

Excellence in Chemistry Research

**Announcing
our new
flagship journal**

- Gold Open Access
- Publishing charges waived
- Preprints welcome
- Edited by active scientists



Meet the Editors of *ChemistryEurope*



Luisa De Cola

Università degli Studi
di Milano Statale, Italy



Ive Hermans

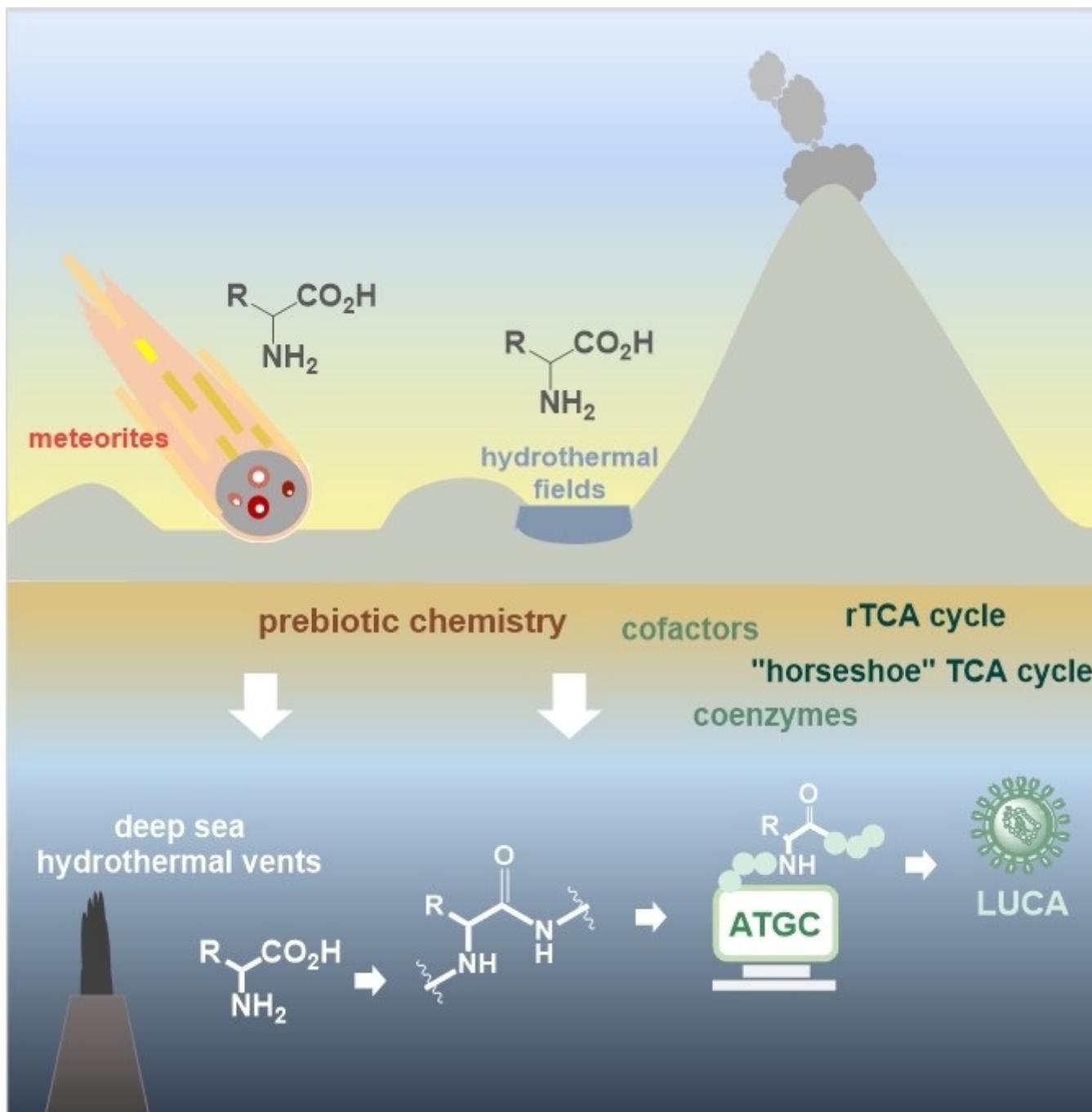
University of
Wisconsin-Madison, USA



Ken Tanaka

Tokyo Institute of
Technology, Japan

On the Evolutionary History of the Twenty Encoded Amino Acids

Andreas Kirschning*^[a]

Abstract: α -Amino acids are essential molecular constituents of life, twenty of which are privileged because they are encoded by the ribosomal machinery. The question remains open as to why this number and why this 20 in particular, an almost philosophical question that cannot be conclusively resolved. They are closely related to the evolution of the genetic code and whether nucleic acids, amino acids, and peptides appeared simultaneously and were available under prebiotic conditions when the first self-sufficient complex molecular system emerged on Earth. This report focuses on prebiotic and metabolic aspects of amino acids and proteins

starting with meteorites, followed by their formation, including peptides, under plausible prebiotic conditions, and the major biosynthetic pathways in the various kingdoms of life. Coenzymes play a key role in the present analysis in that amino acid metabolism is linked to glycolysis and different variants of the tricarboxylic acid cycle (TCA, rTCA, and the incomplete horseshoe version) as well as the biosynthesis of the most important coenzymes. Thus, the report opens additional perspectives and facets on the molecular evolution of primary metabolism.

1. Introduction

Reflections on the origin of life are speculative per se. And fundamental questions such as a) how the 20 proteinogenic amino acids were selected and b) in what temporal sequence they are likely to have appeared on the world stage, open up the possibility of deductively delving far into the darkness of the past.^[1] With respect to proteins, there is a growing body of evidence that the modern genetically encoded amino acid alphabet is evolutionarily highly optimized. These include the structural chemical diversity of the side chains^[2] and the number twenty, which is found in all kingdoms of life and also in viruses, so that their roots must lie far back in the past. Indeed, there has been substantial progress in understanding amino acid evolution^[3] and one scenario argues for an intensive selection process and the reduction to twenty amino acids, while the alternative suggests that new amino acids appeared successively over time and were fitted into the repertoire of protein synthesis?

The Hadean eon represents the time when the earth first formed and abiotic synthesis prevailed (physical-chemical evolution). The following eon of the Archean (about 3,500 billion years ago) is described as the age of bacteria and archaea. By this time, template-based peptide and protein synthesis as well as biosynthetic pathways to individual amino acids must have been established (biological evolution) (Figure 1).

How life may have arisen is now essentially shaped by two hypotheses: The first ("RNA world")^[5] holds that the first living molecules were RNA molecules capable of carrying both catalytic and heritable information. The second hypothesis,

which has gained strong support, especially recently, is called "metabolism first" (Figure 1).^[6] It states that metabolic reactions preceded genetic information carriers. The reactions proceeded in self-sustaining cycles, each growing into an ever larger cycle. Occasionally, these reactions split into two independent routes, so that evolution to new compounds or more efficient pathways could have developed from these mixtures of reactions.

The present report does not aim at advocating one of these two propagated theories. Rather, it follows up on an earlier article that highlighted the relationship of coenzymes and cofactors to the RNA world theory.^[7] This report now addresses the evolutionary interplay of metabolism and the emergence of coenzymes and cofactors, by specifically focusing on the metabolism of amino acids, which feeds primarily through sugar metabolism and the citric acid cycle and ultimately leads to the question of the emergence of the genetic code. Thus, this report is not in opposition to the RNA world theory but complements the earlier article and expands the evolutionary view of coenzymes and cofactors within the prevailing theory of coevolution of peptides and nucleic acids.

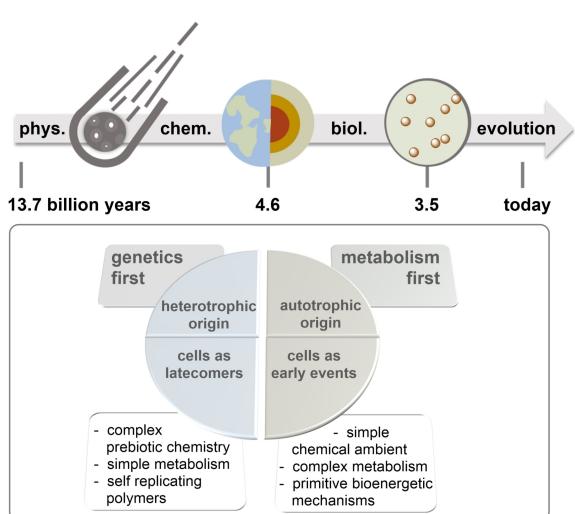


Figure 1. The principal phases of evolution and the two theories including the stages from abiotically formed amino acids to their metabolism as discussed in this report.^[4]

[a] Prof. Dr. A. Kirschning
Institute of Organic Chemistry
Leibniz University Hannover
Schneiderberg 18, 30167 Hannover (Germany)
E-mail: andreas.kirschning@oci.uni-hannover.de

 Selected by the Editorial Office for our Showcase of outstanding Review-type articles <http://www.chemeurj.org/showcase>.

 © 2022 The Authors. Chemistry - A European Journal published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

2. Abiogenesis of Amino Acids and Peptides

2.1. The astrochemical scenario

Meteorites, especially samples from the Murchison meteorite and the martian meteorite ALH84001, as well as in the coma of 67P/Churyumov-Gerasimenko measured by the ROSINA^[8,9] are a fruitful sources for amino acids. Based on their mineral constituents and rock material, they are considered to be similar in geological characteristics to the planetesimals that have survived since the formation of the solar nebula about 4.5 billion years ago.^[10,11] In fact, more than 80 amino acids, both proteinogenic and non-proteinogenic, have been found in meteorites to date in concentration ranges up to about 60 ppm.^[12–14] Twelve of them belong to the proteinogenic group (see Table 1),^[13c] that give rise to hypotheses about the minimum number of amino acids for subsequent steps in molecular evolution.^[15] In particular, the presence of uracil and xanthine helped to identify the collected molecular samples as non-terrestrial in origin based on carbon isotope ratio measurements.^[16] There is only one reference about the existence of dipeptides and diketopiperazines of abiotic origin in a sample of the Murchison^[17,18] and in the meteorite Y-791198.^[18] Recently, it was confirmed that g lycine and other amino acids form in dense interstellar clouds well before they transform into new stars and planets.^[19] Chemically, the combination of the formose reaction and the Strecker reaction have been spotted to have likely played a role but UV photolysis of interstellar ice analogs and electrons mimicking cosmic rays have also been made responsible for the interstellar formation of amino acids.^[20]

Andreas Kirschning studied chemistry at the University of Hamburg and at Southampton University (UK). In Hamburg he joined the group of Prof. Ernst Schaumann and received his PhD in 1989 in the field of organosilicon chemistry. After a postdoctoral stay at the University of Washington (Seattle, USA) with Prof. Heinz G. Floss, he moved to the Technical University of Clausthal in 1991. In 2000 he became full professor at Leibniz University of Hannover. His research interests include all aspects of natural products including mutasynthesis and the use of terpene synthases. Another important aspect of his research addresses the development and combination of enabling technologies in organic synthesis, specifically flow chemistry in combination with inductive heating techniques. Recently, he has devoted himself to questions concerning the origin of life with special reference to the role of coenzymes and cofactors in evolution.



2.2. Prebiotic amino acid and peptide synthesis

It seems rather unlikely that the evolution of life was fed solely, or at all, by the arrival of extraterrestrial organic components, including amino acids, through meteoritic carriers. Several different environments have been proposed as plausible locations for the starting point of the origin of life on planet Earth, and among these, hydrothermal vents and hydrothermal fields^[40] have been favoured. It is accepted that the origin of life did not occur in a single location, as various conditions, some very specific, including heat, light, catalytic surfaces, and reductive environments must have been necessary to establish premetabolic networks.^[21]

Conceivably, products and reactants were transported between individual compartmentally distinct regions of different millieus. Classically, the Strecker synthesis using aldehydes, HCN and ammonia yields α -aminonitriles and eventually α -amino acids. In one variant, aldehydes and HCN are combined in water at neutral pH in the presence of diamidophosphate (DAP) to give *N*-phosphoraminonitriles, which upon treatment with H₂S provide α -aminothioamides.^[22a]

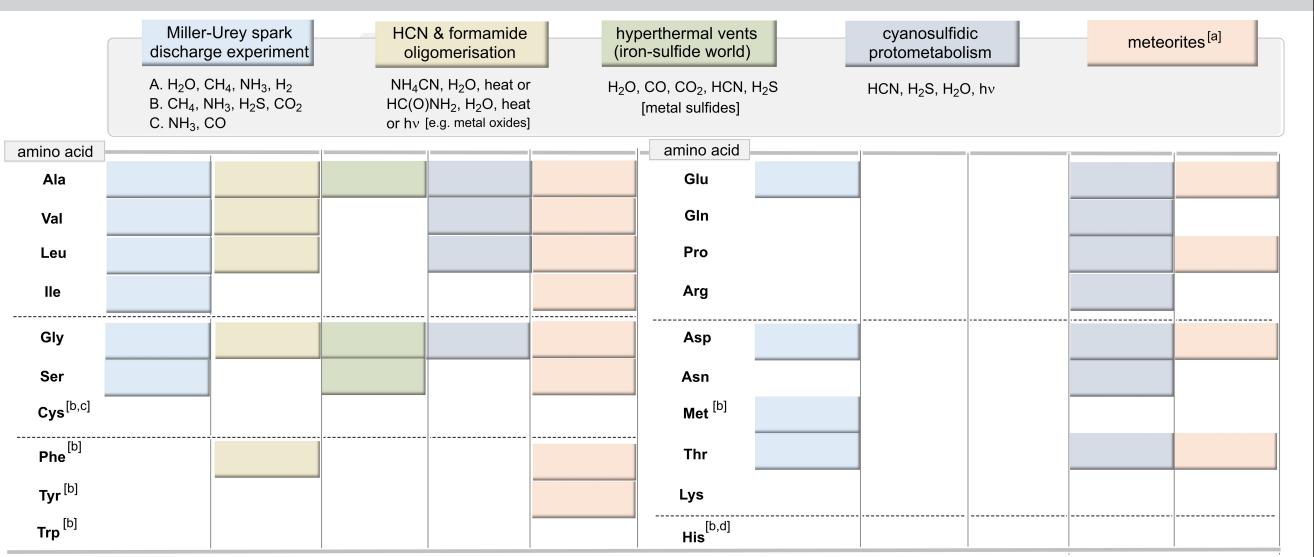
Miller's spark discharge experiments, three of which are listed in Table 1, represented the first attempts to conduct abiotic chemistry under simulated conditions of the primordial Earth.^[23] In a variation of the classic experiment (H₂, H₂O, CH₄, and NH₃) H₂S and CO₂ were added, the outcome was recently reanalyzed (Table 1).^[24] It revealed the presence of other amino acids, particularly methionine and cysteine, although the latter could only be confirmed by analysis of the degradation products.

Oró pointed out that the nucleotide bases are formally oligomers of hydrogen cyanide, and indeed adenine is formed from concentrated aqueous ammonium cyanide under refluxing conditions (Table 1).^[25] In addition, some amino acids were reported to have formed from HCN and ammonia. Closely related to HCN is its formal hydrate formamide. In a series of studies under high energy conditions^[26] with formamide generated from ammonia and CO or formic acid under high temperature conditions led to various carboxylic acids and amino acids as well as urea and carbodiimide. In one case, various meteorite-derived additives served as heterogeneous catalysts.^[27]

The iron–sulfur hypothesis states that deposits of iron sulfide minerals near deep-sea hydrothermal vents^[28] are capable of providing the reductive medium and energy to catalyze complex reaction sequences from simple precursors such as CO and CO₂, HCN and H₂S. The formation of small molecules were identified under simulated conditions in an autoclave; for example ammonia is produced from nitrate under these reducing conditions (FeS, H₂S).^[29]

Methanethiol was also a product from which S-methyl ethanethioate (H₃C–CO–SCH₃) is generated in the presence of nickel and iron sulfide.^[30,31] In the presence of HCN, α -hydroxy and α -amino acids such as glycolate/glycine, lactate/alanine, glycinate/serine, and pyruvic acid are generated. Also 2-oxocarboxylic acids including pyruvic acid react with ammonia in the presence of iron hydroxide or iron sulfide and H₂S to

Table 1. Summary on four major experimental set-ups or protometabolic approaches, their chemical conditions, and listing of amino acids formed in these scenarios as well as amino acids found on carbonaceous meteorites.^[13,14]



[a] Status 2018 according to Ref. [13c]; [b] these amino acids are decomposed by ultraviolet radiation, which prevailed in space and on the early Earth; [c] Pownall and co-workers reported a cysteine synthesis based on a thio-Michael addition of H₂S to an acrylonitrile derivative (ref. [22b]); [d] only imidazole has been detected in carbonaceous meteorites.

form α -amino acids including alanine.^[32] The reaction of α -amino acids with carbonyl sulfide (COS), another important product, resulted in dipeptides and tripeptides (see section).^[33,34] Furthermore, it was reported that short phenylalanine oligopeptides are generated from the monomer when heated to 100 °C in the presence of (Ni,Fe)S and CO.^[35] An attractive aspect of this theory is related to the reductive properties of H₂S, which also plays a key role in Sutherland's cyanosulfidic protometabolism. It is based on the so-called cyanosulfidic chemical homologation process,^[36] in which the key building blocks of RNA, proteins, and lipids are formed only from hydrogen cyanide as the sole source of carbon and nitrogen. Hydrogen sulfide serves as a reducing agent under UV irradiation conditions in the presence of Cu(I)/Cu(II) catalysts and this system effectively acts as a photoredox system.^[37] An important aspect in the light of this article represents the progression of cyanosulfidic protometabolism to amino acids. Without going into details here^[38,39] Sutherland and coworkers found multistep chemical routes leading to the amino acids glycine, alanine, serine, threonine, and proline but also to the more polar representatives asparagine, aspartate, glutamine, and glutamate and finally to the nonpolar amino acids valine and leucine (Table 1).^[36]

None of these four approaches can claim to be universally valid, each has some shortcomings, be it high concentrations required (HCN, DMF), be it long synthetic pathways (cyanosulfidic protometabolism), be it dilution problems and questions about the influence of saline media on the downstream formation of complex molecular systems leading to protometabolic networks (deep-sea hydrothermal vents). The inherent problem of increasing dilution with each reaction step, simply because the individual yields can be far from quantitative, is

overcome by so called wet-dry cycles. They are particularly well realized in hydrothermal fields. These cycles, consisting of repetitive dilution by rain and short-term generated small pond-filling streams, and concentration by evaporation, are also considered essential sites where chemical steps for prebiotic formation of peptides may have occurred.^[40–42]

It is obvious that in different prebiotic milieus several amino acids such as glycine and alanine seem to form inevitably. For other amino acids, such as the two basic amino acids histidine and lysine but also the aromatic amino acids, especially tryptophan, the opposite is true. But model experiments showed that histidine is accessible from imidazole-4-acetaldehyde which in turn is formed from formamidine and erythrose.^[43]

Prebiotic pathways to phenylalanine and tyrosine under primitive early Earth conditions have also been explored, with phenylacetylene playing a central role.^[44] This is thought to form from various hydrocarbons including acetylene at high temperatures, electrical discharges, or by ultraviolet light. H₂S addition and hydrolysis would produce phenylacetaldehyde, which yields phenylalanine by Strecker chemistry. Experimentally, the authors also reported on trace formation of tyrosine. Mentionably, some alkynes such as diacetylene are said to be present on Saturn's moon Titan.^[45] The suggestion that tryptophan can form from indole and pyruvate under Friedel-Crafts conditions with iron-rich saponite followed by transamination under hydrothermal conditions must be considered speculative, also because little is known about whether indole was present in the prebiotic world.^[46] For some amino acids, no convincing synthetic routes under prebiotic conditions have been found so far; e.g. the one proposed for arginine has been questioned by some authors.^[47,48]

Among the five scenarios compiled in Table 1, the studies on meteorites stand out in that they aimed to reveal the full spectrum of products, including non-proteinogenic amino acids. Remarkably, some are found in greater amounts than proteinogenic amino acids, and yet were not later linked to the genetic code (Figure 2).

