Small-molecule autocatalytic networks are universal metabolic fossils

Joana C. Xavier[1*], Stuart Kauffman[2]

1. *Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK, https://orcid.org/0000-0001-9242-8968. 2. Institute for Systems Biology, Seattle, USA.*

Summary

Life and the genetic code are self-referential, and so are autocatalytic networks made of simpler, small molecules. Several origins of life theories postulate autocatalytic chemical networks preceding the primordial genetic code, yet demonstration with biochemical systems lacks. Here, small-molecule reflexively autocatalytic food-generated networks (RAFs) ranging in size from 3 to 619 reactions were found in all of 6683 prokaryotic metabolic networks searched. The average maximum RAF size is 275 reactions for a rich organic medium and 93 for a medium with a single organic cofactor, NAD. In the rich medium, all universally-essential metabolites are produced with the exception of glycerol-1-p (archaeal lipid precursor), phenylalanine, histidine and arginine. The 300 most common reactions, present in at least 2732 RAFs, are mostly involved in amino acid biosynthesis and the metabolism of carbon, 2-oxocarboxylic acid and purines. ATP and NAD are central in generating network complexity, and because ATP is also one of the monomers of RNA, autocatalytic networks producing redox and energy currencies are a strong candidate niche of the origin of a primordial information-processing system. The wide distribution of small-molecule autocatalytic networks indicates that molecular reproduction may be much more prevalent in the Universe than hitherto predicted.

Special Issue: *Emergent phenomena in complex physical and sociotechnical systems: from cells to societies*

*Author for correspondence (joana.xavier@ucl.ac.uk).

†Present address:

Department of Genetics, Evolution and Environment, University

College London, Gower Street, London WC1E 6BT, UK

1 Introduction

2

3 Around four billion years ago or before, in a geochemical setting not far away, a major split occurs: the
4 partition between chemical information storage (a primordial genotype) and chemical function (a primordial
5 phenotype) into different yet collectively self-reproducing molecules. That moment, the origin of a primitive
6 code, is seminal in biology perhaps like no other. Yet, it remains a deep mystery in origins literature [1–5], the
7 complexity of which remains unmatched by any known abiotic system. Whereas the canonical genetic code is
8 the set of rules by which cells translate genotype to phenotype (i.e. triplets of nucleotides in DNA to amino
9 acids in proteins), the process of *coding* (comprising both encoding and decoding by replication, transcription
10 and translation) seems irreducible, involving hundreds of *encoded* proteins. This makes life and coding self-
11 referential, while also requiring all other processes the cell needs for energy, biosynthesis and homeostasis.
12 Today, the smallest self-replicating cells known (highly dependent on more complex life) have genomes with
13 ~112kbp [6]. In each cell there are also hundreds of thousands of RNA molecules [7], millions of proteins and
14 lipids [8,9] and hundreds of millions of small molecules [10]. This is substantial complexity, self-replication
15 and self-organization in a tiny, tiny space. After the origin of coding life was *en route* to the last universal
16 common ancestor (LUCA) [11], bacteria [12,13] and archaea [14]. Before and towards that moment hypotheses
17 abound, all of them sharing the quest for the identity and mechanism of the first self-replicating evolving
18 entities [15–17], also known as the Initial Darwinian Ancestor (IDA) or First Universal Common Ancestor
19 (FUCA) [2]. As mentioned, known minimal self-replicating cells do not exist without the translation,
20 transcription and replication apparatuses, yet strikingly they require few encoded catalytic sequences
21 (enzymes) when provided with a stable environment and supply of nutrients [18–20]. One appealing
22 hypothesis is that a similarly stable environment—notwithstanding important thermal and electrochemical
23 gradients [21–23]—with steady nutrient supply was required at the origin of a primordial information-
24 processing machinery. Small-molecule collectively autocatalytic networks can potentially generate stable and
25 persistent yet evolvable biochemical production, particularly when combined with compartmentalization [24–
26 29]. Therefore, their causal and temporal precedence over a primordial genetic code (a set of rules translating
27 simpler self-replicating genetic molecules and peptides) is parsimonious. Recent work suggested that
28 autocatalysis, in its varied stoichiometric motifs, must be much more widespread in chemical systems than
29 hitherto assumed [30]. In previous work, a particular type of small-molecule non-enzymatic autocatalytic
30 networks was predicted in the metabolism of *Escherichia coli* [31], and later in two prokaryotes regarded as
31 ancient as well as in the global space of oxygen-independent prokaryotic metabolism, producing several
32 amino acids and bases [32]. This type of autocatalytic network is called RAF—reflexively food-generated
33 autocatalytic network—a network in which each reaction is catalysed by a molecule from within the network,
34 and where all molecules can be produced from a set of food molecules by the network itself. The largest RAF
35 found in a biochemical system (with the maximum possible number of reactions) is called maxRAF [33,34].
36 Here, the metabolic networks of 6683 prokaryotes were searched for maxRAFs without encoded enzymatic
37 activity. We show that small-molecule RAFs are universally embedded in prokaryotic metabolic networks,

