The FEBS Journal

FEBS PRESS science publishing by scientists

Evidence for corrin biosynthesis in the last universal common ancestor

Luca D. Modjewski[1], Val Karavaeva[2,3], Natalia Mrnjavac[1], Michael Knopp[1], William F. Martin[1] and Filipa L. Sousa[2]

1 Institute of Molecular Evolution, Faculty of Mathematics and Natural Sciences, Heinrich Heine University Düsseldorf, Germany
2 Department of Functional and Evolutionary Ecology, University of Vienna, Austria
3 Vienna Doctoral School of Ecology and Evolution, University of Vienna, Austria

Keywords
biochemical evolution; cobalamin; cobamide; tetrapyrroles; vitamin B_{12}

Correspondence
F. L. Sousa, Department of Functional and Evolutionary Ecology, University of Vienna, Djerassipl. 1, 1030 Vienna, Austria
Tel: +43 1 4277 76510
E-mail: filipa.sousa@univie.ac.at

Luca D. Modjewski, Val Karavaeva, and Natalia Mrnjavac contributed equally to this article.

(Received 12 August 2024, revised 4 October 2024, accepted 10 December 2024)

doi:10.1111/febs.17367

Corrinoids are cobalt-containing tetrapyrroles. They include adenosylcobalamin (vitamin B_{12}) and cobamides that function as cofactors and coenzymes for methyl transfer, radical-dependent and redox reactions. Though cobamides are the most complex cofactors in nature, they are essential in the acetyl-CoA pathway, thought to be the most ancient CO_2-fixation pathway, where they perform a pterin-to-cobalt-to-nickel methyl transfer reaction catalyzed by the corrinoid iron–sulphur protein (CoFeS). CoFeS occurs in H_2-dependent archaeal methanogens, the oldest microbial lineage by measure of physiology and carbon isotope data, dating corrinoids to ca. 3.5 billion years. However, CoFeS and cobamides are also essential in the acetyl-CoA pathway of H_2-dependent bacterial acetogens. To determine whether corrin biosynthesis was established before archaea and bacteria diverged, whether the pathways arose independently or whether cobamide biosynthesis was transferred from the archaeal to the bacterial lineage (or *vice versa*) during evolution, we investigated phylogenies and structural data for 26 enzymes of corrin ring and lower ligand biosynthesis. The data trace cobamide synthesis to the common ancestor of bacteria and archaea, placing it in the last universal common ancestor of all lifeforms (LUCA),

Abbreviations
ACS, acetyl-CoA synthase; AcsE, Methyl-H_4F:CoFeS methyltransferase; BchE, anaerobic magnesium-protoporphyrin IX monomethyl ester cyclase; BchQ, chlorophyll/bacteriochlorophyll a synthase; BchR, bacteriochlorophyllide d C-12(1)-methyltransferase; BioD, dethiobiotin synthetase; BtuR/CobA/PduO, corrinoid adenosyltransferase; BzaABCDEF, 5-hydroxybenzimidazole synthase; CbiA, cobyrinic acid a, c-diamide synthetase; CbiB, adenosylcobinamide-phosphate synthase; CbiC, precorrin-8X/cobalt-precorrin-8 methylmutase; CbiD, cobalt-precorrin-5B (C1)-methyltransferase; CbiE, cobalt-precorrin-7 (C5)-methyltransferase; CbiF, precorrin-4/cobalt-precorrin-4 C11-methyltransferase; CbiG, cobalt-precorrin 5A hydrolase; CbiH, precorrin-3B C17-methyltransferase/cobalt-factor III methyltransferase; CbiJ, precorrin-6A/cobalt-precorrin-6A reductase; CbiL, precorrin-2/cobalt-factor-2 C20-methyltransferase; CbiT, cobalt-precorrin-6B (C15)-methyltransferase; $CbiX_L$/CbiK/CbiXs, sirohydrochlorin cobaltochelatase; CfbB, Ni-sirohydrochlorin a,c-diamide synthase; CobB, NAD-dependent protein deacetylase/lipoamidase; CobC/CobZ, adenosylcobalamin/alpha-ribazole phosphatase; CobQ, adenosylcobyric acid synthase; CobS, adenosylcobinamide-GDP ribazoletransferase; CobU/CobY, adenosylcobinamide-phosphate guanylyltransferase; CODH, carbon monoxide dehydrogenase; CoFeS, corrinoid iron–sulphur protein; DMB, 5,6-dimethylbenzimidazole; DsrAB, dissimilatory sulphite reductase; EutT, ethanolamine utilization cobalamin adenosyltransferase; FdhA, formate dehydrogenase; Fhs, N10-formyl-H_4F synthetase; FolD, 5,10-methenyl-H_4F cyclohydrolase/dehydrogenase; Ftr, formylmethanofuran:H_4MPT formyltransferase; FwdBD, formylmethanofuran dehydrogenase; Ga, Giga annum; H_4F, tetrahydrofolate; H_4MPT, tetrahydromethanopterin; HBI, 5-hydroxybenzimidazole; HemH/PpfC/CpfC, ferrochelatase; HGT/LGT, horizonal/lateral gene transfer; Hmd, 5,10-methenyltetrahydromethanopterin hydrogenase; ICGs, informational core genes; LUCA, last universal common ancestor; MazG, nucleoside triphosphate diphosphatase; Mch, methenyl-H_4MPT cyclohydrolase; Mer, 5,10-methylenetetrahydromethanopterin reductase; MetF, 5,10-methylene-H_4F reductase; MetF, 5,10-methylenetetrahydrofolate reductase; MeTr, methyltransferase present in some acetogens; MinD, septum site-determining protein; MoCo, molybdenum cofactor; Mtd, F_{420}-dependent methylene-H_4MPT dehydrogenase; ParA, chromosome partitioning protein; PFOR, pyruvate:ferredoxin oxidoreductase,; SAM, *S*-adenosylmethionine; SirB/ShfC, sirohydrochlorin ferrochelatase; SirC, precorrin-2 dehydrogenase; ThiC, phosphomethylpyrimidine synthase; UroIII, uroporphyrinogen III; UroM/SuMT/SirA, uroporphyrin-III C-methyltransferase.

while pterin-dependent methyl synthesis pathways likely arose independently post-LUCA in the lineages leading to bacteria and archaea. Enzymes of corrin biosynthesis were recruited from preexisting ancient pathways. Evolutionary forerunners of CoFeS function were likely Fe-, Ni- and Co-containing solid-state surfaces, which, in the laboratory, catalyze the reactions of the acetyl-CoA pathway from CO_2 to pyruvate under serpentinizing hydrothermal conditions. The data suggest that enzymatic corrin biosynthesis replaced insoluble solid-state catalysts that tethered primordial CO_2 assimilation to the Earth's crust, suggesting a role for corrin synthesis in the origin of free-living cells.

Introduction

Corrinoids are nature's most complex cofactors and coenzymes. They include cobalt-containing cobamides, which differ with respect to the lower ligand: 5,6-dimethylbenzimidazole in cobalamin, and 5-hydroxybenzimidazole, 5-methoxybenzimidazole, 5-met hylbenzimidazole, benzimidazole, 5-methoxy-6-methyl benzimidazole, phenol, *p*-cresol, 2-methylsulfiny ladenine, 2-methylsulfonyladenine, 2-methyladenine or adenosine in other cobamides [1–6]. These cobalt-containing corrins are mainly involved in methyl transfer [7–10] and in radical-based reactions [11], including the generation of deoxyribose nucleotides for DNA by adenosylcobalamin-dependent ribonucleotide reductase [12–14]. Corrinoid-dependent enzymes are also involved in a variety of other processes, such as the biosyntheses of amino acids (Met, [10,15,16]), queuosine [17], bacteriochlorophyll [18], hopanoids [19], polytheonamide [10,20], antibacterials and antivirals [21,22], as well as the detoxification of aromatic and aliphatic chlorinated organics [16,23], and a range of fermentations: amino acids [15,24–27], one-carbon compounds (e.g., methylamine, methanol and methylthiol [16]), ethanolamine and choline [16,28,29], nicotinic acid [16,30], 1,2-diols [16,31–33] and fatty acids [16,26,27]. Corrinoids are required by many organisms but are synthesized by few [34–37]. The biosynthetic pathway in anaerobes entails over 20 enzymatic steps if starting to count from uroporphyrinogen III, the first macrocyclic intermediate and the universal precursor of tetrapyrrole biosynthesis [38], the number of steps to the final product depending on the nature of the lower ligand [39]. Though the genes for corrin synthesis can be laterally transferred across lineages, and some prokaryotes have partial cobalamin synthesis, for example, from cobyric acid to adenosylcobalamin [40], by far the most common strategy to satisfy corrin requirements is to obtain the cofactor from corrin-producing organisms via the environment [34]. The list of corrin auxotrophs includes prokaryotes [35–37,41], algae [37,42], protists [37,43] and humans who strictly require cobalamin for only two enzymatic reactions: the isomerization of methylmalonyl-CoA to succinyl-CoA and the synthesis of methionine from homocysteine [44]. Corrins are thought to be ancient [45,46], and earlier studies traced some corrin-dependent enzymes to the last universal common ancestor, LUCA [47], but the issue of whether the enzymes of corrin biosynthesis themselves trace to LUCA is so far unresolved.

The evidence for corrin antiquity stems from its essential role in the acetyl-CoA pathway, the most ancient among CO_2-fixing pathways [48–51]. In the acetyl-CoA pathway of H_2-dependent acetogens (bacteria) and H_2-dependent methanogens (archaea), the corrinoid iron–sulphur protein (CoFeS) catalyzes a cobamide-dependent cobalt-to-nickel methyl transfer reaction that is essential for acetyl-CoA synthesis [52,53] (Fig. 1A). The geological age of the corrins is given by methanogens, which are the most ancient lineage of microbes by the measure of physiology [46] and carbon isotope data. Methane isotopes [65] provide the earliest evidence for life on Earth, tracing methanogenesis, hence the acetyl-CoA pathway, hence an ancestral CoFeS (judging by the homology between bacterial and archaeal CoFeS), hence corrins, into rocks 3.8 billion years of age. Isotopically ultralight carbon is also reported in 3.8 Ga sediment carbon [66,67], which suggests the presence of the acetyl-CoA pathway [68] without discriminating an acetogen or methanogen source. Phylogenetic studies implicate acetogens and methanogens as the most ancient lineages among modern bacteria and archaea, respectively [47]. Both acetogens and methanogens inhabit H_2-producing hydrothermal systems today [69,70] and

might present windows into the physiology and habitat of the first cells [47,69–72].

In addition to physiology [48–51], isotopes [65–67] and phylogeny [47,69,70], metal catalysts provide an independent line of evidence in favour of the antiquity of the acetyl-CoA pathway. Serpentinizing (H_2-producing) hydrothermal systems synthesize formate [73], an intermediate of the acetyl-CoA pathway [60], and methane, the end product of methanogenesis, from abiotic reactions of H_2 and CO_2 [74]. Though the catalysts that facilitate formate and methane synthesis in vents have not yet been identified *in situ*, native (zero-valent) transition metals are thought to be the catalysts [60], for two reasons. First, Fe^0, Ni^0, Co^0 and their alloys are naturally deposited in serpentinizing hydrothermal vents because of their highly reducing conditions [75]. Second, Fe^0, Ni^0, Ni_3Fe and silica-supported Co-Fe alloys can replace the proteins that synthesize formate, acetate, pyruvate from H_2 and CO_2 via the acetyl-CoA pathway [60–64] (Fig. 1C). In water and in the absence of enzymes, the solid-state metals generate methyl groups from H_2 and CO_2, producing formate, acetate, pyruvate and methane, replacing the function of over 100 enzymes [76]. Furthermore, Ni^0 readily forms amino acids from 2-oxoacids, NH_3 and H_2 [77], and Fe^0 can reduce ferredoxin, functionally substituting for flavin-based electron bifurcation, a complex process that requires multi-enzyme systems capable of coupling exergonic and endergonic reactions [78] and reducing the low-potential physiological donor of electrons in the acetyl-CoA pathway, ferredoxin, with electrons from H_2 [79].

Countering the case for corrin antiquity, its biosynthetic pathway is the longest, and the product is structurally the most complex of any known tetrapyrrole [38,80,81], raising the question of its possible functional precursor during biochemical evolution (Fig. 1). Corrin synthesis itself holds no clear evidence for antiquity, other than its basal position in tetrapyrrole biogenesis [46,82]. While some cofactors such as thiamine or pyridoxal phosphate participate in their own biosynthesis, forming small autocatalytic cycles [71], that is not the case for corrins. The pathway involves the introduction of eight methyl groups to the tetrapyrrole backbone, one of which aids methylene elimination and contraction of the porphyrin to the corrin ring [83,84]. The seven methyl groups present in the corrinoid structure [9] (Fig. 1B) are donated by *S*-adenosylmethionine (SAM), a cofactor of methyl transfers and radical reactions [85], yet SAM does not substitute for cobamide in either the CoFeS reaction [52] or in the pumping reaction at MtrA-H in methanogens [86].

The age of methanogens and the identification of functional abiotic homologues of the acetyl-CoA pathway point to simple environmental precursors of corrin-dependent methyl transfer (Fig. 1C) but do not directly address corrin biosynthesis antiquity. To determine whether corrin synthesis required for the acetyl-CoA pathway traces to the common ancestor of bacteria and archaea – the last universal common ancestor (LUCA) – or was the result of later lateral gene transfers [47], we have investigated the evolutionary history of the enzymes of corrin biosynthesis in acetogens and methanogens [38], including the synthesis of the lower ligand (Fig. 1B) [39] and CoFeS [87,88].

Results and Discussion

There are two pathways of corrin synthesis, an O_2-dependent (aerobic) pathway and an oxygen-sensitive (anaerobic) pathway, where the aerobic pathway requires oxygen for ring contraction and cobalt insertion [38,81]. Only the enzymes of the assumed-to-be-more-ancient anaerobic pathway are of interest in early biochemical evolution for three reasons. First, O_2 is a product of cyanobacterial photosynthesis, which was lacking [89] in the environment at the origin of the acetyl-CoA pathway and methanogens > 3.5 billion years ago [65,82]. Second, both abiotic [90] and enzymatic reactions to pyruvate synthesis along the acetyl-CoA pathway are extremely O_2-sensitive due to the involvement of FeS clusters in electron transfer and other radical-generating reactions [91,92], such that the corrins required for the pathway arose in the absence of O_2 (as with chlorophyll synthesis, the O_2-dependent corrin synthesis reactions arose in response to O_2-dependent inhibition of radical-dependent reactions [93]). Third, acetogens and methanogens themselves, which depend upon CoFeS and the acetyl-CoA pathway for growth [48,49,78], are extremely O_2-sensitive and use the anaerobic corrin synthesis pathway. Because physiology [46], phylogeny [47,69,70], and carbon isotopes [65,66,68] point to acetogens and methanogens as the most ancient bacterial and archaeal lineages, respectively, and because both require cobamide [7,8], the corresponding genes of anaerobic corrin biosynthesis were extracted from the genome of the model acetogen *Moorella thermoacetica* [94] and from the genome of the methanogen *Methanococcus maripaludis* [95].