At this point, it should be noted that several plausible scenarios for the chemical activation of the carboxyl group in amino acids under plausible prebiotic conditions have been described, which enabled the prebiotic formation of peptides.^[49] The reader is referred to several review articles.^[50–52] In principle, it is accepted that a prebiotic chemistry for the formation of peptides must have existed on planet Earth and it is reasonable to consider wet-dry cycle conditions^[53–55] as could have existed in hydrothermal fields.^[40]

Since amino acid metabolism is closely connected with the TCA cycle (tricarboxylic acid; see next section 3.1), it is important to note that experiments on the formation of intermediates of the TCA cycle have also been performed under putative prebiotic conditions, notably by Waddell^[56–58] and Krishnamurthy^[59] and coworkers. For example, citric acid can be generated from oxaloacetate under photolytic conditions, while 2-oxoglutarate and glutamate yield succinic acid as the major photolysis product.^[40]

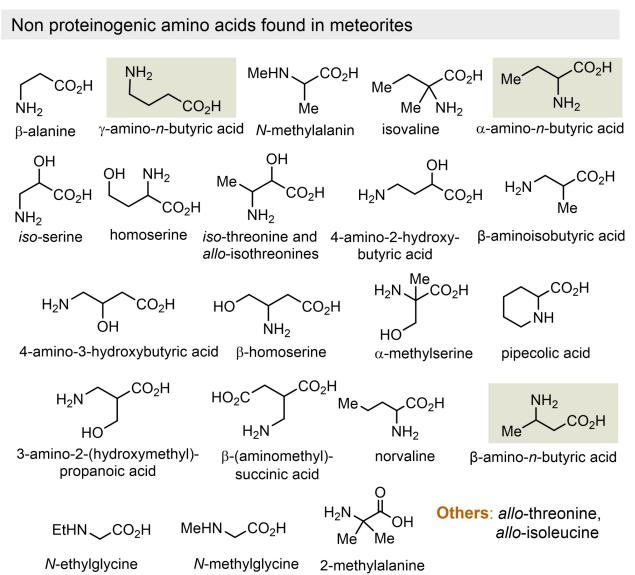


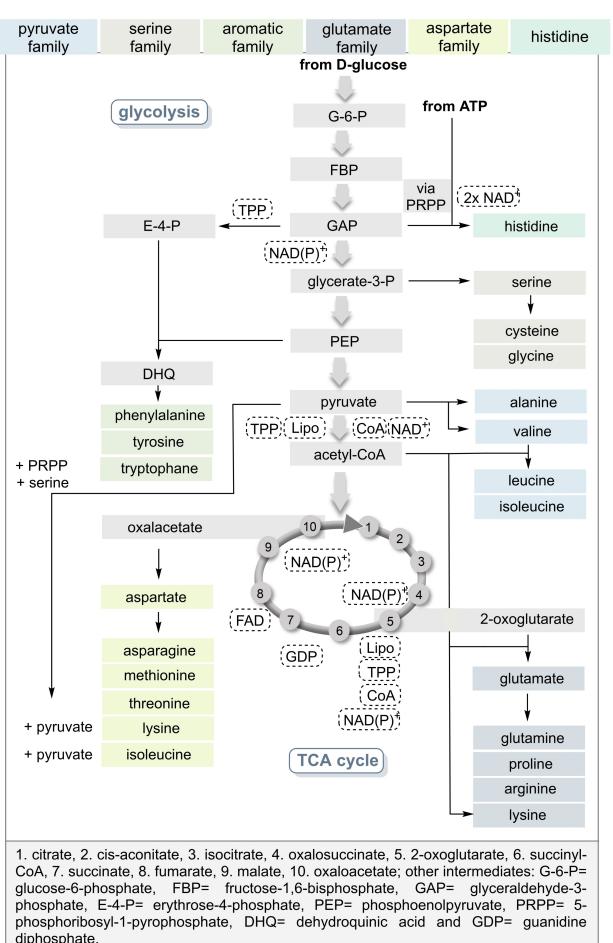
Figure 2. Selected non-coded amino acids found on meteorites and in experiments with electrical discharges mimicking astrochemical conditions (the most abundant amino acids related to the major product glycine are highlighted in grey); additional extraterrestrial organic molecules are listed in the cited literature.^[7]

3. The Biosynthetic Scenario

3.1. Overview on amino acid biosyntheses

The twenty proteinogenic amino acids are biosynthesized from carbohydrate-derived building blocks and these are recruited from glycolysis and the TCA cycle. Histidine differs from the other amino acids in that portions of its backbone originates from ATP; thus, it is the only amino acid directly linked with nucleotide and purine metabolism.^[60,61] Amino acids are typically grouped either based on these building blocks (pyruvate), or amino acids that serve as precursors for other amino acids (aspartate, glutamate, serine). The aromatic amino acids are grouped according to their common chemical structural feature (Scheme 1). The biosynthesis of all proteinogenic amino acids is closely intertwined with the need for coenzymes, which are required both to provide building blocks as well as for the subsequent individual biosyntheses.

At this point the reductive tricarboxylic acid cycle (rTCA cycle; sometimes also referred to as reverse TCA cycle) is worth mentioning.^[62,63] Typically found in bacteria and archaea, the



Scheme 1. Overview of amino acid biosynthesis linked with glycolysis and the TCA cycle. The coenzymes involved (outlined in dashed lines) are assigned to these two metabolic elements, while the coenzymes required for amino acid biosynthesis are listed below Table 2.

TCA cycle basically runs in reverse (Scheme 2). It provides organic building blocks from CO_2 ^[64,65] under highly reductive conditions with the electron donors H_2S or thiosulfate. The rTCA is not only considered ancient, but is also discussed as a plausible candidate for the first autotrophic metabolism with respect to the origin of life.^[66,67]

The two TCA cycles differ in several aspects, for example 2-oxoglutarate:ferredoxin oxidoreductase is a thiamine pyrophosphate (TPP)-dependent enzyme with an additional [4Fe–4S] cluster 18 that transfers CO_2 to succinyl CoA to yield 2-oxoglutarate and CoA (6→5, Scheme 2).^[68] The second step of CO_2 fixation (5→4) is promoted by 2-oxoglutarate carboxylase that uses the mixed anhydride after ATP activation. Interestingly, some facultative chemo-lithoautotrophic thermophiles can switch between both TCA cycles depending on available carbon sources.^[69]

Based on the assumption that the rTCA cycle has prebiotic significance for carbon fixation and that the intermediates oxaloacetate and 2-oxoglutarate have served as precursors for amino acids, a chemical approach to mimic this cycle using metal salts was recently carried out which was coupled with a reductive amination step (see Scheme 10).^[70,71]

3.2. Coenzymes

Coenzymes are small organic non-protein compounds that specifically bind to proteins and actively participate in biotransformations.^[72–75] They have been called vestiges from a prebiotic (RNA) world for several reasons:^[7,76] (1) some of them are structurally closely related to RNA building blocks as exemplified in the adenosine monophosphate handle (AMP, 11), (2) they have, unlike other biomolecules, not undergone structural changes, at least since the emergence of the last unified common ancestor (LUCA), and (3) they have more or less stuck to their primary role, which is to support catalysis. Coenzymes and cofactors can clearly serve as “checkpoints” to validate the plausibility of theories at any stage of molecular

evolution, and as discussed here, this includes the evolution of amino acids and proteins.

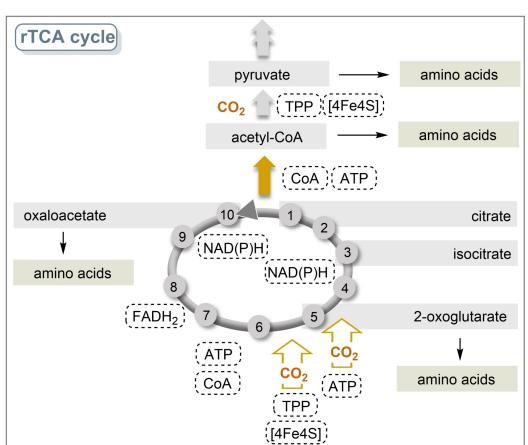
3.2.1. No amino acid biosynthesis without coenzymes

In previously published hypotheses on molecular and biotic evolution, coenzymes have been more or less disregarded. For glycolysis, ATP/ADP and nicotinamide (NAD(P)^+ 12) are required,^[77,78] while the TCA cycle relies on NAD(P)^+ 12, flavins (FAD 13) and guanosine diphosphate (GDP). Another important coenzyme is thiamine pyrophosphate (TPP, 16) which is necessary for the generation of erythrose-4-phosphate (E-4-P) from glyceraldehyde-3-phosphate (GAP 24).

This coenzyme is also found in α -oxoglutarate dehydrogenase and pyruvate dehydrogenase, the latter operates as an entry into the TCA cycle. In anaerobic microorganisms lipoic acid is replaced by ferredoxin 18 as found in pyruvate:ferredoxin oxidoreductase (PFOR), an enzyme also exists in the reverse rTCA cycle (Scheme 2).^[79] From an evolutionary point of view, it is an interesting fact that the pyruvate dehydrogenase complex also shows activity toward 2-oxoglutarate, the substrate of α -oxoglutarate dehydrogenase.^[80] The number of steps as well as the need of coenzymes/cofactors for each amino acid biosynthesis are listed in Table 2.^[81] Nature makes use of only a limited number of coenzymes, while with the exception of methionine, pterine-based coenzymes such as flavins 13 and tetrahydrofolic acid (THF 20), as well as biotin and S-adenosylmethionine, are not part of the list. It comes to a surprise that no metal-based cofactors are involved for the formation of amino acids as metals and metal-based cofactors are thought to have played a key role very early in evolution (see Scheme 5).^[82–84]

The redox coenzyme $\text{NAD(P)}^+/\text{NAD(P)H}$ 12 occurs in each of the known amino acid biosyntheses, with the exception of alanine, aspartate, and asparagine, suggesting that these nicotinamide-based coenzymes probably appeared around the time when the evolution of amino acid biosyntheses took place. A similar assumption can be made for the coenzyme pyridoxal phosphate (PLP, 19) and its amino derivative pyridoxamine phosphate (PLP*). The best documented^[7] and accomplished prebiotic syntheses exist for PLP 19 and nicotinamide 12 with structurally simplified nucleotide element.^[85–87] Pentoses and glyceraldehyde-3-phosphate serve as building blocks for the synthesis of PLP derivatives, while two routes are established to generate nicotinamides, with the biomimetic variant using dihydroxyacetone phosphate and aspartate as precursors.^[88] The latter can react with ribose-1,2-cyclic phosphate to give ribosylated pyridinium salts derived from NAD^+ 12.^[89]

Analysis of the biosynthesis of nicotinamides or PLP supports their ancient nature. Two different biosynthetic pathways are known for each of the two coenzymes. One of each is very simple, with respect to the number of coenzymes or cofactors required for their formation. In prokaryotes, NAD is usually biosynthesized from dihydroxyacetone phosphate and oxaloacetate or aspartate, with their imine derivative serving as the reactive intermediate, so that only ATP and the iron–sulfur

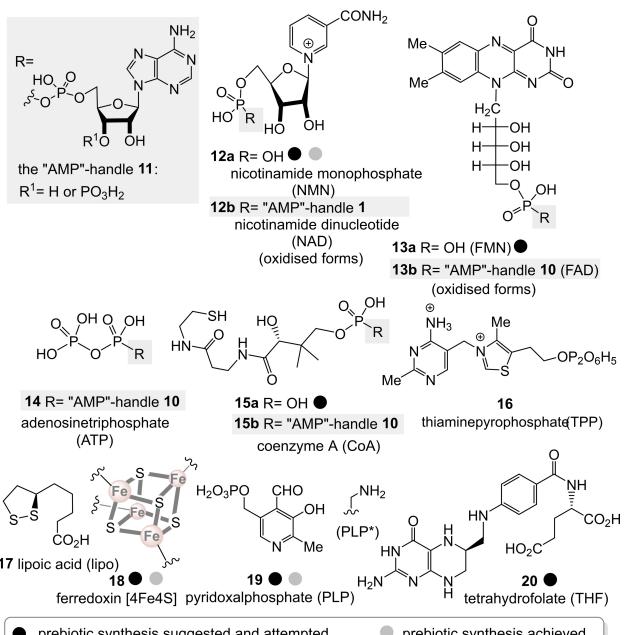


Scheme 2. Overview of reductive TCA cycle with coenzymes and cofactors involved (numbering of intermediates are taken from the TCA cycle in Scheme 1).

Table 2. Top: Biosyntheses of 20 coded amino acids with respect to number of steps and coenzymes required (the glutamate family does not require PLP* but 2-oxoglutarate undergoes reductive amination with NAD(P)H 12 instead). Bottom: Structures of the AMP “handle” 11, coenzymes 12–20 involved in glycolysis, the TCA cycle and amino acid biosynthesis. The list includes the ferredoxin 18 that can exchange lipoic acid 17.^[7]

pyruvate family ¹⁾ amino acid	serine family ²⁾ steps	aromatic family ³⁾ coenzymes	glutamate family ⁴⁾ amino acid	aspartate family ⁵⁾ steps	histidine ⁶⁾ coenzymes
Ala	1	PLP	Glu	1	NADH
Val	4	TPP, NADH, PLP	Gln	2	NADH, ATP
Leu	8	TPP, NADH, NAD ⁺ , PLP	Pro	4	3x NADH, ATP
Ile ^[a]	8	NAD ⁺ , TPP, NADH, PLP	Arg	9	2x NAD ⁺ , 2 ATP, PLP, acetyl-CoA, (carbamoyl-P)
Gly	4	NAD ⁺ , 2 PLP, (THF)	Asp	1	PLP
Ser	3	NAD ⁺ , PLP	Asn	2	PLP, ATP
Cys	5	NAD ⁺ , PLP, (acetyl-CoA)	Met	8	ATP, 2 NADH, 2 PLP, (N ⁵ -Me-THF), (succinyl-CoA)
Phe	10	PLP, NADH, ATP	Thr	6	2 PLP, 2 ATP, 2 NADH
Tyr	10	PLP, NADH, NAD ⁺ , ATP	Lys ^[a]	12	ATP, 3 PLP, 2 NADH, (succinyl-CoA)
Trp	13	PLP, NADH, ATP	His	10	2 NAD ⁺ , PLP

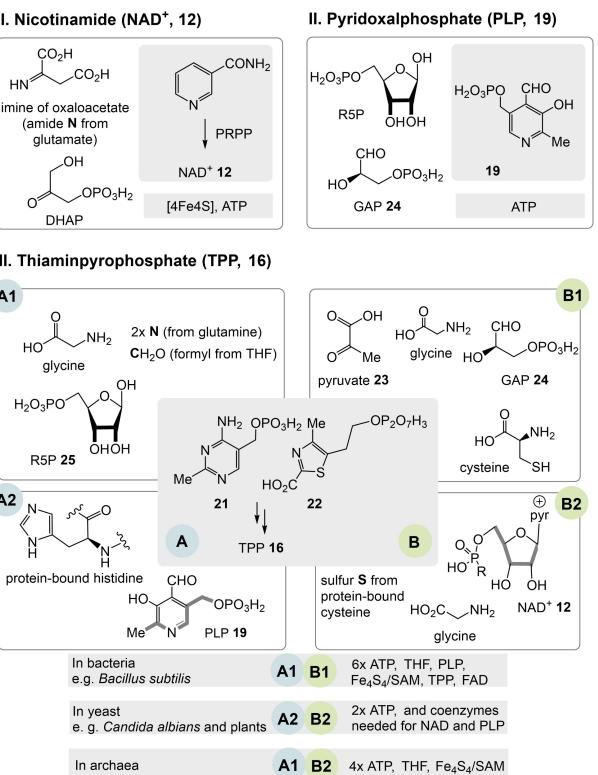
Step counts: 1) from pyruvate, 2) from 3-phosphoglycerate, 3) from E-4-P and PEP, 4) from 2-oxoglutarate, 5) from oxaloacetate, 6) from ATP.



