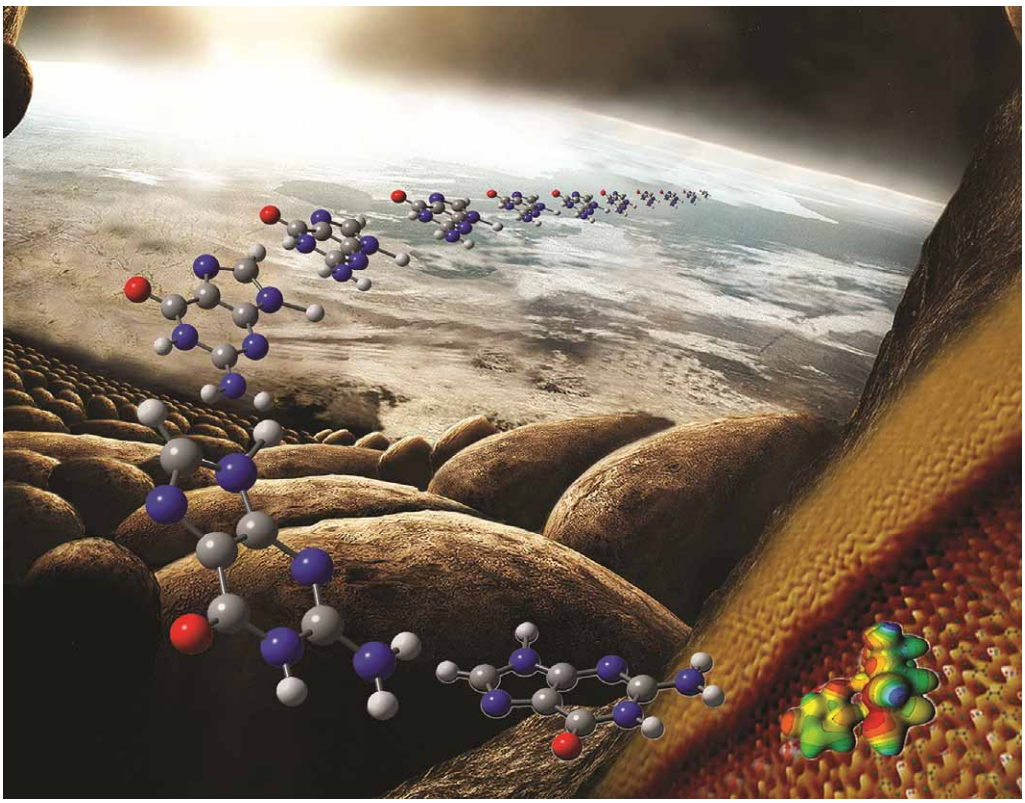
Chem Soc Rev

This article was published as part of the

Prebiotic chemistry themed issue

Guest editors Jean-François Lambert, Mariona Sodupe and Piero Ugliengo

Please take a look at the issue 16 2012 table of contents to access other reviews in this themed issue



Chem Soc Rev

Dynamic Article Links

Cite this: *Chem. Soc. Rev.*, 2012, **41**, 5394–5403

www.rsc.org/csr

TUTORIAL REVIEW

Out of fuzzy chemistry: from prebiotic chemistry to metabolic networks†‡

Juli Peretó*

Received 26th February 2012
DOI: 10.1039/c2cs35054h

The origin of life on Earth was a chemical affair. So how did primitive biochemical systems originate from geochemical and cosmochemical processes on the young planet? Contemporary research into the origins of life subscribes to the Darwinian principle of material causes operating in an evolutionary context, as advocated by A. I. Oparin and J. B. S. Haldane in the 1920s. In its simplest form (*e.g.*, a bacterial cell) extant biological complexity relies on the functional integration of metabolic networks and replicative genomes inside a lipid boundary. Different research programmes have explored the prebiotic plausibility of each of these autocatalytic subsystems and combinations thereof: self-maintained networks of small molecules, template chemistry, and self-reproductive vesicles. This *tutorial review* focuses on the debates surrounding the origin of metabolism and offers a brief overview of current studies on the evolution of metabolic networks. I suggest that a leitmotif in the origin and evolution of metabolism is the role played by catalysers' substrate ambiguity and multifunctionality.

One problem, two approaches, many scenarios

Life emerged very early in the history of planet Earth. The transition from cosmo- and geochemistry to the most primitive life forms took place sometime between 4.1 billion years (Ga) ago—when the first oceans were likely in place—and 3.5 Ga ago—the age of the oldest microfossils.[1] The historical event giving rise to life is currently regarded as an evolutionary process, a succession of increasingly complex stages from prebiotic matter to the first systems exhibiting life-like properties. In its simplest form (*e.g.*, a bacterial cell) life is the outcome of the functional integration of a self-maintained metabolic network, a self-replicative genome and a self-reproductive compartment.

We can approach the origin-of-life problem from two potentially convergent perspectives:

(i) a top-down approach implementing the comparative biochemistry of extant cells, trying to identify their commonalities and reconstruct the universal common ancestor (or universal cenancestor, UC), and

(ii) a bottom-up approach aiming to recreate the most plausible chemical processes leading from simple compounds to supramolecular complexes with biochemical properties, under the conditions found on the primitive Earth. The problem is that the abiotic chemical systems of the highest complexity are very far from the simplest cellular organisation, including the kind the UC might have comprised. Even though most of the finer features remain unknown, the step-wise processes involved in the origin of life are scientifically comprehensible and experimentally reproducible.

The current debates on the origin of life have deep historical and philosophical roots (Lazcano[2] and references therein). In this sense, the scientific discussion on life's emergence does not differ from other intellectual conflicts in biology and has developed through different, sometimes incompatible, models.[3]

Institut Cavanilles de Biodiversitat i Biologia Evolutiva, and Departament de Bioquímica i Biologia Molecular, Universitat de València, C. Jose Beltran 2, 46980 Paterna, Spain.
E-mail: pereto@uv.es; Fax: +34 963543670; Tel: +34 963543666
† Part of the prebiotic chemistry themed issue.
‡ Dedicated to the memory of Lynn Margulis (1938–2011).

Juli Peretó

Juli Peretó is a Tenured Professor of Biochemistry and Molecular Biology at the University of València, and a researcher at the Evolutionary Genetics Unit, Cavanilles Institute for Biodiversity and Evolutionary Biology. His research interests include the evolution of metabolism, the minimal genome concept, and the history of ideas on the natural origin and the artificial synthesis of life. He was formerly the Secretary of the International Society for the Study of the Origin of Life (ISSOL – The International Astrobiology Society) and has been elected as the Second Vice President of ISSOL for the term 2011–2014.

Thus, the abiotic generation of chemical complexity on the primitive Earth combined both endogenous (atmosphere, sea surface, submarine hydrothermal settings) and exogenous (meteorites, comets, interstellar dust particles) sources. Some models differ in terms of their emphasis on minerals (as catalysts and/or stabilisers) and different energy sources (electromagnetic, chemical). The clash between models is even stronger when it comes to the debate on what emerged first, the genetic system or the metabolic one. Whether compartment formation was an early or late phenomenon during life origin is a discussion closely related with the early origin of metabolism. Many advocates of an early emergence of replicators consider the cell as a mere physical compartment for segregating polymers with differential replicative abilities. However, membranes are active players in energetic transductions. Some authors propose that life would be cellular *ab initio*, although there are still many controversial issues, including the chemical nature of the membranes and the kind of primitive molecular transducers of primary energy sources, prior to the existence of intricate protein machines. Experimental approaches on vesicle chemistry point to fatty acids as good candidates for the first amphiphilic protocellular constituents (see Szostak[4] and references therein).

