



## Article

# The Autotrophic Core: An Ancient Network of 404 Reactions Converts H<sub>2</sub>, CO<sub>2</sub>, and NH<sub>3</sub> into Amino Acids, Bases, and Cofactors

Jessica L. E. Wimmer<sup>1,\*</sup>, Andrey do Nascimento Vieira<sup>1</sup>, Joana C. Xavier<sup>1</sup>, Karl Kleinermanns<sup>2</sup>, William F. Martin<sup>1</sup> and Martina Preiner<sup>1</sup>

<sup>1</sup> Department of Biology, Institute for Molecular Evolution, Heinrich-Heine-University of Düsseldorf, 40225 Düsseldorf, Germany; nascima@uni-duesseldorf.de (A.d.N.V.); xavier@hhu.de (J.C.X.); bill@hhu.de (W.F.M.); preiner@hhu.de (M.P.)

<sup>2</sup> Department of Chemistry, Institute for Physical Chemistry, Heinrich-Heine-University of Düsseldorf, 40225 Düsseldorf, Germany; kkleinermanns@yahoo.de

\* Correspondence: jessica.wimmer@hhu.de

**Abstract:** The metabolism of cells contains evidence reflecting the process by which they arose. Here, we have identified the ancient core of autotrophic metabolism encompassing 404 reactions that comprise the reaction network from H<sub>2</sub>, CO<sub>2</sub>, and ammonia (NH<sub>3</sub>) to amino acids, nucleic acid monomers, and the 19 cofactors required for their synthesis. Water is the most common reactant in the autotrophic core, indicating that the core arose in an aqueous environment. Seventy-seven core reactions involve the hydrolysis of high-energy phosphate bonds, furthermore suggesting the presence of a non-enzymatic and highly exergonic chemical reaction capable of continuously synthesizing activated phosphate bonds. CO<sub>2</sub> is the most common carbon-containing compound in the core. An abundance of NADH and NADPH-dependent redox reactions in the autotrophic core, the central role of CO<sub>2</sub>, and the circumstance that the core's main products are far more reduced than CO<sub>2</sub> indicate that the core arose in a highly reducing environment. The chemical reactions of the autotrophic core suggest that it arose from H<sub>2</sub>, inorganic carbon, and NH<sub>3</sub> in an aqueous environment marked by highly reducing and continuously far from equilibrium conditions. Such conditions are very similar to those found in serpentinizing hydrothermal systems.

**Keywords:** chemolithoautotrophy; early metabolism; serpentinizing systems; hydrothermal vents; origins of life

**Citation:** Wimmer, J.L.E.; Vieira, A.; Xavier, J.C.; Kleinermanns, K.; Martin, W.F.; Preiner, M. The Autotrophic Core: An Ancient Network of 404 Reactions Converts H<sub>2</sub>, CO<sub>2</sub> and NH<sub>3</sub> Into Amino Acids, Bases and Cofactors. *Microorganisms* **2021**, *9*, 458. <https://doi.org/10.3390/microorganisms9020458>

Academic Editor: Françoise Bringel

Received: 28 January 2021

Accepted: 19 February 2021

Published: 23 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Biologists have traditionally linked the topic of C1 metabolism to thoughts about life's origins. Haeckel (1902) posited that the first cells probably lived from CO<sub>2</sub> [1], perhaps in a manner similar to organisms discovered by Winogradsky (1888), growing from CO<sub>2</sub> with the help of electrons from inorganic donors [2]. The chemolithoautotrophic lifestyle—converting inorganic carbon into cell mass with inorganic electron donors using chemical energy instead of light—is common among modern microbes that inhabit environments similar to those on the early Earth [3]. Although microbiologists have traditionally favored the view that the first cells were anaerobic autotrophs [4–7], the electric spark experiments of Miller shifted the focus in origins literature from microbiology to nucleic acid chemistry [8]. The facile synthesis of nucleobases from cyanide condensations [9], Spiegelman's in vitro RNA replication experiments using Q $\beta$  replicase [10], and the demonstration that RNA has catalytic activity [11] led to the concept of an RNA world [12] in which RNA molecules became synthesized by abiotic chemical means and then competed with one another for resources (activated ribonucle-

oside triphosphate monomers) via replication [13]. This line of thinking diverted attention away from the more challenging problem concerning the origin of living cells and toward the more tractable problem concerning the origin of nucleic acid monomers [14,15]. However, for critics, the allure of the RNA world concept has caveats, as RNA synthesis, regardless how good, does not alone solve the problem of how living cells arose [16]; *Escherichia coli* is clearly alive; RNA is clearly not.

### 1.1. Metabolism vs. Genetics?

As an alternative to the concept of an RNA world, Wächtershäuser's theory of surface metabolism rekindled the idea of a chemolithoautotrophic start of life and brought energy into the origins debate, positing that the exergonic conversion of iron-sulfur (FeS) minerals to pyrite provided the thermodynamic drive to fuel the origin of biochemical pathways and the first autotrophic cells [17]. Wächtershäuser's theory ignited a "genetics first vs. metabolism first" debate [18–21] that Lipmann had presciently perceived decades in advance [22]. While the theory of surface metabolism interfaced well with catalytic mechanisms in autotrophic cells, it did not interface well with energy conservation in the currency of high-energy phosphate bonds [23] nor did it offer inroads to accounting for the fundamental property of life that the living cell maintains itself in a thermodynamic state that is far from equilibrium. The discovery of deep-sea hydrothermal vents [24] and alkaline hydrothermal vents of serpentinizing systems [25] impacted the origins issue in that they harbor geochemically continuous far from equilibrium conditions that help to define the living state of cells [5].

The idea that life started from CO<sub>2</sub> is appealing, but it only solves half the problem because both for life and for organic synthesis, CO<sub>2</sub> requires a reductant. This is why serpentinizing hydrothermal systems are so interesting in the origins context. Serpentinization synthesizes H<sub>2</sub>, the main energy and electron source of chemolithoautotrophs, from protons and electrons within the Earth's crust through the reduction of H<sub>2</sub>O by Fe(II) minerals. The amount of H<sub>2</sub> generated by serpentinization is substantial, on the order of 16 mmol/kg in some modern systems [26], which is orders of magnitude more H<sub>2</sub> than modern chemolithoautotrophs require for growth [27]. The synthesis of H<sub>2</sub> during serpentinization is a continuous process that has been going on since there was water on Earth [28]. The further researchers explored the properties of serpentinizing systems, the more similarities they revealed to the life process [29], with compartmentation, energy harnessing, catalysis, and the chemical reactions of C1 compounds at vents converging on processes that comprise the core of carbon and energy metabolism in primitive autotrophic cells [30].

From a biological perspective, the genetics first vs. metabolism first debate misses the point because neither by itself is sufficient for life. Countering the genetics first view, cells are made of far more than RNA. Cells consist by dry weight of about 50% protein and 20% RNA, with DNA, lipids, cell wall, reserves, and metabolites making up the rest [31]. Most of the RNA resides in the ribosome, which synthesizes proteins in a process that consumes about 75% of the cell's ATP investment in biosynthetic processes [32], whereby a large portion of the ATP that a cell synthesizes is not used for the synthesis of cell mass—it goes to what is called ATP spilling and maintenance energy [33,34].

Countering the metabolism first view, a handful of small molecules reacting with each other do not qualify as metabolism. A cell is a very complicated chemical system composed of more than 1000 individual partial reactions that harness energy and synthesize building blocks as well as polymers. A decisive property of metabolism is redox balance: the number of electrons that enter the cell in substrates has to be equal to the number of electrons that leave the cell in waste products plus those that remain sequestered in compounds of cell mass; otherwise, metabolism and life come to a halt [35]. Although most reactions in cells are catalyzed by enzymes, enzymes do not perform feats of magic; they just accelerate reactions that tend to occur anyway. Many core metabolic reactions of cells readily take place without enzymes [36–38]. The sum of the chemical re-

actions in the cell (metabolism) runs both the synthesis and the operation of the cellular machinery that produces progeny, harboring a new copy of instructions in DNA, hence heredity, which over generations forms the process called genetics.

Yet the main thing that cells do is neither genetics nor metabolism but energy harnessing because without energy, neither metabolism nor genetics can take place. Metabolism and genetics are merely manifestations of the actions of sustained sets of exergonic chemical reactions over generations. There is a third option in the genetics first vs. metabolism first debate, namely “energy first”, because it is hands-down obvious that energy has to come first [39], for without favorable energetics and energy release, neither genetics, metabolism, nor anything at all will take place, so says the 2nd law of thermodynamics. Cells themselves underscore that view, because the amount of ATP that a cell synthesizes always exceeds the amount of energy required for the synthesis of new cells during growth, often by about a factor of 3 [40] (the converse would violate the 2nd law), underscoring the point that life is an energy-releasing process. For a cell, staying alive means staying far from equilibrium, which it achieves by merely running the exergonic reactions that synthesize ATP: maintenance energy or ATP spilling [33,34]. In low-energy environments, where survival becomes more important than growth [41], maintenance energy becomes the main process of life.

