OPINION

Studies on the origin of life — the end of the beginning

John D. Sutherland

Abstract | Understanding how life on Earth might have originated is the major goal of origins of life chemistry. To proceed from simple feedstock molecules and energy sources to a living system requires extensive synthesis and coordinated assembly to occur over numerous steps, which are governed only by environmental factors and inherent chemical reactivity. Demonstrating such a process in the laboratory would show how life can start from the inanimate. If the starting materials were irrefutably primordial and the end result happened to bear an uncanny resemblance to extant biology — for what turned out to be purely chemical reasons, albeit elegantly subtle ones — then it could be a recapitulation of the way that natural life originated. We are not yet close to achieving this end, but recent results suggest that we may have nearly finished the first phase: the beginning.

Broadly speaking, the origin of life can be approached by thinking from biology down or from chemistry up[1]. From biology down, phylogenetic analysis of gene sequences can be used to plumb the depths and deduce the general nature of the last universal common ancestor (LUCA) of cellular life from its catalogue of genes, but just how relevant is this to the actual origin of life? The latest list of genes thought to be present in LUCA is a long one. The presence of membranes, proteins, RNA and DNA, the ability to perform replication, transcription and translation, as well as harbouring an extensive metabolism driven by energy harvested from ion gradients using ATP synthase[2], reveal that there must have been a vast amount of evolutionary innovation between the origin of life and the appearance of LUCA. Clearly, it is not possible to deduce the precise environment that LUCA inhabited from such a low resolution and incomplete picture, but it is tempting to speculate and assume from its reliance on hydrogen as a reductant — through the action of NiFe hydrogenase — that it lived in a place where hydrogen was plentiful. Furthermore, many of the inferred proteins in LUCA use FeS clusters and other transition-metal-ion-based co-factors; thus, an environment that could furnish these inorganic components is circumstantially implicated[2]. However, what we are really after is the environment and the chemistry associated with the origin of life, not where an organism or collective of organisms way along the evolutionary line lived, but is it possible to get the former from the latter? Unfortunately not, because it is impossible to say whether the predecessors of LUCA inhabited the same environment from day one, or whether life started elsewhere, spread and adapted to the conditions at various locations, and was then killed off everywhere except at LUCA's chosen abode by, for example, a giant impact[3,4].

Considering that our lineage cannot be phylogenetically traced back beyond LUCA to the very origin of life, is there anything about biology's beginnings that we can glean from our downwards look from extant biology? The answer is yes, but a guarded yes. The format of life — its integrated informational, catalytic and compartment-forming subsystems, the *tria juncta in uno* by metabolism[5] — must provide some clues as to the chemical events that created it. The structure of RNA, proteins, lipids and key metabolites gives us synthetic targets or, from a systems point of view, a target mixture.

Biology almost always relies on chemistry that does not proceed efficiently in the absence of catalysis, because this allows chemistry to be regulated by dialling various catalysts up or down. However, most prebiotic chemistry must proceed of its own accord, and this surely suggests that it must generally be different from the underlying chemistry used in biology (although this is not invariably the case; for example, the dismutation of 6,7-dimethyl-8-ribityllumazine can proceed efficiently in the absence of enzymatic catalysis[6]). Nevertheless, despite the inevitable widespread differences between their individual reactions, prebiotic reaction networks ultimately have to transition into biochemical networks; hence, there must be some similarities between the two, if only at the level that practitioners of synthesis would view as strategic.

By approaching the origin of life from chemistry up, there have to be constraints as to what are plausible starting materials and reaction conditions, but defining these constraints can be difficult because of our uncertainties concerning early Earth geochemistry[7]. Synthesis inevitably has to play a major part, but it has to somehow be controlled and coordinated if mixtures of just the right complexity to progress towards life are to be produced. High energy, non-selective chemistry might appear appealing at first, because many (proto-)biomolecules can be produced in one step[8] — especially if one is prepared to analyse down to the parts per million level[9] — but myriad by-products makes their subsequent separation or selective utilization seem impossibly difficult. More plausibly, certain inherently favoured reactions or sequences of reactions might selectively produce key molecules destined for biology. If this were the case, then it should be discoverable through experimental investigation. Accordingly, several years ago, we set out to use experimental chemistry to address two questions. First, are completely different chemistries needed to make the various subsystems? Second, would these chemistries be compatible with each other? Our goal was to investigate the synthesis of nucleotides, amino acids, lipids and other cellular components from simple feedstocks under prebiotically

plausible conditions. We reasoned that any commonality of intermediates and/or (by-) products between subsystems would suggest their linkage from the outset. However, before these results are described, some fundamental aspects of prebiotic chemistry are considered.

The beginning

Systems chemistry synthesis of building blocks. The evidence that certain chemistry was crucial to the emergence of life would be expected to be especially strong for C–C bond forming reactions and synthetic homologations or other elaborations based thereon. This is because early Earth carbon feedstocks were most likely one-carbon compounds, and yet most biomolecules have multiple contiguous carbon atoms. Joining two molecules of the same one-carbon compound together through C–C bond formation is difficult: electrophilic formaldehyde **1** does not easily dimerize (FIG. 1), neither does the weaker electrophile carbon dioxide or the marginally nucleophilic carbon monoxide. By contrast, hydrogen cyanide **2** will dimerize through the attack of its nucleophilic conjugate base upon its weakly electrophilic self, but this reaction is sluggish, and the resultant dimer is more reactive than the monomer thus polymerization ensues[10] (FIG. 1). What is really needed (if we limit our C–C bond forming reaction to two-electron chemistry) is two different one-carbon compounds in the same place at the same time, one of which is a good nucleophile, such as a cyanide anion, and the other a good electrophile, such as formaldehyde **1**. Is this too demanding of the early Earth's environment?

