Title
Hydrogen Metabolism in Syntrophic Microorganisms

Author
Cook, Lauren

Publication Date
2019

UNIVERSITY OF CALIFORNIA

Los Angeles

Hydrogen Metabolism in Syntrophic Microorganisms

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Microbiology, Immunology, and Molecular Genetics

by

Lauren Emma Cook

2019

ABSTRACT OF THE DISSERTATION

Hydrogen Metabolism in Syntrophic Microorganisms

by

Lauren Emma Cook

Doctor of Philosophy in Microbiology, Immunology, and Molecular Genetics

University of California, Los Angeles, 2019

Professor Robert P. Gunsalus, Chair

Acetomicrobium hydrogeniformans OS1 is an obligate anaerobic bacterial species of the phylum *Synergistetes* that generates an unusually high molar ratio of hydrogen from glucose, thus suggesting an undescribed metabolic ability relative to other hydrogen-forming bacteria. Here, the genomic, proteomic, and enzymatic basis of glucose fermentation in *A. hydrogeniformans* is examined. A modified Embden-Meyerhoff pathway was revealed that employs a glyceraldehyde-3-phosphate oxidoreductase enzyme (GAPOR) in place of glyceraldehyde-3-phosphate dehydrogenase (GapA) and glycerate kinase (Pkg). *A. hydrogeniformans* cell extracts exhibited GAPOR as well as pyruvate-ferredoxin oxidoreductase (PFOR) activity in place of GapA, Pgk, pyruvate dehydrogenase (Pdh), or pyruvate formate lyase (Pfl) activity that would generate acetyl-CoA and NADH. Proteomic and enzyme studies revealed that electrons derived from GAPOR and PFOR drive hydrogen formation by a soluble FeFe-type hydrogenase. These *A. hydrogeniformans* findings demonstrate the presence of an unconventional C_6 metabolism that would explain the high hydrogen production. This is the first demonstrated example of GAPOR

in the Bacteria where this enzyme is typically found only in thermophilic archaea. Bioinformatics comparisons to related *Acetomicrobium* strains plus other genera of the phylum *Synergistetes* suggest this modified pathway is also present in other *Synergistetes* species. This dissertation also includes an examination of the genome of a methanogenic archaeon, *Methanospirillum hungatei*.

The dissertation of Lauren Emma Cook is approved.

Peter John Bradley

Beth Ann Lazazzera

Joseph Ambrose Loo

Robert P. Gunsalus, Committee Chair

University of California, Los Angeles

2019

Dedication

This dissertation is dedicated to my grandfather, Richard Remmers, “Pete” to his friends—so Pete to everybody, as the man never met a stranger. I did not get to tell him by two weeks I defended successfully, but then again, he never doubted I would.

Table of Contents

Acknowledgements

The article “Complete genome sequence of Methanospirillum hungatei type strain JF1” (1), published in Standards in Genomic Sciences in 2016, is covered by a Creative Commons Attribution License. This license permits unrestricted reuse of the material with proper attribution. Users have the right to read, download, copy, distribute, print, search, or link to the full texts of these articles.

> RPG and MJM contributed to the conception and design of this project. ALL, HED, ML, OG, NI, and NK were involved in the acquisition and initial analysis of the data; LEC, BC, LR, EM, HM, JRS, NP, HZ, RPG, and MJM were involved in the interpretation of the data. RPG prepared the first draft of the manuscript. All authors were involved in its critical revision and have given final approval of the version to be published and agree to be accountable for all aspects of the work.

The article “Genome Sequence of Acetomicrobium hydrogeniformans OS1” (2), published in the American Society for Microbiology’s Genome Announcements journal in 2018, is covered by a Creative Commons CC BY 4.0 license. This license permits unrestricted reuse of the material with proper attribution. Users have the right to read, download, copy, distribute, print, search, or link to the full texts of these articles.

> RPG was the principal investigator on this study. LEC, SSG, and AI were involved in the interpretation of the data. MM, RST, and MJM performed the initial isolation and metabolic studies of the organism. GW and EAL performed the initial genome sequence and initial analysis therein. LEC wrote the first draft of the manuscript. All authors were

involved in its critical revision and have given final approval of the version to be published and agree to be accountable for all aspects of the work.

Chapter 5 of this dissertation, “Acetomicrobium hydrogeniformans OS1 uses an archaeal-like glycolysis pathway to generate elevated hydrogen levels approaching theoretical limits”, is a manuscript in preparation for publication. The author list for this publication will be as follows: Lauren E. Cook, Janine Fu, Neil Q. Wolford, Farzaneh Sedighian, Joseph A. Loo, Ralph S. Tanner, Rachel R. Ogorzalek Loo, M. J. McInerney, and Robert P. Gunsalus.

RPG was the principal investigator on this study. LEC and RPG were involved in the interpretation of the data. JF, JAL, and RRO performed proteomics. NQW and FS cultured the organism. RST and MJM performed the enzyme activity studies and initially isolated the organism. LEC wrote the first draft of the manuscript.

Permission to reproduce figure 1 from (3) for use in this dissertation has been received from the Copyright Clearance Center. License number 4590300479047, issued May 14, 2019 on behalf of the Microbiology Society.

1. **Gunsalus RP, Cook LE, Crable B, Rohlin L, McDonald E, Mouttaki H, Sieber JR, Poweleit N, Zhou H, Lapidus AL, Daligault HE, Land M, Gilna P, Ivanova N, Kyrpides N, Culley DE, McInerney MJ.** 2016. Complete genome sequence of Methanospirillum hungatei type strain JF1. Standards in genomic sciences **11:**2-2.
2. **Cook LE, Gang SS, Ihlan A, Maune M, Tanner RS, McInerney MJ, Weinstock G, Lobos EA, Gunsalus RP.** 2018. Genome Sequence of Acetomicrobium hydrogeniformans OS1. Genome Announcements **6:**e00581-00518.
3. **Ben Hania W, Bouanane-Darenfed A, Cayol J-L, Ollivier B, Fardeau M-L.** 2016. Reclassification of Anaerobaculum mobile, Anaerobaculum thermoterrenum, Anaerobaculum hydrogeniformans as Acetomicrobium mobile comb. nov., Acetomicrobium thermoterrenum comb. nov. and Acetomicrobium hydrogeniformans comb. nov., respectively, and emendation of the genus Acetomicrobium. International Journal of Systematic and Evolutionary Microbiology **66:**1506-1509.

This dissertation owes considerable debt to the following:

In the Gunsalus Lab at UCLA: Farzaneh Sedighian, who performed cell culture for proteomics; Spencer Gang and Alicia Ihlan, who performed the preliminary genome inventory for *A. hydrogeniformans*, and Robert Gunsalus, for his mentorship, fellowship, advice, and willingness to make of me a functioning scientist.

In the Loo lab at UCLA: Joseph Loo, Rachel Loo, Janine Fu, and Hong Han, for performing quantitative proteomics and providing all the support necessary therein, and fielding constant questions on the data.

At the University of Oklahoma: Matt Maune, Ralph Tanner, and Mike McInerney, who performed the initial isolation and characterization, as well as enzyme activity assays and invaluable advice on culture.

At Washington University: George Weinstock and Elizabeth Lobos, who performed the genome sequencing, curation, and annotation.

Huiying Li at UCLA and Michael Suleski at Temple University, for generous help with any and all things related to informatics, Excel, and anything involving more programming than copying and pasting a macro.

My non-research-affiliated committee members, Beth Lazazzera and Peter Bradley, for their support and mentorship over the years.

The Department of Microbiology, Immunology, and Molecular Genetics at UCLA for awarding me a fellowship in 2016, and for being willing to constantly re-hire me for TAships to help support the lab and add to my teaching experience.

On the home front, my partner (Mike); my mother (Elaine Cook) and father (Mark Cook) and sister (Rachel); my friends, too numerous to name (but I did think of each of you writing this), for their unwavering support and belief over the years. Some of you have been there since I was a little kid who wanted to be a scientist-slash-writer-slash-whatever caught my fancy that week. Thank you for believing in me even when I didn’t believe in myself. My grandfather, to whom I dedicated this work, who bragged about his granddaughter at UCLA to anybody in Oklahoma who would listen.

Vita

Education

Ph.D. Microbiology, Immunology, and Molecular Genetics: University of California, Los Angeles (June 2019)

- Advancement to Candidacy: May 25, 2016
- 2016 Warsaw Fellowship Recipient
- Conference Presentation (Oral)
    - West Coast Bacterial Physiologists Winter 2017
- Conference Presentation (Poster)
    - UCLA-DOE Institute 2015
    - Lake Arrowhead Microbial Genomics September 2016
    - Cedars-Sinai Graduate Student Symposium Fall 2017
    - West Coast Bacterial Physiologists Winter 2017
    - ASMS 2019 Atlanta
- Outreach
    - AWiSE Students' Day 2015, 2016

B.S. Biochemistry (Medicinal Chemistry Concentration): Barrett Honors College, Arizona State University (May 2010)

- Minor: Japanese
- Honors thesis: "Identification of Tumor-Derived HLA-Binding Peptides".
- Presidential Scholar
- SOLUR (School of Life Sciences Undergraduate Research) Grant Recipient (Spring 2008-Spring 2010)

Publications

Gunsalus, Robert P; **Lauren E. Cook**, et al. "Complete genome sequence of *Methanospirillum hungatei* type strain JF1." Standards in Genomic Sciences. 2016 Jan 6;11:2. doi: 10.1186/s40793-015-0124-8. eCollection 2016.

Cook, Lauren E., Spencer S. Gang, Alicia Ihlan, Matthew Maune, Ralph S. Tanner, Michael J. McInerney, George Weinstock, Elizabeth A. Lobos, Robert P. Gunsalus "Genome Sequence of *Acetomicrobium hydrogeniformans* OS1" Genome Announc. Jun 2018, 6 (26) e00581-18; DOI: 10.1128/genomeA.00581-18

Chapter 1. Introduction: The phylum *Synergistetes*, the genus *Acetomicrobium,* and the hydrogen-producing species *Acetomicrobium hydrogeniformans*

Contents

Figures

Figure 1. Phylogenetic tree of Prokaryotic phyla with *Synergistetes* marked.

Figure 2. Neighbor-joining phylogenetic tree based on 16S rRNA gene sequence showing the relationships between Acetomicrobium hydrogeniformans and related representatives of the family *Synergistaceae.* Originally published in Ben Hania et al (1) and used with permission, copyright clearing house license 4590300479047. The preceding figure legend is taken directly from the original paper. *Acetomicrobium hydrogeniformans* is boxed in red, and the genus *Acetomicrobium* is boxed in purple. The genus *Thermovirga* diverged from a common ancestor with *Acetomicrobium*.

Figure 3. Fermentation of glucose to hydrogen gas as seen in *A. hydrogeniformans* OS1, and the combustion of hydrogen gas.

Table 1. General properties of the genus *Acetomicrobium*. Adapted from Ben Hania's review of the genus (1) and Maune's isolation paper for *A. hydrogeniformans* (2). Genome sizes and gene

numbers from IMG (3) for strains *Acetomicrobium flavidum* DSM 20664, *Acetomicrobium thermoterrenum* DSM 13490, *Acetomicrobium mobile* DSM 13181, and *Acetomicrobium hydrogeniformans* ATCC BAA-1850 (OS1).

1 The phylum *Synergistetes,* a recently-described phylum containing *Acetomicrobium hydrogeniformans*

1.1 Taxonomy

The bacterial phylum *Synergistetes* (NCBI:txid508458) was proposed in 2009 and includes species previously grouped into the family *Syntrophomonadaceae* (phylum *Firmicutes*) and the phylum *Deferribacteres.* It was originally defined by 16S rRNA gene phylogeny of 11 type strains corresponding to eight recognized genera: *Aminobacterium*, *Aminomonas*, *Acetomicrobium*, *Dethiosulfovibrio*, *Jonquetella*, *Synergistes*, *Thermovirga*, and *Thermanaerovibrio* (4, 5). For latter reference, *Acetomicrobium was formerly called Anaerobaculum (1).* The phylum was then revised to include the genera *Cloacibacillus*, *Pyramidobacter*, and *Fretibacterium* (5). It currently contains one class (*Synergistia*), one order (*Synergistales*), and one family (*Synergistaceae*) (Figure 1). A prior phylogenetic tree formed of whole genomes of the *Synergistetes* sorted species as the same monophyletic group as the 16S-based trees (6). In 2012 Bhandari and Gupta identified 32 conserved signature indels (CSIs) unique to *Synergistetes* and distributed across a variety of genes that also confirmed the 16S-based phylogeny, both in terms of which species were or were not *Synergistetes* and in terms of how those species are related within the phylum (7). A more recent 16S neighbor-joining phylogenetic tree of the *Synergistetes* shows the genus *Acetomicrobium* split off from the rest of the *Synergistetes* at the phylum root, indicating that within the confines of the phylum it is distinct from the other genera and species (1) (Figure 2). Within the genus *Acetomicrobium*, the species *A. hydrogeniformans* that groups with *A. thermoterrenum* is examined in detail in the following chapters with respect to genomics, fermentative metabolism and the whole cell proteome.

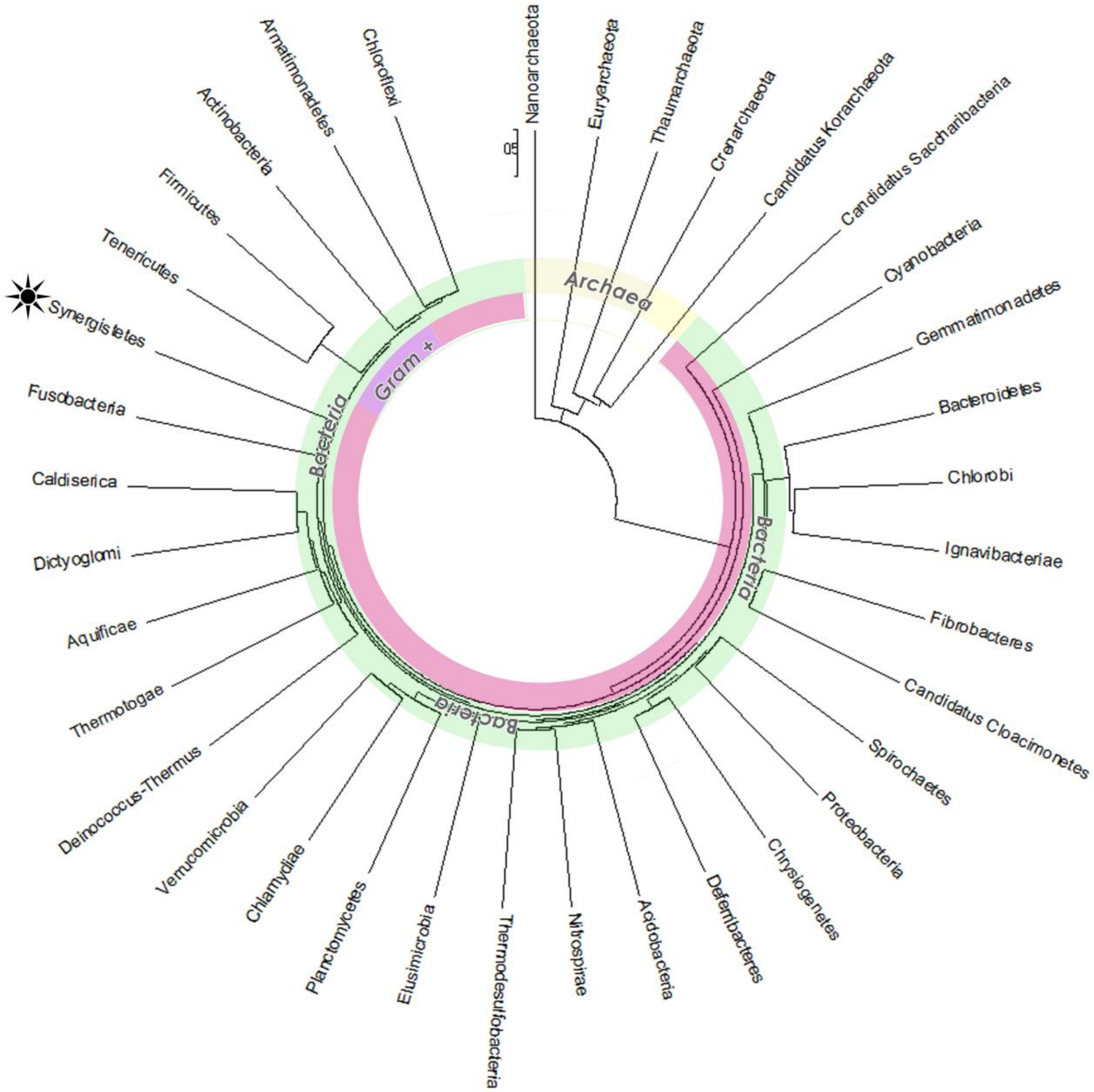

Figure 1. Phylogenetic tree of Prokaryotic phyla with *Synergistetes* marked.

Synergistetes species found in similar environments tend to cluster phylogenetically (i.e., the same *Synergistetes* are found in the same types of environment) (8). Hugenholtz et al refer to the *Synergistetes* as having “one of the most clear-cut phylogenetic placements in the tree of life; a highly reproducible monophyletic main line of descent in the bacterial domain (9)". In other words, the members of the phylum are closely related to each other, with a maximum divergence of 19% difference in 16S sequences when the phylum was defined in 2009 (9) but are phylogenetically removed from outgroup bacteria, forming a distinct clade. In 2014 Jumas-Bilak et al refined the phylum definition to include the additional genera *Cloacibacillus*, *Pyramidobacter*, and *Fretibacterium*, and proposed the inclusion of *Acetomicrobium flavidum* (then phylum *Bacteroidetes*) within the *Synergistetes* based on its 99% 16s sequence similarity with then-genus *Anaerobaculum* species. As *Acetomicrobium flavdium* was described in 1984 (10), it became the earliest-defined member of the new genus. The new *Acetomicrobium-Anaerobaculum* hybrid genus adopted the name '*Acetomicrobium*'. (1, 5) This change was formally made in 2016 (1).

1.2 Physiology and Cell Structure

Despite the wide variety of habitats, the thus-far characterized *Synergistetes* are all anaerobic, Gram-negative-staining, neutrophilic, non-spore-forming, and rod-shaped bacteria (9). Hugenholtz et al speculate this uniformity of physiology across multiple variant habitats results from similar pressures despite those differing environments. That is, as *Synergistetes* species colonize the same ecological niche in their disparate environments (e.g., for amino acid fermentation and turnover), the selection pressures under which they differentiate are quite similar (9). In most species a diderm (two-membrane) cell envelope has been seen under electron microscope (4, 5), including in *Acetomocrobium hydrogeniformans* OS1 (2). This is a phenomenon observed in other Gram-negative bacteria. However, *Synergistetes* species lack the genes for LPS biosynthesis (4, 5), and lack the TolAQR-Pal complex necessary for synthesis of the diderm membrane as seen in other phyla, indicating a potentially different membrane structure-function than that found in the described diderms (9, 11).

1.3 Nutrition and Cell Metabolism

Synergistetes species are generally considered to be amino acid metabolizers. Genes for amino acid transport and metabolism are abundant in these species (~10% of genes) whereas they contained less genes for carbohydrate utilization (~5% of genes) (5). Despite this, certain species, such as *A. flavidum*, are mostly saccharolytic, though they still metabolize amino acids. An early estimate using COG categorization of eleven *Synergistetes* genomes predicted the average phylum abundance of amino acid transport and metabolism genes at 11.8±0.8%, the highest of any bacterial phylum included in the survey, in which the mean of other phyla was 8.2±2.0% genes categorized as same (9).

Despite these genome commonalities the great diversity of the phylum is reflected in the variety of laboratory culture conditions used for the different species and may in turn reflect the great variety of native habitats, even those within the same genus (5). Routine culture of many species necessitates complex media types, thus leaving their nutritional properties relatively unstudied (R. Tanner, personal communication). Besides fermenting amino acids, certain species also metabolize dicarboxylic acids and/or use thiosulfate but no other anaerobic electron acceptors (2).

1.4 Genome Properties

Much of the knowledge of this phylum comes from genomic sequencing studies, attributable to both cultivation difficulties and their relatively low abundance in most environments. Except for select species such as *A. hydrogeniformans*, biochemical and molecular surveys of *Synergistetes* species have not been performed. Analysis is primarily limited to species identification in the metagenomic context. Many additional species even lack a sequenced genome, and are identified as OTUs based on their 16S sequence (5). Excepting the plasmid-containing species *D. peptidovorans* and *T. lienii*, the genomes of *Synergistetes* consist of a circular chromosome of 1.68 to 3.28 Mb in size with a mean 1,000 genes per Mb and with few pseudogenes. Exceptions to this pattern include the oral species *Fretibacterium fastidiosum* with a ratio closer to 500 genes/Mb. Jumas-Bilak et al hypothesize this is due to a higher concentration of pseudogenes and to a "a potential process of genomic reduction related to a narrow niche adaptation". That example aside, genome size and genetic density could not be said to correlate with any habitat. (5)

1.5 Ecology

Synergistetes have been isolated from diverse environments ranging from the human gut and oral cavity (12-15), to anaerobic digesters of both municipal wastewater treatment facilities and novel laboratory prototypes (16, 17), to laboratory artificial rumens (18), hyporheic watershed sediments (19), and pharmaceutical wastewater treatment plants (20). They are widely distributed in a variety of habitats, having been found in 90% of anaerobic habitats surveyed in one study, but within those habitats, are usually a small percentage of the microbial population, around 1% of OTUs detected (8). They are generally anaerobic amino acid metabolizers with unstudied pathways, meaning that to date most knowledge of their biology has come of cultivation-independent methods.

1.6 Roles of Synergistetes in anaerobic environments

Jumas-Bilak et al separated environmental habitats of *Synergistetes* into four general categories: anaerobic digesters, natural springs, natural seawater and sulfur mats, and host-associated microbiota (5). The environmental ubiquity and low abundance of the *Synergistetes* species in anaerobic habitats suggests major ecological roles powerfully played, a potent and necessary contribution in small doses. The *Synergistetes* gene pool may therefore serve as a source for universally-necessary but unstudied anaerobic metabolic processes, especially as involved in amino acid cycling (9). It is yet to be seen if their ubiquity indicates a singular process provided, (i.e., is amino acid metabolism the main 'service' provided to the microbiomes they inhabit) or if *Synergistetes* species specialize in a *variety* of unique but necessary metabolisms in whichever environment they exist. Indeed, in anaerobic sludge digesters and wastewater treatments, as further discussed below, they are confirmed key contributors to amino acid cycling (21).

1.7 Occurrence of *Synergistetes* in anaerobic habitats

Improvements in metagenomics sequence technology have led scientists to conduct microbial population surveys of various aquatic environments of environmental concern. From these data, patterns emerge: the microbial composition of a given environment is impacted by pollution, and the pollution, in turn, can be 'cleansed' by the microbes. The microbial community composition can also serve as an indicator of environmental health and the pollutants therein. The initial metagenomic studies seek patterns indicating an interplay of cause and effect on the environment. In these studies, the community composition of *Synergistetes* OTUs was frequently tied to environmental variables. *Synergistetes* have shown up both in studies of 'natural' water structures (rivers, streams), anthropogenically impacted, as well as artificial water structures (wastewater treatment plants, manmade lagoons).

As many of these metagenomics studies are purely a census, the precise role *Synergistetes* is playing in many of these communities remains to be elucidated. Scientists have made educated guesses of community roles based upon known metabolic pathways in *Synergistetes* and known syntrophic and antagonistic interplay between microbes, but experimental research confirming these guesses is yet forthcoming. An exception to this blind spot is anaerobic fermentation of organic compounds in manmade lagoons and reactors, upon which some metabolic studies of *Synergistetes* have been performed. *Synergistetes* have been shown to break down proteins in anaerobic organic sludge and wastewater. In these environments amino acids are further broken down to volatile fatty acids and, in culture with methanogens, to methane and CO_2 (21).

Synergistetes in the sulfur cycle: Dethiosulfovibrio in 'sulfur mats' and in natural hot springs and seawater

(Note: the genus *Dethiosulfovibrio* is not to be confused with the deltaproteobacteria genus *Desulfovibrio*)*Dethiosulfovibrio* species reduce elemental sulfur and thiosulfate, but not sulfate. In nature the *Dethiosulfovibrio* are found in *Thiodendron* sulfur "mats" (which actually symbiotic multi-species communities, despite being named after one genus) in springs and freshwater, where they may contribute to sulfur cycling (26). *Dethiosulfovibrio* species were detected via culture and metagenomics in various hot springs in Tunisia (27), and *Synergistetes* OTUs were detected in metagenomics studies of the sulfur-methane transition zone of continental margins off Denmark, an ecological niche where methane is oxidized by archaea in syntrophy with sulfur-reducing bacteria (28).

Synergistetes in polluted bodies of water

The hyporheic zone of a riverbed is at the interface of water and ground, a saturated mixture of mud, rocks, and organic matter where river water and local groundwater mix and return to the stream. It is a rich ecological niche with a distinctive biota, and has in the past twenty years become a subject of intense study (29). This is because the hyporheic zone is a flux site and a sink of organic matter wherein cycling of various compounds (carbon, nitrogen, metals, oxygen, phosphorous, etc.) and biogeochemical transformations of metals (and other contaminants of concern) occur. The zone itself contains a gradation of sunlight, oxygen, pH, redox potential, organic matter, and with that the expected attendant microbial species composition variation. (29, 30) It is estimated that in lotic (moving freshwater) environments, 76-96% of respiration occurs in this hyporheic zone, 95% of which is performed by microbes (30). It is a site of great interest to those studying environmental bioremediation and other aspects of microbiological ecology.

As these hyporheic microbial communities serve a variety of ecologically crucial functions, any disturbance of their composition is cause for concern. One may also work in the reverse direction, and use microbial community composition changes to gauge the type of contamination occurring at the river site. Furthermore, within these communities novel microbial methods of bioremediation may yet be discovered, to be applied in a variety of other scenarios, and the *Synergistetes* may well be key players in this.

Synergistetes in polluted rivers in China

Synergistetes bacteria were found in a variety of hyporheic watersheds sampled from the Maozhou river in Shenzhen, China (19). The Maozhou watershed sampling sites were in areas heavily influenced by anthropogenic activities, i.e., due to proximity to human settlements they are exposed to a variety of waste runoffs and therefore are sites in need of bioremediation, especially given the potential for groundwater contamination through the riverbed.

The water treatment infrastructure along the Maozhou is insufficient for the population burden, so a considerable amount of municipal and industrial wastewater is flushed into the river un- or undertreated. Indeed, the river is hazardous by many parameters, including total water nitrogen concentrations up to 37 times the legally ‘safe’ amount for fishing and swimming at all sample sites, and dissolved oxygen levels are below national criteria for same in 37% of sites. Water and sediment levels of manganese, iron, chloride, sulfate, and/or chromium were also beyond acceptable thresholds in at least 85% of sampling sites. Curiously, water-levels of copper, zinc, arsenic, and cadmium were below national thresholds at all sampling sites; however, sediment (i.e. within the hyporheic zone) concentrations of these metals were in most places within the ‘unacceptable’ range. Water levels of lead were not tested, but the sediment was contaminated in 57% of samples. In short, the hyporheic zone in the Maozhou river where *Synergistetes* were discovered is badly polluted, and this ranges from ‘lesser’ levels of pollution (‘upstream’ of settlements) to ‘greater’ levels (‘downstream’ of settlements).

Through deep-sequencing the researchers determined a ‘core’ microbiome for the Maozhou hyporheic zone composed of methanogens and bacteria from the phyla *Proteobacteria, Chloroflexi, Bacteroidetes, Acidobacteria, Synergistetes,* and *Firmicutes*. *Synergistetes* OTUs composed 2.3% of this ‘core’ microbiome. Within the selection of bacteria, there were a variety of Gram-negative anaerobes involved in carbon, nitrogen, and sulfur cycling. Indeed, this ‘core’ microbiome of OTUs found at every testing site consisted only of 155 out of 4446 OTUs (3.5% detected OTUs), and 622 (14%) OTUs were found only in one testing site, showing considerable diversity between sites. More site-specific studies tested the impact of environmental variables (salinity, contaminants, etc.) on this microbiome composition; however, beyond the mention of *Synergistetes* forming a small part of that ‘core’ microbiome, there was no further elaboration of its correlation with any factors (19).

Synergistetes in the severely polluted freshwater Lagos Lagoon, Nigeria

A 2016 study of sediments in Lagos Lagoon, Nigeria—which in areas is badly polluted with hydrocarbons, as well as industrial and municipal waste—detected the greatest enrichment of OTUs classified as *Synergistetes* in the sediment sample from the site most polluted with poly-aeromatic hydrocarbons and heavy metals and with the highest levels of total organic matter and sediment oxygen demand. However, the mention of *Synergistetes* was limited to a figure legend, and so further information on them specifically could not be gleaned from the paper (31). The researchers did note that this most-contaminated site at which *Synergistetes* was most prevalent harbored also a robust community of archaeal acetoclastic methanogens. As will be later elaborated, members of the *Synergistetes* including our genus of interest, *Acetomicrobium*, are acetogens (2), and as such would fit into this ecological niche. However, the authors did not specify more definite taxonomy than the phylum level.

Synergistetes in municipal and industrial wastewaters

Another major cause of water and general environmental pollution is industrial runoff from manufacturing processes. As such, their microbiology is of great interest and extensive research has been performed on waste lagoons and holding tanks associated with factories.

The wastewater runoff of the pharmaceutical manufacturing industry is unique in that it contains in concentrated form biological, hormonal, and medication-based contaminants that in more dilute form pollute municipal waste and water-sources, as well as unique pre-cursors and by-products of the manufacturing process itself. There has been earnest work on the impact of these biologicals on flora and fauna, as well as on human health, and there is keen interest in bioremediation techniques for breaking down these medication metabolites before they are re-incorporated into the environment. As with other forms of bioremediation, microbes are a potential tool. And, in addition to an already-unique complexity, the wastewater in pharmaceutical wastewater treatment plants (pWWTPs) can also contain anti-microbial compounds, to further effect the microbial balance and impact proper functioning.

Synergistetes in pharmaceuticals manufacturing wastewaters

A 2018 study (20) compared microbial compositions in sludge from pWWTPs with communities in sludge from municipal water treatment plants (mWWTPs) and from other forms of industrial water treatment (iWWTPs). The study assessed microbial communities in wastewater from pharmaceutical plants with different specializations, and therefore, different compounds in the wastewater. They found that, in addition to the pharmaceutical samples consistently displaying less microbial diversity, *Thermotogae* and *Synergistetes* were constants in the pharmaceutical samples but not the industrial or municipal samples. Not only were *Synergistetes* prevalent in that small ‘core’ group of OTUs for the various pharmaceutical wastewater types, but other species in the phylum were prevalent in one of the specific environments: *Synergistetes*, along with *Chlorobi* and *Spirochaetes*, were prevalent in the reactors processing wastewater from plants specializing in the fermentation-synthesis of vitamin C (ascorbic acid) (20).

This process, developed in China and still in use by all Chinese manufacturers of the vitamin, begins with a two-step fermentation, wherein sorbitol is fermented into sorbose by *Gluconobacter oxydans* (phylum *Proteobacteria*), and then sorbose is fermented into 2-ketogulonic acid (KGA) by *Ketogulonicigenium vulgare* (phylum *Proteobacteria*) in obligate co-culture with a variety of potential partners (32). KGA is the pre-cursor of vitamin C and is converted into such chemically. Best current evidence suggests that the culture partner for *K. vulgaris* is involved in reducing oxidative stress, and a recent study found that addition of antioxidant glutathione to the monoculture results in improved vitamin C synthesis and *K. vulgaris* growth in monoculture (33, 34). The exact partner(s) used in the plant analyzed by the Zhao group is unknown, although *Bacillus* species (phylum *Firmicutes*) have been popular partners in recent studies. It is in the runoff aggregated from these individual steps *Synergistetes* species found themselves a niche, and in them a potential clue as to *Synergistetes*' role in the community.Note that these communities were assessed in the *sludge* of these plants, and vitamin C, being highly water-soluble, is not well-absorbed into the sludge itself, and so is not a likely contributor to the microbial communities therein. The sludge itself would contain the fat-soluble intermediates and reagents.

Synergistetes in an experimental municipal wastewater plant

Recently, *Synergistetes* bacteria were found in a pilot hydrolysis-aerobic municipal wastewater treatment study (16). While anaerobic sludge reactors provide methane gas as an energy source to offset the energy cost of water purification, its application is limited in that only select climates are warm enough to support its efficient operation year-round, and the actual net energy output has been considerably lower than predicted in calculations. Various engineering efforts to improve the efficiency of anaerobic methane-producing tanks are underway. Bian and colleagues at Tsinghua University tested a prototype hybrid anaerobic hydrolysis/aerobic oxidation processing system, and found that *Synergistetes* bacteria compose 1.3% of the bacteria in the anaerobic hydrolysis culture system. The initial anaerobic hydrolysis step, in which *Synergistetes* are found, adds to the entire process an advantage over current methods in use by degrading a portion of the organics in the system *before* they reach the aeration stage. The aeration itself is energy-intensive, so any anaerobic pre-processing of the organics that cuts down on the aeration duration needed will help to lower energy costs, and the researchers predict the initial anaerobic hydrolytic degradation of organics will improve biodegradability. (16) While a small percentage of the bacterial population, and similar to the percentage of *Synergistetes* found in other studies of the microbial composition of wastewater treatment reactors, the true weight of their contribution to this process is yet unknown. *A. hydrogeniformans* is a known contributor of hydrogen to methanogens; however, the researchers did not detect any methanogens in this particular (hydrolytic) anaerobic chamber. This indicates that in this context the *Synergistetes* must be fulfilling a different community role, perhaps to do with amino acid fermentation, as is its known role in many environments. However, this was not tested.

Synergistetes in oil-production waters

Acetomicrobium hydrogeniformans was first been isolated from the processing water from an oil refinery (2). Along with the *Acetomicrobium* species, *Dethiosulfovibrio peptidovorans*, *Thermovirga lienii*, and other uncultured *Synergistetes* clones have been found in aqueous petroleum reservoirs where they ferment organic acids, as well as the production facilities for natural gas wells (4, 5). Their precise role in these environments is undetermined.

Synergistetes as a possible contributor to uranium bioremediation in contaminated sludge

In a study on the impact of uranium(VI) on the bacterial community composition of anaerobic granular sludge, the *Synergistetes* bacteria total community composition changed only minimally across a range of exposures, while the relative abundances of some other phyla fluctuated widely (17). This study was conducted from an interest in bioremediation of uranium-contaminated environments; specifically, the bacterial reduction of soluble uranium(VI) to insoluble uranium(IV), thereby reducing its dispersion through groundwater. Initial studies found that anaerobic granular sludge from wastewater treatment could convert uranium(VI), but until this study, nobody had looked at the way differing concentrations of uranium(VI) at 24hrs of exposure altered the microbial community composition. The steadiness of *Synergistetes* variety and relative abundance held even though exposure to uranium increased *overall* microbiological diversity, as measured in an increase in detected OTUs (defined in this study as greater than 97% sequence similarity). However, while there was a jump in community diversity between controls (no exposure) and exposure to low level of uranium, there was not a pattern of increasing diversity with increasing uranium concentration. This indicates a finite threshold for the ability of uranium exposure to influence community composition, achieved somewhere up to the 10uM minimum used in the study; however, this was a preliminary study with a limited sample size. A notable anomaly was a dip in community diversity at 40uM exposure beneath that detected in the control group; again, this may be a result of the limited sample size in this study. Limiting our focus only to the *Synergistetes* results, however, shows a phylum only minimally effected by uranium exposure. The biological basis of this is not yet known. A clue may well lie in seeing if diversity *within* the *Synergistetes* phylum is altered, even if the overall level of *Synergistetes*-member bacteria remains constant. This is a metabolically-diverse phylum, and any clues as to the metabolic capabilities of individual members will aid in its classification.

Synergistetes in animal hosts

Synergistetes have also been found in metagenomic surveys of a simulated rumen, in termites, in various ruminants, and in various microenvironments in the human mouth as well as human injury sites. In humans, *Synergistetes* community variations are associated with disease. As with the *Synergistetes* knowledge accumulated in environmental searches, these studies are population surveys, and experimental studies of the role *Synergistetes* play in these diseases are forthcoming. Research funding allocation toward human health concerns, and, specifically, oral microbiology, may encourage mechanistic studies that will supplement the nascent but growing knowledge of this phylum. However, as this is an exceptionally diverse phylum metabolically and environmentally, one must exercise caution in the extent to which that knowledge can be extrapolated to *Synergistetes* that have not yet been directly studied.