### *1.2. Autotrophic Origins and Energy First Link C1 Metabolism to Vents*

Is there an origins option that starts with energy first? Yes, and it is seated firmly in C1 metabolism and autotrophic origins. In 2021, serpentinizing systems have gone a long way to closing the gap between CO<sub>2</sub> and cells. Convergent lines of evidence indicate that reactions of C1 compounds were not only the source of carbon for the first cells but also the source of energy at the origin of the first metabolic reactions. This is because in the reaction of H<sub>2</sub> with CO<sub>2</sub>, the equilibrium lies on the side of the simple reduced carbon compounds that comprise the backbone of carbon metabolism in organisms that use the acetyl Coenzyme A (CoA) pathway of CO<sub>2</sub> fixation—formate, acetate, and pyruvate. The synthesis of these acids from H<sub>2</sub> and CO<sub>2</sub> is exergonic under standard conditions [39], in cells that use the acetyl CoA pathway [6] and under conditions of simulated hydrothermal vents [30]. Hydrothermal vents are generally of interest in modern theories for origins [3] because they present continuously far from equilibrium conditions, with geochemically catalyzed redox reactions and gradients that could be tapped by the first cells for energy harnessing [29].

In hydrothermal systems, both modern and on the early Earth, the key to redox reactions, catalyst synthesis, and the formation of ion gradients, is molecular hydrogen, H<sub>2</sub>, which is generated by the spontaneous geochemical process of serpentinization [28,42,43]. During serpentinization, mineral catalysts awaruite (Ni<sub>3</sub>Fe) and magnetite (Fe<sub>3</sub>O<sub>4</sub>) are formed in situ in serpentinizing hydrothermal vents [44]. These minerals catalyze the synthesis of formate, acetate, and pyruvate as well as methane [30] in the laboratory from H<sub>2</sub> and CO<sub>2</sub> in the presence of only water and the mineral catalyst overnight at 100 °C and only 24 bar. It is likely, but not directly demonstrated, that hydrothermally formed awaruite and magnetite catalyze the synthesis of formate and methane found in the effluent of modern serpentinizing systems [45–48]. Serpentinization also renders the effluent of hydrothermal systems alkaline [48], generating the ion gradients that form at hydrothermal vents.

The synthesis of simple organics from C1 precursors in hydrothermal systems is the only known geochemical process that follows the same chemical route as a modern core pathway of carbon and energy metabolism [30,49]. In addition, modern organisms that use the acetyl CoA pathway for carbon and energy metabolism, acetogens and methanogens, exhibit a physiology that, among known life forms, is most similar to that inferred from genomic reconstructions for the last universal common ancestor of all cells, LUCA [50]. This implicates acetogens and methanogens that lack cytochromes and quinones as very primitive microbial lineages, in line with early predictions from physiolo-

gy [4] and with predictions based on similarities between geochemical and biochemical reactions [36]. It is also consistent with the identification of overlapping autocatalytic networks in the metabolism of acetogens and methanogens that implicate a role for small molecule reaction systems prior to the advent of both protein and RNA [51].

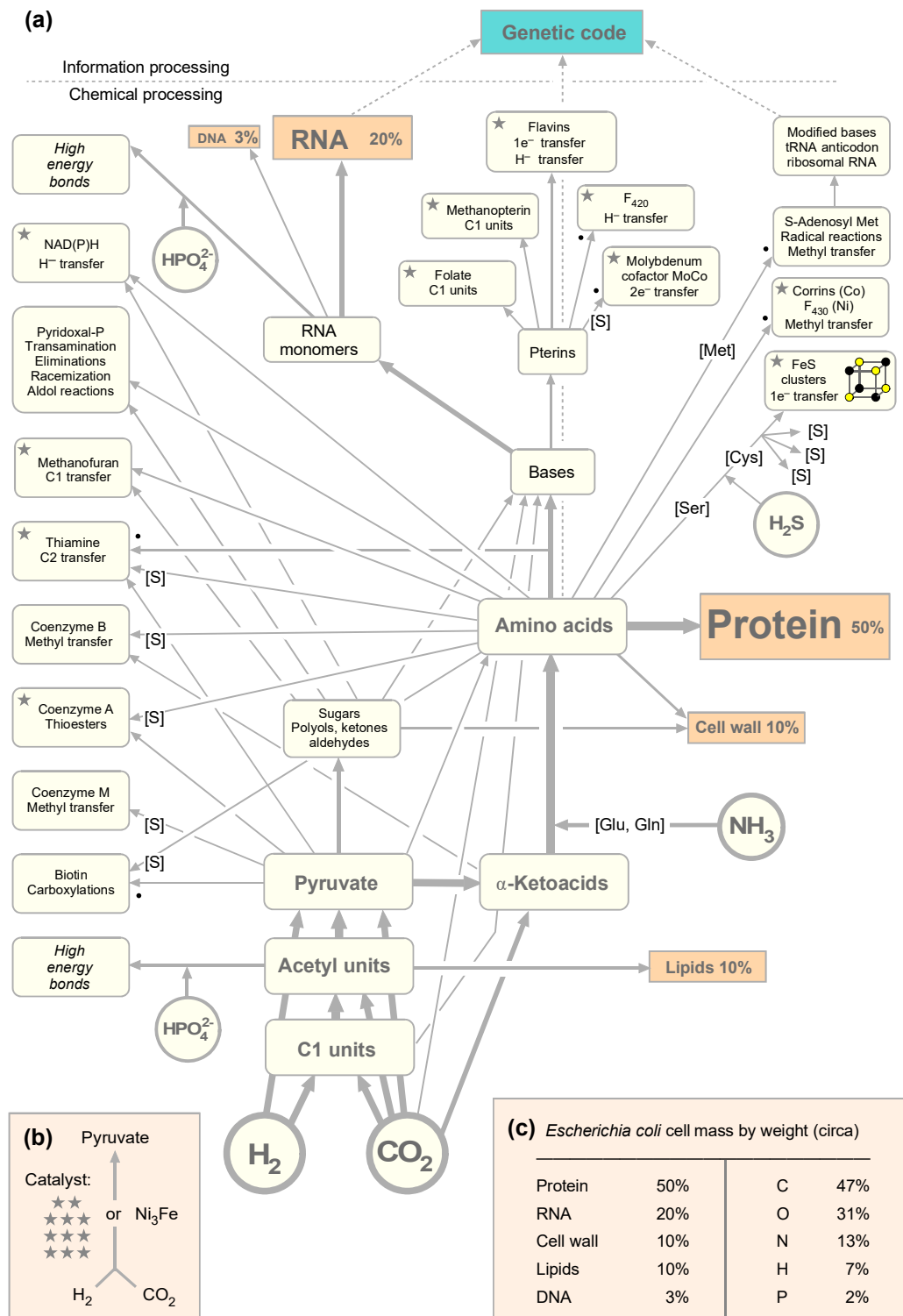
## 2. Methods

### 2.1. Reaction Data Collection

Metabolic reactions were gathered and curated from the Kyoto Encyclopedia of Genes and Genomes (KEGG) reaction database [52] (version December 2020) manually. Synthesis pathways for 46 target compounds (Table S1) were obtained and curated by hand with the help of KEGG pathways [53] and KEGG modules. The 46 target compounds comprise 20 amino acids, four ribonucleoside triphosphates, four deoxyribonucleoside triphosphates, and 18 cofactors shown in Figure 1. KEGG lacked biosynthetic pathway information on iron–sulfur clusters, so these were not included. Although depicted in Figure 1, polymers and the genetic code are also not part of the target set. The reductive acetyl CoA pathway as well as the reverse tricarboxylic acid cycle (rTCA) cycle were added to the reaction set, covering the basal CO<sub>2</sub> fixation along with the gluconeogenesis and pentose phosphate cycle, allowing for the synthesis of key intermediates needed to produce amino acids, nucleic acids, and cofactors from  $\alpha$ -ketoacids, sugars, and aldehydes. Nitrogen fixation pathways were not included, since NH<sub>3</sub> gets incorporated via amino acid synthesis. If a pathway was unavailable in KEGG, it was manually reconstructed based on KEGG pathway maps. The collection of reactions unfolds in a short example: For methanofuran biosynthesis, KEGG module M00935 was used to add reactions R10935, R11038, R11039, R00736, R10902, and R11040. The very last step of producing methanofuran is missing in the module. This reaction from APMF-Glu is depicted in pathway map00680; thus, it was added manually.