To cut a long story short, we think that it is not, because hydrogen cyanide **2** can be converted to formaldehyde **1** by inherently favoured reduction with photochemically generated hydrated electrons and hydrolysis. Furthermore, subsequent cyanohydrin formation and reduction can initiate a Kiliani–Fischer type homologation sequence that builds simple sugars[11,12] (FIG. 1). Other functional groups are also reduced by hydrated electrons: certain α-hydroxy carbonyl compounds are deoxygenated; conjugated double bonds are saturated; and ketones can be reduced to secondary alcohols. Remarkably, when these few reduction reactions are combined with several addition reactions and a dry-state phosphorylation (conditions for which were discovered nearly half a century ago[15,16] but are still being rediscovered[17]), a reaction network leading from hydrogen cyanide **2** (and a few of its derivatives) to the pyrimidine nucleotides, and to precursors to a dozen amino acids and glycerol phosphate lipids, can be defined[18]. The reactions are all high yielding and lead to little else besides biomolecules or their precursors. It is not definitive proof that the building blocks of biology arose in this way, but it is compelling and indicates that the requirements for these reactions to take place should be used to constrain geochemical scenarios on the early Earth. A requirement for ultraviolet irradiation to generate hydrated electrons would rule out deep sea environments. This, along with strong bioenergetic and structural arguments[19], suggests that the idea that life originated at vents[20] should, like the vents themselves, remain 'In the deep bosom of the ocean buried'. The chemistry places certain demands on the environment of the early Earth: for example, the high concentrations of certain species through evaporation of solutions. Supporters welcome these demands as constraints that

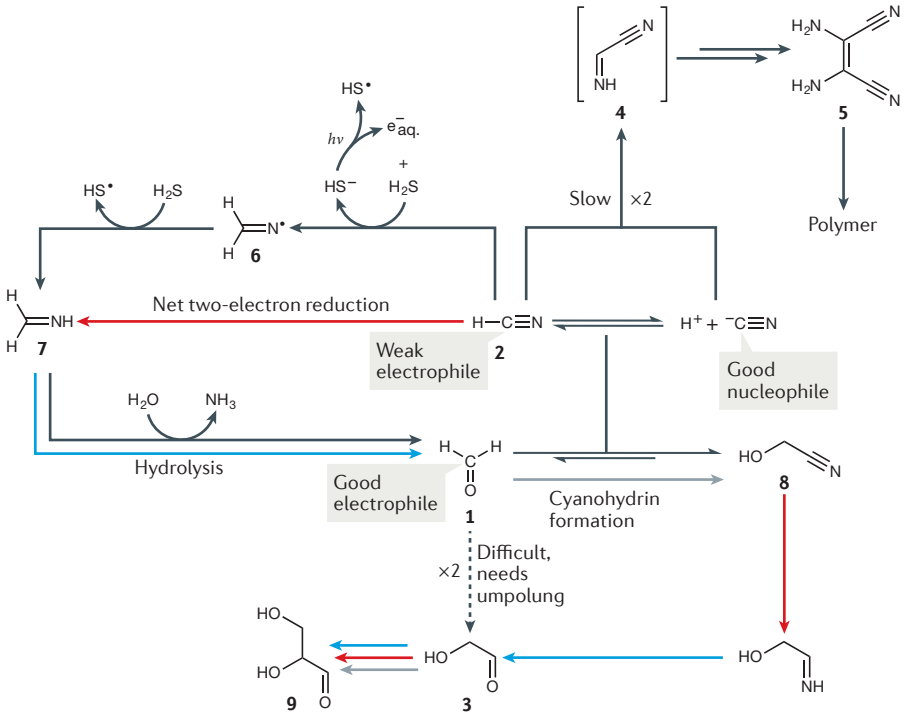


Figure 1 | **Difficulties of one-carbon compound dimerization and how to achieve the equivalent.** The dimerization (indicated by '×2') of geochemically plausible one-carbon feedstock molecules, such as formaldehyde **1** and hydrogen cyanide **2**, by C–C bond forming reactions is difficult for fundamental chemical reasons. Neither of these compounds can react with themselves in their natural electrophilic form: for dimerization to occur, partial polarity reversal or umpolung is required. For formaldehyde **1**, this is difficult to achieve prebiotically and the dimer glycolaldehyde **3** is thus normally considered elusive. For hydrogen cyanide **2**, umpolung is easily achieved, because the pK_a is reasonably low (~9) and the conjugate base, cyanide anion is a good nucleophile. However, hydrogen cyanide **2** itself is only a weak electrophile; therefore, its reaction with the cyanide anion is slow[10]. The dimer **4** that results is a better electrophile than hydrogen cyanide **2**, which is also the case for a trimer that is produced by further addition of **2**. The tetramer **5** can be transiently observed (and is a key purine precursor) but it is also a reasonable electrophile, leading to polymerization. Highly efficient formation of the formaldehyde dimer **3** can occur in an indirect way when hydrogen cyanide **2** undergoes reduction with hydrated electrons[11,12]. These potent yet selective reducing agents can be easily formed when anions, such as hydrosulfide (HS^-), are irradiated with ultraviolet light[13] (indicated by *hν*). The conjugate acid of hydrosulfide, hydrogen sulfide, is a general acid ($pK_a \approx 7$) and efficiently protonates the incipient radical anion formed by addition of a hydrated electron to **2**, giving a relatively stable iminyl radical **6** (REF. 14) (thus hydrogen cyanide is ideally suited to be reduced by hydrated electrons). Hydrogen atom transfer from hydrogen sulfide to **6** then gives formaldehyde imine **7**, which hydrolyses to formaldehyde **1**. At this point, the good electrophilicity of **1** coupled with the good nucleophilicity of the cyanide anion results in rapid formation of the cyanohydrin, glycolonitrile **8**. Further hydrated electron reduction of **8** and hydrolysis of the resultant imine gives glycolaldehyde **3**. The conversion of formaldehyde **1** to glycolaldehyde **3** in this roundabout manner is a Kiliani–Fischer type homologation, repetition of which gives glyceraldehyde **9**, a key (proto)metabolic intermediate towards nucleic acids, proteins and lipids.

help refine primitive Earth scenarios[21]. Detractors view them as unacceptable but must surely then demonstrate that other scenarios can be equally productive[19,22].