[a] Ile and Lys can also be biosynthesized within the aspartate and glutamate families.

cofactor are involved (Scheme 3, case I).^[90–92] Remarkably, the [4Fe4S] cluster 18 does not act as a redox cofactor here but rather as a Lewis acid. The simpler of the two biosynthetic route towards PLP 19 starts from GAP 24 and ribose-5-phosphate and ATP 14 is the only coenzyme needed (Scheme 3, case II).^[93]

The biosynthesis of methionine requires THF 20 as a methyl-transfer acting coenzyme.^[94,95] In recent years, evidence has been accumulated that in anaerobic archaea and bacteria the iron–sulfur corrinoid protein functions as a methyl group



Scheme 3. Summary of NAD⁺ 12, PLP 19 and TPP 16 biosyntheses (for TPP three routes to the two key intermediates 21 and 22: A1–B1, A2–B2 and A1–B2).^[103]

transfer system (see Scheme 5) in methionine biosynthesis,^[96] indicating that THF and its variants likely appeared later.

3.2.2. Why is TPP a likely “latecomer”?

In the present context, TPP 16 is a particularly interesting coenzyme. It is directly involved in the biosynthesis of the three hydrophobic amino acids valine, leucine and isoleucine, the maintenance of the TCA cycle (dehydrogenase complexes; see above) and for the formation of erythrose-4-phosphate (E-4-P), a precursor of aromatic amino acids.^[97–99]

Nature has evolved three different biosynthetic pathways to TPP 16, all of which are terminated by a substitution reaction between hydroxymethyl–pyrimidine phosphate (HMP–P, 21) and hydroxyethylthiazole phosphate (HET–P, 22) and phosphorylation (Scheme 3, case III).^[100] Two independent biosyntheses were found for each of the two fragments, with either ribose 5-phosphate (R5P, 25) and glycine (A1) or alternatively protein-bound histidine and PLP 19 serving as precursors to access HMP–P 21 (A2). The components for the biosynthetic generation of HET–P 22 are either glycine, cysteine, pyruvate 23 and GAP 24 (B1) or alternatively glycine and the five C atoms of ribose found in NAD⁺ 12 (B2).

This analysis is eye-opening. Overall, three amino acids function as building blocks in the three TPP biosynthetic routes, namely glycine, cysteine and histidine. More significant, how-

ever, is the selection and function of the coenzymes required for the biosynthesis of fragments 21 and 22. First, bacterial biosynthesis B1 to 22 reveals that this route, requires TPP for its own generation (pyruvate 23 + GAP 24 yields 1-deoxy-xylulose-5-phosphate), making it an evolutionary irrelevant one.^[101]

Particularly surprising is the fact that the A2 and B2 biosynthetic subroutes use the coenzymes PLP and NAD⁺ as building blocks, a little-known role for these coenzymes. This would imply that these two sub-routes to TPP are evolutionarily younger than those of PLP 19 or NAD⁺ 12, so that biosynthetically TPP can be regarded to be a "latecomer".^[102] Consequently, this would mean that the known biosynthetic pathways to branched aliphatic α -amino acids and aromatic amino acids are also more recent.

3.3. Could amino acids biosynthesized without the coenzyme TPP 16?

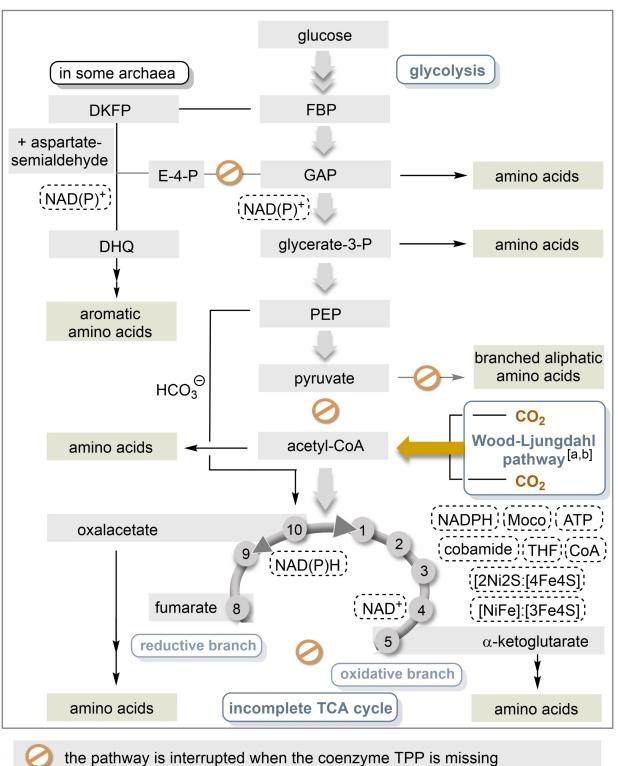
In view of the above discussion, it seems reasonable to perform a "Gedanken experiment" aimed at a TPP-free metabolism for the TCA cycle and the biosynthesis of aliphatic and aromatic amino acids. This "experiment" must ensure that the altered metabolism closely follows existing biosynthetic pathways, simply in the spirit of François Jacob, who pointed out that evolution behaves like a tinkerer and does not constantly invent new pathways.^[104] The execution of this "experiment" finds justification in today's world as it is suspected that *Borrelia burgdorferi* does not rely on TPP.^[105] This organism does not have the TCA cycle, does not perform oxidative phosphorylation, and there are no known pathways for de novo biosynthesis of carbohydrates and amino acids. But the exact metabolism that can fully explain the lack of thiamine-dependent enzymes is still unknown.

3.3.1. Glycolysis and TCA cycle

First we focus on the biotransformation of 2-oxoglutarate to succinyl-CoA (Scheme 1). Are there any older variants of the TCA cycle that do not use TPP 16? A non-cyclic putative evolutionary precursor of the TCA cycle, also known as the incomplete „horseshoe“ TCA, has been found in the strictly anaerobic bacterium *Elusimicrobium minutum*^[106] and the hyperthermophilic archaeum *Ignicoccus hospitalis* (Scheme 4).^[107]

Here, the link between 2-oxoglutarate and fumarate is interrupted, by absence of α -oxoglutarate dehydrogenase and succinate dehydrogenase, so that an oxidative half-cycle to α -oxoglutarate and a reductive half-cycle to fumarate remain intact. The reductive branch of the incomplete TCA cycle is initiated by the interconversion of oxaloacetate to malate, and fumarate.

As a consequence, the horseshoe TCA does not require TPP 16. In many anaerobes such as *E. minutum* it serves for the biosynthesis of various, but not all, proteinogenic amino acids.^[106] The organism is able to form glutamate, glutamine and proline (glutamate family), aspartate, cysteine and threo-



Scheme 4. Amino acid biosynthesis without the coenzyme TPP 16: The incomplete („horseshoe“) TCA cycle contains a reductive (via oxaloacetate) and an oxidative branch (via citrate). The Wood-Ljungdahl C1 fixation pathway provides acetyl-CoA without the need of TPP (DKFP = 6-deoxy-5-ketofructose-1-phosphate, amino acids are listed in Scheme 1). [a] The list of coenzymes and cofactors were established in the acetogen *moorella thermoacetica*; [b] for a prebiotic version of acetogenesis see Ref. [70, 71].

nine (aspartate family), glycine, alanine and serine (pyruvate family), as well as lysine and histidine. However, it relies on the external influx of arginine, asparagine and methionine, as well as the aromatic and alkyl-branched amino acids that, noticeably also depend on TPP 16 for their biosynthesis.

Pyruvate dehydrogenase is the second enzyme to be considered in more detail. Acetogens use the Wood-Ljungdahl or reductive acetyl-CoA pathway, a blueprint for a primal form of C1 fixation.^[107] It proceeds under strictly anaerobic conditions in which two equivalents of CO₂ are fixed and eventually converted to acetyl-CoA. It consists of two branches in which CO₂ is first reductively converted to formate (methyl branch) or to CO (carbonyl branch).^[108]

In the former, CO₂ reduction is brought about by the pterin-based molybdenum cofactor (Moco 28), supposedly an ancient redox system.^[109] In bacteria, the resulting formic acid is next activated by ATP 14 and transferred to the second pterin-derived coenzyme THF 20 (in archaea, it is the variant tetrahydromethylpterin (THMPT)), where the methyl group is formed via a series of reduction steps. From there it is finally transferred to the corrinoid cobamide. The other CO₂ reduction in the CO branch is catalyzed by carbon monoxide dehydrogenase (CODH), which contains NiFe centers 26. Subsequently, the metal-bound carbon monoxide is condensed with coenzyme A

and the methyl group donated by the cobamide eventually leads to acetyl-CoA, catalyzed by acetyl-CoA synthase (ACS; 27) (Scheme 5).^[110]

Considering that the “experiment” aims to circumvent the use of TPP, then the Wood-Ljungdahl pathway reveals some problems, since the need for pterin-derived coenzymes Moco, THF, and THMPT, respectively, creates a new roadblock because the pterin scaffolds likely appeared at a later point in time on the stage. Is it, then, that a prebiotically generated activated acetic acid was recruited instead?

In one case experiments to test the iron–sulfur hypothesis yielded S-methyl ethanethioate in the presence of nickel and iron sulfide.^[30,31] Similarly acetate and pyruvate 23 are formed from H₂ and CO₂ by the minerals greigite (Fe₃S₄), magnetite (Fe₃O₄) and awaruite (Ni₃Fe).^[111] This would imply that the prebiotic provision of building blocks fed the first self-

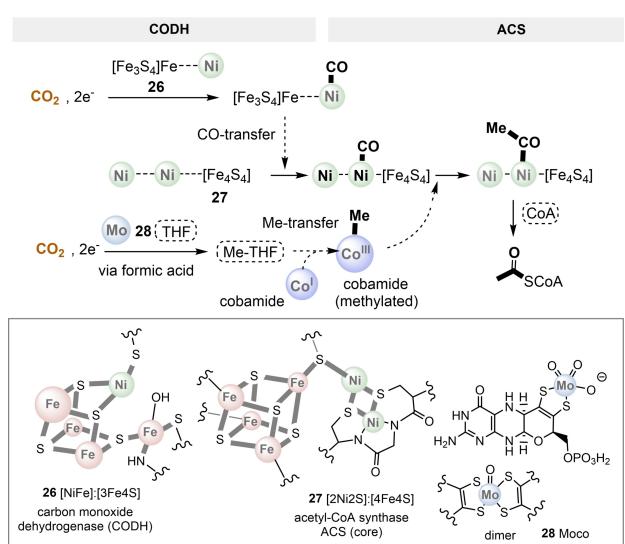
sustaining metabolic systems, and thus pre-metabolism and later metabolism were based on chemical continuity.^[112,113]

3.3.2. Aromatic and branched aliphatic amino acids

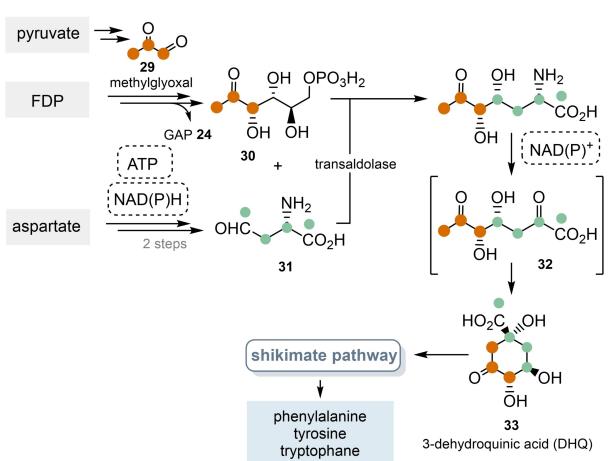
Next, we turn to the branched aliphatic as well as the aromatic amino acids. Erythrose-4-phosphate (E-4-P) and phosphoenol-pyruvate (PEP) act as entries into the shikimate pathway, relevant for aromatic amino acids and E-4-P is formed by a TPP-dependent transketolase.^[114]

An alternative, TPP-free entrance into the shikimate pathway was found in some archaea such as *Methanocaldococcus jannaschii*, where 6-deoxy-5-ketofructose-1-phosphate 30 (from pyruvate 23 via methylglyoxal 29) and L-aspartate semialdehyde 31 serve as gateway building blocks (Scheme 6).^[115–117] Their aldolase-mediated coupling followed by oxidative deamination leads to diketocarboxylic acid 32, which is converted to 3-dehydroquinic acid 33 by an intramolecular aldol reaction. Our knowledge of this biosynthetic pathway, however, is still too incomplete to make definite statements about its role in evolution.^[118]

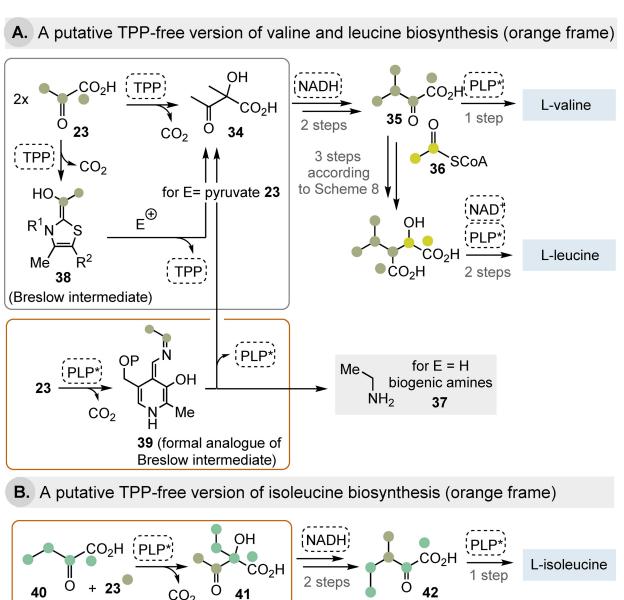
The biosyntheses of the branched aliphatic proteinogenic amino acids valine, leucine, and isoleucine also depend on the acyl anion transfer reagent TPP, starting from 2-oxocarboxylic acids, in case of valine and leucine this is pyruvate 23 and for isoleucine its 2-oxo-butyrate 40 (Scheme 7). The first important products of this acyl transfer process are 34 and 41, respectively and the Breslow intermediate 38 is the key intermediate of this process.^[119] It can be theoretically deduced that the coenzyme pyridoxamine phosphate (19, PLP*) should also be able to



Scheme 5. Structures of cofactors and their role in acetogenesis with CODH and ACS.



Scheme 6. Alternative biosynthetic pathway to DHQ 33 in the archaea *Methanocaldococcus jannaschii*.



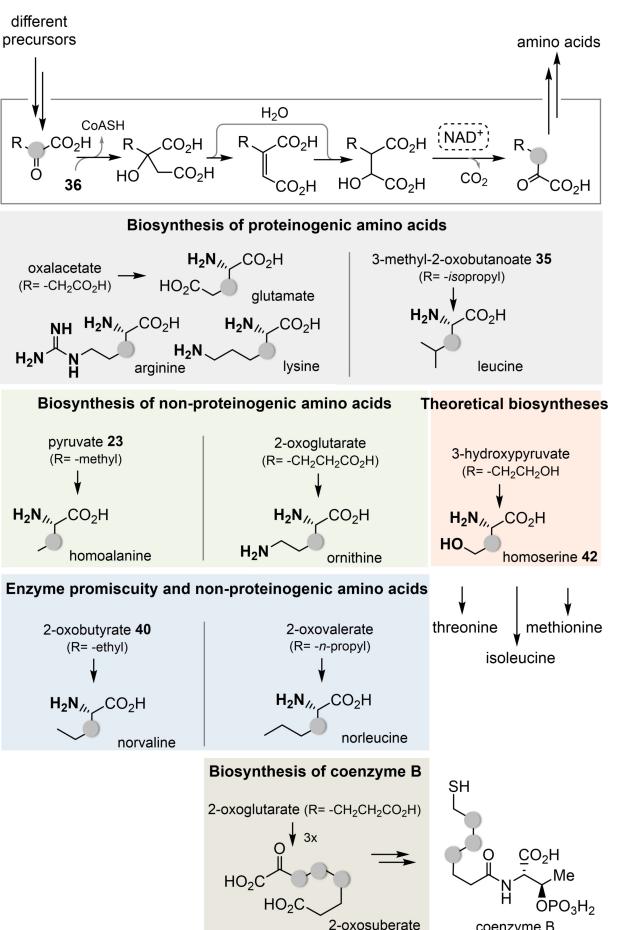
Scheme 7. A. TPP-mediated biosynthesis of valine and leucine and the hypothetical PLP-mediated analogous conversion (in orange frame) to intermediate 34; B. Analogous hypothetical reaction sequence to isoleucine from 2-oxobutyrate 40 via intermediate 41 (orange frame) and 2-oxocarboxylate 42 (PLP* = pyridoxamine phosphate).

mediate such an acyl “Umpolung” if one follows the mechanistically known PLP-mediated decarboxylation of amino acids to amines **37** (Scheme 7). Condensation of an 2-oxocarboxylic acid with PLP* produces an imine that can be in tautomeric equilibrium with a second imine. If, instead of a proton, pyruvate **23** acts as an electrophile, followed by tautomerization of the resulting imine and subsequent hydrolysis of PLP*, coupling products **34** and **41**, respectively, would be formed by an alternative pathway while making PLP* available for transaminations occurring later in the biosynthesis. The glycine cleavage system (GCS) provides an argument that a type **39** intermediate can be captured by electrophiles other than the proton, in the GCS case it is electrophilic sulfur in the oxidized form of lipoic acid.^[120]

Why has nature not realized this alternative, although PLP/PLP* **19** seems to be a simpler carrier of acyl anions compared to TPP **16** according to this mechanism. Is it an ancient, evolutionarily buried role of PLP and PLP*? Or has this role just not been found yet? Are there chemically or biosynthetically simpler approaches to TPP **16** or to simplified analogs unknown to us to date? Or are the three branched aliphatic amino acids perhaps indeed evolutionary “latecomers”? Chemically, the TPP-mediated conversion to **34** or **41** is expected to be more efficient than the conversion with PLP* proposed here, since the latter proceeds via several tautomeric imine intermediates, each prone to hydrolysis under aqueous conditions. So their may have been a quest for a more “robust” acyl anion methodology. This “experiment” shows that it is worthwhile to think about hidden chemical roles of coenzymes.^[121]

3.3. Do amino acids exist that could have been preferred over the coded ones?

The ancient TCA cycle is closely linked to amino acid biosynthesis;^[122] could it have been the starting point for other α -amino acids based on recurring patterns of chemical sequences? Such a sequence is found for 2-oxoglutarate, formed from oxaloacetate, which consists of an initial aldol-like addition of acetyl-CoA **36** followed by water elimination and re-addition of water with opposite regiocontrol. Next NAD⁺-mediated oxidation leads to the (often spontaneous) loss of the original carboxyl group. In essence, 2-oxocarboxylate is transformed into a homologous 2-oxocarboxylate by which a methylene group is formally inserted between the keto function and the substituent R (Scheme 8).^[123] This sequence is also found in the biosyntheses of glutamate, arginine and the non-proteinogenic amino acid ornithine (all from oxaloacetate), lysine (from 2-oxoglutarate) and leucine (from 3-methyl-2-oxobutanoic acid **35**). In a remarkable iterative process, 2-oxoglutarate is extended three times (via 2-oxoadipate and 2-oxopimelate) to 2-oxosuberate, en route to coenzyme B as well as in the biosynthesis of the coenzyme biotin (for coenzyme B see Scheme 8).^[124] Further theoretical examples, not realized by Nature, could be formulated for the proteinogenic amino acids threonine, isoleucine, and methionine, which are principally accessible from homoserine (**42**), which in turn could form from



Scheme 8. Nature's homologation concept of 2-oxocarboxylates (C atoms marked in grey highlight the homologation process).

serine or 3-hydroxypyruvate. However, homoserine is biosynthesized from aspartate via the semialdehyde **31**. The naturally occurring, non-proteinogenic amino acids ornithine and homoalanine are also formed from the aspartate/oxaloacetate and alanine/pyruvate couples,^[125] via this homologation sequence. The last precursor en route to homoalanine is α -oxobutyrate **40**, for which, however, two other biosynthetic alternatives from aspartate and threonine, respectively, are known.^[126]

This four-step homologation sequence is likely very ancient because a) only the coenzyme nicotinamide is required, b) it is found in the TCA cycle and its incomplete „horseshoe“ variant and c) in principle should be reproducible under prebiotic reaction conditions. This suggests the existence of ancient protoenzymes with lower substrate specificity (see Figure 3C).