According to the Oparin–Haldane proposals, it is assumed that heterotrophy was primordial, *i.e.* the first organisms gained energy and matter from a *primitive soup* rich in organic compounds—either of terrestrial or extraterrestrial origin—ready to be catabolised. Indeed, during recent decades scientists have accumulated evidence favouring the contributions of volcanic, atmospheric and cosmic chemistries to the abiotic inventories of the early Earth. Laboratory simulations and analyses of meteorites have been of paramount importance in this context. Nowadays, systems chemistry approaches open new ways forward (Szostak[4] and Powner and Sutherland[5] and references therein). Instead of looking for robust simple synthetic reactions (*e.g.*, Oro's pentamerisation of HCN to give adenine) or multi-component reactions yielding intractable product and by-product mixtures (*e.g.*, Miller's tar produced by electric discharges on simple combinations of gases), systems chemistry focuses on multi-component reactions and sequential one-pot multi-step synthetic sequences triggering molecular synergisms. In fact, moderately "dirty" initial mixtures lead to a simplification in the product outcome. This approach has been successful with nucleotide and lipid synthesis[5] and looks highly promising in the search for prebiotically plausible vesicle growth and reproduction and polynucleotide replication.[4]

At any rate, prebiotic chemists are well aware that one of the major problems they face is the geochemical relevance of the reactions under scrutiny. Our knowledge about the earliest terrestrial environments is so fragmentary and incomplete that deciding which component of abiotic chemistry—*i.e.*, cosmo- or geochemical compounds or processes—is prebiotically relevant will always be puzzling. It is generally accepted that the emergence of suprachemical subsystems exhibiting basic life-like properties, including self-maintained chemical networks, self-replicative polymers, and self-reproductive vesicles (Fig. 1) appear to have been a necessary step during chemical evolution or the prebiotic phase. Eventually, the harmonious articulation of those three prebiotic subsystems within the same functional framework could be considered as the beginning of biological evolution (reviewed by Szathmáry *et al.*[6]).

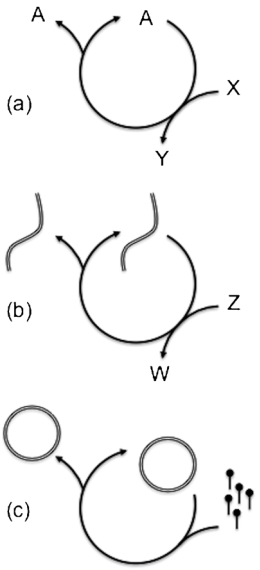


Fig. 1 Three autocatalytic susbsystems relevant for the origin of life. (a) Small molecules A catalyse their own synthesis using as external substrates X and generating waste products Y. (b) Template chemistry allows the generation of copies of double stranded polymers using activated monomers Z and generating waste products W. (c) Amphiphilic molecules self-assemble in vesicles with bilayer membranes. Adding more amphiphiles induces the growing and spontaneous division of vesicles.

Some authors prefer to take a radically different perspective by assuming that life started from some primitive anabolic routes, and an autotrophic origin, rather than a heterotrophic scenario is postulated (reviewed by Peretó[3]). In this case, the synthesis of organic matter from CO_2 requires the energy and electrons released from geochemical redox reactions, usually involving mineral compounds. In these cases, the authors usually invoke the submarine hydrothermal setting to provide the temperature and chemical gradients required by the autotrophic models.

Origins of genetics and metabolism

Some authors equate the origin of life to the origin of first replicators.[3] Theoretical insights and the discovery of catalytic RNAs (ribozymes) brought renewed interest in the idea of an early emergence of genetic polymers. This is one of the fundamental postulates of the *RNA world hypothesis*. According to this model, RNA served both as a genetic polymer and catalyst in the primitive metabolism. This hypothetical stage of evolution apparently solved the "chicken-and-egg" paradox of the origin of life: which came first, the informational polymers or the catalysts needed for their replication? Empirical approaches, such as *in vitro* RNA evolution, support the chemical plausibility of an RNA world. Evidence that RNA itself catalyses the peptide bond formation in extant ribosomes reinforces the antiquity and centrality of ribozymes. Moreover, the most universally conserved genes encode for RNA-related functions. The reader is referred to other recent reviews on the emergence of genetic polymers and the RNA world model.[7]

We may be very close to one of the key postulates of the RNA world scenario experimentally speaking, namely the

discovery of an RNA that could catalyse its own template-directed replication; notwithstanding, an explanation for the prebiotic emergence of a polymer as chemically complex as RNA remains elusive. This hurdle has led some authors to contemplate alternative models that favour the self-organisation of chemical networks as the material and energetic cradle of self-replicative polymers. This scenario, often known as *metabolism first*, emphasises the absence of a primordial genetic polymer.[3] How could a highly ordered polymer possibly emerge without a plausible and efficient way of funneling energy through the system? Not to mention the availability of the energetically and chirally adequate monomers and an efficient way to condensate them. Several authors have advocated the appearance of primitive chemical networks as a first step towards the synthesis of replicators. Would those self-organised prebiotic processes provide a protometabolic scaffold for the emergence of genetics? This paper reviews the debates on the origin of metabolism. On the other extreme of the timeline, we witness the extraordinary complexity of contemporary metabolic networks and the outstanding ability of extant microorganisms to adapt to many different environmental conditions. This review also summarizes some of the current efforts to understand the evolution of metabolisms from a systemic perspective.

Metabolic cycles

The concept of autocatalysis is a cornerstone to the understanding of how life works and of its origins, regardless of whether the attention is focused on metabolism, replicative polymers or reproductive compartments. Autocatalysis refers to the catalysis of a reaction (or sequence of reactions) by one or more of its products.[8] Furthermore, metabolism can be considered one of the central features of life and elucidating its origins has taken centre stage since Oparin started to develop his model. Theoretical constructs such as Gánti's chemoton,[9] Maturana and Varela's autopoiesis[10] and Rosen's (*M*,*R*) systems[11] put metabolism at the core of understanding life. These—and other theoretical endeavours—gravitate around the notion of metabolic closure, the idea of metabolism as a self-constructing machine. As stressed by Gánti,[9] metabolic networks might contain an autocatalytic core of small molecules, in addition to the autocatalytic cycles performed by macromolecular components (*i.e.* nucleic acids and proteins) and the autocatalytic reproduction of membranes.

In autotrophic cells metabolic pathways that convert CO_2 into cell chemical components may represent autocatalytic cycles (see below). But what about metabolic autocatalysis in heterotrophic cells? Might one also assume that with a complete set of enzymes and coenzymes inside a cell alone—without any small metabolite—metabolism would not start functioning? This question has been explored through the stoichiometric analysis of genome-wide metabolic models and a theoretically deduced minimal metabolism.[12] The result is that, in all the analysed cases, there is at least a metabolic core to autocatalytically produce ATP. Additionally, in a species-dependent manner, intermediary metabolism may show the autocatalytic production of other metabolites, such as NAD^+, coenzyme A, tetrahydrofolate (THF), quinones, and (in the case of autotrophic cells) sugars. The authors of the aforementioned studies conclude that these autocatalytic cores might represent the ancestral chemical devices necessary to kick-start metabolic networks. Thus, like the enzyme complement of the cell, some metabolites act as catalysts and metabolism regenerate them cyclically.[13]

Simple and autocatalytic cycles

According to Orgel,[14] metabolic cycles can be classified as simple or autocatalytic.§ Simple cycles allow the chemical transformation of substances with the stoichiometric regeneration of one of the reactants (*i.e.*, the feeder). In terms of standard chemical equations, the catalyst is ignored. For example, the transformation x → y, shown in Fig. 2a, requires the participation of a feeder A, which is also a product of the reaction. In this case, the molecule A that is produced replaces the one consumed. Furthermore, an autocatalytic cycle (Fig. 2b) exhibits an additional yield of the feeder A. Zachar and Szathmáry[18] have carefully examined the nature and properties of multiplying entities, including metabolic cycles. For these authors, an autocatalytic system must not only contain $n > 1$ elements among the set of products equivalent to the feeder A but, what is more, the rate of production of A must exceed the rate of its degradation or decay to secondary by-products.

