Biochemical fossils of the ancient transition from geoenergetics to bioenergetics in prokaryotic one carbon compound metabolism

Filipa L. Sousa, William F. Martin *
Institute for Molecular Evolution, University of Düsseldorf, 40225 Düsseldorf, Germany

1. Introduction

When it comes to the evolution of bioenergetic systems, the topic of this special issue, of interest is the question of how bioenergetic systems got started in the first place. Clearly, in order to evolve a bioenergetic system consisting of genes, proteins, cofactors and — in the case of chemiosmotic coupling — membranes, there has to be some preexisting, exergonic geological precursor reaction that underpinned the chemical origin of those genes, proteins and cofactors. At some point there was a transition from ‘geoenergetics’ to bioenergetics, and there hence existed in the very first life forms some core energy releasing reaction that was harnessed so as to allow energy to be conserved in a chemical currency that could be used to promote metabolic reactions that otherwise were sluggish. It would improve our understanding of early evolution immensely to have a better understanding of what that spontaneous geoenergetic reaction was, what the nature of the first bioenergetic reactions was, and the relationship between those two kinds of reactions. Thanks to advances in understanding subsurface energy-releasing chemical reactions that occur in the Earth’s hydrothermal systems [4,121], paired with advances in understanding the energetics of anaerobic microbes [26], geochemists and biologists are now finding more common ground for discussion on such questions than ever before. Both sides are talking about redox chemistry, metals, and the exergonic reduction of CO2 with electrons stemming from hydrogen and iron.

In an early, and insightful, survey of bioenergetics in anaerobes, Decker et al. [37] suggested, based on comparative biochemistry, that methanogens and acetogens are the most ancient forms of energy metabolism among extant microbes: they are strict anaerobes, they tend to lack cytochromes, and they satisfy their carbon and energy needs from the reduction of CO2 with H2 substrates that would have been abundant on the early Earth. Forty years later, the basic reasoning behind the idea that anaerobic autotrophs are ancient is still modern [105], it still has many virtues, and the underlying reasons have become much more detailed [56,61,113]. In addition, geological findings independently came to support the antiquity of methanogens because biological methane production was found to go back at least 3.4 billion years [193] and geochemical reactions similar to the core bioenergetic reactions of acetogens and methanogens have been found to occur spontaneously at hydrothermal vents [100,169].

As an alternative to acetate or methane formation, Wächtershäuser [196] suggested that pyrite formation from Fe2+ and H2S was the first source of biological energy. But the pyrite theory did not forge a clear...
link to modern microbial physiology, nor did it take into account the vexing ubiquity of chemiosmotic coupling among modern cells [114]. From our standpoint, having a link to modern microbes is important, because very many different possible sources of energy for early biochemical systems can be envisaged, including polyphosphates [12], photochemical ZnS oxidation [125,126], ultraviolet light, and other possibilities [36,91]. But one cannot meaningfully address the ancestral possibilities [36,91]. But one cannot meaningfully address the ancestral possibilities [36,91].

In contrast to the issue of core bioenergetic reactions, a great deal of attention has been given to the issues of (i) whether the first organisms were thermophiles [137,181,182] or not [20,58,62], and ii) and whether they were autotrophs [37,124,197] or not [112,134]. The issue of which pathways they actually used to make a living in the sense of carbon and energy metabolism [97,115] has somehow been of secondary importance. Now is a good time to invigorate the question of the earliest bioenergetic systems, for two reasons. First, newer findings document eyebrow-raising similarities between the bioenergetic reactions of anaerobic autotrophs and geochemical reactions that occur spontaneously at some types of hydrothermal vents [121], an exciting development. Second, electron bifurcation has recently been discovered [104], a mechanism of energy conservation that explains how it is possible for acetogens and methanogens to reduce CO₂ with electrons from H₂, even though the first segment of the reaction sequence is energetically uphill [26]. Electron bifurcation is a major advance in understanding the bioenergetics of anaerobes in general, and of anaerobic autotrophs in particular.

Methanogens and acetogens replenish their ATP pool with a rotor–stator type ATPase that harnesses ion gradients generated during the reduction of CO₂ with H₂ with the involvement of the acetyl-CoA pathway [26]. Among the six CO₂ fixation pathways known, the acetyl-CoA pathway, or Wood–Ljungdahl pathway [107,219], is the only one known that occurs in both archaea and bacteria [16,61]. This and other lines of evidence suggest that it is the most ancient of the six [60,61,114]. In hydrogenotrophic methanogens and acetogens, the acetyl-CoA pathway is simultaneously linked to a pathway of energy metabolism, because these organisms obtain their energy from the reduction of CO₂ to methane and acetate respectively, using H₂ as the electron donor. This is clearly an ancient redox couple for energy metabolism [56,99,105]. In comparisons of the acetyl-CoA pathway in acetogens and methanogens, the use of different cofactors for methyl synthesis from CO₂ stands out: tetrahydrofolate (H₄F) in acetogens versus tetrahydrodihydroxyacetonephosphatase (H₄MPT) in methanogens [49,88,93,109]. The differences in the cofactors are of particular interest because folate is mostly been investigated in the laboratory of Robert H. White [39,118,224] in addition to providing the methyl groups for C₁ units for amino acid, cofactor and nucleotide biosynthesis [109,165,224] in addition to providing the methyl groups for metabolites and ribosome methylation so that the genetic code will work [39,118,180].

In archaea, the situation concerning C₁ carriers—a topic that has mostly been investigated in the laboratory of Robert H. White [35,67,201–210,212]—is more diverse, as recently summarized by de

<table>
<thead>
<tr>
<th>Genes/ accession</th>
<th>Seq. length</th>
<th>No.</th>
<th>Enzyme name</th>
<th>% I*</th>
<th>% S*</th>
<th>Enzyme name</th>
<th>No.</th>
<th>Genes/ accession</th>
<th>Seq. length</th>
</tr>
</thead>
<tbody>
<tr>
<td>fdhA Moth_2312</td>
<td>899</td>
<td>1</td>
<td>Formate dehydrogenase</td>
<td>11*</td>
<td>19*</td>
<td>Formyl-methanofuran dehydrogenase</td>
<td>7</td>
<td>fmdC</td>
<td></td>
</tr>
<tr>
<td>fdhB Moth_2314</td>
<td>707</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fhs Moth_0109</td>
<td>559</td>
<td>2</td>
<td>10-Formyl-H₄F synthetase</td>
<td>10</td>
<td>16</td>
<td>Formyl transferase</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>folD Moth_1516</td>
<td>280</td>
<td>3</td>
<td>5,10-Methylene-H₄F cyclohydrolase/dehydrogenase</td>
<td>11</td>
<td>17</td>
<td>5,10-Methylene-H₄F-MPT (H₄MPT)</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>metE Moth_1191</td>
<td>306</td>
<td>5</td>
<td>5,10-Methylene-H₄F reductase</td>
<td>13</td>
<td>22</td>
<td>5,10-Methylene-H₄F-MPT reductase</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cdh α/acsD Moth_2960</td>
<td>323</td>
<td></td>
<td>Corrinoid iron–sulfur protein (CFeSP)</td>
<td>25*</td>
<td>40*</td>
<td>Corrinoid iron–sulfur protein</td>
<td>384</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cdh γ/acsC Moth_1201</td>
<td>446</td>
<td>13</td>
<td>CO dehydrogenase/acetyl-CoA synthase</td>
<td>25*</td>
<td>37*</td>
<td>CO dehydrogenase/acetyl-CoA synthase</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cdh (α/β)/acsB Moth_1202</td>
<td>729</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cdh δ Moth_1198</td>
<td>323</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cdh γ Moth_1201</td>
<td>446</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The comparison of the enzymes that catalyze the different steps of the Wood–Ljungdahl in acetogens and methanogens. Numbers (No.) refer to the steps presented in Fig. 3. Genes, accessions numbers and sequence length are given using the example of Moebius thermoacetica and Methanothermobacter marburgensis as summarized by Fuchs [61]. The % Identity (I) and % Similarity (S) values refer to the average between all blast hits (see Materials and methods). *— average values; †The values correspond to the most similar subunits of the complexes. The line that separates reaction 6 and 12 from reaction 13 was moved relative to Fuchs’s [61] notation because the CoFE small subunit (Moth_1198) also participates in reaction 13.
 Crécy-Lagard et al. [35], who point out that the archaea, including non-methanotrophic forms, generally tend to possess methanopterin or methanopterin-related C1 carriers. Exceptions to this rule are the halophiles, which possess H₄F instead of H₄MPT [19,135], and Methanosarcina barkeri strain fusaro, which possesses both H₄MPT and H₄F [25]. Here we examine the phylogenetic distribution of genes involved in H₄B biosynthesis and those known so far in H₄MPT biosynthesis among prokaryotic genomes with the aim of exploring the ancestral state of C1 metabolism in the prokaryote common ancestor.

