# Origins of life: Systems chemistry on early Earth

**Article** *in* Nature · June 2009

DOI: 10.1038/459171a · Source: PubMed



**1 author:**

Jack Szostak
Howard Hughes Medical Institute
**367** PUBLICATIONS **36,680** CITATIONS

SEE PROFILE

NEWS & VIEWS

ORIGINS OF LIFE

# Systems chemistry on early Earth

Jack W. Szostak

**Understanding how life emerged on Earth is one of the greatest challenges facing modern chemistry. A new way of looking at the synthesis of RNA sidesteps a thorny problem in the field.**

It is well established that the evolution of life passed through an early stage in which RNA played central roles in both inheritance and catalysis[1] — roles that are currently played by DNA and protein enzymes, respectively. But where did the RNA come from?

Experiments reported by Powner *et al.*[2] (page 239 of this issue) provide fresh insight into the chemical processes that might have led to the emergence of information-coding nucleic acids on early Earth.

For 40 years, efforts to understand the prebiotic synthesis of the ribonucleotide building blocks of RNA have been based on the assumption that they must have assembled from their three molecular components: a nucleobase (which can be adenine, guanine, cytosine or uracil), a ribose sugar and phosphate. Of the many difficulties encountered by those in the field, the most frustrating has been the failure to find any way of properly joining the pyrimidine nucleobases — cytosine and uracil — to ribose[3] (Fig. 1a). The idea that a molecule as complex as RNA could have assembled spontaneously has therefore been viewed with increasing scepticism. This has led to a search for alternative, simpler genetic polymers that might have preceded RNA in the early history of life.

But Powner *et al.*[2] revive the prospects of the 'RNA first' model by exploring a pathway for pyrimidine ribonucleotide synthesis in which the sugar and nucleobase emerge from a common precursor (Fig. 1b). In this pathway, the complete ribonucleotide structure forms without using free sugar and nucleobase molecules as intermediates. This central insight, combined with a series of additional innovations, provides a remarkably efficient solution to the problem of prebiotic ribonucleotide synthesis.

The key to Powner and colleagues' approach was to overcome the deeply ingrained prejudice that carbon–oxygen chemistry (which leads to sugar formation) and carbon–nitrogen chemistry (which leads to nucleobase formation) should be kept separate for as long as possible. One does not have to look far to find the source of this prejudice. Incubation of formaldehyde — a simple carbon–oxygen compound — in alkaline solution rapidly yields a mixture of dozens of sugars[3], which subsequently react to yield an intractable tar of insoluble products. Similarly, simple carbon–nitrogen compounds, derived from cyanide and ammonia, react with each other to generate not only the standard nucleobases, but also many other compounds. It is perfectly reasonable to expect that uncontrolled mixing of these two complex processes would lead to a chemical combinatorial explosion: the synthesis of millions of different organic compounds, of which the desired biological precursor molecules would be a vanishingly small fraction. But in a remarkable example of 'systems chemistry', in which reactants from different stages of a pathway are allowed to interact, Powner *et al.*[2] show that phosphate tames the combinatorial explosion, allowing oxygenous and nitrogenous reactants to interact fruitfully.

The authors' path to RNA begins with the same starting materials used in many recent studies of prebiotic chemistry, but differs in the order in which they are combined. When the structurally simplest sugar, glycolaldehyde, reacts with the simplest derivative of cyanide and ammonia, cyanamide, a complex mixture of undesired compounds is formed. But Powner *et al.* add a third ingredient — phosphate — to the mix. In their reaction, phosphate acts as both a pH buffer and a catalyst, thereby short-circuiting the network of possible unwanted reactions and leading instead to the fast, efficient synthesis of a key intermediate known as 2-aminooxazole (Fig. 1b).

One of the goals of those developing theories of prebiotic chemistry is to identify geochemically plausible means of purifying key intermediates away from contaminants that might cause trouble in later reactions. The remarkable volatility of 2-aminooxazole suggests that it could be purified by sublimation, as it undergoes cycles of gentle warming from the sun, cooling at night (or at higher altitudes) and subsequent condensation. The compound would thus behave as a kind of organic snow, which could accumulate as a reservoir of material ready for the next step in RNA synthesis.