38   working with combinations of all the different possible trophic modes known in the simplest life forms. We
39   also expose and analyse the centrality of the universal metabolic currencies—the NAD/ATP pair—in
40   autocatalysis in prokaryotic metabolism, and reflect on the implications of these findings for future
41   experiments on the emergence of sufficient and stable chemical diversity for the origin of protocells.
42
43

Results and Discussion

45

Small-molecule autocatalysis is a universal metabolic fossil

47

48   The metabolisms of three prokaryotes—*Escherichia coli* [31], *Methanococcus maripaludis* and *Moorella*
49   *thermoacetica* [32]—were previously explored for small-molecule autocatalytic networks, as well as the global
50   space of oxygen-independent prokaryotic metabolism [32]. In all of those networks, RAFs were found without
51   the aid of enzymatic catalysis. The distribution and composition of small-molecule RAFs in metabolic
52   networks made up of different reaction combinations can inform on the nature, emergence and evolution of
53   self-replicating chemical systems. Here we ask which of all 6683 prokaryotic genomes with metabolic
54   annotations in KEGG contain small-molecule RAFs (full list, including an additional 110 genomes not
55   annotated, in Supplementary Data 1). With a rich medium (elsewhere referred to as food set) containing
56   inorganic catalysts, organic carbon and cofactors and small-molecule catalysis annotations as used before (See
57   Methods, Supplementary Data 2-3 and [32]), all genomes contained RAFs, ranging in maximum size
58   (maxRAF) from 3 to 619 reactions (Fig. 1a; Supplementary Data 4). We note that the smallest maxRAF possible
59   contains 3 reactions classified as spontaneous, and these are allowed to occur in all simulations (see Methods).
60   The average maxRAF size obtained with a rich medium is 275 reactions.

61

62   Previous work exposed the centrality of NAD in generating complex autocatalytic networks based on
63   prokaryotic metabolism [32,34], and therefore here we proceeded to ask if maxRAFs could be found just with
64   this one organic cofactor in the growth medium of all prokaryotes. In a medium with just NAD and no other
65   organic cofactors, maxRAFs ranged in size from 3 to 232 reactions, with the same smallest RAF possible
66   consisting of spontaneous reactions, and with average maxRAF size of 93 reactions (Fig. 1b; Supplementary
67   Data 5).

68

69   The nature of metabolism is such that a very small number of molecules catalyses a large number of different
70   reactions [32,35] (see Conclusion). Furthermore, reactions fall within classes assigned to the enzymes known
71   to catalyse them; these are Enzyme Commission (E.C.) numbers, and the seven umbrella categories comprise
72   redox reactions, functional transfers, hydrolyses, non-hydrolytic eliminations or additions, isomerizations,
73   ligations and (recently classified) translocations. In many of these, the enzyme is the pocket where the small-
74   molecule cofactor is the *de facto* catalyst. Regardless, the number of E.C. annotations in a genome is a proxy

75 for the chemical complexity predicted to occur in the metabolism of a single cell. The maxRAF size grows

76 linearly with the number of E.C. annotations assigned to each genome, both in the archaeal and bacterial

77 domains (Fig. 1). This indicates that the structure of autocatalysis embedded in metabolism is such that it is

78 able to grow at least linearly with the addition of new possible chemical routes.

79

80 **Figure 1 – Size of largest autocatalytic networks (maxRAFs) obtained vs. number of different Enzyme**

81 **Commission (EC) annotations for 6683 prokaryotic metabolic networks.** a) maxRAF size with a rich medium

82 containing organic cofactors, metals, and organic carbon. b) maxRAF size when medium contains only

83 inorganic cofactors, organic carbon and NAD. Blue represents bacterial networks and orange archaeal ones.

84

85 Of 5994 reactions annotated with small-molecule catalysts, 1070 are used at least once in a prokaryotic

86 maxRAF with a rich organic medium (Supplementary Data 4). But how many reactions are very often used in

87 several maxRAFs? We found that the most prevalent 300 reactions appear in at least 2732 maxRAFs, and

88 participate in pathways all across metabolism, majorly in amino acid biosynthesis, carbon metabolism, 2-

89 oxocarboxylic metabolism, purine metabolism, the glycolysis/gluconeogenesis trunk and pyruvate

90 metabolism (Fig. 2). These are quite central pathways in prokaryotic metabolism. Even though very few, if

91 any, genetic sequences are universally conserved [36], the analysis of prokaryotic metabolism done here and

92 elsewhere reveals that chemistry, and small-molecule autocatalysis, are heavily conserved and a universal

93 feature of life as we know it on Earth.

94

95 **Figure 2 – Annotations for the 300 metabolic reactions most prevalent in prokaryotic maxRAFs.** The boxes are

96 sized proportionally to the number of reactions annotated to each functional category (number after the

97 comma), taking into account the annotations of the 300 reactions most prevalent in all prokaryotic maxRAFs

98 obtained for 6683 prokaryotic metabolic networks with a rich medium.