Corrinoids are not generally viewed as likely products of prebiotic synthesis because of their complex structure and biosynthesis [38,39,96], which is outlined in Fig. 2 (and in Fig. S1; see Supporting Information

for a description of the initial biosynthetic steps). Of note, the initial steps, from Uro-III to the formation of sirohydrochlorin, are common to siroheme biosynthesis [99], and homologues of CbiA and CbiX are involved in F_{430} biosynthesis [100,101], highlighting the duplications that occurred during tetrapyrrole evolution. At a late stage of the pathway (Fig. 2), adenosylation (catalyzed by BtuR/CobA) introduces an adenosyl moiety as the upper ligand at this stage of the adenosylcobalamine synthesis pathway [38]. In the cobamides that are employed in the CoFeS reaction, the adenosyl upper ligand is missing, the corresponding Co coordination site being occupied by water or by the methyl group during the methyl transfer

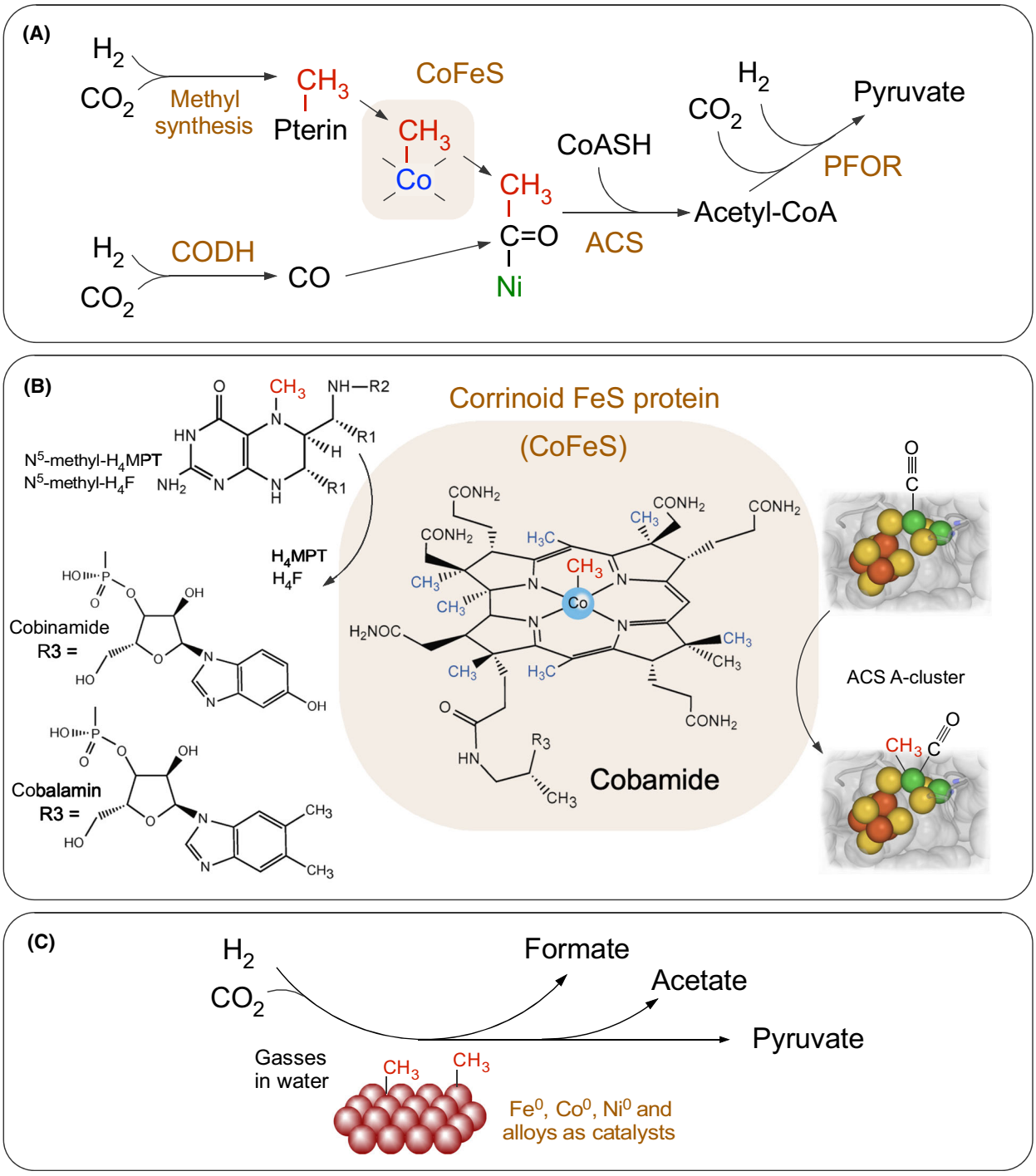


Fig. 1. The role of corrins in the acetyl-CoA pathway. (A) The acetyl-CoA pathway, redrawn from references [49, 54]. Methyl synthesis: pterin-dependent methyl synthesis in the acetyl-CoA pathway of bacteria and archaea from H_2 and CO_2 [49]. ACS, acetyl-CoA synthase [53]; CODH, carbon monoxide dehydrogenase [55]; CoFeS, corrinoid iron–sulphur protein [52]; PFOR, pyruvate:ferredoxin oxidoreductase [54]. The methyl group is highlighted in red, the corrinoid cobalt atom in blue, the proximal nickel atom of the ACS active site [53] in green and enzyme names or steps are given in sepia letters. (B) Corrinoid-dependent methyl transfer from a pterin cofactor to the active site metal cluster (A-cluster) of acetyl-CoA synthase (ACS) (PDB ID: 7ZKJ [56]) in the acetyl-CoA pathway of acetogens and methanogens. The corrinoid iron–sulphur protein CoFeS catalyzes the methyl-transfer reaction, in some cases aided by a separate methyltransferase (MeTr) [52,57]. H_4F and H_4MPT: tetrahydrofolate and tetrahydromethanopterin, respectively. R1 = –H in H_4F and $–CH_3$ in H_4MPT.For the chemical structure of R2,in H4F and H4MPT, , see[58]. Methylcobalamin is shown in the base-off configuration [59], with –R3 replacing the nucleotide loop. Methyl groups derived from *S*-adenosylmethionine (SAM) are shown in blue. Fe atoms are shown in orange, S atoms in yellow and Ni atoms in green. Lower left: Examples of two cobamide lower ligands (–R3 in A); 5-hydroxybenzimidazole (HBI) is found in the cobinamides of some methanogens [39], while 5,6-dimethylbenzimidazole (DMB, the lower ligand base of cobalamin) is found in some bacteria [52]. (C) Native metals catalyze formate, acetate and pyruvate synthesis from H_2 and CO_2 over Ni^0, Fe^0, Co^0, Ni_3Fe and other alloys [60–64], identifying a nonenzymatic, geochemical precursor of corrin-dependent methyl transfer shown in panels A and B. More is known about the enzymatic reaction mechanisms [53] than about the metal-catalyzed forerunners [60–64]. The catalyst-bound methyl groups in panel C are probable intermediates in the reaction mechanism, as methanol and methane are sometimes observed as products [60,61]. In some experiments, Fe_3S_4 and Fe_3O_4 are effective as catalysts, but it cannot be excluded that these catalysts are reduced to Fe^0 at some sites during the reaction with H_2.

reaction [52]. Accordingly, some methanogens (and other archaea) lack BtuR/CobA homologues [36,102]. The biosynthetic route depicted in Figs 2 and 3 follows that described for adenosylated intermediates [38,96], the corresponding genes, except for BtuR/CobA, being present in methanogens and acetogens that are included in this study.

The anaerobic synthesis of the cobalamin lower ligand dimethylbenzimidazole (DMB) is performed by the *bzaABCDE* gene products [39], starting from 5-aminoimidazole ribotide (AIR), an intermediate of thiamine and purine biosynthesis. In some organisms, only the product of *bzaF*, a radical SAM enzyme homologous to *bzaA*, *bzaB* and *thiC* products, is present [39,109], generating 5-hydroxybenzimidazole (HBI) as the cobamide lower ligand (Fig. 1C). Other organisms employ different intermediates of lower ligand synthesis [39], and the lower ligands of cobamides isolated from acetogens [7] and methanogens vary across species [8]. Here we analysed only the biosynthesis of the lower ligand DMB since it is found in both archaeal and bacterial organisms and has a large impact in marine settings [110]. Moreover, it was shown that at least in *Salmonella enterica*, the incorporation of DMB or of adenine as lower ligands is performed by the same enzymes, and, in DMB-limiting conditions, adenine can replace DMB [111].

Homologies trace corrin synthesis to the acetogen-methanogen common ancestor

Alignments of structures generated from amino acid sequences with AlphaFold [103,105] show that the structures of enzymes for methanogen and acetogen cobamide synthesis superimpose well in all cases except one, indicating homology of the corrin synthesis pathways in methanogens and acetogens (Fig. 3, left panel). The same is true for the two enzymes of the carbonyl branch of the acetyl-CoA pathway, acetyl-CoA synthase (ACS) and carbon monoxide dehydrogenase (CODH) (Fig. 3, lower right). In addition, the subunits of the corrinoid iron–sulphur protein (CoFeS) that binds cobamide also appear homologous in the methanogen–acetogen comparison.

CoFeS catalyzes two cobamide-dependent methyltransfer reactions, one from the pterin cofactor to the cobalt atom of the cobamide cofactor and a second from Co to the active-site Ni atom of ACS, in some cases aided by an additional, separate methyltransferase in acetogens [52,57,112,113]. The two CoFeS subunits of *Methanococcus maripaludis* [88] and *M. thermoacetica* [87] were scored as present in a target genome if the genome had best hits for both subunits, a condition fulfilled in many cases. CoFeS showed a similar phylogenetic distribution as ACS (EC 2.3.1.169) and CODH (EC 1.2.7.4) (Fig. 4), which were previously known to be homologous across the bacterial–archaeal divide based on sequence similarity [47,114]. In methanogens, CoFeS is often part of a larger CODH/ACS complex [107]. The CoFeS subunits of *Methanococcus maripaludis* and *M. thermoacetica* show 31.4% and 37.8% amino acid sequence identity in global pairwise comparisons (methanogenic CODH/ACS complex subunits delta and gamma, respectively). The sequence homology of acetogenic and methanogenic CoFeS was previously reported [14,42]. Their structural alignments using AlphaFold-modelled structures [103,105] (Fig. S2) showed TM-scores of 0.931 (methanogen CODH/ACS complex

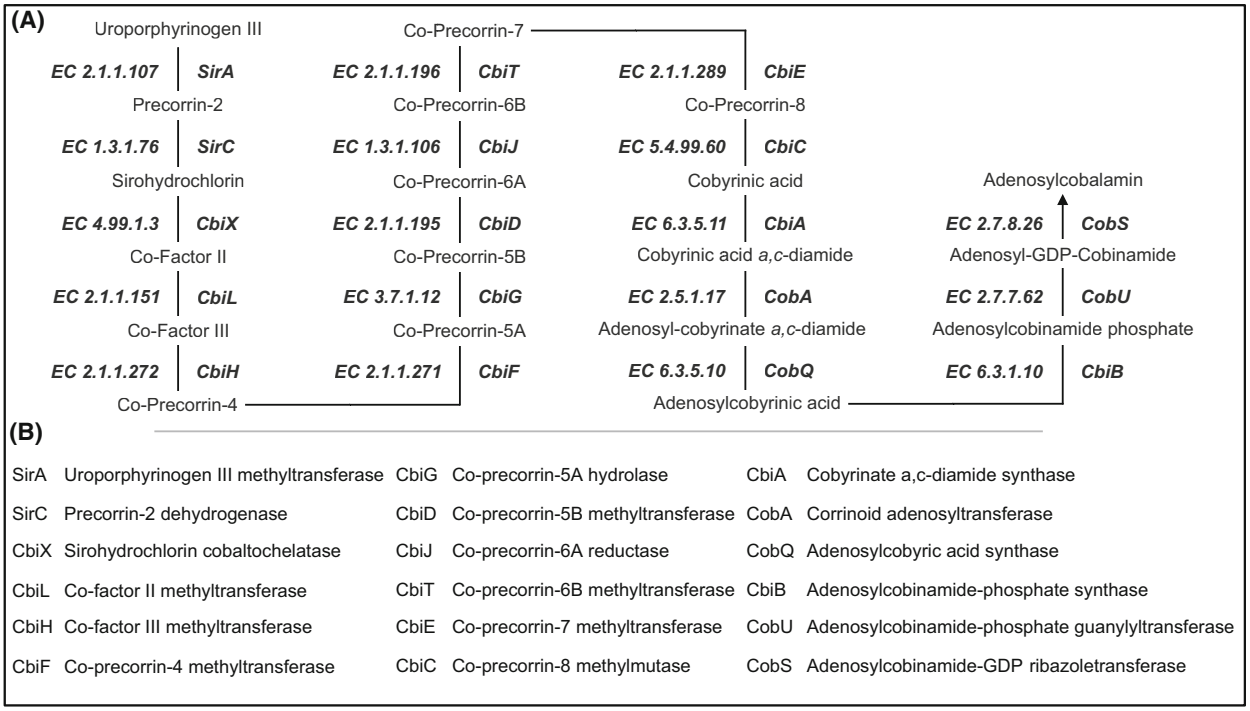
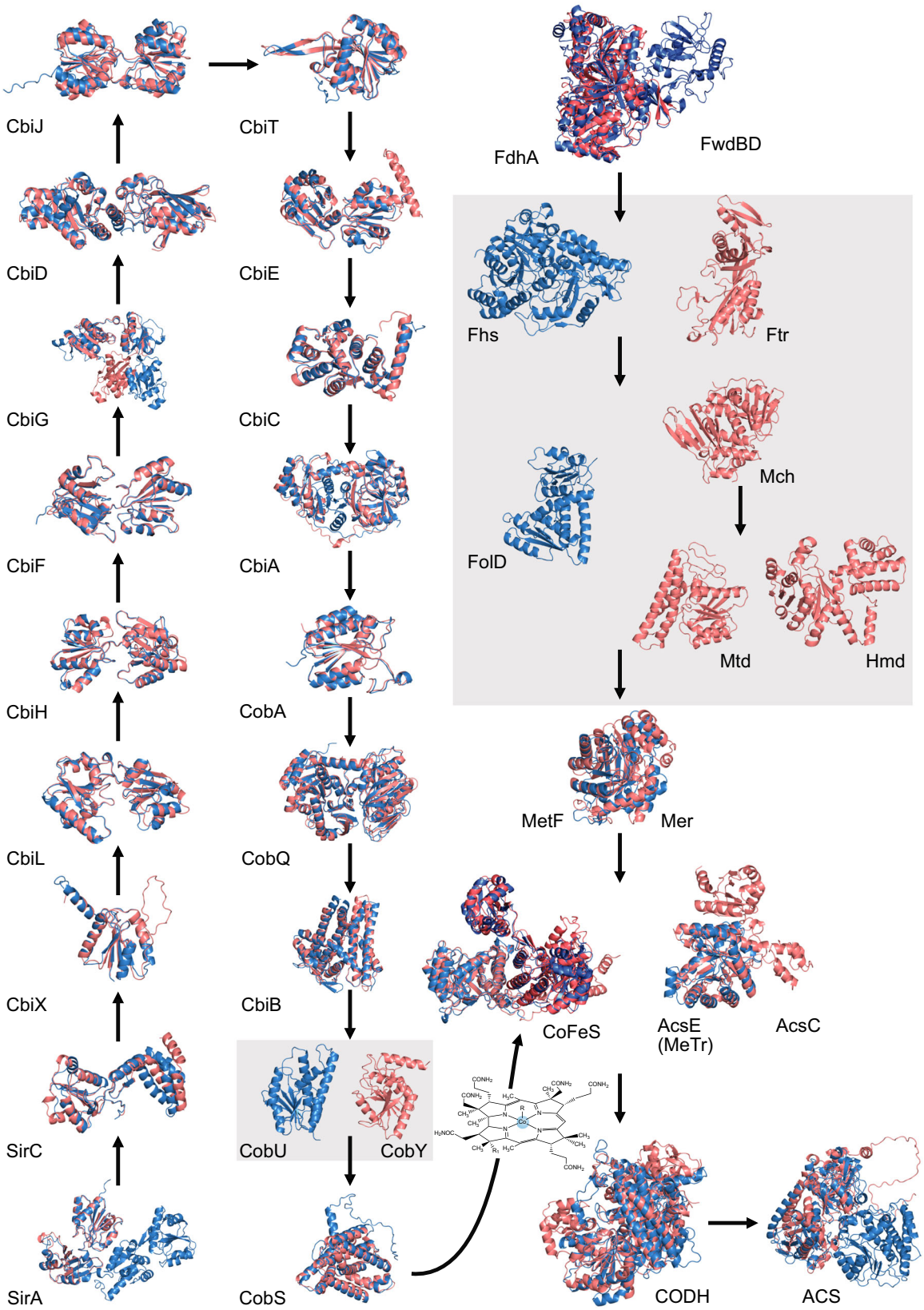