2. Materials and methods

2.1. Data

Genomes of 1606 prokaryotes (117 archaea and 1489 bacteria) were downloaded from RefSeq database (v03.2012) [152]. Literature searches on the biosynthesis of the different pterins were performed. Homologous proteins involved in the different folate and pterin biosynthesis were identified by BLAST [3] within the data set of downloaded genomes using the proteins from [24,35,46,63,69,70,78,90,103,116,117,144,160,162,172,183,186,187,200,220]. The BLAST lists were filtered for E values better than 10⁻³⁰ and amino acid identities ≥ 30%. To account for fused genes, the BLAST results were parsed and a gene classified according to its highest similarity hit. If a gene presented the highest homology with a fused one (e.g., folBK), the presence of both genes (in this case, folB and folK) was considered.

Homologous proteins involved in the acetyl-CoA pathway were identified by BLAST [3] within the data set of downloaded genomes using the proteins from [61] and filtered for E values better than 10⁻¹⁰ and amino acid identities ≥ 20% (Table 1). Protein pairs from organisms where the Wood-Ljungdahl pathway is present were globally aligned using the Needleman–Wunsch algorithm with needle program (EMBOSS package) [159].

2.2. Sequence alignments and phylogenetic analysis

Proteins were aligned using MUSCLE [45] using its default parameters. Statistical testing was done using the program SEQBOOT (PHYLP 3.695 package) [54] by resampling the data sets 100 times. For construction of phylogeny using maximum-likelihood, FastTree 2.1.7 [150] was used with the WAG + G model and four rate categories. A majority extended consensus tree was created with consense (PHYLIP 3.695 package) [54]. Alignments and trees are available upon request.

2.3. Structural analysis

SCOP domain annotations [129] were retrieved by scanning each sequence from Table 1 against the Hidden–Markov Models Library available at the SUPERFAMILY resource [65]. To retrieve the closest available tertiary structure for each family, a BLAST search using the genes from Table 1 as queries was performed against amino acid sequences of protein structures deposited at the Protein Data Bank [17]. The best hits were downloaded and screened for membership in the corresponding protein family according to functional annotation and sequence similarity (no E-value cut-off was employed in order to accommodate differences in substitution rates across sequences). This provided related structures for nearly all of the enzymes numbered 1–13 shown in Fig. 3. That is, there was a structure available in PDB for a protein carrying the same functional annotation as the query, except for i) the B subunit of Methanobrevibacter formylmethyleneum dehydrogenase, and ii) the B subunit of Methanobrevibacter formylmethyleneum dehydrogenase, and iii) the β, α, and γ subunits of cadalalcohol dehydrogenase.

We then compared structures in two steps. In the first step, we asked whether the acetogen and the methanogen enzymes are related by comparing the structures of the proteins corresponding to the acetogen and methanogen enzymes pairwise using DaliLite version 3.1 with the pairwise option [84]. The DALI algorithm uses a weighted sum of similarities of intra-molecular distances to infer structural similarities. DALI-Z scores above 2 are taken as evidence for significant structural similarity in pairwise structural comparisons [84]. Structural alignments were manually checked with Pymol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

In the second step, we asked: to which protein families in PDB do the acetogen and methanogen structures are most closely related. This was done by submitting the coordinates of the PDB entries corresponding to reactions 1–13 in Fig. 4 to a structural similarity search at the DALI server to find its structural homologues. This revealed whether the acetogen and methanogen structures are more similar to each other than they are to other structures represented in PDB or whether they just share common folds. Results are summarized in the Supplemental Material Tables 1 to 15.

3. Results and discussion

3.1. The cofactor pathways

Cofactors play an important role in metabolism, doing most of the necessary catalysis. For example, in some pyridoxal phosphate (PLP) dependent decarboxylases, the PLP cofactor can be responsible for enhancing the catalytic efficiency by 10¹⁰ while the protein itself contributes with just an additional 10⁶ fold [223]. The pathways of H₄F, H₄MPT and molybdopterin biosynthesis are shown in Fig. 1. Guanosine-5′-triphosphate (GTP) is the starting point for the synthesis of all pterin branches (Fig. 1). Folate and methanopterin in addition to the pterin moiety also contain an aminobenzoic acid moiety. Moreover, these structural analogs share a common pterin precursor, 6-hydroxymethyl-7,8-dihydropteroate (dihydropteroate) (H₄MPT) and the presence of a non-pterin group, L-glutamate in case of folate and phosphoribosyl pyrophosphate (PRPP) in the case of methanopterin (Fig. 1A-B).

The bacterial chorismate branch of the pterin pathway entails two sequential steps for the formation of pABA from chorismate. Using the amine group of glutamine as donor, chorismate is first aminated to 4-amino-4-deoxy-chorismate by pabaA and pabB [162] and then converted to pABA by 4-amino-4-deoxychorismate lyase (pabC) [69]. Independent events of fusion of these genes occurred and combinations of pabAB, pabBC or even pabaABC have been reported [24,200]. Although methanogens can incorporate PABA into methanopterin if provided in the culture medium [205,212], genes from this pathway are absent and instead of pABA, methanogens incorporate 4-hydroxybenzoic acid (HB) into methanopterin [212].

The synthesis of the pterin moiety of methanopterin and folate share a common intermediate 6-HMDP and several routes for its formation are found in both domains [34,35,109] (Fig. 1A). The most well-characterized (here called 6-HMDP branch) is the bacterial route where GTP is converted to 6-HMDP by the folE (or folE2)/folB/folK route. First, GTP is hydrolyzed to 7,8-dihydropteroate triphosphate (H₄NTP) by two evolutionary unrelated enzymes, GTP cyclohydrolase IA (folE) [222] (Zinc-dependent) and GTP cyclohydrolase IB (folE2) (Zinc-independent) [166]. The phosphatases responsible for the removal of the triphosphate motif via a 7,8-dihydropteroate monophosphate (H₂NMP) intermediate are still unidentified and the resulting 7,8-dihydropteroate

---

Fig. 1. Biosynthesis of folate and pterin derivatives, redrawn from [35,67,171,186,210]. A) Synthesis of molybdopterin and 6-HMDP (last folate and methanopterin precursor) from GTP. Blue arrows represent the bacterial 6-HMDP branch and red arrows the alternative 6-HMDP archaea branch. Dashed arrows represent steps where no enzyme has been yet identified. B) Folate (blue) and methanopterin (red) biosynthesis from 6-HMDP. C) Molybdopterin biosynthesis from GTP. The black arrows represent the two steps required for molybdopterin biosynthesis common to both prokaryotic domains.
(H₂Neo) is then converted to 6-hydroxymethyl-7,8-dihydropterin (6-HMDP) by a 7,8-dihydropterin aldolase (folB) [81]. Alternatively, pyruvovinylhydropterin synthase paralogs PTPS-III [41,149] and archaeal-specific PTPS-VI [144] can catalyze the direct conversion of H₂NTP (or H₂NMP) to 6-HMD. The last step is the diphosphorylation of 6-HMD to form 6-HMDP by a diphosphokinase (folK) [187,188].