True instances of metabolic (enzymatic) cycles, both simple and autocatalytic, can be found in extant metabolic networks. Thus, the tricarboxylic acids (TCA) or Krebs cycle, the ornithine (or urea) cycle, the pentose phosphate pathway, and the synthesis of thymidylate are examples of simple metabolic cycles. Each turn of the TCA cycle completely oxidises acetate to CO_2, consuming and regenerating a molecule of oxaloacetate (OAA), the feeder of the pathway (Fig. 3a). The ornithine cycle allows the synthesis of urea from HCO_3^-, NH_3 and the α-amino group of aspartate (Fig. 3b) regenerating the consumed ornithine at the end of the course of reactions. The complete oxidation of glucose-6-phosphate to CO_2 (and the corresponding reduction of $NADP^+$) may occur through the pentose phosphate pathway in which the produced pentoses (ribulose-5-phosphate) are stoichiometrically converted into glucose-6-phosphate (Fig. 3c). The methylation of uridine (from deoxyuridine monophosphate or uridylate, dUMP) in carbon 5 generates thymidylate (TMP), a precursor of TTP, which is an essential building block of DNA replication. The metabolic origin of the methyl group is mainly the side chain of the amino acid Ser, which is converted to Gly by the

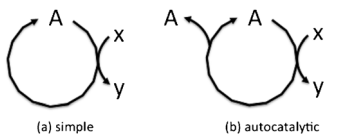


Fig. 2 According to Orgel,[14] metabolic cycles can be simple (a) or autocatalytic (b).

§ In the literature the terminology on catalytic cycles is diverse: Orgel's autocatalytic cycles are referred to as *circuit cycles* by Eschenmoser[15] to differentiate them from reaction *catalysts* or *genetic* reaction cycles, whose prototypical example is template chemistry with nucleic acid polymers. Also Morowitz *et al.*[16] distinguishes between *network autocatalytic* and *template autocatalytic*. For a systematic characterisation of catalytic cycles see Blackmond[17] and Zachar and Szathmáry.[18]

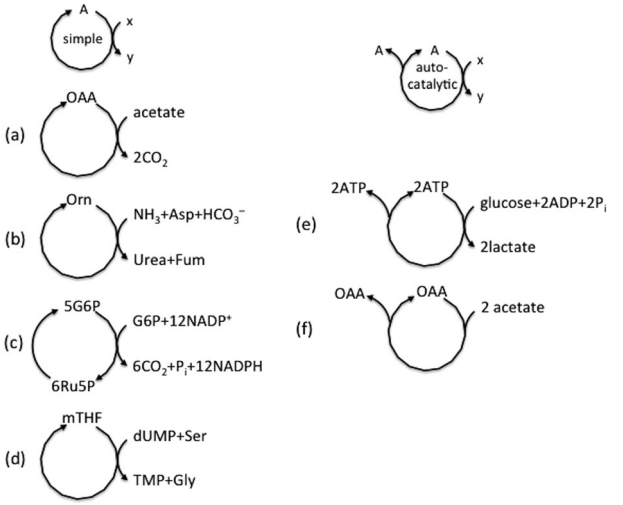


Fig. 3 Examples of metabolic cycles. (a) TCA cycle. (b) Ornithine cycle. (c) Pentose phosphate pathway. (d) Biosynthesis of TMP. (e) Glycolysis. (f) Glyoxylate cycle. Abbreviations: OAA, oxaloacetate; Fum, fumarate; G6P, glucose-6-phosphate; Ru5P, ribulose-5-phosphate; mTHF, N^5,N^{10}-methylene-tetrahydrofolate.

Ser hydroxymethyl transferase enzyme. In this process the side chain is transferred to tetrahydrofolate (THF) to yield N^5,N^{10}-methylene-THF, the mono-carbon unit donor participating in the thymidylate synthase reaction. The regeneration of the charged version of THF restarts the cycle (Fig. 3d).

It is also easy to find examples of autocatalytic cycles in extant cells. One obvious instance is ATP production by glycolysis. Phosphorylation of sugars (*e.g.*, glucose) by ATP is needed to initiate the standard glycolytic pathway. For the sake of simplicity, let us suppose that a redox balance is reached by the synthesis of lactate from pyruvate at the end of the pathway (*i.e.*, lactate fermentation). Stoichiometrically, two ATPs are needed for the preparatory phase of the pathway, whereas in the three-carbon compounds phase, four ATPs are produced. Thus, when one molecule of glucose is converted into two molecules of lactate, the pathway itself consumes two ATPs, whereas two additional ATPs constitute the net energetic yield (Fig. 3e). The glyoxylate cycle allows cell growth, since two-carbon molecules (*e.g.*, acetate) can serve as precursors for the net synthesis of four-carbon metabolites. Actually, it represents a by-pass of the two oxidative decarboxylations in the TCA cycle and can be represented as the stoichiometric synthesis of one OAA molecule from two acetate molecules (Fig. 3f).

Autotrophic cycles

The autotrophic mode of metabolism is required for planetary life, as we know it. Metabolic pathways leading to the net synthesis of cell carbon-based chemicals from CO_2 in autotrophic organisms provide biochemical food to the other consumers of organic molecules (heterotrophic organisms). There are six different ways of fixing biological carbon: the Calvin–Benson cycle, the Arnon cycle (reductive TCA— rTCA— or reverse Krebs cycle), the dicarboxylate–4-hydroxybutyrate cycle, the 3-hydroxypropionate–4-hydroxybutyrate cycle, the 3-hydroxybutyrate bi-cycle, and the Wood–Ljungdahl (reductive acetyl-CoA) pathway. The biochemical details of the

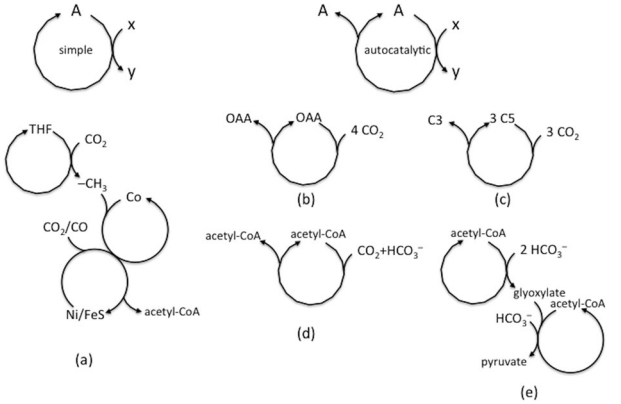


Fig. 4 The autotrophic cycles. (a) The net synthesis of acetyl-CoA from CO/CO_2 (Wood–Ljungdahl pathway). (b) The Arnon or reverse TCA cycle. (c) The Calvin–Benson cycle. (d) The hydroxybutyrate cycles. (e) The 3-hydroxybutyrate bi-cycle. Abbreviations: Co, Co-containing corrinoid; Ni/FeS, Ni–FeS cluster; C3, triose; C5, pentose.