Fig. 2. The anaerobic cobalamin biosynthesis pathway. (A) The schematic shows the abbreviations and EC numbers of enzymes involved in the biosynthesis of adenosylcobalamin as found in the KEGG database, starting with the conversion of uroporphyrinogen III to precorrin-2 and the early insertion of cobalt. Lower ligand synthesis is shown in Fig. S1. (B) In previously published literature, other abbreviations like UroM [97] or SuMT [98] have been used to refer to uroporphyrinogen III methyltransferase (SirA). Abbreviations and full names of enzymes found in this article are displayed in the schematic. A short description of the steps and citations for the individual reactions is given in the Suppporting Information.

subunit delta, acetogen CoFeS small subunit) and 0.836 (methanogen CODH/ACS complex subunit gamma, acetogen CoFeS large subunit), normalized by the length of the respective acetogen protein. These scores indicate that the structures share a common fold, confirming their homology [115]. The methanogen corrinoid-iron sulphur protein CoFeS (AcsC) also shares a common fold with the methyl-H_4F:CoFeS methyltransferase (AcsE), which aids acetogen CoFeS in the methyl transfer step from the pterin to the corrin cobalt (Fig. 3), as previously reported [52].

The active corrins in the acetyl-CoA pathway of acetogens and methanogens accept methyl groups at the open Co coordination site. In adenosylcobalamin,

Fig. 3. Structurally homologous and divergent enzymes in anaerobic cobalamin synthesis and the acetyl-CoA pathway. Left panel: cobalamin synthesis pathway. Right panel: the acetyl-CoA pathway. Structures from acetogens (blue) and methanogens (salmon) are shown superimposed [103], while non-homologous structures are shown adjacent. Adjacent structures of the same colour point to non-homologous enzymatic variants in the same prokaryotic domain, such as in the case of Hmd. Hmd is the Fe-only hydrogenase of archaea that directly reduces methenyl-H_4MPT to methylene-H_4MPT, is expressed under nickel limitation, uses a unique iron-guanylylpyridinol cofactor and is not homologous to any enzymes of the pathways shown here [104]. Grey shading highlights non-homologous enzymes in methanogen-acetogen comparisons. Structures were modelled with AlphaFold [105], except for FwdBD (PDB ID: 5T5I [106]) and CODH/ACS (PDB IDs: 3CF4 and 1MJG [107,108]). Methanogen ACS and methyl synthesis were modelled based on sequences from *Methanothermobacter marburgensis*. The acetogen methyl branch is from *Moorella thermoacetica*. In cobalamin biosynthesis, methanogen sequences are from *Methanococcus maripaludis* (and *Methanosarcina barkeri* in the case of CobA, which is specific to adenosylcobalamin synthesis and missing in many methanogens; see text) and acetogen sequences from *M. thermoacetica*. In multimers, only the structurally homologous subunit was shown for clarity. TM-scores of structural alignments can be found in Table S1. For the nomenclature of enzymes involved in anaerobic cobalamin synthesis, see the legend and pannel (B) of Fig. 2. See also Svetlitchnaia *et al.* [52] for a structural comparison of CoFeS with structurally similar proteins.


CbiJ
CbiT
FdhA
FwdBD
CbiD
CbiE
Fhs
Ftr
CbiG
CbiC
Mch
FolD
CbiF
CbiA
Mtd
Hmd
CbiH
CobA
MetF
Mer
CbiL
CobQ
CoFeS
AcsE
(MeTr)
AcsC
CbiX
CbiB
SirC
CobU
CobY
SirA
CobS
CODH
ACS

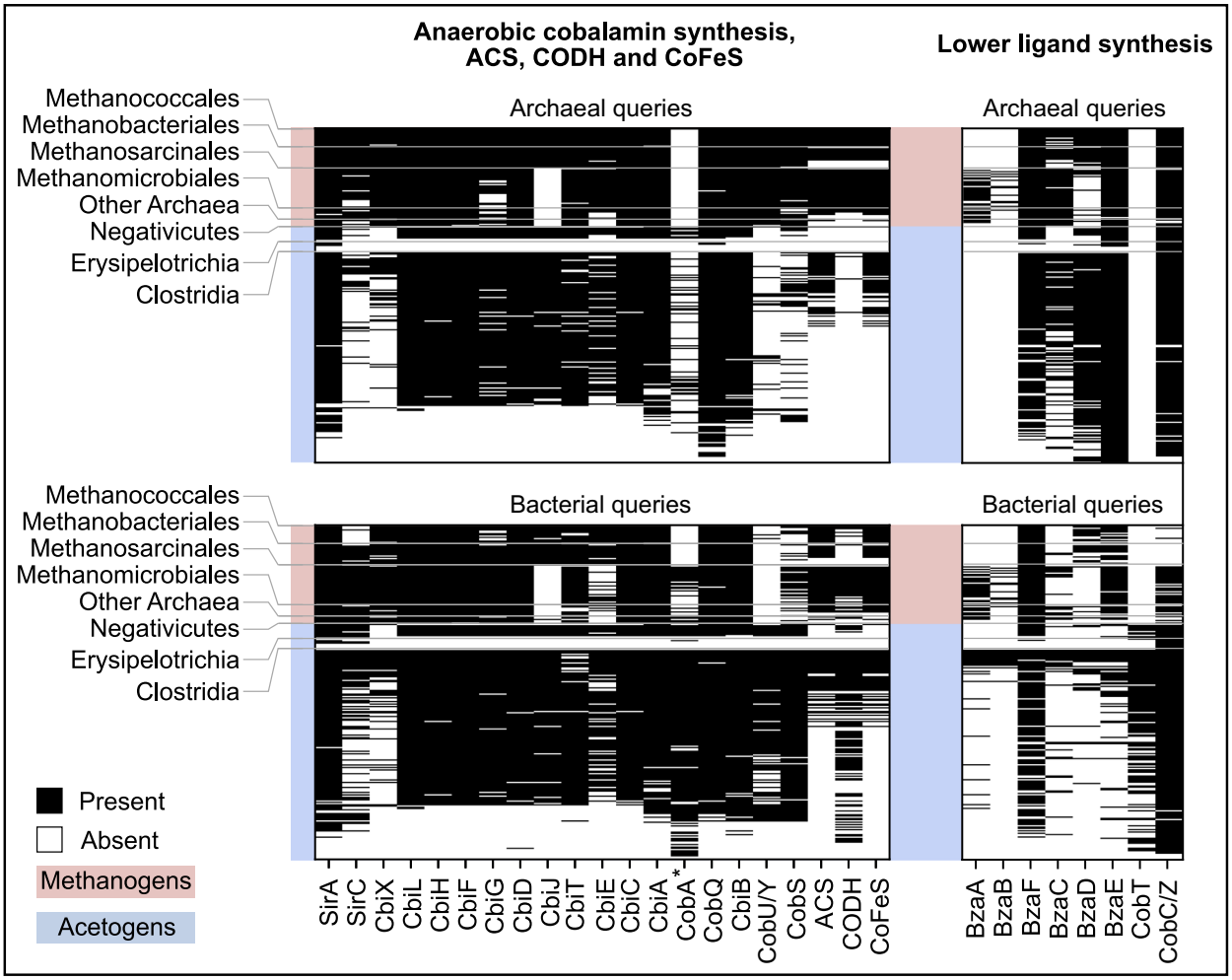


Fig. 4. Presence-absence pattern of enzymes of the anaerobic biosynthesis of cobalamin, acetyl-CoA synthase, carbon monoxide dehydrogenase, CoFeS and lower ligand synthesis enzymes in methanogens and acetogens. Best BLAST hits for all archaeal and bacterial query sequences (see Materials and methods) were scored in data sets of 80 methanogens and 189 acetogens. Analysed enzymes are displayed on the *X*-axis, and groups of methanogens (red) and acetogens (blue) are shown on the *Y*-axis. Black indicates presence and white indicates absence of the enzyme. The column (*) labelled "CobA" covers homologs of CobA, PduO and EutT. Nonorthologous alternatives of CobU (CobY) and CobC (CobZ) [36] were also included.

the central cobalt is coordinated by an upper ligand adenosyl moiety that is introduced by the corrinoid adenosyltransferase BtuR/CobA [116], which can also be replaced by two alternative enzymes, PduO [117] or EutT [118]. BtuR/CobA was generally missing in methanogen genomes from the *Methanococcales* (including our query species *Methanococcus maripaludis*, Fig. 4) and *Methanobacteriales*, which contain hydrogenotrophic species. This is consistent with a lack of adenosylcobalamin in several studied species [7]. CobU (adenosylcobinamide-phosphate guanylyltransferase) and its non-orthologous alternative CobY [36] lack structural homology across the methanogen-acetogen divide (Fig. 3) and, like CobS (adenosylcobinamide-GDP ribazoletransferase), are not widely distributed in reciprocal comparisons among acetogens and methanogens represented here (Fig. 4). Among the enzymes involved in the synthesis of the cobalamin lower ligand, only BzaF, which catalyzes the synthesis of 5-hydroxybenzimidazole (HBI) [39], is well conserved across the bacterial–archaeal divide, reflecting natural variation in lower ligand synthesis observed in methanogens and acetogens [7,8] relative to *Eubacterium limosum*, the species in which lower ligand synthesis was identified [39]. Archaeal and bacterial CobT proteins are not homologous, as previously reported [119]. The distribution of the genes for enzymes of corrin biosynthesis across

the bacterial–archaeal divide at the 25% amino acid sequence identity threshold is shown in Fig. 4.

Independent origins of enzymatic methyl synthesis in acetogens and methanogens

The methyl synthesis branch of the acetyl-CoA pathway provides a continuous supply of methyl groups for the CoFeS reaction in acetogen and methanogen carbon assimilation. The enzymes of methyl synthesis and those underpinning the cofactors synthesis, tetrahydrofolate (H_4F) in bacteria and tetrahydromethanopterin (H_4MPT) in archaea, were previously found to be non-homologous across the bacterial–archaeal divide [50] by the measure of amino acid sequence conservation. Because the methanogen enzyme formylmethanofuran dehydrogenase reveals structural homology to formate dehydrogenase of the acetogen pathway [116], which was not detected at the amino acid sequence level, we reinvestigated the potential common ancestry of acetogen and methanogen methyl synthesis pathways on the basis of structural homology. Identifiers for genes of the methyl branch of the acetyl-CoA pathway for the methanogen *Methanothermobacter marburgensis* and the acetogen *M. thermoacetica* were taken from previous studies [49,94]. Archaeal formylmethanofuran dehydrogenase (FwdBD) and bacterial formate dehydrogenase (FdhA) were not reinvestigated as structural similarity has already been shown [104]. Alignments of structures modelled with AlphaFold [103,105] showed that the reduction of the pterin-bound methylene to a methyl group is also catalyzed by enzymes that share a common fold, with a TM-score > 0.67 (Fig. 3). The electron donor for methylene-H_4MPT reductase in methanogens (Mer) is the cofactor F_{420}; the donor for the acetogen 5,10-methylene-H_4F reductase (MetF) is unclear [120].

The remaining two steps of methyl synthesis in acetogens, N^{10}-formyl-H_4F synthetase (Fhs) and 5,10-methenyl-H_4F cyclohydrolase/dehydrogenase (FolD), correspond to three steps in methanogens, catalyzed by formylmethanofuran:H_4MPT formyltransferase (Ftr), methenyl-H_4MPT cyclohydrolase (Mch) and F_{420}-dependent methylene-H_4MPT dehydrogenase (Mtd). These conversions are catalyzed by structurally non-homologous enzymes in acetogens and methanogens (TM-score < 0.35 in all cases) (Fig. 3). The non-homology of these methyl synthesis enzymes in acetogens and methanogens, their use of different redox cofactors at individual steps, in addition to the divergent syntheses of their corresponding pterin cofactors, indicates that the pathways arose independently in the acetogen and methanogen lineages, respectively [50]. This in turn suggests that the common ancestor of acetogens and methanogens was dependent upon a supply of chemically reactive methyl groups provided by the environment [47,50] and that the use of environmental methyl groups in acetyl and methane synthesis is more ancient than the pterin-dependent biochemical pathways used to synthesize the methyl groups. Such geochemically supplied reactive methyl groups could include methanol, a soluble substrate for the acetyl-CoA pathway in methanogens and acetogens [49] and a synthesis product from H_2 and CO_2 in geochemical analogues of the acetyl-CoA pathway [60], hydrothermal methyl sulphide [71,121], methyl thioacetate [122] or methyl groups bound to solid-state metal catalysts that synthesize methyl groups from H_2 and CO_2 [60,62–64].

The corrin synthesis pathways in acetogens and methanogens consist of homologous enzymes (Figs 3 and 4) that provide the cofactor for the corrin-binding methyltransferase, CoFeS, which is also homologous (Fig. S2), as are the enzymes of the acetyl synthesis branch of the acetyl-CoA pathway (CODH/ACS) (Fig. 3). In the methyl synthesis branch, the reduction of CO_2 to a formyl group (in methanogens) or formate (in acetogens) and the reduction of pterin-bound methylene to a methyl group are both steps catalyzed by structurally homologous enzymes across the acetogen-methanogen divide (Fig. 3, Fig. S2). Two steps of methyl synthesis are not homologous in the acetogen/methanogen comparison, indicating that these two steps "resulted from convergent developments" [48], after the divergence of the acetogen and methanogen lineages.