Some archaea have an Fe(II)-dependent GTP cyclohydrolase IB (MptA) homologue to folE2 that converts GTP to a cyclic 7,8-dihydropterin 2′,3′-cyclic phosphate (cPMP) intermediate [70]. cPMP is subsequently converted to H₂NMP by a Fe(II) dependent-cyclic phosphodiesterase (MptB) [116]. Recently, two alternative archaeal specific enzymes (MptD and MptE) were identified that replace folB and folK respectively [35]. The route via folE2, PTPSVI, PTPSIII, Mpta, MptB, MptD, Mpte constitutes the alternative 6-HMDP branch.

6-HMDP is the branching point between folate and methanopterin biosynthesis (Fig. 1B). The condensation of 6-HMDP with pABA is performed by a dihydropteroate synthase (folP) leading to the formation of 7,8-dihydropteroate [11,178]. A dihydrofolate synthase (folC) catalyzes the formation of dihydrofolate (H₂-folate), adding glutamate to 7,8-dihydropteroate at expenses of ATP. Some organisms have a fused folP-folC gene [103] while in other cases, the fusion is between folB and folK [35]. The conversion of H₂-folate to H₂-folate and/or folate is performed by unrelated dihydrofolate syntheses (folA and folM). The route folP/folC/folA and/or folM complete the branch to folate.

In contrast to the folate pathway, in methanopterin biosynthesis (Fig. 1C), the common intermediate 6-HMDP first reacts with a β-D-ribofuranosylaminobenzene-5-phosphate (RFAP) molecule to form 7,8-dihydropterin-6-yl-4-(β-D-ribofuranosyl)aminobenzene-5′-phosphate [212]. RFAP can be synthesized by the condensation of PRPP with either pABA or 4-hydroxybenzoic acid (HB) [44,157,212]. In both cases, the first reaction is catalyzed by a common enzyme, an RFAP-synthase that is one of the indicators of methanopterin synthesis [15,44,136,157,172].

In the first case, RFAP is directly synthesized by the condensation of pABA with PRPP catalyzed by RFAP synthase [44,157]. In the second case, after an initial condensation of PRPP and HB derived from chorismate or 3-dehydroquinate [212], two additional enzymes, MJ0815 (a possible ATP-grasp enzyme) and MJ0929 (a multifunctional adenylosuccinate lyase), are possibly necessary for the complete RFAP synthesis [212], involving the hitherto unique biological conversion of a phenol to an aniline [212].

The condensation between RFAP with 6-HMDP is performed by the dihydropteroate synthase (MptB gene) [Early MPT] [67,220]. A series of additional steps (two of them, SAM dependent) are necessary to convert this product into the final H₄MPT although the enzymes responsible for these steps are yet to be identified [67]. The route via RFAP-synthase/MptH constitutes what is called here the early methanopterin branch.

The biosynthesis of molybdopterin, the first step consists in the conversion of GTP into the stable cyclic pyranopterin monophosphate (cPMP) intermediate (Fig. 1C) [171]. The formation of the four-carbon atoms of the pyranose-ring by the insertion of the C8 atom of the purine base between the 2′ and 3′ ribose carbon atoms is catalyzed by moaA and moaC via a 5-adenosyl methionine (SAM)-dependent radical mechanism [76,77]. The previously adenylylated (by moeb [101]) molybdopterin synthase complex (moaA and moaE) catalyzes the synthesis of the molybdopterin eddithiolate by incorporating two sulfur atoms into cPMP [214]. Despite the name, molybdopterin itself lacks metal ions. Those are only incorporated in a later stage of the MoCo pathway [171]. The molybdopterin branch consists of the moa/moaC/moaD/moaE route.

3.2. Distributions of pathways across genomes

The genes involved in the different pterin biosynthetic pathways were identified in our data set of 1606 complete sequenced genomes (see Materials and methods). Fig. 2 shows the distributions of the genes for molybdopterin, H₄F and H₄MPT biosynthesis for bacterial and archaeal groups. Each column corresponds to a gene and each line to a taxonomic group. The genes are organized by the five different pathways above mentioned (alternative 6-HMDP branch, early MPT branch, 6-HMDP branch, folate branch, and molybdopterin branch). The dots within each row are colored according to the proportion of organisms within the taxon that have the gene coding for the enzyme. Genes for the molybdopterin biosynthesis are widely distributed among both prokaryotic domains and it has remained highly conserved also in eukaryotes [199].

For most of the enzymes involved in the synthesis of the common 6-HMDP intermediate of folate or methanopterin biosynthesis, a domain specific pattern is observed. For instance, although different routes for the synthesis of 6-HMDP within archaeal organisms can be found, for example folE2/PTPSVI/MptE in thermococci, folE/PTPSVI/MptE in sulfolobales, or MptA/MptB/MptD/MptE in methanogens, in the present sample, the typically bacterial folE/folB/folK route is only present in a few sulfolobales and one member of the thermanautes. Even in these cases, a fusion between the folB and folK has occurred and the genes possibly result from a recent gene acquisition from bacteria [35]. Within bacteria, the opposite picture emerges, with only a few deferriribacteres, thermotogae and deltaproteobacteria representatives having genes that allow a partial synthesis of 6-HMDP via the archaean route.

Comparing the early MPT branch with the two folate branches, this pattern is even more pronounced. Most archaeal organisms have the RFAP synthase responsible for the condensation of RFAP with pABA or HB, while in bacteria the folP/folC/folA or folM route is the preferential one. The exceptions among archaea are haloarchaea and thermoplasmatiales, which are probably due to lateral gene acquisition from bacteria [132]. There are also gene transfers for H₄MPT biosynthesis into bacteria, for example in the case of the proteobacterium Methyllobacterium extorquens, where H₄MPT is used for methanol oxidation [29,74].

3.3. Deep divergence and independent origins of unrelated genes

The genes for biosynthesis of molybdopterin are universally distributed within bacteria and archaea suggesting that they were present in their common ancestor. In contrast, there is a clear separation of the distribution of the genes from the folate and the methanopterin pathways. The H₄MPT synthesis genes appear to be conserved throughout all archaea, where those for H₄F are rare and sparsely distributed. Conversely, H₄F synthesis is present in bacteria, where methanopterin synthesis is lacking or rare, as shown in Fig. 2.

Notably, as can be seen in Figs. 1 and 2, many enzymes of pterin synthesis that lead to the same end product are the result of independent evolutionary processes. The parallel origin of enzymes that are i) ancestral for archaea and bacteria respectively, but ii) different in the two groups suggests that when the enzymes arose, they were selected to accelerate preexisting, similar and spontaneous reactions that predate the enzymes themselves. That is, it suggests that the basic underlying chemistry of the pathway is older than the enzymes that catalyze it. That is another way of saying that enzymes do not perform feats of magic, they just accelerate and add specificity to reactions that tend to occur anyway. After all, prior to the origin of genes and proteins, the chemical reactions that led to the origin of translation were by necessity spontaneous and/or catalyzed by compounds in the environment. Thus there had to be chemical reactions taking place continuously in the environment before genes arose.