different pathways can be found in the study of Fuchs.[19] In their study on the presence of autocatalytic pathways in extant living forms, Kun *et al.*[12] incorporated a reconstructed network of the cyanobacterium *Synechocystis*. When the model was allowed to work in autotrophic mode, they found the highest number of autocatalytic cycles from among all the analysed networks, including the one represented by the Calvin–Benson cycle producing sugars (from CO_2). Here, I represent the six autotrophic pathways in terms of the kind of the catalytic cycles (simple or autocatalytic) they show (Fig. 4). All the pathways, except the reductive acetyl-CoA pathway, are autocatalytic in the sense that the net product they generate represents a stoichiometric increase in the feeder (or a molecule equivalent to it) (compare Fig. 4a with Fig. 4b–e). Thus, the Arnon cycle (Fig. 4b) generates a C4 molecule (OAA) from four CO_2 molecules and regenerates OAA for the following turn of the cycle. The net product of carbon fixation in the Calvin–Benson cycle is a triose C3 (Fig. 4c). Although the substrate necessary to initiate the cycle is a pentose C5, in metabolic terms C3 and C5 are fully equivalent, since there is a complete set of enzymes performing the non-oxidative pentose-phosphate pathway that allows the interconversion of C3 and C5. In terms of carbon stoichiometries we can state the different phases of the cycle as follows:

$$\text{Carboxylative phase: } 3C5 + 3CO_2 \rightarrow 6C3$$

$$\text{Regenerative phase: } 5C3 \rightarrow 3C5$$

$$\text{Net production: } 3CO_2 \rightarrow C3$$

Both hydroxybutyrate cycles (the dicarboxylate–4-hydroxybutyrate cycle, and the 3-hydroxypropionate–4-hydroxybutyrate cycle) are variations on a theme: the net synthesis of acetyl-CoA from CO_2 and the regeneration of acetyl-CoA as the feeder of the cycle (Fig. 4d). The enzymatic differences between the two versions of the cycle are not relevant to this discussion. Finally, the 3-hydroxybutyrate bi-cycle can synthesise a C3 compound (pyruvate) from HCO_3^- (Fig. 4e) through the concatenation of two autocatalytic cycles. Although the feeder of both cycles

is acetyl-CoA, the two net products (glyoxylate and pyruvate) are its equivalents: glyoxylate is converted into pyruvate and pyruvate can easily be converted into acetyl-CoA by oxidative decarboxylation. Thus, these five autotrophic pathways qualify as autocatalytic cycles, since all of them synthesise the feeder, or an immediate precursor, from CO_2.[18]

The reductive acetyl-CoA pathway represents quite a different case (Fig. 4a). The net synthesis of acetyl-CoA from CO_2 is a linear pathway that proceeds through the consecutive action of three simple catalytic cycles. The cycles use cofactors as catalytic reagents: THF, Co-containing corrinoid, and a Ni–FeS cluster.[20] The net product of C fixation in this case (acetyl-CoA) does not meet the criterion of being equivalent to any of the regenerated feeders, thus the reductive acetyl-CoA pathway is not an autocatalytic cycle. This fact could have fascinating evolutionary implications (see below).

Autocatalysis and the origin of metabolism

Among abiotic non-enzymatic cycles, the only autocatalytic cycle with empirical support is the formose reaction. Discovered in the 19th century, this reaction corresponds to the polymerisation of formaldehyde and gives rise to a complex mixture of sugars. The presence of minerals has been explored as a means of decreasing the heterogeneity of the products, and of increasing the stability of some of them, namely those most prebiotically relevant, such as pentoses.[21]

Other cases of abiotic cycles of potential interest include:

i. A simple cyclic production of HCN tetramer from HCN in the presence of formaldehyde, as described by Alan Schwartz and Goverde.[22]

ii. The 'sugar model' proposed by Arthur Weber, based on the reaction of glyceraldehyde with NH_3 in aqueous solution.[23] In this system, the pyruvaldehyde formation rate is enhanced if the heterogeneous product mixture of nitrogen compounds is subsequently added to a fresh solution of glyceraldehyde. Although the underlying mechanism is unknown, an autoinductive, rather than an autocatalytic, reaction cycle is invoked.[17]

iii. Albert Eschenmoser's glyoxylate scenario is a hypothetical, yet very attractive, autocatalytic network starting with glyoxylate and 2,3-hydroxyfumarate (the 'formal' hydrates of CO) with the potential to generate a good variety of building blocks. Some of its postulates are now under experimental scrutiny.[24]

On more theoretical grounds, those who support the early emergence of metabolism suppose that self-organising chemical networks emerged spontaneously, before a template chemistry of genetic polymers had become established on the primitive Earth. The traditional dualism between models supporting "metabolism-first" and "genetics-first" may also be related to different notions of life: either as an emergent property of complex self-sustained systems or as the outcome of an evolutionary process based on the heredity of genetic records.[3,25] Clearly one of the main goals of origins research is to shed light on the origin of autocatalytic cycles under plausible prebiotic conditions. However, the main difference between primordial 'metabolic' and 'genetic' cycles is, at the long run, the vast potential of the structural and functional diversity of self-replicative polymers subject to the action of natural selection as compared to the autocatalytic generation of small organic molecules.[15] It is this huge difference in the evolvability of autocatalytic chemical networks that constrains their prevalence in a prebiotically plausible scenario.[26] To date, besides the formose reaction and a few other aforementioned candidates of chemical networks, protometabolic cycles remain in the domain of purely hypothetical chemistry.

One of the major limitations to studying the origin of metabolism is the discontinuity between prebiotic reactions and metabolic pathways believed to be present in the UC. Phylogenomic reconstructions based on extant metabolic diversity cannot be extrapolated beyond the origin of protein biosynthesis, likely in the RNA/peptide world.[27] Thus, some authors argue that prebiotic synthetic pathways would be completely different from the biosynthetic pathways of the UC.[28] In fact, besides a few cases, the prebiotic reactions described to date are very different from known metabolic pathways. For instance, the abiotic synthesis of amino acids, following the Strecker and Bucherer–Bergs mechanisms, greatly differs from any of the biosynthetic reactions operating in modern microorganisms. Nevertheless, several cases of parallelism between abiotic reactions and metabolic, enzyme-catalysed steps might indicate chemical determinism. Eschenmoser and Loewenthal[29] used the term *chemomimetic biosynthesis* in reference to the enzymatic mechanism for riboflavin synthesis, which resembles the organic process taking place *in vitro* without enzymes. For these authors, this is a clear example of how an enzyme (riboflavin synthase) co-opts and optimises a previously existing, non-enzymatic chemical transformation. There are other examples of chemomimetic processes (Costanzo *et al.*,[30] Lazcano[25a] and references therein), including: (i) the synthesis of intermediates of the purine *de novo* biosynthesis, 5-aminoimidazole-4-carboxamide (AICA) and 5-formamido-imidazole-4-carboxamide (fAICA), upon heating formamide in the presence of montmorillonites, albeit in the metabolic version the enzymes use the ribonucleotide-5′-monophosphate derivatives of AICA and fAICA as precursors of inosine-5′-monophosphate; (ii) the synthesis of orotic acid from aspartic acid and urea, and the decarboxylation of orotic acid yielding uracil; (iii) the reductive amination of 2-oxoglutarate to give glutamate; (iv) the UV-light-induced cyclisation of δ-aminolevulinic acid to yield pyrrole; and (v) the synthesis of acetic acid from CO and CH_3SH catalysed by NiS/FeS. This panorama led Lazcano and Miller[28] to propose that metabolism was actually invented from prebiotic processes that eventually vanished in UC biochemistry.

It may be argued that, so far, we have not explored all the possible prebiotic processes relevant for a primitive metabolism (or protometabolism)[31] and the above list of examples represents a rather limited set of instances of non-enzymatic organic chemistry anticipating extant metabolic pathways. In this light, approaches like systems chemistry could explore other chemical pathways revealing new chemomimetic processes out of fuzzy or less robust transformations. In the meanwhile, and despite the scarcity of experimental evidence, some authors have extrapolated chemical determinism to constitute the entire core of metabolic pathways. Hence, the quasi universal set of chemical transformations that extant organisms undergo to use carbon and energy sources would be a reflection of the oldest metabolic

map, built before the UC existed. Below, we summarise diverse discussions on primitive metabolisms:

i. As argued by de Duve, the replacement of (non-enzymatic) protometabolism by (ribozymatic and/or protein catalysed) metabolism required the emergence of catalysts by a selection process based on their usefulness under certain conditions. Thus, "only those enzymes that fitted within existing chemistry were conserved, from which it follows that protometabolism must have prefigured metabolism. The two are *congruent*; they followed similar pathways" (de Duve,[31] p. 19). Following this line of argument (the congruence principle initially proposed by de Duve in 1993) would give us access to the oldest evolutionary steps connecting prebiotic chemistry and protometabolism. The prevailing protometabolic chemistry would consist of unspecific reactions, under the control of rather inefficient catalysts, such as mineral surfaces and short non-encoded peptides (de Duve's *multimers*). Thus, the first enzymatic catalysts under natural selection (either ribozymes or, later on, genetically encoded protein enzymes) would play a key role in the transition from a "dirty protometabolism" (or *gemisch*) to a "cleaner metabolism". Although the author does not exclude the possibility that an enzyme generating new chemical conditions appeared by chance in the protometabolic landscape, such newcomers would be surpassed by chemical determinism and evolutionary continuity.