In all pathways collected, oxygen-dependent reactions were either replaced with an anaerobic alternative if possible or omitted if not. This was the case for the synthesis of dimethylbenzimidazole, which is a precursor for cobamide. Although an anaerobic synthesis pathway for this precursor is known, starting from 5-aminoimidazole ribotide (short AIR) [54], several other intermediaries are not implemented in KEGG yet. Neither was there an anaerobic alternative for the production of 2-phospholactate as a precursor in the F<sub>420</sub> synthesis pathway available. For both precursors, dimethylbenzimidazole and 2-phospholactate, as well as reduced ferredoxin (involved in the reductive acetyl CoA pathway) and reduced flavodoxin (in the rTCA cycle), we assume them to be producible in an unknown way in early metabolism. Assuming the reactions in question arose before the genetic code, the according proteins were presumably replaced by an alternative at that early period. Three reactions were constructed manually, because they appear as a dashed line in KEGG pathways with no corresponding reaction identification number. The reactions named *RMAN1-3* are presumed to be incomplete, since only the key compounds were listed. Two reactions are affected within tetrahydromethanopterin synthesis and the very last step was within methanofuran synthesis. Involved chemical elements such as molybdenum, sulfur donors, cobalt, and nickel were assumed to be present in the environment. During curation of the final reaction set, redundant reactions occurring in multiple syntheses were reduced to a single occurrence, such as the reaction chorismate  $\rightleftharpoons$  prephenate that occurs in both tyrosine and phenylalanine syntheses.

For the detection of autocatalytic cycles within cofactor biosynthetic pathways, catalysis rules (indicating which cofactors are used as catalysts in each reaction) were gathered from [51] (Supplemental Dataset S1A) [51]. Autocatalysis is assumed if a target is needed as a catalyst within its own biosynthetic pathway.



**Figure 1.** (a) A general map of core metabolism. The arrows in the map do not cover every atom in every cofactor, amino acid, or base, showing main mass contributions instead. A dot indicates that radical S-adenosyl methionine (SAM) enzymes are involved in the biosynthetic pathway leading to the product. [S] indicates that sulfur is incorporated in the biosynthetic pathway. (b) Cofactors indicated by a star are required in the pathway from  $\text{H}_2$  and  $\text{CO}_2$  to pyruvate in either acetogens or methanogens or both. (c) The composition of cells in terms of its main components and elemental contributions to dry weight (from [31]).

## 2.2. Visualization of the Autotrophic Core Network

An undirected metabolic network showing the autotrophic core was generated in simple interaction format (sif) using a custom Python script. The resulting network consists of the given 404 metabolic reactions and 380 involved compounds. The bipartite network was visualized using CytoScape [55] v. 3.8.0. One partition class corresponds to reaction nodes (diamonds), the other one corresponds to compound nodes (circle-shaped). The latter were sized according to their node degree. Target compounds were colored in blue, whereas reaction nodes are depicted smaller and in gray.

## 2.3. Different Core Reaction Sets Based on Distinct Identification Approaches

Two additional reaction datasets were used to determine their intersection with the 404 reactions of the autotrophic core. The LUCA set, containing 355 genes, was identified via the phylogenetic approach [50], translating to 163 metabolic reactions and the ancient 'reflexively autocatalytic food-generated' (RAF) set with 172 reactions (from [51] Figure 4). The intersection between the autotrophic core, LUCA, and the ancient RAF was determined by examining which reactions overlap in the respective analyzed datasets. In addition, the overlap of reactions between all three datasets was determined. The intersection for each comparison (Figure S1) is available in Supplemental Table S3A and the initial reaction lists are in Table S3B.

## 2.4. Statistical Analysis

A contingency table for each highly connected compound (Table 1) was built, comparing the compound frequency in the autotrophic core with the frequency in the global prokaryote anaerobic network consisting of 5994 reactions (from [51] S1A). A significant enrichment of compound frequency in the autotrophic core compared to the global prokaryotic set was observed for *p*-values smaller than 0.05. One-tailed Fisher tests were performed using the package `scipy.stats` in Python 3.6 (Table S4).

**Table 1.** Highly connected nodes.

Compound	Frequency
H <sub>2</sub> O	125
ATP	77
H <sup>+</sup>	76
P <sub>i</sub>	66
ADP	55
CO <sub>2</sub>	49
Glutamate	44
PP <sub>i</sub>	37
NAD <sup>+</sup>	37
NADP <sup>+</sup>	35
NADPH	34
NADH	33
2-Oxoglutarate	24
Pyruvate	22
NH <sub>3</sub>	21

## 3. Results

### 3.1. The Autotrophic Core of Biosynthesis Requires 19 Cofactors

For the purpose of this paper, let us assume for the sake of argument that life really did start from exergonic reactions of H<sub>2</sub> and CO<sub>2</sub> along the acetyl CoA pathway. Why do we assume the acetyl CoA pathway as the starting point of CO<sub>2</sub> fixation? Among the six known pathways of CO<sub>2</sub> fixation [6,56,57], it is the only one that occurs in both bacteria and archaea, the only one that traces in part to LUCA [50], and it is the only one that has

been shown in the laboratory to produce acetate and pyruvate from  $H_2$  and  $CO_2$  without enzymes, using only hydrothermal minerals as catalysts [30]. In that sense, it is the obvious choice as the starting point for metabolic evolution investigations based upon current laboratory evidence. The horseshoe (incomplete) rTCA cycle follows in Figure 1 because it is the pathway that autotrophs using the acetyl CoA pathway employ to generate C4 and C5 precursors for amino acid and other syntheses [6,36,58]. Although the incomplete horseshoe the rTCA cycle occurs in bacteria and archaea, it is fed by the acetyl CoA pathway, which is the only pathway of  $CO_2$  fixation that is known to occur in bacteria and archaea. The other five are known to operate in only one domain [6,56]. The rTCA cycle is also an ancient pathway [59,60], and most of its reactions also operate in the laboratory in the absence of enzymes provided that pyruvate and glyoxylate are supplied as starting material, but the non-enzymatic reaction sequence operates in the oxidative direction, that is, in the absence of  $H_2$  and  $CO_2$  [37]. The acetyl CoA pathway, the rTCA cycle, and the dicarboxylate/4-hydroxybutyrate cycle, which is a derivative of the rTCA cycle and occurs only in archaea, employ  $O_2$  sensitive enzymes, an ancient trait [6]. The other three  $CO_2$  fixation pathways have no  $O_2$ -sensitive enzymes and are typically found in aerobes, occur in only one domain each, and they arose more recently in evolution, using enzymes co-opted from preexisting pathways [6].

We also assume that the first living things on the path to cells we recognize today required the universal amino acids and bases of life, the modern synthesis of which requires in turn cofactors as catalysts. We asked: How big, exactly, is the set of reactions required for the synthesis of the building blocks of cells and the cofactors needed to make them? This gives us an impression of how challenging it would be to generate the main compounds of life at origins, with or without enzymes. We started by sketching out Figure 1a, in which the main pathways of biosynthesis in anaerobes and the amounts of main biosynthetic end products are summarized. Cofactors are usually not present in amounts that would contribute appreciably to cell mass, but they are required as catalysts. Along the acetyl CoA pathway, there are differences in the methanogenic and acetogenic versions [61].

If we look at the cofactors required to get from  $H_2$  and  $CO_2$  to pyruvate in acetogens and methanogens [6,62], we find that methanofuran, NAD(P)H, corrins, coenzyme A, thiamine, flavins,  $F_{420}$ , three pterins—folate, methanopterin, and the molybdenum cofactor MoCo—as well as FeS clusters, a prosthetic group of proteins that we count as a cofactor here, are required. That is a substantial cofactor requirement, not to mention the enzymes that hold those cofactors in place for function in the pathway. The mere requirement for those 11 complicated organic cofactors would appear to make the proposition that C1 metabolism from  $CO_2$  to pyruvate could be the first pathway [36] seem downright absurd, were it not for the recent observation that a bit of metal, awaruite ( $Ni_3Fe$ ), or a piece of iron oxide magnetite ( $Fe_3O_4$ ), can also catalyze the synthesis of pyruvate from  $H_2$  and  $CO_2$  [30] overnight at 100 °C in water. As a set of chemical reactions, the acetyl CoA pathway is older than the genes that encode its enzymes [58], and it is also older than the cofactors required by those enzymes.