Synergistetes in the healthy human microbiome

Synergistetes are most prevalent in humans in the digestive tract, and one study found *Synergistetes* 16s sequences in the vaginal microbiome. (4, 5, 35, 36) In a 2012 study of healthy human adults *Synergistetes* were found in at least one upper digestive tract site (including mouth) in 58% of subjects and in feces in 8.8% of subjects. Of note, here, the *Synergistetes* distribution violates the tendency for this phylum to be ubiquitous and in low abundance: many healthy individuals completely lacked detectable *Synergistetes* at any of ten tested sites in their digestive tracts, but, for those individuals in whom *Synergistetes* were found, they were found in multiple testing sites, and represented up to 10% of the community at those sites. Those sites did follow a pattern: the *Synergistetes* species were prevalent in supragingival sites, saliva, and the palatine tonsil, but were less abundant in tongue dorsum, keratinized gingiva, and feces. This is reflected in the 58% prevalence of *Synergistetes* in at least one mouth site vs the 8.8% prevalence in stools (37). Many of these sites are spatially close, but the tissues themselves are quite different, suggesting again a highly specialized metabolic role for these *Synergistetes*. How this observation squares with the fact that many healthy humans completely lack these *Synergistetes* is yet to be determined, and must in some capacity be attributed to the ability of other microbes to perform the *Synergistetes* role in the human. These were all, insofar as could be determined, healthy subjects; a discussion of the precise way in which *Synergistetes* distribution is correlated with pathology follows later.

The prevalence of *Synergistetes* species in the human mouth means they have featured in many oral microbiome studies. A FISH-based study of subgingival oral biofilms found *Synergistetes* prevalent in the outer layer, adjacent polymorphonuclear leukocytes. The researchers suggest a potential role for *Synergistetes* in mediating host-biofilm interactions (38, 39).

The majority of *Synergistetes* human pathology research has been conducted in the context of dentistry, and these studies have been organized into their own section. However, *Synergistetes* species have been found in infection sites elsewhere in the human body, including soft tissue infections, sepsis (blood samples), and peritoneal abscess. Their precise roles in pathology are yet unknown (5).

Synergistetes involved in human dental pathologies

In a study of periodontitis *Synergistetes* species were found in both healthy and control subjects; however, *different Synergistetes* were found in healthy and control subjects. It was also found that the ‘bleeding index’ (the extent of gum bleeding upon probing) correlated positively with the variety of *Synergistetes* OTUs present, and one OTU, specifically, could be strongly correlated with presence only at diseased sites (as compared to both the gums of healthy patients and healthy gum patches in those with periodontitis) (13). Later studies confirmed that the presence of *Synergistetes* species correlated with high probing depth and bleeding on probing, both measures of disease progression (40), further cementing interest in this phylum from a medical perspective. Specifically, *Synergistetes* cluster A (an OTU designated in the paper) was present in clinical samples from patients with apical periodontitis (12).

Peri-implantitis is a distinct form of inflammation and bone loss occurring around dental implants, sometimes occurring to such extent that the implant fails. A recent study outlines contradictory results as regards the prevalence of *Synergistetes* bacteria in peri-implantitis as compared to healthy implants: an initial 2010 study found *Synergistetes* only in the diseased implants of a cohort, but a study three years later by the same group contradicted this, finding *Synergistetes* also in the healthy implants (14). The most recent study, published in 2018, detected *Synergistetes* bacteria only in one diseased implant (of ten), and concluded therefore (in conjunction with the earlier study finding *Synergistetes* equally distributed among healthy and diseased samples) that *Synergistetes* presence does not correlate with peri-implantitis. The authors attribute this discrepancy to the evolution and refinement of metagenomics techniques, which occurred at a rapid pace in the few years intervening (15). Contrast this with the above-mentioned results that placed *Synergistetes* as a prominent part of the unique *periodontitis* microbiota. Indeed, this is consistent with more recent findings that periodontitis and peri-implantitis microbiota differ more greatly that previously thought (15). So, while *Synergistetes* bacteria may not figure largely in the pathology of peri-implantitis, specifically, they still figure in certain oral pathologies (i.e. periodontitis) and are therefore of great interest in the study of oral health.

Synergistetes in the gut of non-human animals

The first *Synergistetes* species described, *S. jonesii*, was isolated in the goat rumen (41). This initial association of *Synergistetes* with ruminants has held strong through years of study, as more ruminants were found to harbor members of this phylum.

Synergistetes species were detected in the foregut of the world's sole known ruminant bird, the hoatzin, a bird of the South American tropics also aptly known as the stinkbird. The microbiota of this anomalous fellow is well-studied and compared to the biota of mammalian ruminants (mostly domestic cows) by those researching evolutionary convergence (42). That said, while *Synergistetes* were found in the hoatzin foregut, they, along with *Spirochaetes* and *Verrucomicrobia* species, constituted a higher percentage community composition in the cow rumen than the hoatzin. But, this 'higher' percentage was still around the low community average of 1% seen in other environments (42, 43). *Synergistetes* OTUs were also detected in the guts of a variety of termite species. They have also been found in the guts of wood-eating cockroaches and black tiger shrimp. These *Synergistetes* species group together phylogenetically as they group in their host species (5).

While *Synergistetes* are usually present in microbial communities in low abundances, a curious exception was found: they were on average 23% of the OTUs detected in the ceca of a population of wild Western capercaillie (*Tetrao urogallus*), or wood grouse. No *Synergistetes* OTUs were detected in the domesticated population (44). Despite these two avian species having different digestive tracts, *Synergistetes* species found in the (non-ruminating) wood grouse cecum group phylogenetically with the species found in the foregut of the (ruminating) hoatzin and *S. jonesii* from ruminating mammals (42, 44). Wild wood grouse have a varied diet consisting of berries (especially blueberries, a preferred delicacy), insects, buds, leaves, and grasses, all in season in the northern European forests they inhabit, and their microbiota changes with seasonal diet, including the proportion of *Synergistetes.* Domestic grouse had zero detectable *Synergistetes.* Wienemann et al hypothesize one reason for the high mortality of domestic grouse released into the wild may be their differing microbiota and subsequent inability to metabolize a wild diet, especially in winter, when cecum fermentation is necessary to wring all possible energy out of coniferous foliage and detoxify the resins therein (44). The role of *Synergistetes,* specifically, in this process was not studied, but following the example of their role in other ecosystems they are probably vital-though-few.

Domesticated beef cows in the United States, also, were shown to have widely varying *Synergistetes* community compositions based upon diet. FDA scientists switched a population of cattle from their 'traditional' stockyard diet of steam-flaked corn to a diet of distiller grains, a by-product of ethanol production. Cows fed on the distiller grains had a dramatic increase in the fecal community composition of *Synergistetes* (45). This is a proof-of-concept that *Synergistetes* composition can change with diet, which lends credence to the idea that like microbial composition changes in the wood grouse are diet-related. The presence-or-absence of *Synergistetes* may become an indicator of animal lifestyle. The correlation of *Synergistetes* community composition with health or other effects on the animals is yet to be discovered.

***Synergistetes* in a simulated rumen**

A study of microbial composition in a RUSITEC rumen simulation found that *Synergistetes* (along with *Bacteroidetes*) bacteria were enriched in a sample infected with *cpa* toxin-positive *Clostridium perfringens* after 8 days post-infection (or, at 15 days post the setup of the rumen;

the experiment was set with 7 days un-infected to allow the artificial rumen to settle, and 8 days of infection/'experimental' time) (18). *Lentisphaerae (phylum) was enriched at day 12 and 15 in infected samples, and Coprococcus (genus) and Firmicutes (phylum) were enriched at day 12 and 15 in non-infected samples.* Culture amino acid levels, pH, short-chain fatty acids (SCFA), redox potential, and ammonia levels remained stable throughout the trial, and putrescine was the only biogenic amine found to have decreased by day 15. The mechanism behind the enrichment of *Synergistetes* in *C. perfringens*-infected fermenters is unknown. However, the paper's conclusion was that excepting these minor fluctuations in phyla numbers, the rumen microbiota remained relatively stable during infection; the enrichment of *Synergistetes*, specifically, was not discussed as a leading point as it was not the focus of the paper. Though minor within the scope of this paper, it is of potential interest to those researchers studying *Synergistetes*, and so it is here noted. The paper did not include detail of which species within *Synergistetes* fluctuated, only that the total OTUs measured within that phylum fluctuated as a percentage of the total rumen microbiota.

2 The Genus Acetomicrobium

The genus *Acetomicrobium (*NCBI:txid49894) was most recently amended in 2016 (1) to bring the total number of named species to four. *Acetomicrobium flavidum*, originally classified in the phylum *Bacteroidetes* when discovered in 1984, has since been re-classified as a *Synergistetes* and has been given the genus name of the three species previously classified as *Anaerobaculum*. The genus now includes *A. flavidum*, *Acetomicrobium hydrogeniformans*, *Acetomicrobium mobile*, and *Acetomicrobium thermoterrenum*. A former *Acetomicrobium* member, *Caldicoprobacter faecalis*, was re-classified as a *Firmicute*. (1, 5) (Figure 2). Like *A. hydrogeniformans, A. thermoterrenum* was discovered in oil production water while *A. mobile* was discovered in an anaerobic wool-scouring wastewater treatment lagoon (2, 46).

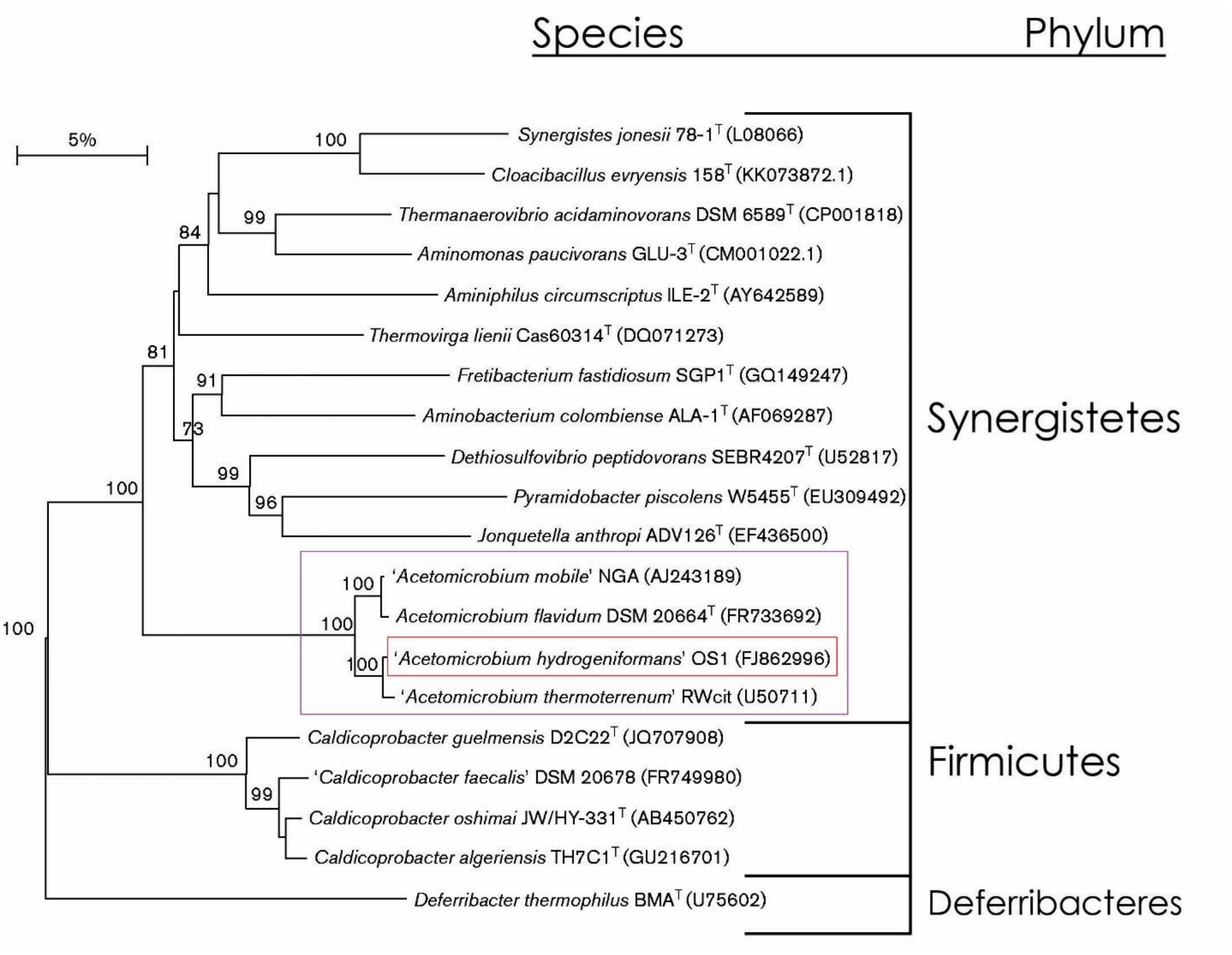

Figure 2. Neighbor-joining phylogenetic tree based on 16S rRNA gene sequence showing the relationships between *Acetomicrobium flavidum* and related representatives of the family *Synergistaceae*. Bootstrap values of 70 % or higher (based on 2000 repetitions) are shown at branch nodes. Bar, 5 substitutions per 100 nt. A phylogenetic tree was reconstructed with the TREECON program, using the neighbor-joining method (1). Tree topology was evaluated by a bootstrap analysis using 2000 resampling of the sequences (2). Its topology was also supported using the maximum-parsimony and maximum-likelihood methods. The results indicate that *Acetomicrobium flavidum* pertains to the family *Synergistaceae*, phylum *Synergistetes*, having *Anaerobaculum mobile* (99.9 % similarity) (3), *Anaerobaculum hydrogeniformans* (96.8 % similarity) (4) and *Anaerobaculum thermoterrenum* (96.4 % similarity) (5), as its closest phylogenetic relatives.

Originally published in Ben Hania et al (1) and used with permission, copyright clearing house license 4590300479047. The preceding figure legend is taken directly from the original paper. *Acetomicrobium hydrogeniformans* is boxed in red, and the genus *Acetomicrobium* is boxed in purple. The genus *Thermovirga* diverged from a common ancestor with *Acetomicrobium*.

In 2014 Jumas-Bilak et al proposed the inclusion of *Acetomicrobium flavidum* (then phylum *Bacteroidetes*) in *Synergistetes* based on its 99% 16s sequence similarity with then-genera *Anaerobaculum* species. As *Acetomicrobium flavdium* was defined in 1984 it became the earliest-defined member of the new genus, and so, the new *Acetomicrobium-Anaerobaculum* hybrid genus adopted the name ‘*Acetomicrobium*’. (1, 5) A 2016 16S neighbor-joining phylogenetic tree of the Synergistetes shows the genus *Acetomicrobium* splitting off from the rest of the *Synergistetes* at the phylum root, indicating that within the confines of the phylum it is distinct from the other species (1) (Figure 2). *A. hydrogeniformans* groups with *A. thermoterrenum* within the genus *Acetomicrobium.*

Some general properties of these strains are summarized in Table 1, which is adapted from Ben Hania’s review of the genus (1) and Maune’s isolation paper for *A. hydrogeniformans* (2). *Acetomicrobium* species have genomes that range from 44 to 51% GC mol%, and are generally around 2.0 Mb. Their optimum growth temperatures fall within 55-60°C, and optimum pH ranges acidic-to-neutral, from 6.6 to 7.6. *A. thermoterrenum* and *A. hydrogeniformans,* which group phylogenetically (figure 2), thrive at 10g/l NaCl, and, in the case of *A. hydrogeniformans*, cannot survive at salt concentrations below 0.8g/l NaCl. *A. mobile* thrives at the more modest 0.08g/l NaCl, and *A. flavidum* thrives at 0g/l, though it can tolerate concentrations up to 40g/l. As *Synergistetes*, they are all amino acid fermenters. They are also capable of fermenting some sugars. All four species can grow on D-fructose, D-glucose, pyruvate, and L-tartrate. *A. flavidum* and *A. hydrogeniformans* can grow on maltose, A. *thermoterrenum* and *A. hydrogeniformans* can grow on mannose, and all the species *but A. hydrogeniformans* can grow on glycerol.

3 Acetomicrobium hydrogeniformans, a copious hydrogen-producer

In 2012 Matthew Maune and Ralph Tanner isolated *Acetomicrobium hydrogeniformans* OS1 (DSM 22491 = ATCC BAA-1850) (until recently known as *Anaerobaculum hydrogeniformans* (1)) from wastewater from oil production in Alaska. It is the type strain for *A. hydrogeniformans* (2). During their pure- and co-culture metabolism studies they found that it produces molecular hydrogen from glucose fermentation at a rate approaching the theoretical maximum of 4 mol H_2/mol glucose (Figure 3). In clarified raw sewage, a cheap and abundant feedstock, *A. hydrogeniformans* OS1 generates 3.1 to 4.5 mmol H_2/L. (1, 2, 47) The high mole ratio of hydrogen from glucose and high partial pressures of hydrogen reached during fermentation (>10%) demonstrates the potential of OS1 for biohydrogen production (2).

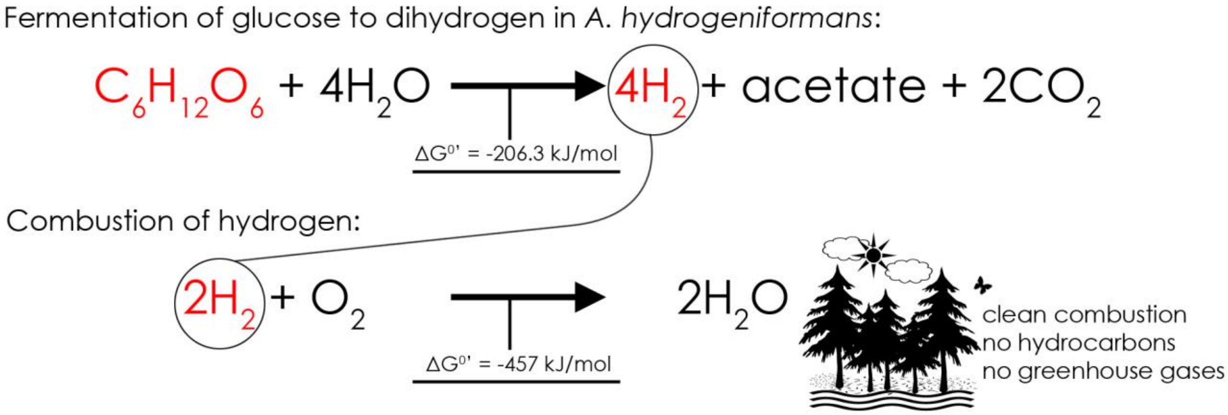

Figure 3. Fermentation of glucose to hydrogen gas as seen in *A. hydrogeniformans* OS1, and the combustion of hydrogen gas.

Molecular hydrogen is an ideal energy source given that its combustion yields water as opposed to greenhouse gases (Figure 3), and while research on hydrogen fuel cells and biohydrogen is underway, current methods of producing hydrogen gas are energy-intensive and ultimately not carbon neutral (48-50). Indeed, researchers are also examining methods of fuel production that would simultaneously break down industrial and agricultural waste products, but these strategies are currently limited to the laboratory due to financial and technical factors (48). The discovery of such an efficient hydrogen-producing organism that can work from a cheap waste product is exciting. The high mole ratio of hydrogen from glucose and high partial pressures of hydrogen reached during fermentation (>10%) demonstrates the potential of OS1 for biohydrogen production and syntrophy (2, 51). Acetate, which is co-produced with H_2 from glucose in *A. hydrogeniformans*, can be converted to the another energy-rich biofuel, methane, by acetotrophic methanogens, as can hydrogen gas itself by hydrogenotrophic methanogens (52).

4 Thesis Summary

The following thesis chapters are:

Chapter 2. “Complete genome sequence of *M. hungatei* type strain JF1”. This a reprint of an article originally published in Standards in Genomic Sciences in 2016.

Chapter 3. “Genome sequence of *A. hydrogeniformans* OS1”. This a reprint of an article originally published in Microbiology Resource Announcements (formerly known as Genome Announcements) in 2018.

Chapter 4. “Detailed analysis of the *A. hydrogeniformans* OS1 Genome”. This chapter an expansion on the report featured in the third chapter. The genome sequencing and annotation provided the basis to reconstruct the metabolism and provides a basic roadmap for subsequent work. It was from this genome annotation that we first glimpsed the metabolic uniqueness of this bacterium, with many genes annotated as being ‘archaeal’ variants.

Chapter 5. “*Acetomicrobium hydrogeniformans* OS1 uses an archaeal-like glycolysis pathway to generate elevated hydrogen levels approaching theoretical limits”, is a draft of a manuscript being prepared for publication. Our initial focus was on the carbon fermentation pathways. We found, in sum, that key electron-liberating enzymes in the pathway utilize archaeal or otherwise unusual-for-bacteria paralogues that allow for capture of low-potential electrons, and hence, for efficient reduction of hydrogen. This was a combination of enzyme activity studies and proteomics studies. Proteomics was performed on *A. hydrogeniformans* OS1 cells grown on a variety of substrates and conditions, and genes expressing abundant protein were matched with reconstructions of potential metabolic pathways. GAPOR and PFOR, ferredoxin-using enzymes, were found by enzyme activity studies to replace NADH-utilizing paralogues in the glycolysis pathway in OS1. While it can be stated with confidence based on bioinformatics analysis that genes are a member of larger enzyme families to which these enzymes belong, we could not with confidence determine which of these predicted genes was coding for that precise enzyme, with that precise substrate.

Chapter 6. “Conclusions and future directions”. Here we ask some questions that arose during the *A. hydrogeniformans* metabolism studies, but have yet to be answered. We also summarize the knowledge accumulated in this study as pertains to the phylum *Synergistetes* and *Acetomicrobium hydrogeniformans*.

Appendix I and II. “A review of the AFOR and PFOR enzyme families” is an extensive review of the AFOR (aldehyde:ferredoxin oxidoreductase) and OFOR (2-oxoacid:ferredoxin oxidoreductase) enzyme families, of which GAPOR and PFOR are members, respectively.

Tables

Table 1. General properties of the genus *Acetomicrobium*. Adapted from Ben Hania’s review of the genus (6) and Maune’s isolation paper for *A. hydrogeniformans* (4). Genome sizes and gene numbers from IMG (7) for strains *Acetomicrobium flavidum* DSM 20664, *Acetomicrobium thermoterrenum* DSM 13490, *Acetomicrobium mobile* DSM 13181, and *Acetomicrobium hydrogeniformans* ATCC BAA-1850 (OS1).

	A. flavidum	***A. thermoterrenum***	***A. mobile***	***A. hydrogeniformans***
GC content mol%	47.1	44	51.5	46.6
Genome size, MB	2.0	2.0	2.2	2.1
Gene count (assembled)	2005	1961	2109	2128
Growth temp optimum (range)	58 (35-65)	55 (28-60)	55-60 (35-65)	55 (40-65)

°C				
pH optimum (range)	7.0 (6.2-8.0)	7.0–7.6 (5.5–8.6)	6.6–7.3 (5.4–8.7)	7.0 (6.0–9.0)
NaCl optimum (range) g/l	0 (0–40)	10 (0–20)	0.08 (0–15)	10 (0.8–70)
Substrate utilization				
D-fructose	+	+	+	+
D-glucose	+	+	+	+
pyruvate	+	+	+	+
L-tartrate	+	+	+	+
Maltose	+	-	-	+
Mannose	-	+	-	+
Glycerol	+	+	+	-

References

1. **Ben Hania W, Bouanane-Darenfed A, Cayol J-L, Ollivier B, Fardeau M-L.** 2016. Reclassification of Anaerobaculum mobile, Anaerobaculum thermoterrenum, Anaerobaculum hydrogeniformans as Acetomicrobium mobile comb. nov., Acetomicrobium thermoterrenum comb. nov. and Acetomicrobium hydrogeniformans comb. nov., respectively, and emendation of the genus Acetomicrobium. International Journal of Systematic and Evolutionary Microbiology **66:**1506-1509.

2. **Maune MW, Tanner RS.** 2012. Description of Anaerobaculum hydrogeniformans sp. nov., an anaerobe that produces hydrogen from glucose, and emended description of the genus Anaerobaculum. Int J Syst Evol Microbiol **62:**832-838.

3. **Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics **25:**2271-2278.

4. **Jumas-Bilak E, Roudière L, Marchandin H.** 2009. Description of 'Synergistetes' phyl. nov. and emended description of the phylum 'Deferribacteres' and of the family Syntrophomonadaceae, phylum 'Firmicutes'. Int J Syst Evol Microbiol **59:**1028-1035

5. **Jumas-Bilak E, Marchandin H.** 2014. The Phylum Synergistetes doi:10.1007/978-3-642-38954-2_384.

6. **Beiko RG.** 2011. Telling the whole story in a 10,000-genome world. Biology Direct **6:**34.

7. **Bhandari V, Gupta RS.** 2012. Molecular signatures for the phylum Synergistetes and some of its subclades. Antonie Van Leeuwenhoek **102:**517-540.

8. **Godon J-J, Morinière J, Moletta M, Gaillac M, Bru V, Delgènes J-P.** 2005. Rarity associated with specific ecological niches in the bacterial world: the 'Synergistes' example. Environmental Microbiology **7:**213-224.

9. **Hugenholtz P, Hooper SD, Kyrpides NC.** 2009. Focus: Synergistetes. Environmental Microbiology **11:**1327-1329.

10. **Soutschek E, Winter J, Schindler F, Kandler O.** 1984. Acetomicrobium flavidum, gen. nov., sp. nov., a thermophilic, anaerobic bacterium from sewage sludge, forming acetate, CO2 and H2 from Glucose. Systematic and Applied Microbiology **5:**377-390.

11. **Gupta RS.** 2011. Origin of diderm (Gram-negative) bacteria: antibiotic selection pressure rather than endosymbiosis likely led to the evolution of bacterial cells with two membranes. Antonie van Leeuwenhoek **100:**171-182.

12. **do Cabo Fernandes C, Rechenberg D-K, Zehnder M, Belibasakis GN.** 2014. Identification of Synergistetes in endodontic infections. Microbial Pathogenesis **73:**1-6.

13. **Vartoukian SR, Palmer RM, Wade WG.** 2009. Diversity and Morphology of Members of the Phylum "Synergistetes" in Periodontal Health and Disease. Applied and Environmental Microbiology **75:**3777-3786.

14. **Koyanagi T, Sakamoto M, Takeuchi Y, Maruyama N, Ohkuma M, Izumi Y.** 2013. Comprehensive microbiological findings in peri-implantitis and periodontitis. Journal of Clinical Periodontology **40:**218-226.

15. **Al-Ahmad A, Muzafferiy F, Anderson AC, Wölber JP, Ratka-Krüger P, Fretwurst T, Nelson K, Vach K, Hellwig E.** 2018. Shift of microbial composition of peri-implantitis-associated oral biofilm as revealed by 16S rRNA gene cloning. Journal of Medical Microbiology **67:**332-340.

16. **Bian X, Gong H, Wang K.** 2018. Pilot-Scale Hydrolysis-Aerobic Treatment for Actual Municipal Wastewater: Performance and Microbial Community Analysis. International Journal of Environmental Research and Public Health **15:**477.

17. **Zeng T, Zhang S, Gao X, Wang G, Lens PNL, Xie S.** 2018. Assessment of Bacterial Community Composition of Anaerobic Granular Sludge in Response to Short-Term Uranium Exposure. Microbial Ecology doi:10.1007/s00248-018-1152-x.

18. **Wetzels SU, Eger M, Burmester M, Kreienbrock L, Abdulmawjood A, Pinior B, Wagner M, Breves G, Mann E.** 2018. The application of rumen simulation technique (RUSITEC) for studying dynamics of the bacterial community and metabolome in rumen fluid and the effects of a challenge with Clostridium perfringens. PLOS ONE **13:**e0192256.

19. **Liao H, Yu K, Duan Y, Ning Z, Li B, He L, Liu C.** 2019. Profiling microbial communities in a watershed undergoing intensive anthropogenic activities. Science of The Total Environment **647:**1137-1147.

20. **Zhao F, Ju F, Huang K, Mao Y, Zhang X-X, Ren H, Zhang T.** 2019. Comprehensive insights into the key components of bacterial assemblages in pharmaceutical wastewater treatment plants. Science of The Total Environment **651:**2148-2157.

21. **Tang Y-Q, Ji P, Hayashi J, Koike Y, Wu X-L, Kida K.** 2011. Tang Y-Q, Ji P, Hayashi J, Koike Y, Wu X-L, Kida K.. Characteristic microbial community of a dry thermophilic methanogenic digester: its long-term stability and change with feeding. Appl Microbiol Biotechnol 91: 1447-1461, vol 91.

22. **Garcia-Vallvé S, Romeu A, Palau J. 2000**. **Horizontal Gene Transfer in Bacterial and Archaeal Complete Genomes. Genome Research 10:1719-1725**.

23. **Nelson KE.** 1999. **Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. Nature 27:323-329**

24. **Deppenmeier U, Johann A, Hartsch T, Merkl R, Schmitz RA, Martinez-Arias R, Henne A, Wiezer A, Baumer S, Jacobi C, Bruggemann H, Lienard T, Christmann A, Bomeke M, Steckel S, Bhattacharyya A, Lykidis A, Overbeek R, Klenk HP, Gunsalus RP, Fritz HJ, Gottschalk G.** 2002. The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. J Mol Microbiol Biotechnol **4:**453-461.

25. **Menes RJ, Muxi L.** 2002. *Anaerobaculum mobile* sp. nov., a novel anaerobic, moderately thermophilic, peptide-fermenting bacterium that uses crotonate as an electron acceptor, and emended description of the genus *Anaerobaculum*. Int J Syst Evol Microbiol **52:**157-164.

26. **Surkov AV, Böttcher ME, Kuever J.** 2012. Sulphur isotope fractionation during the reduction of elemental sulphur and thiosulphate by Dethiosulfovibrio spp. AU Isotopes in Environmental and Health Studies **48:**65-75.

27. **Sayeh R, Birrien JL, Alain K, Barbier G, Hamdi M, Prieur D.** 2010. Microbial diversity in Tunisian geothermal springs as detected by molecular and culture-based approaches. Extremophiles **14:**501-514.

28. **Webster G, Cragg BA, Mathes F, Sass H, Parkes RJ, Weightman AJ, Green CJ, Fry JC, Gorra R, Knab NJ.** 2011. Enrichment and cultivation of prokaryotes associated with the sulphate–methane transition zone of diffusion-controlled sediments of Aarhus Bay, Denmark, under heterotrophic conditions. FEMS Microbiology Ecology **77:**248-263.

29. **Woessner WW.** 2017. Chapter 8 - Hyporheic Zones, p 129-157. *In* Hauer FR, Lamberti GA (ed), Methods in Stream Ecology, Volume 1 (Third Edition) doi:https://doi.org/10.1016/B978-0-12-416558-8.00008-1. Academic Press, Boston.

30. **Stegen JC, Fredrickson JK, Wilkins MJ, Konopka AE, Nelson WC, Arntzen EV, Chrisler WB, Chu RK, Danczak RE, Fansler SJ, Kennedy DW, Resch CT, Tfaily M.** 2016. Groundwater–surface water mixing shifts ecological assembly processes and stimulates organic carbon turnover. Nature Communications **7:**11237.

31. **Obi CC, Adebusoye SA, Ugoji EO, Ilori MO, Amund OO, Hickey WJ.** 2016. Microbial Communities in Sediments of Lagos Lagoon, Nigeria: Elucidation of Community Structure and Potential Impacts of Contamination by Municipal and Industrial Wastes. Frontiers in Microbiology **7:**1213.

32. **Wang W XH.** 2016. Industrial fermentation of Vitamin C. *In* Vandamme E RJ (ed), Industrial Biotechnology of Vitamins, Biopigments, and Antioxidants. Wiley.

33. **Ma Q, Zhang W, Zhang L, Qiao B, Pan C, Yi H, Wang L, Yuan Y-j.** 2012. Proteomic Analysis of Ketogulonicigenium vulgare under Glutathione Reveals High Demand for Thiamin Transport and Antioxidant Protection. PLoS ONE **7:**e32156.

34. **Huang M, Zhang YH, Yao S, Ma D, Yu XD, Zhang Q, Lyu SX.** 2018. Antioxidant effect of glutathione on promoting 2-keto-l-gulonic acid production in vitamin C fermentation system. Journal of Applied Microbiology **125:**1383-1395.

35. **Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Yu W-H, Lakshmanan A, Wade WG.** 2010. The Human Oral Microbiome. Journal of Bacteriology **192:**5002.

36. **The Human Microbiome Project C, Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl AM, FitzGerald MG, Fulton RS, Giglio MG, Hallsworth-Pepin K, Lobos EA, Madupu R, Magrini V, Martin JC, Mitreva M, Muzny DM, Sodergren EJ, Versalovic J, Wollam AM, Worley KC, Wortman JR, Young SK, Zeng Q, Aagaard KM, Abolude OO, Allen-Vercoe E, Alm EJ, Alvarado L, Andersen GL, Anderson S, Appelbaum E, Arachchi HM, Armitage G, Arze CA, Ayvaz T, Baker CC, Begg L, Belachew T, Bhonagiri V, Bihan M, Blaser MJ, Bloom T, Bonazzi V, Paul Brooks J, Buck GA, Buhay CJ, Busam DA, et al.** 2012. Structure, function and diversity of the healthy human microbiome. Nature **486:**207.

37. **Segata N, Haake SK, Mannon P, Lemon KP, Waldron L, Gevers D, Huttenhower C, Izard J.** 2012. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. Genome Biology **13:**R42.

38. **Zijnge V, van Leeuwen MBM, Degener JE, Abbas F, Thurnheer T, Gmür R, M. Harmsen HJ.** 2010. Oral Biofilm Architecture on Natural Teeth. PLOS ONE **5:**e9321.

39. **Vartoukian SR, Palmer RM, Wade WG.** 2007. The division “Synergistes”. Anaerobe **13:**99-106.

40. **Marchesan JT, Morelli T, Moss K, Barros SP, Ward M, Jenkins W, Aspiras MB, Offenbacher S.** 2015. Association of Synergistetes and Cyclodipeptides with Periodontitis. Journal of Dental Research **94:**1425-1431.

41. **Allison MJ, Mayberry WR, McSweeney CS, Stahl DA.** 1992. Synergistes jonesii, gen. nov., sp.nov.: A Rumen Bacterium That Degrades Toxic Pyridinediols. Systematic and Applied Microbiology **15:**522-529.

42. **Godoy-Vitorino F, Goldfarb KC, Karaoz U, Leal S, Garcia-Amado MA, Hugenholtz P, Tringe SG, Brodie EL, Dominguez-Bello MG.** 2011. Comparative analyses of foregut and hindgut bacterial communities in hoatzins and cows. The Isme Journal **6:**531.

43. **Li RW, Connor EE, Li C, Baldwin VIRL, Sparks ME.** 2012. Characterization of the rumen microbiota of pre-ruminant calves using metagenomic tools. Environmental Microbiology **14:**129-139.

44. **Wienemann T, Schmitt-Wagner D, Meuser K, Segelbacher G, Schink B, Brune A, Berthold P.** 2011. The bacterial microbiota in the ceca of Capercaillie (Tetrao urogallus) differs between wild and captive birds. Systematic and Applied Microbiology **34:**542-551.

45. **Rice WC, Galyean ML, Cox SB, Dowd SE, Cole NA.** 2012. Influence of wet distillers grains diets on beef cattle fecal bacterial community structure. BMC Microbiology **12:**25.

46. **Rees GN, Patel BK, Grassia GS, Sheehy AJ.** 1997. *Anaerobaculum thermoterrenum* gen. nov., sp. nov., a novel, thermophilic bacterium which ferments citrate. Int J Syst Evol Microbiol **47:**150-154.

47. **Thauer RK, Jungermann K, Decker K.** 1977. Energy conservation in chemotrophic anaerobic bacteria. Bacteriol Rev **41:**100-180.

48. **Dhar BR, Lee H-S.** 2013. Membranes for bioelectrochemical systems: challenges and research advances. Environmental Technology **34:**1751-1764.

49. **Davila-Vazquez G, Arriaga S, Alatriste-Mondragón F, de León-Rodríguez A, Rosales-Colunga LM, Razo-Flores E.** 2008. Fermentative biohydrogen production: trends and perspectives. Reviews in Environmental Science and Bio/Technology **7:**27-45.

50. **Kapdan IK, Kargi F.** 2006. Bio-hydrogen production from waste materials. Enzyme and Microbial Technology **38:**569-582.

51. **Sieber JR, McInerney MJ, Gunsalus RP.** 2012. Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. Annu Rev Microbiol **66:**429-452.

52. **Sieber JR, McInerney MJ, Gunsalus RP.** 2012. Genomic Insights into Syntrophy: The Paradigm for Anaerobic Metabolic Cooperation. Annual Review of Microbiology **66:**429-452.