Evidence for this hypothesis is still pending, as this would imply that norvaline and norleucine could theoretically have been included in the inventory of encoded hydrophobic amino acids. Both amino acids occur intracellularly, and the enzymes involved, such as isopropylmalate synthetase, are promiscuous and accept different 2-oxoacids (Scheme 7).^[127,128] In addition, some degree of promiscuity of leucyl-tRNA synthetase results in a mischarged norvalyl-tRNA^{Leu} that evades translational proof-

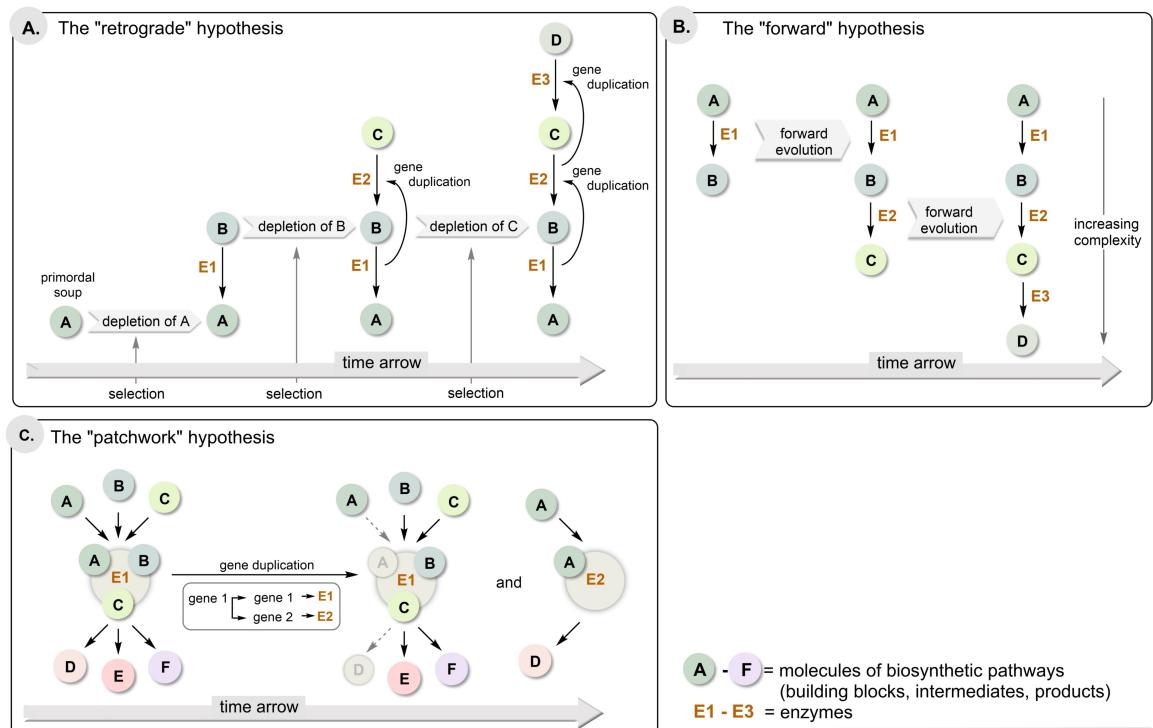


Figure 3. Hypotheses A-C of metabolic evolution.

reading activity and leads to norvaline-containing proteins.^[129] These observations suggest that these chemical ambiguities may have played a role in determining the genetic code (see section 5.4), but norvaline and norleucine were ultimately omitted for whatever reason.

Interestingly, the noncoding amino acids *N*-methylalanine, homoalanine, homoserine, and norvaline are on the list of non proteinogenic amino acids found on meteorites. The Miller-Urey experiments also yielded nonproteinogenic amino acids, including norvaline and norleucine suggesting that they must have been present during the transition to a coded world.^[130] The absence of these *n*-alkyl amino acids was justified by their structural similarity to the versatile proteinogenic amino acid methionine. From a biosynthetic point of view, however, this comparison makes little sense, since the biosynthesis for methionine is among the most complex (see Table 2) and probably appeared late on the stage. Surprisingly, homoalanine and ornithine were also not included into Nature's portfolio of encoded amino acids, although in terms of polarity, homoalanine can be placed between alanine and the branched aliphatic amino acids.

The evolutionary role of ornithine is quite peculiar. Since ornithine is the biosynthetic precursor of arginine, it was suspected that ornithine was initially encoded in the early phase of the genetic code, which was later replaced by arginine. However, the chemical reactivity of ornithine leaves it out of the privileged list, since in its activated form (e.g., when bound to tRNA) it readily cyclizes to the corresponding δ -lactam. And the arginine content of the proteins is much lower than

expected, considering that 6 of the 61 codons for amino acids in the genetic code are for arginine.^[131] Presumably, when the code was expanded, lysine was added with an amino side chain that filled the gap left by the omission of ornithine (see Section 5.3). This is also indicated by the fact that lysine is more abundant than one would expect based on its two codons.

3.4. Three “chicken and egg” problems need to be resolved

Considerations on the biosynthesis of amino acids, their possible timing in evolution, and the special role of coenzymes have obscured some fundamental dilemmas. The generation of the 20 encoded amino acids are catalyzed by enzymes, supported by the presence of coenzymes. Enzymes themselves consist of α -amino acids whose formation they catalyze, so we are facing with a case of causal circularity or, in more general parlance, a “chicken and egg” problem. This also applies to the pairs amino acids/coenzymes and proteins/coenzymes, since the latter are also biosynthesized by enzymes but are themselves essential for various steps in the biosynthesis of amino acids (see Scheme 3).^[132] Clearly, the emergence of a biotic world is inconceivable without the presence of abiotic molecules such as amino acids, small peptides and coenzymes/cofactors, even if they should only be considered in the broader context of a replicable RNA world,^[5,133] or viroid-like system^[134] or as part of a systems chemistry approach.^[135]

4. Hypotheses on Linking Metabolic Pathways with Prebiotic Chemistry

The emergence and refinement of basic biosynthetic pathways enabled primitive organisms to become increasingly independent of exogenous sources. With the in-depth coverage of both abiotic and biotic amino acid syntheses, the assumption can be made that incidental reactivities, both enzymatic and inherently chemical, provide the background against which the recruitment of individual enzymes greatly enhances the function of a slow but pre-existing multistep sequence.^[136] One such entire block of analogous enzyme sequences, was described in Scheme 8.

Why and in which way could finally a metabolism evolve? To this end, several elementary, forward in time hypotheses have been formulated in an attempt to rationalize the evolutionary dynamics that transitioned the prebiotic world into the biotic one.^[131] Drivers of metabolic evolution are thought to be duplication and divergence of genes and enzymes, kinetic optimization of established pathways through fusion of enzymes, and minimization of ATP unit cost, thereby improving thermodynamic efficiency.

4.1. The “retrograde” hypothesis

The “retrograde” theory of evolution, put forward by Horowitz^[137] states that the first living species was a completely heterotrophic organism (Figure 3A). It proliferated at the expense of prebiotically formed organic molecules, for example amino acids or molecules produced under prebiotic conditions with the properties of modern coenzymes and cofactors. The organism will then use up the environmental reserves of A and deplete it to a point where growth is limited. In such an environment, any organism that evolves an enzyme or catalytic system capable of synthesizing a molecule A from precursors B would have a clear selection advantage and would rapidly expand in the environment. This selection process could be repeated for subsequent “generations” until the completion of the biosynthetic pathway known today.

The theory additionally states that further evolution is likely to be based on a random combination of genes. Thus, the simultaneous unavailability of two intermediates (e.g. B and C) would favor a symbiotic association between two mutants, one capable of synthesizing B and the other capable of synthesizing C from other precursors in the environment, leading to the evolution of short reaction chains. Thus, the theory also incorporates the idea of parasitism and symbiosis as driving forces of evolution.^[138,139]

The theory contains aspects that have been critically commented upon. The evolution of metabolic pathways in the reverse direction requires particular environmental conditions in which prebiotic formed organic compounds and potential precursors accumulated, but their presence became depleted over time. In addition, the origin of many other anabolic metabolic pathways cannot be inferred from their backward

evolution because they involve many unstable intermediates. How could these have accumulated in prebiotic and present-day environments?

4.2. The “forward” hypothesis

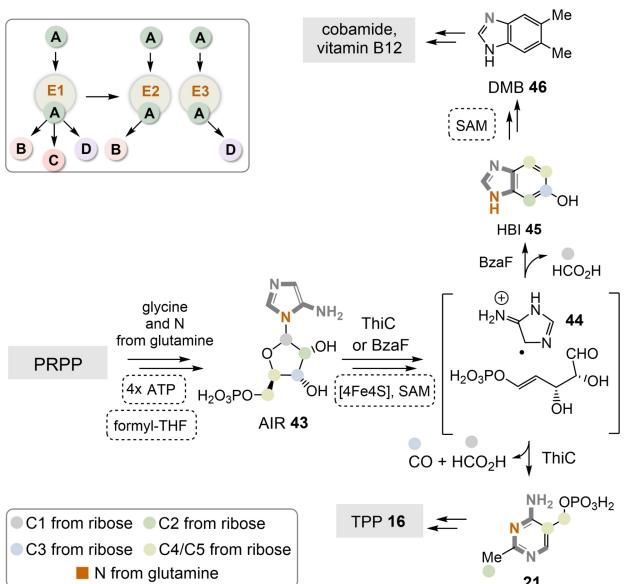
A lesser-known proposal negates the importance of prebiotic compounds in biological evolution.^[140] Here, biosynthesis of end products occurs by forward evolution from simpler precursor molecules (Figure 3B) and enzymes catalyzing earlier steps in a metabolic pathway are older than those acting later. Each intermediate in a biosynthetic pathway must therefore be useful to the organism, since simultaneous evolution of multiple genes in a sequence is too unlikely. The hypothesis seems to be questionable for complex, linear biosynthetic pathways such as those for the purines and the branched-chain amino acids, where the intermediates have no obvious benefit to the organism.^[141]

4.3. The “patchwork” hypothesis

The “patchwork” hypothesis emphasizes the role of primitive enzymes in the evolution of metabolic pathways.^[136,142] These early enzymes were able to react with a broad range of chemically related substrates (Figure 3C).^[143] The catalytic capabilities would still have been low in terms of “turnover”, but allowed metabolism for primitive cells with still small genomes. With the arrival of next-generation enzymes (E2), in which the amino acid sequence is slightly different, substrate specificity and catalytic activity increased. As defined, the “patchwork” concept could not become effective until after the appearance of protein biosynthesis and the rise of enzymes. There is some evidence that the “patchwork” hypothesis had some validity in evolutionary amino acid biosynthesis particularly for threonine, tryptophan, isoleucine, and methionine (three of them belong to the aspartate family).^[60] An example of the “patchwork” hypothesis is the homologation strategy discussed in Scheme 8. Comparative structural and functional analyses revealed that a small number of amino acid substitutions in the active site lead to paralogous proteins. They can recognize substrates with different aliphatic chain lengths as found in the biosynthesis of coenzyme B (CoB) and in the methanogen *Methanocaldococcus jannaschii*, the homocitrate synthase (HCS) is able to accept 2-oxoglutarate, 2-oxoadipate, and 2-oxopimelate.^[124]

The “patchwork” hypothesis can also be applied to the biosynthesis of TPP 16 and imidazole-containing 5,6-dimethylbenzimidazole (DMB, 46) the lower ligand in cobamide and vitamin B12 (Scheme 9).^[144,145] Comparative genomics on bacterial thiamine-pyrimidine synthase (ThiC) revealed the existence of a paralog of thiC, the HBI synthase (BzaF), which is clustered with anaerobic genes for vitamin B12 biosynthesis.

Both enzymes use the same substrate phosphoribosyl-aminoimidazole (AIR, 43) and promote quite different radical chain reactions induced by the coenzyme SAM and cofactor ferredoxin 18.^[146] In the case of TPP biosynthesis, the product is



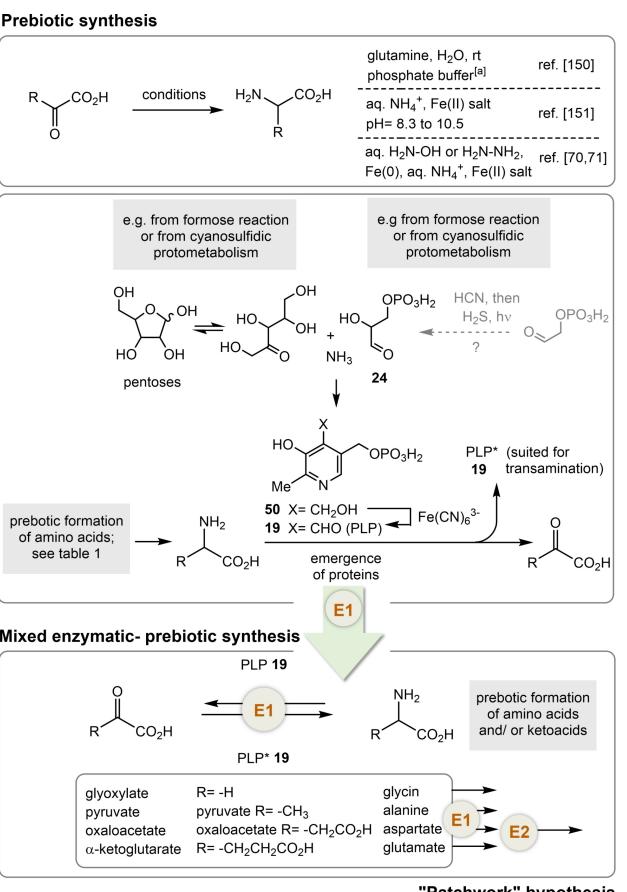
Scheme 9. TPP and DMB-biosynthesis, that can be linked to the “patchwork” hypothesis (Figure 3 C.) (PRPP = 5-phosphoribosyl-1-pyrophosphate).

hydroxymethyl-pyrimidine phosphate (HMP-P, 21)^[101] while in the DMB branch, 5-hydroxybenzimidazole (HBI, 45) is formed first.^[147] The sequence and structural similarities of the two enzymes is reflected in the proposed mechanisms. The key chemical driver for the formation of the subsequent intermediates is the radical cation 44 from where the two routes separate. Subtle changes in the amino acid sequence of the protein determine which of the two completely different products with distinct biological functions is formed. The progenitor protein E1 was still likely to have generated several products B–D from A (Figure 3 C.) as it was unable to control the high reactivity of the intermediate radicals. In the further course, this protein evolved into the two related enzymes E2 and E3, which selectively enabled the formation of the two most important products B and D.

