Biochimica et Biophysica Acta xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbabio

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

ARTICLE INFO

Article history:
Received 26 October 2013
Received in revised form 31 January 2014
Accepted 3 February 2014
Available online xxxx

Keywords:
Pterins
Hydrothermal vents
Origin of life
Methanogens
Acetogens
C1-world

ABSTRACT

The deep dichotomy of archaea and bacteria is evident in many basic traits including ribosomal protein composition, membrane lipid synthesis, cell wall constituents, and flagellar composition. Here we explore that deep dichotomy further by examining the distribution of genes for the synthesis of the central carriers of one carbon units, tetrahydrofolate (H_4F) and tetrahydromethanopterin (H_4MPT), in bacteria and archaea. The enzymes underlying those distinct biosynthetic routes are broadly unrelated across the bacterial–archaeal divide, indicating that the corresponding pathways arose independently. That deep divergence in one carbon metabolism is mirrored in the structurally unrelated enzymes and different organic cofactors that methanogens (archaea) and acetogens (bacteria) use to perform methyl synthesis in their H_4F- and H_4MPT-dependent versions, respectively, of the acetyl-CoA pathway. By contrast, acetyl synthesis in the acetyl-CoA pathway — from a methyl group, CO_2 and reduced ferredoxin — is simpler, uniform and conserved across acetogens and methanogens, and involves only transition metals as catalysts. The data suggest that the acetyl-CoA pathway, while being the most ancient of known CO_2 assimilation pathways, reflects two phases in early evolution: an ancient phase in a geochemically confined and non-free-living universal common ancestor, in which acetyl thioester synthesis proceeded spontaneously with the help of geochemically supplied methyl groups, and a later phase that reflects the primordial divergence of the bacterial and archaeal stem groups, which independently invented genetically-encoded means to synthesize methyl groups via enzymatic reactions. This article is part of a Special Issue entitled: 18th European Bioenergetic Conference.

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 CO_2 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 CO_2 with H_2, 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 Fe^{2+} and H_2S was the first source of biological energy. But the pyrite theory did not forge a clear

☆ This article is part of a Special Issue entitled: 18th European Bioenergetic Conference.

* Corresponding author. Tel.: +49 211 811 3554.
E-mail address: bill@hhu.de (W.F. Martin).

http://dx.doi.org/10.1016/j.bbabio.2014.02.001

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 state of microbial energy metabolism among modern forms unless there are organisms known that actually make a living from such sources. The result is that, despite occasional differences of opinion [133,179], there has never been a heated debate specifically about the nature of the first bioenergetic systems. This might, in part, be due to the circumstance that biological energy conservation generally involves quite complicated molecular machines [1] and there exists a bewildering diversity of routes to consider [5,167], such that the question of which one(s) might be the most ancient is thorny.

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 CO_2 with electrons from H_2, 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 CO_2 with H_2 with the involvement of the acetyl-CoA pathway [26]. Among the six CO_2 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 CO_2 to methane and acetate respectively, using H_2 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 CO_2 stands out: tetrahydrofolate (H_4F) in acetogens versus tetrahydromethanopterin (H_4MPT) in methanogens [49,88,93,109]. The differences in the cofactors are of particular interest because folate is not only central to the acetyl-CoA pathway, it is more generally the universal C1 carrier in bacterial metabolism [118], where it provides C1 units for amino acid, cofactor and nucleotide biosynthesis [109,165,224] in addition to providing the methyl groups for modified bases 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 Crécy-Lagard et al. [35], who point out that the archaea, including non-methanogenic forms, generally tend to possess methanopterin or methanopterin-related C1 carriers. Exceptions to this rule are the halophiles, which possess H_4F instead of H_4MPT [19,135], and *Methanosarcina barkeri* strain fusaro, which possesses both H_4MPT and H_4F [25]. Here we examine the phylogenetic distribution of genes involved in H_4F biosynthesis and those known so far in H_4MPT 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^{-10} 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 (PHYLIP 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 *Moorella* formate dehydrogenase, ii) the B subunit of *Methanothermobacter* formylmethanofuran dehydrogenase, and iii) the β, δ, and γ subunits archaeal CODH/ACS.

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).

Please cite this article as: F.L. Sousa, W.F. Martin, Biochemical fossils of the ancient transition from geoenergetics to bioenergetics in prokaryotic one carbon compound metabolism, Biochim. Biophys. Acta (2014), http://dx.doi.org/10.1016/j.bbabio.2014.02.001