Phosphate continues to have several essential roles in the remaining steps of Powner and colleagues' pathway, in one case causing depletion of an undesired by-product, and in another

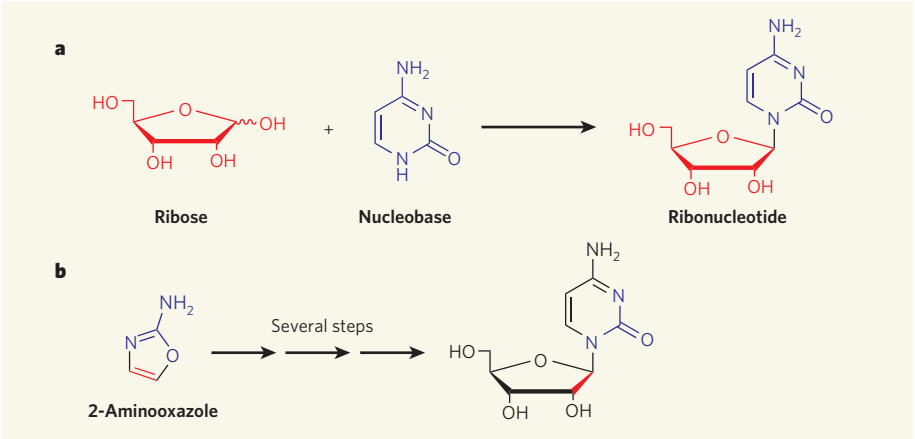


**Figure 1 | Theories of prebiotic syntheses of pyrimidine ribonucleotides.** The idea that RNA might have formed spontaneously on early Earth has inspired a search for feasible prebiotic syntheses of ribonucleotides, the building blocks of RNA. **a**, The traditional view is that the ribose sugar and nucleobase components of ribonucleotides formed separately, and then combined. But no plausible reactions have been found in which the two components could have joined together. **b**, Powner *et al.*[2] show that a single 2-aminooxazole intermediate could have contributed atoms to both the sugar and nucleobase portions of pyrimidine ribonucleotides, so that components did not have to form separately. For a more detailed overview of the pathways depicted here, see Figure 1 on page 239.

saving a critical intermediate from degradation. The penultimate reaction of the sequence, in which the phosphate is attached to the nucleoside, is another beautiful example of the influence of systems chemistry in this set[2] of interlinked reactions. The phosphorylation is facilitated by the presence of urea[4]; the urea comes from the phosphate-catalysed hydrolysis of a by-product from an earlier reaction in the sequence.

The authors wrap up their synthetic tour de force by using ultraviolet light to clean up the reaction mixture. They report that ultraviolet irradiation destroys side products while simultaneously converting some of the desired ribocytidine product to ribouridine (the second pyrimidine component of RNA). The development of this complex photochemistry required remarkable mechanistic insight from Powner and colleagues, who not only correctly predicted that ultraviolet irradiation would destroy the majority of the by-products, but also that the desired ribonucleotides would withstand such treatment.

The authors' careful study[2] of every potentially relevant reaction and side reaction in their sequence is a model of how to develop the fundamental chemical understanding required for a reasoned approach to prebiotic chemistry. By working out a sequence of efficient reactions, they have set the stage for a more fruitful investigation of geochemical scenarios compatible with the origin of life.

Of course, much remains to be done. We must now try to determine how the various starting materials could have accumulated in a relatively pure and concentrated form in local environments on early Earth. Furthermore, although Powner and colleagues' synthetic sequence yields the pyrimidine ribonucleotides, it cannot explain how purine ribonucleotides (which incorporate guanine and adenine) might have formed. But it is precisely because this work opens up so many new directions for research that it will stand for years as one of the great advances in prebiotic chemistry. ■

Jack W. Szostak is in the Howard Hughes Medical Institute and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
e-mail: szostak@molbio.mgh.harvard.edu

1. Joyce, G. F. & Orgel, L. E. in *The RNA World* (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 23–56 (Cold Spring Harbor Laboratory Press, 2006).
2. Powner, M. W., Gerland, B. & Sutherland, J. D. *Nature* **459,** 239–242 (2009).
3. Orgel, L. E. *Crit. Rev. Biochem. Mol. Biol.* **39,** 99–123 (2004).
4. Lohrmann, R. & Orgel, L. E. *Science* **171,** 490–494 (1971).

MOLECULAR MICROBIOLOGY

# A key event in survival

Dave Barry and Richard McCulloch

**The parasitic microorganism *Trypanosoma brucei* evades recognition by its host's immune system by repeatedly changing its surface coat. The switch in coat follows a risky route, though: DNA break and repair.**

Like many other single-celled pathogens, the protozoan *Trypanosoma brucei*, which causes African sleeping sickness in humans, undergoes antigenic variation — that is, it periodically switches its variant surface glycoprotein (VSG), the molecule targeted by host antibodies. But how switching is triggered has remained largely elusive. On page 278 of this issue, Boothroyd *et al.*[1] show that a DNA double-strand break (DSB) upstream of the *T. brucei VSG* gene is the likely primary event in this process. Their results add to the few, albeit crucial, cases in which DSBs trigger developmental processes: these include mating-type switching in yeast, rearrangements of immune-system genes in humans and meiotic cell division to produce sex germ cells[2].