One might challenge the view that the acetyl-CoA pathway is the oldest pathway of CO_2 fixation [48–51]. However, the evidence in favour of its antiquity is unambiguous. There are seven known pathways of CO_2 fixation (the Calvin cycle, the reverse citric acid cycle, the acetyl-CoA pathway, the dicarboxylate/4-hydroxybutyrate cycle, the 3-hydroxypropionate/4-hydroxybutyrate cycle, the 3-hydroxypropionate bicycle and the glycine reductase pathway [123]), but only one of them, the acetyl-CoA pathway, occurs in both bacteria and archaea [49,124]. Moreover, only one of those seven, again only the acetyl CoA pathway, operates *in vitro* to form its main products – formate, acetate and pyruvate [49,90] – from H_2 and CO_2 without enzymes or cofactors, using only metals or minerals that occur in hydrothermal vents as catalysts [60–64,125]. None of the CO_2-fixing reactions of the other six pathways have been shown to operate *in vitro* without enzymes. In addition, the amount of pyruvate

formed via metal-catalyzed, non-enzymatic versions of the acetyl-CoA pathway can reach up to 200 μM [62], which is the physiological concentration of pyruvate in the cytosol of clostridial acetogens growing from H_2 and CO_2 via the acetyl-CoA pathway [126]. Finally, on the point of the antiquity of the acetyl CoA pathway, the same metal catalysts that allow it to unfold without ATP, enzymes, or cofactors convert 2-oxoacids to amino acids at room temperature [77], reduce NAD^+ [127] and reduce ferredoxin with H_2 [79]. The other six pathways consume ATP during CO_2 fixation [49]; the acetyl-CoA pathway is integral to ATP synthesis from H_2 and CO_2 in acetogens and methanogens [78,128]. These findings offer clear experimental evidence that the acetyl-CoA pathway is the most ancient among known CO_2-fixing pathways, and no comparable findings for any of the other six pathways have been reported.

Phylogenies trace corrin synthesis to the acetogen-methanogen common ancestor

To determine whether the corrin pathway of methanogens and acetogens traces to the last universal common ancestor, we constructed alignments and phylogenies based on genes for enzymes of anaerobic cobalamin biosynthesis and lower ligand biosynthesis (see Materials and methods) (Fig. 5). We recovered three classes of phylogenies: (a) bacterial–archaeal monophyly (nine cases: SirA, SirC, CbiX, CbiE, CbiB, CobS, BzaA, BzaB and BzaF), (b) bacterial–archaeal monophyly violated by a few recent transfers of bacterial genes into the same methanogen lineages or by transfers of archaeal genes into the acetogen lineage (seven cases: CbiJ, CbiT, CbiC, CbiA, CobA, CobQ and CobU/Y) and (c) highly interleaved phylogenies suggesting multiple LGT events across the bacterial–archaeal divide (nine cases: CbiL, CbiH, CbiF, CbiG, CbiD, BzaC, BzaD, BzaE and CobC).

The nine cases of domain monophyly trace anaerobic corrin and lower ligand synthesis to the common ancestor of bacteria and archaea. Among those, SirA and SirC are involved in siroheme biosynthesis as well, a cofactor also identified in Weiss *et al.* [47] as being present in LUCA. The seven cases of rarely violated monophyly also trace anaerobic corrin and lower ligand synthesis to the common ancestor of bacteria and archaea, yet with phylogenies punctuated by late lateral gene transfer events, indicated with asterisks in Fig. 5. In some cases, these transfers involved the import of multiple genes from acetogens into the same methanogen species. For instance, the genes of CbiJ, CbiT and CbiC were transferred from acetogens into *Methanomassiliicoccus intestinalis* (GCF_000404225.1), *Methanoplasma termitum* (GCF_000800805.1), *Methanomethylophilus alvus* (GCF_000300255.2) and methanogenic archaeon ISO4-H5 (GCF_001560915.1). A common ancestor of these four archaea acquired the bacterial corrin synthesis genes *cbiJ*, *cbiT* and *cbiC*, apparently in a single event, because the four archaea in question share a common ancestor in a species tree constructed from a concatenated alignment of 23 prokaryotic informational core genes (ICGs) (Fig. S3), constructed mainly from ribosomal proteins and aminoacyl-tRNA synthetases (Table S2). In addition, a deep separation between archaeal and bacterial CbiA/CobB proteins was also recently observed in a phylogeny comprising homologous amidases of the different tetrapyrroles present in 2070 assemblies [129], where, besides CbiA/CobB, the CfbB protein, involved in the amidation of the F_{430} cofactor of methyl-coenzyme M reductase [100,130], and the DsrN protein [131], responsible for the amidation of the siroheme cofactor present in DsrAB proteins [132], are present. This phylogeny clearly shows that the methanogenic CfbB proteins and the DsrN proteins derived from archaeal CbiA/CobB proteins and that archaeal and bacterial CbiA/CobB proteins form two large clades separated by the root. This indicates not only the presence of this protein in LUCA but also the ancestry of CbiA/CobA in relation to F_{430}. A closer inspection of the phylogeny also indicated events of interdomain CbiA/CobA transfers with 11 archaeal sequences found in four small clades inside a clade of 1879 bacterial CbiA/CobB proteins. If indeed the proteins trace back to LUCA, it is expected that, during the almost 4 billion years of evolution, several lateral gene transfers would have occurred, leading to possible synonymous replacements of the original gene by the newly acquired one, as observed in other cases [129,133,134]. Moreover, studies have indicated that an average of 42.5% of genes per species is affected by lateral gene transfer when considering recent transfers [135], and this number raises up to 81% when the oldest transfers are included [136]. Thus, cases in which only a few HGTs are identified can be considered as present in LUCA. Putative LGT events are highlighted in Fig. S3.

The remaining nine trees showed abundant interleaving of bacterial and archaeal corrin synthesis genes, indicating that the homologous genes are, in terms of structure and function, freely interchangeable across the bacterial–archaeal divide. This indicates in turn (a) conservation of function since their divergence from a common ancestral gene and (b) the presence and function of a member of that gene family in the

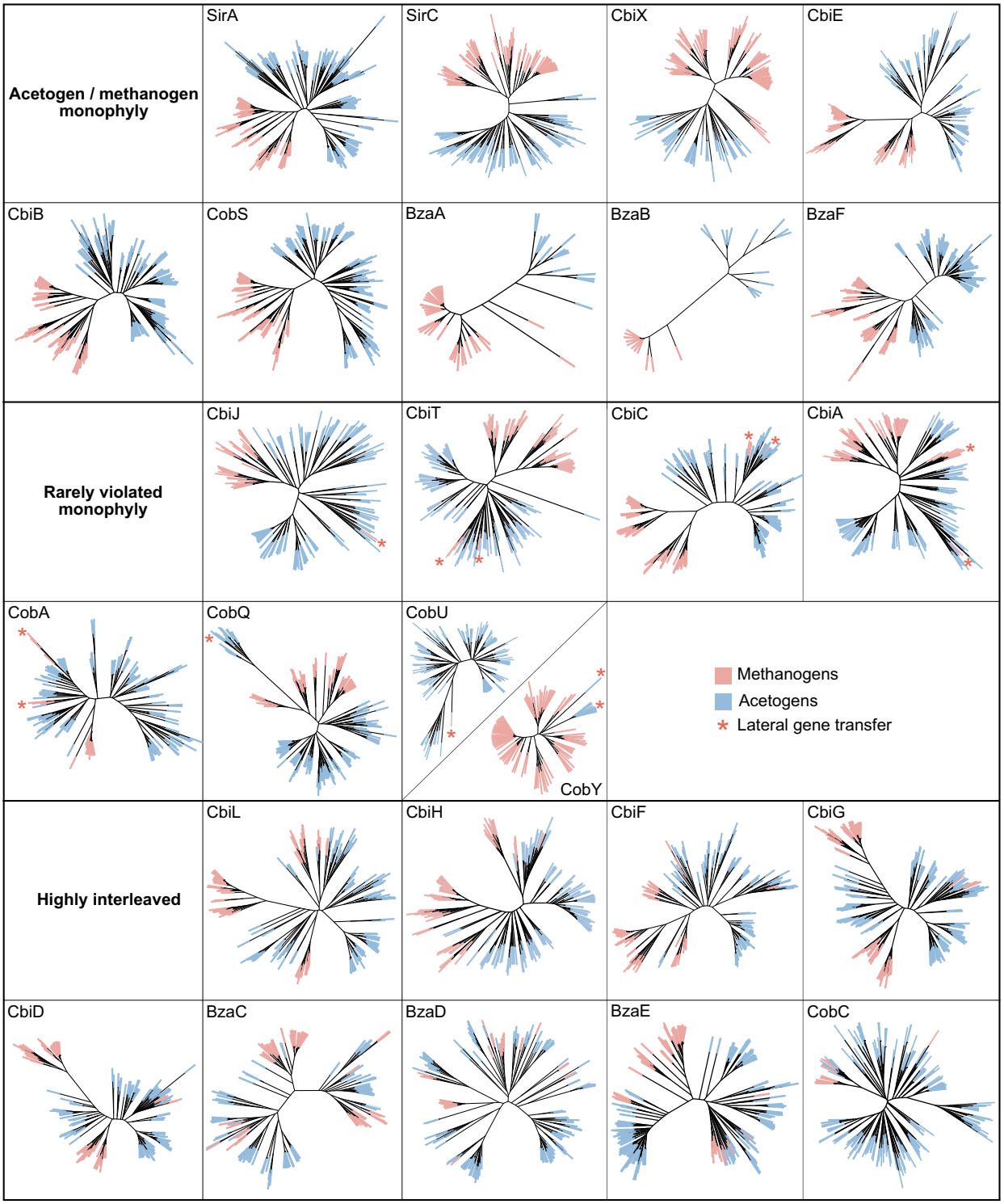


Fig. 5. Unrooted gene trees of enzymes involved in the anaerobic biosynthesis of cobalamin and lower ligand synthesis. The overview shows single-gene trees containing sequences of methanogens (red) and acetogens (blue). Calculations of trees are based on the best BLAST hits of acetogen and methanogen query sequences against genome data sets of methanogens and acetogens. The phylogenies were computed with RAXML version 8.2.12 and are grouped according to their category (monophyly, rarely violated monophyly, highly interleaved) and arranged corresponding to the order displayed in Fig. 2.

bacterial–archaeal common ancestor. Despite the interleaving of methanogens and acetogens in individual trees, the concatenated tree of corrin biosynthesis enzymes recovers domain monophyly (Fig. S4). Given the possible non-specificity of reactions of primordial enzymes, it can be envisaged that CbiE (tracing to LUCA) could have replaced (in LUCA) some of the homologous proteins that show intercalated archaeal-bacterial phylogenies, such as CbiF, CbiH and CbiL. Phylogeny and functional equivalence indicate the presence of an almost complete corrin pathway in the bacterial–archaeal common ancestor with a partial lower ligand biosynthesis.

Cobamide synthesis, a hub in biochemical pathway evolution

None of the intermediates in the 26-enzyme cobamide synthesis pathway are known to serve a biochemical function beyond their role as a pathway intermediate. This would present an evolutionary paradox, because all individual enzymes of the entire pathway would have to be functional before selection could set in – unless, of course, the domains of cobamide synthesis enzymes were recruited from simpler preexisting pathways. To see which pathways might have contributed genes to corrin evolution, we took three approaches.

First, a DIAMOND BLASTP search of cobalamin biosynthesis proteins against a dataset of 404 autotrophic core reactions [137] uncovered homologues of CbiA in F_{430} biosynthesis, homologues of CbiA and CobQ in biotin biosynthesis, and homologues of CbiG and CobQ in histidine metabolism (Table S3). Second, we mapped the sequences of the 355 protein clusters from Weiss *et al.* [47] to InterPro and compared the SUPERFAMILY, TIGR and PFAM assignments with those obtained for cobalamin biosynthesis proteins. We found that the ancient 'P-loop containing nucleoside triphosphate hydrolase' domain, which is the only domain of CobU and BtuR/CobA proteins, is present in 37 different protein families from Weiss *et al.* [47] belonging to seven different functional categories (Fig. 6). CbiT belongs to the large superfamily of *S*-adenosyl-L-methionine-dependent (SAM-dependent) methyltransferases (SSF53335) and is homologous with five protein families of Weiss *et al.* [47], including three involved in RNA modifications. CbiE shares homology with the remaining cobalamin methylases (SirA, CbiF, CbiH and CbiL), indicating ancient diversification of the family. CobQ is a member of the Class I glutamine amidotransferase-like superfamily, which includes a domain of CbiA; the second, nucleotide-binding domain of CbiA is also present in the data of Weiss *et al.* [47] in a cluster annotated as MinD. In total, 12 out of 18 cobalamin biosynthesis proteins belong to protein families that trace to LUCA [47], indicating that their ancestral module was already present in LUCA (Table S4).

Next, we looked for domains shared between cobalamin biosynthesis and domains present in 3939 sequenced genomes of cultured organisms using InterPro domain scans (Fig. 6). This revealed homologues of genes for anaerobic cobalamin biosynthesis in additional ancient pathways. Homologues of cobalamin biosynthesis are involved in amino acid synthesis (tyrosine and histidine) and in cofactor biosynthesis pathways for molybdopterin (MoCo), pyridoxal phosphate and biotin. The class III tetrapyrrole methylases CbiE, CbiF, CbiH, CbiL and SirA are homologous to diphthine synthase, 16S rRNA (cytidine(1402)-2′-O)-methyltransferase and MazG proteins. The class I methyltransferase CbiT shares homology with a large number of SAM-dependent RNA-modifying methyltransferases, which is in accordance with the literature [138]. SirC contains a very common NAD(P)-binding Rossmann-fold domain (Table S5). CobQ is related to BioD (biotin pathway), an enzyme responsible for ring formation [139] and homologous to MinD, CobB and ParA proteins [140]. No homologs were identified for CbiG, CbiJ, CbiD and CbiB. In the pathway for synthesis of the lower ligand, BzaA, BzaB and BzaF are homologous to ThiC, which catalyzes the first committed step of thiamine biosynthesis [141]. Other homologues included enzymes involved in the synthesis of F_{430}, a Ni-containing tetrapyrrole specific to the methane-forming step of methanogenesis [142] whose synthesis shares common steps with cobamide biosynthesis [38].