Chapter 2. Complete genome sequence of *Methanospirillum hungatei* type strain JF1

This is a reprint of:

Gunsalus, R. P., Cook, L. E., Crable, B., Rohlin, L., McDonald, E., Mouttaki, H., Sieber, J. R., Poweleit, N., Zhou, H., Lapidus, A. L., Daligault, H. E., Land, M., Gilna, P., Ivanova, N., Kyrpides, N., Culley, D. E., McInerney, M. J. (2016). "Complete genome sequence of Methanospirillum hungatei type strain JF1". *Standards in Genomic Sciences*, *11*, 2. doi:10.1186/s40793-015-0124-8

Contents

Figures

Additional Figure 1: Best reciprocal protein hits for *M. hungatei* JF1 ORFs with other genomes

Additional Figure 2: Best BLAST hit distribution of *M. hungatei* JF1 ORFs with other genomes

Tables

Table 1: Classification and features of *Methanospirillum hungatei* strain JF1 according to MIGS recommendations published by the genomic standards consortium and the names for life database

Table 2: Project information

Table 3: Genome statistics

Table 4: Number of genes associated with general COG functional categories

1. Abstract

Methanospirillum hungatei strain JF1 (DSM 864) is a methane-producing archaeon and is the type species of the genus *Methanospirillum*, which belongs to the family *Methanospirillaceae* within the order *Methanomicrobiales*. Its genome was selected for sequencing due to its ability to utilize hydrogen and carbon dioxide and/or formate as a sole source of energy. Ecologically, *M. hungatei* functions as the hydrogen- and/or formate-using partner with many species of syntrophic bacteria. Its morphology is distinct from other methanogens with the ability to form long chains of cells (up to 100 μm in length), which are enclosed within a sheath-like structure, and terminal cells with polar flagella. The genome of *M. hungatei* strain JF1 is the first completely sequenced genome of the family *Methanospirillaceae*, and it has a circular genome of 3,544,738 bp containing 3,239 protein coding and 68 RNA genes. The large genome of *M. hungatei* JF1 suggests the presence of unrecognized biochemical/physiological properties that likely extend to the other *Methanospirillaceae* and include the ability to form the unusual sheath-like structure and to successfully interact with syntrophic bacteria.

2. Introduction

Strain JF1 (DSM 864 = ATCC 2790D-5) (1) is the type species for *M. hungatei* and represents the first isolated member of the *Methanospirillaceae* within the order *Methanomicrobiales* (2). The species epithet derives from the Latin and honors Dr. R. E. Hungate, the inventor of methodologies for modern isolation and cultivation of strictly anaerobic bacteria and archaea (3, 4). *M. hungatei* strain JF1 was isolated from a secondary anaerobic sewage treatment digestor in Urbana, Illinois, as part of a study of anaerobic aromatic hydrocarbon metabolism (5). Here, we describe the genome sequence of *M. hungatei* strain JF1, a hydrogen- and formate-utilizing, methane-producing archaean. The genomic data provide insight towards defining the unique genes needed for anaerobic syntrophy (6), which occurs within a phylogenetically diverse range of bacteria, and for classifying genes identified by environmental DNA sequencing projects.

3. Organism Information

3.1 Morphology and physiology

Cells of *Methanospirillum hungatei* strain JF1 are narrow, curved rods (i.e., spirillum shaped) that measure ~0.5 µm by ~7 µm in size (Figure 1, Table 1). The cells are contained within a sheath-like structure that contain one or more cells; the sheath may extend to over 100 µm in length depending on the nutritional conditions (1, 7). Individual cells stain Gram-negative and are weakly motile by polar tufts of flagella. Cells also possess polyphosphate bodies (PPB) or granules located at opposing cell ends (8). Growth and metabolism is strictly anaerobic where hydrogen plus carbon dioxide and/or formate serve as the methanogenic substrate. Acetate is required as the major supply for cell carbon (1, 7). Cells have no other organic nutritional requirements although addition of Casamino Acids or other plant/animal hydrolysis products speeds growth (1). Temperature range for growth is 20-40°C (optimum at 37°C).

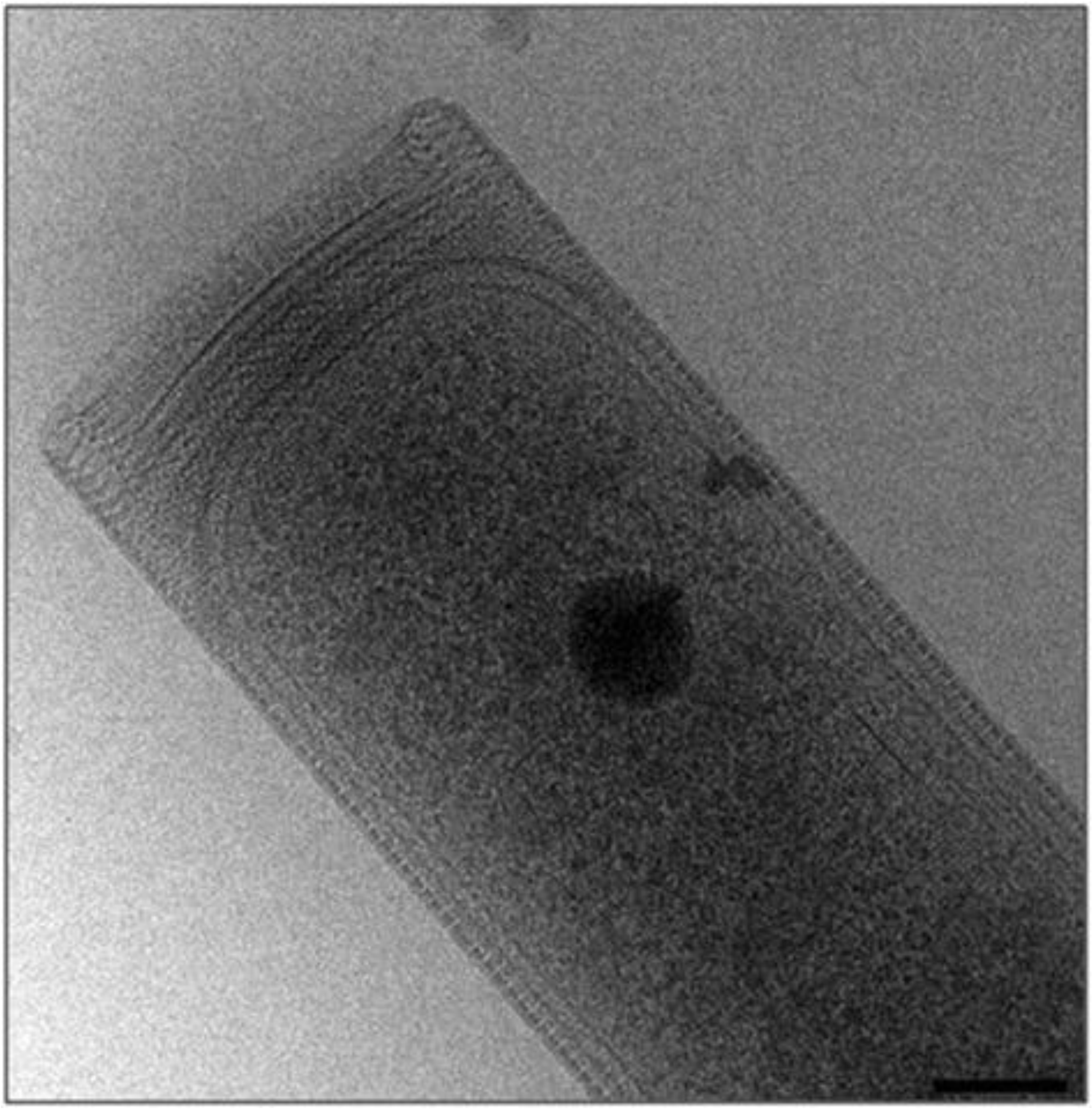

Figure 1: Electron micrograph of *M. hungatei* strain JF1 cells and associated sheath structure. Scale bar corresponds to 100 nm.

MIGS ID	Property	Term	Evidence code[a]
	Current classification	Domain *Archaea*	TAS (9)
		Phylum *Euryarchaeota*	TAS (10)
		Class *Methanomicrobia*	TAS (11)
		Order *Methanomicrobiales*	TAS (12)
		Family *Methanospirillaceae*	TAS (2)
		Genus *Methanospirillum*	TAS (1)
		Species *Methanospirillum hungatei*	TAS (1)
		Type strain JF-1	TAS (1)
	Gram stain	Negative	TAS (1)
	Cell shape	Curved rods 0.5 µM x 7.4 µM	TAS (1)
	Motility	Motile	TAS (1)
	Sporulation	Non-sporulating	TAS (1)
	Temperature range	30°C-40°C	TAS (1)
	Optimum temperature	37°C	TAS (1)
	pH range; Optimum	6.5-10; 7.0-9.0	TAS (2)
	Carbon source	carbon dioxide, formate, acetate	TAS (1)
	Energy source	hydrogen, formate	TAS (1)
	Terminal electron receptor	carbon dioxide	TAS (1)
MIGS-6	Habitat	anaerobic sediments, sewage digesters	TAS (1)
MIGS-6.3	Salinity	fresh to brackish water	TAS (1)
MIGS-22	Oxygen requirement	strict anaerobe	TAS (1)
MIGS-15	Biotic relationship	Syntrophic	TAS (1)
MIGS-14	Pathogenicity	Non-pathogen	TAS (1)
MIGS-4	Geographic location	USA, Urbana, IL	TAS (1)
MIGS-5	Sample collection time	1972	TAS (1)
MIGS-4.1	Latitude	40.109°N	NAS
MIGS-4.2	Longitude	88.204°W	NAS
MIGS-4.4	Altitude	222 m	TAS (1)

Table 1. Classification and features of *Methanospirillum hungatei* strain JF1 according to MIGS recommendations (13) published by the Genomic Standards Consortium (14) and the Names for Life database (15). Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project (16).

Biogenic methane production is important in the global carbon cycle and is used to treat sewage and other organic wastes and to produce biofuel from biomass (17, 18). The degradation of fatty and aromatic acids is often the rate-limiting step in methanogenesis (6). Fatty and aromatic acid degradation is thermodynamically favorable only when hydrogenotrophic methanogens such as *M. hungatei* strain JF1 maintain very low levels of hydrogen and/or formate in a process called syntrophy (17, 19). Members of the genus *Methanospirillum* are often detected in ecosystems where syntrophy is essential (1, 20) and *M. hungatei* strain JF1 is the model partner in syntrophic cocultures of the propionate degrader *Syntrophobacter wolinii* (21), the butyrate degrader *Syntrophomonas wolfei* (22), and the benzoate degraders *Syntrophus buswellii* and 'Syntrophus aciditrophicus' (23, 24).

3.2 Classification and features

The phylogenetic neighborhood of *M. hungatei* strain JF1 is shown in Figure 2 for representative archaeal 16S rRNA sequences belonging to the order *Methanomicrobiales*. The four described *Methanosprillum* species form a well-defined cluster distinct from the other genera within the order where *Methanosprillum lacunae* and *Methanosprillum psychrodurum* form one subgroup and *M. hungatei* plus *Methanosprillum stamsii* form another. All strains of the genus *Methanospirillum* synthesize methane from hydrogen and carbon dioxide, though the ability to use formate is variable. None are able to ferment or respire by using other electron acceptors (i.e., with sulfate, nitrate, or iron). Certain species of other genera within the *Methanomicrobiales* also use formate, and some are reported to also metabolize short chain alcohols.

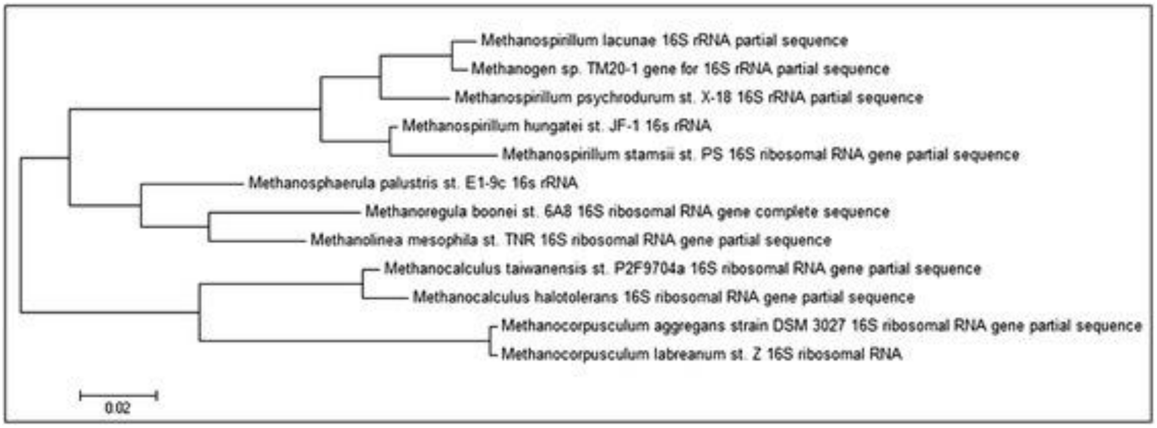

Figure 2: Phylogenetic tree highlighting the position of *Methanosprillulm hungatei* strain JF1 relative to other type strains within the *Methanomicrobiales.* The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (25). The tree with the highest log likelihood (−3033.8513) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 12 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 789 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (26).

The analysis of the four 16S rRNA genes present in the *M. hungatei* JF1 genome reveled nearly identical nucleotide sequences but they differ from one another at two positions (nucleotide positions 937 and 1382) across the 1466 nucleotide length. The previously- published 16S rRNA gene sequences (AY196683 and AB517987) used in phylogenetic investigations were incomplete, *i.e.*, 1271 and 1259 nucleotides, respectively (27, 28).

3.3 Chemotaxonomic data

The cell envelope of this Gram-negative cell wall type includes a surface layer coat, also known as a surface layer protein (SLP), which surrounds the cytoplasmic membrane, and an outermost sheath structure that encapsulates multiple cells, which are arranged in chains up to 0.1 mm in length (1, 8, 29). Cytoplasmic membrane lipids are composed primarily of biphytanyldiglycerol tetraether glycolipids (30). *M. hungatei* strain JF1 lacks *b*-or *c*-type hemes, quinones, and methanophenazine (this study). The DNA G+C content was previously reported with 45 mole% (1).

4. Genome sequencing and annotation

4.1 Genome project history

The *M. hungatei* strain JF1 genome was selected by DOE in 2004 as JGI sequencing project 364479 based on its phylogenetic position, its role in anaerobic decomposition of organic matter, and its ability to grow in co-culture with many syntrophic bacterial species (6). The genome project is deposited in the Genomes OnLine Database (GOLD) (31) as project Id:Gc00350, and the complete genome sequence is deposited in GenBank. Sequencing, finishing, and annotation of the *M. hungatei* genome were performed by the DOE Joint Genome Institute (JGI) (32). A summary of the project information is shown in Table 2.

MIGS ID	Property	Term
MIGS 31	Finishing quality	Finished
MIGS-28	Libraries used	3, 8, 14kb
MIGS 29	Sequencing platforms	Sanger
MIGS 31.2	Fold coverage	14.5X
MIGS 30	Assemblers	PGA
MIGS 32	Gene calling method	Prodical GenePRIMP
	Locus Tag	
	Genbank ID	CP000254
	GenBank Date of Release	March 1, 2006
	GOLD ID	Gc00350
	BIOPROJECT	
MIGS 13	Source Material Identifier	DSM 864T
	Project relevance	Carbon cycle, energy production

Table 2. Project information.

4.2 Growth conditions and genomic DNA preparation

M. hungatei strain JF1 was grown in basal medium under anaerobic conditions at 37° C as previously described (1). High molecular weight genomic DNA was isolated from cell pellets (DSM 864 = ATCC 2790D-5) using the CTAB method described at the JGI's web site (32).

4.3 Genome sequencing and assembly

The genome was sequenced at the Joint Genome Institute (JGI) using a combination of 3 kb, 8 kb, and 40 kb DNA libraries. All general aspects of library construction and sequencing performed are described at the JGI's web site (32). The Phred/Phrap/Consed software package (33) was used to assemble all three libraries and to assess quality (34, 35). Possible miss-assemblies were corrected and gaps between contigs were closed by editing in Consed, custom primer walks, or PCR amplification (Roche Applied Science, Indianapolis, IN). The error rate of completed genome sequence of *M. hungatei* is less than 1 in 50,000. The sequence of *M. hungatei* can be accessed using the GenBank accession number CP000254.

4.4 Genome annotation

Genes were identified using Prodical (36) as part of the Oak Ridge National Laboratory genome annotations pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline (37, 38). The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, and the UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation was preformed within the Integrated Microbial Genomes-Expert Review platform (39, 40). Membrane transport protein analysis was done by IMG with additional analysis by TransportDB (41) TCDB (42) databases. Transcription factor analysis and prediction was by assisted by TBD database (43).

5. Genome Properties

The genome statistics are provided in Table 3 and Figure 3. The genome consists of one circular chromosome of 3,544,738bp with 3,307 predicted genes of which 3,239 are protein-coding genes. Of these, approximately 61% (2,018 genes) were assigned to a putative function while the remaining 37% (1,221 genes) are without assigned functions. The genome is 45.15% G+C and 88.64% coding. The distribution of genes into COGs functional categories is presented in Table 4. Of note, six CRISPER repeats were identified on the chromosome. The *M. hungatei* genome has 51 tRNA genes; 43 have identified functions, which cover all amino acids except His. The genes for histidine biosynthesis from pyruvate are present with the exception that a gene for histidinol phosphate phosphatase (HisN) was not detected. Nutritional studies (1, 7) did not detect histidine auxotrophy, suggesting that *M. hungatei* has undescribed mechanisms for fulfilling the role of HisN and synthesizing His-tRNA.

Attribute	Value	% of Total
Genome size (bp)	3,544,738	100.00%
DNA coding (bp)	3,142,074	88.94%
DNA G+C (bp)	1,600,415	45.15%
DNA scaffolds	1	100%
Total genes	3,307	100.00%
Protein coding genes	3,239	97.94%
RNA genes	68	2.06%
Pseudo genes	99	2.99%
Genes in internal clusters	2172	65.68%
Genes with function prediction	2,018	61.02%
Genes assigned to COGs	1872	56.61%

Genes with Pfam domains	2577	77.93%
Genes with signal peptides	101	3.05%
Genes with transmembrane helices	762	23.04%
CRISPR repeats	6	

Table 3. Genome statistics.

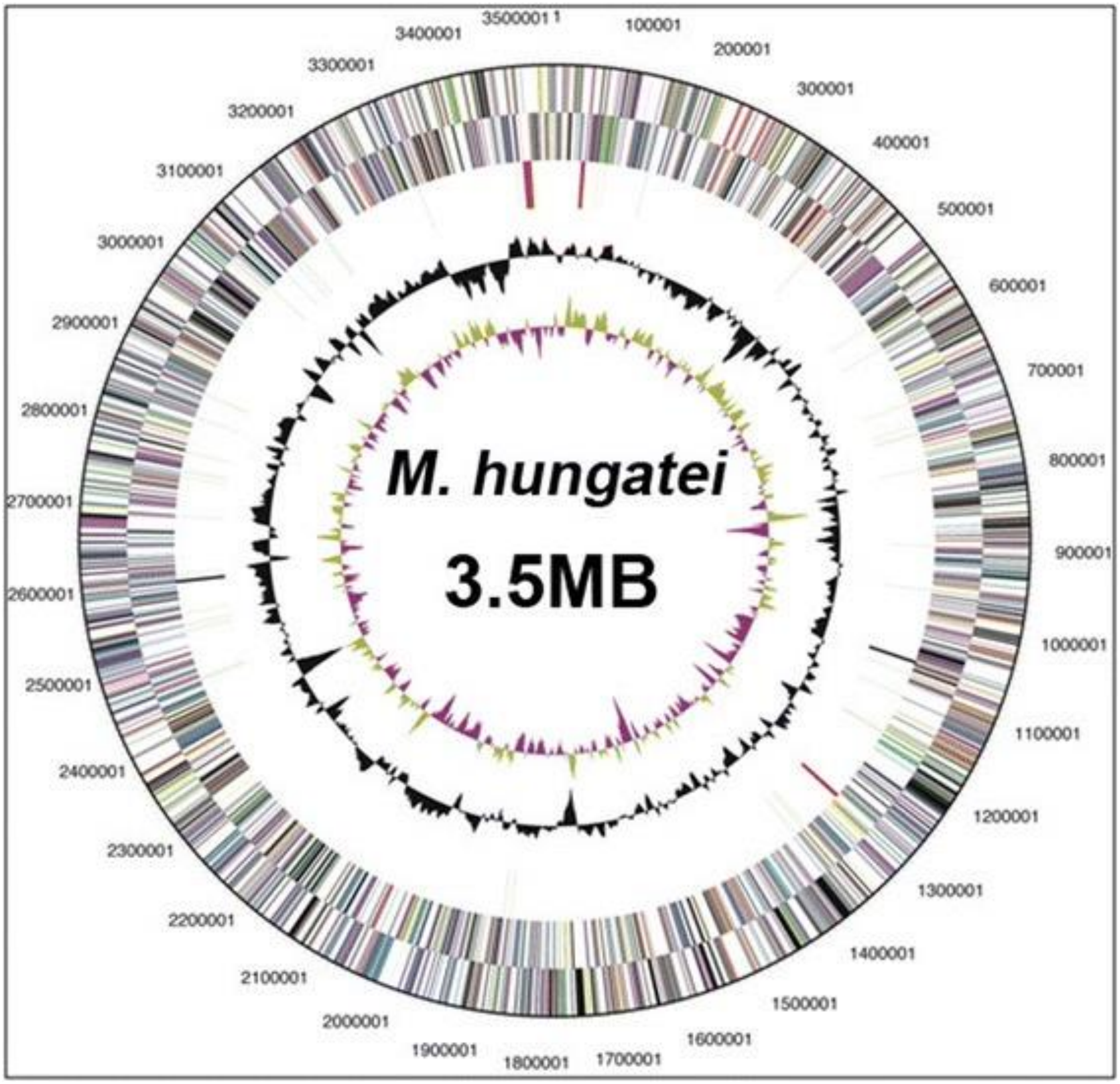

Figure 3: Graphic circular map of the *M. hungatei* JF1 chromosome. The concentric circles from outside to inside indicate: genes on the forward strand, genes on the reverse strand, RNA genes (tRNA's green, .rRNA's red, other RNA's black), GC content, and GC skew

Code	Value	%age	Description
J	180	8.65	Translation, ribosomal structure and biogenesis
A			RNA processing and modification
K	84	4.03	Transcription
L	82	3.94	Replication, recombination and repair
B	8	0.38	Chromatin structure and dynamics
D	16	0.77	Cell cycle control, Cell division, chromosome partitioning
V	53	2.55	Defense mechanisms
T	154	7.4	Signal transduction mechanisms
M	85	4.08	Cell wall/membrane biogenesis
N	54	2.59	Cell motility
U	17	0.82	Intracellular trafficking and secretion
O	89	4.27	Posttranslational modification, protein turnover, chaperones
C	186	8.93	Energy production and conversion
G	59	2.83	Carbohydrate transport and metabolism
E	165	7.93	Amino acid transport and metabolism
F	62	2.98	Nucleotide transport and metabolism
H	162	7.78	Coenzyme transport and metabolism
I	31	1.49	Lipid transport and metabolism
P	147	7.06	Inorganic ion transport and metabolism
Q	16	0.77	Secondary metabolites biosynthesis, transport and catabolism
R	217	10.42	General function prediction only
S	160	7.68	Function unknown
-	1435	43.39	Not in COGs

Table 4. Number of genes associated with general COG functional categories. The total is based on the total number of protein coding genes in the genome.

6. Insights from the genome sequence

6.1 Methanogenesis pathway

The *M. hungatei* JF1 ORFs were organized into pathways where most pathways considered essential for viability of a typical archaeal cell were detected. The methanogenic pathway from hydrogen and carbon dioxide is highly conserved in methanogens and the genes for all the enzymes in the central methanogenic pathway were identified, including a soluble-type heterodisulfide reductase only (Figure 4). The genome contains three gene sets for molybdenum (*fmd*) or tungsten (*fwd*) type formylmethanofuran (MFR) dehydrogenases (Mhun_1981-84, Mhun_1985-94 and Mhun_210612) that catalyze the ferredoxin-dependent first step of carbon dioxide reduction. There are three genes for methenyl–H_4MPT tetrahydromethanopterin (H_4MPT) cyclohydrolyase (Mch: Mhun_0022, Mhun_0444, Mhun _2384), which catalyze the third pathway step.

Single genes encode enzymes for the second, fourth, and fifth pathway steps, formylMFR:tetrahydromethanopterin formyl transferase (Ftr: Mhun_1808), methylene–H_4MPT dehydrogenase (Mtd: Mhun_2255) and methylene –H_4MPT reductase (Mer: Mhun_2257). The latter two enzymes employ reduced cofactor F_{420} as substrate. The remaining two enzymes in the pathway are multi-subunit complexes: H_4MPT Smethyltransferase (Mtr: Mhun_2168-75), and the type I methyl-CoM reductase (Mcr: Mhun_2144-2148). The CoM-S-S-CoB heterodisulfide reductase (Hdr: Mhun_1834-39) so named for the methanogenic co-enzymes M and B, reduces CoM-S-S-CoB hertodisulfide generated by Mcr. The reaction catalyzed by a soluble-type Hdr is likely an electron bifurcation, which couples the energetically favorable reduction of CoM-S-SCoB by formate and/or H_2 with the energetically unfavorable reduction of ferredoxin by formate and/or H_2 (44).

The oxidation of hydrogen or formate is needed to generate reduced ferredoxin and cofactor F_{420} used in several of the above reactions (Figure 4). The oxidation of hydrogen or formate may be accomplished by one or more of the multiple hydrogenase and formate dehydrogenase enzymes. Five nearly identical gene clusters encode soluble formate dehydrogenase (Fdh) enzymes: Mhun_1813-1814, Mhun_1832-1833, Mhun_2020-2021, Mhun_2022-2023, and Mhun_3237-3238. There are two formate/nitrite-type transporters

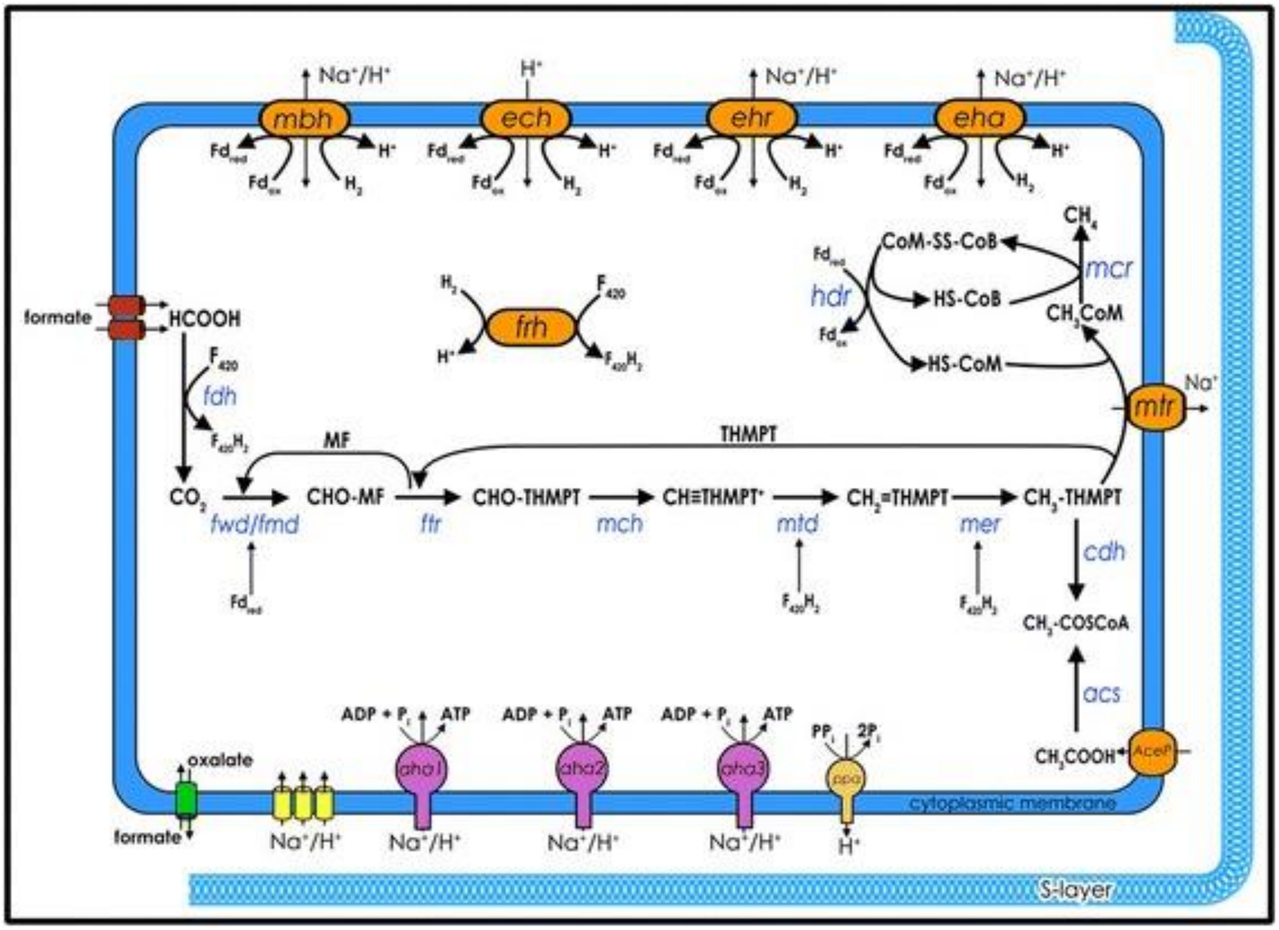

Figure 4: Overview of central metabolism in *M. hungatei* strain JF1. The pathway for methane formation from hydrogen and formate is shown in black with key steps shown with gene/enzyme designations. Membrane proteins involved in energy transduction electron transport, and ion/solute translocation are arranged along the cytoplasmic membrane: archaeal ATP synthase, Aha; formate dehydrogenases, Fdh; hydrogenases (Mbh, Ech, Ehr, Eha Frh); formyl-methanofuran dehydrogenase, Fmd, Fwd; methenyl–H_4MPT tetrahydromethanopterin cyclohydrolyase, Mch; formylMFR:tetrahydromethanopterin formyl transferase, Ftr; methylene–H_4MPT dehydrogenase, Mtd; methylene–H_4MPT reductase, Mer; H_4MPT S-methyltransferase (Mtr; methyl-CoM reductasem Mcr; and CoM-S-S-CoB heterodisulfide reductase, H)

(Mhun_0075, Mhun_1811). The five hydrogenase gene clusters include *echABDDEF* (Mhun_1741-1747), *ehrABCDLS* (Mhun_1817-1822), *ehaABCDEFGHIJK* (Mhun_2094-2106), *frhADGB* (Mhun_2329-2332), and *mbhABCDEFGHIJKLMN* (Mhun_2579-2592). The *ech*, *eha*, *ehr*, and *mbh* gene clusters encode membrane-associated enzymes that likely reduce Fd. These are believed to employ ion gradients (Na^+ or H^+) to assist Fd reduction at low hydrogen levels. The remaining hydrogenase gene cluster (*frh*ADGB) encodes a soluble hydrogenase that reduces F_{420}.

6.2 Transporters, ion movement, and ATP synthesis

M. hungatei JF1 has 352 genes involved in membrane transport as determined by IMG/ER, which constitute 10.64% of the genome. These include 34 multi-component ATP-binding cassette or ABC-type transporter genes plus related but unlinked genes (152 genes in total), sixty genes encoding secondary transporters, twelve genes for ion channels, seven genes for P-ATPases, one H^+ translocating pyrophosphatase (Mvp, H+PPase; Mhun_2414) gene, and four type II secretion systems. A highly unusual feature of the *M. hungatei* genome is the presence of three H^+ or Na^+ -translocating AoA_1-type ATP synthetase gene clusters encoded by 27 genes (Aha1, Mhun_1177-1185; Aha2, Mhun_1757-1765, and Aha3, Mhun_1768-1775). The gene order is conserved relative to the corresponding Aha complex in *Methanosarcina acetivorans* (45). Although it is unknown whether these systems utilize protons or sodium ions, the *M. acetivorans* ortholog is believed to use sodium ions (45). Likewise, the membrane-bound H_4MPT Smethyltransferase (Mtr) is predicted to be sodium dependent. Three genes encode $Na^{+/}H^+$ antiporters (Mhun_0680, Mhun_0841, Mhun_2803) that might maintain ion balance where the last differs by also possessing a Trk domain.

6.3 Cell Biosynthesis

The genome of *M. hungatei* encodes an acetyl-CoA synthase/CO dehydrogenase complex (Cdh; Mhun_0686-0690). The role of Cdh is undefined at this time because *M. hungatei* must acquire acetate supplied in the medium for growth rather than synthesizing acetylCoA from CO_2, which is the usual role of Cdh in hydrogenotrophic methanogens. Uptake of acetate for incorporation into cell material is predicted to occur by the Mhun_0634 *aceP* gene product (45). Five acetyl-CoA synthetase genes are present that could activate acetate to acetyl-CoA. Mhun_0352, Mhun_0567, and Mhun_1721 share > 62% identity at the amino acid level with each other, but only share < 34.2% amino acid identity with Mhun_0592 and Mhun_2392.

M. hungatei has two set of genes that could be used to carboxylate acetyl-CoA to pyruvate (Mhun_2393-2396 and Mhun_0450-0453). Oxaloacetate can be synthesized by carboxylation of pyruvate using pyruvate carboxylase (Mhun_3189-3190) or by conversion of pyruvate to phosphoenol pyruvate by pyruvate dikinase (Mhun_2610 or Mhun_1141) and carboxylation of phosphoenol pyruvate to oxaloacetate by phosphoenol pyruvate carboxylase (Mhun_0174). The genes necessary to convert oxaloacetate to 2-oxoglutarate by the reductive arm of the tricarboxylic acid cycle were detected (malate dehydrogenase, Mhun_1155; fumarate hydratase, Mhun_0089-0090; succinyl-CoA ligase, Mhun_0096-0095; 2-oxoglutarate synthase, Mhun_0091-0094 or Mhun_29922994; and fumarate reductase, Mhun_3052-3053). Complete biosynthetic pathways for the synthesis of all amino acids except histidine from pyruvate, oxaloacetate, and 2-oxoglutarate as the main starting materials were detected.

There are few genomic clues regarding the composition of the *M. hungatei* cell envelope. The genome contains a large number of PDK domain-containing genes (31 genes) as well as TRP domain-containing genes (41 genes). Many of these have transmembrane and/or SP signal elements that would suggest cell envelope associations but it is unknown if any of the proteins are significantly expressed. There are no clear protein candidates for the morphologically defined cell envelope structures containing a surface layer, sheath, and plugs (1, 8).

6.4 Stress

There appear to be few cellular adaptations in *M. hungatei* for stress response. Among those found are defense against oxygen damage: catalase (Mhun_2433), peroxidase (Mhun_2733), manganese/iron superoxide dismutase (Mhun_2974), heavy metal resistance (Mhun_1348, Mhun_3034), drug resistance (Mun_0598, Mhun_1195) and heat shock (Mhun_2436).

6.5 Regulation and signal transduction

The *M. hungatei* genome contains a typical set of archaeal RNA polymerase genes and one BRE recognition factor (TFB) analogous to eukaryotic transcription initiating factor B (TFIIB) (Mhun_2481; Tfb) plus two TATA-box binding proteins or TBP's that confer promoter recruitment and specificity (Tbp1, Mhun_0568 and Tbp2, Mhun_0593). There are ~65 DNA-binding transcription factors identified that modulate gene expression. These belong to a variety of protein families common to bacteria but include few regulatory proteins typical of eukaryotes (e.g., homeodomain-like, zinc finger, SRF-like, or p53-like proteins). There are numerous bacterial-type two-component regulatory systems including 82 histidine kinase-type sensor transmitters, 41 response regulatory proteins, and 18 receiver-only domain proteins. Of the 82 histidine kinases, 55 are soluble and 27 are membrane-associated. They are generally unlinked genetically and thus do not suggest an interacting partner in sensory transduction.

6.6 Motility and taxis

M. hungatei JF1 possesses multiple archaeal-type flagella filaments at the cell ends (1, 8), now termed archaealla that resemble bacterial type IV pili (46, 47). The genome contains one *flhGFHIJ* gene cluster (Mhun_0102-0105) encoding a basal body structure. Three FlaB-type pili genes make up the archaella filaments (Mhun_1238, Mhun_3139, Mhun_3140). Although little is known about the chemotactic abilities of *M. hungatei*, other than movement towards an essential nutrient, acetate (48), there are multiple chemosensory genes present in the genome. These include 3 CheA, 4 CheB, 4 CheC, 1 CheD, 3 CheR, 1 CheY, and 14 CheW, genes plus 27 genes encoding MCP sensory proteins (methyl accepting chemotaxis proteins) that detect unknown attractants and/or repellants. Twelve MCPs are membrane-associated and 15 MCPs are soluble.

Multiple genes (~11 paralogs) are also present in the *M. hungatei* JF1 genome for archaeal-type pili like those seen in *Methanococcus maripaludus*, *Haloferax volcanii*, and *Sulfollobus acidocaldarius* (49). These archaeal proteins, distinct from the bacterial pili-type proteins, were previously annotated as hypothetical genes (e.g., Mhun_0297). The *H. volcanii* pili proteins provide adhesion to surfaces and the orthologs in *M. hungatei* JF1 may function in cell-cell adhesion or in cell-cell communication, although such appendages have not been previously observed in EM micrographs. All but one of the eleven *M. hungatei* JF1 paralogs are in clusters of 2 to 3 genes each and often with ABCtype transport genes.