ii. The bottom-up continuity approach of de Duve is mirrored in the top-down work of comparative biochemists. Most strategies based on the comparative functional analysis of complete genomes result in a limited set of metabolic enzymes in the UC gene repertoire, albeit with a high degree of sequence redundancy indicative of the role likely played by ancestral gene duplications.[27] Several authors have discussed on the earliest stages of metabolic complexity comparing the repertoire of universally conserved enzymatic structural motifs and the metabolic abilities of extant microorganisms. A core of conserved bacterial proteins—related to RNA metabolism and connected to purine/histidine biosynthesis, riboflavin cofactor biosynthesis and branched-chain amino acids biosynthesis—has been likened to an archive of life's earliest stages.[25a,32]

iii. Based on chemical arguments and the universality of some biochemical processes, Morowitz[33] has proposed the rTCA cycle as the primordial metabolic core. From this early network of transformations, metabolism would emerge in a radial manner, adding successive layers of complexity to the original rTCA cycle. Similar arguments on chemistry and universality, as well as continuity, have been used by Meléndez-Hevia and coauthors.[34] Notwithstanding, they argue against the primitiveness of the rTCA and postulate that the main skeleton of primordial metabolism was glycolysis and a horseshoe, open Krebs cycle, irradiating from it to all other metabolic pathways.

iv. Wächtershäuser (ref. 35 and references therein) proposed a hypothetical autocatalytic carbon fixation process, mediated by transition metal complexes and driven by the electrons obtained from the anaerobic synthesis of pyrite from H_2S and FeS. The set of postulated reactions is a non-enzymatic version of rTCA functioning on the surface of the nascent pyrite crystal. Other presumed processes in this protometabolism would be the accumulation of lipids synthesised from activated acetic acid, the peptide synthesis from carbon fixation, and the formose reaction based on CO, all these processes operating on mineral surfaces (*i.e.*, surface metabolism hypothesis). New evidence points to molecular synergism between the organic products of CO fixation (amino acids Gly and Ala) and the process of further CN^- and CO fixation under aqueous conditions at high temperature and pressure.[36] All in all, this model has been related to the chemical setting and temperature conditions of submarine hydrothermal vents. Meanwhile, Russell and Martin (Martin[37] and references therein) have found inspiration in alkaline hydrothermal vents (like Lost City) for a model that invokes a key energetic role for H_2, produced by serpentinisation, and catalysis by transition metals, during the emergence of the earliest protometabolic process, *i.e.*, the synthesis of acetate from CO_2, as a geochemical precursor of the extant biochemical Wood–Ljungdahl pathway.

Primordial heterotrophy or autotrophy?

In the late 19th century, hypotheses on the origin of life were dominated by the idea of the early emergence of microorganisms able to fix carbon, not unlike extant photosynthesisers. Thus, as early as 1868, Haeckel suggested that the simplest protoplasmic substances originated from the fixation of inorganic carbonates, before their differentiation into individual primitive organisms, which he called *monera*. However, by 1924 it was all clear to the biochemist Oparin that the complexity of carbon fixation should be preceded by a simpler metabolic mode, namely, heterotrophic consumption of environmental organic matter.[2] In Oparin's time the fact meteorites contained carbon compounds was already known; furthermore, he was well aware of the numerous laboratory-based reports on the organic synthesis of many compounds of biological interest. Three facts caught Oparin's attention: the simplicity of fermentation pathways; their wide distribution among extant living organisms; and the supposed abundance of organic compounds on the early planet. This led him to the conclusion that the "consumption of organic matter is the oldest form of nutrition", advocating a heterotrophic origin of life.[38] In 1929, Haldane arrived at the same conclusion independently: "the first precursors of life found food available in considerable quantities [...]. As the primitive atmosphere contained little or no oxygen, they must have obtained the energy which they needed for growth by some other process than oxidation—in fact, by fermentation."[39] Today we know that primordial heterotrophy could be diverse among different anoxic transformations of high-energy abiotic organic compounds, which is not necessarily fermentation, as microbiologists understand it nowadays.[28] The strong experimental and observational support for the presence of carbon molecules on the primitive Earth is the cornerstone of the heterotrophic origin of life.

Following the Haeckelian tradition, there is also a school of thought defending a radically different scenario, *i.e.*, the primordial appearance of autotrophic pathways. Several authors invoke the emergence of reaction networks leading from CO or CO_2 to the diversity of organic molecules necessary for life. Regarding pathway type, as discussed above, Morowitz[33] and Wächtershäuser[35] advocate a primitive version of the rTCA cycle, whereas Russell and Martin propose a non-enzymatic synthesis of acetate from CO_2—similar to the contemporary Wood–Ljungdahl pathway (reviewed by Martin[37]).

Although some partial reactions of Wächtershäuser's model have found experimental support, the main process—namely, the fixation of CO_2 into the intermediates of the rTCA cycle—remains elusive. However, the reduction of dinitrogen or nitrate to ammonia, the synthesis of α-hydroxy acids and α-amino acids from CO, the synthesis of activated acetic acid thioester from CO and H_2S, or the synthesis of peptides have been demonstrated experimentally. At any rate, the efficiency of FeS and H_2S as reducing agents under strict anoxic conditions at moderately high temperatures (100 ± 20 °C) constitute an additional plausible source of organic compounds on the primitive Earth. Meanwhile, the hypothesis of a primitive version of the reductive synthesis of acetate from CO and CO_2—albeit conceptually systematic—only has circumstantial geochemical support.

Current phylogenomic studies have not drawn any sound conclusion on whether the UC was endowed with an autotrophic metabolism, and still less can be concluded on the primitiveness of autotrophy.[27] Even the apparently more approachable question as to which is the oldest pathway of carbon fixation remains unanswered.[19] Thus, considering their phylogenetic distribution, clearly the Calvin–Benson cycle, the rTCA cycle, and the hydroxypropionate bicycle seem idiosyncratic to bacteria. Conversely, the dicarboxylate–4-hydroxybutyrate and the hydroxypropionate–4-hydroxybutyrate cycles appear to be restricted to the Archaea domain. Finally, only the Wood–Ljungdahl pathway shows a wider distribution, since it is found in both prokaryotic domains, among strictly anaerobic species of Bacteria and Archaea: in methanogenic and sulfate reducing Euryarchaeota, acetogenic Firmicutes, some Spirochaetes, many δ-proteobacteria, and in the annamox bacteria of Planctomycetes.[19] Furthermore, it is the shortest and simplest (synthesis of a C2 compound from monocarbon molecules) autotrophic pathway and, if H_2 is available, demands less energy for biomass production.[40] In addition, it is the only pathway that can simultaneously fix carbon and synthesise ATP by substrate level phosphorylation if acetyl-CoA is converted to acetate.[20] Although the finer details are still unknown, the Wood–Ljungdahl pathway must also be linked to a chemiosmotic energy conservation mechanism.[19,20] Finally, the experimental simulations of C–C bond formation from CO and CH_3SH using NiS/FeS as catalysts indicate the antiquity of the Wood–Ljungdahl-like pathway for carbon fixation.[41]

There is another aspect supporting the simplicity of this autotrophic pathway, compared to the other five known pathways, namely, the Wood–Ljungdahl pathway is linear, resulting from the connection of three simple catalytic cycles, instead of being a true autocatalytic cycle (Fig. 4a), as discussed before. The participation of cofactors, such as organometallic catalysts, as members of the catalytic cycle, instead of organic intermediates, could speak in favour of its antiquity. As argued by Eschenmoser,[15] before the emergence of catalysts' autocatalytic replication and their incorporation in a cycle, the chemical environment was probably explored and exploited by non-autocatalytic catalysts. But some of those, such as low molecular mass organometallic catalysts or mineral surfaces, never became members of autocatalytic cycles. This could be the case of the Wood–Ljungdahl pathway catalysed by the precursors of the cofactors THF, corrinoid, and the Ni–FeS cluster if we accept the hypothesis of the primitiveness of catalysis by cofactors or their ancestors.