99

100 The linear growth of maxRAF sizes with the expanding biochemical space (Fig. 1) indicates that small-

101 molecule maxRAFs can potentially generate sufficient chemical diversity—nucleotides, amino acids and other

102 essential small organic catalysts—necessary for the emergence of small-polymers that would then lead to a

103 primordial genetic code in a single localized system. We therefore proceeded to ask which of the universal

104 molecules required for the essential coding machinery in prokaryotic biomass can be produced by small-

105 molecule RAFs embedded in prokaryotic metabolism.

106

107 **Metabolic currencies required before coding**

108

109 To further investigate the potential of RAFs in individual metabolic maps, we asked which networks

110 produced molecules considered universally-essential for a primordial genetic code: amino acids, nucleobases,

111 lipid precursors for archaea and bacteria respectively [37] and cofactors. We note that cofactors were all

112 provided in the rich medium, but not always used (Fig. 3). Strikingly, the medium with just one organic

cofactor, NAD, allowed for the production of the bacterial lipid precursor but not the archaeal, all RNA bases (UTP and CTP in only 10 metabolic networks) and only four amino acids – alanine, asparagine, aspartic acid and cysteine. As before [34] we note that NAD is converted to ATP in a reversible reaction used by 94% of all maxRAFs (both in the Rich or NAD-only media) - the Nicotinic Acid Mononucleotide Adenylyltransferase reaction. The implications of this interconversion in autocatalysis were previously explored [34], but here we expand on these with a broader perspective. This reaction is the simplest link in metabolism between ATP and NAD, the most prevalent biochemical currencies (energy and redox, respectively), which are essential in any known minimal metabolism. The appearance of the pair of currencies, both containing adenine in its structure, is a particularly interesting moment in the origins' path, not only because ATP is also a genetic monomer in RNA chains, but also because both NAD and ATP are cofactors in a very large portion of the metabolic space [32]. Redox and energy currencies allow for coupling between different parts of biochemical networks, one of the most striking features of metabolism: energy is stored in the ATP molecule, and reductive power in the NADH molecule, and those are delivered to endergonic and anabolic steps respectively, which significantly increase network complexity. These represent a major innovation in chemical function before the origin of coding, one that is another deep question in origins [2]. We return to the question of the origin of energetic and redox couplings in the Conclusion.

Energy and redox currencies are not the only essential cofactors in a minimal metabolism, though, and the simulation with NAD as the single organic cofactor in the medium confirms that there are other very important catalytic bottlenecks in prokaryotic biochemistry as we know it (Fig. 2). When NAD is the only organic cofactor added to the medium, the majority of amino acids cannot be produced, and the signal for all deoxyribonucleotides other than dATP is lost (dATP can be produced from ATP and formate). This indicates that there is a vast number of reactions not included in biochemical databases that were of utmost importance in prebiotic chemistry, or, that there is a vast number of catalysis rules hitherto unknown. Promisingly, recent experiments have shown the potential and ubiquitousness of metals in non-enzymatic catalysis [38–40]. Also, other organic catalysts are known to be efficient in the absence of the respective enzyme [41–45] and some have been produced abiotically (reviewed in [46]).

The results with NAD as the only organic cofactor in the medium indicate that very important pieces in the puzzle of the origin of metabolism and life are missing, but also that when redox and energy coupling emerged, significant chemical organization could also emerge. Assuming that RNA arose before ATP is nearly impossible, as ATP is a monomer of RNA (the activated form). RNA polymerases only work with NTPs and most prebiotic chemists agree that NTPs precede RNA chains [47,48] as NMPs cannot be polymerized without strong condensing agents. Assuming that once AMP (ADP or ATP) arose it only participated in RNA polymerization seems naïve. It is possible and quite plausible that the first currencies invented in prebiotic chemistry were simpler than the pair NAD/ATP, which is universal in prokaryotes and therefore emerging front and centre in our results. Yet, for the origin of genetic polymers concurrent with sufficient amino acids for peptides and a primordial code, the pair together with other organic cofactors seems to be essential (Fig. 3).

Phil. Trans. R. Soc. A.

151

Figure 3 – Production of universally-essential molecules by maxRAFs obtained in 6683 prokaryotes. Molecules
are organized in classes of amino acids, bases, cofactors (all provided in the simulation with rich medium) and
lipid precursors. The size of the bars corresponds to the percentage of maxRAFs that can produce the given
molecule in the two sets of simulations (grey in rich medium, black in a medium with a single organic
cofactor, NAD).

Conclusion

In the elusive path towards the origin of the first cells, for functional and selectable polymers to arise
monomers must have been produced steadily and consistently. Autocatalytic chemical networks made of
small-molecules are not only a solution to that requirement, extensively fulfilled by distinct organic syntheses
even if in non-compatible conditions (reviewed in [15,49]); autocatalytic networks can provide closure [50], a
localized emergence of self-referential reproduction in the Universe. Autocatalytic chemical networks open
doors to self-organization, persistence, order, and catalytic function expressed in a primitive phenotype that
can evolve upon growth and modification [25–27,29].