6.7 Comparison to other archaeal genomes

The 3.54 MB *M. hungatei* JF1 genome is the largest within the order *Methanomicrobiales* that have been sequenced thus far including *Methanosphaerula palustris* (2.92 MB) and *Methanocorpusculum labreanum* (1.80 MB). The *M. hungatei* JF1 genome is also among the largest within the *Archaea* domain: only three species sequenced thus far, belonging to the genus *Methanosarcina* (i.e., *Methanosarcina acetivorans*, 5.75 MB; *Methanosarcina barkeri*, 4.87 MB; and *Methanosarcina mazei*, 3.83 MB), plus one halophile, *Haloarcula marismortaui* (4.27 MB), exceed it in size. The large genome of *M. hungatei* JF1 suggests the presence of unrecognized biochemical/physiological properties that likely extend to the other *Methanospirillaceae* and include the ability to form the unusual sheath-like structure and to successfully interact with syntrophic bacteria.

When *M. hungatei* ORFs were compared pair-wise to individual microbial genomes (50, 51), best reciprocal BLAST hits revealed closest associations to the taxonomically related archaea: *Methanoculleus marisnigri* (1395 reciprocal gene hits), *Methanosarcina acetivorans* (1203), and *Methanosarcuna barkeri* (1150), and extending to *Haloquadratum walsbyi* (657) (Additional File 1). Thus, approximately 650 to 1,200 genes are similar and well-conserved across these 17 archaeal species whereby the remaining genes (ca. 1700 genes) represent a novel complement within the *M. hungatei* genome. Interestingly, seven of the next thirteen closest matches are bacterial species among which are many syntrophic microorganisms that likely grow in close association with *M. hungatei*. Strikingly, *Syntrophobacter fumaroxidans* strain MPOB exhibited 634 best reciprocal BLAST hits.

In another comparison, the best BLAST hit to any microbial gene product was determined (Additional File 2) and showed 1; 167; 277; and 142 ORFs closest hits in the genomes of *Methanoculleus marisnigri, Methanocorpusculum labreanum, and Methanosarcina barkeri*, respectively. Notably three bacterial genomes, '*Syntrophus* aciditrophicus', *Syntrophobacter fumaroxidans,* and *Nostoc spp.* gave 21-19 best BLAST hits each, suggesting the possibility of lateral gene transfer events from these potential syntrophic partners. The occurrence of *Nostoc*-related genome sequences raises interesting questions concerning microbial interactions and lateral gene transfer with methanogens present in complex microbial communities (52).

7. Extended insights

The large genome of *M. hungatei* JF1 suggests the presence of unrecognized biochemical/physiological properties that likely extend to the other *Methanospirillaceae* and include the ability to form the unusual sheath-like structure and the ability to successfully interact with syntrophic bacteria. A number of genes may have been acquired by lateral gene transfer from its syntrophic partners or other microorganisms present in complex microbial communities. Also of particular note are multiple genes for archaeal type IV pili that may function in cell-cell adhesion or cell-cell communication and genes for multiple hydrogenases and formate dehydrogenases to metabolize hydrogen and formate generated by its syntrophic partners. The core machinery of *M. hungatei* to produce methane from hydrogen and carbon dioxide and/or formate is typical of other hydrogenotrophic methanogens, except that *M. hungatei* has genes for three H^+ or Na^+-translocationg A_OA_1-type ATP synthases. *M. hungatei* has four 16S ribosomal RNA genes that each differ at two positions. Further understanding of the novel compliment of *M. hungatei* genes will likely provide a more thorough understanding of the multispecies interactions involved in syntrophy and the synthesis of complex structures such as the *M. hungatei* sheath, which is shared by multiple cells.

8. Conclusions

We report here an inventory of the genomic features of the methane-producing anaerobic archaeon, *Methanospirillum hungatei* strain JF1 (DSM 864), and describe its phylogenetic relationship to its neighbors. We further identify from *M. hungatei*'s sizable genome examples of genes involved in anaerobic syntrophy, and as the type strain of the *Methanospirillum*, suggest potential universal qualities of this genus. We hope this report aids and stimulates further study of this fascinating organism.

Competing Interests

The authors declare they have no competing interests.

Authors' Contributions

RPG and MJM contributed to the conception and design of this project. ALL, HED, ML, OG, NI, and NK were involved in the acquisition and initial analysis of the data; LEC, BC, LR, EM, HM, JRS, NP, HZ, RPG, and MJM were involved in the interpretation of the data. RPG prepared the first draft of the manuscript. All authors were involved in its critical revision and have given final approval of the version to be published and agree to be accountable for all aspects of the work.

Acknowledgements

This work was performed under the auspices of the US Department of Energy Office of Science Biological and Environmental Research Program. We also acknowledge support by the US Department of Energy Contract No. DE-FG02-96ER20214, from the Chemical

Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences (MJM) for syntrophy functions, Department of Energy Grant DE-FG03-86ER13498 (RPG) the UCLA-DOE Institute of Genomics and Proteomics Grant DE-FC02-02ER63421 (RPG), and the National Science Foundation grant NSF 1244566 (MJM and RPG) for methanogen biochemistry.

Supplemental Figures

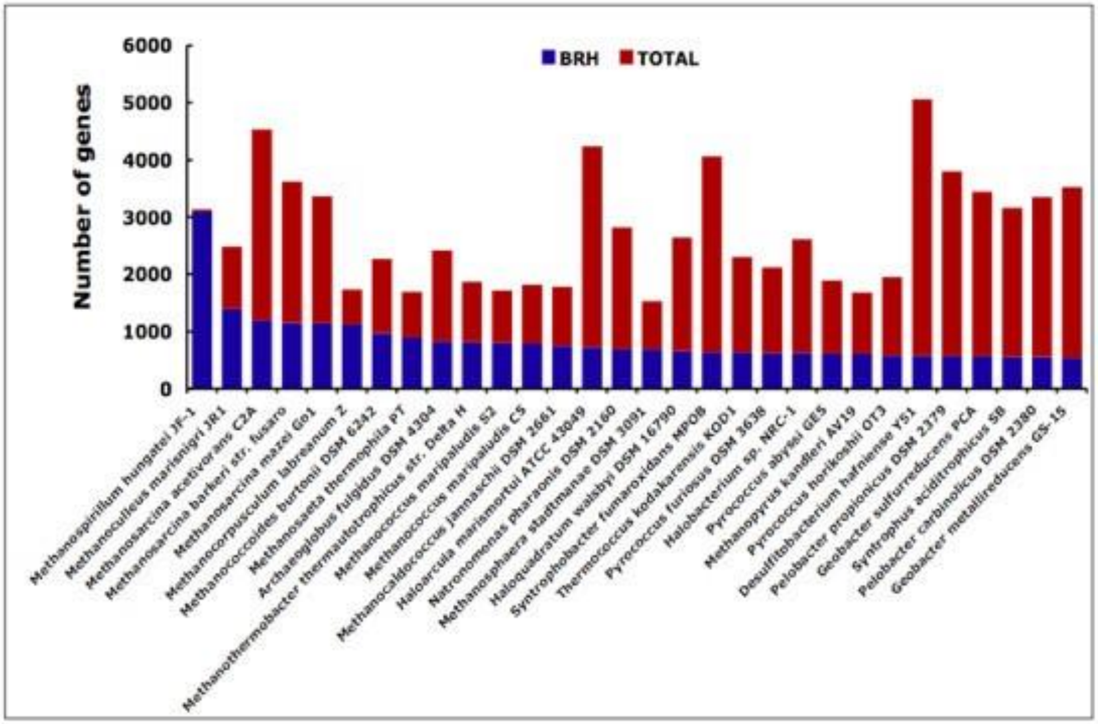

Figure 1S Best reciprocal protein hits for *M. hungatei* JF1 ORFs with other genomes.

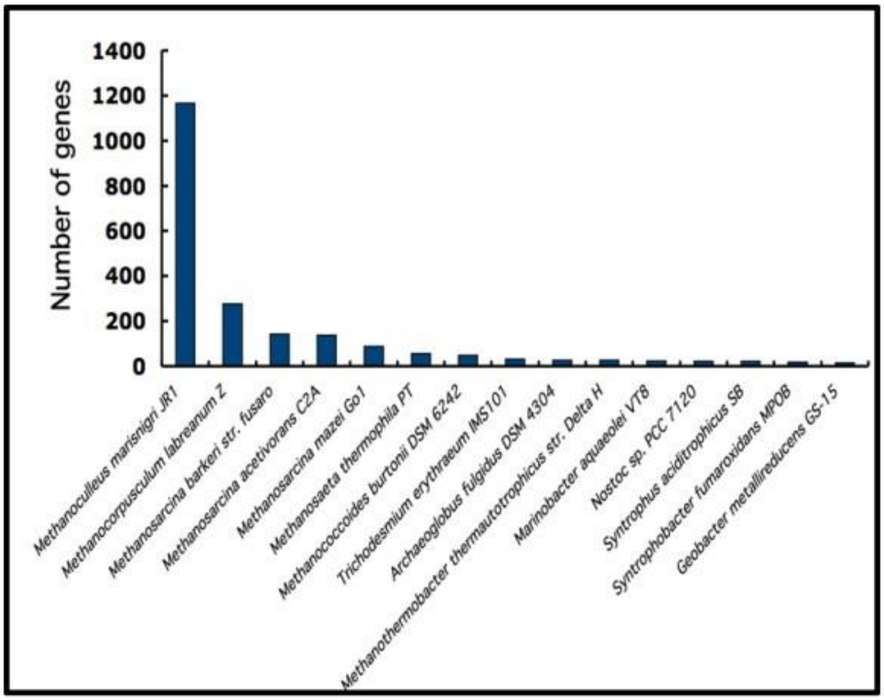

Figure S2 Best BLAST hit distribution of *M. hungatei* JF1 ORFs with other genomes.

References

1. **Ferry JG, Smith PH.** 1974. *Methanospirillum*, a new genus of methanogenic bacteria, and characterization of *Methanospirillum hungatii* sp. nov. International Journal of Systematic Bacteriology **24:**465-469.

2. **Boone DR, Whitman W.B., and Koga, Y.** 2001. "Family III. Methanospirillaceae fam. nov." In: Bergey's Manual of Systematic Bacteriology, 2nd ed, vol Volume 1: The Archaea and the deeply branching and phototrophic Bacteria, p 264, Springer-Verlag, New York.

3. **Hungate RE.** 1950. The anaerobic mesophilic cellulolytic bacteria. Bacteriol Rev **14:**1-49.

4. **Wolfe RS.** 2011. Techniques for cultivating methanogens. Methods Enzymol **494:**1-22.

5. **Ferry JG, Wolfe RS.** 1976. Anaerobic degradation of benzoate to methane by a microbial consortium. Arch Microbiol **107:**33-40.

6. **McInerney MJ, Struchtemeyer CG, Sieber J, Mouttaki H, Stams AJ, Schink B, Rohlin L, Gunsalus RP.** 2008. Physiology, ecology, phylogeny, and genomics of microorganisms capable of syntrophic metabolism. Ann N Y Acad Sci **1125:**58-72.

7. **Ferry JG, Wolfe RS.** 1977. Nutritional and biochemical characterization of Methanospirillum hungatii. Appl Environ Microbiol **34:**371-376.

8. **Toso DB, Henstra AM, Gunsalus RP, Zhou ZH.** 2011. Structural, mass and elemental analyses of storage granules in methanogenic archaeal cells. Environ Microbiol **13:**2587-2599.

9. **Woese CR, Kandler O, Wheelis ML.** 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA **87:**4576-4579.

10. **Garrity GM, Holt JG, Whitman WB, Keswani J, Boone DR, Koga Y, Miller TL, Stetter KO, Zellner G, Chong SC, Huber H, Huber G, Ferry JG, Ollivier**

B, Mah RA, Sowers KR, Zhilina TN, Baker CC, Romesser JA, Grant WD, Patel GB, McGenity TJ, Kamekura M, Ventosa A, Kobayashi T, Oren A, Montalvo-Rodríguez R, Vreeland RH, Tindall BJ, Huber R, Xu Y, Zhou P, Reysenbach A-L, Langworthy TA, Tian X, Zillig W, Kobayashi T, Hafenbradl D. 2001. Phylum AII. Euryarchaeota phy. nov, p 211-355. *In* Boone DR, Castenholz RW, Garrity GM (ed), Bergey's Manual® of Systematic Bacteriology: Volume One : The Archaea and the Deeply Branching and Phototrophic Bacteria doi:10.1007/978-0-387-21609-6_17. Springer New York, New York, NY.

11. **Garrity GM, Bell JA, Lilburn T.** 2015. The Revised Road Map to the Manual. Bergey's Manual of Systematics of Archaea and Bacteria doi:doi:10.1002/9781118960608.bm00023
12. **Balch WE, Fox GE, Magrum LJ, Woese CR, Wolfe RS.** 1979. Methanogens: reevaluation of a unique biological group. Microbiol Rev **43:**260-296.
13. **Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, Ashburner M, Axelrod N, Baldauf S, Ballard S, Boore J, Cochrane G, Cole J, Dawyndt P, De Vos P, DePamphilis C, Edwards R, Faruque N, Feldman R, Gilbert J, Gilna P, Glockner FO, Goldstein P, Guralnick R, Haft D, Hancock D, Hermjakob H, Hertz-Fowler C, Hugenholtz P, Joint I, Kagan L, Kane M, Kennedy J, Kowalchuk G, Kottmann R, Kolker E, Kravitz S, Kyrpides N, Leebens-Mack J, Lewis SE, Li K, Lister AL, Lord P, Maltsev N, Markowitz V, Martiny J, et al.** 2008. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol **26:**541-547.
14. **Field D, Amaral-Zettler L, Cochrane G, Cole JR, Dawyndt P, Garrity GM, Gilbert J, Glöckner FO, Hirschman L, Karsch-Mizrachi I, Klenk H-P, Knight R, Kottmann R, Kyrpides N, Meyer F, San Gil I, Sansone S-A, Schriml LM, Sterk P, Tatusova T, Ussery DW, White O, Wooley J.** 2011. The Genomic Standards Consortium. PLOS Biology **9:**e1001088.
15. **Garrity G.** Names for Life Browser Tool takes expertise out of the database and puts it right in the browser., p 9.
16. **Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G.** 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics **25:**25-29.

17. **Schink B.** 1997. Energetics of syntrophic cooperation in methanogenic degradation. Microbiol Mol Biol Rev **61:**262-280.

18. **McInerney MJ, Sieber JR, Gunsalus RP.** 2009. Syntrophy in anaerobic global carbon cycles. Current Opinion in Biotechnology **20:**623-632.

19. **Sieber JR, McInerney MJ, Gunsalus RP.** 2012. Genomic Insights into Syntrophy: The Paradigm for Anaerobic Metabolic Cooperation. Annual Review of Microbiology **66:**429-452.

20. **Qiu YL, Sekiguchi Y, Imachi H, Kamagata Y, Tseng IC, Cheng SS, Ohashi A, Harada H.** 2004. Identification and isolation of anaerobic, syntrophic phthalate isomer-degrading microbes from methanogenic sludges treating wastewater from terephthalate manufacturing. Appl Environ Microbiol **70:**1617-1626.

21. **Boone DR, Bryant MP.** 1980. Propionate-Degrading Bacterium, Syntrophobacter wolinii sp. nov. gen. nov., from Methanogenic Ecosystems. Appl Environ Microbiol **40:**626-632.

22. **McInerney MJ, Bryant MP.** 1981. *Syntrophomonas wolfei* gen. nov. sp. nov., an anaerobic, syntrophic, fatty acid-oxidizing bacterium. Appl Environ Microbiol **41:**1029-1039.

23. **Jackson BE, V.K. Bhupathiraju, R.S. Tanner, C.R. Woese, and M.J. McInerney.** 1999. *Syntrophus aciditrophicus* sp. nov., a new anaerobic bacterium that degrades fatty acids and benzoate in syntrophic association with hydrogen-utilizing microorganisms. Arch Microbiol **171:**107-114.

24. **Mountfort D. BW, Krumholz L, Bryant M.,.** 1984. *Syntrophus buswellii* gen. nov., sp. nov.: a benzoate catabolizer from methanogenic ecosystems. International Journal of Systematic Bacteriology **34:**2.

25. **Tamura K, Nei M.** 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol **10:**512-526.

26. **Tamura K, Stecher G, Peterson D, Filipski A, Kumar S.** 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **30:**2725-2729.

27. **Iino T, Mori K, Suzuki K.** 2010. Methanospirillum lacunae sp. nov., a methane-producing archaeon isolated from a puddly soil, and emended descriptions of the genus Methanospirillum and Methanospirillum hungatei. Int J Syst Evol Microbiol **60:**2563-2566.

28. **Wright AD, Pimm C.** 2003. Improved strategy for presumptive identification of methanogens using 16S riboprinting. J Microbiol Methods **55:**337-349.

29. **Zeikus JG, Bowen VG.** 1975. Fine Structure of Methanospirillum hungatii, vol 121, p 373-380.

30. **Kushwaha SC, Kates M, Sprott GD, Smith IC.** 1981. Novel polar lipids from the methanogen Methanospirillum hungatei GP1. Biochim Biophys Acta **664:**156-173.

31. **Pagani I LK, Jansson J, Chen IM, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC. 2012**. **The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata**. **Nucleic Acids Res, :D571-D579**.

32. **Anonymous.** DOE Joint Genome Institute. http://www.jgi.doe.gov. Accessed

33. **Anonymous.** Phred/Phrap/Consed software package. http://www.phrap.com. Accessed

34. **Gordon D, Abajian C, Green P.** 1998. Consed: a graphical tool for sequence finishing. Genome Res **8:**195-202.

35. **Ewing B, Green P.** 1998. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. Genome Research **8:**186-194.

36. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119.

37. **Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC.** 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat Methods **7:**455-457.

38. **Anonymous.** GenePRIMP. http://geneprimp.jgi-psf.org. Accessed

39. **Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics **25:**2271-2278.

40. **Anonymous.** IMG-ER. http://img.jgi.doe.gov/er. Accessed

41. **Anonymous.** TransportDB. http://www.membranetransport.org/. Accessed

42. **Anonymous.** TCDB. http://www.tcdb.org/. Accessed

43. **Anonymous.** TBD. http://www.transcriptionfactor.org. Accessed

44. **Welte C, Deppenmeier U.** 2014. Bioenergetics and anaerobic respiratory chains of aceticlastic methanogens. BBA **1837:**1130-1147.

45. **Rohlin L, Gunsalus RP.** 2010. Carbon-dependent control of electron transfer and central carbon pathway genes for methane biosynthesis in the Archaean, Methanosarcina acetivorans strain C2A. BMC Microbiol **10:**62.

46. **Jarrell KF, Albers SV.** 2012. The archaellum: an old motility structure with a new name. Trends Microbiol **20:**307-312.

47. **Thomas NA, Bardy SL, Jarrell KF.** 2001. The archaeal flagellum: a different kind of prokaryotic motility structure. FEMS Microbiol Rev **25:**147-174.

48. **Migas J, Anderson KL, Cruden DL, Markovetz AJ.** 1989. Chemotaxis in Methanospirillum hungatei. Appl Environ Microbiol **55:**264-265.

49. **Esquivel RN, Xu R, Pohlschroder M.** 2013. Novel archaeal adhesion pilins with a conserved N terminus. J Bacteriol **195:**3808-3818.

50. **Plugge CM, Henstra AM, Worm P, Swarts DC, Paulitsch-Fuchs AH, Scholten JC, Lykidis A, Lapidus AL, Goltsman E, Kim E, McDonald E, Rohlin L, Crable BR, Gunsalus RP, Stams AJ, McInerney MJ.** 2012.

Complete genome sequence of Syntrophobacter fumaroxidans strain (MPOB(T)). Stand Genomic Sci **7:**91-106.

51. **Sieber JR, Sims DR, Han C, Kim E, Lykidis A, Lapidus AL, McDonnald E, Rohlin L, Culley DE, Gunsalus R, McInerney MJ.** 2010. The genome of Syntrophomonas wolfei: new insights into syntrophic metabolism and biohydrogen production. Environ Microbiol **12:**2289-2301.

52. **Deppenmeier U, Johann A, Hartsch T, Merkl R, Schmitz RA, Martinez-Arias R, Henne A, Wiezer A, Baumer S, Jacobi C, Bruggemann H, Lienard T, Christmann A, Bomeke M, Steckel S, Bhattacharyya A, Lykidis A, Overbeek R, Klenk HP, Gunsalus RP, Fritz HJ, Gottschalk G.** 2002. The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. J Mol Microbiol Biotechnol **4:**453-461.

Chapter 3. Genome Sequence of *Acetomicrobium hydrogeniformans* OS1

This is a reprint of:

Cook, L. E., Gang, S. S., Ihlan, A., Maune, M., Tanner, R. S., McInerney, M. J., Weinstock, G., Lobos, E. A., and Gunsalus, R. P. (2018). Genome Sequence of Acetomicrobium hydrogeniformans OS1. *Genome announcements*, *6*(26), e00581-18.

Contents

1 Genome Note

Acetomicrobium hydrogeniformans strain OS1^T, until recently known as *Anaerobaculum hydrogeniformans* (1), of the phylum *Synergistetes* is capable of producing almost four molecules of hydrogen per glucose molecule, the theoretical maximum (2, 3). It ferments other substrates including amino acids, dicarboxylic acids and other sugars, and can respire using several sulfur compounds. NaCl is an absolute requirement for growth, which suggests a sodium-based energy strategy (2, 4, 5). Although it can reduce elemental sulfur, thiosulfate and L-cysteine to sulfide, the pathways are unknown. The genome is currently divided among eight scaffolds, one of which contains 2,103,414 bp and seven of which are small scaffolds ranging from 7,345 bp to 507 bp. The genome lacks genes needed for respiration of sulfate and nitrate, and genes for *b*- and *c*-type cytochromes. The metabolic strategy appears to be limited primarily to fermentation of simple sugars, amino acids and certain dicarboxylic acids in addition to an as-yet undescribed respiratory mode for elemental sulfur or thiosulfate. Despite the ability to ferment sugars, the Emden-Meyerhof-Parnas, pentose phosphate, and Entner-Doudoroff pathways appear to be incomplete. It also lacks a complete set of TCA genes. OS1^T possesses two gene clusters with hydrogenase-related functions, to form molecular hydrogen. It also possesses genes for two putative formate dehydrogenases and a formate hydrogenlyase complex. It has multiple genes for stress adaptation, including those for motility, pili, osmotic adaptation, heavy-metal resistance, multi-drug export and toxin/antitoxin synthesis. It has a twenty-one-gene cassette for the synthesis of a carboxysome-like shell structure. Although motility was not previously reported (2), motility can be observed in wet mounts of cultures and negative-stained cells exhibit polar flagella (R. Tanner, personal communication). The genome contains a complete inventory of genes needed to make flagella along with multiple genes for chemoreception and chemotaxis. The genome also contains a complement of ten *pil/pul* genes needed for synthesis and assembly of a Type IV pilus apparatus. It has a relatively small number of primary transcription factors (43 genes), and six predicted two-component regulatory systems, including the aforementioned chemotaxis signaling systems. Thus, OS1 appears to have adopted a minimal regulatory strategy, although the substrate range is considerable with respect to the variety of sugars, amino acids and dicarboxylates used (2). Little is known about translational or post-translational control operating in this or other *Synergistetes* species. Strikingly, the genome lacks genes for the prototypical bacterial FoF$_1$-type ATP synthase, but possesses instead three clusters of genes encoding either archaeal-like or vacuolar-like A/V ATP synthases (img.doe.gov). It also has other archaeal-like annotated genes, including those acting in glycolysis; of note, we found four putative candidates for glyceraldehyde-3-phosphate oxidoreductase, a ferredoxin-utilizing enzyme with archaeal properties. Other genes with archaeal properties include those related to formyl methanofuran dehydrogenase A, B, C, D and

E subunits, methanyltetrahydromethanopterin cyclohydrolase, and tetrahydro-methanopterin N-formyl transferase. These genes may have been acquired by lateral gene transfer (6-8).

Accession number

This whole-genome shotgun project has been deposited in GenBank under the accession no. ACJX00000000.

Acknowledgements

These sequence data were produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov) in collaboration with the user community.

This work was supported by Department of Energy contract DE-FG02-08ER64689 (to R.P.G., R.S.T., and M.J.M.), UCLA-DOE Institute award DE-FC02-02ER63421 (to R.P.G.), and National Science Foundation grant NSF 1515843 (to R.P.G. and M.J.M.).

References

1. **Ben Hania W, Bouanane-Darenfed A, Cayol J-L, Ollivier B, Fardeau M-L.** 2016. Reclassification of Anaerobaculum mobile, Anaerobaculum thermoterrenum, Anaerobaculum hydrogeniformans as Acetomicrobium mobile comb. nov., Acetomicrobium thermoterrenum comb. nov. and Acetomicrobium hydrogeniformans comb. nov., respectively, and emendation of the genus Acetomicrobium. International Journal of Systematic and Evolutionary Microbiology **66:**1506-1509.

2. **Maune MW, Tanner RS.** 2012. Description of Anaerobaculum hydrogeniformans sp. nov., an anaerobe that produces hydrogen from glucose, and emended description of the genus Anaerobaculum. Int J Syst Evol Microbiol **62:**832-838.

3. **Thauer RK, Jungermann K, Decker K.** 1977. Energy conservation in chemotrophic anaerobic bacteria. Bacteriol Rev **41:**100-180.

4. **Menes RJ, Muxi L.** 2002. *Anaerobaculum mobile* sp. nov., a novel anaerobic, moderately thermophilic, peptide-fermenting bacterium that uses crotonate as an electron acceptor, and emended description of the genus *Anaerobaculum*. Int J Syst Evol Microbiol **52:**157-164.

5. **Buss KA, Ingram-Smith C, Ferry JG, Sanders DA, Hasson MS.** 1997. Crystallization of acetate kinase from Methanosarcina thermophila and prediction of its fold [In Process Citation]. Protein Sci **6:**2659-2662.

6. **Garcia-Vallvé S, Romeu A, Palau J. 2000**. **Horizontal Gene Transfer in Bacterial and Archaeal Complete Genomes. Genome Research 10:1719-1725**.

7. **Nelson KE.** 1999. **Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. Nature 27:323-329**

8. **Deppenmeier U, Johann A, Hartsch T, Merkl R, Schmitz RA, Martinez-Arias R, Henne A, Wiezer A, Baumer S, Jacobi C, Bruggemann H, Lienard T, Christmann A, Bomeke M, Steckel S, Bhattacharyya A, Lykidis A, Overbeek R, Klenk HP, Gunsalus RP, Fritz HJ, Gottschalk G.** 2002. The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. J Mol Microbiol Biotechnol **4:**453-461.

Chapter 4. Detailed Analysis of the *Acetomicrobium hydrogeniformans* OS1 Genome

Contents

Figure 1. Electron micrograph of *A. hydrogeniformans* strain OS1. Negatively-stained cell image was obtained as described (1).

Figure 2. Neighbor-joining phylogenetic tree based on 16S rRNA gene sequence showing the relationships between *Acetomicrobium flavidum* and related representatives of the family *Synergistaceae*. Bootstrap values of 70 % or higher (based on 2000 repetitions) are shown at branch nodes. Bar, 5 substitutions per 100 nt. A phylogenetic tree was reconstructed with the TREECON program, using the neighbor-joining method (2). Tree topology was evaluated by a bootstrap analysis using 2000 resampling of the sequences (3). Its topology was also supported using the maximum-parsimony and maximum-likelihood methods. The results indicate that *Acetomicrobium flavidum* pertains to the family *Synergistaceae,* phylum *Synergistetes*, having *Anaerobaculum mobile* (99.9 % similarity) (4), *Anaerobaculum hydrogeniformans* (96.8 % similarity) (5) and *Anaerobaculum thermoterrenum* (96.4 % similarity) (6), as its closest phylogenetic relatives.

Originally published in Ben Hania et al (1) and used with permission, copyright clearing house license 4590300479047. The preceding figure legend is taken directly from the original paper. *Acetomicrobium hydrogeniformans* is boxed in red, and the genus *Acetomicrobium* is boxed in purple. The genus *Thermovirga* diverged from a common ancestor with *Acetomicrobium*.

Figure 3. A metabolic reconstruction of the *A. hydrogeniformans* OS1 genome content.

Table 1. Classification and general features of *A. hydrogeniformans* strain OS1 by MIGS recommendations.

Table 2. Genome sequence project information of *A. hydrogeniformans* OS1 according to the MIGS recommendations.

Table 3. Genome Statistics.

Table 4. Number of genes associated with the general COG functional categories.

Supplementary material

Table S1: Hydrogenase and dehydrogenase-containing gene clusters in OS1. Includes a membrane-bound [NiFe] hydrogenase associated with a formate hydrogenlyase complex, and a soluble [FeFe] hydrogenase.

Table S2. List of genes potentially involved in the pentose phosphate pathway in *A. hydrogeniformans* OS1. The pathway is incomplete.

Table S3: List of genes potentially involved in the Wood-Ljungdahl pathway in A. hydrogeniformans OS1. OS1 has a potential complete set of genes for pathway I, the version of the pathway found in acetogenic bacteria with *Moorella* species as the model. OS1 is lacking any paralogues of genes for two enzymes in pathway II (as found in methanogenic archaea and modeled on *Methanosarcina* species). Enzymes and genes in table are named for the respective names given in the model organism as listed above. Genes that are a 3/4 match for EC number are marked with an asterisk. Unmarked genes are a 4/4 match for the EC number.

Table S4. List of genes involved in solute transport in *A. hydrogeniformans* OS1.
Table S5. List of genes involved in energy storage in *A. hydrogeniformans* OS1.
Table S6. List of genes involved in shell structure in *A. hydrogeniformans* OS1.
Table S7. List of genes involved in motility and chemotaxis in *A. hydrogeniformans* OS1.
Table S8. List of genes involved in pili formation in *A. hydrogeniformans* OS1.
Table S9. List of RNAP and sigma factor genes in *A. hydrogeniformans* OS1.
Table S10. List of genes involved in two component signal transduction in *A. hydrogeniformans* OS1.
Table S11. Genes with archaeal gene/protein annotations.

1. Abstract

Acetomicrobium hydrogeniformans strain $OS1^T$, until recently known as *Anaerobaculum hydrogeniformans* (7), of the phylum *Synergistetes* is capable of producing almost four molecules of hydrogen per glucose molecule, the theoretical maximum (5, 8). It ferments other substrates including amino acids, dicarboxylic acids and other sugars, and can respire using several sulfur compounds. NaCl is an absolute requirement for growth, which suggests a sodium-based energy strategy (4, 5, 9). Although it can reduce elemental sulfur, thiosulfate and L-cysteine to sulfide, the pathways are unknown. The genome is currently divided among eight scaffolds, one of which contains 2,103,414 bp and seven of which are small scaffolds ranging from 7,345 bp to 507 bp. The genome lacks genes needed for respiration of sulfate and nitrate, and genes for *b*- and *c*-type cytochromes. The metabolic strategy appears to be limited primarily to fermentation of simple sugars, amino acids and certain dicarboxylic acids in addition to an as-yet undescribed respiratory mode for elemental sulfur or thiosulfate. Despite the ability to ferment sugars, the Emden-Meyerhof-Parnas, pentose phosphate, and Entner-Doudoroff pathways appear to be incomplete. It also lacks a complete set of TCA and oxidative pentose-phosphate pathway genes. *A. hydrogeniformans* $OS1^T$ possesses two gene clusters with hydrogenase-related functions, to form molecular hydrogen. It also possesses genes for two putative formate dehydrogenases and a formate hydrogenlyase complex. It has multiple genes for stress adaptation, including those for motility, pili, osmotic adaptation, heavy-metal resistance, multi-drug export and toxin/antitoxin synthesis. It has a twenty-one-gene cassette for the synthesis of a carboxysome-like shell structure. Although motility was not previously reported (5), motility can be observed in wet mounts of cultures and negative-stained cells exhibit polar flagella (R. Tanner, personal communication). The genome contains a complete inventory of

genes needed to make flagella along with multiple genes for chemoreception and chemotaxis. The genome also contains a complement of ten *pil/pul* genes needed for synthesis and assembly of a Type IV pilus apparatus. It has a relatively small number of primary transcription factors (43 genes), and six predicted two-component regulatory systems, including the aforementioned chemotaxis signaling systems. Thus, *A. hydrogeniformans* appears to have adopted a minimal regulatory strategy, although the substrate range is considerable with respect to the variety of sugars, amino acids and dicarboxylates used (5). Little is known about translational or post-translational control operating in this or other *Synergistetes* species. Strikingly, the genome lacks genes for the prototypical bacterial FoF_1-type ATP synthase, but possesses instead three clusters of genes encoding either archaeal-like or vacuolar-like A/V ATP synthases (10). It also has other archaeal-like annotated genes, including those acting in glycolysis; of note, we found four putative candidates for glyceraldehyde-3-phosphate oxidoreductase, a ferredoxin-utilizing enzyme with archaeal properties. Other genes with archaeal properties include those related to formyl methanofuran dehydrogenase A, B, C, D and E subunits, methanyltetrahydromethanopterin cyclohydrolase, tetrahydro-methanopterin N-formyl transferase, and a gene for the synthesis of wyosine. These genes may have been acquired by lateral gene transfer (11-13).

2. Introduction

Acetomicrobium hydrogeniformans strain OS1^T (DSM 22491 = ATCC BAA-1850), until recently known as *Anaerobaculum hydrogeniformans* (7), is the type strain for *A. hydrogeniformans* (5). It is of particular interest given its association with pipeline bio-corrosion (14) and its potential role in industrial production of alternative biofuels (5). It also serves as a model for pure culture and co-culture metabolism studies for members of the phylum *Synergistetes* in anaerobic waste digesters (5, 15) as well as the human microbiome (16, 17). It is capable of producing almost four molecules of hydrogen per glucose molecule at metabolically favorable rates, making it a candidate for economic generation of hydrogen gas (5). Hydrogen yields approach the theoretical maximum of 4 moles of H_2 per mole glucose degraded predicted for dark fermentation of glucose (8). *A. hydrogeniformans* ferments other substrates including other sugars, amino acids, and dicarboxylic acids, and respires several sulfur compounds as electron acceptors. The high mole ratio of hydrogen from glucose and high partial pressures of hydrogen reached during fermentation (>10%) demonstrates the potential of OS1 for biohydrogen production and syntrophy (5, 18). H_2 is a clean-burning fuel that can be made from renewable sources and its combustion does not produce greenhouse gasses. Acetate, which is co-produced with H_2 from glucose, can be converted to methane, another energy-rich biofuel made by acetotrophic methanogens (19). An understanding of the *A. hydrogeniformans* OS1 genome is central to deciphering the metabolic machinery involved in efficient biohydrogen production and as the model for the physiology of other members of the *Synergistetes*. Here, we present a summary of the genome properties of *A. hydrogeniformans* strain OS1 including processes for nutrient uptake, central catabolic metabolism, motility, transcription, and stress adaptation.

3. Classification and features

Morphology and physiology

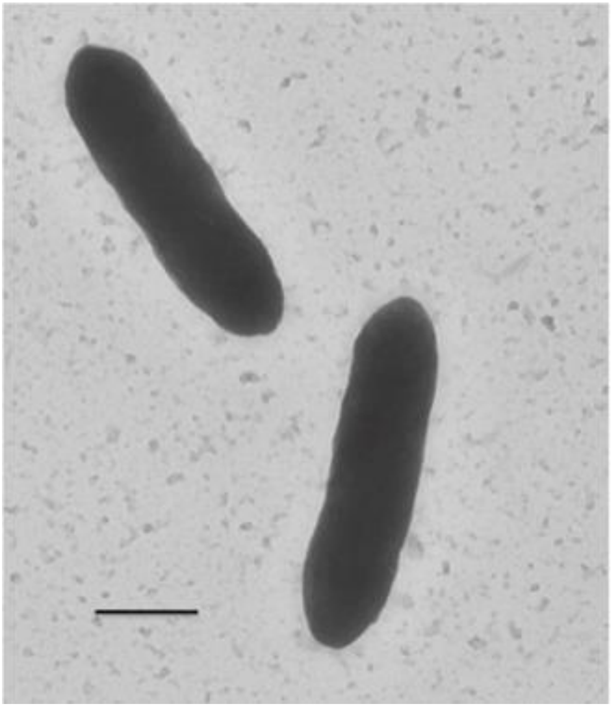

Figure 1. Electron micrograph of *A. hydrogeniformans* strain OS1. Negatively-stained cell image was obtained as described (1).