Almost half of all enzymatic reactions depend on at least one cofactor, either metals or organic complexes that complement and increase the catalytic toolkit of protein enzymes, playing essential roles in redox chemistry, group transfer or bond formation.[42] Except for hydrolases, enzymatic reaction classes contain at least 30% of cofactor-dependent enzymes. The most remarkable dependence is shown by oxidoreductases (more than 80% of the redox enzymes use organic cofactors). In fact, hydride transfer is an essential mechanism that is performed exclusively by cofactor-dependent enzymes.[42] Centrality, universality and constancy of cofactors in extant metabolism would suggest that they became part of the chemistry of life early on. Since the classical discussions introduced by Handler, King and White III,[43] among others, there is general agreement on a protometabolic role of cofactor-like molecules, *i.e.*, these reactive compounds would precede the establishment of full-fledged enzymes as suggested by the fact that many cofactors are reactive in the absence of proteins. In fact, the presence of ribonucleotide moieties in many cofactors has been used as an argument in favour of their adoption when the metabolism was still catalysed by ribozymes,[43c,44] thereby expanding the rather limited chemical capacity of catalytic RNAs. Extending White arguments[43c] we could consider extant enzymes as the third phase of metabolic evolution, preceded by a first stage with a prominent role of cofactors and a second one dominated by ribozymes and cofactors working together.

Among the participating cofactors in the Wood–Ljungdahl pathway, THF and corrinoid have also been considered as descendants of folate and tetrapyrrole-like molecules present in the primordial soup (reviewed by Holliday *et al.*[45]). The case of transition metals in the Ni–FeS cluster has an even better correlation with abiotic processes.[20,35,37] In summary, the wide phylogenetic distribution of the Wood–Ljungdahl pathway, its simplicity and low energetic cost, the participation of potentially old reagents and its non-autocatalytic nature might all be indicative of the early incorporation of the reductive net synthesis of acetyl-CoA from CO_2 to metabolism, and was likely the first autotrophic pathway.

Towards a systemic view of metabolic evolution

Metabolism in extant organisms is a chemical machinery of astonishing complexity and adaptability. For instance, the genome of a typical free-living bacterium feeding on organic matter, like *Escherichia coli*, has the potential to express enzymes catalysing *ca.* 1000 chemical reactions, and the functions of almost 20% of its genes have yet to be annotated. Indeed, a single *E. coli* cell, growing exponentially under standard aerobic conditions, contains 100 million metabolite molecules and 2.4 million of enzyme molecules.[46] The density of proteins inside a cell seems to be optimised to maximise the speed of biochemical reactions: enzyme-catalysed transformations are hardly rate-limiting for a cell, which appears to be more dependent on the kinetics of physical processes like molecular diffusion and protein folding.[47] To gain an understanding of

the harmonious steady state flux balance still represents a challenge to biochemists, biotechnologists, and systems biologists. Even more challenging is to try to delineate the processes that led from presumably small dirty protometabolic networks of chemical transformations to the sophisticated metabolic networks found in modern cells. Nevertheless, despite the emphasis on enzymatic specificity we learn from textbooks and the whole history of enzymology, it is clear that metabolic accuracy has its limits and stochastic factors, broad substrate specificity, and catalytic promiscuity¶ contribute to a somewhat fuzzy chemistry of metabolism, making it essential for evolution to occur.[48,49]

There are several models describing how metabolic pathways have diversified after protein enzymes became established (critically reviewed by Szathmáry *et al.*,[6] Holliday *et al.*[45] and Peretó[50]). It is generally assumed that the *patchwork* or *recruitment model*, which proposes that metabolism was initially performed by a small repertoire of enzymes with low specificity, offers the most comprehensive explanation. The oldest metabolic pathways may have been assembled by the recruitment of primitive enzymes showing remarkable substrate ambiguity and catalytic promiscuity. Specificity, specialisation and efficiency in metabolism were explored and expanded by duplication and divergence of the genes coding for metabolic enzymes (Peretó[50] and references therein). However, this must have occurred without completely eroding some of the ancestral characteristics of the primitive catalysers as many modern enzymes also exhibit broad substrate specificity and multifunctionality (Copley[51] and references therein). In fact, the co-option of latent enzymatic activities has been tested through protein resurrection experiments (for a recent example using secondary plant metabolism enzymes, see Huang *et al.*[52]) and genetic studies of microorganisms (see below). Enzyme recruitment has successfully been invoked to explain the evolution of many pathways, including the urea and TCA cycles, the autotrophic pathways, and the biosynthesis of amino acids, cofactors and membrane phospholipids (see Peretó[50] for the original references).

The plausibility of the transition from a few sloppy, multipurpose enzymes to a wider array of more specialised activities has been the object of theoretical models and simulations. By applying the "survival of the fastest" premise to their kinetic models, Kacser and Beeby[53] showed that primitive cells endowed with a small number of messy catalysers inevitably gave rise to the proliferation of more proficient enzymes, if the selective pressure favoured growth rate. What is more, multifunctionality seems deeply rooted in the very origin of metabolic closure, and hence in the origin of life, as suggested by theoretical models based on Rosen's (M,R) systems.[13]

Recent easy access to massive data on metabolic processes present in organisms from the three domains of life has boosted the development of new systemic approaches to the evolution of metabolism. Such approaches reach well beyond the arbitrary borders of individual metabolic pathways that researchers have delineated during the historical development of biochemistry. Thus, phylogenetic methods and functional analyses of genome-wide metabolic reconstructions allow us to test the predictions of the patchwork model regarding the mosaic distribution of homologous protein domains throughout contemporary metabolic networks.[54] Furthermore, it is remarkable that the distribution of kinetic constant values is dependent on the metabolic context, the enzymes of the central intermediate metabolism (*e.g.*, glycolysis, TCA cycle, and pentose phosphate pathway) being those with higher catalytic (k_{cat}) and specificity constants, whereas those of the secondary metabolism exhibit lower kinetic constant values,[55] that is, they are relatively slower and less specific. Conversely, catalytic promiscuity seems to be more widespread in central, amino acid and lipid metabolism but less abundant in secondary metabolism, although this conclusion is constrained by the accuracy and completeness of annotated promiscuous activities in current databases.[56]

At any rate, the uneven distribution of substrate ambiguity and catalytic promiscuity in the metabolic networks could reflect the evolutionary history of enzymes but could also be of adaptive value to organisms as a reservoir of metabolic innovations. In experiments performed with a collection of knockout mutants of *E. coli*, 20% of auxotrophs were rescued by the overexpression of multifunctional suppressors, either isozymes, enzymes (or transporters) showing substrate ambiguity or promiscuous enzymes.[57] In other words, extant metabolic networks are plastic and evolvable, thanks to the intrinsic structural and functional flexibility of enzymes[48,58] allowing organisms to successfully navigate the messy[48] underground[59] of metabolism.