At face value, there are three possible interpretations for the foregoing observation that archaea and bacteria differ in their C1 metabolism: i) The common ancestor of prokaryotes was archaean in C1 metabolism and the common ancestor of bacteria reinvented C1 metabolism by evolving the genes for folate synthesis, thereby replacing the preexisting
methanopterin pathway, and the repertoire of enzymes in the bacterial lineage that previously interacted with methanopterin C1 carriers were able to accommodate the new cofactor. ii) The common ancestor of prokaryotes was bacterial in C1 metabolism and the common ancestor of archaea reinvented C1 metabolism so as to replace the folate pathway, the converse scenario to (i) above. iii) The common ancestor of the prokaryotes had neither the folate nor the methanopterin biosynthetic pathway and its C1 metabolism was therefore more primitive, operating with C1 carriers that were spontaneously synthesized in the environment, or both. Scenarios (i) and (ii) have the problem that the gradual de novo evolution of a new and independent C1 carrier pathway in the presence of an existing one does not seem very likely. Scenario (iii) might seem radical, because it implies the existence of C1 carriers and/or chemically accessible C1 units before the advent of genes and proteins. But it is clear that neither genes nor proteins could have arisen without a preexisting, continuous and spontaneous synthesis of C1 units underpinning the synthesis of the building blocks for both.

Scenario (iii) furthermore implies the existence of a common prokaryotic ancestor that was not too different from the progenote that Woese and Fox [217] originally had in mind, namely a non-free-living entity that had a primitive ribosome and the genetic code, but not much more than that in terms of supporting biosynthetic pathways. However, the progenote was originally also seen as the direct progenitor of eukaryotes [216–218]. Subsequent findings indicated the presence of mitochondria in the eukaryote common ancestor [47,127,173,176,195] and the branching of eukaryote informational genes from within the archaeal lineage, rather than as a sister to it [33,92,215], such that today, eukaryotes are more readily understood as having arisen via symbiosis between fully-fledged prokaryotes, with the energetics of mitochondria having played the critical role in that transition [98]. In that sense, the progenote is better seen from today’s standpoint as the common ancestor of bacteria and archaea only (reviewed in [215]). A progenote that was depauperate in metabolic genes does not seem unreasonable from the standpoint of genomes, given the very few genes that are universal to all prokaryotes [95,179]. Clearly, there was a phase in early evolution during which an ancestral stock of genes and proteins was being invented (the dawn of enzymes), and it follows that ‘some’ of those inventions occurred subsequent to the divergence of the ancestral bacterial and archaeal lineages from the progenote. The question of how many ‘some’ are, and which pathways were affected remains unanswered here. Folate and methanopterin biosynthesis genes appear to belong to that class. Our findings with regard to H4F and H4MPT biosynthesis gene differences are

---

Fig. 2. Distribution of the genes involved in folate and pterin biosynthesis among 1606 prokaryotic genomes. The left part of the figure represents the organization of the selected taxonomic groups from 1606 completed sequenced genomes (117 archaeal and 1489 eubacterial). The right part of the figure represents the proportion of genomes within a taxon where the gene is present. Each column represents a different gene. Homologous proteins involved in the several steps of, the alternative 6-HMDP branch (folE2, PTPSVI, PTPSIII, MptA, MptB, MptD, MptE), Early MPT (RFAP, MptF), bacterial 6-HMDP branch (folE, folB, folK), 6-HMDP to folates branch (folD, folC, folA, folM) and molydbopterin (moaA, moaB, moaC, moaD) biosynthesis pathway were identified by BLAST. *—Some of the genes are fused together. The BLAST results were filtered for E values better than 10⁻10 and amino acid identities of at least 30%. The genes of molybdopterin biosynthesis are widely distributed within the two considered domains. On the contrary, there is a clear separation of the distribution of the genes from the folate and the methanopterin pathways. Note that sequences similar to the bacterial genes for pABA synthesis are found in almost all archaea, but biochemical evidence indicates that methanogens, as a prominent archaeal group, do not synthesize pABA by the bacterial route [148,121].
based on a broader sample than that available to Maden [109], who came to a very similar conclusion, namely that “there are few, if any, close homologues to enzymes of folate biosynthesis among Archaea that utilize H₄MPT, even in the early part of the pathway.”

As an alternative to the present interpretations, a popular theory has it that many of the molecular differences between archaea and bacteria are the result of adaptation to “energy stress” in the archaeal lineage [194]. The present data on C₁ carriers are distinctly at odds with that.
theory. Acetogens live from an even lower free energy change than methanogens [38,190], hence if energy stress was influencing the nature of the cofactors, acetogens and methanogens should use the same type of cofactors and organisms with less energy stress would use different ones. Not so. The differences in C1 carrier synthesis reflect phylogenetic divergence of the lineages that invented folate and methanopterin biosynthesis, not convergent adaptation in response to energy stress.

3.4. More than cofactors

The manifestations of the acetyl-CoA pathway in methanogens and acetogens differ in far more features than just their C1 carriers alone (Fig. 3). In the main, the acetyl-CoA pathway involves two segments: methyl synthesis and acetyl synthesis [61,155,156]. Fig. 3 is redrawn from Fuchs [61] in such a way as to underscore the differences in the methyl synthesis and acetyl synthesis [61,155,156]. Formate dehydrogenase, a Mo- or W-utilizing enzyme, provides the entry point. The Mo(0) or W(0) are reduced to Mo(VI) or W(IV), (transition-metal) catalysis. The first step of methyl synthesis in acetogens entails NADPH-dependent formate dehydrogenase, a Mo- or W-utilizing enzyme, while the first step of the methanogen pathway involves formylmethanofuran dehydrogenase, which requires MoCo, ferredoxin and methanofuran as cofactors. Only the MoCo-binding domain of these two enzymes are related at the level of sequence similarity, and none of the remaining enzymes of the methyl synthesis segment of the pathway are sequence related in comparison of acetogens and methanogens (Table 1). This is true for formyl-pterin synthesis (steps 2 vs. 8, numbered in the figure), the cyclohydrolase reaction that gives rise to the pterin bound methenyl moiety (steps 3 vs. 9), its reduction to a methylene bridge, NADH-dependent in acetogens, dependent on cofactor F420 (a flavin relative) in methanogens (steps 4 vs. 10) or the reduction to the level of methyl-pterin that is ferredoxin dependent in acetogens but F420-dependent in methanogens (steps 5 vs. 11). The foregoing reactions are all catalyzed by proteins that lack sequence similarity and have different size and subunit composition (Table 1). Only the MoCo binding subunits of the first step are related at the level of sequence similarity using BLAST [83].

The proteins that catalyze the methyltransferase reactions — steps 6 and 12, encoded by Moth_1197 in Moorella and MTBMA_c02910 and MTBMA_c02910 in Methanothermobacter — share no readily detectable sequence similarity using BLAST, but they are related at the level of structure by virtue of sharing a common TIM-barrel fold [155]. These methyltransferases connect the methyl synthesis segment with the acetyl synthesis segment in two ways, functionally by transferring the methyl group from the cobalt atom in methyl-cobinamide to a nickel atom at the A-site of CODH, a rare metal-to-metal methyl transfer reaction [185], this activity is provided by CODH/ACS itself (Holger Dobbek, personal communication).

In stark contrast to methyl synthesis, the acetyl synthesis segment of the pathway (step 13) is catalyzed by the highly conserved and highly homologous subunits of CODH/ACS in acetogens and methanogens (Table 1). Also in stark contrast to the methyl synthesis segments, which entail five different organic cofactors each in acetogens and methanogens, all of the catalysis in the acetyl synthesis segment is provided by metals, Ni and Fe (Fig. 3). The mechanism of the CODH/ACS reaction, which is redrawn here from Ragsdale [156], involves the Fd-dependent reduction of CO2 to CO at the C-Cluster of CODH, CO binding to Ni at the A-cluster of the ACS subunit, methyl transfer to Ni from CoFeSP, carbonyl insertion to form the Ni-bound acetyl group, and thiolytic removal from the enzyme via the mercapto group of coenzyme A to yield acetyl-CoA.