A. hydrogeniformans strain $OS1^T$ is a member of the phylum *Synergistetes,* a Gram negative-staining, rod-shaped bacterium (Figure 1, Table 1) (5). It grows as non-spore-forming motile cells with an average size ranging from 0.4-0.5 µm in width to 1.7-2.7 µm in length (5). *A. hydrogeniformans* is strictly anaerobic and grows at an optimal temperature of 55°C with a range of 40°C-65°C. It grows in defined culture up to 7% NaCl, which is considerably higher than NaCl concentrations that other species in the *Acetomicrobium* genus tolerate (5, 7). Unlike the other species, NaCl is an absolute requirement for *A. hydrogeniformans* growth which suggests a sodium-based energy strategy (5, 9). Optimal growth occurs at 1% NaCl at pH 7.5 and with a pH range of 6-9. Strain $OS1^T$ can reduce elemental sulfur, thiosulfate, and L-cysteine to sulfide, but cannot reduce sulfate, sulfite, or nitrate. Like other currently-known members of the genus *Acetomicrobium, A. hydrogeniformans* strain $OS1^T$ can reduce crotonate to butyrate when glucose is present, but cannot ferment crotonate alone. In pure culture, $OS1^T$ consumed 14.1 mmol/L glucose, and produced 47.2 mmol/L H_2 and 19.7 mmol/L acetate (5). When grown in co-culture with *Methanothermobacter thermautotrophicus* as the hydrogen-oxidizing methanogen, OS1 consumed 8.4 mmol/L glucose and produced 17.6 mmol/L acetate and 8.1mmol/L methane. The genetic potential for syntrophic metabolism of other substrates is currently unknown.

A. hydrogeniformans strain OS1^T was isolated from Alaskan oil pipeline water production stream and cultured anaerobically at 50°C in 1% NaCl with a 100% nitrogen gas phase. It was the dominant culturable heterotrophic microorganism detected (5). Colonies in agar medium were small, circular, smooth, and yellowish in color.

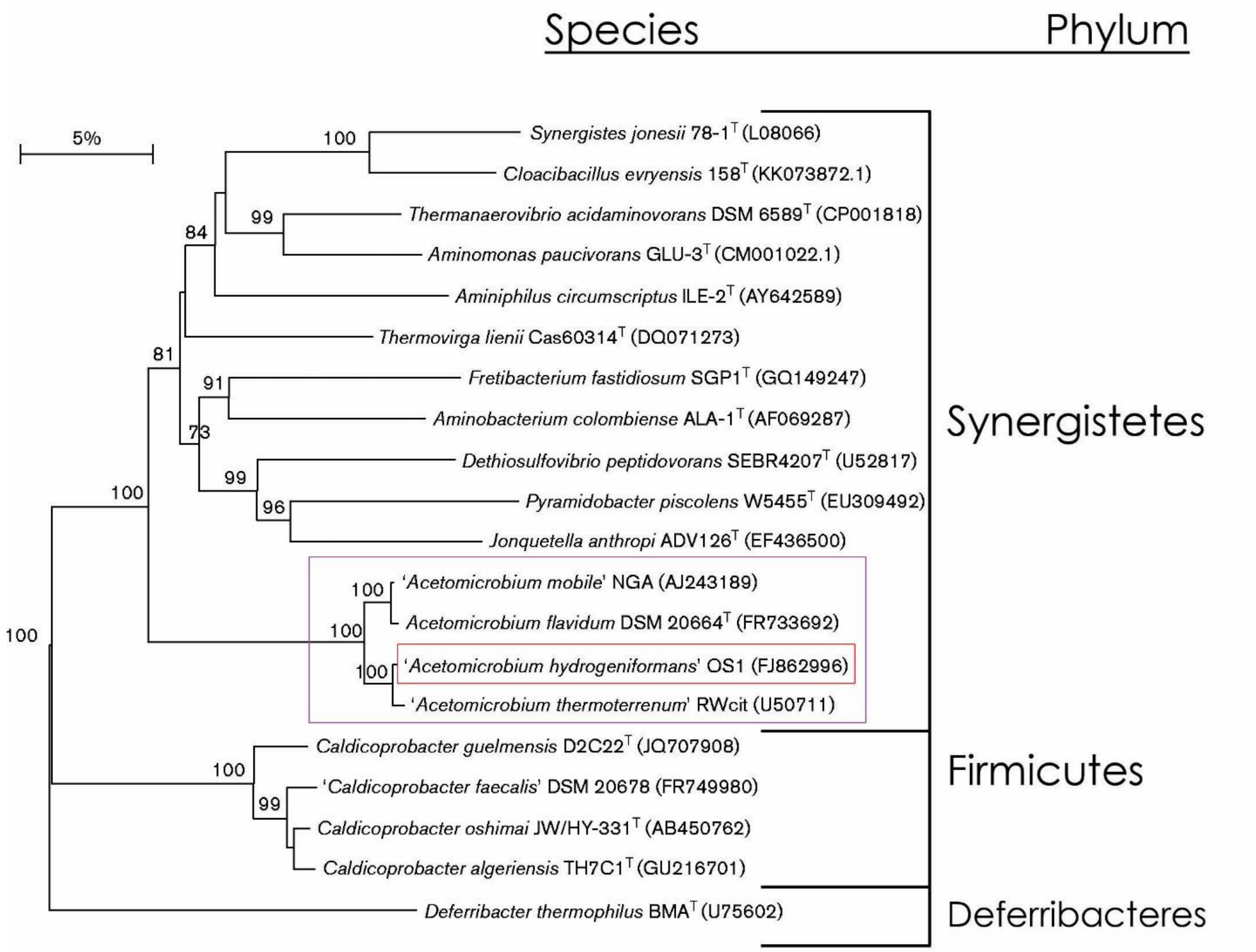

Figure 2. Neighbor-joining phylogenetic tree based on 16S rRNA gene sequence showing the relationships between *Acetomicrobium flavidum* and related representatives of the family *Synergistaceae*. Bootstrap values of 70 % or higher (based on 2000 repetitions) are shown at branch nodes. Bar, 5 substitutions per 100nt. A phylogenetic tree was reconstructed with the TREECON program, using the neighbor-joining method (2). Tree topology was evaluated by a bootstrap analysis using 2000 resampling of the sequences (3). Its topology was also supported

using the maximum-parsimony and maximum-likelihood methods. The results indicate that *Acetomicrobium flavidum* pertains to the family *Synergistaceae*, phylum *Synergistetes*, having *Anaerobaculum mobile* (99.9 % similarity) (4), *Anaerobaculum hydrogeniformans* (96.8 % similarity) (5) and *Anaerobaculum thermoterrenum* (96.4 % similarity) (6), as its closest phylogenetic relatives.

Originally published in Ben Hania et al (1) and used with permission, copyright clearing house license 4590300479047. The preceding figure legend is taken directly from the original paper. Acetomicrobium hydrogeniformans *is boxed in red, and the genus* Acetomicrobium *is boxed in purple. The genus* Thermovirga *diverged from a common ancestor with* Acetomicrobium.

16S rRNA gene sequence analysis

The phylogeny of *A. hydrogeniformans* OS1^T is shown in figure 2. The analysis of the two 16S rRNA genes present in the *A. hydrogeniformans* OS1 genome revealed identical nucleotide sequences over the entire length of the 1409 bp nucleotide sequence. A 16S neighbor-joining phylogenetic tree of the *Synergistetes* shows the genus *Acetomicrobium* split off from the rest of the *Synergistetes* at the phylum root, indicating that within the confines of the phylum it is distinct from the other genera and species (7) (figure 2). Within the genus *Acetomicrobium*, the species *A. hydrogeniformans* groups with *A. thermoterrenum.* All species in the genus ferment simple sugars to produce hydrogen. Members of the *Synergistetes* including *Acetomicrobium* species catabolize amino acids, and other cultured members of the phylum *Synergistetes* use dibcarboxylates, but the latter property has not been tested with *Acetomicrobium* species (5). All four described *Acetomicrobium* species reduce thiosulfate, elemental sulfur, and cystine, but not sulfate or sulfite, to hydrogen sulfide (5).

Chemotaxonomy

The genome of *A. hydrogeniformans* OS1 has a G+C content of 46.6 mol% (IMG JGI). Cellular fatty acid methyl ester (FAME) analysis revealed that the dominant fatty acid is iso-$C_{15:0}$ (90.5%), followed by iso-$C_{11:0}$ (4.4%), iso-$C_{13:0}$ 3-OH (3.3%), and iso-$C_{13:0}$ (1.8%) (5). Polar lipids detected included diphosphatidylgylcerol, phosphatidylglycerol, and phosphatidylethanolamine, other unknown phospholipids, and several aminophospholipids.

4. Genome sequencing and annotation

Genome project history

A. hydrogeniformans OS1 (DSM 22491) was selected for sequencing based on its high hydrogen yield (5) and its association with microbially-induced oil pipeline corrosion (14). The genome project is deposited in the Genomes On Line Database (GOLD) as project ID Gi0029125 (20) and the sequence is deposited in GenBank (NZ_ACJX00000000.3; NCBI Project ID 33123). Sequencing, finishing, and annotation were performed by the Genome Institute at Washington University (WGC) as described above. A summary of the project information is shown in Table 2.

Growth conditions and DNA isolation

A. hydrogeniformans strain OS1 was grown in the previously-described basal medium with glucose (5). High molecular weight DNA was isolated from cell pellets using the CTAB method described by the Joint Genome Institute (21).

Genome sequencing, assembly and annotation

A Newbler draft assembly was generated on *Acetomicrobium hydrogeniformans* OS1 DNA sample/isolate AAAB-OS1-16445, using Newbler version MapAsmResearch-10/14/2011. The input for this assembly was 100X of Illumina 100bp paired-end sequences generated on the Hiseq (Version 3) platform. The genome was screened for contamination and sent to The Washington Genome Institute's (TGI) manual sequence improvement group (MSI). Automated and manual finishing approaches were performed to improve the draft assembly by resolving obvious miss-assemblies, making joins between contigs, editing the consensus sequence to correct sequence errors, trimming irrelevant sequences on the ends of contigs, and ordering and orienting the contigs where possible. Annotation was performed on the improved sequence assembly at WGC.

The TGI's gene annotation process includes generating both *ab initio* and evidence-based (BLAST) predictions. Coding sequences were predicted using GeneMark (22) and Glimmer3 (23). Intergenic regions not spanned by GeneMark and Glimmer3 were blasted against NCBI's non-redundant bacterial (NR) database and predictions generated based on protein alignments. Loci were then defined by clustering predictions within the same reading frame. The best prediction at each locus was selected by comparing all predictions against the best evidence (non-redundant bacterial, NR and Pfam) (24) and resolving overlaps between adjacent coding genes. tRNA genes were determined using tRNAscan-SE (25) and non-coding RNA genes by RNAmmer (26) and Rfam (27).

The final gene set was processed through KEGG (28) to assign metabolic pathways, psortB to predict subcellular localization, and Interproscan (29) to find functional domains to determine possible function. Gene product names were determined by BER (http://sourceforge.net).

5. Genome Properties

A list of genome statistics is provided in Table 3. The genome consists of 2,123,925 bp of which 91.79% is predicted to code for proteins (2,068 ORFs), and the G+C content is 46.53%. Of these, approximately 84.5% (1,797 genes) were assigned to a putative function while the remaining 15.5% (271 genes) are without assigned functions. The distribution of genes into

COGs functional categories is presented in Table 4. The genome has 47 tRNA genes and two copies each of the 16S and 23S rRNA genes.

***A. hydrogeniformans* OS1 has no extrachromosomal elements**

The original annotation of *A. hydrogeniformans* OS1 reported a single circular plasmid of 5,386 kb in size containing seven ORFs (OS1_ 2124-02130) annotated as two bacteriophage replication A proteins, phage protein C, bacteriophage scaffolding protein D, a capsid/F protein, a major spike protein (G-protein), and a microvirus H pilot protein. This sequence matches precisely that of the viral plasmid used as a control in sequencing and has therefore been determined to be contamination by the sequencing facility. *A. hydrogeniformans* OS1, therefore, has no extrachromosomal elements.

6. Central metabolism

The *A. hydrogeniformans* OS1 ORFs were organized into pathways, and most pathways considered essential for growth were detected. A metabolic reconstruction of the major metabolic and cell functions is shown in Figure 3. A description of the major features follows.

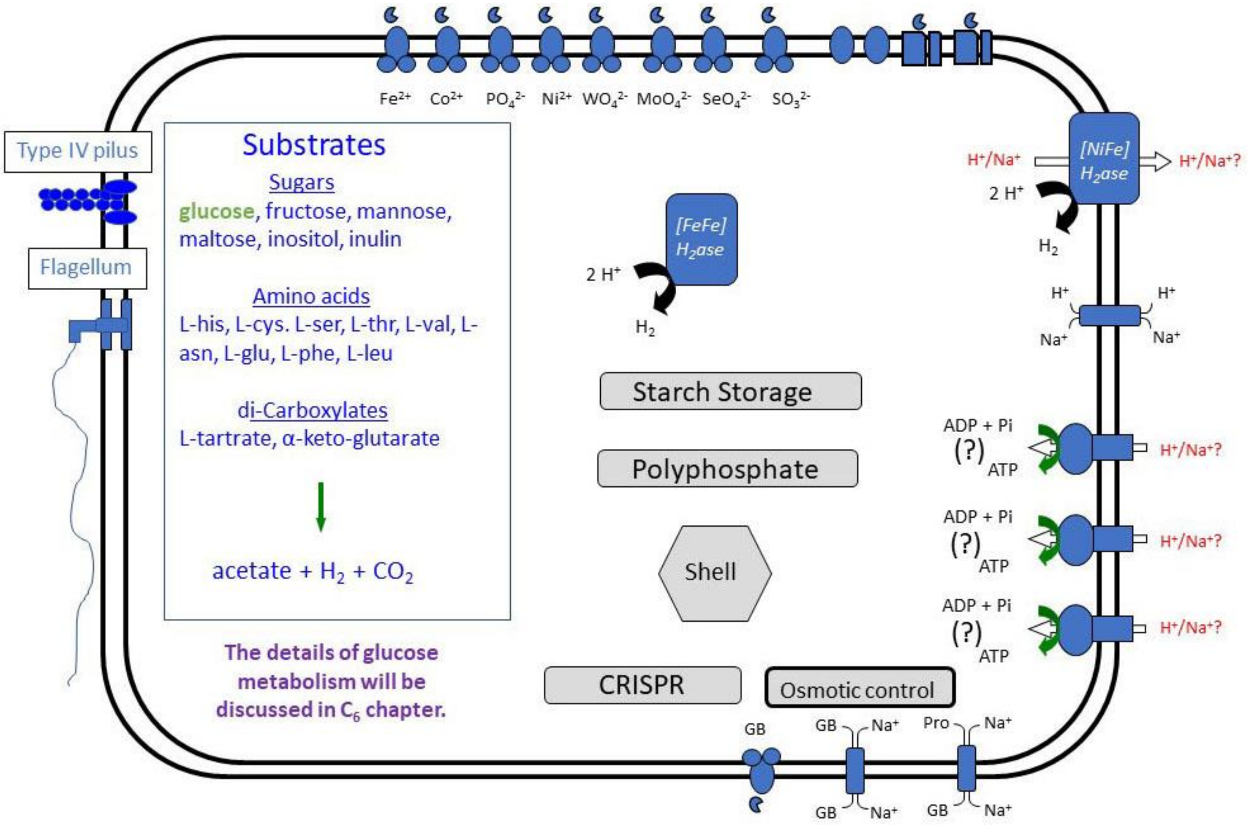

Figure 3. A metabolic reconstruction of the *A. hydrogeniformans* OS1 genome content.

Fermentation and energy generation

Although *A. hydrogeniformans* OS1 can reduce elemental sulfur, thiosulfate, and L-cysteine to sulfide, the pathways are unknown. The genome lacks genes needed for respiration of sulfate or nitrate genes, genes for b- or c-type cytochromes, and genes for ubi- or menaquinone synthesis. The metabolic strategy appears to be primarily limited to fermentation of simple sugars, amino acids and certain dicarboxyilic acids and an as-yet undescribed respiratory mode for elemental sulfur or thiosulfate. However, a dissimilatory sulfite reductase was present (OS1 01021) plus one ETF gene cluster (OS1_0897-0898).

Glycolysis and Entner-Doudoroff pathways

Pathways for glycolysis and the Entner-Doudoroff pathway appear incomplete in *A. hydrogeniformans* OS1, although the organism is known to ferment simple sugars and some amino acids (5). ATP formation by substrate-level phosphorylation would occur by the conversion of acetyl-phosphate to acetate by acetate kinase (ack), which in turn is derived from acetyl-CoA by *phosphotransacetylase* (pta).

A. hydrogeniformans lacks a complete set of genes for known TCA cycles

A. hydrogeniformans OS1 does not appear to have a complete suite of genes for the TCA cycle. Of the necessary genes it only has paralogues for isocitrate dehydrogenase (OS1_0231) and fumarate hydratase (OS1_0871-0872).

A. hydrogeniformans lacks a complete set of genes for known (oxidative/catabolic) pentose phosphate pathways

Genes are outlined in table S2. The oxidative arm of the pentose-phosphate pathway generates NAD(P)H which is in turn used as a reducing agent in reductive biosynthesis pathways. An organism lacking this pathway must find another mechanism to generate NAD(P)H or a manner of compensating. From this pathway comes also ribose-5-phosphate (R5P), a precursor for nucleotide synthesis, and erythrose-4-phosphate (E4P), a precursor in the synthesis of aromatic amino acids. The method of NADH generation in *A. hydrogeniformans* is unknown.

A. hydrogeniformans may have a complete set of genes for the Wood-Ljungdahl pathway

There are two versions of the Wood-Ljungdahl pathway: one as found in acetogenic bacteria (pathway I), and one as found in methanogenic archaea (pathway II). Pathway numbering comes from the metabolic pathway aggregation database MetaCyc. A survey of the *A. hydrogeniformans* genome reveals potential genes to fill every step in pathway I, but missing paralogues for two important genes in the second pathway, an H_2-forming methylene-H4MPT dehydrogenase and an F_{420}-dependent methylene-H4MPT reductase (table S3). At the point in pathway I in which either an NAD-utilizing or a ferredoxin-utilizing methylenetetrahydrofolate reductase can be used, it is the NAD-utilizing paralogue found in the OS1 genome (OS1_1179).

Hydrogenases and Dehydrogenases

Pyruvate Dehydrogenase (PDH) and the Pyruvate Dehydrogenase Complex (PDC)

A. hydrogeniformans has genes for all the components of the pyruvate dehydrogenase complex (PDC) as known in *E. coli* (table S1). This includes pdhAB (adjacent on genome), the alpha and beta parts of the E1 subunit (OS1_ 0088 and 0089), pdhC, the dihydrolipoamide acetyltransferase E2 subunit (OS1_0084), and pdhD, the dihydrolipoamide dehydrogenase subunit (OS1_0776).

Hydrogenases

A. hydrogeniformans OS1 possesses two gene clusters with probable hydrogenase-related functions (OS1_0245-0251, and OS1_ 0493-0502). The first appears to be a membrane-associated [NiFe] hydrogenase and may be part of a formate hydrogenlyase complex, while the latter appears to be a soluble [FeFe] hydrogenase. There are also two genes with putative formate dehydrogenase activity. (table S1) No genes were found for pyruvate formate-lyase.

Solute transport systems

An inventory of membrane transporters revealed multiple genes for ABC-type, secondary, TRAP-type, and P-type transporters (Figure 3, Table S3). These include uptake systems for sugars, amino acids and peptides, dicarboxylates, and trace metals (Ni, Co, Mo, W, Se and Fe) needed for electron transfer proteins and other fermentation enzymes. Together, these transport-related genes comprise approximately 15.08% of the genes present in the genome.

ABC-type transport systems

A. hydrogeniformans OS1 contains 110 ABC-type transport-associated genes (Table S3), divided into 28 multi-component systems plus several orphan gene fragments. There are two ABC-type transport systems for cobalt (OS1_ 0324-00328; 01472-01474), two ferric iron (Fe^{3+}) transport systems (OS1_ 0465-00466; OS1_ 1107-01109), one for molybdate (OS1_ 1146-01147), one for cobalt-tungstate (OS1_ 0476-00478), a Zn^{2+}/Mn^{2+} transport system (OS1_ 0901-00903), a nickel system (OS1_ 1008-01010), four branched-chain amino acid transport systems (OS1_ 0371-00375; OS1_ 0422-00423; OS1_ 1325-01329; OS1_ 1845-01848), four oligopeptide transporters (OS1_ 0805-00807 , OS1_ 0919-0922, OS1_ 01008-0001010, OS1_ 1956-1959). Several additional amino acid and sugar transport systems are also present (Table S3).

TRAP-type transport systems

There are 23 genes associated with TRAP-type transport systems divided into 7 multipart systems plus two orphan proteins (Table S3). These include five di- or tri-carboxylate uptake systems. Representative solute systems are depicted in the metabolic reconstruction of *A. hydrogeniformans* OS1 (Figure 3).

Na^+/H^+ antiporters

Also predicted are two multi-subunit Na^+/H^+ antiporters (OS1_0251-0258; OS1_1049-1056) and two orphan genes with homology to the NhaC subunit (OS1_00271, OS1_00341). OS1_0251-0258 has the gene order MnhEFGx(A)BCD; the A subunit is in parenthesis because the genome annotation was 'hypothetical protein' but BLAST of the protein sequence reveals subunit A to be the closest match of known proteins. The x, OS1_0255, is annotated as "Uncharacterized MnhB-related membrane protein" with a domain of unknown function. This operon therefore has all of the Mnh-associated genes (A-G) used to build the Na/H antiporter as found in *Staphylococcus aureus* (30). The OS1_1049-1056 cluster is predicted to be the genes MnhEFGxBCDD, with the x again as annotated for OS1_0255. Further studies would have to be conducted to confirm if the deletion of subunit A and the duplication of subunit D is an accurate prediction. OS1_1049 and OS1_1050 (the duplicated D subunits) are both predicted 487aa but are only 32% identical.

Other transport pathways

Nineteen secondary transport systems are present plus one voltage-gated ion (OS1_ 0824), one vacuolar-type H^+-translocating pyrophosphatase (OS1_ 1397) and one sodium transporting OAA decarboxylase cluster (OS1_ 0873-00877).

Cellular adaptation and stress response

A. hydrogeniformans OS1 has multiple genes for stress adaptation that include those for motility, pili, osmotic adaptation, heavy metal resistance, multi-drug export, and toxin/antitoxin synthesis. There are multiple glycine, choline, and carnatine/betaine/proline transport systems along with genes for N-acetyl-beta-lysine synthesis and/or metabolism (OS1_ 0721-0722), heavy metal resistance (OS1_ 0012), heavy metal transport (OS1_ 0068) and an ABC-type transport system involved in resistance to organic solvents (OS1_ 1594-01595). There is also one twenty-one gene CRISPR locus (Figure 3) and a gene predicted to code an H_2O_2-forming NADH oxidase (OS1_1113). The genome lacks genes for catalase, superoxide dismutase, peroxidase, and the SoxRS-type response. However, two peroxiredoxin genes which may metabolize hydrogen peroxide (OS1_ 1723-01724) and a nitroreductase (OS1_ 0976) are present.

Energy Storage

A glycogen/starch storage-utilization four gene cluster (OS1_ 0505 to- OS1_ 0508) enables energy conservation when excess carbon is present (Figure 3, Table S4).

Carboxysome Shells

A. hydrogeniformans genome has a twenty-one gene cassette (OS1_ 1070 to OS1_ 1090) for synthesis of a carboxysome-like shell structure including five shell proteins, four ethanolamine-utilization proteins, four propanediol dehydratase proteins and related enzymes for glycerol/ethanolamine/1-propanediol metabolism (Figure 3, Table S5).

Motility, taxis and pili

A. hydrogeniformans genome contains a complete inventory of genes needed to make flagella (Figure 3). Motility was not previously reported (5) but negative-stained *A. hydrogeniformans* OS1 cells exhibit polar flagella (Figure 2), and cells exhibit motility in wet-mount slides (Ralph Tanner, personal communication). These genes include flagellar structural proteins for the basal-body, motor, hook, filament, L, P and M-rings) (i.e., FlgABCDGHIKLM, FliCDEFGHIJKLMNPQRS and MotAB). No FlgEFJ or FliOT genes were detected. Flagella biogenesis genes include a σ28 and an anti-σ28 factor (*flgM*), although no *E. coli*-type master switch proteins *flhCD* are present.

The *A. hydrogeniformans* OS1 genome also contains multiple genes for chemoreception and chemotaxis: *cheC (3 copies)* and *cheW (3 copies),* plus one copy each for *cheB*, *cheD*, *cheR*, and *cheY* (Table S6). No *cheA*, *cheZ* or *cheV* genes were evident. There are four genes for membrane-bound MCP's (methyl-accepting chemotaxis proteins), plus a gene for a soluble-type MCP with chemoreceptor domains for environmental signal detection (Table S6). The environmental/metabolic signals of the chemotactic response in *A. hydrogeniformans* OS1 are unknown.

The *A. hydrogeniformans* OS1genome also contains a complement of ten *pil/pul* genes needed for synthesis and assembly of an undescribed type IV pilus apparatus. These include PilY1 and PilT genes for tip-associated adhesion, and retraction proteins. No pre-pilin genes are apparent (Figure 3, Table S7).

Regulation and signal transduction

The *A. hydrogeniformans* OS1 genome contains a prototypical bacterial RNA core polymerase ($\alpha\beta\beta'\omega$: *rpoA*, *rpoB*, *rpoC*, *rpoZ*) along with four sigma factors and two anti-sigma factors that confer promoter specificity (Table S8). These include three general housekeeping $\sigma70$ factor (*rpoD*), a flagella biosynthesis factor (*fliA/WhiG*: OS1_ 1260), a FlgM anti-sigma factor (OS1_ 1490), and a Ser/Thr protein kinase anti-sigma factor gene (OS1_ 0503). Interestingly, one of these anti-sigma regulatory factor genes (the Ser/Thr protein kinase; OS1_0503) is located in a putative iron-only-type hydrogenase gene cluster. No 54σ-type factors (*rpoN*) or 32σ-type factors (*rpoH*) similar to that found in nitrogen or heat shock control in *E. coli* and other bacteria were detected, nor does the genome contain 54σ-interacting transcriptional regulators. *A. hydrogeniformans* OS1 also has a relatively small number of primary transcription factors (53 genes) containing a helix-turn-helix motif for DNA binding. Thus, it appears to have adopted a minimal regulatory strategy, although the substrate range is significant with respect to sugars, amino acids and dicarboxylates used (5). There are six two-component regulatory systems based on the presence of 6 histidine kinase-type sensor transmitter genes and 8 receiver-only domain protein genes (Table S9) in the genome. Three of these are genetically linked to a cognate two-component protein member as its potential interacting partner in signal transduction. Little is known about translational or post-translational control operating in this or other *Synergistetes* species.

Archaeal-like genes

The genome possesses three clusters of genes encoding either archaeal-like or vacuolur-like A/V ATP synthases as noted above (indistinguishable informatically), plus thirteen other archaeal-type annotated genes (Table S10). These include genes related to methanogenic Coenzyme F_{390} synthtase, Factor F_{420}-reducing hydrogenase-like alpha and beta subunits, formyl methanofuran dehydrogenase A, B, C, D and E subunits, methanyltetrahydromethanopterin cyclohdrolase, tetrahydro methanopterin N-formyl transferase, an archaeal flavoprotein, archaeal fructose 1,6-bisphosphatase and an archaeal phosphogycerate mutase. It also has other archaeal-like annotated genes, including those acting in glycolysis; of note, we found four putative candidates for glyceraldehyde-3-phosphate oxidoreductase, a ferredoxin-utilizing enzyme with archaeal properties. These genes may have been acquired by lateral gene transfer (11-13). Curiously, there is also a gene for the synthesis of wyosine used in a modification of tRNAs found in eukaryotes and archaea but not bacteria (OS1_1106).

Despite this seeming affinity for archaeal gene paralogues, the version of the Wood-Ljungdahl pathway OS1 may have is closest to that found in acetogenic bacteria, and not that found in methanogenic archaea (Table S3).

A. hydrogeniformans lacks the bacterial FoF1-type ATP Synthase

Strikingly, the *A. hydrogeniformans* OS1 genome lacks genes for the prototypical bacterial type FoF1 type ATP synthase (Figure 3). Instead, it has multiple gene clusters for three complete archaeal-or-vacuolur-type (V/A type) ATP synthase operons (OS1_ 0767-00775; OS1_ 1788-01795; OS1_ 1897-01905). Which type of ATP synthase these are (i.e. V or A type, or: ion-pumping only or ATP synthesis activity with same) cannot be determined informatically.

7. Discussion

A. hydrogeniformans OS1 has an 'ancient' metabolism with many archaea-like genes

Phylogenetic studies and metabolic reconstructions indicate that the most ancient microbes, leading back to LUCA, were acetogenic and methanogenic, H_2-dependent autotrophs using CO_2 as a terminal electron acceptor. This is supported by a phylogenetic tree that links LUCA to archaeal methanogens and acetogenic *Clostridia* bacteria (and *Moorella*, a sub-division formed by organisms formerly lumped into *Clostridia*) (31). Fe-S centers and transition metal utilization (including group 6 elements molybdenum and tungsten), also, are considered hallmarks of ancient metabolism, relics of a proto-life that arose from spontaneous geothermal chemical interactions, such as would be found at submarine hydrothermal vents. At these hydrothermal vents, the water is hot and rich in H_2, an environment in which spontaneous synthesis of organic molecules is possible. Indeed, such spontaneous organic syntheses as may have fostered proto-life have been observed at the marine vents. (31)

While *A. hydrogeniformans* OS1 has metabolic hallmarks like those of these early enzymes, such as predicted tungsten/molybdenum enzymes, its metabolism not H_2-dependent, and, indeed, growth is *halted* at 0.2 atm of H_2. It, instead, *produces* H_2. It grows syntrophically with methanogen *Methanothermobacter*, which keeps the partial pressure of hydrogen low by processing the H_2 released by *A. hydrogeniformans*. (personal communication, Ralph Tanner) It makes sense that it would co-evolve with H_2-requiring organisms in a mutually beneficial relationship, and within that confined space, horizontal gene transfer could occur. Further research would have to be conducted.

The Wood-Ljungdahl pathway, evidence of which is found in *A. hydrogeniformans*'s genome, requires molecular hydrogen for an electron donor to fix CO_2. *A. hydrogeniformans* produces H_2; it could perhaps recycle some of this waste product into CO_2 fixation. The authors of a 2016 *Nature Microbiology* phylogenetics paper (henceforth the "Weiss" paper) suggest that the last universal common ancestor (LUCA) utilized the Wood-Ljungdahl pathway and was a CO_2-and-N_2-fixing, H_2-dependent anaerobic thermophile autotroph (31). If this is the case, the Wood-Ljungdahl pathway is ancient indeed and has shown its evolutionary fitness within the type of environment the LUCA would have inhabited. Prior to this study, a 2011 review of carbon dioxide fixation pathways predicted the following:

> "It appears that the formation of an activated acetic acid from inorganic carbon represents the initial step toward [the initiation of the first, ancestral] metabolism [in LUCA]. Consequently, biosyntheses likely started from activated acetic acid and gluconeogenesis preceded glycolysis." (32)

This first prediction, at least, aligns with the conclusions reached in the Weiss paper: the first reactions spanning the transition from proto-life to a 'life' with a self-contained metabolism involved these acetic acid-producing, carbon fixing pathways. Weiss further refines this to be the

Wood-Ljungdahl pathway, specifically. With the incorporation of the Wood-Ljungdahl pathway we therefore see the appearance of yet another aspect of ‘ancient’ metabolism in *A. hydrogeniformans,* in conjunction with the ferredoxin-utilizing glycolysis enzymes, the A/V-type ATP synthase, and the [FeFe] hydrogenase. Curiously, the version of the Wood-Ljungdahl pathway we see in *A. hydrogeniformans* most closely resembles that found in acetogenic bacteria, not archaea; this is the opposite from what we see in the carbon utilization pathway, as will be outlined in the next chapter.

A. hydrogeniformans may utilize a version of the Wood-Ljungdahl pathway (reductive acetyl-CoA pathway) found in acetogenic bacteria to synthesize acetyl-CoA from H_2 and CO_2

Some acetogenic microbes (the category in which OS1 belongs) and methanogens utilize the Wood-Ljungdahl pathway, also known as the reductive acetyl-CoA pathway. In this pathway molecular hydrogen is utilized as an electron donor, while carbon dioxide acts as an electron acceptor and the carbon therein is used for biosynthesis (i.e. it is a carbon-fixing pathway, or carbon dioxide fixation pathway). Acetyl-CoA is the final product of the pathway, which is then utilized in various other metabolic pathways. The Wood-Ljungdahl pathway is also capable of acing in reverse, oxidizing organic compounds to produce electrons for other reducing reactions. (33) It is one of six CO_2-fixation pathways found in currently-extant prokaryotes, the most common of which is the reductive pentose-phosphate pathway, or the Calvin-Benson cycle (32). The reductive pentose-phosphate cycle is also the means by which all plants perform fixation of CO_2. Fuchs notes that the other five carbon fixation pathways known in prokaryotes all “pivot on acetyl-coenzyme A, the turntable of metabolism, demanding a gluconeogenic pathway starting from acetyl-coenzyme A and CO_2 (32).”

There are two versions of the Wood-Ljungdahl pathway: one as found in acetogenic bacteria (pathway I), and one as found in methanogenic archaea (pathway II). In pathway I, acetyl-CoA for acetate production is produced. In pathway II, acetyl-CoA for carbohydrate synthesis is produced. Pathway numbering comes from the metabolic pathway aggregation database MetaCyc. In *Moorella thermoacetica* (formerly *Clostridium thermoaceticum*), the carbon dioxide and low-potential electrons (bound to ferredoxin) from the breakdown of oxalate by oxalate:ferredoxin oxidoreductase (a member of the OFOR family of enzymes, further discussed in the next chapter) are fed into the Wood−Ljungdahl pathway (pathway I) for acetogenesis (34-36). In both versions of the pathway, CO_2 begins the pathway, either taken up by the cell itself or produced as a metabolic by-product by another cellular process, and acetyl-CoA is the final product.

Searches for possible enzymes to fill in each step in the pathway were conducted based on pfams found in enzyme paralogues characterized in *Moorella* species (acetogenic bacteria) for pathway I and *Methanosarcina* species (methanogenic archaea) for pathway II. For double-checking, EC numbers were used. A gene was considered a possible paralogue if it matched at least three of the four decimals in the EC numbering scheme, meaning that the gene belongs to a family that acts upon the class of target molecule, but the specific substrate may be different. These genes that are a 3/4 match for EC number are marked with an asterisk in table S3. Unmarked genes are a 4/4 match for the EC number. This latitude was given with mind to the variability in enzymes across species, and with personal experience in the ability of the gene labeling algorithms to accurately predict a class of enzyme but weakness in the prediction of the specific substrate within that class of molecule, especially for 'archaeal' genes, data about which is still lacking in the enzyme-signature determining algorithms. (This observation is documented in the carbon pathway chapter as pertains to identifying GAPOR.) The Wood-Ljungdahl pathway I may therefore be another way *A. hydrogeniformans* produces acetate, from the fixation of CO_2, and one way it offloads molecular hydrogen, which depresses its growth, but the evidence in favor of this assertion is loose and must be examined with more rigorous methods. Even given the latitude of the search criteria, pathway II was not found represented, and therefore it can be stated with reasonable confidence that OS1 is not likely to utilize a version of Wood-Ljungdahl pathway II.

8. Conclusion

The *A. hydrogeniformans* OS1 genome sequence is of particular interest given its ability to produce hydrogen gas and its association with oil pipeline corrosion (5). Pathways for sugar, amino acid and dicarboxylate metabolism, and thiosulfate and sulfur reduction were either incomplete or absent (Figure 3), suggesting the involvement of unusual and/or unpredicted genes/proteins. If it performs respiration, *A. hydrogeniformans* OS1 would rely on archaeal systems for ATP synthesis from chemiosmotic gradients. The metabolic reconstruction also provides a framework to examine the phylogeny of related *Acetomicrobium* species plus other members of the phylum *Synergistetes* for which little experimental data exist.

Accession number(s). This Whole Genome Shotgun project has been deposited in GenBank under the accession no. [ACJX]00000000. The OS1_ here referenced correspond to genome taxon ID 2517487012 at the Integrated Microbial Genomes database.

9. Acknowledgements

The initial genomics work is documented in the following paper and this chapter owes considerable debt to the authors: Cook, L. E., Gang, S. S., Ihlan, A., Maune, M., Tanner, R. S., McInerney, M. J., Weinstock, G., Lobos, E. A., and Gunsalus, R. P. (2018). Genome Sequence

of Acetomicrobium hydrogeniformans OS1. *Genome announcements*, *6*(26), e00581-18. doi:10.1128/genomeA.00581-18

This work was supported by Department of Energy contract DE-FG02-08ER64689 (to RPG, RST and MJM), the UCLA-DOE Institute award DE-FC02-02ER63421, and the National Science Foundation grant NSF 1515843 (to RPG and MJM).