From protometabolism to metabolism. The term cyanosulfidic is used to describe the chemistry we have uncovered because of its reliance on hydrogen cyanide and hydrosulfide as an optimal hydrated electron source. The reaction network is referred to as protometabolic systems chemistry because we propose that it might have foreshadowed the subsequent evolution of metabolism after life first emerged[18]. As previously discussed, the reactions of modern metabolism should not be generally expected to reflect those of prebiotic chemistry. However, prebiotic chemical networks could have shaped the general layout of metabolism by providing a portfolio of products connected through synthetic logic, to which biology became addicted. Progressive depletion of the various products in the environment would have then driven the evolution of a series of biosynthetic pathways to these essential compounds using different individual reactions. It is most likely that the synthetic logic was the same, however, with selection for overall biosynthetic network efficiency arriving at the same general solution as that dictated by prebiotic chemical expediency.

To better appreciate what is meant by this, consider the role in extant biology of the triose phosphates, glyceraldehyde-3-phosphate **10** and dihydroxyacetone phosphate **11**, as precursors to ribonucleotides, lipids and certain amino acids such as valine **12** (FIG. 2a). The stereospecific enzyme-catalysed addition of a two-carbon unit from another carbohydrate to glyceraldehyde-3-phosphate **10** gives ribose-5-phosphate **13** (REF. 23). The phosphate group of the latter means that the furanose form predominates in solution, rather than the otherwise preferred pyranose forms, making subsequent (ribo-furano-) nucleotide assembly more efficient. The two-electron reduction of one or the other of the interconvertible triose phosphates stereospecifically generates either enantiomer of glycerol-1-phosphate **14**. Complex enzyme-catalysed functional group manipulations and rearrangements generate the methyl groups of valine **12** from the terminal carbon atoms of two molecules of glyceraldehyde-3-phosphate **10** via pyruvate **15** (REFS 24–26). The chemistry underlying these various transformations

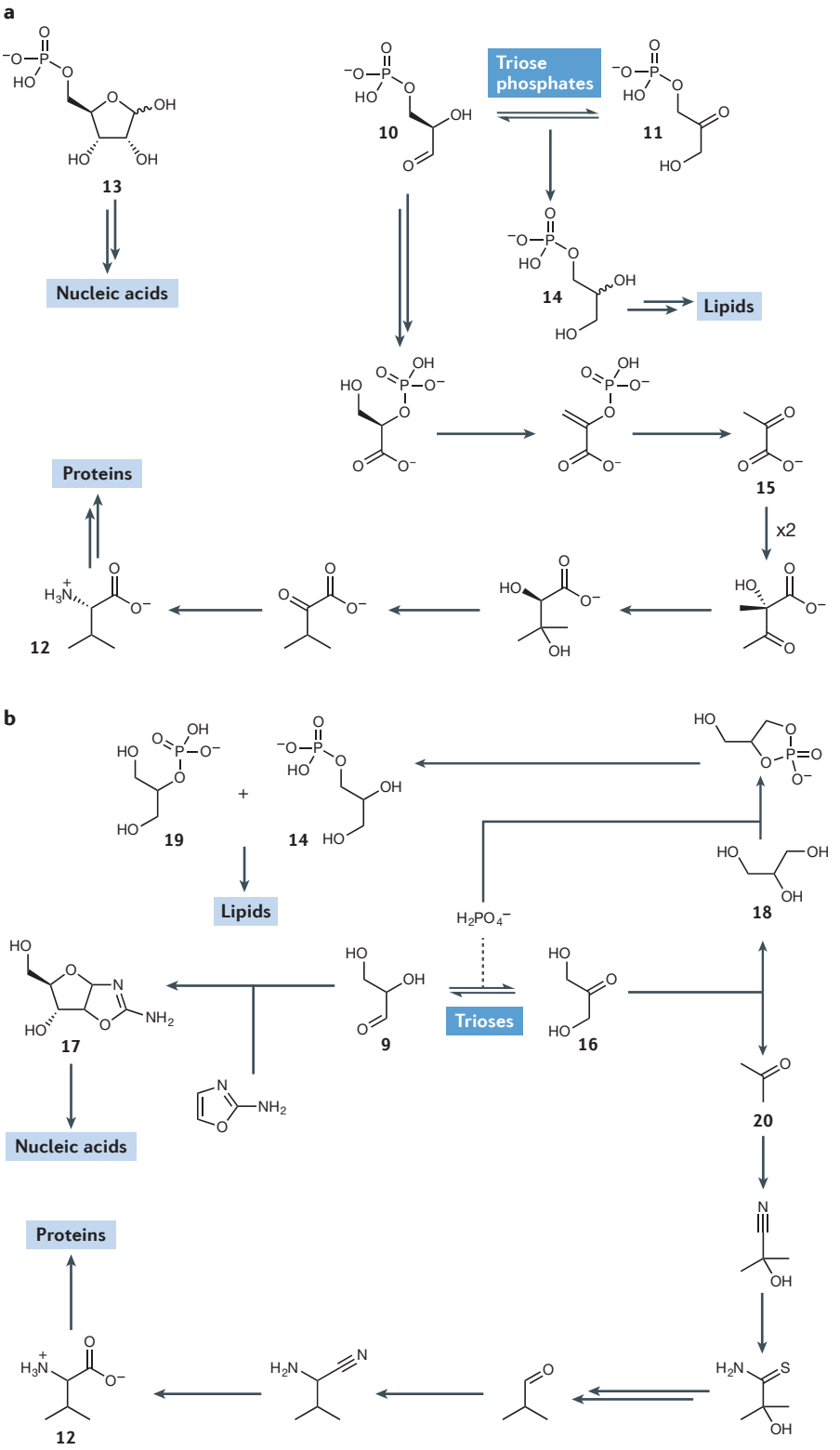


Figure 2 | **Underlying similarities between extant metabolism and cyanosulfidic protometabolism.** For both extant metabolism (panel **a**) and protometabolism (panel **b**), three-carbon sugars or their phosphate derivatives (grey shading) serve as precursors of lipids (through reduction), nucleic acids (through addition of a two-carbon unit) and amino acids, such as valine **12** (through multistep pathways). The specific reactions are different in the two cases, but such differences are to be expected because prebiotic reactions have to proceed of their own accord, whereas most biochemical reactions require enzyme catalysis. However, the strategy or logic behind the branching of the product tree, to use synthetic chemistry parlance, is the same in both cases.