4.4. Mixed origin of metabolic pathways

Later the “patchwork” hypothesis was extended by including prebiotic chemistry and combining it with the appearance of the first enzymes.^[113,148] It was assumed that prebiotically generated molecules should be chemically quite stable. These were complemented by molecules derived from existing metabolic pathways in cells for which stability was not a mandatory requirement. The expansion of the metabolic repertoire should have occurred by gene duplication and should have produced non-specific catalysts. Supposedly, these early proteins were formed by non-enzymatic reactions.^[149,21c]

The so called mixed-origin approach can be exemplified for the biomimetic trans- or reductive amination of 2-oxo-carboxylic acids (Scheme 10). The non-enzymatic transamination of glyoxylate with glutamine as amino donor is the earliest



Scheme 10. Merging prebiotic chemistry with concepts of metabolic evolution exemplified for transamination of 2-oxocarboxylic acids (E₁, E₂ = enzymes).

example of such a prebiotic study and the reversal of transamination namely from glycine to glyoxylate in the presence of formaldehyde has also been reported.^[150,151] However, this reaction exhibits an unfavorable equilibrium for the formation of imines in the aqueous medium. It can be circumvented by using an stoichiometric amount of hydrazine (or hydroxylamine) and added metallic iron to create a reductive environment.^[70,71]

En route to the biotic world, a possible prebiotic (biomimetic)^[152] generation of PLP 19 pushed the system, because it is known that coenzymes alone can drive their chemistry (Scheme 10, top).^[7] The transamination chemistry became more efficient with the emergence of the first enzyme E1, a transaminase capable of binding PLP, but with low substrate specificity. According to the “patchwork” hypothesis, the promiscuous character gradually disappeared after the formation of next-generation enzymes, promoted by gene duplication (e.g. E2). In fact, PLP-dependent transaminases are known for their broad specificity.^[153]

4.5. "Shell" hypothesis

As a final example, the so-called "shell" hypothesis is discussed, which specifically addresses the reductive citric acid (rTCA) cycle. According to this hypothesis, it should have led to an "energy amphiphile" core, being the starting point for the formation of new molecules. These propagate in a network-like manner and constitute molecular shells over previously formed metabolic nuclei (Figure 4, top).^[154a] This hypothesis assumes that the prebiotic chemical processes are "imprinted" on modern metabolism as relics.^[155] Accordingly, metabolic biogenesis manifested itself in a hierarchy of nested reaction networks of increasing complexity.^[154] It starts with shell A, which includes glycolysis and fatty acid biosynthesis.^[156] This was superseded by the introduction of nitrogen originated from amino acids in shell B and eventually sulfur in shell C. Consequently, purines, pyrimidines, and many other cofactors or coenzymes formed, as evolutionary "latecomers". In parallel, the rTCA underwent a change via the transition of the bidirectional to finally the TCA cycle. In this scenario, current enzymes are replaced by naturally occurring minerals or metal ions – a concept that has also been proposed for other metabolic cycles and networks.^[29,157,158]

However, one can consider a modified account by taking into account the close evolutionary relationship between amino acid metabolism and the availability of selected coenzymes (Table 2)^[131] and combining this with the horseshoe TCA cycle, the ancient form of the TCA cycle (see Scheme 4), which would substantially modify Morowitz's hypothesis (Figure 4, bottom). NADH-dependent reductions and PLP-mediated transamina-

tions are central to the metabolism of most amino acids. The simple amino acids glycine,^[159] alanine and aspartate are synthesized from the corresponding 2-oxocarboxylic acids in step by transamination, first chemically (see Scheme 10, top) and later after its arrival by the coenzyme PLP/PLP* **19** (Scheme 10, bottom). Elements of the original "shell" hypothesis found support when the rTCA cycle was recently reactivated under putative prebiotic conditions using metal salt promoters.^[70,71]

The idea of primordial metabolic cycles has generated controversial debates.^[160] Orgel pointed out that abiotic reactions proceed at low yields, i.e., the more reaction steps that occur in linear succession, the more catastrophically the overall yield decreases. This is a particularly fatal problem in metabolic cycles because the substrate concentration for the first step depends entirely on the yields of the following steps.

5. Retro-Bioanalytical Approaches to the Origin of Amino Acid Metabolism

5.1. The last unified common ancestor (LUCA)

The transition from the abiotic to the biotic world can also be analyzed in the inverse direction, starting from today's amino acid metabolism and its coding at the genetic level. Among other early "life forms",^[134] LUCA, a central theoretical model from which were the three kingdoms of Life (bacteria, archaea, and eucarya) evolved, needs to be covered. Using the mapped genomic diversity of the biotic world,^[161] a comprehensive phylogenetic reconstruction of the metabolic abilities of LUCA was reported by Martin et al.^[161b] The analysis suggests that LUCA was an autotrophic, thermophilic, anaerobic prokaryote, living in hydrothermal vents, relying on the Wood-Ljungdahl pathway (see Scheme 5).

Consequently, LUCA must have already used the entire alphabet of the 20 encoded amino acids as well as most of the coenzymes and cofactors. Thus, LUCA cannot resolve the problems of causal circularity mentioned in Section 3.4 but must have arisen long after the appearance of the first life forms.^[162,163]

5.2. The chemical space and physicochemical properties of proteinogenic amino acids

An evaluative approach employs the analysis of the chemical space filled by the various functionalized side chains of proteinogenic amino acids as well as their ability to form polar interactions with the surrounding medium.^[164] Recently published *in silico* studies on hypothetical peptide sets were performed with amino acid compositions comprising 3–19 amino acids, and a total of 1913 structurally distinct α -amino acids were included in the repertoire. The adaptive value of their combined physicochemical properties compared to those of the modern set consisting of the known twenty amino acids

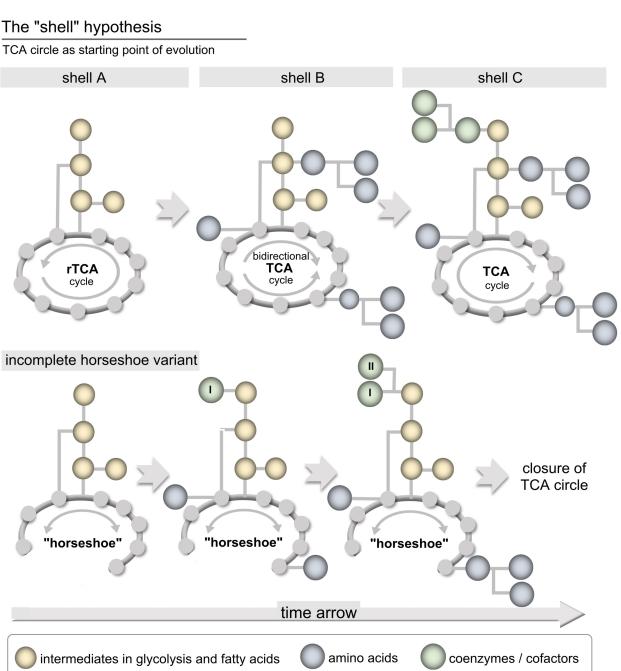


Figure 4. Morowitz "shell" hypothesis (top) and adaption when the incomplete horseshoe variant would be included (bottom). Coenzymes labeled with numbers I and II appeared early in evolution, for example NAD^+ and PLP.

were determined. As a major result it was found that such hypothetical sets, which included the encoded amino acids, are particularly adaptive.

It was concluded that each time a coded amino acid appeared on the scene, it provided adaptive value. Each selection step could have helped to expand the evolving set of amino acids, leading to an increase in the number of encoded amino acids in the emerging alphabet. Why then did life stop exploring chemical space further with the 20 encoded amino acids? It was suggested that property space was already well-explored at this stage. As soon as the evolving organisms acquired one or more amino acids from the modern alphabet, their abilities improved, so that natural selection becomes visible already at this molecular level.

Supposedly “latecomers” such as the aromatic amino acids phenylalanine, tyrosine, and tryptophan, but also cysteine, exert a strong influence on the structural rigidity of peptides and proteins. The starting point of another study was therefore the replacement of the aromatic amino acids from an enzyme, specifically dephospho coenzyme kinase, by lysine.^[165] One finding was that the two-step process catalyzed by the enzyme was not completely deactivated, particularly the second step. This presumed decoupling of activity and structure was taken as an indication that early life could get along for a time without highly structured proteins. Only in the further course of evolution did the repertoire of amino acids complete itself with more strongly structuring amino acids.^[166]

The hypothesis that active proteins with a much smaller repertoire of amino acids are conceivable has also been tackled by Akanuma et al.^[167] A folded, soluble and stable supposedly ancient nucleoside kinase with reduced catalytic activity, can arise from an alphabet of only 13 amino acids, that lacks, for example, Cys, Phe, Ile, Met, Gln, Thr, and Trp.^[168] The authors selected the remaining thirteen amino acids based on the assumption that LUCA was likely a thermophilic organism that required thermostable primordial proteins.^[169] These conceptual *de novo* strategies form the basis for the emergence of an early protein world and provide thoughts on the origin of the genetic code.

5.3. Histidine and cationic amino acids

Among the 20 encoded amino acids histidine stands out in many respects. It has been most often discussed in terms of a retro-bioanalytical rational to explain the selection of the twenty encoded amino acids and the emergence of the genetic code, despite the fact that so far, no convincing conditions for the abiotic synthesis of this amino acid have been found.^[43,170] Moreover, the presence of imidazole-containing purines in meteorites is well documented, but histidine is conspicuously absent in carbonaceous chondrites (see Table 1).

Amino acids with positively charged side chains, particularly lysine, arginine, and histidine, are considered key players in the co-evolution between early proteins and early nucleic acids because of their ability to stabilize RNA by acting as an early chaperone and their role in expanding the catalytic repertoire

of the molecular world at that time (RNA world theory).^[5] It was argued that in the biotic world, the imidazole side chain often acts as a catalytic site in enzyme reactions by mediating general acid-base catalysis in evolutionarily highly conserved catalytic triads.^[171]

The biosynthesis of histidine is unusual in that it is the only one that uses a nucleotide as the starting building block. If the currently favored RNA-world theory is correct, then this would consequently induce a direct connection of histidine to the purines, as is also the case for so many coenzymes.^[7,76] Despite the fact that the biosynthesis is quite long and consists of ten linear steps, many of which are hydrolytic or condensation reactions, only the simple coenzymes pyridoxamine (PLP*, 19) and NAD⁺ 12 are actually needed. In all histidine-synthesizing organisms, the pathway is unbranched consisting of nine intermediates and of eight distinct proteins, encoded by eight genes. Analysis of the structure and organization as well as the phylogenetic analyses of the *his* genes, which included the role of gene fusions,^[172] suggest that histidine biosynthesis must be very ancient and that it may have existed already well before the emergence of LUCA.^[173]

These analyses focused primarily on the *hisA* and *hisF* genes because they provide partial evidence for the “retrograde” hypothesis on the origin and evolution of metabolic pathways (Figure 5). The genes are paralogous and are arranged in tandem in the same operon and the two associated proteins catalyze two sequential steps in the same biosynthetic pathway. Histidine may thus play a key role in elucidating the transition from abiotic to biotic evolution and how the RNA world was fused with the emerging world of amino acids, proteins and coenzymes.

Finally, if cationic amino acids were indeed absent in a primitive RNA world, complexes of negatively charged amino acid residues, such as those found in aspartate and glutamate, with divalent metal cations (e.g. Mg²⁺, Fe²⁺) may have taken over the role of basic amino acids in primitive prebiotic peptides, it has been speculated.^[174] It was also found that removal of divalent Mg and its replacement by the prebiotic metal iron(II) leads to an expansion of the catalytic repertoire of RNA.

An interesting approach to elucidate the currently strongly favored theory of coevolution of peptides and nucleic acids was achieved by cationic protopeptides.^[175] In particular depsipeptides and polyesters can interact directly with RNA, leading to mutual stabilization. Interestingly, RNA is able to increase the lifetime of such protopeptides.

5.4. The genetic code as guideline

5.4.1. Crick's “frozen accident” theory

Francis Crick published the first thoughts on the selection of 20 encoded amino acids, called the “frozen accident theory”. It states that it would have been just as possible for another group of (originally abiotically formed) 20 amino acids to cross the “finish line” first, but after the first representatives had been

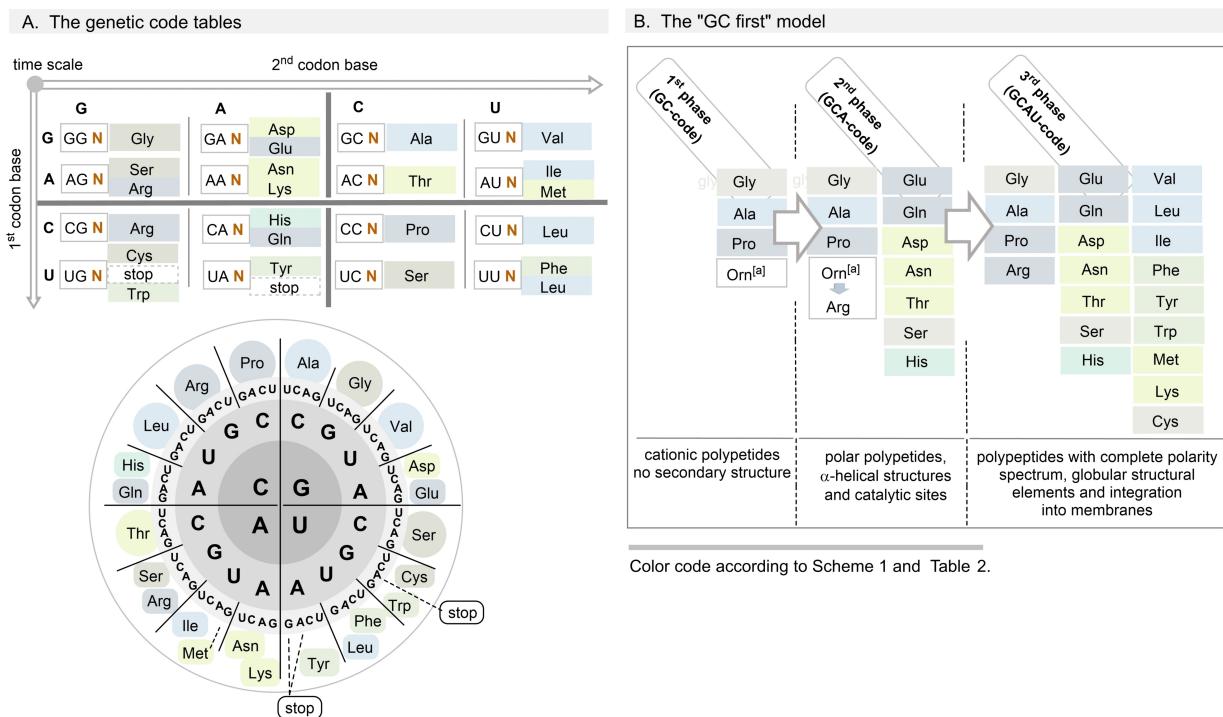


Figure 5. A. The genetic code table (adopted and slightly modified from Ref. [15, 9a]). Thick lines divide the table into quadrants between which a transversion mutation is required to change the encoded amino acid; (orange N = 3rd nucleotide exchanged from G via A, C and U). **B.** Evolution of the amino acid repertoire in protein biosynthesis according to references [188] and [193] ("GC first").

admitted, the door to the exclusive club was closed to the other amino acids present.^[176] According to this theory, the process of freezing took place when life had developed a certain degree of complexity. Due to its high evolutionary potential, this system was successful in competition with all other systems and has therefore survived as the only universal system until today. Before that time, changes in protein sequences resulting from changes in the code could thus be tolerated.

However, Crick's model does not explain why some codons can be grouped according to the physicochemical properties of their respective amino acids.^[177] It seems likely that amino acids were not chosen primarily for their ability to support catalysis in the form of their oligomers, since (metal) complexes with highly effective cofactors for carrying out diverse reactions have likely already existed.^[5]

Would more rational ties between amino acids and the formation of the genetic code be conceivable instead?^[178] If the discussions about the origin of life yielded a considerable variety of different proposals and hypotheses (besides the RNA world hypothesis,^[5] also the thioester world proposed by De Duve,^[179] the peptide model of Commeyras,^[180] the sugar model of Weber^[181] and others^[133]), then the picture of the formation of the genetic code discussed until today appears equally confusing as is shown below.^[182]

5.4.2. The "UA first" model

Based on the assumption that RNA must have preceded DNA, a primordial genetic code beginning with U and A was proposed.^[183] Accordingly, codons for Lys, Met, Ile, Asn, Tyr, Leu, Phe, and a stop codon arose in the first evolutionary phase. However, most of these amino acids are based on many biosynthetic steps requiring many coenzymes, and therefore can be considered rather evolutionary latecomers from a metabolic point of view. For several valid reasons, the alternative starting point with G and C as the initial letters of the prehistoric code has found greater popularity.^[184] In this case, it is not so much the biosyntheses of the nucleic acid building blocks that are used for primary analysis, but those of the amino acids.

5.4.3. The "glycine first" model

Among the 20 canonical amino acids of the biological coding system, the amino acid glycine is one of the most abundant in prebiotic experiments (see Table 1). Bernhardt, Patrick as well Tamura assumed that coding triplets are point mutated, and glycine (codon: GG N) served as the starting point.^[159] This was followed by serine, aspartic acid and/or glutamic acid – small hydrophilic amino acids. At this stage, this would have given rise to short, water-soluble peptides. Evolution of the code is thought to have occurred by duplication and mutation of tRNA

sequences, resulting in a radiation of codon assignment from the top left corner outward (Figure 5 A). In this way, small hydrophobic peptides or mixed peptides would have been added. Gradually, longer polypeptides would have formed that contained a hydrophobic core for folding and stability.

For Francis and others^[185] the genetic code table also served as a starting point and they came to a modified conclusion. Based on quantitative analysis of abiotically produced amino acids, it was concluded that the original genetic code consisted of only the four GNC triplets encoding for Gly, Asp/Glu, Ala, and Val (top row, Figure 5 A).

5.4.4. The "GC first" model

Hartman and Smith proposed that the first set of amino acids should be linked to the GC code, so that the coding stretch consisted only of the nucleobases guanine and cytosine (Figure 5 B).^[186] The "GC first" model, in which no amino acid but these two nucleotides act as conceptual starting point, overlaps strongly with the "glycine first" model, since glycine is encoded by the four triplets GG N. As a consequence, glycine, alanine, proline, and an amino acid with a positively charged side chain were the first for template-controlled peptide biosynthesis.^[186] Budisa and co-workers proposed ornithine (see section 3.3) which in time was later exchanged for arginine (Figure 5 B).

This hypothesis follows biochemical assumptions and this initial selection therefore focuses on the interaction and eventual stabilization of polyanionic RNA by the resulting early peptides. It enabled the later co-evolution of peptide sequences and the translation apparatus. A crucial aspect that the authors included in their considerations is the necessity of protein folding in the context of the "RNA world theory". It could have started with elongated and rigid peptides, with proline serving as the key amino acid. The first repertoire had to include positively charged amino acids that could dock to negatively charged RNA bodies. In addition, alanine as the simplest chiral amino acid supplemented the initial list of coded amino acids. The next evolutionary step in the code was the purine base adenine, which established a GCA code by allowing access of additional, now polar, amino acids, namely aspartate and glutamate, asparagine and glutamine, threonine, serine, and histidine.