Tables

Table 1: Classification and general features of *A. hydrogeniformans* strain OS1 by MIGS recommendations (37)

Table 1			
MIGS ID	**Property**	**Term**	**Evidence code**
	Current classification	Domain *Bacteria*	TAS (5)
		Phylum *Synergistetes*	TAS (5)
		Class *Synergistia*	TAS (5)
		Order *Synergistales*	TAS (5)
		Family *Synergistaceae*	TAS (5)
		Genus *Acetomicrobium*	TAS (5, 7)
		Species *Acetomicrobium hydrogeniformans*	TAS (5, 7)
MIGS-7	Subspecific genetic lineage (strain)	OS1	TAS (5)
MIGS-12	Reference for biomaterial	Maune and Tanner 2012	TAS (5)
	Gram stain	Gram-negative	TAS (5)
	Cell shape	rod-shaped	TAS (5)
	Motility	non-motile	TAS (5)
	Sporulation	non-sporulating	TAS (5)
	Temperature range	Moderately thermophilic, 40-65C	TAS (5)
	Optimum temperature	55C	TAS (5)
	Salinity	optimum growth at 1%	TAS (5)
MIGS-22	Relationship to oxygen	obligate anaerobe	TAS (5)
	Carbon source	In pure culture: D-fructose, D-glucose, pyruvate, L-tartrate, Casamino acids, L-asparagine, L-cysteine, L-histidine, L-serine, L-threonine, L-valine, alpha-ketoglutarate, L-glutamate, Malonate, Maltose, D-mannose, Inositol, Tryptone, L-leucine, L-phenylalanine, inulin. In co-culture:	TAS (5)

		glucose.	
	Energy metabolism	Chemo-organotroph	TAS (5)
MIGS-6	Habitat	saline oil production waters	TAS (5)
MIGS-6.2	pH	optimum 7.0	TAS (5)
MIGS-15	Biotic relationship	free living, syntrophy	TAS (5)
MIGS-14	Known pathogenicity	None	TAS (5)
MIGS-16	Specific host	None	
MiGS-14	Biosafety level	1	TAS [ref 7]
MIGS-19	Trophic level	producer	
MIGS-23.1	Isolation	oil production fluids	TAS (5)
MIGS-4	Geographic location	Alaska, USA	TAS (5)
MIGS-5	Time of sample collection	1996	TAS (5)
MIGS-4.1	Latitude	70º 19' North	
MIGS-4.2	Longitude	149º 35' West	
MIGS-4.3	Depth	ground level	
MIGS-4.4	Altitude	ground level	
Evidence codes - TAS: Tracible Author Statement (i.e., a direct report exists in the literature). Evidence codes are from the Gene			
Onotology project [ref 8]			

Table 2. Genome sequence project information of *A. hydrogeniformans* OS1 according to the MIGS recommendations (37).

MIGS ID	**Property**	**Term**
MIGS-31	Finishing quality	Draft
MIGS-29	Sequencing Platforms	Illumina
MIGS-30	Assemblers	Newbler version MapAsmResearch-10/14/2011
MIGS-32	Gene calling method	GeneMark and Glimmer3
	GenBank Date of Release	15-Nov-15
	GOLD ID	Gi0029125
	NCBI Project ID	33123
	Database: IMG	10679
MIGS-13	Source material identifier	DSM 22491
	Project relevance	Pipeline corrosion / Biofuels

Table 3. Genome Statistics [a]

Attribute	**Number**	**% of Total**
Genome size (bp)	2,123,925	100.00%
DNA coding region (bp)	1,949,595	91.79%
DNA G+C content (bp)	988,228	46.53%
DNA Scaffolds	8	
Extrachromasomal elements	0*	
Total genes	2,128	100.00%
RNA genes	60	2.82%
rRNA operons	6	
rRNA genes	6	0.28%
5S rRNA	2	0.09%
16S rRNA	2	0.09%
23S rRNA	carc	0.09%
tRNA genes	47	2.21%
Other RNA genes	7	0.33%
Protein-coding genes	2,068	97.18%
Pseudo genes	0	0.00%
Genes with function prediction	1,797	84.45%
Genes in paralog clusters	649	30.50%
Genes assigned to COGs	1,790	84.12%
Genes assigned Pfam domains	1,841	86.51%
Genes with signal peptides	56	2.63%
Genes with transmembrane helices	465	21.85%
CRISPR repeats	1	

[a] data from IMG

*The original annotation predicted one extrachromosomal element. That was contamination; it is a viral plasmid used as a control in sequencing.

Table 4. Number of genes associated with the general COG functional categories

Code	**Value**	**%age**	**Description** [a]
J	146	7.42	Translation, ribosomal structure and biogenesis
A	0	0	RNA processing and modification
K	82	4.17	Transcription
L	101	5.13	Replication, recombination and repair
B	0	0	Chromatin structure and dynamics
D	33	1.68	Cell cycle control, cell division, chromosome partitioning
Y	0	0	Nuclear structure
V	16	0.81	Defense mechanisms
T	55	2.79	Signal transduction mechanisms
M	112	5.69	Cell wall/membrane/envelope biogenesis
N	62	3.15	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	42	2.13	Intracellular trafficking, secretion, and vesicular transport
O	61	3.1	Posttranslational modification, protein turnover, chaperones
C	178	9.04	Energy production and conservation
G	127	6.45	Carbohydrate transport and metabolism
E	247	12.55	Amino acid transport and metabolism
F	63	3.2	Nucleotide transport and metabolism
H	93	4.73	Coenzyme transport and metabolism
I	37	1.88	Lipid transport and metabolism
P	99	5.03	Inorganic ion transport and metabolism
Q	28	1.42	Secondary metabolites biosynthesis, transport, and catabolism
R	226	11.48	General function prediction only
S	160	8.13	Function unknown
n/a	338	15.88	Not in COGs

[a] data from IMG

Supplementary Materials:

Table S1 Hydrogenase and dehydrogenase-containing gene clusters in OS1. Includes a membrane-bound [NiFe] hydrogenase associated with a formate hydrogenlyase complex, and a soluble [FeFe] hydrogenase.

[NiFe] Hydrogenase / Formate hydrogenlyase complex **(Formate dehydrogenase + hydrogenase)** **(Membrane-bound)**	
Locus Tag	**Gene Product Name**
0245	Formate hydrogenlyase subunit 6/NADH:ubiquinone oxidoreductase 23 kD subunit (chain I)
0246	Formate hydrogenlyase subunit 4
0247	membrane-bound hydrogenase subunit alpha
0248	Respiratory-chain NADH dehydrogenase, subunit
0249	membrane-bound hydrogenase subunit mbhJ
0250	hypothetical protein
0251	multicomponent Na+:H+ antiporter subunit D
0252	multicomponent Na+:H+ antiporter subunit C
0253	multicomponent Na+:H+ antiporter subunit B
0254	hypothetical protein
0255	Predicted subunit of the Multisubunit Na+/H+ antiporter
0256	multicomponent Na+:H+ antiporter subunit G
0257	multicomponent Na+:H+ antiporter subunit F
0258	multicomponent Na+:H+ antiporter subunit E
[FeFe] Hydrogenase **(Soluble)**	
Locus Tag	**Gene Product Name**
0492	hydrogenase maturation protease
0493	NAD-reducing hydrogenase large subunit
0494	NAD-reducing hydrogenase small subunit
0495	formate dehydrogenase major subunit
0496	NADH-quinone oxidoreductase subunit F
0497	NADP-reducing hydrogenase subunit HndB
0498	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
0499	NADH-quinone oxidoreductase subunit E
0500	hypothetical protein
0501	hypothetical protein
0502	Iron only hydrogenase large subunit, C-terminal domain
0503	Anti-sigma regulatory factor (Ser/Thr protein kinase)
0504	hypothetical protein
Potential formate dehydrogenase I	

0295	NADH-quinone oxidoreductase subunit E	
0296	NADH-quinone oxidoreductase subunit F	
0297	formate dehydrogenase major subunit	
Potential formate dehydrogenase II		
0965	NADH-quinone oxidoreductase subunit E	
0966	NADP-reducing hydrogenase subunit HndB	
0967	NADH-quinone oxidoreductase subunit F	
0968	formate dehydrogenase major subunit	
Pyruvate Dehydrogenase		
Locus Tag	Gene	Pyruvate Dehydrogenase Complex Subunits
0089	pdhA	E1 subunit alpha
0088	pdhB	E1 subunit beta
0084	pdhC	E2 (dihydrolipoamide acetyltransferase)
0776	pdhD	E3 (dihydrolipoamide dehydrogenase)

Table S2. List of genes potentially involved in the pentose phosphate pathway in *A. hydrogeniformans* OS1. The pathway is incomplete.

EC #	**Gene Name**	**Enzyme name**	**pfams**	**OS1 gene count**	**OS1 Locus Tags**
1.1.1.49	Zwf	NADP$^+$-dependent glucose-6-phosphate dehydrogenase	PF00479, PF02781	0	-----
3.1.1.31	Pgl	6-phosphogluconolactonase	PF10282	0	-----
1.1.1.44	Gnd	6-phosphogluconate dehydrogenase	PF00393, PF03446	0	-----
5.1.3.1	**Rpe**	**ribulose-phosphate 3-epimerase**	**PF00834**	**1**	**0621**
5.3.1.6	**rpiB**	**ribose-5-phosphate isomerase B**	**PF02502**	**1**	**1736**
	rpiA	ribose-5-phosphate isomerase A	PF06026	0	------
2.2.1.1	**tktA**	**transketolase 1**	**PF00456, PF02779, PF02780**	**1**	**1374**
	tktB	**transketolase 2**	**PF00456, PF02779, PF02780**	**1**	**1374**
2.2.1.2	**talB**	**transaldolase B**	**PF00923**	**1**	**0322**

Table S3: List of genes potentially involved in the Wood-Ljungdahl pathway in A. hydrogeniformans OS1. OS1 has a potential complete set of genes for pathway I, the version of the pathway found in acetogenic bacteria with *Moorella* species as the model. OS1 is lacking any paralogues of genes for two enzymes in pathway II (as found in methanogenic archaea and modeled on *Methanosarcina* species). Enzymes and genes in table are named for the respective names given in the model organism as listed above. Genes that are a 3/4 match for EC number are marked with an asterisk. Unmarked genes are a 4/4 match for the EC number.

Pathway I (acetogenic bacteria)			
EC#	**Enzyme**	**Gene**	**Locus Tag**
1.17.1.10	NADP-dependent formate dehydrogenase	fdh	0297
			0495
			0968
6.3.4.3	formyltetrahydrofolate synthetase (Formate-tetrahydrofolate ligase)	ftfL	1605
			2031
3.5.4.9	methenyltetrahydrofolate cyclohydrolase [multifunctional]	folD	0644
1.5.1.5	methenyltetrahydrofolate cyclohydrolase [multifunctional]		
BRANCH - choose one of the following			
1.5.7.1	methylenetetrahydrofolate reductase (ferredoxin)	?	none
1.5.1.20	methylenetetrahydrofolate reductase (NAD)	metF	1179
BRANCH CLOSE			
2.1.1.258	methyltetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase	acsE	*1178
2.3.1.169	carbon monoxide dehydrogenase/acetyl-CoA synthase sub. Beta	acsB	*0608
1.2.7.4	carbon monoxide dehydrogenase/acetyl-CoA synthase sub. Alpha	acsA	0607
Pathway II (methanogenic archaea)			
EC	**Enzyme**	**Gene**	**Locus Tag**
1.2.7.12	formylmethanofuran dehydrogenase	fwdC	0645
		fwdA	0653
		fwdB	0654
		fwdD	0655
		fwdE	0803
2.3.1.101	formylmethanofuran:H4MPT formyltransferase	ftr	0647
3.5.4.27	methenyl-H4MPT cyclohydrolase	mch	0644
1.12.98.2	H2-forming methylene-H4MPT dehydrogenase	hmd	NONE

1.5.98.2	F420-dependent methylene-H4MPT reductase	mer	NONE
2.1.1.258	methyltetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase	cdhE	*1178
2.3.1.169	carbon monoxide dehydrogenase/acetyl-CoA synthase sub. Beta	cdhB	0608
1.2.7.4	carbon monoxide dehydrogenase/acetyl-CoA synthase sub. Alpha	cdhA	0607
2.3.1.169	acetyl-CoA decarbonylase	cdhC	*0863

Table S4. List of genes involved in solute transport in *A. hydrogeniformans* OS1.

Transporter Classification Family Number	**Transporter Classification Family Name**	**Gene Count**	**System Count**
TC:1.A.1	The Voltage-gated Ion Channel (VIC) Superfamily	1	1
TC:1.A.23	The Small Conductance Mechanosensitive Ion Channel (MscS) Family	1	1
TC:1.A.30	The H- or Na-translocating Bacterial Flagellar Motor/ExbBD Outer Membrane Transport Energizer (Mot/Exb) Superfamily	5	2
TC:1.A.33	The Cation Channel-forming Heat Shock Protein-70 (Hsp70) Family	2	1
TC:1.A.35	The CorA Metal Ion Transporter (MIT) Family	1	1
TC:1.B.17	The Outer Membrane Factor (OMF) Family	3	3
TC:1.B.22	The Outer Bacterial Membrane Secretin (Secretin) Family	1	1
TC:1.B.33	The Outer Membrane Protein Insertion Porin (Bam Complex) (OmpIP) Family	1	1
TC:1.B.42	The Outer Membrane Lipopolysaccharide Export Porin (LPS-EP) Family	2	2
TC:2.A.1	The Major Facilitator Superfamily (MFS)	5	5
TC:2.A.10	The 2-Keto-3-Deoxygluconate Transporter (KDGT) Family	2	2
TC:2.A.15	The Betaine/Carnitine/Choline Transporter (BCCT) Family	1	1
TC:2.A.21	The Solute:Sodium Symporter (SSS) Family	6	6
TC:2.A.22	The Neurotransmitter:Sodium Symporter (NSS) Family	4	2
TC:2.A.23	The Dicarboxylate/Amino Acid:Cation (Na or H) Symporter (DAACS) Family	4	4

TC:2.A.25	The Alanine or Glycine:Cation Symporter (AGCS) Family	4	4
TC:2.A.26	The Branched Chain Amino Acid:Cation Symporter (LIVCS) Family	2	1
TC:2.A.27	The Glutamate:Na Symporter (ESS) Family	1	1
TC:2.A.35	The NhaC Na H Antiporter (NhaC) Family	3	3
TC:2.A.38	The K+ Transporter (Trk) Family	2	2
TC:2.A.39	The Nucleobase:Cation Symporter-1 (NCS1) Family	1	1
TC:2.A.4	The Cation Diffusion Facilitator (CDF) Family	1	1
TC:2.A.40	The Nucleobase:Cation Symporter-2 (NCS2) Family	2	2
TC:2.A.44	The Formate-Nitrite Transporter (FNT) Family	1	1
TC:2.A.47	The Divalent Anion:Na Symporter (DASS) Family	1	1
TC:2.A.49	The Chloride Carrier/Channel (ClC) Family	1	1
TC:2.A.5	The Zinc (Zn2+)-Iron (Fe2+) Permease (ZIP) Family	1	1
TC:2.A.51	The Chromate Ion Transporter (CHR) Family	2	1
TC:2.A.55	The Metal Ion (Mn2+-iron) Transporter (Nramp) Family	1	1
TC:2.A.56	The Tripartite ATP-independent Periplasmic Transporter (TRAP-T) Family	23	10

Table S5. List of genes involved in energy storage in *A. hydrogeniformans* OS1.

Locus tag	**Gene Product Name**
0505	alpha-glucan phosphorylases
0506	glycogen/starch synthases, ADP-glucose type
0507	glucose-1-phosphate adenylyltransferase
0508	hypothetical protein

Table S6. List of genes involved in shell structure in *A. hydrogeniformans* OS1.

Locus tag	**Gene Product Name**
1070	ethanolamine utilization protein
1071	ethanolamine utilization protein
1072	carbon dioxide concentrating mechanism carboxysome shell protein
1073	ethanolamine utilization protein
1074	propanediol dehydratase large subunit
1075	propanediol dehydratase medium subunit
1076	propanediol dehydratase small subunit
1077	molecular chaperone
1078	hypothetical protein

1079	carbon dioxide concentrating mechanism carboxysome shell protein
1080	carbon dioxide concentrating mechanism carboxysome shell protein
1081	propanediol utilization protein
1082	ethanolamine utilization protein EutJ family protein
1083	hypothetical protein
1084	BMC domain protein
1085	carbon dioxide concentrating mechanism carboxysome shell protein
1086	ATP:cob(I)alamin adenosyltransferase
1087	NAD-dependent aldehyde dehydrogenases
1088	Predicted NADH:ubiquinone oxidoreductase, subunit RnfC
1089	carbon dioxide concentrating mechanism carboxysome shell protein
1090	Glycerol dehydrogenase and related enzymes

Table S7. List of genes involved in motility and chemotaxis in *A. hydrogeniformans.*

Che-family

Locus tag	**Gene name**	**Gene Product Name**
1255	*cheB*	CheB
0429	*cheC*	CheC
1155		CheC
1257		CheC-like
1258	*cheD*	CheD
1906	*cheR*	CheR
0537	*cheW*	CheW
0568		CheW
1256		CheW-like
1245	*cheY*	CheY
2123		CheY

Methyl-Accepting Chemotaxis Proteins

Locus tag	**Notes**
0541	Orphan – soluble
0646	Orphan - membrane

0972	Membrane
0981	
2083	
0378	

Table S8. List of genes involved in pili formation in *A. hydrogeniformans.*

Locus tag	**Gene name**	**Gene Product Name**
0512	*pulO/pilD*	PulO/PilD
0819	*pilM*	PilM
0828	*pilN*	PilN
0829	pilM	PilM
0831	pilY1	PilY1
0833	pulG	PulG
0834	pilV	PilV
0835	pulG	PulG
0837	pilT	PilT
0838	pulE/pilB	PulE/PilB

Table S9. List of RNA Polymerase and sigma factor genes in *A. hydrogeniformans.*

Locus tag	**Gene name**	**Gene Product Name**
0503		Anti-sigma regulatory factor (Ser/Thr protein kinase)
1260	*fliA*	RNA polymerase sigma factor, FliA/WhiG family
1275	*rpoD1*	RNA polymerase sigma factor, sigma-70 family
1399	*rpoD2*	RNA polymerase sigma factor, sigma-70 family
1490	*flgM*	flagellar biosynthesis anti-sigma factor FlgM
1433	*rpoB*	DNA-directed RNA polymerase, beta subunit
1434	*rpoC*	DNA-directed RNA polymerase, beta' subunit
1470	*rpoA*	DNA-directed RNA polymerase, alpha subunit
1774	*rpoD2*	RNA polymerase sigma factor, sigma-70 family
1869	*rpoZ*	DNA-directed RNA polymerase, omega subunit

Table S10. List of two-component signal transduction genes in *A. hydrogeniformans.*

Locus tag	**Gene Product Name**	**Notes**
0498	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase.	
0561	Histidine kinase./Histidine kinase-DNA gyrase B-, and HSP90-like ATPase.	with 0561 (cheY with HTH)
0748	Signal transduction histidine kinase	with 0749 (cheY with wing h wing)
0948	Signal transduction histidine kinase	with 0949 (cheY wing h wing) and 2 P ABC uptake genes
1256	Chemotaxis protein histidine kinase and related kinases	has CheW-like domain plus kinase and htp, etc in che cluster
2122	Signal transduction histidine kinase	with 2123 (CheY, atpase, DNA binding) on small contig

Table S11. Genes with archaeal gene/protein annotations.

Locus tag	**Gene Product Name**
0493	Factor F_{420}-reducing hydrogenase-like, alpha subunit
0494	Factor F_{420}-reducing hydrogenase-like, beta subunit
0644	methanyltetrahydromethanopterin cyclohdrolase
0645	formylmethanofuran dehydrogenase subunit C
0653	formylmethanofuran dehydrogenase subunit A
0654	formylmethanofuran dehydrogenase subunit B
0655	formylmethanofuran dehydrogenase subunit D
0803	formylmethanofuran dehydrogenase subunit E
0851	archaeal fructose 1,6-bisphosphatase
0639	archaeal phosphogycerate mutase
1168	archaeal phosphogycerate mutase
1540	MoaD family protein archaeal
2018	Coenzyme F_{390} synthatase
0647	tetrahydro methanopterin N-formyl transferase
1148	F_{420} gamma glutamyl ligase
2136	archaeal fructose 1,6-bisphosphatase
0651	archaeal flavoprotein
0767-00775	V/A type ATP synthase operon 1

1788-01795	V/A type ATP synthase operon 2
1897-01905	V/A type ATP synthase operon 3
1106	Wyosine [tRNA(Phe)-imidazoG37] synthetase
0269	Aldehyde: ferredoxin oxidoreductase
1030	Aldehyde: ferredoxin oxidoreductase
1044	Aldehyde: ferredoxin oxidoreductase
1674	Aldehyde: ferredoxin oxidoreductase

References

1. **Toso DB, Henstra AM, Gunsalus RP, Zhou ZH.** 2011. Structural, mass and elemental analyses of storage granules in methanogenic archaeal cells. Environ Microbiol **13:**2587-2599.

2. **Nei M, Saitou N.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution **4:**406-425.

3. **Felsenstein J.** 1985. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution **39:**783-791.

4. **Menes RJ, Muxi L.** 2002. *Anaerobaculum mobile* sp. nov., a novel anaerobic, moderately thermophilic, peptide-fermenting bacterium that uses crotonate as an electron acceptor, and emended description of the genus *Anaerobaculum*. Int J Syst Evol Microbiol **52:**157-164.

5. **Maune MW, Tanner RS.** 2012. Description of Anaerobaculum hydrogeniformans sp. nov., an anaerobe that produces hydrogen from glucose, and emended description of the genus Anaerobaculum. Int J Syst Evol Microbiol **62:**832-838.

6. **Rees GN, Patel BK, Grassia GS, Sheehy AJ.** 1997. *Anaerobaculum thermoterrenum* gen. nov., sp. nov., a novel, thermophilic bacterium which ferments citrate. Int J Syst Evol Microbiol **47:**150-154.

7. **Ben Hania W, Bouanane-Darenfed A, Cayol J-L, Ollivier B, Fardeau M-L.** 2016. Reclassification of Anaerobaculum mobile, Anaerobaculum thermoterrenum, Anaerobaculum hydrogeniformans as Acetomicrobium mobile comb. nov., Acetomicrobium thermoterrenum comb. nov. and Acetomicrobium hydrogeniformans

comb. nov., respectively, and emendation of the genus Acetomicrobium. International Journal of Systematic and Evolutionary Microbiology **66:**1506-1509.

8. **Thauer RK, Jungermann K, Decker K.** 1977. Energy conservation in chemotrophic anaerobic bacteria. Bacteriol Rev **41:**100-180.

9. **Buss KA, Ingram-Smith C, Ferry JG, Sanders DA, Hasson MS.** 1997. Crystallization of acetate kinase from Methanosarcina thermophila and prediction of its fold [In Process Citation]. Protein Sci **6:**2659-2662.

10. **Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics **25:**2271-2278.

11. **Garcia-Vallvé S, Romeu A, Palau J. 2000**. **Horizontal Gene Transfer in Bacterial and Archaeal Complete Genomes. Genome Research 10:1719-1725**.

12. **Nelson KE.** 1999. **Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. Nature 27:323-329**

13. **Deppenmeier U, Johann A, Hartsch T, Merkl R, Schmitz RA, Martinez-Arias R, Henne A, Wiezer A, Baumer S, Jacobi C, Bruggemann H, Lienard T, Christmann A, Bomeke M, Steckel S, Bhattacharyya A, Lykidis A, Overbeek R, Klenk HP, Gunsalus RP, Fritz HJ, Gottschalk G.** 2002. The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. J Mol Microbiol Biotechnol **4:**453-461.

14. **Liang R, Grizzle RS, Duncan KE, McInerney MJ, Suflita JM.** 2014. Roles of thermophilic thiosulfate-reducing bacteria and methanogenic archaea in the biocorrosion of oil pipelines. Front Microbiol **5:**89.

15. **Chen S, Liu X, Dong X.** 2005. Syntrophobacter sulfatireducens sp. nov., a novel syntrophic, propionate-oxidizing bacterium isolated from UASB reactors. Int J Syst Evol Microbiol **55:**1319-1324.

16. **Horz H-P, Citron DM, Warren YA, Goldstein EJC, Conrads G.** 2006. Synergistes Group Organisms of Human Origin. Journal of Clinical Microbiology **44:**2914-2920.

17. **Vartoukian SR, Palmer RM, Wade WG.** 2007. The division “Synergistes”. Anaerobe **13:**99-106.

18. **Sieber JR, McInerney MJ, Gunsalus RP.** 2012. Genomic Insights into Syntrophy: The Paradigm for Anaerobic Metabolic Cooperation. Annual Review of Microbiology **66:**429-452.

19. **Ferry JF.** 2015. Acetate Metabolism in Anaerobes from the Domain Archaea. Life **5:**1454-1471.

20. **Pagani I LK, Jansson J, Chen IM, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC. 2012**. **The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata**. **Nucleic Acids Res, :D571-D579**.

21. **Anonymous.** DOE Joint Genome Institute. http://www.jgi.doe.gov. Accessed

22. **Borodovsky M, Mills R, Besemer J, Lomsadze A.** 2003. Prokaryotic gene prediction using **GeneMark and GeneMark.hmm.** , **Curr Protoc Bioinformatics**

23. **Delcher AL, Harmon D, Kasif S, White O, Salzberg SL.** 1999. **Improved microbial gene identification with GLIMMER. Nucleic Acids Res, 27:4636-4641**.

24. **Finn R, Tate J, Mistry J, Coggill P, Sammut S, Hotz HR, Ceric G, Forslund K, Eddy S, Sonnhammer E, Bateman A.** 2008. The Pfam protein families database. Nucleic Acids Res **36:**281-288.

25. **Lowe TM, Eddy SR. 1997**. **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1:955-964**.

26. **Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research **35:**3100-3108.

27. **Griffiths-Jones S, Moxon R, Marshall M, Khanna A, Eddy SR, Bateman, A.** 2005. **Rfam: annotating non-coding RNAs in complete genomes**. **Nucleic Acids Res, 33:D121-124**.

28. **Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M.** 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res,.

29. **Quevillon E, Silventoinen V, Pillai S, Harte N, N. M, R. A, Lopez R.** 2005. InterProScan: protein domains identifier. Nucleic Acids Res, doi: Jul 1;33(Web Server issue):W25-8.

30. **Hiramatsu T, Kodama K, Kuroda T, Mizushima T, Tsuchiya T.** 1998. A putative multisubunit Na+/H+ antiporter from Staphylococcus aureus. Journal of bacteriology **180:**6642-6648.

31. **Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF.** 2016. The physiology and habitat of the last universal common ancestor. Nature Microbiology **1:**16116.

32. **Fuchs G.** 2011. Alternative Pathways of Carbon Dioxide Fixation: Insights into the Early Evolution of Life? Annual Review of Microbiology **65:**631-658.

33. **Borrel G, Adam PS, Gribaldo S.** 2016. Methanogenesis and the Wood–Ljungdahl Pathway: An Ancient, Versatile, and Fragile Association. Genome Biology and Evolution **8:**1706-1711.

34. **Gibson MI, Chen PY-T, Johnson AC, Pierce E, Can M, Ragsdale SW, Drennan CL.** 2016. One-carbon chemistry of oxalate oxidoreductase captured by X-ray crystallography. Proceedings of the National Academy of Sciences **113:**320.

35. **Gibson MI, Brignole EJ, Pierce E, Can M, Ragsdale SW, Drennan CL.** 2015. The Structure of an Oxalate Oxidoreductase Provides Insight into Microbial 2-Oxoacid Metabolism. Biochemistry **54:**4112-4120.

36. **Pierce E, Becker DF, Ragsdale SW.** 2010. Identification and Characterization of Oxalate Oxidoreductase, a Novel Thiamine Pyrophosphate-dependent 2-Oxoacid Oxidoreductase That Enables Anaerobic Growth on Oxalate. Journal of Biological Chemistry **285:**40515-40524.

37. **Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, Ashburner M, Axelrod N, Baldauf S, Ballard S, Boore J, Cochrane G, Cole J, Dawyndt P, De Vos P, DePamphilis C, Edwards R, Faruque N, Feldman R, Gilbert J, Gilna P, Glockner FO, Goldstein P, Guralnick R, Haft D,**

Hancock D, Hermjakob H, Hertz-Fowler C, Hugenholtz P, Joint I, Kagan L, Kane M, Kennedy J, Kowalchuk G, Kottmann R, Kolker E, Kravitz S, Kyrpides N, Leebens-Mack J, Lewis SE, Li K, Lister AL, Lord P, Maltsev N, Markowitz V, Martiny J, et al. 2008. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol **26:**541-547.

Chapter 5. *Acetomicrobium hydrogeniformans* OS1 uses an archaeal-like glycolysis pathway to generate elevated hydrogen levels approaching theoretical limits

Contents

Figures

been biochemically or structurally characterized. Tree was generated by MEGA6 (1). Genes utilized for the tree's creation are listed in table S1.

Tables

Table 1.1-1.4. The *A. hydrogeniformans* gene/protein candidates for glucose fermentation by the Embden-Meyerhof-Parnas (EMP) pathway, pentose phosphate (PP) pathway and Entner-Doudoroff (ED) pathway. Gene/protein designations are according to EcoCyc/Metacyc nomemclature (2).

Table 2. Activity of glucose pathway enzymes in *A. hydrogeniformans OS1*. Enzyme activity is expressed in µmol substrate utilized per min per mg protein. Preparation of cell extracts and enzyme assays are described in Materials and Methods. BDL, below detection limit; MV, methyl viologen; BV, benzyl viologen. Standard deviations are given when activity was measured with cell-free extracts prepared from different cultures.

Table 3. Hydrogenase, formate dehydrogenase, and formate hydrogenlyase activities in *A. hydrogeniformans OS1* broken cells. Activity is expressed in umol substrate utilized per min per mg protein. Individual assays were performed in *A. hydrogeniformans* broken cells and in the particulate and soluble cell fractions. The BV assays and preparation of cell extracts are described in Materials and Methods, and preparation of cell extracts are described in Materials and Methods. Abbreviations: BDL, below detection limit; MV, methyl viologen; BV, benzyl viologen.

Table 4. Gene/protein candidates for hydrogenases and dehydrogenases in *A. hydrogeniformans.* Hydrogenases include a membrane-bound [NiFe]-type hydrogenase in conjunction with a formate dehydrogenase, forming a formate hydrogenlyase complex, and a soluble [FeFe] type hydrogenase. Gene/proteins for unspecified dehydrogenase subunits with homology to NADH-ubiquinone oxidoreductase are also included. Dehydrogenases include the formate dehydrogenase in the aforementioned putative formate hydrogenlyase complex; two predicted formate dehydrogenases, and a predicted unspecified dehydrogenase with homology to NADH-ubiquinone oxidoreductase.

Table 5. Putative electron-carrying proteins containing a 2Fe-2S and/or 4Fe-4S type cluster. This includes various doxins and iron-sulfur-containing proteins without another evident enzymatic classification less than 600 amino acids. Includes a sum of the total cell protein dedicated to these putative electron carriers under each growth condition.

Table 6. Properties and abundance of the *A. hydrogeniformans* AFOR-family proteins (GAPOR family). *M. maripaludis* GAPOR sequence from strain C6 genome (IMG gene ID 641284211) and *P. furiosus* GAPOR sequence from strain DSM 3638 (IMG gene ID 638172975) as recorded at the Integrated Microbial Genomes database (3). Percent identity matrix generated by Clustal2.1 (4).

Table 7. Properties and abundances of the *A. hydrogeniformans* OFOR-family proteins.

Supplementary materials

Table S1. The protein sequences used to generate the phylogenetic tree in figure 4 are listed. (accession numbers):

1. Abstract

Acetomicrobium hydrogeniformans OS1 is an obligate anaerobic bacterial species of the phylum *Synergistetes* that was previously isolated from an oil production water stream and ferments sugars, amino acids, and dicarboxylic acids. During metabolism of glucose it generates an unusually high molar ratio of hydrogen, thus suggesting an undescribed metabolic ability relative to other hydrogen-forming bacteria. Here, the genomic, proteomic, and enzymatic basis of glucose fermentation in *A. hydrogeniformans* was examined. A modified Embden-Meyerhoff pathway was revealed that employs a glyceraldehyde-3-phosphate oxidoreductase enzyme (GAPOR) in place of glyceraldehyde-3-phosphate dehydrogenase (GapA) and glycerate kinase (Pkg). *A. hydrogeniformans* cell extracts exhibited GAPOR as well as pyruvate-ferredoxin oxidoreductase (PFOR) activity in place of GapA, Pgk, pyruvate dehydrogenase (Pdh), or pyruvate formate lyase (Pfl) activity that would generate acetyl-CoA and NADH. Bacterial-type phosphotransacetylase and acetate kinase enzymes then generate acetate and ATP by substrate level phosphorylation. Inspection of the *A. hydrogeniformans* genome revealed several potential hydrogenase gene clusters that would account for hydrogen formation. Proteomic and enzyme studies revealed that electrons derived from GAPOR and PFOR drive hydrogen formation by a soluble FeFe-type hydrogenase. These *A. hydrogeniformans* findings demonstrate the presence of an unconventional C_6 metabolism that would explain the high hydrogen production. This is the first demonstrated example of GAPOR in the Bacteria where this enzyme is typically found

only in thermophilic archaea. Bioinformatics comparisons to related *Acetomicrobium* strains plus other genera of the phylum *Synergistetes* suggest this modified pathway is also present in other *Synergistetes* species.

2. Introduction

Acetomicroibum hydrogeniformans strain OS1, previously known as *Anaerobaculum hydrogeniformans,* is a Gram-negative, non-motile, rod-shaped obligate anaerobe of the phylum *Synergistetes* (5, 6). Isolated from a moderately thermophilic Alaskan oil pipeline water production stream, it is able to ferment a variety of substrates including glucose and certain other sugars, amino acids, and dicarboxylic acids (5). Other species of the genus *Acetomicrobium* exhibit similar temperature and pH growth ranges although their source of isolation and substrate utilization profiles differ (6).

Synergistetes species occur in relatively low numbers across diverse microbial communities. Members of the phylum have been isolated from the termite hindgut, subgingival plaque, intestinal tracks of warm blooded animals, industrial wastewater, and petroleum reservoirs (7). As of May 2019 the following sixteen genera have been formally designated: *Acetomicroibum*, *Aminiphilus*, *Aminivibrio*, *Aminobacterium*, *Aminomonas*, *Aminiphilus*, *Cloacibacillus*, *Dethiosulfovibrio*, *Fretibacterium*, *Jonquetella*, *Lactivibrio*, *Pacaella*, *Pyramidobacter*, *Rarimicrobium*, *Thermanaerovibrio*, and *Thermovirga* (8-11). These genera are relatively unstudied at the either the genomic or metabolic levels and are distantly related to the *Fusobacteria* and *Firmicutes*.

Previous studies by Maune and coworkers revealed that *A. hydrogeniformans* produces H_2 gas at a yield approaching the theoretical maximum for glucose fermentation (historically also called ‘dark’ fermentation as it proceeds in absence of light) with a yield of almost four molecules of H_2 produced per six-carbon sugar molecule consumed (5, 12). The metabolic basis of this ability was unknown. *A. hydrogeniformans* is an attractive model to examine hydrogen production due to its potential applications in microbial-produced alternative fuels from low cost feed-stocks and bio-wastes (5). The high mole ratio of hydrogen from glucose and high partial pressures of hydrogen reached during fermentation (>10%) demonstrates the potential of OS1 for biohydrogen production and syntrophy (5, 13). H_2 is a clean-burning fuel that can be made from renewable sources and its combustion does not produce greenhouse gasses. Acetate, which is co-produced with H_2 from glucose, can be converted to methane, another energy-rich biofuel made by acetotrophic methanogens (14).

The genome sequence of *A. hydrogeniformans* reveals a relatively small genome (2.12 MB) with 2,068 protein coding genes (15). To address the underlying cellular machinery employed during glucose fermentation and hydrogen formation, we first inventoried the genome sequence information and evaluated several alternative metabolic routes. The pathway enzymes used by *A. hydrogeniformans,* fermentation were then evaluated assayed in using broken cell extracts. These data were complemented by proteomic studies to identify the individual proteins involved in C_6 sugar metabolism. The resulting findings reveal an unconventional pathway for fermentation of glucose to acetate and hydrogen production that has not previously been identified in the bacterial kingdom.

3. Methods

Cell growth. *A. hydrogeniformans* strain OS1 (DSM 22491 = ATCC BAA-1850) is the type strain for *A. hydrogeniformans* (5). *A. hydrogeniformans* was cultured anaerobically as previously described using a medium containing per liter: 2 g yeast extract, 10 g NaCl, 10 g TES buffer, 10 ml vitamin solution, 10 ml trace element solution and 10 ml OS1 mineral base (5). The medium was prepared anaerobically with a N_2:CO_2 gas mixture (80:20 v/v), adjusted to pH 7.5, and dispensed into anaerobic serum bottles or tubes prior to sterilization. Before inoculation, a sterile solution of reducing agent (1% v/v final concentration) of a 2.5% mixture of Na_2S/Cys•HCl was added aseptically along with a 5% v/v final volume of sterile glucose solution (4g/50 ml H_2O) (5). Alternatively, filter sterilized pyruvate was substituted for glycose using 0.5 ml stock solution (3g pyruvate /50 ml H_2O). Cultures were incubated at 55°C as previously reported. For enzyme studies, cells were grown in the above medium.