Previously, we showed that the biochemistry of three prokaryotes concealed small-molecule autocatalytic
networks that could in theory generate several of the building blocks of life. Here, we explored the genomes of
all prokaryotes so-far annotated in a widely used biochemical database [51], in total 6683, in the same exercise.
Strikingly, we find that all metabolic networks can, under our set of assumptions (see Methods, Discussion
below and [32]), generate small-molecule autocatalytic networks (Fig. 1a). The universal redox currency NAD,
which is directly converted to ATP, the universal energetic currency, must be present in the medium (food set)
to generate an interesting level of complexity (Fig. 1b; Fig. 3). This result confirms that the emergence of ATP
and NAD is a vital step in prebiotic chemical evolution [52–54]. Their emergence allowed for chemical
coupling between distinct parts of the network and the feasibility of endergonic and unfavourable steps that
were nevertheless advantageous for the chemical phenotype. Even if most likely preceded by simpler
molecules that carried phosphorylating and reductive power, the pair was almost certainly essential at the
time of the emergence of functional polymers required for a primordial code. Adenine, present in both
currencies as well as in other universally-essential cofactors as FAD and CoA [35] is also one of the genetic
monomers.

It is likely that for the complexity of the networks we describe, their size, organization and persistence,
compartmentalization was required [27], and the compartments were themselves actively involved in the
chemical, energetic and structural maintenance of the network [22,23,55]. The extent to which these
compartments were complex, and their roles in the chemical network are still quite unknown and further

188 experimental work in this direction is desired. The maxRAFs here analysed can produce the universal
189 precursor of bacterial lipids, glycerol-3-phosphate, and several routes have been described for prebiotic lipid
190 production (reviewed in [15,49]), as well as micelle and vesicle replication in autocatalytic routes [56].
191
192 Some recent criticism to the use of the RAF framework in the study of early metabolism [57] requires
193 addressing. First, it is important to reinforce that the RAF framework does not force all of the network to be
194 catalysed. Both in previous work [31,32] and here, the introduction of a catalyst named 'spontaneous' allows
195 the model to include reactions that are not catalysed. The model assumes those reactions occur spontaneously
196 because 'spontaneous' is always added in all media simulated. Secondly, the catalyst 'peptide' was not
197 introduced to catalyse reactions not linked to cofactors, it was introduced as an annotation. The catalyst
198 'peptide' is, contrarily to the catalyst 'spontaneous', never added to the medium in simulations relevant to the
199 origin of metabolism [32]. Reactions catalysed by 'peptide' are not annotated to any small-molecule catalyst,
200 and therefore are not part of the small-molecule networks here or elsewhere described as relevant to the origin
201 of metabolism. Thirdly, the cyclic nature of cofactors is taken into account in the RAF model, which uses the
202 full stoichiometry of metabolic reactions where cofactors are modified. For example, in reactions that use ATP,
203 the reaction is modelled with ADP on the products side, and ATP is counted both as a reagent and as a
204 catalyst. Equally, reactions that use NADH are modelled with NAD on the other side, and 'NADs', the high
205 class that equates NAD(P)H, is counted as a catalyst. Those reactions associated to recycled cofactors may be
206 associated to other cofactors or metals in the database we used, the Universal Protein Resource
207 Knowledgebase [58], and those are taken into account as well. Fourthly, in this model, as in biochemistry,
208 every reaction is not controlled by a specific catalyst. In our simulations, among 5723 molecules participating
209 in 5994 annotated reactions, only 47 small-molecules or classes thereof act as catalysts. We use classes as
210 'flavins', 'nads', 'folates' and 'divalent cation' to equate structurally or chemically similar catalysts, which
211 generalizes catalysis even more than what is natural in the cell. Finally, because ATP is a monomer of RNA, in
212 agreement with others we are confident that the origin of ATP before the origin of RNA is quite parsimonious
213 – in fact, it seems to be a necessity, given the difficulty to polymerize NMPs directly to RNA. Because ATP
214 allows for much more chemically than polymerization, it is parsimonious to assume that other chemical
215 reactions occurred in the same chemical niche where ATP originated, before and concurrently with its
216 polymerization with other bases to form RNA strands.
217
218 There are, however, and despite confounding criticism addressed, significant and real weaknesses in the
219 model that we hope to be addressed in the future. The curation of reaction reversibility, and the assignment of
220 different catalysts to reactions running in reverse, are required improvements, which require the annotations
221 of thousands of reactions to be curated manually. One example is the synthesis of cysteine, which occurs in
222 our simulations in small RAFs from pyruvate, ammonia and hydrogen sulphide, in a highly-endergonic
223 reverse direction of a reaction from a common cysteine degradation pathway. Encouragingly, another
224 weakness of the model is also expected to be a strong mitigator of the latter. The oxygen-independent network
225 used here and before represents a myopic portion of the possible chemical space in organic chemistry, a space