either does not take place in the absence of enzyme catalysis or is dominated by deleterious chemistry. For example, the umpoled addition of pyruvate **15** to itself does not occur in the absence of catalysis[26], and glyceraldehyde-3-phosphate **10**, when left to its own devices, undergoes intramolecular phosphate-assisted enolate formation followed by $E1_{cB}$ elimination giving the enol of methylglyoxal[27]. These examples reveal that biology now uses positive and negative catalysis to hold sway over metabolism.

Comparing this partial metabolic network to part of the protometabolic network that we uncovered[18] (FIG. 2b) reveals the strategic similarities. Thus, the trioses glyceraldehyde **9** and dihydroxyacetone **16** similarly serve as precursors of ribonucleotides, lipids and certain amino acids such as valine **12**. The stereoselective addition of a two-carbon enolate equivalent to glyceraldehyde **9** preferentially gives ribo- and arabino-configured pentose amino-oxazoline intermediates **17** constrained to the furanose form by the fused-ring system. Phosphate-catalysed isomerization of glyceraldehyde **9** generates dihydroxyacetone **16**, which is reduced by hydrated electrons to two products in roughly equal amounts. The first, glycerol **18**, then undergoes phosphorylation to give glycerol-phosphate-lipid precursors **14** and **19**. The second, acetone **20**, undergoes a variant of the Kiliani–Fischer homologation chemistry in which deoxygenation precedes the reduction of a nitrile group equivalent. The methyl groups of the resulting valine **12** derive from both terminal carbon atoms of a single molecule of glyceraldehyde **9**. Cyanosulfidic chemistry thus furnishes building blocks for three subsystems of biology that logically derive from three-carbon sugar units. Subsequent metabolic routes to the same building blocks would have to use different feedstocks and chemistry, because levels of hydrogen cyanide **2** would presumably have dropped, and life would now be in the shade or otherwise shielded from ultraviolet irradiation. These early organisms would benefit from their newly acquired metabolism having an obligate reliance on catalysts because of the regulatory capacity that they bestow. Modern metabolism is not a palimpsest, however, because prebiotic chemistry is not completely overwritten, its final products remain and its synthetic logic endures. The resemblance to modern biochemistry might not be obvious to the non-chemist at first, but it is to those with chemical acuity.

The beginning of the end

Assembly of the building blocks. In a way, life can be thought of as an energized assemblage of building blocks; therefore, any synthesis of these building blocks, however all-encompassing, is only a step towards demonstrating, or recapitulating, the origin of life. More synthesis is needed, this time the synthesis of macromolecules and assemblies. Dehydration reactions will be needed to produce biopolymers from the corresponding monomers, and there will be a need for energy sources to drive this. The chemistry can even be energetically wasteful as long as there are abundant sources of energy, and this might even be preferable because it could give rise to another of biology's addictions: energy dissipation. It seems unlikely that the earliest evolutionary wonders of biology were all discovered simultaneously. A stepwise process seems much more plausible, but for each innovation to be retained against the otherwise inexorable drift towards equilibrium, energy would have to be expended endlessly[28]. Combining these ideas with the concept that aliveness need not be all or nothing leads to a diagrammatic depiction that we have found useful to collect and direct our thoughts[1,29] (FIG. 3). In such a diagram, the aliveness of a system is plotted as a function of time or system complexity. Viewed in this way, the prebiotic synthesis of building blocks — to which we have devoted so much of our time — only corresponds to a small increase in the complexity of the system and to no increase in its aliveness (a humbling thought). However, this synthesis is necessary to put the system on the right path, and knowing the steps that have been taken can give some hints as to the nature of the steps that follow, at least up to a point: the necessity–contingency boundary when the synthesis of macromolecules from multiple monomers reaches the stage in which only a fraction of all possible sequence variants can be sampled owing to the number of possible permutations exceeding the number of molecules. At this point, new function is most likely to accrue to the system if it can be easily accessed, because in sequence space there are multiple islands of macromolecules with that function[30]. As an aside, this