For whole cell proteomic studies, *A. hydrogeniformans* was grown in the above medium to mid-log with three successive serial transfers prior to cell harvest. Serum culture bottles were chilled on ice, cells removed and harvested by centrifugation at 7,000 rpm for 20 minutes at 4°C in an IEC bench top centrifuge. Cells pellets were re-suspended in ALS buffer [4% ammonium dodecyl sulfate, 0.1% sodium deoxycholate, 100mM ammonium bicarbonate, 5mM tris(2-carboxyethyl) phosphine (TCEP)] and stored at -70°C until processed for proteomic analysis. Three biological replicates with three technical replicates each were used for all MS data shown in the Results section. In other proteomic studies, cells were grown in *A. hydrogeniformans* standard medium to 33%, 50% and 100% glucose depletion.

Enzyme assays. Enzyme activity was calculated from the slope of rates of absorbance change plotted against at least three different protein concentrations. Means and standard deviations are given when activity was measured with cell-free extracts prepared from different cultures.

Mass Spectrometry. Cell pellets suspended in ALS solution [4% ammonium dodecyl sulfate, 0.1% sodium deoxycholate, 100mM ammonium bicarbonate, 5mM tris(2-carboxyethyl) phosphine (TCEP)] were thawed and incubated at 90°C for 10 minutes. The resulting lysates were subjected to enhanced filter-aided sample preparation (eFASP) (16) for buffer exchange, alkylation and overnight digestion at 37°C with sequencing grade trypsin (Thermo Scientific). The peptide-containing filtrates were acidified and extracted with ethyl acetate to remove deoxycholic acid. The peptide-containing aqueous phase was dried in a vacuum concentrator. Digested peptides were reconstituted in 0.5% acetic acid, desalted with STAGE tips assembled from 3M Empore C18 Solid Phase Extraction Disks (17)], and dried again. Peptides were reconstituted in LC-MS injection buffer (3% acetonitrile, 0.1% formic acid) and quantified by Pierce Quantitative Fluorometric Peptide Assay (Thermo Scientific, Product #23290).

Processed whole cell samples were analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) on an EASY nLC1000 (Thermo Scientific) coupled to a quadrupole orbitrap mass spectrometer (Q-Exactive, Thermo Scientific). Peptides (300 ng) were loaded onto a Acclaim PepMap100 C18 trap column (Thermo Scientific, 75µm x 2cm, 100 Å). Peptides were separated on an Acclaim PepMap RSLC C18 analytical column (Thermo Scientific, 75µm x 25µm, 100Å). The LC buffers were buffer A (0.1% formic acid in water) and buffer B (0.1% formic acid in acetonitrile). Peptides were eluted at 300nL/min using the following gradient of buffer B: 3-8% for 2 min; 8-20% for 48 min; 20-30% for 22 min; 30-50% for 5 min; 50-80% for 2 min; 80% for 10 min.

The mass spectrometer was operated in data-dependent acquisition mode with a 400-1800 *m/z* MS scan acquired at 70,000 x resolution with an AGC target of 1×10^6 and a maximum fill time of 100 msec followed by MS/MS acquisitions on the most abundant precursor ions (up to 10). MS/MS scans were acquired with a dynamic mass range at a resolution of 17,500 and with an AGC target of 1×10^5 and a maximum fill time of 80 msec. Precursor ions were isolated (+/- 2 *m/z*) and activated in the HCD trap with normalized collision energy of 30 and an intensity threshold of 2.5×10^4. Dynamic exclusion was set to 30s with “exclude isotopes” selected.

Peak lists generated from the raw data files by Proteome Discoverer version 2.2 were submitted to Mascot server version 2.5 (Matrix Science) and searched against the Integrated Microbial Genomes and Microbiomes (IMG/M) complete proteome for *A. hydrogeniformans* strain OS1 (Genome ID: 2517487012, released August 6, 2015) supplemented with protein sequences of common contaminants. Trypsin/P was specified as cleavage enzyme with up to 2 missed cleavages considered, precursor mass tolerance set to ±10 ppm and product mass error set to 0.02 Da. Cysteine carboxyamidomethylation and methionine oxidation were set as variable modifications. False discovery rate of the target decoy PSM validator was set to 1% (strict) and 5% (relaxed) using peptide spectrum matches with a maximum delta Cn of 0.05. Protein quantification was based on the average peak area from the extracted-ion chromatogram (XIC) for at most 3 intense tryptic peptides per protein. Normalization across all biological and technical replicates was performed based on the total peptide amount per replicate. 3 biological replicates, each with 3 technical replicates were used for quantification.

Mass spectrometry was performed on two/three biological samples for each growth condition and the results averaged. Error bars represent standard deviation. Protein abundance is expressed as fmol protein detected for each protein detected, and in percent of total *A. hydrogeniformans* cell proteins detected per sample run. The level of detected *A. hydrogeniformans* proteins were expressed in fmol peptides detected per sample run and/or as percent relative to total cell protein.

Bioinformatics. Genes of interest in the *A. hydrogeniformans strain OS1* genome (GenBank accession no. NZ_ACJX00000000.3) were identified using the IMG-JGI genome/gene query platform. Search criteria included by gene and gene product name, by pathway association, by protein family (18), by orthologous protein blast search using the *Escherichia coli* database (19), MetaCyc database, (2). GAPOR searches were based primarily on the *Pyrococcus furiosus* genome information stored in the Integrated Microbial Genomics database (IMG) (3). Two to three criteria were generally needed to thoroughly survey the *A. hydrogeniformans* genome sequence given that the genes had been machine annotated by the GWU annotation pipeline and by the IMG-JGI pipeline (15). Search redundancy was generally used as a measure of thoroughness. Using the manually curated genome results, a reconstruction of *A. hydrogeniformans* core metabolism was performed for the putative glucose utilization pathway to the end products acetate, CO_2 and H_2. The provisional metabolic reconstruction (Figure 2) is further described below in the Results Section. To visualize proteomic data, protein abundance (in fmol detected or as % of total cell proteins) data for glucose utilization proteins as well as for hydrogen formation were labeled by an abbreviated locus tag nomenclature or by the gene/protein name. The *A. hydrogeniformans* locus tags (GenBank accession no.

NZ_ACJX00000000.3) were abbreviated by deleting part of the longer designation prefix by OS1_ XXXX (e.g., GCWU000281_00882 to OS1_0882).

AFOR Protein Family Phylogenetic Tree. Protein sequences for the four *A. hydrogeniformans*-associated AFOR genes (OS1_0269, OS1_1030, OS1_1044, and OS1_1674) were retrieved from the genome stored at the Integrated Microbial Genomics database (3). Archaeal protein sequences for characterized AFORs were obtained from the NCBI genome browser. Tree was modeled after tree in the *Pyrobaculum aerophilum* GAPOR isolation paper by Reher and colleagues (20). The protein sequences used to generate the phylogenetic tree are listed in Table S1. The following method was automatically generated by the MEGA6 program: The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (21). The tree with the highest log likelihood (-24605.9223) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 42 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 425 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (1).

4. Results

Metabolic reconstruction of central metabolism in A. hydrogeniformans strain OS1. As a framework to investigate glucose utilization and hydrogen formation by *A. hydrogeniformans* we first performed a metabolic reconstruction based on genome sequence information (15). The published machine-annotated sequence was first queried for orthologous sugar metabolizing genes/proteins of the Embden-Meyerhoff-Parnas pathway (EMP), the pentose phosphate pathway (PP), and the Entner-Doudoroff pathway (ED) based on the MetaCyc pathway database (2, 22), Materials and Methods). Each pathway appeared to be incomplete for one or more enzymes (Table 1). The unaccounted for enzymes were: EMP pathway, hexokinase (Glk); PP pathway, G-6-P-dehydrogenase (Zwf), NADH-glucose-6-phosphate-1-hydrogenase (Zwf), 6-P-gluconolactonase (Pgl), 6-P-gluconate dehydrogenase (Gnd) and ribose-5-P isomerase A (RpiA); ED pathway, hexokinase (Glk/HK) G-6-P-dehydrogenase, NADH-glucose-6-phosphate-1-hydrogenase (Zwf) and glycerate kinase (GlxK/GarK1).

Since the EMP pathway appeared to be nearly intact in *A. hydrogeniformans*, we then searched for alternative enzymes that would account for the overall conversion of glucose to pyruvate. As no glucose-phosphorylating enzyme that would initiate the first step of sugar breakdown (i.e., hexokinase, Glk) was identified, glucose may enter the cell via a PTS-type uptake system (possibly encoded by OS1_0546, _0755, _0756, _1851; Table 1), and/or, alternatively, by an unidentified ABC-type or ion driven glucose uptake system working with an undescribed/unidentified hexose kinase. Also of note, several gene loci encoding an archaeal-type glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR) were identified (Figure 1A). These would replace the conventional EMP pathway enzymes G-3-P dehydrogenase (GAPDH, or GapA) and phosphoglycerate kinase (Pkg). This GAPOR-type enzyme reaction was also incorporated into a provisional *A. hydrogeniformans* C6 metabolic reconstruction (Figure 2, discussed below). Interestingly, the genome pathway search also revealed paralogous gene/protein candidates for several other fermentation pathway reactions. These included four 6-phosphofrutose kinase (Pfk) candidates, three fructose bisphosphate aldolase (Fba) candidates and eight phosphoglycerate mutase (Gpm) candidates. It was unknown which of the above candidate pathway proteins were synthesized by *A. hydrogeniformans.*

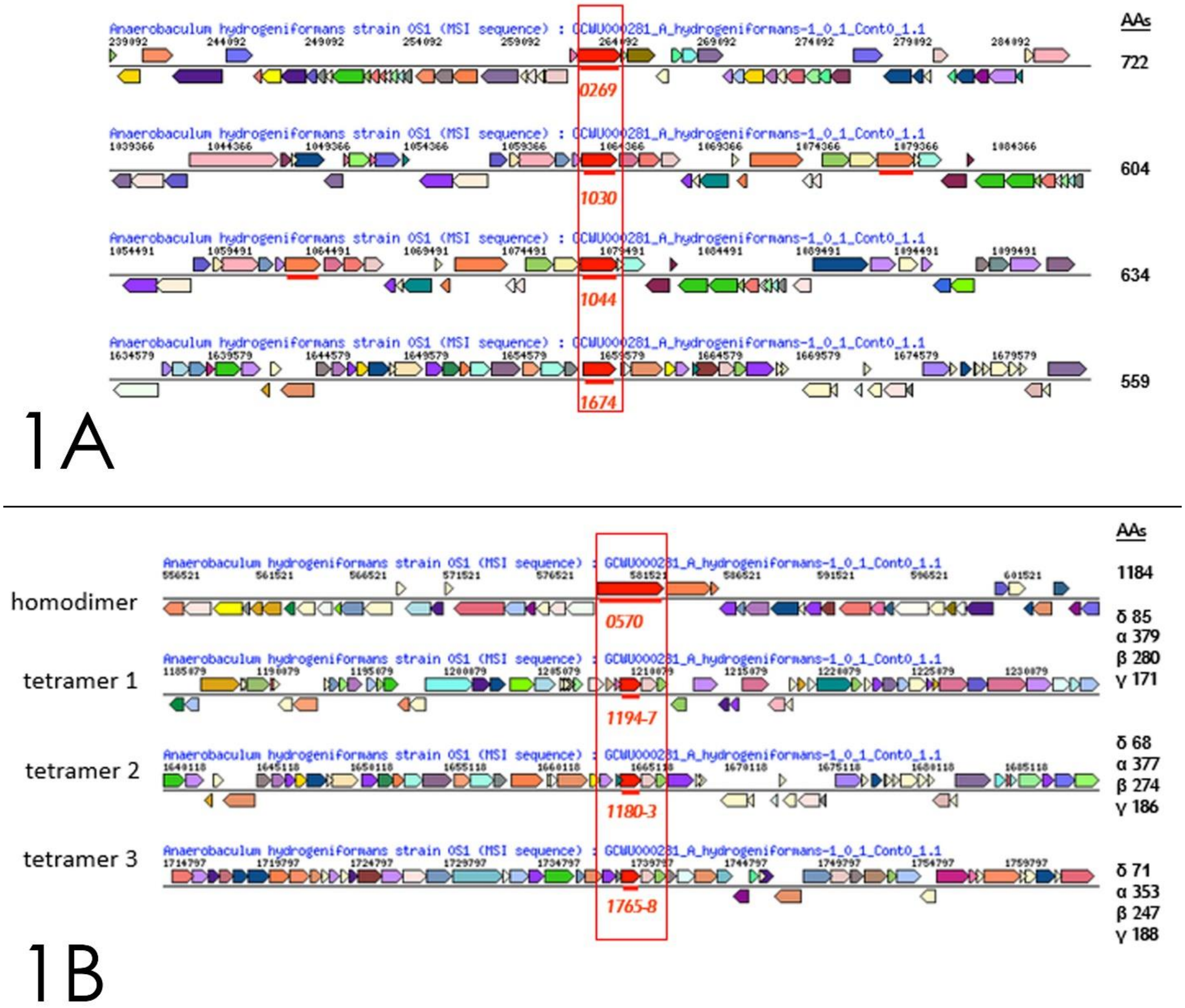

Figure 1. Gene neighborhood of GAPOR- and PFOR-like enzymes present in the *A. hydrogeniformans* genome. A) The four paralogous GAPOR enzymes (*go*r) are aligned relative to their 5' ends. The nearby gene context is also shown. B) The four PFOR-like enzymes. These include three four-gene *por* clusters plus a single large *por* gene aligned relative to the 5' ends along with nearby genes for context and comparison.

Following the formation of pyruvate from glucose, *A. hydrogeniformans* then generates two moles each of acetate and CO_2 plus hydrogen gas as the fermentation pathway end-products (5). The genome inventory search revealed genes for a putative pyruvate dehydrogenase (Pdh) but no genes were identified for either a pyruvate formate lyase (Pfl) or pyruvate decarboxylase (Pdc) that would generate acetyl-CoA from pyruvate. However, four sets of genes for an alternative pyruvate-consuming enzyme, pyruvate:ferredoxin oxidoreductase (PFOR) were identified (Figure 1B, Table 1). Three of the candidates are encoded by individual sets of four adjacent genes, while the fourth candidate is encoded by a single large gene (discussed below). The acetyl-CoA then generated could be converted to acetate by *A. hydrogeniformans* by conventional bacterial phosphotransacetylase (Pta) and acetate kinase (Ack) enzymes, of which three candidates of each were identified in the genome (Table 1.1). No genes for ADP-forming acetyl-CoA synthetase or AMP-forming acetyl-CoA ligase (Acs) were found (23).

Assays for glucose pathway enzymes in A. hydrogeniformans strain OS1. To establish which of the proposed EMP pathway enzymes were present in *A. hydrogeniformans,* we prepared and assayed cell extracts for them plus for several alternative pathway enzymes (Materials and Methods). Activity was detected for all the EMP pathway enzymes (table 1.1, table 2) except for glyceraldehyde-3-phosphate dehydrogenase (Gdh) while phosphoglycerate mutase (Gpm) was not examined. Cell extracts also exhibited high glyceraldehyde-3-phosphate:ferredoxin oxidoreductase (GAPOR) activity that would replace the two of the EMP pathway enzymes, glyceraldehyde-3-P dehydrogenase (GAPDH) and 2-phosphoglycerate kinase (Pgk). Noteworthy, activity of phosphoglycerate kinase was low. We modified the glycolysis pathway reconstruction in *A. hydrogeniformans* to use a ferredoxin-reducing enzyme in place of the NADH-generating enzymes. GAPOR activity has previously only found in thermophilic archaea.

Pyruvate, the end-product of the EMP pathway is converted to acetyl-CoA and finally to acetate by three additional reactions as proposed by the genomic analysis. Pyruvate:ferredoxin oxidoreductase (PFOR) activity was abundant in cell extracts whereas neither pyruvate dehydrogenase (Pdh) nor a pyruvate formate lyase (Pfl) activity were detected that form acetyl-CoA. Then, two following reactions then form acetate and CO_2. Both phospho-transacetylase (Pta) and acetate kinase (Ack) were detected (Table 1). Here, ATP generation in *A. hydrogeniformans* occurs by substrate level phosphorylation via a conventional bacterial acetate kinase reaction. These combined genomic and enzyme studies support the presence of a modified fermentation pathway in *A. hydrogeniformans* whereby an archaeal-like GAPOR enzyme is employed to couple carbon oxidation with ferredoxin reduction and where pyruvate oxidation is accomplished by a bacterial like PFOR enzyme (Figure 2, discussed below).

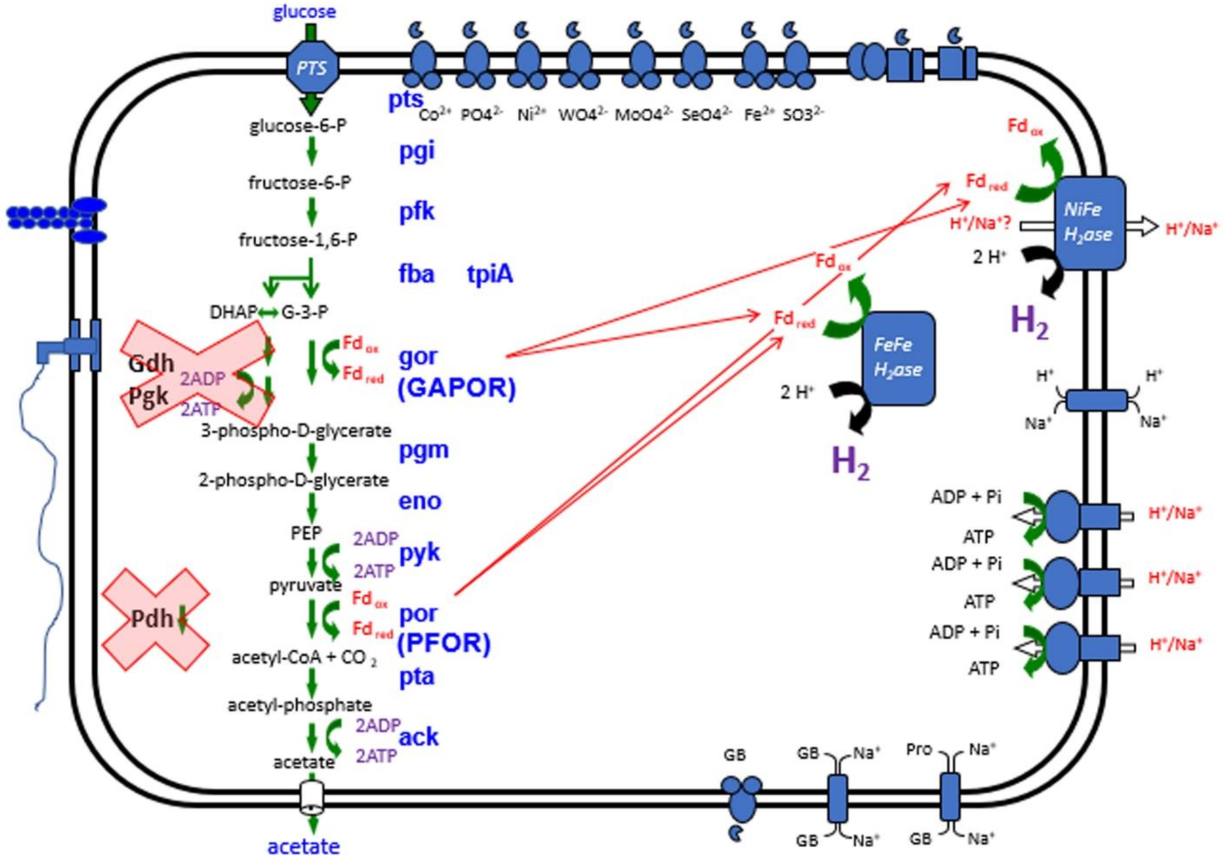

Figure 2. Metabolic reconstruction of glucose fermentation and hydrogen formation by *A. hydrogeniformans.* The indicated gene/enzyme abbreviations are according to the EcoCyc and MetaCyc nomenclatures. PTS, phosphotransferase system; PGI, phosphoglycerate isomerase; PFK, phosphofructokinase; FBA, fructose bisphosphate aldolase; TPI, triose-phosphate isomerase; Gdh,glyceraldehyde-3-phosphate hydrogenase (GAPDH); gor, glyceraldehyde-3-phosphate oxidoreductase (GAPOR); Pdh, pyruvate dehydrogenase; pgm, phosphoglycerate mutase; eno, enolase; pyk, pyruvate kinase; por, pyruvate:ferredoxin oxidoreductase (PFOR); pta, phosphotransacetylase; ack, acetate kinase; ferredoxin, Fd.

Proteomic analysis of the glucose fermentation pathway. To determine which of the multiple protein candidates were employed in the proposed glucose fermentation pathway, we performed quantitative proteomics on cells grown on glucose medium (Materials and Methods). The results are summarized in Table 1.1 and Figure 2. Where a single gene/protein candidate was predicted, each protein was detected and was typically abundant at a level of 0.07-0.3% of the total cell protein. For example, glucose-6-phosphate isomerase (Pgi, OS1_0293), triose-phosphate isomerase (TpiA, OS1_1934), enolase (Eno, OS1_1402) and pyruvate kinase (Pyk, OS1_1621) were 0.2%, 0.3%, 0.08%, and 0.07%, respectively. Of the four 6-phosphofructokinase candidates, only one was abundant (Pfk, OS1_1537). Likewise, for the fructose-biphosphate aldolase paralog (Fba, OS1_1515): it was three times more abundant than for OS1_0322. Of the four identified GAPOR candidates to convert glyceraldehyde-3-phosphate to 3-phospho-D-glycerate, only one (GAPOR1, OS1_0269) was present in significant amounts (Table 1.1). The other three were either synthesized in low abundance (OS1_1030, OS1_1044) or not detected (OS1_1674). Here, OS1_1030 was about one eighth while OS1_0269 was ~1/30th of the OS1_1044 protein level. The sole protein candidate for glyceraldehyde-3-phosphate dehydrogenase candidate (GapA, OS1_1932) was also detected in high amounts although but strikingly, no GapA enzyme activity was detected in cell extracts (Table 2). Lastly, of the eight paralogous phosphoglycerate mutase (Pgm) candidates, two were abundant (OS1_0109 and OS1_0639) whereas all others were low or not detected.

By the metabolic reconstruction, *A. hydrogeniformans* then converts pyruvate to acetyl-CoA and CO_2. Of the four PFOR candidates identified (Figure 1B). Only PFOR1 protein (OS1_0570) was detected in high abundance. Proteins for all other polypeptides candidates comprising tetrameric PFORs, named here as PFOR 2 (OS1_1194-97), PFOR3 (OS1_1680-83) and PFOR 4 (OS1_1765-68), were in low abundance or not detected (Table 1). In the final reactions leading to acetate formation, one phospho-transacetylase (Pta1, OS1_0712) and two acetate kinase candidates (Ack1, OS1_0711 and Ack3, 1570) were abundantly synthesized. These combined bioinformatic, enzyme, and protein data support a proposed model for glucose fermentation whereby *A. hydrogeniformans* employs a GAPOR enzyme in place of the conventional bacterial pathway enzymes, glyceraldehyde-3-phosphate dehydrogenase and phosphoglycerate kinase to generate pyruvate. Unresolved is a need for the cell to abundantly synthesize the GapA and Pgk proteins.

We also performed a proteomic analysis of *A. hydrogeniformans* cells grown with pyruvate in place of glucose (Materials and Methods). Under these alternative growth conditions, cells would require hexose sugar synthesis for cell biosynthetic needs. Interestingly, levels of most all of the fermentation pathway proteins (Table 1) remained constant versus whether cells were grown on pyruvate versus glucose conditions. A similar proteome pattern was also seen when cells were grown with glucose and pyruvate present (Table 1). These results suggest that fermentation pathway enzyme synthesis in *A. hydrogeniformans* is not differentially regulated.

Electron and proton flow to hydrogen in A. hydrogeniformans. We have established a glucose fermentation model in which electrons are captured at two key steps by ferredoxin. We now search for ways those electrons could be used in the production of dihydrogen.

The genome inventory of *A. hydrogeniformans* genes revealed several hydrogenase-like gene clusters that would provide for H_2 formation (Figure 3). One composed of 12-14 genes, would encode a membrane associated Ni-Fe-type hydrogenase (OS1_0245 to _0257), designated here as Hyd1, which may be associated with a formate dehydrogenase to form a formate hydrogenlyase complex. A second cluster of 14 genes (Figure 3B) would encode a soluble Fe-Fe only hydrogenase (OS1_0492 to _0504), designated Hyd2. When we assayed broken-cell suspensions of *A. hydrogeniformans* for hydrogenase activity using either methyl viologen (MV) or benzyl viologen (BV) as the mediator dye): activity was detected with each substrate (Table 3). In contrast, no activity was detected using either NADH:MV or NADPH:MV to measure oxidoreductase showed low activity. These results suggest that a confrucating-type hydrogenase enzyme was not used for hydrogen formation in *A. hydrogeniformans*. A role for reduced ferredoxin rather than NADH (NADPH) or Fd:NADH (NADPH) as the electron donor(s) is therefore established where the hydrogenase-dependent hydrogen production would regenerate oxidized Fd needed in subsequent rounds of glucose oxidation by the Fd-dependent GAPOR and PFOR enzymes.

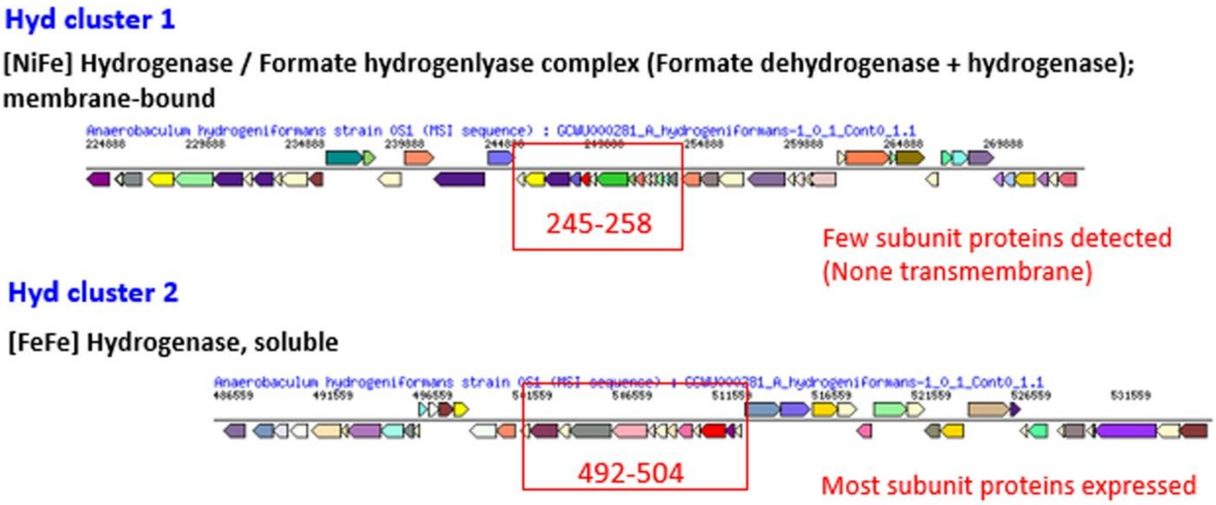

Figure 3. Gene neighborhoods of the NiFe-type and the FeFe only-type hydrogenase gene clusters present in the *A. hydrogeniformans* genome. A) NiFe-type hydrogenase gene cluster for Hyd1, and B) FeFe-only type hydrogenase gene cluster for Hyd2. Other genes upstream and downstream are shown in their orientation and scale relative to the hydogenase genes listed in Table 3.

Inspection of the proteomic data for *A. hydrogeniformans* cells grown with glucose revealed a subset of the *A. hydrogeniformans* Hyd2 (FeFe, soluble) hydrogenase polypeptides (OS1_0493, 0494, 0495, 0496, 0497, 0499) that were expressed at levels. These were approximately ten-fold higher on average than those associated with Hyd1 (NiFe, membrane-associated) (OS1_0246, 0247, 0248, 0249) (Table 4). Strikingly, cells grown with pyruvate in place of glucose exhibited a 2 to 7-fold reduction in the OS1_0493, 0494, 0495, 0496, 0497, 0499 polypeptide levels.

When cells were cultured on both glucose *and* pyruvate, the above polypeptide levels were still reduced relative glucose culture alone, suggesting that glucose or a resulting carbon pathway intermediate acts to induce or de-repress Hyd2 protein synthesis. Levels of Hyd1 polypeptides remained low under all conditions, indicating that the Hyd2 hydrogenase provides the primary role in hydrogen formation.

The *A. hydrogeniformans* gene inventory also revealed several putative NADH-quinone oxidoreductase protein subunits annotated as subunits F, E, B, and major (OS1_0295 to _0297 and _0965-0969) along with possible sodium proton antiporter proteins (OS1_1049-1057; Table 3). Only the OS1_0296 and _0297 proteins were made in significantly high amounts relative to the others and under any of the condition examined. Likewise, several NADH-quinone oxidoreductase like polypeptides were detected (Table 3). A role for these gene products in *A. hydrogeniformans* metabolism is unknown. The genome inventory did not reveal any anaerobic respiratory enzymes including a *complete* NADH-dehydrogenase complex, bc1 complex, nitrate or nitrite reductase, sulfate reductase, TMAO or DMSO reductase, lactate dehydrogenase, or cytochrome oxidase. This lack of a known respiratory complex supports our hypothesis that *A. hydrogeniformans* generates ATP via substrate-level phosphorylation.

Proteins for iron and flavin electron transfer in A. hydrogeniformans. Genes for several ferredoxins and other iron and/or flavin containing proteins were identified (Table 5). These included a number of predicted 2Fe-2S and/or 4Fe-4S containing proteins along with “redoxins” including an annotated ferredoxin (OS1_1855), flavorubredoxin (OS1_0882), thioredoxin (OS1_1363), two glutaredoxins (OS1_0294 and OS1_0674), rubredonin (OS1_1378), two peroxiredoins (OS1_0917 and _1723) and a multimeric flavodoxin (OS_1696). Of these thirty examined, only four proteins were detected in high abundance. These were an annotated thiol peroxidase (OS_0709), a flavorubredoxin (OS_0882), a thioredoxin 1 (OS1_1363) and a glutaredoxin-like protein (OS1_0294) at 0.38%, 0.27%, 0.95% and 0. 67% of the total cell proteins in glucose-grown cells, respectively. Most increased in abundance when cells were grown with pyruvate. All remained elevated when both pyruvate and glucose were used. Their potential roles in electron transfer and/or oxidative stress response remain to be examined. Also identified was a putative EtfAB electron transfer flavoprotein complex (OS_0897-0898). However, none of the EtfAB proteins were expressed.

5. Discussion

A. hydrogeniformans employs a non-conventional glycose fermentation pathway. Based on a combination of bioinformatic, proteomic and biochemical experiments, the *A. hydrogeniformans* the reactions for glucose fermentation to acetate, CO_2, and H_2 were assigned and evaluated. Paralogous candidates were identified for several conventional EMP pathway steps and all others employed a single enzyme based on the genomic and proteomic findings (Table 1, Figure 2). The most likely phosphofructokinase (Pfk), fructose-biphosphate aldolase (Fba), and phosphoglycerate mutase (Pgm) protein assignments were also made from the multiple paralogs identified. Unaccounted for from the usual EMP pathway was a hexokinase (Glk) gene/protein. Here, a putative PTS system or as yet unidentified kinase would provide this function. An archaeal-like GAPOR enzyme functions in place of the GAPDH (GapA) and phosphoglycerate kinase (Pgk) pathway reactions as no GAPDH and little phosphoglycerate kinase activity was detected (discussed below: Figure 2). Following formation of pyruvate, a pyruvate:ferredoxin oxidoreductase (PFOR) enzyme is then employed to generate acetyl-CoA and CO_2 rather than by use of the alternative enzymes pyruvate dehydrogenase, pyruvate formate lyase, or pyruvate carboxylase. This conclusion was based on lack of protein candidates, and/or lack of detected enzyme activity in cell extracts. Acetyl-CoA is then converted to acetate by conventional bacterial-type acetate kinase (Ack) and phosphotransacetylase (Pta) enzymes. *A. hydrogeniformans* appears to synthesize ATP solely by substrate level phosphorylation.

It is unknown why a glyceraldehyde-3-phosphate dehydrogenase-like protein (GAPDH) was synthesized in high abundance in *A. hydrogeniformans* cells when its enzymatic activity was not detected. Likewise, it is unclear why a phosphoglycerate kinase–like protein was synthesized when its function is theoretically covered by GAPOR (Table 2).

There may be a hint for the activity of this GAPDH protein in the very same archaea that model the GAPOR-utilizing glycolysis pathway we see in *A. hydrogeniformans*. Mukund et al (24) report increased GAPDH activity in *P. furiosus* when cells were grown on pyruvate versus glucose. They hypothesized that GAPDH in *P. furiosus* functions physiologically during gluconeogenesis. It is unknown if the *A. hydrogeniformans* GAPDH-like enzyme operates during gluconeogenesis in a similar manner to that from *P. furiosus.* The function of the phosphoglycerate kinase-like protein, which was also abundant in *A. hydrogeniformans*, is also unknown. It may also function during gluconeogenesis to generate an additional molecule of ATP in lieu of reduced Fd via the GAPOR reaction. This would conceivably allow cells to adjust or fine tune carbon and electron flow.

Multiple Pfk, Fba, and Pgm paralogs were revealed by the *A. hydrogeniformans* genome analysis (Table 1.1, Figure 2). Roles for the weakly expressed candidates are presently unknown but may be due to specificity for other substrates. Several pyruvate dehydrogenase-like proteins (locus tags OS1_0084, 0088, 0089, and 0776) were synthesized (Tables 1.4), but no enzyme activity was detected (Table 2). It was unclear if these proteins were inactive under the conditions used, or if they act on alternative substrates with different cellular functions. For example, they may be needed during biosynthesis of cell intermediates when C_6 sugar precursors are unavailable. Interestingly, growth of *A. hydrogeniformans* with pyruvate in place of glucose did not alter the level of these enzyme candidates (Tables 1.1-1.4).

Lastly, of the three Pta and three Ack candidates (Table 1), *A. hydrogeniformans* had a single abundant Pta protein that would provide for the phsopho-transacetlyase activity detected. Two abundant Ack proteins were observed and it is unknown if one or both are functional in forming acetate and ATP. Roles for the other paralogs are also unclear but may related to the metabolism of higher chain-length CoA intermediates generated by fermentation of other substrates.

***A. hydrogeniformans* encodes multiple AFOR/GAPOR family genes.** This study demonstrates the first example of a GAPOR enzyme activity in a bacterial species. *A. hydrogeniformans* was shown to have four paralogous candidates for the enzyme while proteomic studies establish that only OS1_0269 was made in significant amounts (i.e., by ~100-fold) relative to the other three candidates (Table 1). GAPOR enzymes have previously been found only in anaerobic archaea species with an exception of the microaerobic hyperthermophilic archaeon, *P. aerophilum.* The *P. aerophilum* GAPOR remained active under an atmosphere containing 0.2% O_2 (20) in contrast to the oxygen-sensitive GAPOR's from other archaea. GAPOR activity in *A. hydrogeniformans*, also, was maintained in an atmosphere of up to 0.2% O_2 (data not shown).

The four GAPOR candidates identified in this study (Figure 1) belong to the aldehyde: ferredoxin oxidoreductase protein superfamily called AFOR that currently consists of five semi-characterized enzyme types named AFOR, GAPOR, FOR, wor4, and wor5 (See thesis Appendix materials table 1). All the AFOR family enzymes consist of a single subunit, but in different stoichiometries, from monomers (GAPOR) to homotetramers (FOR). The *A. hydrogeniformans* AFOR-family proteins range from 60.5 kDa to 81.8 kDa in size, and their stoichiometry is unknown.

Protein sequence alignment comparisons of the four *A. hydrogeniformans* GAPOR-like proteins reveal weak sequence conservation between them or to the previously described archaeal GAPORs from *Methanococcus maripaludus* and *Pyrococcus furiosus* (Table 6). These protein sequence identities only range from the low 20% up to 53%. Alignment of the GAPOR primary amino acid sequences reveals that all the *A. hydrogeniformans* AFOR genes are more similar to the *Pyrococcus furiosus* GAPOR than that from *M. maripaludis*, but of the candidates, OS1_0296 is most like the *M. maripaludus* non-phosphorylating GAPOR (Table 6). *M. maripaludus and A. hydrogeniformans* have in common that they are not hyperthermophiles, as all other currently-documented GAPOR-utilizers: *M. maripaludus* is a mesophile (optimum growth temperature 38°C), whereas *A. hydrogeniformans* is a moderate thermophile (optimum growth temperature 50°C), as are the other *Acetomicrobium* species (5, 6, 25). The *Acetomicrobium* species differ, however, in their salt tolerances.