226 which can in theory generate tens of thousands of reactions from just a handful of starting molecules [28] (and
227 where many millions more of molecules are known). This means that many chemical routes and catalysts are
228 not taken into account in this work, routes which most likely are able to overcome highly endergonic steps
229 taken in our simulations. The accurate computational prediction and assignment of small-molecule catalysis to
230 individual reactions is still infeasible, and the use of known biochemistry reveals much about the fundamental
231 principles and network structure while opening doors to future work. When more is revealed about general
232 rules of catalysis and complex chemical systems, it is most likely that different, larger and more versatile
233 autocatalytic networks will be found, and confirmed experimentally.
234
235 The fact that small-molecule autocatalytic chemical networks can be found universally in prokaryotic
236 metabolism is a strong indication that they were the initial step of molecular reproduction in the Universe. In
237 the clear wording of H. Morowitz , "*In any case the problem must begin at the beginning. In my view the*
238 *beginning is the network in small molecule space*"[59]. Recent work identified universal core autocatalytic
239 motifs and analysed their kinetic viability [30]. In these motifs every molecule is an autocatalyst, and each
240 motif contains a fork of one molecule to two, amplifying the number of autocatalysts. Among five such motifs,
241 each higher numbered is better at survival. It is now possible to examine the set of metabolic transformations
242 in any prokaryotic map to access the distribution of the five motifs, their location in metabolic maps and their
243 roles in prebiotic and early biotic evolution. Autocatalysis in a non-equilibrium system can increase the
244 concentrations of some metabolites. Do motifs increase concentrations of metabolites that would be low in the
245 absence of other catalysts? If early life required a stable source of building blocks as amino acids and
246 nucleotides, we expect to see such a correlation. Also, the distribution of core motifs in metabolic maps could
247 be compared to random or abiotic chemistry. The global network used here uses significant levels of external
248 catalysis (by metal ions or organic catalysts produced in yet unknown prebiotic routes), but as mentioned in
249 [30] the core autocatalytic motifs should not be altered. In accordance with the suggestion of an unappreciated
250 abundance of autocatalysis in chemistry based on core motifs [30], and with findings of high chemical
251 diversity in meteorites [60] and other astrological bodies [61], our results indicate that molecular reproduction
252 started over four billion years ago, much earlier than the last universal common ancestor [62] with small-
253 molecule autocatalytic networks in constrained yet favourable [63] geochemical settings on Earth, and perhaps
254 much earlier all over the Universe.
255

Additional Information

Data Accessibility

259 The list of prokaryotes used in this study is provided in ESM Supplementary Data 1. Metabolic annotations for
260 each genome are available in KEGG [51]. The full oxygen-independent metabolic network annotated with
261 small-molecule catalysis with some reaction directions curated is available in ESM Supplementary Data 2, and

262 the food sets are available in Supplementary Data 3. All maxRAFs obtained here are provided in ESM
263 (Supplementary Data 4 and 5). The RAF algorithm used was the same as published in [32].
264

265 **Authors' Contributions**
266 JCX conceived and designed the study, collected data, performed data analysis and visualization and wrote
267 the first draft of the manuscript. SK contributed in the study conception, data interpretation and analysis and
268 in the writing of the final manuscript. All authors read and approved the manuscript.
269

279
280

Methods

281

282

283 **Data Preparation**
284

285 Genome annotations to E.C. numbers were obtained from the KEGG database [51]. The list of prokaryotic
286 genomes used here, the corresponding number of E.C. annotations and number of reactions for each is
287 available in Supplementary Data 1. Whereas the full list of 6793 prokaryotes is provided, the results for 110
288 genomes with no E.C. annotations were not taken into account in results analyses. In the annotation of
289 metabolic reactions, the generalist categories "Metabolic pathways", "Biosynthesis of secondary metabolites",
290 "Biosynthesis of antibiotics" and "Microbial metabolism in diverse environments" were excluded.
291

292 The set of oxygen-independent catalysis-annotated reactions used here is the same as in previous work, where
293 small-molecule catalysis rules were obtained from Uniprot [58] and from some reactions themselves [32] with
294 small exceptions. The modifications done here include a curation of the directions for some
295 deoxyribonucleotides syntheses reactions (identifiers in the KEGG database are R02022, R02020, R02023,
296 R02017, R02018, R02019, R02024), which were previously taken in the wrong direction as stated in KEGG; the
297 curated network is available as Supplementary Data 2. Furthermore, thioredoxin (previously absent from all

Phil. Trans. R. Soc. A.

298 food sets used) was now added in the rich media to allow for dNTP production; the rich media used here is
299 therefore Food Set 3 in [32] with the single addition of thioredoxin, and the NAD media used here is Food Set
300 2 in the same study with the single addition of NAD (Supplementary Data 3). Both media tested contain the
301 'spontaneous' and 'pooling' catalysts (but not 'peptide') as before. When building individual metabolic
302 networks for running the maxRAF algorithm, the full set of reactions annotated to each genome was
303 considered, with the addition of 147 spontaneous reactions allowed to always occur.
304
305 **maxRAF calculation**
306 The algorithm for the calculation of maxRAFs used here was published in previous work [32]. All RAFs
307 obtained are available in Supplementary Data 4 and 5.
308
309