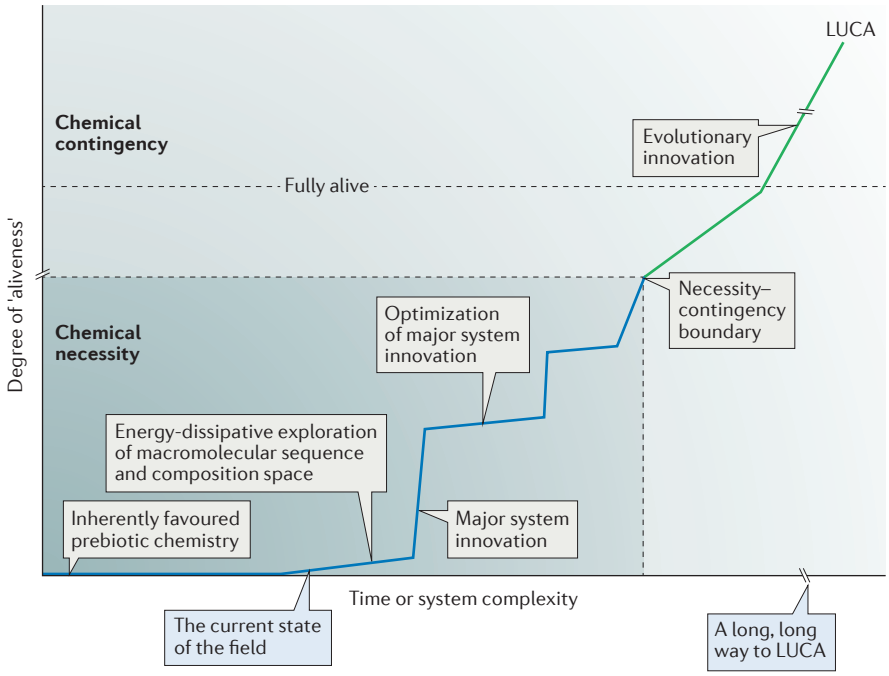


Figure 3 | **Transition to life: onwards and upwards.** The transition of a system from the inanimate state to the animate is envisioned as an increase in 'aliveness' over time. We (and others[29]) prefer to consider this transition as a series of steps, rather than a single step, following the prelude of prebiotic chemistry[1]. Equilibrium is death, which means some sort of coupling of energy dissipation to maintain the system continuously out of equilibrium throughout the transition is envisaged, but when we first started contemplating this, we could not see a way in which this might be achieved, hence the somewhat nebulous picture. Also shown is the necessity–contingency boundary beyond which material limitations prevent full exploration of the sequence space of macromolecules assembled from different monomeric building blocks; therefore, chemical determinism can no longer be relied on as a source of innovation, and further improvements have to be chanced upon instead.

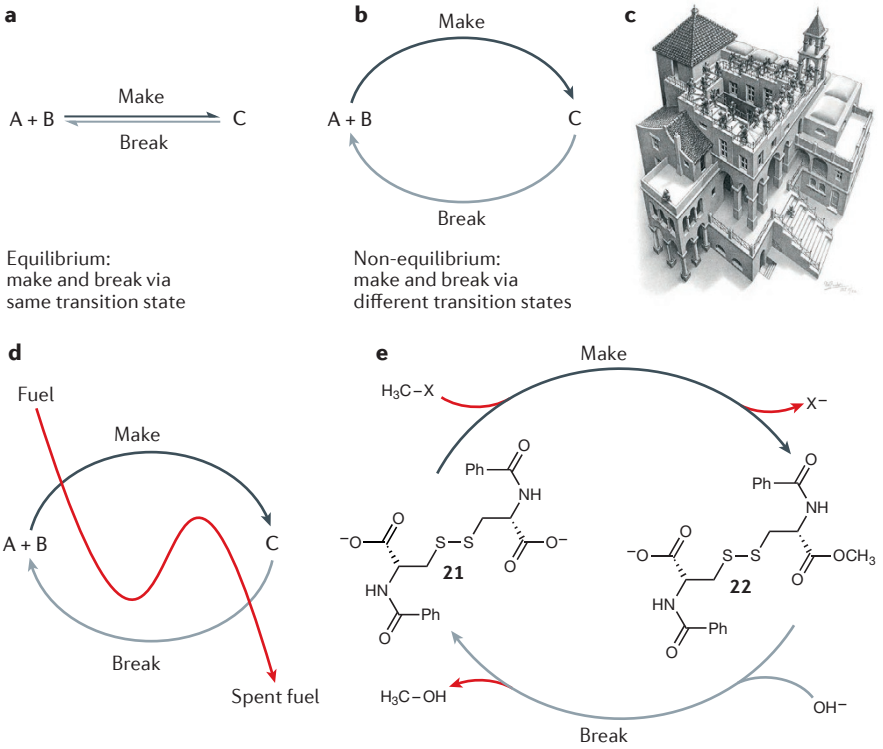
suggests that if the tape were to be replayed, a biology ostensibly similar to ours might emerge with proteins and RNA assembled from (mostly) the same monomers performing similar functions because of chemical necessity, but a closer look would reveal sequences radically different to those used in our biology because of chemical contingency. Getting back on track, regardless of the sequences discovered through functional selection, the actual synthesis of macromolecules presents major challenges to the system because of the need for repeated joining reactions — oligomerizations or ligations — in water.

In the case of RNA, not only must phosphodiester links be repeatedly forged (if we assume joining reactions involving P–O bond formation), but they must ultimately connect the 5′-oxygen of one nucleotide to the 3′-oxygen, and not the 2′-oxygen, of the next nucleotide. 2′,5′-Linkages can be tolerated functionally at low levels in certain RNAs[31], but they are not inheritable in a sequence-specific manner, and for most intents and purposes, extant biology uses 3′,5′-linkages. Although we have demonstrated that 3′,5′-linkages can be preferentially formed by prebiotically selective 2′-*O*-acetylation and ligation of those oligonucleotides with 3′-phosphate termini in mixtures of oligonucleotides with 2′- and 3′-phosphate termini[32], the synthetic selectivities and preferences are not enough to explain how RNA with all 3′,5′-linkages might first have been produced.

While mulling this over, and inspired by the art of Escher, the brilliant work of van Esch, Eelkema and co-workers[33] on energy-dissipative cycling reactions came to our attention (BOX 1). In this work, methylation of a carboxylate group lowers the aqueous solubility of the dipeptide derivative **21** such that it assembles into stacks and thence fibres. Saponification of the methyl ester **22** in the alkaline medium regenerates the carboxylate group but is balanced by remethylation, as long as the supply of the methylating agent is maintained. Although a cycle between carboxylate **21** and methyl ester **22** operates, it is unidirectional and not an equilibrium, because the hydrolysis and methylation proceed via different mechanisms ($B_{Ac}2$ versus S_N2 (retro-$B_{Al}2$)). Essentially, energy that is liberated through the conversion of the methylating agent to methanol is dissipated to maintain the methyl ester **22** and the fibres formed therefrom in an out-of-equilibrium state. In the work of van Esch, Eelkema and colleagues, the same