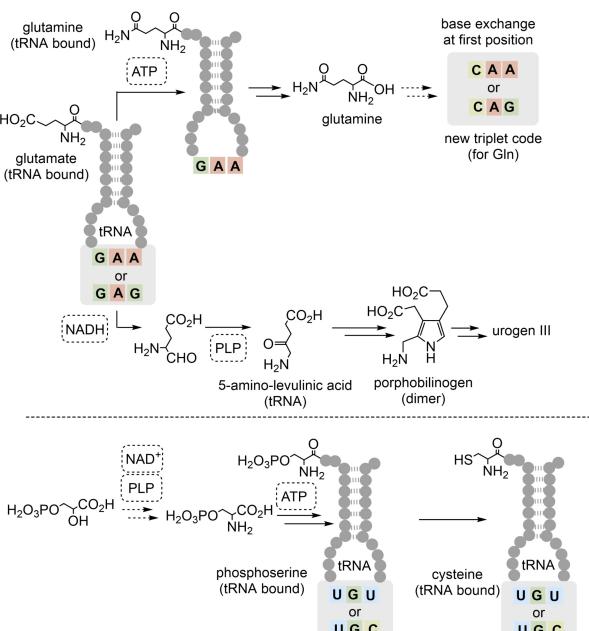
Finally, uracil and consequently the hydrophobic amino acids methionine, leucine, isoleucine, valine, phenylalanine, tyrosine, tryptophan, cysteine and lysine came into play many of which are formed via complex biosynthetic pathways. Based on these considerations, it was possible to develop a stepwise model correlating the gradual recruitment of the number of encoded amino acids with the hierarchy of protein folding, that eventually yielded α -helical structures, and the establishment of protein tertiary structures.^[187]

Recently, Westhof et al. presented an alternative circular representation of the genetic code table, choosing an asymmetric distribution of codons (Figure 5A, bottom).^[188] In it, there is a clear separation between GC-rich 4-codon boxes and AU-

rich 2:2-codon and 3:1-codon boxes. Within this presentation, the multiplicity and complexity of nucleotide modifications, particularly in the anticodon loop, are usefully separated. They correlate well with the need to stabilize AU-rich codon-anticodon pairs. Westhof also regards the GC pair as the starting point, which is gradually extended by A/U together with tRNA modifications and the modification of enzymes (see below: chapter 5.4.7 and Scheme 11).

5.4.5. The RNY hypothesis

Another concept, primarily proposed by Shepherd, states that primitive tRNA translations were based on the repetition of coding RNY triplets. These are remnants of an ancient code that is still detectable today, although it has been largely replaced by the present universal code.^[189] R stands for purines, Y for pyrimidines, while N can be either a purine or pyrimidine. In fact, these triplets outnumber RNR triplets fourfold in genes for extant proteins. The excess of RNY codons is likely due to a preference for the corresponding tRNAs rather than the remnants of an ancestral genetic code. RNY triplets encode Asn, Ser, Thr, Ile, Asp, Gly, Ala, and Val, with no basic amino acids among them. It has been postulated that once the primordial genetic code reached the RNY code, the elimination of any amino acid was strongly inhibited at this stage and therefore the genetic code was already to some extent frozen.^[190] But the necessity of the listed amino acids to stabilize RNA in a prebiotic world pose a problem for the validity of the hypothesis.^[191]



Scheme 11. tRNA-linked amino acid biosynthesis of glutamine and 5-amino-levulinic acid in the thermophilic archaeon *Methanopyrus kandleri* (Mka) (top); tRNA-linked amino acid biosynthesis of cysteine found in *Methanobacteriales* species (bottom).

5.4.6. The theory of co-evolution of the genetic code

The theory of co-evolution of the genetic code postulates that prebiotic synthesis did not provide all twenty amino acids and therefore some of them had to come from the co-evolving amino acid biosynthetic pathways.^[192] The addition of new amino acids increased protein versatility, improving the trajectory for perfecting catalytic capabilities in some enzymes and reducing combined transcription and translation errors to the <0.0003 range.^[193] Finally, against a background of decreasing errors, the noise introduced by the insertion of an additional amino acid with complete codon assignment would result in an overly large selective disadvantage that far outweighed the advantage of a new amino acid side chain. Accordingly, the expansion of the code to include additional amino acids with full codon assignment had to come to a halt. In contrast, the argument has been advanced that the number of encoded amino acids may have been greater due to the plethora of amino acids available from the chemical environment.^[194] The number of amino acids may also have varied at different times until the selection process was complete according to Crick's "frozen accident" theorem. This argument is supported by the analysis of some contemporary proteins composed of less than 20 different amino acids and by the experiments described in section 5.2.^[195]

5.4.7. Biosynthetic steps on tRNA-linked amino acids

Another feature that may have come into play at a later stage in the evolution of the genetic code is tRNA-linked amino acid biosynthesis. These include hydrolytically sensitive asparagine and glutamine. Most bacteria and all archaea arrive at Gln–tRNAGln by the transamidation pathway (Scheme 11)^[196] which could then have led to the expansion of the genetic code (for glutamine: CAA and CAG after replacing the first base G with C).^[197]

Biosynthesis with tRNA-bound glutamate is still found today, for example in the thermophilic archaeon *Methanopyrus kandleri* (Mka). Glu–tRNA is thereby converted to semialdehyde by Glu–tRNA reductase.^[198] From there a PLP-dependent biotransformation leads to 5-amino-levulinic acid and hence to tetrapyrroles and urogen III a, biosynthetic precursor of heme.^[199]

Additional examples of tRNA-linked amino acid biosynthesis are found exclusively in methanogenic archaea of the *Methanobacteriales* species, there especially for the formation of cysteine. Starting from D-3-phosphoglycerate, O-phospho-L-serine is first biosynthesized, which is loaded onto tRNA^{Cys} by ATP activation and then converted to tRNA-bound cysteine at this site.^[200]

With this perspective in mind, it was postulated that some early biosynthetic steps on tRNA-linked amino acids existed in the pre-LUCA era and that these tRNA-associated biotransformations were then replaced by today's tRNA-independent biosyntheses. Consequently, these disappeared from organisms living today. This hypothesis does not include the fact that coenzymes

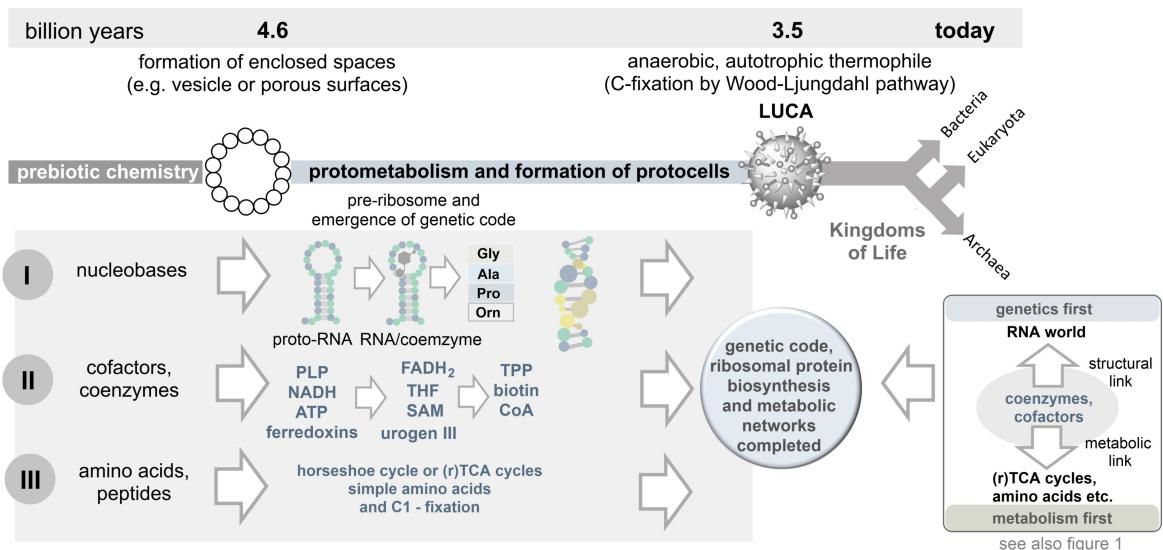
had to be present, such as NADPH 12 and PLP 19 into consideration that are involved in heme biosynthesis, a cofactor. Therefore, it is to suggest that the timing of coenzyme appearance should also be included into genetic code evolution.^[201]

5.4.8. Plausibility check exemplified by the "GC first" model

Even though the number of hypotheses and theories about the origin of the genetic code is even larger than the ideas presented here, there is still no reliable solution to the riddle. Are the criteria used to determine which genetic code is suitable for the primitive genetic code inadequate? The RNY code, the AU code, and the GC code model are mainly based on the codon pattern of the genes existing today, the RNA world theory of the origin of life and nucleotide metabolism, the stability of the RNA secondary structure, and the simplicity of the code compared to the universal genetic code, respectively. Prebiotic amino acid chemistry, the evolution of early amino acid metabolism, the transition or "hybridization" between the two, and the occurrence and role of coenzymes such as PLP 19 or other chemical mediators and catalysts have possessed a marginal prominence. Therefore, the different scenarios show some weaknesses. This can be illustrated by the choice of the first four amino acids in the "GC-first" model (section 5.4.4 and Figure 5 B). Serine appears in the second phase, while glycine in this scenario is at the very beginning of the evolution of the genetic code (see also "glycine first" model). This contradicts the biosynthesis known today, which starts from glycerate-3-phosphate and biotransformation to serine, which in turn serves as a precursor for glycine, a reversal of the relationship with the said model. It therefore can be argued that glycine had to be recruited from the prebiotic environment. The second proposed early amino acid is alanine, which is formed in one step by transamination from pyruvate 23, is a metabolically reasonable choice. The third amino acid is proline, which is biosynthetically formed from glutamate in four steps (Table 2). However, glutamate comes into play later in phase two of the "GC first" model, again a chronological logic that does not quite follow the timing from a biosynthetic perspective. One may argue that in *Methanococcus jannaschii* ornithine serves as precursor in proline biosynthesis with NAD⁺ being required as coenzyme.^[202] Although ornithine is named as one of the first amino acids encoded in the GC first theory (see Figure 5B), the fundamental problem persists because ornithine is also biosynthesized from glutamate.

6. Summary and Outlook

The present article reveals that the molecular evolution of amino acids up to the genetic code could not have succeeded in isolation without also considering those of nucleic acids (although not discussed in detail here) and of coenzymes and cofactors. These three lineages must have arisen at an early stage of evolution, not without influence from each other, but



Scheme 12. Three parallel lineages of molecular and metabolic evolution: nucleic acids (I), coenzymes/cofactors (II) and proteins and metabolic networks (III) – a hypothetical model as well as structural and functional relationships of coenzymes/cofactors to the “genetics first” and “metabolism first” theories.

in direct interplay and mutual influence, as outlined in Scheme 12.^[203] In particular, the two transitional phases from the prebiotic world to the first premetabolic systems and protocells, as well as the transition to LUCA, both marked with large arrows, appear shrouded in fog. LUCA, regarded to be autotrophic, thermophilic, anaerobic prokaryote that lived in hydrothermal vents and used the Wood-Ljungdahl pathway (section 5.1)^[161] likely evolved from a bewildering variety of different premetabolic forms of protocells that are very difficult to grasp or describe.

In-depth studies of highly parasitic life forms, such as the marine hyperthermophilic and chemolithoautotrophic archaea *Nanoarchaeum equitans*,^[204] that still survive with reduced genomes and an incomplete metabolism but recruit all metabolic building blocks and nutrients from outside are interesting models for pre-LUCA forms of Life. The difference to such parasitic organisms would be that these ancient forms would have to take up the building blocks from their abiotic environment.

Thus, the journey recounted in this report ends with a take-home message that the evolutionary development of metabolism specifically of amino acids and proteins can and should be considered only in the context of the origin of nucleotides and coenzymes and cofactors and not in isolation. In a nutshell, metabolism (“metabolism first”) and the evolutionary relation to replication (“genetic first”) should not be considered alone (Figure 1). Rather, coenzymes and cofactors are evident to be an important link between the two, in simplified terms structurally with nucleic acids^[7] and functionally with metabolism (scheme 12).

Acknowledgements

I thank my students and Dr. Carsten Zeilinger (Leibniz University Hannover) who provided important suggestions and ideas. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Keywords: amino acids • coenzymes • genetic code • metabolism • origin of life

- [1] Considerations on the origin of life is typically approached using the hypothetical-deductive method: a) A. Świeżyński, *Int. J. Astrobiol.* **2016**, *15*, 291–299; b) G. Wächterhäuser, *J. Theor. Biol.* **1997**, *187*, 483–449; c) A. Kirschning, T. A. C. Reydon, *Beilstein J. Org. Chem.* **2015**, *11*, 893–896.
- [2] In rare cases, the codon UGA stands for selenocysteine, and the codon UAG for pyrrolysine; both of which are not covered here.
- [3] a) K. Ruiz-Mirazo, C. Briones, A. de la Escosura, *Chem. Rev.* **2014**, *114*, 285–366; b) S. D. Fried, K. Fujishima, M. Makarov, I. Cherepashuk, K. Hlouchova, *J. R. Soc. Interface* **2022**, *19*: 20210641.
- [4] The RNA world theory as well as the emergence of homochirality are not in the focus of this review.
- [5] a) W. Gilbert, *Nature* **1986**, *319*, 618; b) G. F. Joyce, *Nature* **2002**, *418*, 214–221; c) M. P. Robertson, G. F. Joyce, *Cold Spring Harbor Perspect. Biol.* **2012**, *4*, a003608.
- [6] F. A. Anet, *Curr. Opin. Chem. Biol.* **2004**, *8*, 654–659.
- [7] A. Kirschning, *Angew. Chem. Int. Ed.* **2020**, *60*, 6242–6269.