References

311 1. Kun Á, Radványi Á. 2018 The evolution of the genetic code: Impasses and challenges. *BioSystems*. **164**,
312   217–225. (doi:10.1016/j.biosystems.2017.10.006)
313 2. Preiner M *et al.* 2020 The future of origin of life research: Bridging decades-old divisions. *Life* **10**, 20.
314   (doi:10.3390/life10030020)
315 3. Marshall P. 2021 Biology transcends the limits of computation. *Prog Biophys Mol Biol*
316   (doi:10.1016/j.pbiomolbio.2021.04.006)
317 4. Carter CW. 2017 Coding of class I and II aminoacyl-tRNA synthetases. *Adv Exp Med Biol* **966**, 103–148.
318   (doi:10.1007/5584_2017_93)
319 5. Carter CW, Wills PR. 2021 The Roots of Genetic Coding in Aminoacyl-tRNA Synthetase Duality. *Annu*
320   *Rev Biochem*. **90**, 349–373. (doi:10.1146/annurev-biochem-071620-021218)
321 6. Bennett GM, Moran NA. 2013 Small, smaller, smallest: The origins and evolution of ancient dual
322   symbioses in a phloem-feeding insect. *Genome Biol Evol* **5**, 1675–1688. (doi:10.1093/gbe/evt118)
323 7. Mackie GA. 2013 RNase E: At the interface of bacterial RNA processing and decay. *Nat Rev Microbiol*
324   **11**, 45–57. (doi:10.1038/nrmicro2930)
325 8. Milo R. 2013 What is the total number of protein molecules per cell volume? A call to rethink some
326   published values. *BioEssays* **35**, 1050–1055. (doi:10.1002/bies.201300066)
327 9. Neidhardt FC, Edwin Umbarger. 1996 Chemical Composition of Escherichia coli. In *Escherichia coli*
328   *and Salmonella: Cellular and Molecular Biology*, p. Table 1. ASM Press.
329 10. Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD. 2009 Absolute metabolite
330   concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat Chem Biol* **5**, 593–
331   599. (doi:10.1038/nchembio.186)
332 11. Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF. 2016 The
333   physiology and habitat of the last universal common ancestor. *Nat Microbiol* **1**, 16116.
334   (doi:10.1038/nmicrobiol.2016.116)

335   12.   Xavier JC, Gerhards RE, Wimmer JLE, Brueckner J, Tria FDK, Martin WF. 2021 The metabolic network
336   of the last bacterial common ancestor. *Commun Biol* **4**, 413. (doi:10.1038/s42003-021-01918-4)

337   13.   Coleman GA, Davín AA, Mahendrarajah TA, Szánthó LL, Spang A, Hugenholtz P, Szöllsi GJ, Williams
338   TA. 2021 A rooted phylogeny resolves early bacterial evolution. *Science* **372**.
339   (doi:10.1126/science.abe0511)

340   14.   Williams TA, Szöllosi GJ, Spang A, Foster PG, Heaps SE, Boussau B, Ettema TJG, Martin Embley T.
341   2017 Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc Natl Acad*
342   *Sci U S A* **114**, E4602–E4611. (doi:10.1073/pnas.1618463114)

343   15.   Ruiz-Mirazo K, Briones C, de la Escosura A. 2014 Prebiotic Systems Chemistry: New Perspectives for
344   the Origins of Life. *Chem Rev* **114**, 285–366. (doi:10.1021/cr2004844)

345   16.   Duim H, Otto S. 2017 Towards open-ended evolution in self-replicating molecular systems. *Beilstein J*
346   *Org Chem.* **13**, 1189–1203. (doi:10.3762/bjoc.13.118)

347   17.   Ameta S, Matsubara YJ, Chakraborty N, Krishna S, Thutupalli S. 2021 Self-reproduction and darwinian
348   evolution in autocatalytic chemical reaction systems. *Life* **11**, 308. (doi:10.3390/life11040308)

349   18.   Gil R, Silva FJ, Peretó J, Moya A. 2004 Determination of the Core of a Minimal Bacterial Gene Set.
350   *Microbiol Mol Biol Rev* **68**, 518–537. (doi:10.1128/MMBR.68.3.518)

351   19.   Xavier JC, Patil KR, Rocha I. 2014 Systems biology perspectives on minimal and simpler cells.
352   *Microbiol Mol Biol Rev* **78**. (doi:10.1128/MMBR.00050-13)

353   20.   Choe D, Kim SC, Palsson BO, Cho BK. 2019 Construction of minimal genomes and synthetic cells. In
354   *Minimal Cells: Design, Construction, Biotechnological Applications* (eds AR Lara, G Gosset), pp. 45–67.
355   Springer International Publishing. (doi:10.1007/978-3-030-31897-0_2)

356   21.   Baross JA, Hoffman SE. 1985 Submarine hydrothermal vents and associated gradient environments as
357   sites for the origin and evolution of life. *Orig Life Evol Biosph* **15**, 327–345. (doi:10.1007/BF01808177)

358   22.   Martin W, Baross J, Kelley D, Russell MJ. 2008 Hydrothermal vents and the origin of life. *Nat Rev*
359   *Microbiol* **6**, 805–814. (doi:10.1038/nrmicro1991)