The remarkable capacity of microorganisms to innovate has also been addressed from the systems biology perspective.[60] Computational models and experimental evolution under lab conditions show that microorganisms survive many deletions of metabolic enzyme-coding genes. This robustness of their metabolic webs is consistent with the extraordinary ability of free-living microorganisms to cope with environmental changes in terms, for instance, of degrading new carbon sources or xenobiotics. Although the evolutionary underpinnings of such metabolic robustness remain elusive,[60] redundancy of usually latent enzymatic circuits enables them to remain viable under a wide range of environments. Counterexamples can be found in microorganisms adapted to intracellular life, like parasites and mutualist endosymbionts. In this case, the reductive evolution observed in their genomes causes a drastic loss of robustness. However, the chemical constancy of the intracellular environment compensates for this metabolic fragility (Belda *et al.*[61] and references therein). Wagner[60] suggests that metabolic networks, regulatory circuits and macromolecules show a phenotypic diversity based on their mutational robustness, and that this is the fuel of most evolutionary innovations. In their exploration into the evolution of metabolic networks, Rodrigues and Wagner[62] found that many different genotypes might be functionally equivalent, whereas small changes in genotypes sharing the same phenotype gave rise to remarkable phenotypic diversification. Thus, natural selection could easily explore the different solutions to the same kind of problem in the phenotypic landscape of microorganisms. This combination

¶ In this paper I refer to substrate ambiguity as the lack of strict specificity in an enzyme recognising a series of more or less related substrates. There is enzymatic or catalytic promiscuity (or multifunctionality) when in addition to the native activity (*i.e.*, the one emerged under selective pressures), the enzyme shows other activities that are, in principle, physiologically irrelevant.

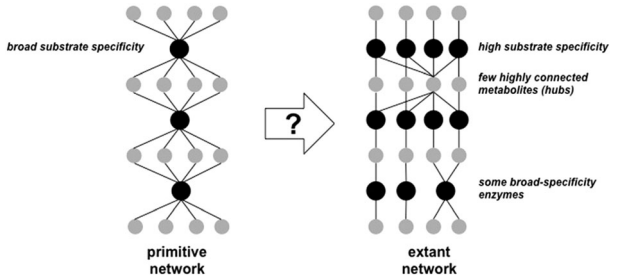


Fig. 5 Topological aspects of primitive networks and extant metabolic networks. In protometabolic networks catalysers (black points) exhibited broad substrate specificity (substrates and products of reactions represented by grey points). In present day metabolisms many enzymes show high substrate specificity, whereas some other exhibit broad-specificity. As a consequence, most metabolites are poorly connected with enzymes although a few of them (*e.g.*, coenzymes) are highly connected (hubs).

of plasticity and robustness seems to be a fundamental property of metabolic networks, essential to both adaptation and evolution.

Let us finish with a cautionary tale. The application of network thinking to metabolism may provide insights into metabolic evolution[63] although researchers are not always using the best graph projections or the most appropriate ways of defining network interactions in metabolism.[64] The proliferation of studies based on simplistic approaches to metabolic complexity pose the risk that many irrelevant conclusions on the structure, dynamics and evolution of metabolism may be drawn. As proposed by Solé and co-workers[64] the more realistic and informative way of representing metabolic networks is by using bipartite graphs. In these graphs two types of nodes are considered, namely metabolites and enzymes, and the edges only connect nodes of different kinds. In most studies unipartite graphs are used, where the nodes are metabolites (or enzymes), and the edges, enzymes (or metabolites). In fact, these unipartite graphs are the corresponding metabolite or enzyme projections of a bipartite graph (for definitions, examples and original references, see Montañez *et al.*[64]). In the last decade, topological analysis has suggested some remarkable properties of metabolic networks, also found in other natural and artificial webs, like small-worldness, scale-freeness and hierarchical modularity.[63] But these properties of metabolic networks, however attractive and realistic they might sometimes seem, are seriously affected by the use of unipartite graphs and may actually be artefactual. For a critical review on the weaknesses of the metabolic network analysis, the reader is addressed to Lima-Mendez and van Helden.[65] Thus, the study of metabolic networks should use more natural representations of them, like bipartite graphs, in order to portray and comprehend metabolism from an evolutionary systems-biology perspective. The final goal would be to understand how small networks of pretty inefficient enzymes became big networks of more proficient enzymes, with some metabolites acting as hubs (Fig. 5).

Conclusions

Oparin[38] said that scientists explore the origin of life "like two parties of workers boring from the two opposite ends of a tunnel", following the chemically oriented bottom-up strategy or the biologically based top-down approach. There is no doubt that the emergence and evolution of metabolism is one of the central debates in origin-of-life research, but also one that has been addressed less systematically from an experimental viewpoint. A key question is how the first autocatalytic cycles became incorporated into the chemistry of life. Unfortunately, we only have one example of a true autocatalytic abiotic cycle (the formose reaction), and a great deal of research lies ahead in the field of systems chemistry, focusing on the identification and characterisation of small self-sustained chemical networks. At any rate, current models indicate that some chemical dirtiness has accompanied metabolism since its inception and during its evolution, from small protometabolic networks to the metabolic complexity of extant cells. The absence of absolute specificity and the multifunctionality of catalysers appear to be compulsory qualities in models representing both the origin of metabolism and the evolution of metabolic pathways. The emergence and expansion of metabolic complexity during the post-enzymatic era must be addressed from phylogenomic and systemic perspectives. Thus, today, Oparin's two parties of workers need to use the tools of systems chemistry and systems biology.

Acknowledgements

Financial support by EU grant ST-FLOW and MICINN (BFU2009-12895-C02-01) is gratefully acknowledged. Thanks are given to José Aguilar-Rodríguez, Antonio Lazcano, Kepa Ruiz-Mirazo and the anonymous referees for helpful comments on the manuscript.

Notes and references

1 D. Wacey, M. R. Kilburn, M. Saunders, J. Cliff and M. D. Brasier, *Nat. Geosci.*, 2011, **4**, 698–702.
2 A. Lazcano, *Cold Spring Harbor Perspect. Biol.*, 2010, **2**, a2002089.
3 J. Peretó, *Int. Microbiol.*, 2005, **8**, 23–31; I. Fry, *Origins Life Evol. Biospheres*, 2011, **41**, 3–16.
4 J. W. Szostak, *Philos. Trans. R. Soc. London, Ser. B*, 2011, **366**, 2894–2901.
5 M. A. Powner and J. D. Sutherland, *Philos. Trans. R. Soc. London, Ser. B*, 2011, **366**, 2870–2877.
6 E. Szathmáry, M. Santos and C. Fernando, *Top. Curr. Chem.*, 2005, **259**, 167–211.
7 T. Cech, *Cell*, 2009, **136**, 599–602; A. Ricardo and J. W. Szostak, *Sci. Am.*, 2009, **301**, 54–61; M. P. Robertson and G. F. Joyce, *Cold Spring Harbor Perspect. Biol.*, 2010, **2**, a003608.
8 O. Taran and G. von Kiedrowski, *et al.*, in *Encyclopedia of Astrobiology*, ed. M. Gargaud, Springer, Heidelberg, 2011, vol. 1, pp. 128–129.
9 T. Gánti, *The Principles of Life*, Oxford University Press, Oxford, 2003. This edition contains the English translation of Gánti's main works originally published in Hungarian together with many footnotes and two explanatory chapters on the biological and philosophical significance of those works by E. Szathmáry and J. R. Griesemer, respectively.
10 H. R. Maturana and F. Varela, *De máquinas y seres vivos*, Editorial Universitaria, Santiago de Chile, 1973, (in Spanish); H. R. Maturana and F. Varela, *Autopoiesis and Cognition: the Realisation of the Living*, D. Reidel Publishing Company, Dordrecht, 1980.
11 R. Rosen, *Life Itself: a Comprehensive Inquiry into the Nature, Origin and Fabrication of Life*, Columbia University Press, New York, 1991; J. C. Letelier, M. L. Cárdenas and A. Cornish-Bowden, *J. Theor. Biol.*, 2011, **286**, 100–113 and references therein. In this paper, the authors analyse the concept of *metabolic closure*