By the foregoing count, it takes 11 cofactors to synthesize pyruvate in the modern pathway, whereby we have not counted the two steps requiring pyruvoyl enzymes at decarboxylation steps in the CoA (pantothenate) synthesis pathway; the pyruvoyl cofactor is synthesized from serine residues in the polypeptide chain of the enzyme [63]. Very surprisingly, only three additional cofactors (biotin, pyridoxal phosphate, and SAM) are required for the synthesis of the 11 other cofactors plus the main nucleosides of nucleic acids and the 20 amino acids, whereby only two more (coenzyme M and coenzyme B) are required specifically in the methanogenic pathway of energy conservation. That makes a total of 19 cofactors (counting NAD and NADP separately as well as the flavins flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) along with two corrins  $F_{420}$  and cobamide) to support their own synthesis plus the synthesis of four ribonucleoside triphosphates, four deoxyribonucleoside triphosphates, and 20 amino ac-

ids. The genetic code and polymers are not included in the autotrophic core. In total, that makes a list of 47 target compounds (19 cofactors, 8 nucleotides, and 20 amino acids; Table S1) that would be required to synthesize the substance of cells, as summarized in Figure 1.

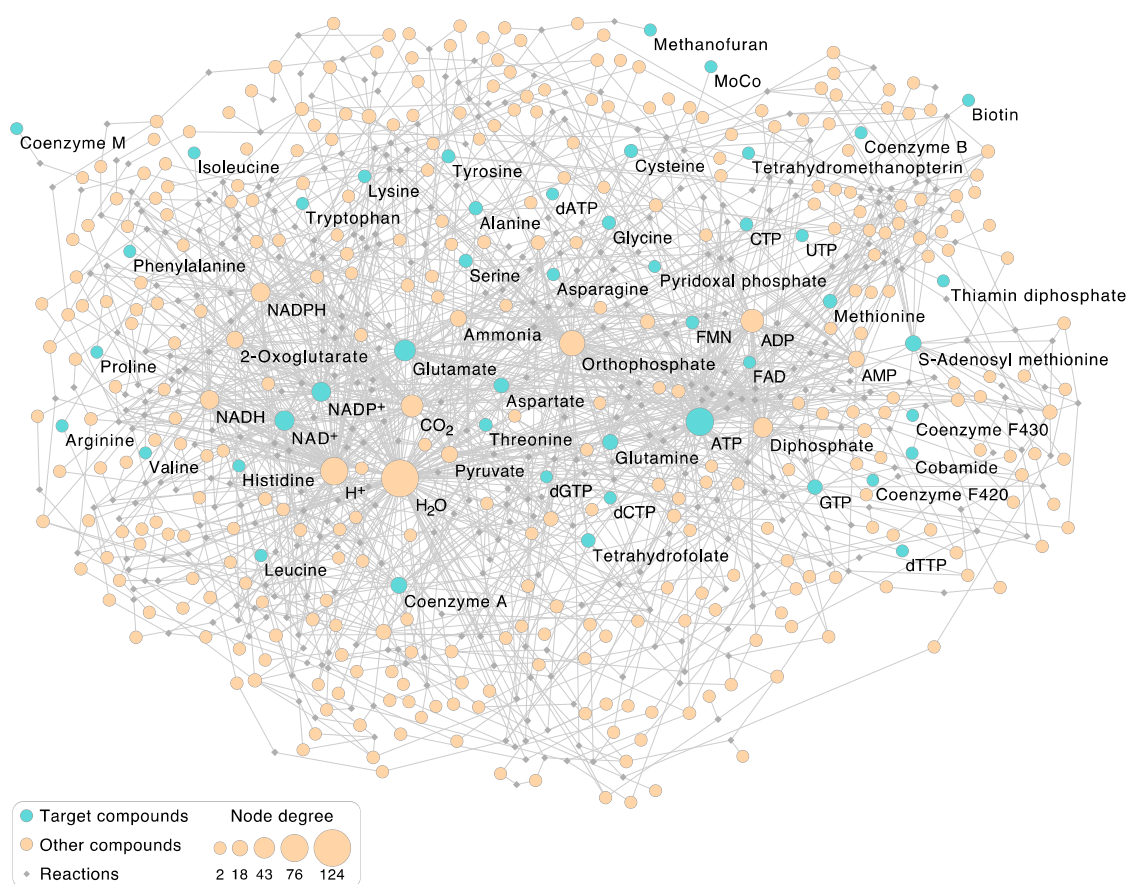
The starting point of Figure 1a is H<sub>2</sub> and CO<sub>2</sub>. Critics of autotrophic origins will be quick to point out that cyanide chemistry can readily give rise to amino acids and bases under laboratory conditions [14,64], such that we need not worry about Figure 1. However, in reply, we would be equally quick to point out that there are 415 distinct reactions in microbial metabolism involving CO<sub>2</sub> as a substrate in either the forward or reverse direction [65], but there are no reactions known to us in which cyanide serves as a main source of carbon in core anabolic metabolism. Some bacteria can convert CN<sup>-</sup> to CO<sub>2</sub> and NH<sub>3</sub> or formate and NH<sub>3</sub> for growth [66,67], because CO<sub>2</sub>, formate, and NH<sub>3</sub> readily enter metabolism, whereas cyanide does not. In other words, CO<sub>2</sub> directly enters and exits the organic chemistry of the cell substance at 415 reactions, where cyanide puts up a zero. We interpret the fact that cyanide has nothing to do with modern metabolism as a clear metabolic fossil: no role for cyanide in modern metabolism indicates that cyanide had nothing to do with primordial metabolism either, or was at best <1/415th as important as CO<sub>2</sub>. In that sense, the main message of Figure 1 is the overall scheme, the metabolism of cells, not that it contains amino acids and bases as products.

That brings us to nitrogen. If carbon did not enter metabolism via cyanide, then the same must be true for nitrogen. If not via cyanide, how did N enter metabolism? All the amino acids, bases, and cofactors contain nitrogen (except coenzyme M). Nitrogen enters metabolism as NH<sub>3</sub> (NH<sub>4</sub><sup>+</sup> is very unreactive) with N atoms replacing O atoms in amino acids, either via an acyl phosphate intermediate in the glutamine synthase reaction or via reductive aminations of 2-oxoacids [36,68]. An exception is the carbamoyl phosphate synthase reaction, in which NH<sub>3</sub> reacts with carboxyphosphate to form carbamate in pyrimidine and arginine biosynthesis. Of course, NH<sub>3</sub> is synthesized from N<sub>2</sub> by nitrogenase to make it available for incorporation into organic compounds. However, N does not enter metabolism as N<sub>2</sub>; it enters metabolism as NH<sub>3</sub>, which is why we selected NH<sub>3</sub> as the source of nitrogen in Figure 1. Similarly, sulfur enters metabolism as H<sub>2</sub>S in cysteine synthesis from serine, either via serine activation as *O*-acetylserine or *O*-phosphoserine [69]. N and S enter metabolism as dissolved gasses (NH<sub>3</sub> and H<sub>2</sub>S) via amino acid synthesis [36]. In cells that live from H<sub>2</sub> and CO<sub>2</sub>, C, N, S, and electrons (H<sub>2</sub>) enter metabolism as gasses.

### 3.2. Enzymatic Reactions in the Autotrophic Core

Figure 1a depicts the relationships among reactions that underpin the core synthesis of cells from H<sub>2</sub>, CO<sub>2</sub>, and NH<sub>3</sub>, but it does not depict the reactions themselves. To find out which, what kind of, and how many reactions are required to synthesize 18 cofactors, 8 nucleotides, and 20 amino acids from H<sub>2</sub>, CO<sub>2</sub>, NH<sub>3</sub>, and H<sub>2</sub>S, we turned to KEGG pathways using Figure 1 as a framework to identify the reactions and enzymes that catalyze them. The metabolic network for the 404 reactions (Table S2) that comprise the autotrophic core is shown in Figure 2.





**Figure 2.** The autotrophic core network of 404 reactions underlying Figure 1. The undirected bipartite graph comprises 404 reaction nodes (displayed as gray diamonds) and 380 compound nodes (circles). The 46 target compounds are colored blue; other compounds involved in the reactions appear orange. Target compounds correspond to the core compounds in Figure 1. Each compound participating in a reaction is connected to the respective reaction node with an edge. Compounds are sized according to node degree (number of reactions the compound takes place in). For example, H<sub>2</sub>O appears either as reactant or product in 125 reactions and is the most frequent compound in the 404 reactions (see also Table 1). In primordial metabolic processes, before the existence of enzymes, a more limited spectrum of compounds than those in Figure 1 was provided by the environment. Compound nodes are labeled if they are targets or if the node degree is  $\geq 20$ . Note that FeS clusters are not included in this figure since their synthesis cannot be reconstructed using KEGG. The network contains only L-amino acids.