The shared homologies of *cob* enzymes reflect links to cofactor synthesis, amino acid synthesis and RNA modification (Fig. 6), functions that were clearly present in LUCA [47,137], indicating that the 26-step cobamide synthesis pathway was assembled from protein domains involved in ancient preexisting pathways. Cobalamin synthesis genes were, however, also recruited later in evolution, for example, in chlorophyll biosynthesis (BchE, BchR and BchQ) of photosynthetic bacteria [38]. In addition, all tetrapyrrole class II chelatases derive from a common ancestor [143,144], that evolved through duplication to give rise to $CbiX_L$, CbiK, HemH (PpfC/CpfC) and SirB (ShfC). Both *Methanococcus maripaludis* and *M. thermoacetica* use the short version of the enzyme ($CbiX_s$), having retained the primordial class II tetrapyrrole chelatase and further underscoring the common ancestry of the pathways in these organisms. Cobamide biosynthesis

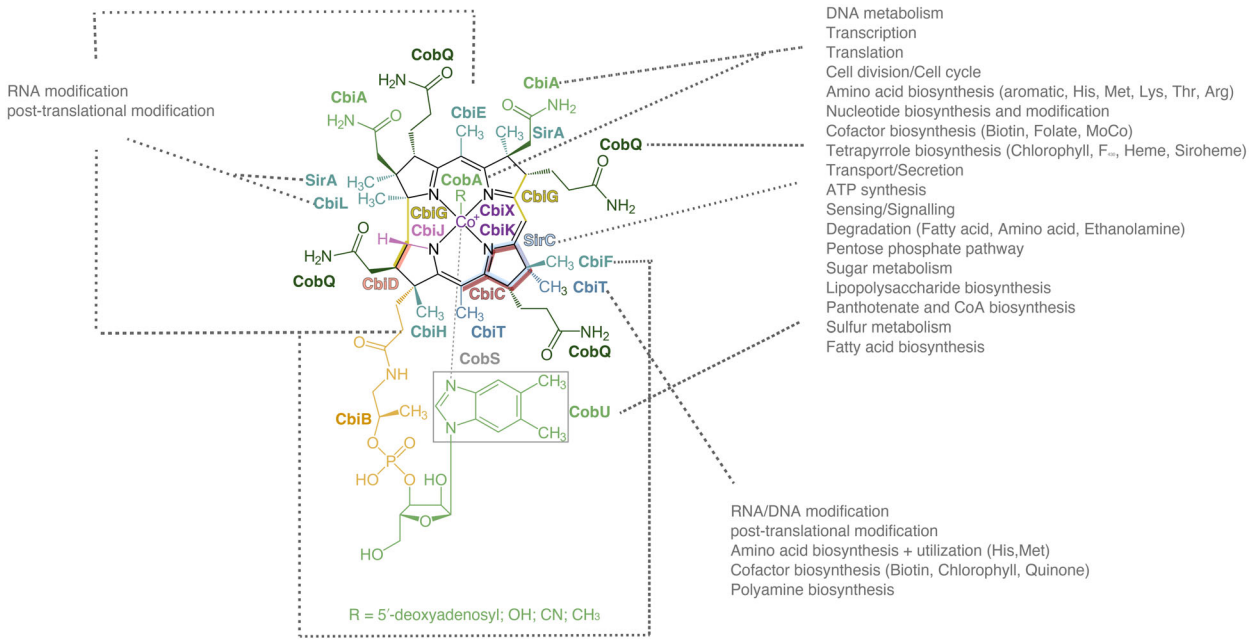


Fig. 6. Homologs of cobalamin biosynthesis enzymes in other pathways. The chemical structure of adenosylcobalamin is coloured according to the enzymes that are involved in the synthesis of the respective region of the cobalamin structure. Dashed arrows point to KEGG pathway categories linked to identified homologs of the corresponding enzymes.

represents a hub of functional domains that was assembled early in the evolution of biochemical pathways and tapped for useful activities during lineage divergence.

The acetogen–methanogen common ancestor and LUCA

The data indicate that corrin biosynthesis was present in the acetogen-methanogen common ancestor. This traces it to LUCA, because there are only three possibilities concerning the relationship between the common ancestor of bacteria and archaea and LUCA. (a) Bacteria arose from a specific lineage of archaea, making LUCA an archaeon. (b) Archaea arose from a specific lineage of bacteria, making LUCA a bacterium. (c) Neither domain arose from the other, such that LUCA was the common ancestor of bacteria and archaea sampled here. No current molecular phylogenetic studies support an origin of one prokaryotic domain from the other or one member of one prokaryotic domain from the other. Though debates exist about LUCA's organizational state [145,146], very few studies directly address the physiology of LUCA [47] and numerous studies support possibility 3 [47,147,148], which, given the homologies (Fig. 3) and phylogenies (Fig. 5) presented here, would place corrin biosynthesis in the common ancestor of bacteria and archaea, hence in LUCA.

The presence of corrin synthesis in LUCA contrasts with the lack of homologous methyl synthesis branches across the bacterial and archaeal versions of the acetyl-CoA pathway that delivers methyl groups for CoFeS (Fig. 3), suggesting that LUCA utilized geochemically supplied methyl groups [47,50]. A recent study reported that LUCA possessed a nearly complete acetyl-CoA pathway [149], although the genes they identified (in their supplemental information) were specifically for the archaeal acetyl-CoA pathway and were often present in fewer than 10% of the bacterial plus bacterial-metagenomic lineages sampled, indicating transfer of the archaeal genes to bacteria, rather than presence of the archaeal pathway in LUCA, which their method inferred. Many such archaea-to-bacteria transfers are known for the archaeal methyl synthesis pathway, starting with the O_2-dependent bacterial methylotroph *Methylobacterium extorquens*, where genes for the entire methyl synthesis pathway and methanopterin cofactor synthesis were acquired by transfer from archaea [150]. Several other studies uncovered further examples in which individual genes of the archaeal methyl synthesis and methanopterin pathways had been transferred from archaea to bacterial recipients [50,151–154], which are

often O_2-dependent methylotrophs and methanotrophs. Lateral gene transfer is a general hurdle to identifying genes that were present in LUCA [47]. The reactions of methyl synthesis in Fig. 3 are reversible. The presence of homologous enzymes generating formate (FdhA, FwdBD) could reflect formate utilization (generation of CO_2 and reductant) as reductant in early serpentinizing environments [70]; the presence of homologous enzymes for synthesis of methylene-pterins (MetF, Mer) (Fig. 3) could reflect methyl-dependent generation of methylene-H_4F and methylene-H_4MPT as C1 units for biosyntheses.

Implications for early biochemical evolution

Earlier gene-based investigations of corrin biosynthesis evolution [80,155–157] have missed the role of corrinoids in CoFeS and the acetyl-CoA pathway, even though their function in ancient pathways has long been evident [7,8,48,49]. The present findings trace corrinoids to the earliest phases of biochemical evolution, to a time when some essential biochemical functions were performed by proteins, some were performed by cofactors alone and some were still performed by solid-state metals [62,64,77,90]. Isotope data trace the ultralight isotope signature of the acetyl-CoA pathway [68] back to > 3.5 Ga [65,66], consistent with the structural (Fig. 3) and phylogenetic (Fig. 5) evidence for the age of corrins.

The present data suggest a sequence of events as outlined in Fig. 7. Under the conditions of serpentinizing hydrothermal vents, transition metals alone convert H_2 and CO_2 into formate, acetate and pyruvate in water [60,62,63,158]. Methyl synthesis and transfer can occur on solid-state catalyst surfaces (Fig. 7A). The exergonic nature of carbon and energy metabolism from H_2-dependent CO_2 fixation can, in principle, energetically support the origin of genes and proteins with the help of substrate-level phosphorylation [71,159]. Proteins, once present, can organize Fe and Ni into active sites through ligation to cysteine sulfhydryls and perform the functions of ferredoxin [79], CODH and ACS [60,61], such that acetyl and C≡O synthesis become independent of the environmental surface but remain tied to solid-state catalysts for methyl group supply from H_2 and CO_2 (Fig. 7B). In this stable intermediate configuration, enzymatic carbon metabolism remains tethered to catalysts on the Earth's crust. More ancient than enzymatic methyl synthesis, corrin synthesis could supply soluble and diffusible Co-bound methyl groups, freeing ACS and acetyl synthesis from contact with the solid state and allowing acetyl synthesis to migrate to the aqueous phase (Fig. 7C). Pterin-dependent enzymatic methyl synthesis from H_2 and CO_2 freed methyl synthesis from solid-state catalysts, involving folate in the lineage leading to acetogens and methanopterin in the lineage leading to methanogens [50] (Fig. 7D). This required the origin of additional complexes essential for methanogens and acetogens, namely, hydrogenases and flavin-based electron bifurcation [78], the precursor of which was Fe^0-dependent ferredoxin reduction with H_2 under hydrothermal conditions [79].

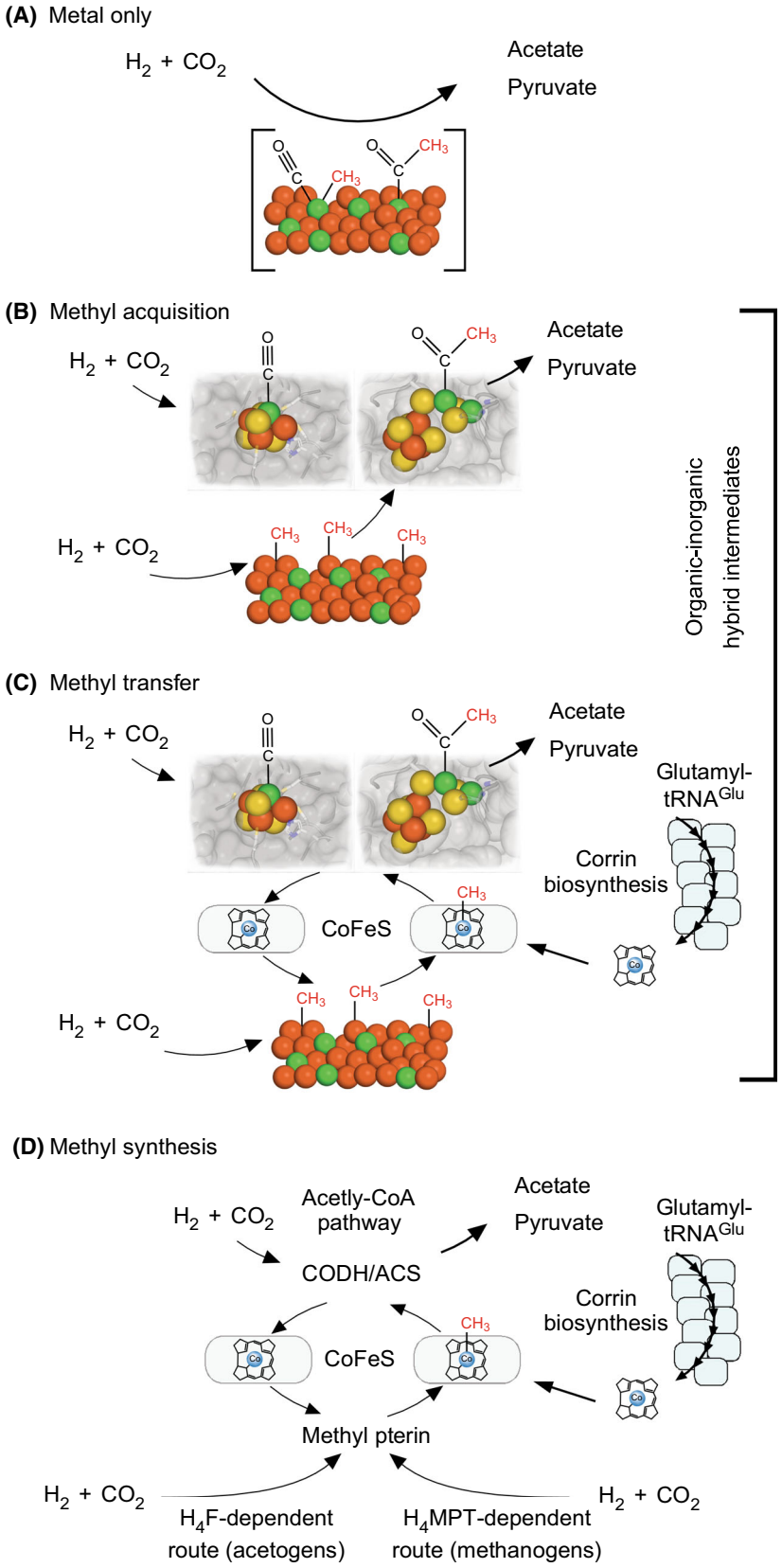
A critical observer might maintain that even enzymatic corrin synthesis is 'just too complicated' to be ancient [80], but the argument fails. Why? Corrins are far less complicated than ribosomes, with 1/2000th the molecular mass [160], and ribosome synthesis had to be in place before any of the homologous protein-coding genes for corrin biosynthesis arose. Complexity is not an argument against antiquity *per se*, as the ribosome attests. At biochemical origins, complex enzymatic pathways require functional catalytic precursors. The present findings establish a catalytic continuity between natural geochemical environments [75] and the requirement for Fe, Ni and Co in enzymes and corrin cofactors of the acetyl-CoA pathway, which might foster the identification of additional intermediate states in early biochemical evolution.

Materials and methods

Data

A dataset of 80 genomes of methanogens and 189 genomes of acetogens from the Reference Sequence database

Fig. 7. Early phases of biochemical evolution in serpentinizing hydrothermal vents. (A) Transition metals alone convert H_2 and CO_2 into acetate and pyruvate; methyl synthesis and transfer occur on the solid-state catalyst surface. (B) Still tied to solid-state catalysts for methyl supply, proteins incorporate Fe- and Ni-atoms into their active sites, allowing the synthesis of carbon monoxide and acetyl to become soluble and independent of the environmental surface. (C) The origin of corrin synthesis generates a supply of mobile, cobalt-bound methyl groups, releasing ACS from the solid state and enabling acetyl synthesis to migrate to the aqueous phase. (D) H_2-dependent methyl formation from CO_2 also migrated to the aqueous phase with the origin of enzymatic methyl synthesis after the archaeal–bacterial divergence.


(A) Metal only
H2 + CO2
Acetate
Pyruvate
(B) Methyl acquisition
H2 + CO2
Acetate
Pyruvate
H2 + CO2
Organic-inorganic hybrid intermediates
(C) Methyl transfer
H2 + CO2
Acetate
Pyruvate
Glutamyl-tRNAGlu
Corrin biosynthesis
CoFeS
H2 + CO2
(D) Methyl synthesis
H2 + CO2
Acetly-CoA pathway
CODH/ACS
Acetate
Pyruvate
Glutamyl-tRNAGlu
Corrin biosynthesis
CoFeS
Methyl pterin
H2 + CO2
H4F-dependent route (acetogens)
H4MPT-dependent route (methanogens)
H2 + CO2

(September 2016) [161] was analyzed. The methanogens belong to five archaeal orders (including "Other Archaea"), and three classes of acetogens are represented. A taxonomy report can be found in Table S6. Sequences of enzymes involved in the anaerobic biosynthesis of cobalamin (Fig. 2; based on [38]), as well as acetyl-CoA synthase (ACS), carbon monoxide dehydrogenase (CODH) and CoFeS of *M. thermoacetica* (acetogen representative) and *Methanococcus maripaludis* (methanogen representative), were acquired from the KEGG (Kyoto Encyclopedia of Genes and Genomes) [162] database in June 2023 (KEGG organisms *mtho*/*mta* and *mmp*, respectively). *M. thermoacetica* sirohydrochlorin cobaltochelatase CbiX was downloaded from GenBank (TYL17187.1). Acetogen sequences of enzymes involved in lower ligand synthesis were also acquired from the Reference Sequence database and GenBank. Methanogen query sequences were selected from BLAST results against the acetogen set of sequences. Sequences of enzymes involved in the methyl synthesis in *M. thermoacetica* [94] were acquired from the IMG/M database [163] and UniProt [164]. For *Methanothermobacter marburgensis*, sequences (as described in Ref. [49]) were downloaded from GenBank. A dataset of 3939 genomes downloaded from NCBI (November 2019) was analyzed for the presence of homologues of the biosynthesis of cobalamin but excluding the aerobic cobalamin biosynthesis pathway. All analyzed sequences, including the CobA alternatives PduO/EutT and the corresponding organisms, are listed in Table S7.