360   23.   Lane N. 2017 Proton gradients at the origin of life. *BioEssays* **39**, 1600217. (doi:10.1002/bies.201600217)

361   24.   Kauffman SA. 1971 Cellular homeostasis, epigenesis and replication in randomly aggregated
362   macromolecular systems. *J Cybern* **1**, 71–96. (doi:10.1080/01969727108545830)

363   25.   Vasas V, Fernando C, Santos M, Kauffman S, Szathmáry E. 2012 Evolution before genes. *Biol Direct* **7**,
364   1. (doi:10.1186/1745-6150-7-1)

365   26.   Markovitch O, Lancet D. 2012 Excess mutual catalysis is required for effective evolvability. *Artif Life*
366   **18**, 243–266. (doi:10.1162/artl_a_00064)

367   27.   Hordijk W, Naylor J, Krasnogor N, Fellermann H. 2018 Population dynamics of autocatalytic sets in a
368   compartmentalized spatial world. *Life* **8**, 33. (doi:10.3390/life8030033)

369   28.   Wolos A, Roszak R, Zadlo-Dobrowolska A, Beker W, Mikulak-Klucznik B, Spólnik G, Dygas M,
370   Szymkuc S, Grzybowski BA. 2020 Synthetic connectivity, emergence, and self-regeneration in the
371   network of prebiotic chemistry. *Science* **369**. (doi:10.1126/science.aaw1955)

372   29.   Ameta S, Matsubara YJ, Chakraborty N, Krishna S, Thutupalli S. 2021 Self-reproduction and darwinian
              Phil. Trans. R. Soc. A.

373 evolution in autocatalytic chemical reaction systems. *Life* **11**, 308. (doi:10.3390/life11040308)

374 30. Blokhuis A, Lacoste D, Nghe P. 2020 Universal motifs and the diversity of autocatalytic systems. *Proc*
375 *Natl Acad Sci U S A* **117**, 25230–25236. (doi:10.1073/pnas.2013527117)

376 31. Sousa FL, Hordijk W, Steel M, Martin WF. 2015 Autocatalytic sets in E. coli metabolism. *J Syst Chem* **6**,
377 4. (doi:10.1186/s13322-015-0009-7)

378 32. Xavier JC, Hordijk W, Kauffman S, Steel M, Martin WF. 2020 Autocatalytic chemical networks at the
379 origin of metabolism. *Proc R Soc B Biol Sci* **287**, 20192377. (doi:10.1098/rspb.2019.2377)

380 33. Hordijk W, Steel M. 2004 Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *J*
381 *Theor Biol* **227**, 451–461. (doi:10.1016/j.jtbi.2003.11.020)

382 34. Steel M, Xavier JC, Huson DH. 2020 The structure of autocatalytic networks, with application to early
383 biochemistry. *J R Soc Interface* **17**, 20200488. (doi:10.1098/rsif.2020.0488)

384 35. Xavier JC, Patil KR, Rocha I. 2017 Integration of biomass formulations of genome-scale metabolic
385 models with experimental data reveals universally essential cofactors in prokaryotes. *Metab Eng* **39**,
386 200–208. (doi:10.1016/j.ymben.2016.12.002)

387 36. Lagesen K, Ussery DW, Wassenaar TM. 2010 Genome update: the 1000th genome--a cautionary tale.
388 *Microbiology* **156**, 603–8. (doi:10.1099/mic.0.038257-0)

389 37. Sojo V, Pomiankowski A, Lane N. 2014 A Bioenergetic Basis for Membrane Divergence in Archaea and
390 Bacteria. *PLoS Biol* **12**, e1001926. (doi:10.1371/journal.pbio.1001926)

391 38. Sousa FL, Preiner M, Martin WF. 2018 Native metals, electron bifurcation, and CO2 reduction in early
392 biochemical evolution. *Curr Opin Microbiol* **43**, 77–83. (doi:10.1016/j.mib.2017.12.010)

393 39. Preiner M *et al.* 2020 A hydrogen-dependent geochemical analogue of primordial carbon and energy
394 metabolism. *Nat Ecol Evol* **4**, 534–542. (doi:10.1038/s41559-020-1125-6)

395 40. Muchowska KB, Varma SJ, Moran J. 2019 Synthesis and breakdown of universal metabolic precursors
396 promoted by iron. *Nature* **569**, 104–107. (doi:10.1038/s41586-019-1151-1)

397 41. Argueta EA, Amoh AN, Kafle P, Schneider TL. 2015 Unusual non-enzymatic flavin catalysis enhances
398 understanding of flavoenzymes. *FEBS Lett* **589**, 880–884. (doi:10.1016/j.febslet.2015.02.034)

399 42. Metzler DE, Snell EE. 1952 Deamination of serine. I. Catalytic deamination of serine and cysteine by
400 pyridoxal and metal salts. *J Biol Chem* **198**, 353–61.