Other than supplying a greater level of detail than Figure 1, and showing the relative size of nodes, the network itself in Figure 2 is not hugely informative, but some of its properties are. Keeping in mind that Figure 2 comprises the marrow of modern metabolism, hence reactions that were present in life's common ancestor, we first asked what the most highly connected metabolites are. The fifteen most common metabolites are given in Table 1. The most common compound in the autotrophic core is by far H<sub>2</sub>O. As stated above, water is the solvent of life's chemistry and its most common reaction partner. Proponents of the RNA world generally view water as a poison for origins, because it promotes the hydrolysis of RNA [70]. However, the host rocks of serpentinizing hydrothermal systems are replete with environments of low water activity, mainly because water is consumed by rock in the serpentinization process [71,72]. Furthermore, fluctuating water activities that occur during serpentinization can be conducive to polymerization reactions [71]. Life counteracts the hydrolysis problem by coupling nucleic acid and protein polymerization reactions to exergonic reactions via ATP synthesis and hydrolysis such that polymer synthesis vastly outpaces hydrolysis [58]. Accordingly, ATP is the second most common reactant in the autotrophic core (Table 1), right before protons. Protons are of course normally bound to water as H<sub>3</sub>O<sup>+</sup>, although they are not counted as water here. Protons arise in hydride transfer reactions involving NADH and NADPH

which yield NAD<sup>+</sup> and NADP<sup>+</sup>, respectively. The frequency of protons in the network mainly reflects the frequency of NAD(P)H-dependent redox reactions in the autotrophic core (Table 1).

Among reactions that involve the formation or alteration of bonds with carbon atoms, the most common carbon-containing compound in the autotrophic core is, fittingly, CO<sub>2</sub>, which underscores the CO<sub>2</sub>-dependent nature of core metabolism. Among the 404 reactions in the core, 49, or every eighth reaction, involves CO<sub>2</sub>. This can be seen as physiological evidence in favor of autotrophic origins. The next most common carbon backbone in the core is glutamate, which is the main workhorse of nitrogen metabolism. Glutamate arises as a product in amidotransferase reactions involving glutamine as an amino donor and in transamination reactions that produce 2-oxoglutarate, which is also among the top 15 reactants in the core. ATP hydrolysis products P<sub>i</sub> and PP<sub>i</sub> round out the list as well as pyruvate, which links the acetyl CoA pathway to sugar synthesis and the reverse citric acid cycle [73] and is a common starting point for cofactor synthesis in the autotrophic core (Figure 2). Last among the top fifteen is NH<sub>3</sub>, which is often donated to biosynthetic reactions from glutamine via an amidotransferase [74] during the enzymatic reaction, without being released as free NH<sub>3</sub> in the cytosol.

We identified five autocatalytic cycles in the network, that is, cofactors that are required for their own biosynthesis: pyridoxal phosphate and thiamine, whose biosyntheses were previously identified as autocatalytic cycles [36], plus ATP, NAD, and NADP. Though not contained within our set, Davidson recently reported that coenzyme A is required for activation of the complex that synthesizes the active moiety of decarboxylating pyruvoyl enzymes, which are involved in CoA biosynthesis [63]. That would make a sixth autocatalytic cycle.

### 3.3. Comparison of the Autotrophic Core With LUCA's Genes and Ancient Autocatalytic Sets

Other recent papers have addressed the nature of ancient metabolism by looking at phosphate-independent reactions among all KEGG reactions [75], the properties of thioester-dependent reactions [76] or chemical investigations of metabolic reactions without enzymes [30,37,49,77]. A different approach has been to focus on evidence for the nature of ancient microbial metabolism that is recorded in the genomes and metabolism of bacteria and archaea. A phylogenetic approach to ancient microbial metabolism uncovered 355 genes present in bacteria and archaea trace to LUCA on the basis of vertical intradomain inheritance as opposed to archaeal–bacteria transfer [50]. Autocatalytic networks called RAFs, for reflexively autocatalytic food-generated networks, have been identified in the metabolism of anaerobic autotrophs, with an ancient RAF of 172 genes that overlaps in the metabolism of H<sub>2</sub>-dependent acetogens and methanogens [51]. Do these sets overlap with the autotrophic core, and if so, how?

A comparison of these three sets (Table S3; Figure S1) reveals that among the 404 reactions of the autotrophic core, only 24 are represented among the 355 genes (6%) that trace to LUCA. That low degree of overlap is not surprising for two reasons. First, only a fraction of genes that trace to LUCA by phylogenetic criteria were involved in amino acid or cofactor biosynthesis, most being involved in ribosomal biogenesis or other categories. Second, only 3% of all genes shared by bacteria and archaea were not subjected to bacterial–archaeal transfers by the measure of phylogenetic trees [50], which is a criterion that played no role in the construction of Figure 1. However, it is very noteworthy that all of the cofactors shown in Figure 1, with the exception of the archaeal-specific cofactors CoM and CoB, do trace to LUCA via phylogeny, because enzymes that trace to LUCA possessed those cofactor requirements for activity [50]. In that sense, there is excellent agreement between the physiology of LUCA as inferred from phylogeny and the present autotrophic core, their commonality being cofactors, organic catalysts that are smaller and involved in a greater number of reactions than any individual enzyme.

Among the 172 reactions present in the ancient autocatalytic network shared by acetogen and methanogen RAFs [51], 81 (47%) are present in the autotrophic core. This

substantial overlap also makes sense, because all cells use the same amino acids and because both this study and Xavier et al. [51] focused on bacteria and archaea that use the acetyl CoA pathway, which by itself involves almost all of the cofactors shown in Figure 1 as it operates in bacteria and archaea. That is again noteworthy, because even though pyruvate, the central C3 product of the acetyl CoA pathway [6], can be obtained from H<sub>2</sub> and CO<sub>2</sub> using only simple minerals as catalysts [30], the biological pathway requires about a dozen enzymes and cofactors. These cofactors trace to LUCA [50], are well represented in RAFs [51], and comprise the basal foundation of the ancient autotrophic core (Figure 1b). Clearly, in early metabolism, cofactors and the catalytic minerals that were their inorganic precursors were very important [78]. Although self-evident, this indicates that there existed a vectorial progression in metabolic evolution that centered around the nature of catalysts: from transition metal minerals to organic cofactors to enzymes, each adding specificity and rate enhancement to exergonic reactions that tend to occur anyway. The retention of transition metal centers in some enzymes, such as carbon monoxide dehydrogenase, acetyl CoA synthetase, hydrogenases, or nitrogenase, suggests that microbes have been unable to invent catalysts that can perform the same reactions without the help of electrons in the *d*-orbitals of transition metals.

The comparison with 5994 anaerobic prokaryotic reactions (see S1A in [51]) tells us which compounds are enriched in the autotrophic core. Table S4 shows that this is true for ATP (and ADP plus P<sub>i</sub>), CO<sub>2</sub>, glutamate, pyruvate, and 2-oxoglutarate. This suggests a more crucial role of these compounds in the origin of the core subsequent to later evolution in anaerobes, reflecting a process of carbon backbone elongation from CO<sub>2</sub> at the heart of the core as a supply of precursors for cofactor and amino acid biosynthesis, the latter being the starting point for nucleotide biosynthesis [78].

#### 4. Conclusions

It is human nature to wonder about the origin of life, which is an issue that is among the most debated of all scientific questions. However, in comparison to questions concerning the existence of dark matter or how consciousness works, the origins process lies in the ancient past, and its events are only accessible through inference. Debates within the origins field can be fierce and have a long history. They hinge upon definitions about what qualifies as being alive, what one assumes to be the habitat that brought forth the first biochemical reactions, what came first, small molecule metabolism and proteins or nucleic acids and genetics, what the nature of first energy source(s) was that the early life forms harnessed in order to grow, and what kinds of chemical compounds existed before the first energy-releasing reactions germane to modern metabolism started taking place [21]. The literature harboring those debates is generally exhausting, because the same arguments resurface over and over again. The more broadly one reads the literature on early evolution, the more one gets the impression that scientists not only do not agree about origins and the nature of the first forms of life, but worse, that scientists know little about early evolution, leaving the topic open to unconstrained speculation and argument. That puts the origins field at risk of defining scientific progress in the units of debate preparation and presentation skills rather than units of empirical findings that are linked to the explanandum (real life); it also risks vulnerability to criticisms about the role of dogma in science.

Biologists tend to hold that there are traces of early evolution preserved in metabolism itself [4,6]. While there is no obvious proof for that conjecture, the nature of basic building blocks of life is dramatically well conserved across all cells [79]. All life forms we know use proteins made of amino acids, nucleic acids made of purines, pyrimidines, sugars, and phosphate. That means that the first forms of life from which all modern forms descend had that core chemistry in place, in addition to the universal genetic code to transfer information from nucleic acids to protein at the ribosome. This adds direly needed constraints to the origins problem. By looking at metabolism from a comparative standpoint, one can distill insights into the nature of early cells.