Sequence homology search in methanogens and acetogens

The representative archaeal and bacterial sequences were scored in methanogens and acetogens with DIAMOND version 2.0.1.139 [165,166]. The homology search was performed using an expectation value threshold of 1×10^{-10} and an identity threshold of 25%. For each enzyme, the best sequence hit was selected per genome for further analyses. INTERPROSCAN (version 5.66-9.8) [167] was run on the retrieved hits to obtain further functional annotations. Global sequence alignments of methanogen and acetogen CoFeS subunits were calculated based on the Needleman-Wunsch algorithm [168].

Synteny analysis

Genomes containing best-hit sequences were used for the synteny analysis, which was used in phylogenetic analysis. The best hits were mapped to their gene location files (feature tables), and a syntenic neighbourhood was defined using a window of two upstream and two downstream proteins in the neighbourhood of the proteins of interest (same chromosome). The identified syntenic neighbourhood for each best hit was extracted and mapped onto the phylogenies with the corresponding accessions, protein types, PFAM annotations and gene descriptions for a detailed phylogenetic analysis and better distinction of paralogous sequences.

Expanded sequence homology searches

DIAMOND (version 2.18) was also run against the dataset from Wimmer *et al.* [137] with an identity cutoff of 25% and *E*-value cutoffs of 10^{-10}, 10^{-8} and 10^{-5}, and the best hits were further analyzed. In addition, INTERPROSCAN (version 5.66-9.8) [167] was run on the dataset of 3939 genomes and on the sequences from the 355 protein families mapped to LUCA from Weiss *et al.* [47]. All hits that shared the same SUPERFAMILY annotations as cobalamin biosynthesis enzymes were further analysed.

Computation of multiple sequence alignments and phylogenetic trees

Gene trees containing sequences of methanogens and acetogens for 26 analysed enzymes involved in the anaerobic biosynthesis of cobalamin and lower ligand synthesis were computed based on the best hit selection of the homology search or from the synteny analysis. For that, multiple sequence alignments were made using MAFFT L-INS-I version 7.471 [169] and unrooted gene trees were reconstructed with RAXML version 8.2.12 [170] using the PROTCATWAG model and a random starting seed (*12345*). The same pipeline was used to compute a concatenated tree of methanogens and acetogens with best hits for ≥ 17 enzymes. Single sequence alignments of the analysed enzymes were combined to obtain a large alignment for tree reconstruction. MAFFT and RAXML were used to calculate a species tree of all 80 methanogens present in the dataset based on a concatenated alignment combining sequences of prokaryotic informational core gene families (ICGs; [93]), which were determined by using the verticality value [171], the number of genomes and the taxonomic distribution of a protein family. All 80 methanogens are represented by at least one sequence in 23 different prokaryotic informational core gene families (Table S2). The trees were visually modified with the use of the INTERACTIVE TREE OF LIFE [172].

Structural alignments of enzymes

In order to compare the structures of methanogenic and acetogenic enzymes involved in anaerobic cobalamin synthesis and methyl synthesis in the acetyl-CoA pathway, structures were either obtained from the Protein Data Bank (PDB, [173]) or modelled using AlphaFold version 2.2.0 [105]. Structural alignments and TM-scores were determined using US-ALIGN version 20230609 [103]. Alignments were visually inspected, and figures were prepared using PYMOL version 2.5.7 (The PyMOL Molecular Graphics

System, Version 2.5.7; Schrödinger, LLC (http://www.pymol.org/pymol)]). In those cases when the structures overlapped over one or more domains, but not the full-length protein (usually reflected in significant differences in the TM scores normalized by the lengths of the two proteins), the overlapping domains were separately aligned to obtain a more accurate TM-score. This allowed us to determine structural homology for the superimposed domains, in line with the modular nature of protein evolution.

Acknowledgements

We thank Katharina Trost for calculating and providing informational core genes. Computational infrastructure and support were provided by the Centre for Information and Media Technology at Heinrich Heine University Düsseldorf and by the Life Science Computer Cluster (LiSC) at the University of Vienna. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 101018894). For funding, WFM thanks the ERC (101018894), the Deutsche Forschungsgemeinschaft (MA 1426/21-1) and the Volkswagen Foundation (Grant 96_742), For funding, FLS thanks the Wiener Wissenschafts-, Forschungs- und Technologiefonds (grant agreement VRG15-007) and the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation program (grant agreement 803768).

Conflict of interest

The authors declare no conflict of interest.

Author contributions

WFM designed research; LDM, VK, MK, NM, WFM and FLS performed research; LDM, VK, MK, NM, WFM and FLS analyzed data; and LDM, VK, MK, NM, WFM and FLS wrote the paper.

Data availability statement

Alignments, phylogenies, commands, synteny script and the dataset of 3939 genomes are available at Figshare: https://doi.org/10.6084/m9.figshare.25765518.

References

1 Matthews RG, Koutmos M & Datta S (2008) Cobalamin-dependent and cobamide-dependent methyltransferases. *Curr Opin Struct Biol* **18**, 658–666.
2 Crofts TS, Seth EC, Hazra AB & Taga ME (2013) Cobamide structure depends on both lower ligand availability and CobT substrate specificity. *Chem Biol* **20**, 1265–1274.
3 Cheong CG, Escalante-Semerena JC & Rayment I (2001) Structural investigation of the biosynthesis of alternative lower ligands for cobamides by nicotinate mononucleotide: 5,6-dimethylbenzimidazole phosphoribosyltransferase from *Salmonella enterica*. *J Biol Chem* **276**, 37612–37620.
4 Rondon MR, Trzebiatowski JR & Escalante-Semerena JC (1997) Biochemistry and molecular genetics of cobalamin biosynthesis. *Prog Nucleic Acid Res Mol Biol* **56**, 347–384.
5 Renz P, Blickle S & Friedrich W (1987) Two new vitamin B-12 factors from sewage sludge containing 2-methylsulfinyladenine or 2-methylsulfonyladenine as base component. *Eur J Biochem* **163**, 175–179.
6 Kräutler B, Kohler HPE & Stupperich E (1988) 5′-Methylbenzimidazolyl-cobamides are the corrinoids from some sulfate-reducing and sulfur-metabolizing bacteria. *Eur J Biochem* **176**, 461–469.
7 Stupperich E & Kräutler B (1988) Pseudo vitamin B_{12} or 5-hydroxybenzimidazolyl-cobamide are the corrinoids found in methanogenic bacteria. *Arch Microbiol* **149**, 268–271.
8 Stupperich E, Eisinger HJ & Kräutler B (1988) Diversity of corrinoids in acetogenic bacteria. P-cresolylcobamide from *Sporomusa ovata*, 5-methoxy-6-methylbenzimidazolylcobamide from *Clostridium formicoaceticum* and vitamin B_{12} from *Acetobacterium woodii*. *Eur J Biochem* **172**, 459–464.
9 Mathur Y & Hazra AB (2022) Methylations in vitamin B_{12} biosynthesis and catalysis. *Curr Opin Struct Biol* **77**, 102490.
10 Gruber K, Puffer B & Kräutler B (2011) Vitamin B_{12}-derivatives—enzyme cofactors and ligands of proteins and nucleic acids. *Chem Soc Rev* **40**, 4346–4363.
11 Buckel W & Golding B (1996) Glutamate and 2-methyleneglutarate mutase: from microbial curiosities to paradigms for coenzyme B_{12}-dependent enzymes. *Chem Soc Rev* **25**, 329–337.
12 Jordan A, Torrents E, Jeanson C, Eliasson R, Hellman U, Wernstedt C, Barbé J, Gibert I & Reichard P (1997) B_{12}-dependent ribonucleotide reductases from deeply rooted eubacteria are structurally related to the aerobic enzyme from *Escherichia coli*. *Proc Natl Acad Sci USA* **94**, 13487–13492.
13 Harder J (1993) Ribonucleotide reductases and their occurrence in microorganisms: a link to the RNA/DNA transition. *FEMS Microbiol Rev* **12**, 273–292.
14 Tauer A & Benner SA (1997) The B12-dependent ribonucleotide reductase from the archaebacterium *Thermoplasma acidophila*: an evolutionary solution to

the ribonucleotide reductase conundrum. *Proc Natl Acad Sci USA* **94**, 53–58.
15 Banerjee RV & Matthews RG (1990) Cobalamin-dependent methionine synthase. *FASEB J* **4**, 1450–1459.
16 Matthews RG (2009) Cobalamin- and corrinoid-dependent enzymes. In Metal-Carbon Bonds in Enzymes and Cofactors (Sigel A, Sigel H & Sigel RKO, eds), pp. 53–114. The Royal Society of Chemistry, London.
17 Dowling DP, Miles ZD, Köhrer C, Maiocco SJ, Elliott SJ, Bandarian V & Drennan CL (2016) Molecular basis of cobalamin-dependent RNA modification. *Nucleic Acids Res* **44**, 9965–9976.
18 Gough SP, Petersen BO & Duus JO (2000) Anaerobic chlorophyll isocyclic ring formation in *Rhodobacter capsulatus* requires a cobalamin cofactor. *Proc Natl Acad Sci USA* **97**, 6908–6913.
19 Schmerk CL, Welander PV, Hamad MA, Bain KL, Bernards MA, Summons RE & Valvano MA (2015) Elucidation of the *Burkholderia cenocepacia* hopanoid biosynthesis pathway uncovers functions for conserved proteins in hopanoid-producing bacteria. *Environ Microbiol* **17**, 735–750.
20 Parent A, Guillot A, Benjdia A, Chartier G, Leprince J & Berteau O (2016) The B_{12}-radical SAM enzyme PoyC catalyzes valine C_{β}-methylation during polytheonamide biosynthesis. *J Am Chem Soc* **138**, 15515–15518.
21 Bridwell-Rabb J, Li B & Drennan CL (2022) Cobalamin-dependent radical *S*-adenosylmethionine enzymes: capitalizing on old motifs for new functions. *ACS Bio Med Chem Au* **2**, 173–186.
22 Wang Y & Begley TP (2020) Mechanistic studies on CysS – a vitamin B_{12}-dependent radical SAM methyltransferase involved in the biosynthesis of the tert-butyl group of cystobactamid. *J Am Chem Soc* **142**, 9944–9954.
23 Banerjee R & Ragsdale SW (2003) The many faces of vitamin B12: catalysis by cobalamin-dependent enzymes. *Annu Rev Biochem* **72**, 209–247.
24 Zelder O, Beatrix B, Leutbecher U & Buckel W (1994) Characterization of the coenzyme-B12-dependent glutamate mutase from *Clostridium cochlearium* produced in *Escherichia coli*. *Eur J Biochem* **226**, 577–585.
25 Kreimeyer A, Perret A, Lechaplais C, Vallenet D, Medigue C, Salanoubat M & Weissenbach J (2007) Identification of the last unknown genes in the fermentation pathway of lysine. *J Biol Chem* **282**, 7191–7197.
26 Mancia F, Keep NH, Nakagawa A, Leadlay PF, McSweeney S, Rasmussen B, Bö Secke P, Diat O & Evans PR (1996) How coenzyme B12 radicals are generated: the crystal structure of methylmalonyl-coenzyme A mutase at 2 A resolution. *Structure* **4**, 339–350.
27 Takahashi-Iñiguez T, García-Hernandez E, Arreguín-Espinosa R & Flores ME (2012) Role of vitamin B12 on methylmalonyl-CoA mutase activity. *J Zhejiang Univ Sci B* **13**, 423–437.
28 Kaplan BH & Stadtman ER (1968) Ethanolamine deaminase, a cobamide coenzyme-dependent enzyme. I. Purification, assay, and properties of the enzyme. *J Biol Chem* **243**, 1787–1793.
29 Kofoid E, Rappleye C, Stojiljkovic I & Roth J (1999) The 17-gene ethanolamine (eut) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins. *J Bacteriol* **181**, 5317–5329.
30 Beatrix B, Zelder O, Linder D & Buckel W (1994) Cloning, sequencing and expression of the gene encoding the coenzyme B12-dependent 2-methyleneglutarate mutase from *Clostridium barkeri* in *Escherichia coli*. *Eur J Biochem* **221**, 101–109.
31 Forage RG & Foster MA (1982) Glycerol fermentation in *Klebsiella pneumoniae*: functions of the coenzyme B12-dependent glycerol and diol dehydratases. *J Bacteriol* **149**, 413–419.
32 Toraya T (2002) Enzymatic radical catalysis: coenzyme B_{12}-dependent diol dehydratase. *Chem Rec* **2**, 352–366.
33 Seyfried M, Daniel R & Gottschalk G (1996) Cloning, sequencing, and overexpression of the genes encoding coenzyme B_{12}-dependent glycerol dehydratase of *Citrobacter freundii*. *J Bacteriol* **178**, 5793–5796.
34 Sokolovskaya OM, Shelton AN & Taga ME (2020) Sharing vitamins: cobamides unveil microbial interactions. *Science* **369**, eaba0165.
35 Shelton AN, Seth EC, Mok KC, Han AW, Jackson SN, Haft DR & Taga ME (2019) Uneven distribution of cobamide biosynthesis and dependence in bacteria predicted by comparative genomics. *ISME J* **13**, 789–804.
36 Rodionov DA, Vitreshak AG, Mironov AA & Gelfand MS (2003) Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. *J Biol Chem* **278**, 41148–41159.
37 Zhang Y, Rodionov DA, Gelfand MS & Gladyshev VN (2009) Comparative genomic analyses of nickel, cobalt and vitamin B12 utilization. *BMC Genomics* **10**, 78.
38 Bryant DA, Hunter CN & Warren MJ (2020) Biosynthesis of the modified tetrapyrroles – the pigments of life. *J Biol Chem* **295**, 6888–6925.
39 Hazra AB, Han AW, Mehta AP, Mok KC, Osadchiy W, Begley TP & Taga ME (2015) Anaerobic biosynthesis of the lower ligand of vitamin B_{12}. *Proc Natl Acad Sci USA* **112**, 10792–10797.
40 Lawrence JG & Roth JR (1996) Evolution of coenzyme B_{12} synthesis among enteric bacteria: evidence for loss and reacquisition of a multigene complex. *Genetics* **142**, 11–24.