3.5. Structural comparisons

The example of the methylpterin:CoFeS methyltransferase, steps 6 and 12, where sequence similarity is lacking but structural similarity is present ([155,225]), is a prescient reminder that the structural features of homologous proteins are more conserved than their amino acid sequences [30,141]. However, shared structural similarity can either indicate an orthologous kind of common ancestry for two proteins (they are each others’ closest relatives at the level of structural similarity), or it can also indicate paralogous common ancestry, attributable for example to a shared common fold between two distantly related and distinct protein families. We thus undertook a search for similarities among enzymes of the Wood–Ljungdahl pathway at the level three-dimensional protein structures. Using the available structures for enzymes from the Moorella and Methanothermobacter pathways, or for the structures that those proteins most closely related to them, structural comparisons were performed (see Materials and methods). The results are summarized in Fig. 4.

We start the comparison between steps 1 and 7 (Fig. 4a). Moorella fdhA identifies Escherichia coli fdhH (PDB ID: 1AA6; [21]). E. coli fdhH has a four domain αβ ωβ structure containing selenocysteine, two molybdopterin cofactors and one Fe4S4 cluster. The M. thermoacetica sequence has 4 SCOP domains [129] belonging to the superfamilies 2Fe–2S ferredoxin-like, 4Fe–4S ferredoxins, formate dehydrogenase/DMSO reductase domains 1–3 and ADC-like (β-barrel domain). Like Moorella fdhA, E. coli fdhH finds, by sequence similarity [83], a domain of Methanothermobacter fmdB which belongs to the formate dehydrogenase/DMSO reductase domains 1–3 SCOP superfamily (thick dotted arrow in Fig. 4a). There are no structures available for fmdB. The Methanothermobacter fmdC sequence has two domains belonging to the SCOP superfamilies α subunit of glutamate synthase, C-terminal domain and the ADC-like proteins. FmdC, the molybdenum containing protein, shares 35% amino acid sequence identity to the tungsten-containing enzyme, fmdD. The Archaeoglobus fmdC (PDB ID: 2K18) structure is more similar to the VCP-like ATPase (Z-score of 8.8, rsmd 2.8, 99/185 amino acids) than it is to fdhH (Z-score of 7.6, rsmd 15.9, 108/146 amino acids), as indicated by the black arrow to VCP-like ATPase versus the grey arrow to fdhH (Fig. 4a). FmdE from Desulfitobacterium hafniense DCB–2 (PDB ID: 2GLZ) defines the FmdE-like SCOP superfamily [10] but it has no significant structural hits in PDB to other protein families using DALI. There are currently no structures in PDB for Moorella fdhB proteins. By sequence search using MOTH_2314 as query in PDB, the best hits were with the small chain of glutamate synthase (E-value 10−43) from Azospirillum brasilense (PDB ID: 2VDC chain G; [32]), and with dihydroyprimidine dehydrogenase (E-value 10−28) from Sus scrofa (PDB ID: 1GTB; [42]). The Moorella sequence has 4 SCOP domains belonging to the superfamilies Nqo1C-terminal domain-like, nucleotide-binding domain, and alpha-helical ferredoxin (twice). Thus, beyond the homologous MoCo-binding
domains, the acetogen and methanogen enzymes share a common fold, but are not structural orthologues.

For reactions 2 and 8 (Fig. 4b), formyl-H$_4$F synthetase, Fhs from *M. thermoacetica* [154] (PDB ID: 1EG7), belongs to the P-loop containing nucleoside triphosphate hydrolases SCOP superfamily (Nitrogenase iron protein-like family). The structure of formylmethanofuran:H$_4$PMT formyltransferase from *Methanopyrus kandleri*, Ftr (PDB ID: 1FTR) has an $\alpha$/$\beta$ sandwich fold and the 2 domains define the SCOP family formylmethanofuran:H$_4$PMT formyltransferase [111]. The two structures show no structural similarity (Z-score below 2) but they each find structural homologs in PDB (Fig. 4b). Thus, the methanogen and acetogen enzymes are structurally unrelated.

Turning to reactions 3 and 4, the FolD enzyme of *Moorella* is bifunctional. For reaction 3, the methylene-H$_4$F cyclohydrolase domain of *Moorella* found FolD from *Campylobacter jejuni* (PDB ID: 3P2O), as did reaction 4, catalyzed by the methylene-H$_4$F dehydrogenase domain. FolD has 2 $\alpha$/$\beta$ fold domains belonging to the aminoacid dehydrogenase-like SCOP superfamily in the N-terminal domain and NAD(P)-binding Rossmann-fold domains aminoacid dehydrogenase-like, in the C-terminal domain. The cyclohydrolase domain (reaction 3) detects significant structural similarity to malate oxidoreductase (Z-score 16.2, rmsd 3.1, 217/373 amino acids) but no significant structural similarity to the methenyl-H$_4$MPT cyclohydrolase, mch (PDB ID: 1QLM), from *M. kandleri* [66], which catalyzes reaction 9 and has two domains with a
new αβ fold, defining the methenyl-H₄MPT cyclohydrolase SCOP superfamily (Fig. 4c). For reaction 4, the methylene-H₄F dehydrogenase domain does share low pairwise structural similarity (Z-score 5.2, rmsd 3.4, 102/282 amino acids) to the methylene-H₄MPT dehydrogenase (reaction 10) from M. kandleri, Mtd (PDB ID: 1QV9), which has an αβ-domain followed by an α-helix bundle and a short β-sheet [73] and defines the F₄₂₀-dependent methylene-H₄MPT dehydrogenase SCOP family. But Mtd shares much greater structural similarity (Z-score 10.7, rmsd 2.7, 121/258 amino acids) to the MtaC methyltransferase SCOP superfamily than it does to the folate dependent enzyme (Fig. 4c). Thus, in the comparison of reactions 3 and 9, the acetogen and methanogen enzymes are structurally unrelated, while in the comparison of reactions 4 and 10, the acetogen and methanogen enzymes share a common fold but are not structural orthologs.

For reactions 5 and 11 (Fig. 4d), the Moorella methylene-H₄Freductase returned MetF from Thermus thermophilus HB8 (PDB ID: 1V93). The MetF structure contains a TIM-barrel and belongs to the FAD-linked oxidoreductase SCOP superfamily. The corresponding methanogen enzyme, methylene-H₄MPT reductase from Methanothermobacter thermautotrophicus, Mer (PDB ID: 1F07) has a TIM barrel fold with a non-prolyl cysteine bond [174] and belongs to the bacterial luciferase-like (F₄₂₀-dependent oxidoreductase) SCOP superfamily. Structurally, MetF is just as similar to betaine homocysteine methyltransferase (Z-score 15.8, rmsd 3.5, 238/348 amino acids) as it is to Mer (Z-score 15.7, rmsd 3.6, 225/312 amino acids), but Mer is much more similar to F₄₂₀-dependent alcohol dehydrogenases (Z-score 34, rmsd 2.4, 293/330 amino acids), F₄₂₀-dependent glucose-6-phosphate dehydrogenases and other members of the bacterial luciferase family [9] than it is to MetF. It should be recalled that in another acetogen, Acetobacterium woodii, which in contrast to Moorella lacks cytochromes, a different type of methylene-H₄F reductase exists [85] that is clustered in the genome next to a flavoprotein and an RnfC subunit [146] that is lacking in Moorella. Because of the very exergonic nature of the methylene-H₄F reductase reaction, it has

---

**Fig. 5.** Distribution of the genes from the WL-pathway. The left part of the figure represents the organization of the selected taxonomic groups from 1606 completed sequenced genomes (117 archaeal and 1489 eubacterial). The right part of the figure represents the proportion of genomes within a taxon where the gene is present. Each column represents a different gene and numbers correspond to the reactions presented in Table 1. Homologous proteins involved in the several steps of the acetogen type (2: Fhs, 3/4: folD, 5: metF, 6: ascE/MetF, 13: cdhδ/acsD, cdhγ/acsC, and cdhα(αβ)/acsB) and methanogen type (8: Ftr, 9: mch, 10: mtd, 11: mer, 12: cdhγ/acsC and cdhδ/acsD, 13: cdhα1−4, cdhε1−2 and cdhβ) were identified by BLAST (E-value threshold of 10⁻¹⁰ and amino acid identities of at least 20%).
been suspected to be a site of energetic coupling [226], via a mechanism that might involve electron bifurcation [146]. Thus, the Moorella and Methanotrophobacteria methylene-pterin reductase enzymes are related, but the methanogen enzyme is structurally more similar to the luciferase superfamily than it is to MetF and there exists some diversity among bacterial methylene-H4F reductases as well [85].

