polymerization of adenosine cyclic 2',3'-phosphate. *J. Mol. Evol.* **2,** 303-316 (1973).
4. Bowler, F. R., Chan, C. K. W., Duffy, C. D., Gerland, B., Islam, S., Powner, M. W., Sutherland, J. D. & Xu, J. Prebiotically plausible oligoribonucleotide ligation facilitated by chemoselective acetylation. *Nat. Chem.* **5,** 383-389 (2013).
5. Ritson, D. & Sutherland, J. D. Prebiotic synthesis of simple sugars by photoredox systems chemistry. *Nat. Chem.* **4,** 895-899 (2012).
6. Ritson, D. J. & Sutherland, J. D. Synthesis of aldehydic ribonucleotide and amino acid precursors by photoredox chemistry. *Angew. Chem. Int. Ed.* **52,** 5845-5847 (2013).