Box 1 | **Chemical cycling and energy dissipation**

A simple equilibrium can be viewed as cycling, but in the context of the origin of life, equilibrium is death (panel **a**). More preferable would be the cycle shown in panel **b**, in which bond making and bond breaking steps proceed via different transition states, but realizing such a process cyclically on its own is not possible because the relative stabilities of starting materials and products are fixed. Ideally, a situation would exist in which both bond making and bond breaking are energetically downhill, reminiscent of Escher's wonderful 'Ascending and Descending' and the Penrose stairs (panel **c**), but these are, of course, illusory, and thermodynamics cannot be deceived. The seminal contribution of van Esch and colleagues was to show how energy dissipation could be coupled to a cycle to render it continuously operable in one direction as long as fuel is supplied[33] (panel **d**). The molecular means by which they did this is shown in panel **e**. Panel **c** is a reproduction of M.C. Escher's "Ascending and Descending" © 2016 The M.C. Escher Company - The Netherlands. All rights reserved. www.mcescher.com.



product was continually (re)formed, but an interesting situation would arise if multiple products were possible. This is because the selectivity for a particular product could be attributed to a preference for it in the outbound synthetic step or to a preference for the hydrolysis of the other products in the return step. In either case, the particular product should become enriched if a mixture of products was subjected to the energy-dissipative cycling process, especially if there was selectivity for the product in both steps.

In the case of RNA, it has been known for many years that 2′,5′-linkages are far more hydrolytically labile than 3′,5′-linkages[34–36]. Furthermore, the 2′,3′-cyclic phosphate that initially results from such cleavage undergoes further hydrolysis to 2′- and 3′-monophosphates with a moderate preference for the latter[37]. Thus, if this hydrolysis chemistry were to be combined with the acetylation–ligation chemistry that we uncovered[32], then it might be possible for RNA with a mixture of both linkage isomers to undergo conversion into RNA with just 3′,5′-linkages, and then for sequences to be shuffled by hydrolysis, crossover and ligation (FIG. 4) as foreseen by Lehman's recombination model[38]. If such conversion and subsequent shuffling is proven to be possible under plausible conditions, then the hydrolytic instability of RNA, so long viewed as an Achilles heel[39], might instead be viewed as an absolute necessity for energy-dissipative cycling. Following the same sort of clue, might other biological (macro)molecules or subsystems be amenable to synthesis that is made extremely selective by the additional energy expenditure afforded by cycling?