Reactions 6 and 12 entail the methyltransferase that catalyzes the transfer of the pterin-bound methyl group to cobamid in CoFeS. Moorella MeTr (PDB ID: 4djch chain A, B) consists of a TIM-barrel [225] and belongs to the dihydropterotate synthetase-like (thetysyladephydrofolate — utilizing methyltransferases) SCOP superfamily. Because of translational fusions, Moorella MeTr links steps 6 and 12 to CODH/ACS, reaction 13 (Fig. 4e). There is considerable structural information for the bacterial enzymes, including Moorella ([143,225]). There are no structures of the archaeal AcsD subunit, but in sequence comparisons at SUPERFAMILY [65] it returns the same SCOP superfamily as Moorella MeTr and it is well known that the archaeal and bacterial CODH/ACS are homologous [156,175], although Cdhα from M. barkeri (PDB ID: 3CF4 chain G; [64]) does not have a homologue in the Moorella enzyme. For Cdhα from M. barkeri (PDB ID: 3CF4 chain A) the structure belongs to the acetyl-CoA synthase SCOP superfamily [64]. These corrinoid-dependent methyltransferases, and methionine synthase (MetH) all share a common TIM-barrel fold and are related [14,155]. They are furthermore related to the methyltransferase system of CoFeSP [225].

Thus, through structural comparison using the DALI server [84] we found common folds among some enzymes from the methyl synthesis branch of the acetyl-CoA pathway in acetogens and methanogens (Fig. 4) where amino acid similarity was lacking (Table 1), in particular we found shared TIM barrel containing protein families. Even in the case of the methyl-pterin dehydrogenases (Fig. 4c), DALI alignments find higher similarity scores for the proteins to other families. Structural evolutionary studies of the TIM barrel fold have delivered arguments both in favor of divergent [53] and convergent evolution [102] of this domain. Moreover, even studies supporting divergent evolution do not reach a consensus of which group(s) of TIM barrel containing protein families shared a common ancestor ([6,31,130,131,158,213,227]). Thus, these weak structural similarities are consistent with the view that the two methyl-synthesis branches of the Wood–Ljungdahl pathway evolved independently in the ancestors of Moorella and Methanotrophobacter (Fig. 4). Moreover, because of the patterns of sequence conservation that we observe for the acetogen and methanogen versions of the WL among prokaryotes more generally (Fig. 5) the conclusion can be extended that, these weak structural similarities are consistent with the view that the two methyl-synthesis branches of the Wood–Ljungdahl pathway evolved independently in the ancestors of acetogens and methanogens, even though there are some common domains among enzymes. The exceptions are the corrinoid dependent methyltransferases, which are all interrelated [155].

In the acetogen–methanogen comparison of the acetyl-CoA pathway, the acetyl synthesis segment and the methyltransferase system that donates methyl groups to CODH/ACS are clearly homologous and, as previously proposed ([185,225]), evolutionarily related. The remaining enzymes of methyl synthesis segment are not related (except for a few shared domains). The methyl-pteridine:cofactor methyltransferases and corresponding domains in CoFeS and CODH/ACS belong to the same SCOP superfamily. Svetitchina et al. [185] have suggested that the more complicated methionine synthase family arose from the methyltransferase system of CoFeSP in the acetyl-CoA pathway, which would further point to the antiquity of this CO2-fixation route.

3.6. Discriminating between some alternative theories

The present findings permit discrimination between some competing alternatives for the evolution of the WL-pathway (Fig. 6). Braakman and Smith [22,23] suggested that the acetyl-CoA pathway was present in the last common ancestor of all cells (‘Luca’, which we take to mean the last common ancestor of prokaryotes), with subsequent vertical inheritance of the pathway into the bacterial and archaeal lineages, which generates the expected patterns of homology that are shown in Fig. 6a. Nitschke and Russell [133] suggested that a denitrifying methanotrophic WL-pathway was present in ‘Luca’ and that methanogenesis arose late and altogether independently from acetogenesis, which generates the expected patterns of homology that are shown in Fig. 6b. Ferry and House [56] suggested that the acetyl synthesis segment of the WL pathway was present in the last common ancestor, as did Martin and Russell [115] who, like Sousa et al. [179], furthermore suggested that the methyl synthesis segments arose independently in the acetogen and methanogen lineages. That view generates the expected patterns of homology that are shown in Fig. 6c. The observation from sequence (Table 1) and structural comparisons (Fig. 4) is summarized in Fig. 6d.

The recent investigations of Braakman and Smith [22,23] clearly predict that the enzymes of methyl synthesis in the WL-pathway should be homologous and related in acetogen-methanogen comparisons, which is however not the case. How can that be? Their “phylometabolic” method infers gene presence or absence based upon sequence annotation data rather than being based upon sequence similarity or structure similarity data. It was likely the reliance upon database annotations that led them to infer the presence of a complete WL-pathway in ‘Luca’ [22,23], a conclusion that, however, incompatible with both earlier studies [109] and with the sequence, structure and cofactor comparisons (Table 1, Figs. 3, 4) reported here.

The suggestion of Nitschke and Russell [133] that the acetyl-CoA pathway and in particular its manifestation in methanogenesis is not ancient has several problems. The schematic phylogenetic tree in their Fig. 2 that depicts methanogens arising very late in evolution is incompatible with current studies of deep phylogeny (reviewed by [215]). Furthermore, and as pointed out previously [179], the denitrifying methanotrophy model requires oxidizing conditions (that is, high concentrations of NO and/or nitrite) in the ocean; but under even the most mildly oxidizing conditions, the accumulation of reduced organic compounds at the vent ocean interface is no longer thermodynamically favorable [119]. Finally, the biological process of methanogenesis has an observable homologue at hydrothermal vents in the form of continuous geochemical synthesis of methane and other organic compounds in serpentinizing systems [52,100,151]. By contrast, even under today’s oxic atmosphere, geochemical methane oxidation, which the model of Nitschke and Russell [133] requires, has so far not been reported, likely owing to the very high bond energy of the C–H bond in methane [168], suggesting that denitrifying methanotrophy is a product of biological evolution, not the source of its origin.

In summary, the current data from sequence and structural comparisons are most compatible with a scenario in which acetyl synthesis was present in ‘Luca’, while the methyl synthesis segment arose later and independently in the stem lineages leading to acetogens and methanogens (Fig. 6d). That observation is furthermore consistent with the distribution of genes for H4F and H4MPT cofactor biosynthesis in Fig. 2.