- [8] a) M. A. Sephton, *Nat. Prod. Rep.* **2002**, *19*, 292–311; b) K. I. Öberg, *Chem. Rev.* **2016**, *116*, 9631–9663; c) K. Altwegg, H. Balsiger, A. Bar-Nun, J.-J. Berthelier, A. Bieler, P. Bochsler, C. Bröios, U. Calmonte, M. R. Combi, H. Cottin, J. De Keyser, F. Dhooghe, B. Fiethé, S. A. Fuselier, S. Gasc, T. I. Gombosi, K. C. Hansen, M. Haessig, A. Jäckel, E. Kopp, A. Korth, L. Le Roy, U. Mall, B. Marty, O. Mousis, T. Owen, H. Rème, M. Rubin, T. Sémon, C.-Y. Tzou, J. H. Waite, P. Wurz, *Sci. Adv.* **2016**, *2*, e1600285.
- [9] J. L. Bada, D. P. Glavin, G. D. McDonald, L. Becker, *Science* **1998**, *279*, 362–365.
- [10] A. M. Shaw, *Astrochemistry: From Astronomy to Astrobiology*; John Wiley & Sons: New York, NY, USA, **2007**.
- [11] D. S. Lauretta, H. Y. McSween, *Meteorites and the Early Solar System II*; University of Arizona Press: Tucson, AZ, USA, **2006**.
- [12] a) M. A. Sephton, *Astron. Geophys.* **2004**, *45*, 2–8; b) T. Koga, H. Naraoka, *Sci. Rep.* **2017**, *7*: 636.
- [13] a) J. Cronin, G. Cooper, S. Pizzarello, *Adv. Space Res.* **1995**, *15*, 91–97; b) J. E. Elsila, J. C. Aponte, D. G. Blackmond, A. S. Burton, J. P. Dworkin, D. P. Glavin, *ACS Cent. Sci.* **2016**, *2*, 370–379; c) A. S. Burton, E. L. Berger, *Life* **2018**, *8*, 14.
- [14] J. E. Elsila, N. M. Johnson, D. P. Glavin, J. C. Aponte, J. P. Dworkin, *Meteorit. Planet. Sci.* **2021**, *56*, 586–600.
- [15] A. Ruf, L. L. S. d'Hendecourt, P. Schmitt-Kopplin, *Life* **2018**, *8*, 18.
- [16] a) Y. Oba, Y. Takano, H. Naraoka, N. Watanabe, A. Kouchi, *Nat. Commun.* **2019**, *10*, 4413; b) H. Naraoka, Y. Yamashita, M. Yamaguchi, F.-R. Orthous-Daunay, *ACS Earth Space Chem.* **2017**, *1*, 540–550.
- [17] Y. Oba, H. Naraoka, *Meteorit. Planet. Sci.* **2006**, *41*, 1175–1181.
- [18] A. Shimoyama, R. Ogasawara, *Origins Life Evol. Biospheres* **2002**, *32*, 165–179.
- [19] K. Hadraoui, H. Cottin, S. L. Ivanovski, P. Zapf, K. Altwegg, Y. Benilan, N. Biver, V. Della Corte, N. Fray, J. Lasue, S. Merouane, A. Rotundi, V. Zakharov, *A&A* **2019**, *630*, A32.
- [20] a) M. P. Bernstein, J. P. Dworkin, S. A. Sandford, G. W. Cooper, L. J. Allamandola, *Nature* **2002**, *416*, 401–403; b) W. Portugal, S. Pilling, P. Boduch, H. Rothard, D. P. P. Andrade, *MNRAS* **2014**, *441*, 3209–3225.
- [21] a) S. N. Semenov, L. J. Kraft, A. Ainla, M. Zhao, M. Baghbanzadeh, V. E. Campbell, K. Kang, J. M. Fox, G. M. Whitesides, *Nature* **2016**, *537*, 656–660; b) S. Islam, M. W. Pownall, *Chem.* **2017**, *2*, 470–501; c) N. Kitadai, S. Maruyama, *Geoscience Front.* **2018**, *9*, 1117–1153.
- [22] a) K. Ashe, C. Fernandez-García, M. K. Corpinot, A. J. Coggins, D.-K. Bučar, M. W. Pownall, *Commun. Chem.* **2019**, *2*, 23; b) C. S. Foden, S. Islam, C. Fernández-García, L. Maugeri, T. D. Sheppard, M. W. Pownall, *Science* **2020**, *370*, 865–869.
- [23] a) S. L. Miller, *Science* **1953**, *117*, 528–529; b) S. L. Miller, H. C. Urey, *Science* **1959**, *130*, 245–251.
- [24] a) E. T. Parker, H. J. Cleaves, J. P. Dworkin, D. P. Glavin, M. Callahan, A. Aubrey, L. Lazcano, J. L. Bada, *Proc. Acad. Sci. USA* **2011**, *108*, 5526–5531; b) E. T. Parker, H. J. Cleaves, M. P. Callahan, J. P. Dworkin, D. P. Glavin, A. Lazcano, J. L. Bada, *Origins Life Evol. Biospheres* **2011**, *41*, 201–212; c) J. L. Bada, *Chem. Soc. Rev.* **2013**, *7*, 2186–2196.
- [25] a) J. Oró, *Biochem. Biophys. Res. Commun.* **1960**, *2*, 407–412; b) J. Oró, A. P. Kimball, *Arch. Biochem. Biophys.* **1961**, *94*, 217–227; c) J. Oró, S. S. Kamat, *Nature* **1961**, *190*, 442–443.
- [26] H. Yamada, M. Hirobe, T. Okamoto, *Yakugaku Zasshi* **1980**, *100*, 489–492.
- [27] R. Saladino, G. Botta, M. Delfino, E. Di Mauro, *Chem. Eur. J.* **2013**, *19*, 16916–16922.
- [28] W. Martin, M. J. Russell, *Phil. Trans. R. Soc. B* **2007**, *362*, 1887–1925.
- [29] E. Blöchl, M. Keller, G. Wächtershäuser, K. O. Stetter, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 8117–8120.
- [30] C. Huber, G. Wächtershäuser, *Science* **1997**, *276*, 245–247.
- [31] G. Wächtershäuser, M. W. W. Adams; In J. Wiegel (ed.), *Thermophiles: The Keys to Molecular Evolution and the Origin of Life*. **1998**, 47–57.
- [32] C. Huber, G. Wächtershäuser, *Science* **2006**, *314*, 630–632.
- [33] C. Huber, G. Wächtershäuser, *Tetrahedron Lett.* **2003**, *44*, 1695–1697.
- [34] a) C. Huber, G. Wächtershäuser, *Science* **1998**, *281*, 670–672; b) L. Leman, L. Orgel, M. R. Ghadiri, *Science* **2004**, *306*, 283–286.
- [35] C. Huber, W. Eisenreich, S. Hecht, G. Wächtershäuser, *Science* **2003**, *301*, 938–940.
- [36] J. D. Sutherland, *Angew. Chem. Int. Ed.* **2016**, *55*, 104–121; *Angew. Chem.* **2016**, *128*, 108–126.
- [37] a) C. Sagan, B. N. Khare, *Science* **1971**, *173*, 417–420; b) B. N. Khare, C. Sagan, *Nature* **1971**, *232*, 577–579.
- [38] B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy, J. D. Sutherland, *Nat. Chem.* **2015**, *7*, 301–307.
- [39] D. D. Sasselov, J. P. Grotzinger, J. D. Sutherland, *Sci. Adv.* **2020**, *6*, eaax3419.
- [40] D. Deamer, *Assembling Life “How can Life Begin on Earth and Other Habitable Planets?* Oxford University Press **2019**.
- [41] a) H. M. Fares, A. E. Marras, J. M. Ting, M. V. Tirrell, C. D. Keating, *Nature Commun.* **2020**, *11*, 5423; b) D. S. Ross, D. Deamer, *Life* **2016**, *6*, 28.
- [42] I. Mamajanov, P. J. MacDonald, J. Ying, D. M. Duncanson, G. R. Dowdy, C. A. Walker, A. E. Engelhart, F. M. Fernández, M. A. Grover, N. V. Hud, F. J. Schork, *Macromolecules* **2014**, *47*, 1334–1343.
- [43] a) C. Shen, L. Yang, S. Miller, J. Oró, *Origins Life Evol. Biospheres* **1987**, *17*, 295–305; b) C. Shen, L. Yang, S. L. Miller, J. Oró, *J. Mol. Evol.* **1990**, *31*, 167–174.
- [44] N. Friedmann, S. L. Miller, *Science* **1969**, *166*, 766–767.
- [45] Z. Guennoun, A. Coupeaud, I. Couturier-Tamburelli, N. Pietri, S. Coussan, J. P. Aycard, *Chem. Phys.* **2004**, *300*, 143–151.
- [46] B. Ménez, C. Pisapia, M. Andreani, F. Jamme, Q. P. Vanbellingen, A. Brunelle, L. Richard, P. Dumas, M. Réfrégiers, *Nature* **2018**, *564*, 59–63.
- [47] D. Yoshino, K. Hayatsu, E. Anders, *Cosmochim. Acta* **1971**, *35*, 927–938.
- [48] H. J. Cleaves, A. D. Aubrey, J. L. Bada, *Origins Life Evol. Biospheres* **2009**, *39*, 109–126.
- [49] E.-i. Imai, H. Honda, K. Hatori, A. Brack, K. Matsuno, *Science* **1999**, *283*, 831–834.
- [50] B. M. Rode, *Peptides* **1999**, *20*, 773–786.
- [51] G. Danger, R. Plasson, R. Pascal, *Chem. Soc. Rev.* **2012**, *41*, 5416–5429.
- [52] M. Frenkel-Pinter, M. Samanta, G. Ashkenasy, L. J. Leman, *Chem. Rev.* **2020**, *120*, 4707–4765.
- [53] D. L. Rohlfsing, *Science* **1976**, *193*, 68–70.
- [54] A. Brack, *Chem. Biodiversity* **2007**, *4*, 665–679.
- [55] J.-F. Lambert, *Origins Life Evol. Biospheres* **2008**, *38*, 211–242.
- [56] a) T. G. Waddell, B. S. Henderson, R. T. Morris, C. M. Lewis, A. G. Zimmermann, *Origins Life* **1987**, *17*, 149–153.
- [57] T. G. Waddell, S. K. Geevarghese, B. S. Henderson, R. M. Pagni, J. S. Newton, *Origins Life* **1989**, *9*, 603–607.
- [58] T. G. Waddell, T. J. Miller, *Origins Life* **1992**, *21*, 219–223.
- [59] M. Yadav, S. Pulletikurti, J. R. Yerabolu, R. Krishnamurthy, *Nat. Chem.* **2022**, *14*, 170–178.
- [60] M. A. Bromke, *Metabolites* **2013**, *3*, 294–311.
- [61] M. E. Winkler, S. Ramos-Montaño, *EcoSal* **2009**, *3*, 10.1128/ecosal-plus.3.6.1.9.
- [62] M. C. Evans, B. B. Buchanan, D. I. Arnon, *Proc. Natl. Acad. Sci. USA* **1966**, *55*, 928–934.
- [63] B. B. Buchanan, D. I. Arnon, *Photosynth. Res.* **1990**, *24*, 47–53.
- [64] M. Hügler, C. O. Wirsén, G. Fuchs, C. D. Taylor, S. M. Sievert, *J. Bacteriol.* **2005**, *187*, 3020–3027.
- [65] G. Fuchs, *Annu. Rev. Microbiol.* **2011**, *65*, 631–658.
- [66] A. Becerra, M. Rivas, C. García-Ferris, A. Lazcano, J. Peretó, *Int. Microbiol.* **2014**, *17*, 91–97.
- [67] N. Kitadai, M. Kameya, K. Fujishima, *Life* **2017**, *7*, 39.
- [68] P. Y.-T. Chen, B. Li, C. L. Drennan, S. J. Elliott, *Joule* **2019**, *3*, 595–611.
- [69] T. Nunoura, Y. Chikaraishi, R. Izaki, T. Suwa, T. Sato, T. Harada, K. Mori, Y. Kato, M. Miyazaki, S. Shimamura, K. Yanagawa, A. Shuto, N. Ohkouchi, N. Fujita, Y. Takaki, H. Atomi, K. Takai, *Science* **2018**, *359*, 559–563.
- [70] K. B. Muchowska, S. J. Varma, E. Chevallot-Beroux, L. Lethuillier-Karl, G. Li, J. Moran, *Nat. Ecol. Evol.* **2017**, *1*, 1716–1721.
- [71] K. B. Muchowska, S. J. Varma, J. Moran, *Nature* **2019**, *569*, 104–107.
- [72] M. Richter, *Nat. Prod. Rep.* **2013**, *30*, 1324–1345.
- [73] J. D. Fischer, G. L. Holliday, S. A. Rahman, J. M. Thornton, *J. Mol. Biol.* **2010**, *403*, 803–824.
- [74] M. E. Webb, A. Marquet, R. R. Mendel, F. Rébeillé, A. G. Smith, *Nat. Prod. Rep.* **2007**, *24*, 988–1008.
- [75] M. Richter, *Nat. Prod. Rep.* **2013**, *30*, 1324–1345.
- [76] H. B. White III, *J. Mol. Evol.* **1976**, *7*, 101–104.
- [77] T. Convay, *FEMS Microbiol. Rev.* **1992**, *9*, 1–27.
- [78] R. M. Daniel, M. J. Danson, *J. Mol. Evol.* **1995**, *40*, 559–563.
- [79] A. Katsyv, M. C. Schoelmerich, M. Basen, V. Müller, *FEBS Open Bio* **2021**, *11*, 1332–1342.
- [80] J. T. Beatty, H. Gest, *J. Bacteriol.* **1981**, *148*, 584–593.
- [81] H. E. Umberger, *Annu. Rev. Biochem.* **1978**, *47*, 533–606.
- [82] Y. Li, N. Kitadai, R. Nakamura, *Life* **2018**, *8*, 46.
- [83] L. Belmonte, S. S. Mansy, *Elements* **2016**, *12*, 413–418.
- [84] N. Kitadai, R. Nakamura, M. Yamamoto, K. Takai, N. Yoshida, Y. Oono, *Sci. Adv.* **2019**, *5*, eaav7848.
- [85] M. J. Dowler, W. D. Fuller, L. E. Orgel, R. A. Sanchez, *Science* **1970**, *169*, 1320–1321.

- [86] N. Friedmann, S. L. Miller, R. A. Sanchez, *Science* **1971**, *171*, 1026–1027.
- [87] H. J. Cleaves, S. L. Miller, *J. Mol. Evol.* **2001**, *52*, 73–77.
- [88] S. Civis, L. Juha, D. Babankova, J. Cvacka, O. Frank, J. Jehlicka, B. Kralikova, J. Krasa, P. Kubat, A. Muck, M. Pfeifer, J. Skala, J. Ullschmied, *Chem. Phys. Lett.* **2004**, *386*, 169–173.
- [89] H. Kim, S. A. Benner, *Chem. Eur. J.* **2018**, *24*, 581–584.
- [90] G. Magni, A. Amici, M. Emanuelli, N. Raffaelli, S. Ruggieri, *Adv. Enzymol. Relat. Mol. Biol.* **1999**, *73*, 135–182.
- [91] S. O. de Choudens, L. Loiseau, Y. Sanakis, F. Barras, M. Fontecave, *FEBS Lett.* **2005**, *79*, 3737–3743.
- [92] T. Begley, C. Kinsland, R. Mehl, A. Osterman, P. Dorrestein, *Vitam. Horm.* **2001**, *61*, 103–119.
- [93] B. Richts, J. Rosenberg, F. M. Commichau, *Front. Mol. Biosci.* **2019**, *6*, 32.
- [94] A. D. Welch, *Perspect. Biol. Med.* **1983**, *27*, 64–75.
- [95] V. Gorelova, O. Bastien, O. De Clerk, S. Lespinats, F. Rébeillé, D. van Der Straeten, *Sci. Rep.* **2019**, *9*, 5731.
- [96] M. N. Price, A. M. Deutschbauer, A. P. Arkin, *PLoS Genet.* **2021**, *17*, e1009342.
- [97] K. Tittmann, *FEBS J.* **2009**, *276*, 2454–2468.
- [98] V. I. Bunik, A. Tylicki, N. V. Lukashov, *FEBS J.* **2013**, *280*, 6412–6442.
- [99] L. D. Palmer, D. M. Downs, *J. Biol. Chem.* **2013**, *288*, 30693–30699.
- [100] In *E. coli* glycine is substituted by L-tyrosine: T. P. Begley, D. M. Downs, S. E. Ealick, F. W. McLafferty, A. P. Van Loon, S. Taylor, N. Campobasso, H. J. Chiu, C. Kinsland, J. J. Reddick, J. Xi, *Arch. Microbiol.* **1999**, *171*, 293–300.
- [101] a) L. D. Palmer, D. M. Downs, *J. Biol. Chem.* **2013**, *288*, 30693–30699; b) A. Chatterjee, A. B. Hazra, S. Abdelwahed, D. G. Hilmy, T. P. Begley, *Angew. Chem. Int. Ed.* **2010**, *49*, 8653–8656; *Angew. Chem.* **2010**, *122*, 8835–8838.
- [102] A similar discussion of the possible timing of their arrival can be made for other coenzymes, namely biotin and the flavins.
- [103] P. C. Dorrestein, H. Zhai, F. W. McLafferty, T. P. Begley, *Chem. Biol.* **2004**, *11*, 1373–1381.
- [104] F. Jacob, *Science* **1977**, *196*, 1161–1166.
- [105] a) K. Zhang, J. Bian, Y. Deng, A. Smith, R. E. Nunez, M. B. Li, U. Pal, A.-M. Yu, W. Qiu, S. E. Ealick, C. Li, *Nature Microbiol.* **2016**, *2*, 16213; b) D. Downs, *Nat. Microbiol.* **2016**, *2*, 16252; c) M. Kerstholt, M. G. Netea, L. A. B. Joosten, *Ticks Tick Borne Dis* **2020**, *11*, 101386.
- [106] D. P. R. Herlemann, O. Geissinger, W. Ikeda-Ohtsubo, V. Kunin, H. Sun, A. Lapidus, P. Hugenholtz, A. Brune, *Appl. Environ. Microbiol.* **2009**, *75*, 2841–2849.
- [107] a) S. J. Varma, K. B. Muchowska, P. Chatelain, J. Moran, *Nat. Ecol. Evol.* **2018**, *2*, 1019–1024; b) J. Sobotta, T. Geisberger, C. Moosmann, C. M. Scheidler, W. Eisenreich, G. Wächtershäuser, C. Huber, *A Possible Life* **2020**, *10*, 35; c) F. L. Sousa, T. Thiergart, G. Landan, S. Nelson-Sathi, I. A. C. Pereira, J. F. Allen, N. Lane, W. F. Martin, *Phil Trans R Soc B* **2013**, *368*, 0088; d) W. Nitschke, M. J. Russell, *Phil. Trans. R. Soc. B* **2013**, *368*, 0258.
- [108] B. Maden, *Biochem. J.* **2000**, *350*, 609–629.
- [109] a) R. R. Mendel, *J. Biol. Chem.* **2013**, *288*, 13165–13172; b) C. lobbi-Nivol, S. Leimkühler, *Biochim. Biophys. Acta Bioenerg.* **2013**, *1827*, 1086–1101.
- [110] R. H. White, *Vitam. Horm.* **2001**, *61*, 299–337.
- [111] M. Preiner, K. Igashiki, K. B. Muchowska, M. Yu, S. J. Varma, K. Kleinermanns, M. K. Nobu, Y. Kamagata, H. Tüysüz, J. Moran, W. F. Martin, *Nat. Ecol. Evol.* **2020**, *4*, 534–542.
- [112] W. F. Martin, *Front. Microbiol.* **2020**, *11*, 817.
- [113] A. Lazzcano, S. L. Miller, *J. Mol. Evol.* **1999**, *49*, 424–431.
- [114] A. R. Knaggs, *Nat. Prod. Rep.* **2003**, *20*, 119–136.
- [115] a) R. H. White, *Biochemistry* **2004**, *43*, 7618–7627; b) R. H. White, H. Xu, *Biochemistry* **2006**, *45*, 12366–12379; c) R. H. White, *Biochemistry* **2008**, *47*, 5037–5046.
- [116] D. V. Miller, M. Ruhlin, W. K. Ray, H. Xu, R. H. White, *FEBS Lett.* **2017**, *591*, 2269–2278.
- [117] T. Soderberg, *Archaea* **2005**, *1*, 347–352.
- [118] Theoretically, a TPP-free access to 1-deoxy-xylulose-5-phosphate, typically formed by a ketolase from pyruvate **23** + GAP **24**, is conceivable if, alternatively, 6-deoxy-5-ketofructose-1-phosphate **30** reacts with glycolaldehyde phosphate in an aldolase reaction. H. K. Lichtenhaler, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1999**, *50*, 47–65. This reaction has not yet been found.
- [119] a) R. Breslow, *J. Am. Chem. Soc.* **1958**, *80*, 3719–3726; b) M. Pareek, Y. Reddi, R. B. Sunoj, *Chem. Sci.* **2021**, *12*, 7973–7992.
- [120] a) R. Braakman, *PLoS Comp. Biol.* **2012**, *8*, e1002455; b) G. Kikuchi, Y. Motokawa, T. Yoshida, K. Hiraga, *Proc. Jpn. Acad. Ser. B* **2008**, *84*, 246–263; c) Y.-L. Du, K. S. Ryan, *Nat. Prod. Rep.* **2019**, *36*, 430–457.
- [121] PLP is also capable of promoting O₂-promoted oxidations but property may have evolved only when O₂ had appeared on earth: E. R. Hoffarth, K. C. Haatveit, E. Kuatsjah, G. A. MacNeil, S. Saroya, C. J. Walsby, L. D. Eltis, K. N. Houk, M. Garcia-Borràs, K. S. Ryan, *Proc. Nat. Acad. Sci.* **2021**, *118*, 40, e2012591118.
- [122] E. Meléndez-Hevia, T. G. Waddell, M. Cascante, *J. Mol. Evol.* **1996**, *43*, 293–303.
- [123] R. A. Jensen, *Ann. Rev. Microbiol.* **1976**, *30*, 409–425.
- [124] a) R. M. Drevland, Y. Jia, D. R. J. Palmer, D. E. Graham, *J. Biol. Chem.* **2008**, *283*, 28888–28896; b) D. M. Howell, K. Harich, H. Xu, R. H. White, *Biochemistry* **1998**, *37*, 10108–10117.
- [125] H. Xu, Y. Zhang, X. Guo, S. Ren, A. A. Staempi, J. Chiao, W. Jiang, G. Zhao, *J. Bacteriol.* **2004**, *186*, 5400–5409.
- [126] Homoalanine is a known amino acid formed by transamination from α -ketobutyrate **40**. H. E. Umbarger, B. Brown, *J. Bacteriol.* **1957**, *73*, 105.
- [127] α -Isopropylmalate synthetase is able to catalyze the condensations of acetyl-CoA with pyruvate, α -ketobutyrate, α -ketovalerate, or α -keto- β -methylvalerate as well as α -ketoisovalerate: M. Kisumi, M. Sugiura, I. Chibata, *J. Biochem.* **1976**, *80*, 333–339.
- [128] a) G. Kohlhaw, T. R. Leary, H. E. Umbarger, *Biol. Chem.* **1969**, *244*, 2218–2225; b) M. Strassman, L. N. Ceci, *Arch. Biochem. Biophys.* **1967**, *119*, 420–428; c) T. Sai, K. Aida, T. Uemura, *Gen. Appl. Microbiol.* **1969**, *15*, 345–363.
- [129] I. Apostol, J. Levine, J. Lippincott, J. Leach, E. Hess, C. B. Glascock, M. J. Weickert, R. Blackmore, *J. Biol. Chem.* **1997**, *272*, 28980–28988.
- [130] C. Alvarez-Carreño, A. Becerra, A. Lazcano, *Origins Life Evol. Biospheres* **2013**, *43*, 363–375.
- [131] T. H. Jukes, *Biochem. Biophys. Res. Commun.* **1973**, *53*, 709–714.
- [132] A. Kirschning, *Nat. Prod. Rep.* **2021**, *38*, 993–1010.
- [133] Other than the RNA world hypothesis have been proposed: "Protein first": P. Andras, C. Andras, *Med. Hypotheses* **2005**, *64*, 678–688; „metabolism first“: see Ref. [6]; „lipid first“: D. Segré, D. Ben-Eli, D. W. Deamer, D. Lancet, *Origins Life Evol. Biospheres* **2001**, *31*, 119–145.
- [134] a) J. Chela-Flores, *J. Theor. Biol.* **1994**, *166*, 163–166; b) T. O. Diener, *Biol. Direct* **2016**, *11*, 15; c) P. Forterre, D. Prangishvili, *Ann. N. Y. Acad. Sci.* **2009**, *1178*, 65–77; d) F. Broecker, K. Moelling, *Ann. N. Y. Acad. Sci.* **2019**, *1447*, 53–68.
- [135] R. Krishnamurthy, *Acc. Chem. Res.* **2017**, *50*, 455–459.
- [136] R. A. Jensen, *Annu. Rev. Microbiol.* **1976**, *30*, 409–425 and references cited therein.
- [137] N. H. Horowitz, *Proc. Natl. Acad. Sci. USA* **1945**, *31*, 153–157.
- [138] A. G. M. King, *Origins Life* **1977**, *8*, 39–53.
- [139] R. Kurup, P. A. Kurup, *The Origin of Life – Abiogenesis and Symbiosis*. Lap Lambert Academic Publishing **2019**, ISBN 620010007.
- [140] S. Granick, *Ann. N. Y. Acad. Sci.* **1957**, *96*, 292–308.
- [141] R. Fani, M. Fondi, *Phys. Life Rev.* **2009**, *6*, 23–52.
- [142] M. Ycas, *J. Theor. Biol.* **1974**, *44*, 145–160.
- [143] M. Fondi, R. Fani, *Res. Microbiol.* **2009**, *160*, 502–512.
- [144] P. Mehta, S. H. Abdelwahed, M. K. Fenwick, A. B. Hazra, M. E. Taga, Y. Zhang, S. E. Ealick, T. P. Begley, *J. Am. Chem. Soc.* **2015**, *137*, 10444–10447.
- [145] B. Hazra, A. W. Han, A. P. Mehta, K. C. Mok, V. Osadchiy, T. P. Begley, M. E. Taga, *Proc. Natl. Acad. Sci.* **2015**, *112*, 10792–10797.
- [146] K. Yokoyama, E. A. Lilla, *Nat. Prod. Rep.* **2018**, *35*, 660–694.
- [147] D. M. Gagnon, T. A. Stich, A. P. Mehta, S. H. Abdelwahed, T. P. Begley, R. D. Britt, *J. Am. Chem. Soc.* **2018**, *140*, 12798–12807.
- [148] L. Delaye, A. Lazzcano, *Phys. Life Rev.* **2005**, *2*, 47–64.
- [149] K. B. Muchowska, S. J. Varma, J. Moran, *Chem. Rev.* **2020**, *120*, 7708–7744.
- [150] a) H. I. Nakada, S. Weinhouse, *J. Biol. Chem.* **1953**, *204*, 831–836; later the reductive amination of α -ketoglutarate was reported b) H. J. Morowitz, E. Paterson, S. Chang, *Origins Life Evol. Biospheres* **1995**, *25*, 395–399.
- [151] a) L. M. Barge, E. Flores, M. M. Baum, D. G. VanderVelde, M. J. Russell, *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 4828–4833; b) F. S. Mohammed, K. Chen, M. Mojica, M. Conley, J. W. Napoline, C. Butch, P. Pollet, Ramanarayanan Krishnamurthy, C. L. Liotta, *Synlett* **2015**, *28*, 93–97.
- [152] B. E. Burns, Y. Xiang, C. L. Kinsland, F. W. McLafferty, T. P. Begley, *J. Am. Chem. Soc.* **2005**, *127*, 3682–3683.
- [153] F. Steffen-Munberg, C. Vickers, A. Thontowi, S. Schätzle, T. Meinhardt, M. S. Humble, H. Land, P. Berglund, U. T. Bornscheuer, M. Höhne, *ChemCatChem* **2013**, *5*, 154–157.