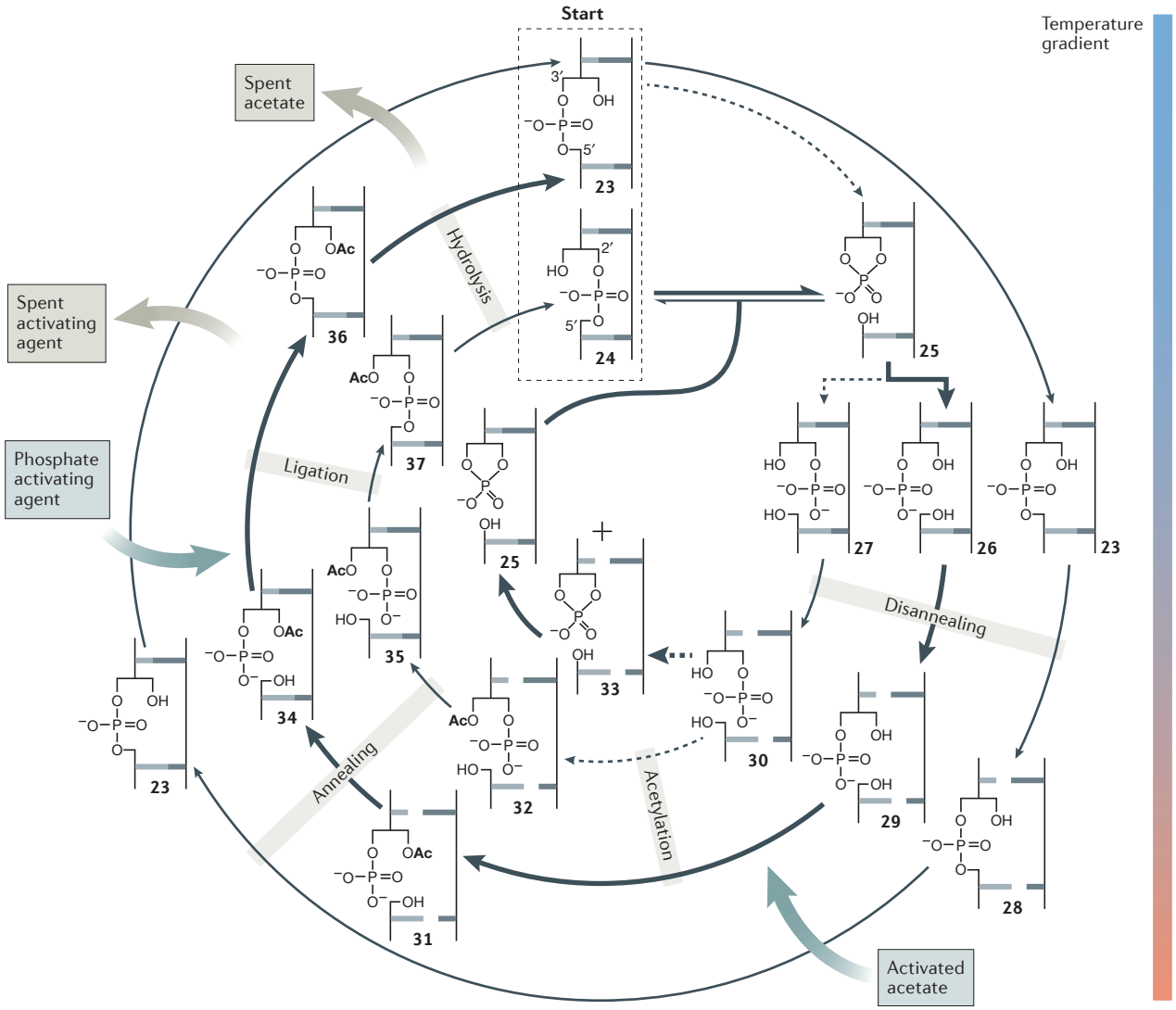


Figure 4 | **Towards RNA cycling.** By piecing together the literature[32,34–37], a scheme for RNA cycling can now be constructed. It starts with a population of duplex oligonucleotides with heterogeneous linkage isomers, represented by **23** and **24**, undergoing hydrolysis. It is known that under mildly alkaline conditions, hydrolysis of the 3′,5′-linkage in duplex **23** is much slower than cleavage of the 2′,5′-linkage in duplex **24**. The cleavage of the 2′,5′-linkage in duplex **24** involves initial reversible isomerization to the nicked duplex **25** with a 2′,3′-cyclic phosphate followed by hydrolysis of the latter to nicked duplexes **26** and **27** with 3′- and 2′-monophosphates, respectively. There is moderate selectivity in the latter hydrolysis for production of duplex **26** with a 3′-phosphate. Heating would then cause disannealing of both the nicked duplexes **26** and **27** and the residual duplex **23**, the former giving strand separated nicked duplexes **29** and **30** and the latter giving the strand separated duplex **28**. Acetylation of oligonucleotides with 2′- and 3′-monophosphate termini takes place on the phosphate group initially, and is followed by an attack of the adjacent hydroxyl group at the acyl carbon or at the phosphorus of the intermediate mixed anhydride. 3′-Phosphate termini are efficiently converted to 2′-acetyl-3′-phosphate termini, whereas 2′-phosphate termini are less efficiently converted to mixtures of 2′-phosphate-3′-acetyl termini and 2′,3′-cyclic phosphate termini. Thus, the strand-separated nicked duplex **29** is efficiently converted to its 2′-acetate **31** and **30** is converted to a mixture of its 3′-acetate **32** and 2′,3′-cyclic phosphate **33**. If crossover is ignored for the moment, strand re-annealing then produces the acetylated nicked duplexes **34** and **35**, the nicked duplex with a 2′,3′-cyclic phosphate **25**, as well as reproducing duplex **23**. The addition of a phosphate activating agent leaves duplex **23** unchanged but ligates the acetylated nicked duplexes **34** and **35** to give the acetylated duplexes **36** and **37**. Under hydrolysis conditions, acetyl groups are removed and phosphodiester bond hydrolysis recommences. Thus, the acetylated duplex **36** is converted into the starting duplex **23**, and **37** is transiently converted into the other starting duplex **24**, before participating in the equilibrium with the nicked duplex that includes a 2′,3′-cyclic phosphate **25**. It is complicated, but the selectivities in the various hydrolysis and acetylation steps all stack up such that it can be anticipated that the 2′,5′-linked duplex **24** will be progressively converted to the 3′,5′-linked duplex **23** with each turn of the cycle (following the bold arrows). Although the 3′,5′-linked duplex **23** is more stable to alkaline conditions than its isomer **24**, it will occasionally undergo hydrolysis, and this, along with crossover at the annealing stage in the cycle, should result in partial sequence shuffling[38]. Consequently, a random population of RNA molecules with heterogeneous linkage isomers might convert over time to a population having only 3′,5′-linkages and then continue to explore sequence space through partial hydrolysis, crossover and religation. This remarkable molecular remodelling — should it be demonstrable — might go a long way to explaining how the RNA that is seen in extant biology came into being. Furthermore, the extensive energy dissipation required would confirm the link between the informational subsystem and metabolism at an early stage. The coloured bar indicates a temperature gradient over which this proposed cycle operates.

Diacyl glycerol phosphate lipids seem to be good candidates, with the correct chain length for vesicle formation potentially being selectable through a combination of acylation and hydrolysis, but proteins are not. The random synthesis of peptides then appears particularly unattractive and very early coded synthesis is suggested. Clearly, we are not yet even at the beginning of the end of our quest to understand it, but the end of the beginning is offering up some very tantalizing clues about the origin of life.

John D. Sutherland is at the MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK.
johns@mrc-lmb.cam.ac.uk