Antigenic switching can occur through several genetic strategies, the most common being the differential activation of an archive of silent genes and pseudogenes. Although only one gene is transcribed, from a specialized expression site, switching occurs when silent genes, or their fragments, are duplicated in the expression site by a gene-conversion process, replacing all or part of the expressed gene. In some pathogens, the expressed gene can be constructed as a mosaic from several archival pseudogenes; such a combinatorial strategy expands the scale of variation enormously, with, for example, five pseudogenes giving rise to hundreds of combinations[3].

*Trypanosoma brucei* has evolved an even more staggeringly complex system. It, too, transcribes a single *VSG* gene, but the sources of sequences that contribute to switching are large and diverse. It has several inactive expression sites, and its archive contains up to 200 *VSG* genes that lie at the ends (telomeres) of a set of mini-chromosomes, as well as a further 1,600 silent genes — of which two-thirds are pseudogenes — on the main chromosomes[4]. The potential for mosaic variation therefore seems beyond estimation. Intact archival genes are duplicated starting from an upstream set of repeat sequences each 70 base pairs (bp) long[5], all the way to sequences at the downstream end of the coding sequence, or, in the case of silent telomeric genes, perhaps to the nearby end of the chromosome. As gene conversion in other organisms is initiated by a DSB in the conversion site, such a break has been proposed also to occur in the *T. brucei VSG*

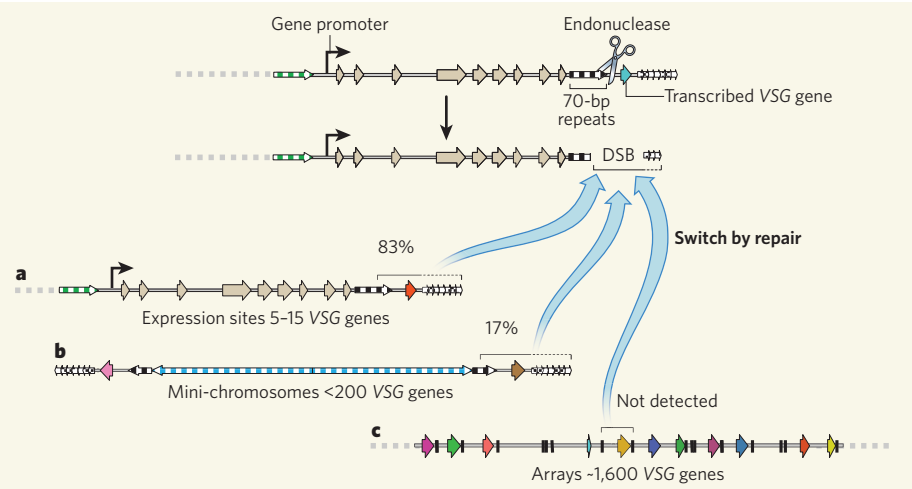


**Figure 1 | Antigenic switching and sources.** Boothroyd *et al.*[1] used an endonuclease enzyme to induce a DNA double-strand break (DSB) adjacent to the 70-bp-repeat region of the active *VSG* gene in *Trypanosoma brucei*. Consequently, the region from the DSB site to the end of the *VSG* gene was deleted. The protozoan filled this gap by a repair process, using silent *VSG* loci on other chromosomes as template. Locations of donor sequences included: (**a**) expression sites (of which there are 5–15 per strain) at the telomeres of the main chromosomes; (**b**) telomeres of some 100 mini-chromosomes found in the *T. brucei* genome; and (**c**) tandemly arrayed *VSG* genes in the main chromosomes. The copied regions stretched from the 70-bp-repeat regions to the telomere, or, for intact genes, to the end of the *VSG*. The frequencies of conversions the authors detected (shown as percentages) differ from those observed during infections with natural strains of *T. brucei*, in which mini-chromosomes dominate as donors. Brackets denote the duplicated region, with dashed sections indicating uncertainty over where the duplication ends. Broad arrows indicate genes; narrow arrows, repetitive DNA sequences (70-bp repeats are shown in black and white). Coloured arrows are different *VSG* genes; grey arrows, genes other than *VSG*.