Here, we have identified 404 reactions that comprise the autotrophic core. It contains five small autocatalytic cycles in which cofactors participate in their own synthesis. The core represents a collection of reactions that underpin the synthesis of RNA and proteins. It was present in the first cells, but it can hardly have arisen all at once. The aqueous synthesis of pyruvate from  $H_2$  and  $CO_2$  using only solid-state metal or metal oxides as a catalyst [30] indicates that the core itself likely started from  $H_2$  and  $CO_2$  and grew outwards from pyruvate while incorporating nitrogen from  $NH_3$ . How complex the core could have become prior to the origin of enzymes is a question for future study. However, let us keep in mind that enzymes just accelerate reactions that tend to occur anyway. It is well known that many enzymatic reactions take place without enzymes [36], although sometimes, the non-enzymatic reaction rates can be so slow as to be irrelevant [80]. However, it was also demonstrated that citric acid cycle reactions [49,81] and a number of reactions involving sugars in central metabolism [77,82] can be catalyzed non-enzymatically. This suggests that a fairly complex system of reactions, yet with far less specificity than that in the core, could have arisen before the advent of genes and proteins.

$H_2O$  is the most common reactant in the autotrophic core, indicating an aqueous environment during its formation. That environment was not only aqueous but also reducing, as revealed by the abundance of redox reactions in the autotrophic core, the central role of  $CO_2$ , and the circumstance that the core's main products (amino acids and nucleic acids) are far more reduced than  $CO_2$ . Furthermore, the number of central reactions depending upon the hydrolysis of high-energy phosphate bonds indicates that the core arose in the presence of a continuous and highly exergonic chemical reaction capable of continuously synthesizing high-energy phosphate bonds, both before and after the origin of enzymes; here, an  $H_2$ -dependent  $CO_2$  reduction to acetate [30] forming acyl phosphate bonds [58] is the proposition.

Thus, the chemical reactions of the autotrophic core suggest that it formed in an aqueous environment that supplied  $H_2$ ,  $CO_2$ , and  $NH_3$ , was highly reducing, and harboring continuously far from equilibrium conditions. Those conditions are very similar to those found in serpentinizing hydrothermal systems [44,77], and furthermore, they are very similar to those inferred from the functions of enzymes that vertically trace to the last universal common ancestor [50,83].

Notwithstanding pyrrolysine [84], selenocysteine [85], and a number of modified bases [86], the lack of fundamental deviation among modern life forms from the core building blocks of life, core information processing, and the core repertoire of cofactors [87] indicates that whatever chemical processes occurred at origin did not give rise to alternative cores with enough staying power to persist to the present. "Still, other cores could have existed" the critic might interject, which is true. "But even if they existed, they are irrelevant", we would counter, because they are disjunct from the biologist's explanandum: the autotrophic core that we can observe in modern life forms.

**Supplementary Materials:** The following are available online at [www.mdpi.com/2076-2607/9/2/458/s1](http://www.mdpi.com/2076-2607/9/2/458/s1), Table S1: List of 47 target compounds of the autotrophic core metabolism, Table S2: Autotrophic core consisting of 404 metabolic reactions needed to synthesize amino acids, nucleic acids, cofactors and intermediate precursors, Table S3: (A) Reaction lists for 404 autotrophic core reactions, 163 LUCA reactions, 172 Core RAF reactions, (B) their respective intersections and (C) intersection of all three reaction sets with functional KEGG annotation, Table S4: Frequency of highly connected nodes among the autotrophic core and the global prokaryote anaerobic network and results of statistical tests for significant compound enrichment using Fisher's exact test, Figure S1: Venn diagram showing the proportion of intersecting reactions of three different core datasets.

**Author Contributions:** Conceptualization, W.F.M., J.L.E.W., J.C.X., M.P.; methodology, J.L.E.W., J.C.X.; formal analysis, J.L.E.W.; data curation, J.L.E.W., J.C.X.; writing—original draft preparation, W.F.M., M.P.; writing—review and editing, M.P., J.L.E.W., W.F.M., A.d.N.V., J.C.X., K.K.; visualization, J.L.E.W. and W.F.M.; supervision, W.F.M.; funding acquisition, W.F.M. and M.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the German Research Foundation (Ma 1426/21-1) as well as the Volkswagen Foundation (VW 96742).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Metabolic data that supports the findings of this study is available in KEGG [52]. The data presented in this study are available in supplementary material.

**Acknowledgments:** We thank Harun Tüysüz, Joseph Moran and Fernando Tria for discussions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Haeckel, E. *Natürliche Schöpfungs-Geschichte. Gemeinverständliche Wissenschaftliche Vorträge über die Entwicklungslehre Zehnte verbesserte Auflage. Zweiter Theil: Allgemeine Stammesgeschichte*; Georg Reimer Verlag: Berlin, Germany, 1902.
2. Winogradsky, S. *Beiträge zur Morphologie und Physiologie der Bakterien. H. 1. Zur Morphologie und Physiologie der Schwefelbakterien*; A. Felix: Leipzig, Germany, 1888.
3. Madigan, M.T.; Martinko, J.M.; Bender, K.S.; Buckley, D.H.; Stahl, D.A. *Brock Biology of Microorganisms*, 15th ed.; Pearson Education: Harlow, UK, 2019.
4. Decker, K.; Jungermann, K.; Thauer, R.K. Energy production in anaerobic organisms. *Angew. Chem. Int. Ed.* **1970**, *9*, 138–158, doi:10.1002/anie.197001381.
5. Baross, J.A.; Hoffmann, S.E. Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life. *Orig. Life Evol. Biosph.* **1985**, *15*, 327–345.
6. Fuchs, G. Alternative pathways of carbon dioxide fixation: Insights into the early evolution of life? *Annu. Rev. Microbiol.* **2011**, *65*, 631–658, doi:10.1146/annurev-micro-090110-102801.
7. Kowallik, K.V.; Martin, W.F. The origin of symbiogenesis: An annotated English translation of Mereschkowsky's 1910 paper on the theory of two plasma lineages. *Biosystems* **2021**, *199*, 104281, doi:10.1016/j.biosystems.2020.104281.
8. Miller, S.L. A production of amino acids under possible primitive earth conditions. *Science* **1953**, *117*, 528–529, doi:10.1126/science.117.3046.528.
9. Oró, J.; Kimball, A.P. Synthesis of purines under possible primitive earth conditions. I. Adenine from hydrogen cyanide. *Arch. Biochem. Biophys.* **1961**, *94*, 217–227, doi:10.1016/0003-9861(61)90033-9.
10. Haruna, I.; Spiegelman, S. Specific template requirements of RNA replicases. *Proc. Natl. Acad. Sci. USA* **1965**, *54*, 579–587, doi:10.1073/pnas.54.2.579.
11. Cech, T.R. The chemistry of self-splicing enzymes. *Science* **1986**, *236*, 1532–1539, doi:10.1126/science.2438771.
12. Gilbert, W. The RNA world. *Nature* **1986**, *319*, 618, doi:10.1038/319618a0.
13. Lincoln, T.A.; Joyce, G.F. Self-sustained replication of an RNA enzyme. *Science* **2009**, *323*, 1229–1232, doi:10.1126/science.1167856.
14. Powner, M.W.; Gerland, B.; Sutherland, J.D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **2009**, *459*, 239–242, doi:10.1038/nature08013.
15. Sutherland, J.D. The origin of life—Out of the blue. *Angew. Chem. Int. Ed.* **2016**, *55*: 104–121, doi:10.1002/anie.201506585.
16. Shapiro, R. Small molecule interactions were central to the origin of life. *Q. Rev. Biol.* **2006**, *81*, 105–126, doi:10.1086/506024.
17. Wächtershäuser, G. Before enzymes and templates: Theory of surface metabolism. *Microbiol. Rev.* **1988**, *52*, 452–484.
18. Orgel, L.E. The implausibility of metabolic cycles on the prebiotic earth. *PLoS Biol.* **2008**, *6*, e18–e19, doi:10.1371/journal.pbio.0060018.
19. Kamminga, H. Historical perspective: The problem of the origin of life in the context of developments in biology. *Orig. Life Evol. Biosph.* **1988**, *18*, 1–11, doi:10.1007/BF01808777.
20. Pascal, R.; Pross, A. Stability and its manifestation in the chemical and biological worlds. *Chem. Commun.* **2015**, *51*, 16160–16165, doi:10.1039/C5CC06260H.
21. Preiner, M.; Asche, S.; Becker, S.; Betts, H.C.; Boniface, A.; Camprubi, E.; Chandru, K.; Erastova, V.; Garg, S.G.; Khawaja, N.; et al. The future of origin of life research: Bridging decades-old divisions. *Life* **2020**, *10*, doi.org/10.3390/life10030020.
22. Lipmann, F. Projecting backward from the present stage of evolution of biosynthesis. In *The Origin of Prebiological Systems and of Their Molecular Matrices*; Fox, S.W., Ed.; Academic Press: New York, NY, USA, 1965; pp. 259–280.
23. Ferry, J.G.; House, C.H. The stepwise evolution of early life driven by energy conservation. *Mol. Biol. Evol.* **2006**, *23*, 1286–1292, doi:10.1093/molbev/msk014.
24. Corliss, J.B.; Dymond, J.; Gor, L.I.; Edmond, J.M.; Herzen, R.P. Von; Bal, R.D.; Green, K.; Williams, D.; Bainbri, A.; Crane, K.; et al. Submarine thermal springs on the Galapagos rift. *Science* **1979**, *203*, 1073–1083, doi:10.1126/science.203.4385.1073.
25. Kelley, D.S.; Karson, J.A.; Blackman, D.K.; Früh-Green, G.L.; Butterfield, D.A.; Lilley, M.D.; Olson, E.J.; Schrenk, M.O.; Roe, K.K.; Lebon, G.T.; et al. An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30° N. *Nature* **2001**, *412*, 145–149, doi:10.1038/35084000.