doi:10.1038/s41570-016-0012
Published online 18 Jan 2017

1. Sutherland, J. D. The origin of life — out of the blue. *Angew. Chem. Int. Ed.* **55**, 104–121 (2016).
2. Weiss, M. C. *et al.* The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **1**, 16116 (2016).
3. Sleep, N. H. & Zahnle, K. Refugia from asteroid impacts on early Mars and the early Earth. *J. Geophys. Res.* **103**, 28529–28544 (1998).
4. Abramov, O. & Mojzsis, S. J. Microbial habitability of the Hadean Earth during the late heavy bombardment. *Nature* **459**, 419–422 (2009).
5. Gánti, T. *The Principles of Life* (Oxford Univ. Press, 2003).
6. Paterson, T. & Wood, H. C. S. Deuterium exchange of C-methyl protons in 6,7-dimethyl-8-D-ribityl-lumazine, and studies of the mechanism of riboflavin biosynthesis. *J. Chem. Soc. D* 290–291 (1969).
7. Powner, M. W. & Sutherland, J. D. Prebiotic chemistry: a new *modus operandi*. *Phil. Trans. R. Soc. B* **366**, 2870–2877 (2011).
8. Miller, S. L. A production of amino acids under possible primitive Earth conditions. *Science* **117**, 528–529 (1953).
9. Parker, E. T. *et al.* Primordial synthesis of amines and amino acids in a 1958 Miller H_2S-rich spark discharge experiment. *Proc. Natl Acad. Sci. USA* **108**, 5526–5531 (2011).
10. Sanchez, R. A., Ferris, J. P. & Orgel, L. E. Studies in prebiotic synthesis II. Synthesis of purine precursors and amino acids from aqueous hydrogen cyanide. *J. Mol. Biol.* **80**, 223–253 (1967).
11. Ritson, D. & Sutherland, J. D. Prebiotic synthesis of simple sugars by photoredox systems chemistry. *Nat. Chem.* **4**, 895–899 (2012).
12. 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).
13. Sauer, M. C., Crowell, R. A. & Shkrob, I. A. Electron photodetachment from aqueous anions. 1. Quantum yields for generation of hydrated electron by 193 and 248 nm laser photoexcitation of miscellaneous inorganic anions. *J. Phys. Chem. A* **108**, 5490–5502 (2004).
14. Zard, S. Z. Iminyl radicals: a fresh look at a forgotten species (and some of its relatives). *Synlett* **1996**, 1148–1154 (1996).
15. Lohrmann, R. &. Orgel, L. E. Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents. *Science* **171**, 490–494 (1971).
16. Schoffstall, A. M. Prebiotic phosphorylation of nucleosides in formamide. *Orig. Life* **7**, 399–412 (1976).
17. Burcar, B. *et al.* Darwin's warm little pond: a one-pot reaction for prebiotic phosphorylation and the mobilization of phosphate from minerals in a urea-based solvent. *Angew. Chem. Int. Ed.* **55**, 13249–13253 (2016).
18. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **7**, 301–307 (2015).
19. Jackson, J. B. Natural pH gradients in hydrothermal alkali vents were unlikely to have played a role in the origin of life. *J. Mol. Evol.* **83**, 1–11 (2016).
20. Martin, W. F., Sousa, F. L. & Lane, N. Energy at life's origin. *Science* **344**, 1092–1093 (2014).
21. Springsteen, G. Reaching back to jump forward: recent efforts towards a systems-level hypothesis for an early RNA world. *ChemBioChem* **16**, 1411–1413 (2015).
22. Sojo, V., Herschy, B., Whicher, A., Camprubí, E. & Lane, N. The origin of life in alkaline hydrothermal vents. *Astrobiology* **16**, 181–197 (2016).
23. Schenk, G., Duggleby, R. G. & Nixon, P. F. Properties and functions of the thiamin diphosphate dependent enzyme transketolase. *Int. J. Biochem. Cell Biol.* **30**, 1297–1318 (1998).
24. Chipman, D. M., Duggleby, R. G. & Tittmann, K. Mechanisms of acetohydroxyacid synthases. *Curr. Opin. Chem. Biol.* **9**, 475–481 (2005).
25. Wong, S. H., Lonhienne, T. G., Winzor, D. J., Schenk, G. & Guddat, L. W. Bacterial and plant ketol-acid reductoisomerases have different mechanisms of induced fit during the catalytic cycle. *J. Mol. Biol.* **424**, 168–179 (2012).
26. Pirrung, M. C., Holmes, C. P., Horowitz, D. M. & Nunn, D. S. Mechanism and stereochemistry of α,β-dihydroxyacid dehydratase. *J. Am. Chem. Soc.* **113**, 1020–1025 (1991).
27. Richard, J. P. Acid–base catalysis of the elimination and isomerization reactions of triose phosphates. *J. Am. Chem. Soc.* **106**, 4926–4936 (1984).
28. Pascal, R., Pross, A. & Sutherland, J. D. Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics. *Open Biol.* **3**, 130156 (2013).
29. Kuhn, H. Model consideration for the origin of life. *Naturwissenschaften* **63**, 68–80 (1976).
30. Jiménez, J. I., Xulvi-Brunet, R., Campbell, G. W., Turk-MacLeod, R. & Chen, I. A. Comprehensive experimental fitness landscape and evolutionary network for small RNA. *Proc. Natl Acad. Sci. USA* **110**, 14984–14989 (2013).
31. Engelhart, A. E., Powner, M. W. & Szostak, J. W. Functional RNAs exhibit tolerance for non-heritable 2′–5′ versus 3′–5′ backbone heterogeneity. *Nat. Chem.* **5**, 390–394 (2013).
32. Bowler, F. R. *et al.* Prebiotically plausible oligoribonucleotide ligation facilitated by chemoselective acetylation. *Nat. Chem.* **5**, 383–389 (2013).
33. Boekhoven, J., Hendriksen, W. E., Koper, G. J., Eelkema, R. & van Esch, J. H. Transient assembly of active materials fueled by a chemical reaction. *Science* **349**, 1075–1079 (2015).
34. Usher, D. A. & McHale, A. H. Hydrolytic stability of helical RNA: a selective advantage for the natural 3′,5′-bond. *Proc. Natl Acad. Sci. USA* **73**, 1149–1153 (1976).
35. Usher, D. A. Early chemical evolution of nucleic acids: a theoretical model. *Science* **196**, 311–313 (1977).
36. Rohatgi, R., Bartel, D. P. & Szostak, J. W. Nonenzymatic, template-directed ligation of oligoribonucleotides is highly regioselective for the formation of 3′–5′ phosphodiester bonds. *J. Am. Chem. Soc.* **118**, 3340–3344 (1996).
37. Kuusela, S. & Lönnberg, H. Metal ion-promoted hydrolysis of uridine 2′,3′-cyclic monophosphate: effect of metal chelates and uncomplexed aquo ions. *J. Phys. Org. Chem.* **5**, 803–811 (1992).
38. Lehman, N. A recombination-based model for the origin and early evolution of genetic information. *Chem. Biodivers.* **5**, 1707–1717 (2008).
39. Pace, N. R. Origin of life-facing up to the physical setting. *Cell* **65**, 531–533 (1991).

Acknowledgements
This work was supported by the Medical Research Council (No. MC_UP_A024_1009), and a grant from the Simons Foundation (No. 290362 to J.D.S.). J.D.S. thanks members of his group for helpful discussions and suggestions.

Competing interests statement
The author declares no competing financial interests.

FURTHER INFORMATION
'Ascending and Descending' by M. C. Escher: http://www.mcescher.com/gallery/recognition-success/ascending-and-descending/