- [154] a) H. J. Morowitz, *Complexity* **1999**, *4*, 39–53; b) G. Caetano-Anolles, L. S. Yafremava, H. Gee, D. Caetano-Anolles, H. S. Kim, J. E. Mittenthal, *Int. J. Biochem. Cell Biol.* **2009**, *41*, 285–297.
- [155] a) A. Eschenmoser, *Chem. Biodiversity* **2007**, *4*, 554–573; b) S. A. Benner, H.-J. Kim, E. Biondi, *Life* **2019**, *9*, 84; c) S. Harrison, N. Lane, *Nat. Commun.* **2018**, *9*, 5176; d) A. Wołos, R. Roszak, A. Żądło-Dobrowolska, W. Beker, B. Mikulak-Klucznik, G. Spółnik, Maw Dygas, S. Szymkuć, B. A. Grzybowski, *Science* **2020**, *369*, eaaw1955.
- [156] The evolution of fatty acids and the fatty acid synthase type II is not discussed here.
- [157] a) E. Smith, H. J. Morowitz, *The Origin and Nature of Life on Earth: the Emergence of the Fourth Geosphere*; 1st edn.; Cambridge University Press: New York, NY, **2016**; b) H. J. Morowitz, J. D. Kostelník, J. Yang, G. D. Cody, *Proc. Nat. Acad. Sci.* **2000**, *97*, 7704–7708; c) E. Smith, S. J. Morowitz, *Proc. Nat. Acad. Sci.* **2004**, *101*, 13168–13173; d) R. T. Stubbs, M. Yadav, R. Krishnamurthy, G. Springsteen, *Nat. Chem.* **2020**, *12*, 1016–1022.
- [158] a) M. Ralser, *Biochem. J.* **2018**, *475*, 2577–2592; b) K. B. Muchowska, E. Chevallot-Beroux, J. Moran, *Bioorg. Med. Chem.* **2019**, *27*, 2292–2297.
- [159] Arguments positioning glycine as one of the early coded amino acids during the emergence of the genetic code: a) H. S. Bernhardt, W. M. Patrick, *J. Mol. Evol.* **2014**, *78*, 307–309; b) H. S. Bernhardt, W. P. Tate, *Biol. Direct* **2008**, *3*, 53; c) K. Tamura, *J. Mol. Evol.* **2015**, *81*, 69–71.
- [160] L. E. Orgel, *PLoS Biol.* **2008**, *6*, 5–13.
- [161] a) M. C. Weiss, F. L. Sousa, N. Mrnjavac, S. Neukirchen, M. Roettger, S. Nelson-Sathi, W. F. Martin, *Nat. Microbiol.* **2016**, *1*, 16116; b) W. F. Martin, F. L. Sousa, *Cold Spring Harbor Perspect. Biol.* **2016**, *8*, a01812.
- [162] A. Nasir, E. Romero-Severson, J. M. Claverie, *Trends Microbiol.* **2020**, *28*, 959–967.
- [163] A. Cornish-Bowden, M. L. Cárdenas, *J. Theor. Biol.* **2017**, *434*, 68–74.
- [164] a) M. Ilardo, M. Meringer, S. Freeland, B. Rasulev, H. J. Cleaves II, *Sci. Rep.* **2015**, *5*, 9414; b) V. Kubyshkin, N. Budisa, *Biotechnol. J.* **2017**, *12*, 1600097; c) M. Ilardo, R. Rose, M. Meringer, B. Rasulev, N. Grefenstette, J. Stephenson, S. Freeland, R. J. Gillams, C. J. Butch, H. J. Cleaves II, *Sci. Rep.* **2019**, *9*, 12468.
- [165] M. Makarov, J. Meng, V. Tretyachenko, P. Srb, A. Březinová, V. G. Giacobelli, L. Bednárová, J. Vondrášek, A. K. Dunker, K. Hlouchová, *Protein Sci.* **2021**, *30*, 1022–1034.
- [166] J. Yan, J. Cheng, L. Kurgan, V. N. Uversky, *Cell. Mol. Life Sci.* **2020**, *77*, 2423–2440.
- [167] a) K. U. Walter, K. Vamvaca, D. Hilvert, *J. Biol. Chem.* **2005**, *280*, 37742–37746; b) M. M. Müller, J. R. Allison, N. Hongdilokkul, L. Gaillon, P. Kast, W. F. van Gunsteren, P. Marlière, D. Hilvert, *PLoS Genet.* **2013**, *9*, e1003187.
- [168] R. Shibue, T. Sasamoto, M. Shimada, B. Zhang, A. Yamagishi, S. Akanuma, *Sci. Rep.* **2018**, *8*, 1227.
- [169] S. Akanuma, S. Yokobori, Y. Nakajima, M. Bessho, A. Yamagishi, *Evolution* **2015**, *69*, 2954–2962.
- [170] a) M. C. Maurel, J. Ninio, *Biochimie* **1987**, *69*, 551–553; b) C. Shen, T. Mills, J. Oró, *J. Mol. Evol.* **1990**, *31*, 175–179; c) D. H. White, J. C. Erickson, *J. Mol. Evol.* **1980**, *16*, 279–290; d) C. Shen, A. Lazcano, J. Oró, *J. Mol. Evol.* **1990**, *31*, 445–452.
- [171] S.-M. Liao, Q.-S. Du, J.-Z. Meng, Z.-W. Pang, R.-B. Huang, *Chem. Cent. J.* **2013**, *7*, 44.
- [172] I. Yanai, Y. I. Wolf, E. V. Koonin, *Genome Biol.* **2002**, *3*, 0024.1.
- [173] a) R. Fani, M. Brilli, M. Fondi, P. Liò, *BMC Evol. Biol.* **2007**, *7 Suppl 2*, S4; b) R. Fani, P. Liò, A. Lazcano, *J. Mol. Evol.* **1995**, *41*, 760–774.
- [174] C. Hsiao, I.-C. Chou, C. D. Okafor, J. C. Bowman, E. B. O'Neill, S. S. Athavale, A. S. Petrov, N. V. Hud, R. M. Wartell, S. C. Harvey, L. D. Williams, *Nat. Chem.* **2013**, *5*, 525–528.
- [175] M. Frenkel-Pinter, J. W. Haynes, A. M. Mohyeldin, C. Martin, A. B. Sargon, A. S. Petrov, R. Krishnamurthy, N. V. Hud, L. D. Williams, L. J. Leman, *Nat. Commun.* **2020**, *11*, 3137.
- [176] a) F. H. Crick, *J. Mol. Biol.* **1968**, *38*, 367–379; b) L. E. Orgel, *J. Mol. Biol.* **1968**, *38*, 381–393; c) L. R. de Pouplana, A. G. Torres, Á. Rafels-Ybern, *Life* **2017**, *7*, 14.
- [177] A. Gospodinov, D. Kunnev, *Life* **2020**, *10*, 81.
- [178] a) F. Scossa, A. R. Fernie, *Comput. Struct. Biotechnol. J.* **2020**, *18*, 482–500; b) A. Becerra, *J. Mol. Evol.* **2021**, *89*, 183–188.
- [179] C. De Duve, ' Blueprint for a Cell: The Nature and Origin of Life', Patterson, Burlington, N. C., **1991**.
- [180] R. Pascal, L. Boiteau, A. Commeyras, *Top. Curr. Chem.* **2005**, *259*, 69.
- [181] A. L. Weber, *Origins Life Evol. Biospheres* **2001**, *31*, 71.
- [182] a) M. Barbieri, *BioSystems* **2019**, *185*, 104024; b) V. Kubyshkin, C. G. Acevedo-Rocha, N. Budisa, *BioSystems* **2018**, *164*, 16–25; c) H. Seligmann, G. Warthi, *Comput. Struct. Biotechnol. J.* **2017**, *15*, 412–424.
- [183] A. Jimenez-Sánchez, *J. Mol. Evol.* **1995**, *41*, 712–716.
- [184] Hyperthermophilic archaea such as *P. fumarii* exhibit higher GC content, particularly at the third nucleic acid within the codon, likely due to a selection pressure for more hydrogen bonding to increase DNA thermal stability.
- [185] a) B. R. Francis, *J. Mol. Evol.* **2013**, *77*, 134–158; b) M. Eigen, R. Winkler-Oswatitsch, *Naturwissenschaften* **1981**, *68*, 217–228; c) K. Ikebara, Y. Omori, R. Arai, A. Hirose, *J. Mol. Evol.* **2002**, *54*, 530–538.
- [186] a) H. Hartman, T. Smith, *Life* **2014**, *4*, 227–249; b) T. Smith, H. Hartman, *FEBS Lett.* **2015**, *589*, 3499–3507.
- [187] a) V. Kubyshkin, N. Budisa, *Int. J. Mol. Sci.* **2019**, *20*, 5507; b) V. Kubyshkin, N. Budisa, *Curr. Opin. Biotechnol.* **2019**, *60*, 242–249.
- [188] H. Grosjean, E. Westhof, *Nucl. Acids Res.* **2016**, *44*, 18020–8040.
- [189] J. C. W. Shepherd, *Trends Biochem. Sci.* **1984**, *9*, 8–10.
- [190] M. V. José, G. S. Zamudio, M. Palacios-Pérez, J. R. Bobadilla, S. T. Farías, *Origins Life Evol. Biospheres* **2015**, *45*, 77–83.
- [191] T. H. Jukes, *J. Mol. Evol.* **1996**, *42*, 377–381.
- [192] a) J. T.-F. Wong, *BioEssays* **2005**, *27*, 416–425; b) G. K. Philip, S. J. Freeland, *Astrobiology* **2011**, *11*, 235–240.
- [193] M. J. T. Wong, P. M. Bronskill, *J. Mol. Evol.* **1979**, *13*, 115–125.
- [194] A. L. Weber, S. L. Miller, *J. Mol. Evol.* **1981**, *17*, 273–284.
- [195] Y. Lu, S. J. Freeland, *J. Theor. Biol.* **2008**, *250*, 349–361.
- [196] a) M. Ibba, D. Söll, *Annu. Rev. Biochem.* **2000**, *69*, 617–650; b) K. Sheppard, J. Yuan, M. J. Hohn, B. Jester, K. M. Devine, D. Söll, *Nucl. Acids Res.* **2008**, *36*, 1813–1825.
- [197] M. Di Giulio, *J. Mol. Evol.* **2002**, *55*, 616–622.
- [198] J. Moser, S. Lorenz, C. Hubschwerlen, A. Rompf, D. Jahn, *J. Biol. Chem.* **1999**, *274*, 30679–30685.
- [199] J. Moser, W.-D. Schubert, V. Beier, I. Bringemeier, D. Jahn, D. W. Heinz, *EMBO J.* **2001**, *20*, 6583–6590.
- [200] A. Sauerwald, W. Zhu, T. A. Major, H. Roy, S. Palioura, D. Jahn, W. B. Whitman, J. R. Yates 3rd, M. Ibba, D. Söll, *Science* **2005**, *307*, 1969–1972.
- [201] X. Long, H. Xue, J. T.-F. Wong, *Evol. Bioinformatics* **2020**, *16*, 1–11.
- [202] a) W. L. Muth, R. N. Costilow, *J. Biol. Chem.* **1974**, *249*, 7463–7467; b) M. Raupner, R. H. White, *J. Bacteriol.* **2001**, *183*, 5203–5205.
- [203] Latest publications on the coevolution of RNA and peptides as well as on RNA-self replication see a) F. Müller, L. Escobar, F. Xu, E. Węgrzyn, M. Nainyté, T. Amatov, C.-Y. Chan, A. Pichler, T. Carell, *Nature* **2022**, *605*, 279–284; b) R. Mizuuchi, T. Furabayashi, N. Ichihashi, *Nat. Commun.* **2022**, *13*: 1460.
- [204] H. Huber, S. Burggraf, T. Mayer, I. Wyschkony, R. Rachel, K. O. Stetter, *Int. J. Syst. Evol. Microbiol.* **2000**, *50*, 2093–2100.

Manuscript received: May 8, 2022

Accepted manuscript online: June 21, 2022

Version of record online: July 28, 2022