26. Charlou, J.L.; Donval, J.P.; Fouquet, Y.; Jean-Baptiste, P.; Holm, N. Geochemistry of high H<sub>2</sub> and CH<sub>4</sub> vent fluids issuing from ultramafic rocks at the Rainbow hydrothermal field (36°14' N, MAR). *Chem. Geol.* **2002**, *191*, 345–359, doi:10.1016/S0009-2541(02)00134-1.
27. Thauer, R.K.; Kaster, A.-K.; Seedorf, H.; Buckel, W.; Hedderich, R. Methanogenic archaea: Ecologically relevant differences in energy conservation. *Nat. Rev. Microbiol.* **2008**, *6*, 579–591, doi:10.1038/nrmicro1931.
28. Sleep, N.H.; Bird, D.K.; Pope, E.C. Serpentinite and the dawn of life. *Philos. Trans. R. Soc. B* **2011**, *366*, 2857–2869, doi:10.1098/rstb.2011.0129.
29. Martin, W.; Baross, J.; Kelley, D.; Russell, M.J. Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* **2008**, *6*, 805–814, doi:10.1038/nrmicro1991.
30. Preiner, M.; Igarashi, K.; Muchowska, K.B.; Yu, M.; Varma, S.J.; Kleinermanns, K.; Nobu, M.K.; Kamagata, Y.; Tüysüz, H.; Moran, J.; et al. A hydrogen dependent geochemical analogue of primordial carbon and energy metabolism. *Nat. Ecol. Evol.* **2020**, *4*, 534–542, doi:10.1038/s41559-020-1125-6.
31. Schönheit, P.; Buckel, W.; Martin, W.F. On the origin of heterotrophy. *Trends Microbiol.* **2016**, *24*, 12–25, doi:10.1016/j.tim.2015.10.003.
32. Harold, F.M. *The Vital Force: A Study of Bioenergetics*; WH Freeman: New York, NY, USA, 1986.
33. Russell, J.B.; Cook, G.M. Energetics of bacterial growth: Balance of anabolic and catabolic reactions. *Microbiol. Rev.* **1995**, *59*, 48–62.
34. Russell, J.B. The energy spilling reactions of bacteria and other organisms. *J. Mol. Microbiol. Biotechnol.* **2007**, *13*, 1–11, doi:10.1159/000103591.
35. Van Dijken, J.P.; Scheffers, W.A. Redox balances in the metabolism of sugars by yeasts. *FEMS Microbiol. Rev.* **1986**, *32*, 199–224, doi:10.1111/j.1574-6968.1986.tb01194.x.
36. Martin, W.F.; Russell, M.J. On the origin of biochemistry at an alkaline hydrothermal vent. *Philos. Trans. R. Soc. B* **2007**, *362*, 1887–1925, doi:10.1098/rstb.2006.1881.
37. Muchowska, K.B.; Varma, S.J.; Moran, J. Synthesis and breakdown of universal metabolic precursors promoted by iron. *Nature* **2019**, *569*, 104–107, doi:10.1038/s41586-019-1151-1.
38. Ralser, M. An appeal to magic? The discovery of a non-enzymatic metabolism and its role in the origins of life. *Biochem. J.* **2018**, *475*, 2577–2592, doi:10.1042/BCJ20160866.
39. Thauer, R.K.; Jungermann, K.; Decker, K. Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol. Rev.* **1977**, *41*, 100–180.
40. Stouthamer, A.H. Energy-yielding pathways. In *The Bacteria*; Gunsalus, I.C., Ed.; Academic Press: New York, NY, USA, 1978; Volume 6, pp. 389–462.
41. Hoehler, T.M.; Jørgensen, B.B. Microbial life under extreme energy limitation. *Nat. Rev. Microbiol.* **2013**, *11*, 83–94, doi:10.1038/nrmicro2939.
42. Arndt, N.T.; Nisbet, E.G. Processes on the young earth and the habitats of early life. *Annu. Rev. Earth Planet. Sci.* **2012**, *40*, 521–549, doi:10.1146/annurev-earth-042711-105316.
43. Russell, M.J.; Hall, A.J.; Martin, W. Serpentinization as a source of energy at the origin of life. *Geobiology* **2010**, *8*, 355–371, doi:10.1111/j.1472-4669.2010.00249.x.
44. Preiner, M.; Xavier, J.C.; Sousa, F.; Zimorski, V.; Neubeck, A.; Lang, S.Q.; Greenwell, H.C.; Kleinermanns, K.; Tüysüz, H.; McCollom, T.M.; et al. Serpentinization: Connecting geochemistry, ancient metabolism and industrial hydrogenation. *Life* **2018**, *8*, 41, doi:10.3390/life8040041.
45. Proskurowski, G.; Lilley, M.D.; Seewald, J.S.; Früh-Green, G.L.; Olson, E.J.; Lupton, J.E.; Sylva, S.P.; Kelley, D.S. Abiogenic hydrocarbon production at lost city hydrothermal field. *Science* **2008**, *319*, 604–607, doi:10.1126/science.1151194.
46. Lang, S.Q.; Butterfield, D.A.; Schulte, M.; Kelley, D.S.; Lilley, M.D. Elevated concentrations of formate, acetate and dissolved organic carbon found at the Lost City hydrothermal field. *Geochim. Cosmochim. Acta* **2010**, *74*, 941–952, doi:10.1016/j.gca.2009.10.045.
47. Konn, C.; Charlou, J.L.; Holm, N.G.; Mousis, O. The production of methane, hydrogen and organic compounds in ultramafic-hosted hydrothermal vents of the Mid-Atlantic Ridge. *Astrobiology* **2015**, *15*, 381–399, doi:10.1089/ast.2014.1198.
48. Schrenk, M.O.; Brazelton, W.J.; Lang, S.Q. Serpentinization, carbon, and deep life. *Rev. Mineral. Geochem.* **2013**, *75*, 575–606, doi:10.2138/rmg.2013.75.18.
49. Varma, S.J.; Muchowska, K.B.; Chatelain, P.; Moran, J. Native iron reduces CO<sub>2</sub> to intermediates and endproducts of the acetyl-CoA pathway. *Nat. Ecol. Evol.* **2018**, *2*, 1019–1024, doi:10.1038/s41559-018-0542-2.
50. Weiss, M.C.; Sousa, F.L.; Mrnjavac, N.; Neukirchen, S.; Roettger, M.; Nelson-Sathi, S.; Martin, W.F. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **2016**, *1*, 16116, doi:10.1038/nrmicrobiol.2016.116.
51. Xavier, J.C.; Hordijk, W.; Kauffman, S.; Steel, M.; Martin, W.F. Autocatalytic chemical networks preceded proteins and RNA in evolution. *Proc. R. Soc. Lond. B* **2020**, *287*, doi:10.1098/rspb.2019.2377.
52. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30, doi:10.1093/nar/28.1.27.
53. Qiu, Y.Q. KEGG pathway database. In *Encyclopedia of Systems Biology*; Dubitzky, W., Wolkenhauer, O., Cho, K.H., Yokota, H., Eds.; Springer: New York, NY, USA, 2013; pp. 1068–1069.