41 Gómez-Consarnau L, Sadcheva R, Gifford SM, Cutter LS, Fuhrman JA, Sanudo-Wilhelmy SA & Moran MA (2018) Mosaic patterns of B-vitamin synthesis and utilization in a natural marine microbial community. *Environ Microbiol* **20**, 2809–2823.
42 Bunbury F, Helliwell KE, Mehrshani P, Davey MP, Salmon DL, Holzer A, Smirnoff N & Smith AG (2020) Responses of a newly evolved auxotroph of *Chlamydomonas* to B_{12} deprivation. *Plant Physiol* **183**, 167–178.
43 Torrents E, Trevisiol C, Rotte C & Hellman U (2006) *Euglena gracilis* ribonucleotide reductase: the eukaryote class II enzyme and the possible antiquity of eukaryote B_{12} dependence. *J Biol Chem* **281**, 5604–5611.
44 Kolhouse JF & Allen RH (1977) Recognition of two intracellular cobalamin binding proteins and their identification as methylmalonyl-CoA mutase and methionine synthetase. *Proc Natl Acad Sci USA* **74**, 921–925.
45 Eschenmoser A (1988) Vitamin B_{12}: experiments concerning the origin of its molecular structure. *Angew Chem Int Ed* **27**, 5–39.
46 Decker K, Jungermann K & Thauer RK (1970) Energy production in anaerobic organisms. *Angew Chem Int Ed* **9**, 138–158.
47 Weiss M, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S & Martin WF (2016) The physiology and habitat of the last universal common ancestor. *Nat Microbiol* **1**, 16116.
48 Fuchs G & Stupperich E (1985) Evolution of autotrophic CO_2 fixation. In Evolution of Prokaryotes (Schleifer KH & Stackebrandt E, eds), pp. 235–251. Academic Press, London.
49 Fuchs G (2011) Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Annu Rev Microbiol* **65**, 631–658.
50 Sousa FL & Martin WF (2014) Biochemical fossils of the ancient transition from geoenergetics to bioenergetics in prokaryotic one carbon compound metabolism. *Biochim Biophys Acta* **1837**, 964–981.
51 Martin WF (2020) Older than genes: the acetyl-CoA pathway and origins. *Front Microbiol* **11**, 817.
52 Svetlitchnaia T, Svetlichnyi V, Meyer O & Dobbek H (2006) Structural insights into methyltransfer reactions of a corrinoid iron-sulfur protein involved in acetyl-CoA synthesis. *Proc Natl Acad Sci USA* **103**, 14331–14336.
53 Can M, Abernathy MJ, Wiley S, Griffith C, James CD, Xiong J, Guo Y, Hoffman BM, Ragsdale SW & Sarangi R (2023) Characterization of methyl- and acetyl-Ni intermediates in acetyl CoA synthase formed during anaerobic CO_2 and CO fixation. *J Am Chem Soc* **145**, 13696–13708.
54 Ragsdale SW (2008) Enzymology of the Wood-Ljungdahl pathway of acetogenesis. *Ann N Y Acad Sci* **1125**, 129–136.
55 Biester A, Grahame DA & Drennan CL (2024) Capturing a methanogenic carbon monoxide dehydrogenase/acetyl-CoA synthase complex via cryogenic electron microscopy. *Proc Natl Acad Sci USA* **121**, e2410995121.
56 Ruickoldt J, Basak Y, Domnik L, Jeoung JH & Dobbek H (2022) On the kinetics of CO_2 reduction by Ni, Fe-CO dehydrogenases. *ACS Catal* **12**, 13131–13142.
57 Goetzl S, Jeoung JH, Hennig SE & Dobbek H (2011) Structural basis for electron and methyl-group transfer in a methyltransferase system operating in the reductive acetyl-CoA pathway. *J Mol Biol* **411**, 96–109.
58 Maden BE (2000) Tetrahydrofolate and tetrahydromethanopterin compared: functionally distinct carriers in C1 metabolism. *Biochem J* **350**, 609–629.
59 Ragsdale SW, Lindahl PA & Münck E (1987) Mössbauer, EPR, and optical studies of the corrinoid/iron-sulfur protein involved in the synthesis of acetyl coenzyme A by *Clostridium thermoaceticum*. *J Biol Chem* **262**, 14289–14297.
60 Preiner M, Igarashi K, Muchowska KB, Yu M, Varma SJ, Kleinermanns K, Nobu MK, Kamagata Y, Tüysüz H, Moran J *et al.* (2020) A hydrogen-dependent geochemical analogue of primordial carbon and energy metabolism. *Nat Ecol Evol* **4**, 534–542.
61 Varma SJ, Muchowska KB, Chatelain P & Moran J (2018) Native iron reduces CO_2 to intermediates and end-products of the acetyl-CoA pathway. *Nat Ecol Evol* **2**, 1019–1024.
62 Beyazay T, Ochoa-Hernandez C, Song Y, Belthle KS, Martin WF & Tüysüz H (2023) Influence of composition of nickel-iron nanoparticles for abiotic CO_2 conversion to early prebiotic organics. *Angew Chem Int Ed Engl* **62**, e202218189.
63 Song Y, Beyazay T & Tüysüz H (2023) Effect of alkali- and alkaline-earth-metal promoters on silica-supported Co-Fe alloy for autocatalytic CO_2 fixation. *Angew Chem Int Ed Engl* **63**, e202316110.
64 Belthle KS, Martin WF & Tüysüz H (2024) Synergistic effects of silica-supported iron-cobalt catalysts for CO_2 reduction to prebiotic organics. *ChemCatChem* **16**, e202301218.
65 Ueno Y, Yamada K, Yoshida N, Maruyama S & Isozaki Y (2006) Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* **440**, 516–519.
66 Tashiro T, Ishida A, Hori M, Igisu M, Koike M, Mejean P, Takahata N, Sano Y & Komiya T (2017) Early trace of life from 3.95 Ga sedimentary rocks in Labrador, Canada. *Nature* **549**, 516–518.
67 Mojzsis SJ, Arrhenius G, McKeegan KD, Harrison TM, Nutman AP & Friend CRL (1996) Evidence for life on Earth before 3,800 million years ago. *Nature* **384**, 55–59.

68 Blaser MB, Dreisbach LK & Conrad R (2013) Carbon isotope fractionation of 11 acetogenic strains grown on H_2 and CO_2. *Appl Environ Microbiol* **79**, 1787–1794.

69 Mei R, Kaneko M, Imachi H & Nobu MK (2023) The origin and evolution of methanogenesis and Archaea are intertwined. *PNAS Nexus* **2**, pgad023.

70 Colman DR, Kraus EA, Thieringer PH, Rempfert K, Templeton AS, Spear JR & Boyd ES (2022) Deep-branching acetogens in serpentinized subsurface fluids of Oman. *Proc Natl Acad Sci USA* **119**, e2206845119.

71 Martin WF & Russell MJ (2007) On the origin of biochemistry at an alkaline hydrothermal vent. *Philos Trans R Soc Lond B Biol Sci* **362**, 1887–1925.

72 Schöne C, Poehlein A, Jehmlich N, Adlung N, Daniel R, von Bergen M, Scheller S & Rother M (2022) Deconstructing *Methanosarcina acetivorans* into an acetogenic archaeon. *Proc Natl Acad Sci USA* **119**, e2113853119.

73 Lang SQ, Butterfield DA, Schulte M, Kelley DS & Lilley MD (2010) Elevated concentrations of formate, acetate and dissolved organic carbon found at the Lost City hydrothermal field. *Geochim Cosmochim Acta* **74**, 941–952.

74 Etiope G (2017) Abiotic methane in continental serpentinization sites: an overview. *Procedia Earth Planet Sci* **17**, 9–12.

75 Schwander L, Brabender M, Mrnjavac N, Wimmer JLE, Preiner M & Martin WF (2023) Serpentinization as the source of energy, electrons, organics, catalysts, nutrients and pH gradients for the origin of LUCA and life. *Front Microbiol* **14**, 1257597.

76 Mrnjavac N, Wimmer JLE, Brabender M, Schwander L & Martin WF (2023) The moon-forming impact and the autotrophic origin of life. *ChemPlusChem* **88**, e202300270.

77 Kaur H, Rauscher SA, Werner E, Song Y, Yi J, Kazöne W, Martin WF, Tüysüz H & Moran J (2024) A prebiotic Krebs cycle analog generates amino acids with H_2 and NH_3 over nickel. *Chem* **10**, 1528–1540.

78 Buckel W & Thauer RK (2013) Energy conservation via electron bifurcating ferredoxin reduction and proton/Na^+ translocating ferredoxin oxidation. *Biochim Biophys Acta* **1827**, 94–113.

79 Brabender M, Henriques-Pereira DP, Mrnjavac N, Schlikker ML, Kimura ZI, Sucharitakul J, Kleinermanns K, Tüysüz H, Buckel W, Preiner M *et al.* (2024) Ferredoxin reduction by hydrogen with iron functions as an evolutionary precursor of flavin-based electron bifurcation. *Proc Natl Acad Sci USA* **121**, e2318969121.

80 Holliday GL, Thornton JM, Marquet A, Smith AG, Rebeille F, Mendel R, Schubert HL, Lawrence AD & Warren MJ (2007) Evolution of enzymes and pathways for the biosynthesis of cofactors. *Nat Prod Rep* **24**, 972–987.

81 Warren MJ, Raux E, Schubert HL & Escalante-Semerena JC (2002) The biosynthesis of adenosylcobalamin (vitamin B_{12}). *Nat Prod Rep* **19**, 390–412.

82 Martin WF & Sousa FL (2015) Early microbial evolution: the age of anaerobes. *Cold Spring Harb Perspect Biol* **8**, a018127.

83 Wada K, Harada J, Yaeda Y, Tamiaki H, Oh-Oka H & Fukuyama K (2007) Crystal structures of CbiL, a methyltransferase involved in anaerobic vitamin B biosynthesis, and CbiL in complex with S-adenosylhomocysteine – implications for the reaction mechanism. *FEBS J* **274**, 563–573.

84 Frank S, Brindley AA, Deery E, Heathcote P, Lawrence AD, Leech HK, Pickersgill RW & Warren MJ (2005) Anaerobic synthesis of vitamin B_{12}: characterization of the early steps in the pathway. *Biochem Soc Trans* **33**, 811–814.

85 Broderick JB, Duffus DR, Duschene KS & Shepard EM (2014) Radical S-adenosylmethionine enzymes. *Chem Rev* **114**, 4229–4317.

86 Aziz I, Kayastha K, Kaltwasser S, Vonck J, Welsch S, Murphy BJ, Kahnt J, Wu D, Wagner T, Shima S *et al.* (2024) Structural and mechanistic basis of the central energy-converting methyltransferase complex of methanogenesis. *Proc Natl Acad Sci USA* **121**, e2315568121.

87 Lu WP, Schiau I, Cunningham JR & Ragsdale SW (1993) Sequence and expression of the gene encoding the corrinoid/iron-sulfur protein from *Clostridium thermoaceticum* and reconstitution of the recombinant protein to full activity. *J Biol Chem* **268**, 5605–5614.

88 Maupin-Furlow J & Ferry JG (1996) Characterization of the cdhD and cdhE genes encoding subunits of the corrinoid/iron-sulfur enzyme of the CO dehydrogenase complex from *Methanosarcina thermophila*. *J Bacteriol* **178**, 340–346.

89 Slotznick SP, Johnson JE, Rasmussen B, Raub TD, Webb SM, Zi JW, Kirschvink JL & Fischer WW (2022) Reexamination of 2.5-Ga "whiff" of oxygen interval points to anoxic ocean before GOE. *Sci Adv* **8**, eabj7190.

90 Mrnjavac N, Schwander L, Brabender M & Martin WF (2024) Chemical antiquity in metabolism. *Acc Chem Res* **57**, 2267–2278.

91 Ragsdale SW (2003) Pyruvate ferredoxin oxidoreductase and its radical intermediate. *Chem Rev* **103**, 2333–2346.

92 Lu Z & Imlay JA (2021) When anaerobes encounter oxygen: mechanisms of oxygen toxicity, tolerance and defense. *Nat Rev Microbiol* **19**, 774–785.

93 Mrnjavac N, Schwander L, Brabender M & Martin WF (2024) The radical impact of oxygen on prokaryotic evolution—enzyme inhibition first, uninhibited essential biosyntheses second, aerobic respiration third. *FEBS Lett* **598**, 1692–1714.