3.7. Demands on spontaneous synthesis — asking for miracles?

As Ferry and House [56] have suggested, and as we have also suggested but for different reasons [115,179], substrate level energy conservation entailing thioester synthesis and acyl phosphate synthesis as it occurs in acetogens or some methanogens grown on CO [161] could be a sustained source of harnesable chemical energy in chemical and early biological evolution. But that, in turn, requires a sustained source of geochemical methyl groups before and during of genes and proteins [114]. How might such
compounds have been formed? McCollom et al. [120] have shown that the gas–water shift reaction:

\[
\text{CO}_2 + \text{H}_2 \rightarrow \text{CO} + \text{H}_2\text{O}
\]  

(1)

is a ready source of reactive C1 compounds (CO and formate) under simulated high-temperature (ca. 250 °C) hydrothermal vent conditions in the laboratory both with and without added iron. Over periods of hours to days, several percent of the carbon is reduced to Fischer–Tropsch type reaction products [122]. Also relevant in this context is a new enzyme of CO2 reduction in *A. woodii* that has recently been discovered, a hydrogen dependent carbon dioxide reductase that catalyzes the reduction of CO2 with H2 to formic acid. The enzyme is Mo-dependent and otherwise has no organic cofactors, only FeS clusters [140,170]. And Lost City today generates formate, at least in micromolar amounts [100], and methane — generated from CO2 via geochemical carbon.
reduction processes [121] — in millimolar amounts [151]; methane is also observed in terrestrial serpentinizing systems [52]. Thus the inference that the progenote arose and existed in an environment that was continuously supplied with chemically accessible C1 compounds is not fundamentally problematic (that is, it is compatible with present observations). Regarding methyl groups specifically, Loison et al. [108] used simulated hydrothermal conditions to react CO (ca. 40 μM) and H2S in the presence of Ni2+ at 90 °C for hours to days. They obtained mixtures of de novo synthesized products replete with methyl groups including methylsulfide, dimethylsulfide, methylethylsulfide, formate, acetate, propionate and many short branched-chain compounds, the latter indicating that the synthesis was not a standard Fischer–Tropsch type reaction. Reactions of these types serve to exemplify the kind of geochemical synthesis of reactive C1 compounds that we have in mind.

The synthesis of formyl pterins from CO2 requires an uphill energy investment (Fig. 3). But the energetics of the acetyl-CoA pathway are such that the energy rich thioester can, in principle, be synthesized continuously, provided that there is an environmental source of methyl groups, for example methyl sulfide [79,86], although abiotic methyl sulfide has not been reported from hydrothermal vents so far. Though not shown in Fig. 3, energy conservation in both methanogens and acetogens requires chemiosmotic ion pumping, because there is not enough energy in the pathway starting from H2 and CO2 to support carbon and energy metabolism via substrate level phosphorylation alone [55,61,115,128,190]. Also not shown in Fig. 3, the operation of the pathway in both acetogens and methanogens depends upon flavin-dependent electron bifurcation [26] in order to generate low potential ferredoxins with electrons from H2.

And what about the abiotic synthesis of pterins themselves? Sutherland and Whitfield [184] suggested a synthesis based on HCN. However, Heinz et al. [80] heated equimolar amounts of lysine, alanine, and glycine at 180 °C for several hours and obtained various pterins, one of which is shown in Fig. 6e. In addition to pterins, they also obtained several flavins. This type of synthesis has been repeated very often with similar results [189], and while one might complain that those conditions have little to do with alkaline hydrothermal vents, the observation raises a point worth mentioning, namely that enzymes do not create new reactions, they optimize existing ones [50]. For example, the reaction in Fig. 6f shows the biosynthesis of riboflavin, a rather complicated reaction catalyzed by the enzyme riboflavin synthase [57]. However, the reaction also occurs spontaneously at neutral pH without the help of enzymes [138]. That is of interest when one discusses flavin-based electron bifurcation as a possibly very ancient mechanism in early biochemistry [82]. It might seem virtually impossible at first sight that such complicated cofactors as flavins could arise without immense catalytic help. But such spontaneous, non-enzymatic flavin-generating reactions suggest that such compounds are more “natural” than one might think. Perhaps cofactor structures have something genuinely “predisposed” about them [184]. Even a cofactor as complicated as vitamin B12 can have many structurally component that seems to spontaneously fall into place. Eschenmoser [51] wrote about B12: “...the A/D-ring junction, regarded as the main obstacle to a chemical vitamin B12 synthesis at the outset, is in fact a structural element that is formed readily and in a variety of ways from structurally appropriate precursors [...] the same holds for other specific structural elements of the vitamin B12 molecule, including the characteristic arrangement of double bonds in the corrin chromophore, the special dimension of the macrocyclic ring of the corrin ligand, the specific attachment of the nucleotide loop to the propionic acid side chain of ring D, and the characteristic constitutional arrangement of the side chains around the ligand periphery (which vitamin B12 shares with all uroporphinoid cofactors). All these outwardly complex structural elements are found to ‘self-assembly’ with surprising ease under structurally appropriate preconditions; the amount of ‘external instruction’ required for their formation turns out to be surprisingly small in view of the complexity and specificity of these structural elements.”

3.8. A role for geochemical methyl groups in translation

In modern hydrothermal systems, methane of abiogenic origin is common [52,151]. The reducing power to convert CO2 into methane comes from the process of serpentinization [121], a series of geochemical reactions in which Fe2+ in the submarine crust reduces H2O in hydrothermal systems to H2 and inorganic carbon to abiogenic CH4, which can be present in amounts up to 10 mM in vent effluent [151]. Serpentinization has been occurring since there was liquid water on Earth [8,177]. Although methyl compounds of abiogenic origin have not been reported in hydrothermal effluent so far (Tom McCollom, personal communication), methane synthesis in serpentinizing systems is a rather sure indicator that methyl groups are being generated as intermediates in hydrothermal systems [121,169]. Accordingly, our suggestion for the source of methyl groups to feed acetyl thioester synthesis during the phase of evolution prior to the origin of either methyl synthesis branch of the Wood–Ljungdahl pathway is that these were spontaneously synthesized geochemically in serpentinizing hydrothermal systems. Are some methyl groups relics from the ancient past?

The ribosome of both archaea and bacteria is methylaed, particularly around the peptidyl transferase site [75]. In addition to the ribosome, most tRNAs contain modified bases, and a number of those modified bases are shared by bacteria and archaea [28,143]. These modifications include the introduction of sulfur atoms, thiomethyl groups, acetyl groups and isoprene groups and the like [143], but by far the most common modifications of both the ribosomal RNA and tRNA bases are methylations. One view has it that RNA methylation is a comparatively late appearance in chemical evolution [59] and might represent a kind of intermediate state in the transition from RNA to DNA as the genetic material [147]. The prevalence and universal conservation of some sulfur atoms, thiomethyl groups, acetyl groups and many methyl groups in RNA modifications would not fit very well in that view. An alternative view has it that RNA methylations and base modifications are holdovers from the chemical environment where the RNA world, the genetic code, the progenote and life arose — a chemically reactive and far from equilibrium environment rich in sulfur and replete with chemically reactive methyl groups [114,179]. In that view, the modern enzymatic introduction of such RNA base modifications would be a means to recreate the ancestral state. In other words, conserved methylations in ribosome–tRNA interactions might be a window into the workings of the ribosome at a time when bases were synthesized spontaneously, without extensive help from genes.

In metabolism, the methyl groups that are introduced into tRNA and tRNA come from S-adenosyl methionine, SAM [13,27,221] and are transferred to the 2’-OH of ribose or the bases themselves by base modifying enzymes, many of which belong to the radical SAM family [13,27,145]. Even the synthesis of other methyl-carriers such as methanopterins, require radical SAM enzymes [209]. Radical SAM enzymes have FeS clusters and generate a radical intermediate in the reaction mechanism [71]. FeS clusters and radical enzyme mechanisms are likely ancient biochemical traits [72,96] and their involvement would be consistent with the general antiquity of base methylations in RNA–tRNA interactions. The five amino acid methylations in the active site of methyl-CoM reductase [48] might also be seen as similar relics from environmental chemistry.