and compare, among others endeavours in the understanding of life, the autopoiesis concept and the models of chemoton and (*M*,*R*) systems.
12 A. Kun, B. Papp and E. Szathmáry, *Genome Biol.*, 2008, **9**, R51.
13 A. Cornish-Bowden and M. L. Cárdenas, *Chem. Biodiversity*, 2007, **4**, 2396–2406.
14 L. E. Orgel, *PLoS Biol.*, 2008, **6**, e18.
15 A. Eschenmoser, *Origins Life Evol. Biospheres*, 2007, **37**, 309–314.
16 H. J. Morowitz, J. D. Kostelnik, J. Yang and G. D. Cody, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 7704–7708.
17 D. G. Blackmond, *Angew. Chem., Int. Ed.*, 2009, **48**, 386–390.
18 I. Zachar and E. Szathmáry, *BMC Biol.*, 2010, **8**, 21.
19 G. Fuchs, *Annu. Rev. Microbiol.*, 2011, **65**, 631–658.
20 S. W. Ragsdale and E. Pierce, *Biochim. Biophys. Acta*, 2008, **1784**, 1873–1898.
21 H.-J. Kim, A. Ricardo, H. Illangkoon, M. J. Kim, M. A. Carrigan, F. Frye and S. A. Benner, *J. Am. Chem. Soc.*, 2011, **133**, 9457–9468.
22 A. W. Schwartz and M. Goverde, *J. Mol. Evol.*, 1982, **18**, 351–353.
23 A. Weber, *Origins Life Evol. Biospheres*, 2007, **37**, 105–111.
24 A. Eschenmoser, *Angew. Chem., Int. Ed.*, 2011, **50**, 12412–12472; V. N. Sagi, V. Punna, F. Hu, G. Meher and R. Krishnamurthy, *J. Am. Chem. Soc.*, 2012, **134**, 3577–3589.
25 (*a*) A. Lazcano, *Origins Life Evol. Biospheres*, 2010, **40**, 161–167; (*b*) K. Ruiz-Mirazo, J. Peretó and A. Moreno, *Origins Life Evol. Biospheres*, 2004, **34**, 323–346.
26 V. Vasas, E. Szathmáry and M. Santos, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 1470–1475.
27 A. Becerra, L. Delaye, S. Islas and A. Lazcano, *Annu. Rev. Ecol. Evol. Syst.*, 2007, **38**, 361–379.
28 A. Lazcano and S. L. Miller, *J. Mol. Evol.*, 1999, **49**, 424–431.
29 A. Eschenmoser and E. Loewenthal, *Chem. Soc. Rev.*, 1992, **21**, 1–16.
30 G. Costanzo, R. Saladino, C. Crestini, F. Ciciriello and E. Di Mauro, *BMC Evol. Biol.*, 2007, **7**(Suppl. 2), S1.
31 C. de Duve, *Singularities: Landmarks on the Pathway of Life*, Cambridge University Press, Cambridge, 2005.
32 A. Danchin, G. Fang and S. Noria, *Proteomics*, 2007, **7**, 875–889; G. Caetano-Anollés, H. S. Kim and J. E. Mittenthal, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 9358–9363.
33 (*a*) H. J. Morowitz, *Beginnings of Cellular Life*, Yale University Press, New Haven, 1992; (*b*) H. J. Morowitz, *Complexity*, 1999, **4**, 39–53; (*c*) E. Smith and H. J. Morowitz, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 13168–13173.
34 E. Meléndez-Hevia, N. Montero-Gómez and F. Montero, *J. Theor. Biol.*, 2008, **252**, 505–519.
35 G. Wächtershäuser, *Prokaryotes*, 2006, **1**, 275–283.
36 C. Huber, F. Kraus, M. Hanzlik, W. Eisenreich and G. Wächtershäuser, *Chem.–Eur. J.*, 2012, **18**, 2063–2080.
37 W. F. Martin, *Biol. Direct*, 2011, **6**, 36.
38 A. I. Oparin, *Proiskhozhedenie Zhizni*, Moscow, 1924. Reprinted and translated in: J. D. Bernal, *The Origin of Life*, Weidenfeld and Nicolson, London, 1967.
39 J. B. S. Haldane, *Rational. Ann.*, 1929, 3–10.
40 N. R. Boyle and J. A. Morgan, *Metab. Eng.*, 2011, **13**, 150–158.
41 J. Peretó, A. M. Velasco, A. Becerra and A. Lazcano, *Int. Microbiol.*, 1999, **2**, 3–10.
42 J. D. Fischer, G. L. Holliday, S. A. Rahman and J. M. Thornton, *J. Mol. Biol.*, 2010, **403**, 803–824.
43 (*a*) P. Handler, in *Proc. 5th Int. Congr. Biochem*, ed. A. I. Oparin, Macmillan, New York, 1961, pp. 149–157; (*b*) G. A. M. King, *BioSystems*, 1980, **13**, 23–45; (*c*) H. B. White III, *J. Mol. Evol.*, 1976, **7**, 101–104; (*d*) H. B. White III, *The Pyridine Nucleotide Coenzymes*, ed. J. Everse, B. Anderson and K.-S. You, Academic Press, New York, 1982, pp. 1–17.
44 V. R. Jadhav and M. Yarus, *Biochimie*, 2002, **84**, 877–888.
45 G. L. Holliday, J. M. Thornton, A. Marquet, A. G. Smith, F. Rébeillé, R. Mendel, H. L. Schubert, A. D. Lawrence and M. J. Warren, *Nat. Prod. Rep.*, 2007, **24**, 972–987.
46 B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien and J. D. Rabinowitz, *Nat. Chem. Biol.*, 2009, **5**, 593–599.
47 K. A. Dill, K. Ghosh and J. D. Schmit, *Proc Natl. Acad. Sci. U. S. A.*, 2011, **108**, 17876–17882.
48 D. S. Tawfik, *Nat. Chem. Biol.*, 2010, **6**, 692–696.
49 O. Khersonsky and D. S. Tawfik, *Annu. Rev. Biochem.*, 2010, **79**, 471–505.
50 J. Peretó, in *Origins and Evolution of Life: An Astrobiological Perspective*, ed. M. Gargaud, P. López-García and H. Martin, Cambridge University Press, Cambridge, 2011, ch. 18, pp. 270–287.
51 S. D. Copley, *J. Biol. Chem.*, 2012, **287**, 3–10.
52 R. Huang, F. Hippauf, D. Rhorbeck, M. Haustein, K. Wenke, J. Feike, N. Sorrelle, B. Piechulla and T. J. Barkman, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 2966–2971.
53 H. Kacser and R. Beeby, *J. Mol. Evol.*, 1984, **20**, 38–51.
54 R. R. Copley and P. Bork, *J. Mol. Biol.*, 2000, **303**, 627–641.
55 A. Bar-Even, E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik and R. Milo, *Biochemistry*, 2011, **50**, 4402–4410.
56 P. Carbonell, G. Lecointre and J.-L. Faulon, *J. Biol. Chem.*, 2012, **286**, 43994–44004.
57 W. M. Patrick, E. M. Quandt, D. B. Swartzlander and I. Matsumura, *Mol. Biol. Evol.*, 2007, **24**, 2716–2722.
58 N. Tokuriki and D. S. Tawfik, *Science*, 2009, **324**, 203–207.
59 R. D'Ari and J. Casadesús, *BioEssays*, 1998, **20**, 181–186.
60 A. Wagner, *Trends Genet.*, 2011, **27**, 397–410.
61 E. Belda, F. J. Silva, J. Peretó and A. Moya, *PLoS One*, 2012, **7**, e30652.
62 J. F. M. Rodrigues and A. Wagner, *PLoS Comput. Biol.*, 2009, **5**, e1000613.
63 T. Yamada and P. Bork, *Nat. Rev. Mol. Cell Biol.*, 2009, **10**, 791–803.
64 R. Montañez, M. A. Medina, R. V. Solé and C. Rodríguez-Caso, *BioEssays*, 2010, **32**, 246–256.
65 G. Lima-Mendez and J. van Helden, *Mol. Biosyst.*, 2009, **5**, 1482–1493.