94 Pierce E, Xie G, Barabote RD, Saunders E, Han CS, Detter JC, Richardson P, Brettin TS, Das A, Ljungdahl LG *et al*. (2008) The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*). *Environ Microbiol* **10**, 2550–2573.
95 Hendrickson EL, Kaul R, Zhou Y, Bovee D, Chapman P, Chung J, Conway de Macario E, Dodsworth JA, Gillett W, Graham DE *et al*. (2004) Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis*. *J Bacteriol* **186**, 6956–6969.
96 Dailey HA, Dailey TA, Gerdes S, Jahn D, Jahn M, O'Brian MR & Warren MJ (2017) Prokaryotic heme biosynthesis: multiple pathways to a common essential product. *Microbiol Mol Biol Rev* **81**, e00048-16.
97 Videira MAM, Lobo SAL, Sousa FL & Saraiva LM (2020) Identification of the sirohaem biosynthesis pathway in *Staphylococcus aureus*. *FEBS J* **287**, 1537–1553.
98 Dailey HA & Medlock AE (2022) A primer on heme biosynthesis. *Biol Chem* **403**, 985–1003.
99 Stroupe ME, Leech HK, Daniels DS, Warren MJ & Getzoff ED (2003) CysG structure reveals tetrapyrrole-binding features and novel regulation of siroheme biosynthesis. *Nat Struct Biol* **10**, 1064–1073.
100 Zheng K, Ngo PD, Owens VL, Yang XP & Mansoorabadi SO (2016) The biosynthetic pathway of coenzyme F430 in methanogenic and methanotrophic archaea. *Science* **354**, 339–342 (published correction appears in *Nature* 2017, 545, 40).
101 Moore SJ, Sowa ST, Schuchardt C, Deery E, Lawrence AD, Ramos JV, Billig S, Birkenmeyer C, Chivers PT, Howard MJ *et al*. (2017) Corrigendum: elucidation of the biosynthesis of the methane catalyst coenzyme F_{430}. *Nature* **545**, 116.
102 Buan NR, Rehfeld K & Escalante-Semerena JC (2006) Studies of the CobA-type ATP:Co(I)rrinoid adenosyltransferase enzyme of *Methanosarcina mazei* strain Gö1. *J Bacteriol* **188**, 3543–3550.
103 Zhang C, Shine M, Pyle AM & Zhang Y (2022) US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat Methods* **19**, 1109–1115.
104 Shima S, Huang G, Wagner T & Ermler U (2020) Structural basis of hydrogenotrophic methanogenesis. *Annu Rev Microbiol* **74**, 713–733.
105 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A *et al*. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589.
106 Wagner T, Ermler U & Shima S (2016) The methanogenic CO_2 reducing-and-fixing enzyme is bifunctional and contains 46 [4Fe-4S] clusters. *Science* **354**, 114–117.
107 Gong W, Hao B, Wei Z, Ferguson DJ, Tallant T, Krzycki JA & Chan MK (2008) Structure of the $\alpha_2\varepsilon_2$ Ni-dependent CO dehydrogenase component of the *Methanosarcina barkeri* acetyl-CoA decarbonylase/synthase complex. *Proc Natl Acad Sci USA* **105**, 9558–9563.
108 Doukov TI, Iverson TM, Seravalli J, Ragsdale SW & Drennan CL (2002) A Ni-Fe-Cu center in a bifunctional carbon monoxide dehydrogenase/acetyl-CoA synthase. *Science* **298**, 567–572.
109 Gagnon DM, Stich TA, Mehta AP, Abdelwahed SH, Begley TP & Britt RD (2018) An aminoimidazole radical intermediate in the anaerobic biosynthesis of the 5,6-dimethylbenzimidazole ligand to vitamin B_{12}. *J Am Chem Soc* **140**, 12798–12807.
110 Wienhausen G, Dlugosch L, Jarling R, Wilkes H, Giebel HA & Simon M (2022) Availability of vitamin B_{12} and its lower ligand intermediate α-ribazole impact prokaryotic and protist communities in oceanic systems. *ISME J* **16**, 2002–2014.
111 Anderson PJ, Lango J, Carkeet C, Britten A, Kräutler B, Hammock BD & Roth JR (2008) One pathway can incorporate either adenine or dimethylbenzimidazole as an alpha-axial ligand of B_{12} cofactors in *Salmonella enterica*. *J Bacteriol* **190**, 1160–1171.
112 Doukov TI, Seravalli J, Stezowski JJ & Ragsdale SW (2000) Crystal structure of a methyltetrahydrofolate- and corrinoid-dependent methyltransferase. *Structure* **8**, 817–830.
113 Hu SI, Pezacka E & Wood HG (1984) Acetate synthesis from carbon monoxide by *Clostridium thermoaceticum*. Purification of the corrinoid protein. *J Biol Chem* **259**, 8892–8897.
114 Adam PS, Borrel G & Gribaldo S (2018) Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc Natl Acad Sci USA* **115**, E1166–E1173.
115 Xu J & Zhang Y (2010) How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* **26**, 889–895.
116 Vitols E, Walker GA & Huennekens FM (1966) Enzymatic conversion of vitamin B_{12s} to a cobamide coenzyme, α-(5,6-dimethylbenzimidazolyl) deoxyadenosylcobamide (adenosyl-B_{12}). *J Biol Chem* **241**, 1455–1461.
117 Johnson CLV, Buszko ML & Bobik TA (2004) Purification and initial characterization of the *Salmonella enterica* PduO ATP:Cob(I)alamin adenosyltransferase. *J Bacteriol* **186**, 7881–7887.
118 Buan NR, Suh S & Escalante-Semerena JC (2004) The eutT gene of *Salmonella enterica* encodes an oxygen-labile, metal-containing ATP:corrinoid adenosyltransferase enzyme. *J Bacteriol* **186**, 5708–5714.
119 Jeter VL, Mattes TA, Beattie NR & Escalante-Semerena JC (2019) A new class of

phosphoribosyltransferases involved in cobamide biosynthesis is found in methanogenic archaea and cyanobacteria. *Biochemistry* **58**, 951–964.
120 Öppinger C, Kremp F & Müller V (2022) Is reduced ferredoxin the physiological electron donor for MetVF-type methylenetetrahydrofolate reductases in acetogenesis? A hypothesis. *Int Microbiol* **25**, 75–88.
121 Reeves EP, McDermott JM & Seewald JS (2014) The origin of methanethiol in midocean ridge hydrothermal fluids. *Proc Natl Acad Sci USA* **111**, 5474–5479.
122 Kitadai N, Nakamura R, Yamamoto M, Okada S, Takahagi W, Nakano Y, Takahashi Y, Takai K & Oono Y (2021) Thioester synthesis through geoelectrochemical CO_2 fixation on Ni sulfides. *Commun Chem* **4**, 37.
123 Sánchez-Andrea I, Guedes IA, Hornung B, Boeren S, Lawson CE, Sousa DZ, Arren BE, Claassens NJ & Stams AJM (2020) The reductive glycine pathway allows autotrophic growth of *Desulfovibrio desulfuricans*. *Nat Commun* **11**, 5090.
124 Berg IA (2011) Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. *Appl Environ Microbiol* **77**, 1925–1936.
125 Beyazay T, Belthle KS, Farès C, Preiner M, Moran J, Martin WF & Tüysüz H (2023) Ambient temperature CO_2 fixation to pyruvate and subsequently to citramalate over iron and nickel nanoparticles. *Nat Commun* **14**, 570.
126 Furdui C & Ragsdale SW (2000) The role of pyruvate ferredoxin oxidoreductase in pyruvate synthesis during autotrophic growth by the Wood-Ljungdahl pathway. *J Biol Chem* **275**, 28494–28499.
127 Henriques Pereira DP, Leethaus J, Beyazay T, do Nascimento Vieira A, Kleinermanns K, Tüysüz H, Martin WF & Preiner M (2022) Role of geochemical protoenzymes (geozymes) in primordial metabolism: specific abiotic hydride transfer by metals to the biological redox cofactor NAD^+. *FEBS J* **289**, 3148–3162.
128 Müller V, Chowdhury NP & Basen M (2018) Electron bifurcation: a long-hidden energy-coupling mechanism. *Annu Rev Microbiol* **72**, 331–353.
129 Neukirchen S, Pereira IAC & Sousa FL (2023) Stepwise pathway for early evolutionary assembly of dissimilatory sulfite and sulfate reduction. *ISME J* **17**, 1680–1692.
130 Thauer RK (2019) Methyl (alkyl)-coenzyme M reductases: nickel F-430-containing enzymes involved in anaerobic methane formation and in anaerobic oxidation of methane or of short chain alkanes. *Biochemistry* **58**, 5198–5220.
131 Debussche L, Thibaut D, Cameron B, Crouzet J & Blanche F (1990) Purification and characterization of cobyrinic acid a,c-diamide synthase from *Pseudomonas denitrificans*. *J Bacteriol* **172**, 6239–6244.
132 Lübbe YJ, Youn HS, Timkovich R & Dahl C (2006) Siro(haem)amide in *Allochromatium vinosum* and relevance of DsrL and DsrN, a homolog of cobyrinic acid a,c-diamide synthase, for sulphur oxidation. *FEMS Microbiol Lett* **261**, 194–202.
133 Loy A, Duller S, Baranyi C, Mußmann M, Ott J, Sharon I, Béjà O, Le Paslier D, Dahl C & Wagner M (2009) Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes. *Environ Microbiol* **11**, 289–299.
134 Koonin EV, Makarova KS & Aravind L (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55**, 709–742.
135 Dmitrijeva M, Tackmann J, Matias Rodrigues JF, Huerta-Cepas J, Coelho LP & von Mering C (2024) A global survey of prokaryotic genomes reveals the eco-evolutionary pressures driving horizontal gene transfer. *Nat Ecol Evol* **8**, 986–998.
136 Dagan T, Artzy-Randrup Y & Martin W (2008) Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc Natl Acad Sci USA* **105**, 10039–10044.
137 Wimmer JLE, Vieira AN, Xavier JC, Kleinermanns K, Martin WF & Preiner M (2021) The autotrophic core: an ancient network of 404 reactions converts H_2, CO_2, and NH_3 into amino acids, bases, and cofactors. *Microorganisms* **9**, 458.
138 Schubert HL, Blumenthal RM & Cheng X (2003) Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Sci* **28**, 329–335.
139 Sirithanakorn C & Cronan JE (2021) Biotin, a universal and essential cofactor: synthesis, ligation and regulation. *FEMS Microbiol Rev* **45**, fuab003.
140 Huang L, Qin Y, Yan Q, Lin G, Huang L, Huang B & Huang W (2015) MinD plays an important role in *Aeromonas hydrophila* adherence to *Anguilla japonica* mucus. *Gene* **565**, 275–281.
141 Settembre E, Begley TP & Ealick SE (2003) Structural biology of enzymes of the thiamin biosynthesis pathway. *Curr Opin Struct Biol* **13**, 739–747.
142 Wongnate T, Sliwa D, Ginovska B, Smith D, Wolf MW, Lehnert N, Raugei S & Ragsdale SW (2016) The radical mechanism of biological methane synthesis by methyl-coenzyme M reductase. *Science* **352**, 953–958.
143 Romão CV, Ladakis D, Lobo SAL, Carrondo MA, Brindley AA, Deery E, Matias PM, Pickersgill RW, Saraiva LM & Warren MJ (2011) Evolution in a family of chelatases facilitated by the introduction of active site asymmetry and protein oligomerization. *Proc Natl Acad Sci USA* **108**, 97–102.
144 Brindley AA, Raux E, Leech H, Schubert HL & Warren MJ (2003) Identification and characterization

of a small 13-15 kDa "ancestral" cobaltochelatase ($CbiX^S$) in the archaea. *J Biol Chem* **278**, 22388–22395.
145 Forterre P (2001) Genomics and early cellular evolution. The origin of the DNA world. *C R Acad Sci III* **324**, 1067–1076.
146 Di Giulio M (2011) The last universal common ancestor (LUCA) and the ancestors of archaea and bacteria were progenotes. *J Mol Evol* **72**, 119–126.
147 Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K *et al.* (2016) A new view of the tree of life. *Nat Microbiol* **1**, 16048.
148 Mahendrarajah TA, Moody ERR, Schrempf D, Szantho LL, Dombrowski N, Davin AA, Pisani D, Donoghue PCJ, Szöllősi GJ, Williams TA *et al.* (2023) ATP synthase evolution on a cross-braced dated tree of life. *Nat Commun* **14**, 7456.
149 Moody ERR, Álvarez-Carretero S, Mahendrarajah TA, Clark JW, Betts HC, Dombrowski N, Szánthó LL, Boyle RA, Daines S, Chen X *et al.* (2024) The nature of the last universal common ancestor and its impact on the early Earth system. *Nat Ecol Evol* **8**, 1654–1666.
150 Chistoserdova L, Vorholt JA, Thauer RK & Lidstrom ME (1998) C1 transfer enzymes and coenzymes linking methylotrophic bacteria and methanogenic Archaea. *Science* **281**, 99–102.
151 Chistoserdova L, Jenkins C, Kalyuzhnaya MG, Marx CJ, Lapidus A, Vorholt JA, Staley JT & Lidstrom ME (2004) The enigmatic planctomycetes may hold a key to the origins of methanogenesis and methylotrophy. *Mol Biol Evol* **21**, 1234–1241.
152 Kalyuzhnaya MG, Korotkova N, Crowther G, Marx CJ, Lidstrom ME & Chistoserdova L (2005) Analysis of gene islands involved in methanopterin-linked C1 transfer reactions reveals new functions and provides evolutionary insights. *J Bacteriol* **187**, 4607–4614.
153 Chistoserdova L & Kalyuzhnaya MG (2018) Current trends in methylotrophy. *Trends Microbiol* **26**, 703–714.
154 Adam PS, Borrel G & Gribaldo S (2019) An archaeal origin of the Wood-Ljungdahl H4MPT branch and the emergence of bacterial methylotrophy. *Nat Microbiol* **4**, 2155–2163.
155 Jin X, Yang Y, Cao H, Gao B & Zhao Z (2022) Eco-phylogenetic analyses reveal divergent evolution of vitamin B_{12} metabolism in the marine bacterial family '*Psychromonadaceae*'. *Environ Microbiol Rep* **14**, 147–163.
156 Kirschning A (2022) On the evolution of coenzyme biosynthesis. *Nat Prod Rep* **39**, 2175–2199.
157 Kuznetsov S, Milenkin A & Antonov I (2022) Translational frameshifting in the chlD gene gives a clue to the coevolution of the chlorophyll and cobalamin biosyntheses. *Microorganisms* **10**, 1200.
158 Belthle KS, Beyazay T, Ochoa-Hernandez C, Miyazaki R, Foppa L, Martin WF & Tüysüz H (2022) Effects of silica modification (Mg, Al, Ca, Ti, and Zr) on supported cobalt catalysts for H_2-dependent CO_2 reduction to metabolic intermediates. *J Am Chem Soc* **144**, 21232–21243.
159 Sousa FL, Thiergart T, Landan G, Nelson-Sathi S, Pereira IAC, Allen JF, Lane N & Martin WF (2013) Early bioenergetic evolution. *Philos Trans R Soc Lond B Biol Sci* **368**, 20130088.
160 Bowman JC, Petrov AS, Frenkel-Pinter M, Penev PI & Williams MD (2020) Root of the tree: the significance, evolution, and origins of the ribosome. *Chem Rev* **120**, 4848–4878.
161 Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S *et al.* (2022) Database resources of the national center for biotechnology information. *Nucleic Acids Res* **50**, D20–D26.
162 Kanehisa M & Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30.
163 Chen IMA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, Hajek P, Ritter SJ, Webb C, Wu D *et al.* (2023) The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Res* **51**, D723–D732.
164 The UniProt Consortium (2023) UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res* **51**, D523–D531.
165 Buchfink B, Xie C & Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59–60.
166 Buchfink B, Reuter K & Drost HG (2021) Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* **18**, 366–368.
167 Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240.
168 Needleman SB & Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**, 443–453.
169 Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780.
170 Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.
171 Nagies FSP, Brueckner J, Tria FDK & Martin WF (2020) A spectrum of verticality across genes. *PLoS Genet* **16**, e1009200.

172 Letunic I & Bork P (2021) Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, W293–W296.
173 Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN & Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. The anaerobic cobalamin biosynthesis pathway.

Fig. S2. Structural alignments (25) of the two CoFeS subunits of *M. maripaludis* (mmp0980 and mmp0981, red) and *M. thermoacetica* (moth_1198 and moth_1201, blue).

Fig. S3. Species tree of 80 analyzed methanogens.

Fig. S4. Concatenated tree of the anaerobic route to cobalamin in methanogens and acetogens.

Table S1. TM-scores for structural alignments of archaeal and bacterial enzymes of the methyl branch of the acetyl-CoA pathway and the anaerobic cobalamin synthesis pathway.

Table S2. Prokaryotic informational core genes (ICG) and their functions. The genes were used to calculate a species tree comprising all analysed methanogens (Fig. S3).

Table S3. Diamond blast hits of cobalamin biosynthesis proteins vs the dataset of 404 autocatalytic core reactions.

Table S4. Homologues of cobalamin biosynthesis proteins found in 355 protein families traced to LUCA (Weiss *et al.* 2016). Sorted by functional category.

Table S5. Homologues of cobalamin synthesis proteins that share SUPERFAMILY annotation found in the dataset of 3939 genomes.

Table S6. Accession numbers and taxonomy of acetogen and methanogen genomes analysed in this work. The genomes are a subset of a Reference Sequence (RefSeq) data set of 2016.

Table S7. Query sequences used for homology searches and sources for structural comparisons.