54. Hazra, A.B.; Han, A.W.; Mehta, A.P.; Mok, K.C.; Osadchiy, V.; Begley, T.P.; Taga, M.E. Anaerobic biosynthesis of the lower ligand of vitamin B12. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10792–10797, doi:10.1073/pnas.1509132112.
55. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504, doi:10.1101/gr.1239303.
56. Berg, I.A.; Kockelkorn, D.; Ramos-Vera, W.H.; Say, R.F.; Zarzycki, J.; Hügler, M.; Alber, B.E.; Fuchs, G. Autotrophic carbon fixation in archaea. *Nat. Rev. Microbiol.* **2010**, *8*, 447–460, doi:10.1038/nrmicro2365.
57. Hügler, M.; Sievert, S.M. Beyond the Calvin cycle: Autotrophic carbon fixation in the ocean. *Ann. Rev. Mar. Sci.* **2011**, *3*, 261–289, doi:10.1146/annurev-marine-120709-142712.
58. Martin, W.F. Older than genes: The acetyl CoA pathway and origins. *Front. Microbiol.* **2020**, *11*, 817, doi:10.3389/fmicb.2020.00817.
59. Mall, A.; Sobotta, J.; Huber, C.; Tschirner, C.; Kowarschik, S.; Bačnik, K.; Mergelsberg, M.; Boll, M.; Hügler, M.; Eisenreich, W.; et al. Reversibility of citrate synthase allows autotrophic growth of a thermophilic bacterium. *Science* **2018**, *359*, 563–567, doi:10.1126/science.aao2410.
60. Nunoura, T.; Chikaraishi, Y.; Izaki, R.; Suwa, T.; Sato, T.; Harada, T.; Mori, K.; Kato, Y.; Miyazaki, M.; Shimamura, S.; et al. A primordial and reversible TCA cycle in a facultatively chemolithoautotrophic thermophile. *Science* **2018**, *359*, 559–563, doi:10.1126/science.aao3407.
61. Maden, B.E. Tetrahydrofolate and tetrahydromethanopterin compared: Functionally distinct carriers in C1 metabolism. *Biochem. J.* **2000**, *350*, 609–629, doi:10.1042/0264-6021:3520935.
62. Sousa, F.L.; Martin, W.F. Biochemical fossils of the ancient transition from geoenergetics to bioenergetics in prokaryotic one carbon compound metabolism. *Biochim. Biophys. Acta* **2014**, *1837*, 964–981, doi:10.1016/j.bbabi.2014.02.001.
63. Davidson, V.L. Protein-derived cofactors revisited: Empowering amino acid residues with new functions. *Biochemistry* **2018**, *57*, 3115–3125, doi:10.1021/acs.biochem.8b00123.
64. Canavelli, P.; Islam, S.; Powner, M.W. Peptide ligation by chemoselective aminonitrile coupling in water. *Nature* **2019**, *571*, 546–549, doi:10.1038/s41586-019-1371-4.
65. Xavier, J.C.; Preiner, M.; Martin, W.F. Something special about CO-dependent CO<sub>2</sub> fixation. *FEBS J.* **2018**, *285*, 4181–4195, doi:10.1111/febs.14664.
66. Knowles, C.J. Cyanide utilization and degradation by microorganisms. *Ciba. Found. Symp.* **1988**, *140*, 3–15, doi:10.1002/9780470513712.ch2.
67. Fernandez, R.F.; Dolgih, E.; Kunz, D.A. Enzymatic assimilation of cyanide via pterin-dependent oxygenolytic cleavage to ammonia and formate in *Pseudomonas fluorescens* NCIMB 11764. *Appl. Environ. Microbiol.* **2004**, *70*, 121–128, doi:10.1128/aem.70.1.121-128.2004.
68. Huber, C.; Wächtershäuser, G.  $\alpha$ -hydroxy and  $\alpha$ -amino acids under possible hadean, volcanic origin-of-life conditions. *Science* **2006**, *314*, 630–632, doi:10.1126/science.1130895.
69. Liu, Y.; Beer, L.L.; Whitman, W.B. Sulfur metabolism in archaea reveals novel processes. *Environ. Microbiol.* **2012**, *14*, 2632–2644, doi:10.1111/j.1462-2920.2012.02783.x.
70. Marshall, M. How the first life on Earth survived its biggest threat—Water. *Nature* **2020**, *588*, 210–213, doi:10.1038/d41586-020-03461-4.
71. Do Nascimento Vieira, A.; Kleinermaans, K.; Martin, W.F.; Preiner, M. The ambivalent role of water at the origins of life. *FEBS Lett.* **2020**, *594*, 2717–2733, doi:10.1002/1873-3468.13815.
72. Lamadrid, H.M.; Rimstidt, J.D.; Schwarzenbach, E.M.; Klein, F.; Ulrich, S.; Dolocan, A.; Bodnar, R.J. Effect of water activity on rates of serpentinization of olivine. *Nat. Commun.* **2017**, *8*, 16107, doi:10.1038/ncomms16107.
73. Berg, I.A.; Kockelkorn, D.; Buckel, W.; Fuchs, G. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in archaea. *Science* **2007**, *318*, 1782–1786, doi:10.1126/science.1149976.
74. Raushel, F.M.; Thoden, J.B.; Holden, H.M. The amidotransferase family of enzymes: Molecular machines for the production and delivery of ammonia. *Biochemistry* **1999**, *38*, 7891–7899, doi:10.1021/bi990871p.
75. Goldford, J.E.; Hartman, H.; Smith, T.F.; Segrè, D. Remnants of an ancient metabolism without phosphate. *Cell* **2017**, *168*, 1126–1134.e9, doi:10.1016/j.cell.2017.02.001.
76. Semenov, S.N.; Kraft, L.J.; Ainla, A.; Zhao, M.; Baghbanzadeh, M.; Campbell, V.E.; Kang, K.; Fox, J.M.; Whitesides, G.M. Autocatalytic, bistable, oscillatory networks of biologically relevant organic reactions. *Nature* **2016**, *537*, 656–660, doi:10.1038/nature19776.
77. Muchowska, K.B.; Varma, S.J.; Moran, J. Nonenzymatic metabolic reactions and life's origins. *Chem. Rev.* **2020**, *120*, 7708–7744, doi:10.1021/acs.chemrev.0c00191.
78. Preiner, M.; Xavier, J.C.; Vieira, A.D.N.; Kleinermaans, K.; Allen, J.F.; Martin, W.F. Catalysts, autocatalysis and the origin of metabolism. *Interface Focus* **2019**, *9*, 20190072, doi:10.1098/rsfs.2019.0072.
79. Morowitz, H.J.; Kostelnik, J.D.; Yang, J.; Cody, G.D. The origin of intermediary metabolism. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7704–7708, doi:10.1073/pnas.110153997.
80. Wolfenden, R. Benchmark reaction rates, the stability of biological molecules in water, and the evolution of catalytic power in enzymes. *Annu. Rev. Biochem.* **2011**, *80*, 645–667, doi:10.1146/annurev-biochem-060409-093051.

81. Muchowska, K.B.; Varma, S.J.; Chevallot-Beroux, E.; Lethuillier-Karl, L.; Li, G.; Moran, J. Metals promote sequences of the reverse Krebs cycle. *Nat. Ecol. Evol.* **2017**, *1*, 1716–1721, doi:10.1038/s41559-017-0311-7.
82. Messner, C.B.; Driscoll, P.C.; Piedrafita, G.; De Volder, M.F.L.; Ralser, M. Nonenzymatic gluconeogenesis-like formation of fructose 1,6-bisphosphate in ice. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 7403–7407, doi:10.1073/pnas.1702274114.
83. Sousa, F.L.; Nelson-Sathi, S.; Martin, W.F. One step beyond a ribosome: The ancient anaerobic core. *Biochim. Biophys. Acta.* **2016**, *1857*, 1027–1038, doi:10.1016/j.bbabi.2016.04.284.
84. Krzycki, J.A. The direct genetic encoding of pyrrolysine. *Curr. Opin. Microbiol.* **2005**, *8*, 706–712, doi:10.1016/j.mib.2005.10.009.
85. Böck, A.; Forchhammer, K.; Heider, J.; Leinfelder, W.; Sawers, G.; Veprek, B.; Zinoni, F. Selenocysteine: The 21st amino acid. *Mol. Microbiol.* **1991**, *5*, 515–520, doi:10.1111/j.1365-2958.1991.tb00722.x.
86. Grosjean, H.; de Crécy-Lagard, V.; Marck, C. Deciphering synonymous codons in the three domains of life: Co-evolution with specific tRNA modification enzymes. *FEBS Lett.* **2010**, *584*, 252–264, doi:10.1016/j.febslet.2009.11.052.
87. Xavier, J.C.; Patil, K.R.; Rocha, I. Integration of biomass formulations of genome-scale metabolic models with experimental data reveals universally essential cofactors in prokaryotes. *Metab. Eng.* **2017**, *39*, 200–208, doi:10.1016/j.ymben.2016.12.002.