401 43. Zabinski RF, Toney MD. 2001 Metal Ion Inhibition of Nonenzymatic Pyridoxal Phosphate Catalyzed
402 Decarboxylation and Transamination. *J Am Chem Soc* **123**, 193–198. (doi:10.1021/ja0026354)

403 44. Barrows LR, Magee PN. 1982 Nonenzymatic methylation of DNA by S-adenosylmethionine in vitro.
404 *Carcinogenesis* **3**, 349–351. (doi:10.1093/carcin/3.3.349)

405 45. Bazhenova TA, Bazhenova MA, Petrova GN, Mironova SA, Strelets V V. 2000 Catalytic behavior of the
406 nitrogenase iron-molybdenum cofactor extracted from the enzyme in the reduction of C2H2 under
407 nonenzymatic conditions. *Kinet Catal* **41**, 499–510. (doi:10.1007/BF02756066)

408 46. Kirschning A. 2021 Coenzymes and Their Role in the Evolution of Life. *Angew Chemie - Int Ed.* **60**,
409 6242–6269. (doi:10.1002/anie.201914786)

410 47. Lin H, Jiménez EI, Arriola JT, Müller UF, Krishnamurthy R. 2022 Concurrent Prebiotic Formation of

411   Nucleoside-Amidophosphates and Nucleoside-Triphosphates Potentiates Transition from Abiotic to

412   Biotic Polymerization. *Angew Chemie* **134**, e202113625. (doi:10.1002/ange.202113625)

413 48. Kim HJ, Benner SA. 2021 Abiotic Synthesis of Nucleoside 5′-Triphosphates with Nickel Borate and

414   Cyclic Trimetaphosphate (CTMP). *Astrobiology* **21**, 298–306. (doi:10.1089/ast.2020.2264)

415 49. Kitadai N, Maruyama S. 2018 Origins of building blocks of life: A review. *Geosci Front* **9**, 1117–1153.

416   (doi:10.1016/j.gsf.2017.07.007)

417 50. Kauffman S. 2019 *A world beyond physics: the emergence and evolution of life*. Oxford University

418   Press.

419 51. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017 KEGG: New perspectives on

420   genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353–D361.

421   (doi:10.1093/nar/gkw1092)

422 52. White HB. 1976 Coenzymes as fossils of an earlier metabolic state. *J Mol Evol* **7**, 101–104.

423   (doi:10.1007/BF01732468)

424 53. Dyson FJ. 1982 A model for the origin of life. *J Mol Evol* **18**, 344–350. (doi:10.1007/BF01733901)

425 54. Fontecilla-Camps JC. 2019 Geochemical Continuity and Catalyst/Cofactor Replacement in the

426   Emergence and Evolution of Life. *Angew Chemie - Int Ed.* **58**, 42–48. (doi:10.1002/anie.201808438)

427 55. Matsuo M, Kurihara K. 2021 Proliferating coacervate droplets as the missing link between chemistry

428   and biology in the origins of life. *Nat Commun 2021 121* **12**, 1–13. (doi:10.1038/s41467-021-25530-6)

429 56. Bachmann PA, Luisi PL, Lang J. 1992 Autocatalytic self-replicating micelles as models for prebiotic

430   structures. *Nature* **357**, 57–59. (doi:10.1038/357057a0)

431 57. Higgs PG. 2021 When Is a Reaction Network a Metabolism? Criteria for Simple Metabolisms That

432   Support Growth and Division of Protocells. *Life* **11**, 966. (doi:10.3390/life11090966)

433 58. UniProt Consortium T. 2018 UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **46**, 2699.

434   (doi:10.1093/nar/gky092)

435 59. Morowitz HJ. 1977 Perspectives on Thermodynamics and the Origin of Life. In *Advances in biological

436   and medical physics*, pp. 151–163. Elsevier. (doi:10.1016/B978-0-12-005216-5.50013-X)

437 60. Schmitt-Kopplin P, Gabelica Z, Gougeon RD, Fekete A, Kanawati B, Harir M, Gebefuegi I, Eckel G,

438   Hertkorn N. 2010 High molecular diversity of extraterrestrial organic matter in Murchison meteorite

439   revealed 40 years after its fall. *Proc Natl Acad Sci U S A* **107**, 2763–2768. (doi:10.1073/pnas.0912157107)

440 61. Postberg F *et al.* 2018 Macromolecular organic compounds from the depths of Enceladus. *Nature* **558**,

441   564–568. (doi:10.1038/s41586-018-0246-4)

442 62. Betts HC, Puttick MN, Clark JW, Williams TA, Donoghue PCJJ, Pisani D. 2018 Integrated genomic and

443   fossil evidence illuminates life's early evolution and eukaryote origin. *Nat Ecol Evol* **2**, 1556–1562.

444   (doi:10.1038/s41559-018-0644-x)

445 63. Wimmer JLE, Xavier JC, Vieira A d. N, Pereira DPH, Leidner J, Sousa FL, Kleinermanns K, Preiner M,

446   Martin WF. 2021 Energy at Origins: Favorable Thermodynamics of Biosynthetic Reactions in the Last

447   Universal Common Ancestor (LUCA). *Front Microbiol* **12**, 3903. (doi:10.3389/fmicb.2021.793664)

448