3.9. Archaea: ancestrally methanogenic?

The ubiquity of H4MPT biosynthesis genes among the archaea, allowing for lateral acquisition of H4F in haloarchaea, suggests that H4MPT was the C1 carrier in the archaeal common ancestor. But was the archaeal common ancestor a methanogen?

Methanogens are clearly ancient [105], biogenic methane is also clearly ancient, going back some 3.5 Ga [193], some phylogenetic analyses implicate methanogens as the ancestral archael group [92,215], methanogens are replete with metal and metal sulfide catalysts
while transferring the methyl moiety from methyl H4MPT to CoM complex of cytochrome-lacking methanogens, which pumps sodium. Good candidates for such ancestral pumping systems are the MtrA energetically independent from the geochemical ion gradient at the vent. The involvement of cytochromes in sulfate reducers prompted Decker et al. [37] to suggest that they are a more recent arrival on the evolutionary stage than methanogens that lack cytochromes, and that argument still seems robust [179]. Also, carbon and energy metabolism in methanogens both involve the acetyl-CoA pathway, but these are separated in sulfate reducers into distinct carbon and energy metabolic routes [139,153], whereby many sulfate reducers use the acetyl-CoA pathway. Separation would allow carbon and energy metabolism to evolve independently in sulfate reducers, providing freedom for diversi-ty in energy metabolic routes [167] including quinone-dependent ion pumping, while the methanogens remained more specialized due to the nature of their coupled carbon and energy metabolism.

When the theory of pyrite energetics was first presented, Wächtershäuser [196] mentioned methanogenesis and sulfur reduction with electrons from H2 as possible primordial energy sources for the archaea, before discarding both in favor of pyrite formation. Methanogenesis was discarded because it required an ‘energized coupler’ to link exergonic steps in methane generation to the endergonic reduction of CO2 and the energized coupler was seen as a product of evo-lution, hence derived, not ancient [196]. Twenty five years later, flavin based electron bifurcation is found to mediate the corresponding steps of methanogenesis, and the ‘energized coupler’ turns out to be FeS clusters in reduced ferredoxin [26,89]. Thus, methanogenesis is a good candidate for ancestral archaeal energy metabolism once again, and as Herrmann et al. [82] point out, reduced ferredoxin has the attributes of an energy currency more primitive than ATP. Furthermore, what once seemed to be an insurmountable leap in complexity, the origin of chemiosmotic coupling — an aspect that Wächtershäuser never integrated well into the otherwise robust FeS theory — can now be readily understood in the context of naturally preexisting proton gradients at alkaline hydrothermal vents [163,164]. Those natural gradients could be harnessed with the advent of genes and proteins, leaving the hardest step for last: the invention of machines to replace the preexisting ion gradient by coupling exergonic reactions to ion pumping [99], and thus become ener-getically independent from the geochemical ion gradient at the vent. Good candidates for such ancestral pumping systems are the MtrA–H complex of cytochrome-lacking methanogens, which pumps sodium while transferring the methyl moiety from methyl H2MPT to CoM [191] and the Rnf complex in cytochrome-lacking acetogens that pumps sodium while reducing NAD+ with electrons from reduced ferredoxin [18]. In both groups, the synthesis of low potential reduced ferredoxin is dependent upon electron bifurcation [26].

In that context, one could consider the sole coupling reaction of methanogens that lack cytochromes — the exergonic transfer of a methyl group from a nitrogen atom to a sulfur atom by the MtrA–H methyltransferase complex — in a more ancient light. At the fringes of a hydrothermal vent where natural ion gradients, harnessed by the universal ATPase, were dissipating, but geochemically generated methyl moieties were still abundant, such a “substrate level pumping” (methyltransferase) reaction might have required a lower level of evolutionary invention than the coupling of ion pumping to electron transfers. Might the methyltransferase reaction be the most ancient ion pumping step in the archaeal lineage? It would fit well with the notion that life arose in an environment rich in geochemical methyl groups and with the idea that methanogenesis is the earliest form of chemiosmotic energy conservation in archaea.

The deep dichotomy of archaea and bacteria is evident at the level of ribosomal protein composition [198], membrane lipid synthesis [94], cell wall constituents [2], and flagellar composition [87] in addition to other traits. That deep dichotomy is further underscored by the distribution of genes for the synthesis of H4F in bacteria and H2MPT in archaea, respectively. Furthermore, the enzymes underlying those distinct biosynthetic routes are broadly unrelated across the bacterial–archaeal divide (Fig. 2), indicating that the corresponding pathways arose independently subsequent to divergence from a progenote-like last univers-al common ancestor, which from the standpoint of gene content appears to have been more or less a geochemically fuelled ribosome-and-code complex. That deep divergence is also mirrored in the different enzymes and cofactors that archaea and bacteria use to perform methyl synthesis for methanogenesis and acetogenesis respectively (Fig. 3). While acetyl synthesis at CODH/ACS is conserved among acetogens and methanogens and involves transition metal catalysis. Thus, the acetyl-CoA pathway, while being the most ancient of known CO2 assimilation pathways [61], reflects two phases in early evolution: an ancient one in the progenote, in which acetyl thioester synthesis proceeded with the help of spontaneously (geochemically) supplied methyl groups, and a later phase that reflects the primordial divergence of the bacterial and archaeal stem groups and their independent inventions of genetically-encoded means to synthesize methyl groups via enzymatic pathways.

4. Conclusions

Of course, the acetyl-CoA pathway is not universal, it was superceded by other routes of CO2 assimilation and energy metabolism that were invented and inherited (both laterally and vertically) during microbial evolution [61,192]. But it appears to be the evolutionary starting point of carbon and energy metabolism in both archaea and bacteria — reflecting a chemistry that i) is more ancient than the genes that catalyze it today, and that ii) has a living fossil ‘sister group’ in the form of geochemical methane synthesis at modern hydrothermal vents.

The ubiquity of H2MPT biosynthesis genes among archaea is compati-ble with the view that the first free-living members of the archaeal domain were methanogens. However, new and potentially more an-cient kinds of energy metabolism involving sulfur are still being discov-ered [123]. Conserved acetyl thioester synthesis in the acetyl-CoA pathway, together with independently invented methyl synthesis path-ways using independently invented pterin C1 carriers, appear to hold clues about the energetic and chemical environment within which the progenote and its descendant stem lineages arose. The prevalence of methyl groups in the chemically modified bases in the ribosome and tRNAs also, in our view, points to the environment in which the progenote navigated the transition from geoenergetics and geosynthesis to bioenergetics and biosynthesis, within the confines of naturally forming inorganic microcompartment at a Hadean hydrothermal vent, one that was rich in reactive methyl groups, a world of one carbon compounds, or a ‘C1 world’, were one so inclined.

Acknowledgements

We thank Robert H. White and Tom McCollom for many helpful comments on an earlier version of the manuscript and Holger Dobek for very helpful information regarding methyltransferase activities and similarities in archaeal CoFeSP. We thank referee 2 for many helpful comments and leads into the literature, and referee 1 for asking us to do the structural comparisons for enzymes of the Wood–Jungendahl pathway. We thank the European Research Council for financial support.
Appendix A. Supplementary data

Supplementary data to this article can be found online at dx.doi.org/10.1016/j.bbabio.2014.02.001.

References


72] S.Q. Lang, D.A. Butterfield, M. Schulte, D.S. Kelley, M.D. Lilley, Elevated concentra-
tions of formate, acetate and dissolved organic carbon found at the Lost City hydro-