Based on this alignment alone, and given that OS1_0269 is the most abundantly expressed AFOR under glycolytic conditions and likely acting as a GAPOR, one would conclude that *A. hydrogeniformans* has a phosphorylating GAPOR similar to that found in *Pyrococcus* *furiosus*. But, this was based on only 53% identity. A phylogenetic analysis of the *A. hydrogeniformans* AFOR genes with multiple archaeal AFORs reveals a more complex picture than the simple protein alignment would convey (figure 4). OS1_1044 and OS1_1674 form their own clade, branching off from the common ancestor to all the broad-spectrum AFORs. This indicates that they might be unique, even perhaps forming a completely unknown class of AFORs that would be utilized under conditions other than the glycolytic or gluconeogenic. OS1_1030, which was expressed at a low level under same, is most related to the *Pyrococcus* and *Thermococcus* broad-spectrum AFORs. These AFORs were documented to catalyze the oxidation of a variety of aldehydes, but to have a preference for amino acid-derived aldehydes (26, 27). Perhaps OS1_1030 has some affinity for glyceraldehyde-3-phosphate. It would be interesting to see its expression levels under amino acid-fermenting conditions. OS1_0269, which was the most abundantly expressed under glycolytic and gluconeogenic growth, appears the most closely related to known GAPORs of the four OS1 AFOR genes. This is not, however, to say that they are similar. Indeed, OS1_0269 is a deep-branching member of a category all its own, having split from the GAPORs with the other AFORs from a common ancestor for the whole group of enzymes, but again split from those other non-GAPOR AFORs at an ancient common ancestor. If it is a GAPOR, it is a unique GAPOR, indicating early horizontal transfer to an ancestor of *A. hydrogeniformans* or divergence from the common ancestor of all prokaryotes.

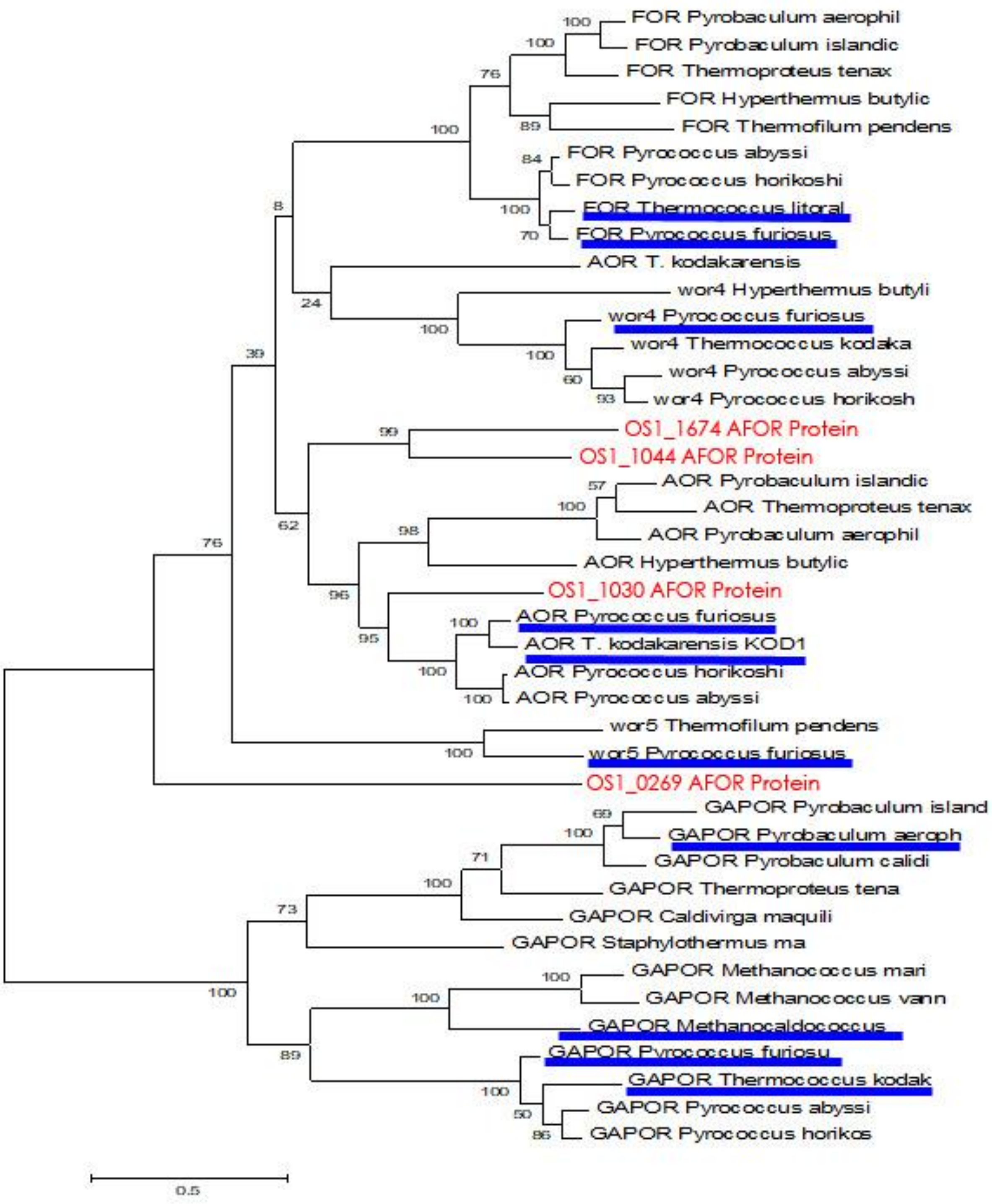

Figure 4. Phylogenetic tree of the archaeal AFORs and the AFOR candidates from *A. hydrogeniformans* OS1. *A. hydrogeniformans* genes are in red. Genes underlined in blue have

been biochemically or structurally characterized. Tree was generated by MEGA6 (1). Genes utilized for the tree's creation are listed in table S1.

Several attempts were made to informatically determine the substrate of the four AFOR-type proteins in *A. hydrogeniformans*, beyond merely comparing total protein sequence percentage identity and phylogenetic distance in comparison to known GAPORs. A protein BLAST of the N-terminal 38-residue sequence unique to the *P. furiosus* GAPOR (using NCBI tools on April 17, 2018), but not its other AFOR proteins, returned GAPOR proteins from three strains of *P. furiosus* at a 100% match with e-values of 4e-18 to 4e-16. The next best match was significantly weaker and with a correspondingly lower e-value and percent amino acid identity: this would be a predicted GAPOR from the bacterium *Thermococcus chitonophagus* that had a 74% identity and e-value of 1e-10. Of the top 45 BLAST hits, all are of the N-terminus of predicted GAPOR sequences from hyper-thermophilic archaea. However, the BLAST of this putative *P. furiosus* GAPOR-specific N-terminal sequence to *A. hydrogeniformans* (at IMG) gave no hits. A BLAST of the N-terminal sequence from *P. furiosus* (24) performed on four *M. maripaludis* genomes (c5, c6, c7, x1) also returned no matches. Likewise, a BLAST of all genomes in IRG/ER yielded a similar result to the above BLASTs using the NCBI database. All top hits were from putative GAPORs form thermophilic archaea. These findings further suggest that the a GAPOR from *A. hydrogeniformans* may be unique, or, at the very least, significantly different from those found in hyperthermophlic archaea. Unfortunately, this N-terminus signature that differentiates GAPOR from other AFOR-family enzymes in *Pyrococcus,* will not be of use in determining the GAPOR from other AFOR-family enzymes in *A. hydrogeniformans*.

Also unknown is the type of metal cofactor present in the *A. hydrogeniformans* GAPOR candidates. Tungsten and molybdenum examples have been identified in the archaeal enzymes:

W in the hyperthermophile GAPORs, Mo in the *M. maripaludus* GAPOR (24, 28). Indeed, when the *M. maripaludus* GAPOR gene was expressed in *E. coli* grown in a medium containing tungsten as well as molybdenum, no enzyme activity was detected. This shows that the *M. maripaludus* GAPOR requires molybdenum to function, and that tungsten can out-compete it for binding or in some way interfere with its ability to function.

***A. hydrogeniformans* employs a homodimeric pyruvate ferredoxin oxidoreductase, PFOR.** Pyruvate:ferredoxin oxidoreductase is a member of the 2-oxyacid:ferredoxin oxidoreductase enzyme superfamily (OFOR) that thus far consists of seven subclasses: OFOR, PFOR, PFOR/PDC, OGOR, IOR, OOR, and VOR (see Table 2 in Appendix materials). All four of the *A. hydrogeniformans* PFOR candidates belong to the OFOR family. The large single subunit type encoded by OS1_0570 was made in highest abundance (Table 1). The other three OFOR-type enzymes encoded by sets of four genes each (OS1_1194-1197, 1680-1683 and 1765-1768) were either in low abundance or not detected. Properties of the *A. hydrogeniformans* PFOR-like proteins are shown in Table 7. Of the four subunit types, the corresponding α, β, γ, δ subunits are similar in size as is their relative gene orders. Also evident is the conservation of their domains within the larger PFOR single subunit polypeptide encoded by OS1_ 0570. Ferredoxin-like proteins that would provide for electron transfer to the hydrogenases are found in the PorD or δ subunits of the tetrameric OFORs. These have yet to be tested experimentally for their roles in generating and transferring low potential electrons used for hydrogen production.

This provides for the energetically favorable attributes of the *A hydrogeniformans* pathway along with GAPOR. The metabolic strategy is superior to NADH-mediated hydrogen reduction as performed by enteric bacteria (29), despite the loss of one ATP-generating step in glycolysis. Unresolved are roles for the OFOR2, OFOR3, and OFOR4 enzymes in *A hydrogeniformans.* These are likely involved in 2-oxyacid metabolism of other fermentations for amino acids (see appendix for details).

***A. hydrogeniformans* generates hydrogen from a low potential electron carrier via Hyd2.** Of the two gene clusters in *A. hydrogeniformans* encoding proposed hydrogenases, a putative soluble FeFe-type hydrogenase designated Hyd2 appears to provide the major enzyme role in hydrogen gas formation. This is based on the approximately ten-fold higher abundance of the encoded proteins relative to those of a membrane-associated NiFe-type enzyme, Hyd1 (Figure 2, Table 4). The subunit properties and cofactor compositions of these enzymes are currently unknown as neither has been isolated and characterized. Since no formate dehydrogenase activity was detected in *A. hydrogeniformans* cells under glycolytic conditions (Table 3), formate does not appear to be an intermediate or end product for disposal of electrons originating from GAPOR or PFOR under same.

The flow of electrons from any of the OFOR enzyme candidates to hydrogen in *A. hydrogeniformans* is likely provided by either the delta unit of the respective enzymes (i.e., OS1_1194, 1680, 1795) that possesses a 2/4Fe 2/4S cluster, or by an analogous section of the OS1_0570-encoded polypeptide central domain. It is yet unclear how the GAPOR enzyme would accomplish electron transfer to hydrogenase. Top candidates for transferring GAPOR-liberated electrons are other ferredoxin/redoxin-like protein candidates identified in the genome search. (Table 4) Based on the proteomic data, the best candidates are an annotated thiol peroxidase (OS1_0709), a flavorubredoxin (OS_0882), a thioredoxin 1 (OS1_1363) and/or a glutaredoxin-like protein (OS1_0294). All were in high abundance (0.38%, 0.27%, 0.95% and 0. 67%, respectively of the total cell protein) in glucose-grown cells. These candidates were identified by a pfam search for proteins containing an iron-sulfur center that were (a) not associated with another known function, such as a hydrogenase function, and (b) smaller than 400AA. We did not find any studies of GAPOR activity that did not use ferredoxin or another electron acceptor in their assays, so it is not clear if GAPOR could deliver electrons directly to the hydrogenase.

These electron-carrier candidates were also highly expressed under pyruvate-only growth and glucose + pyruvate growth, with OS1_0709 (thiol peroxidase) reaching 1.08% total cell protein under pyruvate (gluconeogenesis) growth. OS1_0294, a putative glutaredoxin-like protein of the Yru-B family, jumps from 0.07% total protein under glucose-only growth to 0.31% total protein under pyruvate-only growth, and the soluble thioredoxin (OS1_1363) jumps from 0.09% total protein under glucose-only growth to 0.25% total protein under pyruvate-only growth. In sum, under glycose growth conditions (i.e., glucose-only), these putative electron transferring-proteins compose 1.01% of the total cell protein; under gluconeogenic conditions (pyruvate-only), 2.50% of total cell protein; under glucose + pyruvate growth, 2.26% of total cell protein. (Table 6) This indicates greater need for these electron-transfer capabilities under gluconeogenic conditions, although the subject of the study is hydrogen production under glycolytic conditions. Hydrogen production under gluconeogenic conditions has not been tested.

In either case, the use of the low-potential electron carrier would be energetically superior to use of NADH as the electron carrier. The latter with a higher mid-volt potential would be subject to hydrogenase inhibition as hydrogen gas accumulates. Indeed, this seems to account for the near-record accumulation of hydrogen in the culture head space (5, 6). Future studies are needed to biochemically isolate, identify and characterize the putative electron carrier(s) and hydrogenases used by *A. hydrogeniformans*.

The above strategy of employing a low potential electron carrier would resolve the apparent thermodynamic hurdle encountered at elevated hydrogen levels versus by using NADH as the electron donor (Figure 2). A higher and less favorable mid-volt potential presents significant problems for many fermentative bacteria that employ pyruvate dehydrogenase and/or pyruvate formate lyase type enzymes for pyruvate oxidation. Likewise, employing ferredoxin-utilizing GAPOR in place of NAD-utilizing GAPDH in earlier stages of glucose oxidation may have a similar energetically-beneficial effect for the organism.

Energy conservation in A. hydrogeniformans. Substrate level phosphorylation appears to be the sole energy-conserving reaction employed by *A. hydrogeniformans* under glycolytic growth conditions. Hyd1 may be part of a formate hydrogenlyase complex, but it was not well-expressed under any of the tested conditions (table 4). Two putative Na^+/H^+ antiporters were detected in the genome, but only one was minimally expressed, while the other was not expressed at all (table 4). The first one's ability to contribute considerably to the cell's energy is questionable and further research would have to be done to pinpoint its function. The expressed predicted hydrogenase, Hyd2, does not appear to include a manner of generating a membrane-associated ion gradient, nor is thought to operate by bifurcation based on the enzyme assay data described in the Results section (M. McInerney, personal communication). Potentially, another energy conserving step could involve cell acetate export by a secondary-type symporter that would generate a Na^+/H^+ ion gradient, but the genome did not reveal any obvious candidates.

If substrate level phosphorylation indeed provides the sole supply of energy in *A. hydrogeniformans*, glucose fermentation would produce a net yield of 2 ATP molecules per glucose metabolized assuming one is needed to activate glucose. Phosphoglycerate kinase appears inactive, which deprives the cell of the two ATP it generates. This leaves two ATP-forming kinases, pyruvate kinase and acetate kinase, to generate 2 ATP each for a total of 4 ATP, two of which would be re-invested in the formation the early pathway intermediates glucose-6-phosphate and fructose-1, 6-phosphosphate.

Metabolic implications for related Synergistetes species. Our analysis of genomes of related *Acetomicrobium* species, reveals a likely similar strategy for metabolism of glucose (data not shown). This likewise may apply to more distantly related members of the *Synergistetes*. A preliminary genomic search of the other members of the genus, *Acetomicobium thermoterrenum, Acetomicobium mobile,* and *Acetomicobium flavidum*, reveals analogous sets of genes to *A. hydrogeniformans* for the core pathway for metabolism of glucose (data not shown). These include numbers and types of GAPOR, PFOR, Hyd1, Hyd2, and other glucose pathway enzymes. Thus, a similar modified glycolytic pathway is predicted for these genus members that that would include all gene candidates for the modified EMP pathway and associated reactions leading to hydrogen formation. These findings enlarge the known range of strategies employed by bacteria in nature. They also necessitate a reexamination of the rapidly accumulating genomic and metagenomic data.

6. Acknowledgements

Enzyme analysis was supported by grant DE-FG02-96ER20214 from the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the U.S. Department of Energy to MJM. Genome analysis, metabolic reconstruction, cell cultivation and proteomic data analysis was supported by U.S. Department of Energy grant DE-FG03-

86ER13498 (RPG), the UCLA-DOE Institute DE-FC02-02ER63421 (RPG and JAL) and NSF 1244566 to RPG and MJM.

This paper is a manuscript in preparation with the following authors: Lauren Emma Cook, Janine Fu, Neil Q. Wolford, Fari Sedighian, Joseph A. Loo, Ralph S. Tanner, Rachel R. Ogorzalek Loo, M. J. McInerney, and Robert P. Gunsalus.

7. Tables

Table 1.1. Gene/protein orthologous candidates for the EMP pathway in *A. hydrogeniformans.* Please see Materials and Methods.

Gene	**Product**	**OS1 Locus Tag**	**Average Total Cellular Protein**		
			Glucose	**Pyruvate**	**Glucose & Pyruvate**
pts	phosphocarrier protein	OS1_0546	0.049%	0.039%	0.031%
	IIA component, nitrogen regulator	OS1_0755	0.014%	0.012%	0.030%
	IIC component, fructose	OS1_0756	0.012%	0.008%	0.021%
	IIC component, glucosamine	OS1_1851	0.145%	0.106%	0.145%
pgi	glucose-6-phosphate isomerase	OS1_0293	0.200%	0.525%	0.582%
pfk	Phosphofructokinase	OS1_0077	0.001%	0.000%	0.000%
		OS1_0754	0.003%	0.001%	0.005%
		OS1_1537	0.057%	0.104%	0.092%
		OS1_1613	0.009%	0.011%	0.011%
fba	Fructose bisphosphate aldolase	OS1_0322	0.069%	0.086%	0.147%
		OS1_0699	0.015%	0.016%	0.019%
		OS1_1515	0.225%	0.389%	0.366%
tpiA	triose-phosphate isomerase	OS1_1934	0.296%	0.293%	0.256%
gapA	Glyceraldehyde-3-phosphate dehydrogenase	OS1_1932	5.090%	5.149%	4.835%
pgk	Phosphoglycerate kinase	OS1_1933	4.008%	4.275%	4.123%
gor	Glyceralde-3-phosphate ferredoxin oxidoreductase	OS1_0269	0.274%	0.349%	0.279%
		OS1_1030	0.003%	0.032%	0.044%
		OS1_1044	0.000%	0.008%	0.011%
		OS1_1674	No protein detected		

pgm	Phosphoglycerate mutase	OS1_0109	0.205%	0.307%	0.338%
		OS1_0573	0.013%	0.031%	0.026%
		OS1_0578	0.001%	0.006%	0.004%
		OS1_0639	0.309%	0.386%	0.409%
		OS1_1168	No protein detected		
		OS1_2035	0.008%	0.004%	0.006%
		OS1_2095	0.001%	0.001%	0.001%
		OS1_2105	No protein detected		
eno	Enolase	OS1_1402	0.080%	0.096%	0.102%
pyk	Pyruvate kinase	OS1_1621	0.067%	0.144%	0.138%

Table 1.2. Gene/protein orthologous candidates for the pentose phosphate pathway by *A. hydrogeniformans.* The pathway was based on the MetaCyc database (2). Please see Materials and Methods.

Gene	Product	OS1 Locus Tag	Average Total Cellular Protein		
			Glucose	Pyruvate	Glucose & Pyruvate
zwf	NADP$^+$-dependent glucose-6-phosphate dehydrogenase	No ID	---	---	---
pgl	6-phosphogluconolactonase	No ID	---	---	---
gnd	6-phosphogluconate dehydrogenase	No ID	---	---	---
rpe	ribulose-phosphate 3-epimerase	OS1_0621	0.003%	0.004%	0.004%
rpiB	ribose-5-phosphate isomerase B	OS1_1736	0.045%	0.068%	0.073%
rpiA	ribose-5-phosphate isomerase A	No ID	---	---	---
tktA	transketolase 1	OS1_1374	0.068%	0.131%	0.131%
tktB	transketolase 2	OS1_1374	0.068%	0.131%	0.131%
talB	transaldolase B	OS1_0322	0.069%	0.086%	0.147%

Table 1.3. Gene/protein orthologous candidates for the Entner-Doudoroff pathway in *A. hydrogeniformans.* Please see Materials and Methods.

Gene	Product	OS1 Locus Tag	Average Total Cellular Protein		
			Glucose	Pyruvate	Glucose & Pyruvate
pts	phosphocarrier protein	OS1_0546	0.049%	0.039%	0.031%
	IIA component, nitrogen regulator	OS1_0755	0.014%	0.012%	0.030%
	IIC component, fructose	OS1_0756	0.012%	0.008%	0.021%

	IIC component, glucosamine	OS1_1851	0.145%	0.106%	0.145%
glk	hexokinase	No ID	---	---	---
zwf	NADP$^+$-dependent glucose-6-phosphate dehydrogenase	No ID	---	---	---
edd	6-phosphogluconate dehydrataste	OS1_0422	0.058%	0.069%	0.083%
		OS1_2111	---	---	---
eda/kdgA	KDPG aldolase	OS1_0699	0.015%	0.016%	0.019%
gapA	Glyceraldehyde-3-phosphate dehydrogenase	OS1_1932	5.090%	5.149%	4.835%
garK, glxK	glycerate kinase	No ID	---	---	---
eno	Enolase	OS1_1402	0.080%	0.096%	0.102%
pyk	Pyruvate kinase	OS1_1621	0.067%	0.144%	0.138%

Table 1.4. Gene/protein orthologous candidates for the conversation of pyruvate to acetyl-CoA by *A. hydrogeniformans*. Includes candidates for replacement steps in those with no enzyme activity. The pathway was based on the MetaCyc database (2). Please see Materials and Methods.

Gene	**Product**	**OS1 Locus Tag**	**Average Total Cellular Protein**		
			Glucose	**Pyruvate**	**Glucose & Pyruvate**
pdhA	Pyruvate Dehydrogenase E1 Alpha	OS1_0089	0.000%	0.000%	0.000%
pdhB	Pyruvate Dehydrogenase E1 Beta	OS1_0088	No protein detected		
pdhC	Pyruvate Dehydrogenase E2	OS1_0084	0.001%	0.003%	0.003%
pdhD	Pyruvate Dehydrogenase E3	OS1_0776	0.013%	0.005%	0.006%
Por1	Pyruvate: ferredoxin oxidoreductase	OS1_0570	5.494%	3.501%	3.689%
porD2		OS1_1194	0.003%	0.002%	0.003%
porA2		OS1_1195	0.027%	0.019%	0.021%
porB2		OS1_1196	0.027%	0.023%	0.029%
porG2		OS1_1197	0.028%	0.030%	0.027%
porD3		OS1_1680	0.002%	0.001%	0.002%
porA3		OS1_1681	0.061%	0.023%	0.042%
porB3		OS1_1682	0.096%	0.040%	0.064%
porG3		OS1_1683	0.118%	0.046%	0.063%
porD4		OS1_1765	0.037%	0.016%	0.033%
porA4		OS1_1766	0.386%	0.339%	0.343%
porB4		OS1_1767	0.150%	0.116%	0.129%
porG4		OS1_1768	0.088%	0.073%	0.073%
pta	acetyl-CoA C-acetyltransferase	OS1_0712	0.623%	0.612%	0.613%
		OS1_0713	0.034%	0.015%	0.018%

		OS1_0863	0.019%	0.016%	0.023%
ack	Acetate kinase	OS1_0711	0.193%	0.144%	0.186%
		OS1_0714	0.060%	0.035%	0.043%
		OS1_1570	0.196%	0.174%	0.223%

Table 2. Activity of glucose fermentation pathway enzymes in *A. hydrogeniformans OS1*.

Designation	Enzyme	Assay condition	Specific activity (nmol/min/mg) ± std dev
Glk/Hk	Hexose kinase	ATP	9.4 ± 0.7
		ADP	4.2 ± 1.2
		none	1.5 ± 1.1
Pgi	Phosphoglucose isomerase		150
Pfk	Phosphofructokinase	ATP	32.8 ± 11.3
		ADP	BDL
		PPi	BDL
Fba	Fructose-1,6-bisP aldolase		296
TpiA	Triose phosphate isomerase		3,300
GapA	Glyceraldehyde-3-P dehydrogenase	NAD^+	BDL
		$NADP^+$	BDL
GAPOR	Glyceraldehyde-3-P:ferredoxin oxidoreductase		39 ± 13
Pkg	3-Phosphoglycerate kinase	1,3-bis-P glycerate direction	1030
		G3P direction	980
Eno	2-phosphoglycerate enolase		310
	Pyruvate kinase		73
PFOR	Pyruvate:ferredoxin oxidoreductase	MV	1080
Pdh	Pyruvate dehydrogenase	NAD^+	BDL
		$NADP^+$	BDL
Pta	Phosphotransacetylase		725
Ack	Acetate kinase		1270
Acs	Acetyl-CoA synthetase (AMP-forming)		520
Zwf	Glucose dehydrogenase	NAD^+	13
		$NADP^+$	8
Gnd	6-phosphogluconate dehydrogenase		16
KdgK	2-Oxogluconate kinase		40
	Glyceraldehyde oxidoreductase	MV	75
TalB	Transaldolase		22 ± 10
TktAB	Transketolase		18

Activities are expressed in µmol substrate utilized per min per mg protein. Individual assays and preparation of cell extracts are described in Materials and Methods. Abbreviations: BDL: below detection limit; MV: methyl viologen; BV: benzyl viologen.

The activity was calculated from the slope of rates of absorbance change plotted against at least three different protein concentrations. Means and standard deviations are given when activity was measured with cell-free extracts prepared from different cultures.

Table 3. Activity of hydrogenase, formate dehydrogenase, and formate hydrogenlyase in *A. hydrogeniformans OS1*.

Enzyme	Fraction/Assay	Specific activity (nmol/min/mg)
Hydrogenase	Supernatant MV	5,500
	Supernatant BV	10,900
	Pellet MV	7,500
	Pellet BV	19,400
NADH:MV OR		310
NADPH:MV OR		510
Formate dehydrogenase		BDL
NADH:MV OR		310
NADPH:MV OR		510

Activities are expressed in umol substrate utilized per min per mg protein. Individual assays were performed in *A. hydrogeniformans* broken cells and in the particulate and soluble cell fractions. The BV assays and preparation of cell extracts are described in Materials and Methods, and preparation of cell extracts are described in Materials and Methods. Abbreviations: BDL: below detection limit; MV: methyl viologen; BV: benzyl viologen

The activity was calculated from the slope of rates of absorbance change plotted against at least three different protein concentrations. Means and standard deviations are given when activity was measured with cell-free extracts prepared from different cultures.

Table 4. Gene/protein candidates for hydrogenases and dehydrogenases in *A. hydrogeniformans.* Hydrogenases include a membrane-bound [NiFe]-type hydrogenase in conjunction with a formate dehydrogenase, forming a formate hydrogenlyase complex, and a soluble [FeFe] type hydrogenase. Gene/proteins for unspecified dehydrogenase subunits with homology to NADH-ubiquinone oxidoreductase are also included. Dehydrogenases include the formate dehydrogenase in the aforementioned putative formate hydrogenlyase complex; two predicted formate dehydrogenases, and a predicted unspecified dehydrogenase with homology to NADH-ubiquinone oxidoreductase.

Gene Product Name	**Locus Tag**	**Average Total Cellular Protein**		
		Glucose	**Pyruvate**	**Glucose & Pyruvate**
Hyd1 [Ni Fe]-type hydrogenase + formate dehydrogenase (formate hydrogenlyase)				
Formate hydrogenlyase subunit 6/NADH:ubiquinone oxidoreductase 23 (chain I)	OS1_0245	0.003%	0.002%	0.002%
Formate hydrogenlyase subunit 4	OS1_0246	0.010%	0.016%	0.013%
membrane-bound hydrogenase subunit alpha	OS1_0247	0.072%	0.091%	0.085%
Respiratory-chain NADH dehydrogenase, subunit	OS1_0248	0.038%	0.043%	0.046%
membrane-bound hydrogenase subunit mbhJ	OS1_0249	0.033%	0.033%	0.038%
hypothetical protein	OS1_0250	0.000%	0.000%	0.000%
multicomponent Na+:H+ antiporter subunit D	OS1_0251	0.005%	0.007%	0.009%
multicomponent Na+:H+ antiporter subunit C	OS1_0252	0.005%	0.006%	0.008%
multicomponent Na+:H+ antiporter subunit B	OS1_0253	0.001%	0.004%	0.005%
hypothetical protein	OS1_0254	0.004%	0.006%	0.008%
Predicted subunit of the Multisubunit Na+/H+ antiporter	OS1_0255	0.000%	0.000%	0.000%
multicomponent Na+:H+ antiporter subunit G	OS1_0256	0.007%	0.013%	0.011%
multicomponent Na+:H+ antiporter subunit F	OS1_0257	0.006%	0.009%	0.010%
multicomponent Na+:H+ antiporter subunit E	OS1_0258	0.014%	0.022%	0.020%
Hyd2 [Fe Fe] type soluble hydrogenase + formate dehydrogenase (formate hydrogenlyase)				
hydrogenase maturation protease	OS1_0492	0.010%	0.000%	0.000%
NAD-reducing hydrogenase large subunit	OS1_0493	1.973%	0.080%	0.155%

NAD-reducing hydrogenase small subunit	OS1_0494	0.520%	0.007%	0.015%
formate dehydrogenase major subunit	OS1_0495	1.377%	0.057%	0.123%
NADH-quinone oxidoreductase subunit F	OS1_0496	0.967%	0.020%	0.065%
NADP-reducing hydrogenase subunit HndB	OS1_0497	0.174%	0.007%	0.016%
Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase	OS1_0498	0.002%	0.000%	0.000%
NADH-quinone oxidoreductase subunit E	OS1_0499	0.600%	0.015%	0.038%
hypothetical protein	OS1_0500	0.001%	0.000%	0.000%
hypothetical protein	OS1_0501	0.040%	0.001%	0.002%
Iron only hydrogenase large subunit, C-terminal domain	OS1_0502	0.033%	0.002%	0.003%
Anti-sigma regulatory factor (Ser/Thr protein kinase)	OS1_0503	0.006%	0.000%	0.000%
hypothetical protein	OS1_0504	0.009%	0.000%	0.000%
Na$^+$/H$^+$ antiporter 1				
multicomponent Na+:H+ antiporter subunit D	OS1_0251	0.005%	0.007%	0.009%
multicomponent Na+:H+ antiporter subunit C	OS1_0252	0.005%	0.006%	0.008%
multicomponent Na+:H+ antiporter subunit B	OS1_0253	0.001%	0.004%	0.005%
hypothetical protein*	OS1_0254	0.004%	0.006%	0.008%
Uncharacterized MnhB-related membrane protein	OS1_0255	No protein detected		
multicomponent Na+:H+ antiporter subunit G	OS1_0256	0.007%	0.013%	0.011%
multicomponent Na+:H+ antiporter subunit F	OS1_0257	0.006%	0.009%	0.010%
multicomponent Na+:H+ antiporter subunit E	OS1_0258	0.014%	0.022%	0.020%
Na$^+$/H$^+$ antiporter 2				
multicomponent Na+:H+ antiporter subunit D	OS1_1049	No protein detected		
multicomponent Na+:H+ antiporter subunit D	OS1_1050			
multicomponent Na+:H+ antiporter subunit C	OS1_1051			
multicomponent Na+:H+ antiporter subunit B	OS1_1052			
Uncharacterized MnhB-related	OS1_1053			

membrane protein				
multicomponent Na+:H+ antiporter subunit G	OS1_1054			
multicomponent Na+:H+ antiporter subunit F	OS1_1055			
multicomponent Na+:H+ antiporter subunit E	OS1_1056			
Predicted Dehydrogenase				
NADH-quinone oxidoreductase subunit F	OS1_1027	0.000%	0.002%	0.002%
NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G)	OS1_1028	0.000%	0.000%	0.000%
Fe-S-cluster-containing hydrogenase component 2	OS1_1029	0.000%	0.004%	0.004%
Potential Formate Dehydrogenase I				
NADH-quinone oxidoreductase subunit E	OS1_0295	0.012%	0.011%	0.016%
NADH-quinone oxidoreductase subunit F	OS1_0296	0.177%	0.235%	0.266%
formate dehydrogenase major subunit	OS1_0297	0.132%	0.185%	0.164%
Potential Formate Dehydrogenase II				
NADH-quinone oxidoreductase subunit E	OS1_0965	0.000%	0.000%	0.000%
NADP-reducing hydrogenase subunit HndB	OS1_0966	0.000%	0.000%	0.000%
NADH-quinone oxidoreductase subunit F	OS1_0967	0.000%	0.000%	0.000%
formate dehydrogenase major subunit	OS1_0968	0.000%	0.000%	0.000%

*BLAST indicates homology to subunit A.

Table 5. Putative electron-carrying proteins containing a 2Fe-2S and/or 4Fe-4S type cluster. This includes various redoxins and iron-sulfur-containing proteins without another evident enzymatic classification less than 600 amino acids. Includes a sum of the total cell protein dedicated to these putative electron carriers under each growth condition.

Gene Product Name	**Locus Tag**	**AA**	**TM Helices**	**Average Total Cellular Protein**		
				Glucose	**Pyruvate**	**Glucose & Pyruvate**
thiol peroxidase, atypical 2-Cys peroxiredoxin	OS1_0709	171	0	0.376%	1.085%	0.914%
Flavorubredoxin	OS1_0882	401	0	0.269%	0.405%	0.439%
thioredoxin 1	OS1_1363	106	0	0.095%	0.250%	0.160%
Glutaredoxin-like protein, YruB-family	OS1_0294	81	0	0.067%	0.315%	0.303%
hypothetical protein	OS1_0291	311	0	0.040%	0.123%	0.135%
electron transfer flavoprotein alpha subunit apoprotein	OS1_0898	326	0	0.027%	0.015%	0.020%
Glutaredoxin-like protein, YruB-family	OS1_0674	82	0	0.026%	0.031%	0.036%
electron transfer flavoprotein beta subunit	OS1_0897	260	0	0.020%	0.013%	0.015%
Multimeric flavodoxin WrbA	OS1_1696	185	0	0.017%	0.034%	0.048%
hypothetical protein	OS1_2054	379	0	0.012%	0.117%	0.106%
small redox-active disulfide protein 2	OS1_0335	81	0	0.010%	0.027%	0.018%
Uncharacterized conserved protein, DUF362 family	OS1_1810	382	0	0.009%	0.009%	0.009%
4Fe-4S binding domain-containing protein	OS1_0581	252	0	0.009%	0.014%	0.009%
4Fe-4S dicluster domain-containing protein	OS1_0268	127	0	0.007%	0.007%	0.008%
4Fe-4S dicluster domain-containing protein	OS1_0783	171	0	0.005%	0.010%	0.007%
peroxiredoxin (alkyl hydroperoxide reductase subunit C)	OS1_1724	118	0	0.004%	0.014%	0.009%
2Fe-2S iron-sulfur cluster binding domain-containing protein	OS1_0785	110	0	0.004%	0.007%	0.005%
BFD-like [2Fe-2S] binding domain-containing protein	OS1_0782	96	0	0.004%	0.006%	0.006%
protein of unknown function (DUF4139)	OS1_0462	499	1	0.003%	0.006%	0.005%
Uncharacterized 2Fe-2 and 4Fe-4S clusters-containing protein, contains DUF4445	OS1_0583	610	0	0.001%	0.002%	0.001%

domain						
Uncharacterized conserved protein, DUF362 family	OS1_1059	434	0	0.001%	0.004%	0.004%
Peroxiredoxin	OS1_1723	47	0	0.001%	0.001%	0.000%
Peroxiredoxin	OS1_0917	167	1	0.000%	0.003%	0.002%
4Fe-4S dicluster domain-containing protein	OS1_1104	261	3	0.000%	0.001%	0.000%
hypothetical protein	OS1_2019	439	0	No protein detected		
ferredoxin	OS1_1855	60	0	No protein detected		
Rubredoxin	OS1_1378	52	0	No protein detected		
4Fe-4S binding domain-containing protein	OS1_1103	65	0	No protein detected		
hypothetical protein	OS1_0822	369	0	No protein detected		
hypothetical protein	OS1_0823	473	0	No protein detected		
hypothetical protein	OS1_1483	156	0	No protein detected		
Sum of total cell protein percentages for each growth category				**% of Total Cell Protein**		
				Glucose	**Pyruvate**	**Glucose & Pyruvate**
Total Cell Protein Dedicated to Small Electron Carriers				1.007%	2.498%	2.257%

Table 6 Properties and abundance of the *A. hydrogeniformans* AFOR-family proteins (GAPOR family). *M. maripaludis* GAPOR sequence from strain C6 genome (IMG gene ID 641284211) and *P. furiosus* GAPOR sequence from strain DSM 3638 (IMG gene ID 638172975) as recorded at the Integrated Microbial Genomes database (3). Percent identity matrix generated by Clustal2.1 (4).

Locus Tag	**Predicted Size (AA)**	**Predicted MW (kDa)**	**Homology with other AFOR-like genes (percent identity matrix)**					
			OS1 Genes – internal comparison				**GAPORs from other organisms**	
			0269	**1030**	**1044**	**1674**	***M. maripaludis***	***P. furiosus***
OS1_0269	722	81.81	----	26.48%	25.16%	24.09%	20.11%	28.62%
OS1_1030	604	64.94	26.48%	-----	38.10%	36.94%	21.85%	53.92%
OS1_1044	634	69.22	25.16%	38.10%	-----	40.60%	17.90%	40.27%
OS1_1674	559	60.49	24.09%	36.94%	40.60%	-----	20.38%	37.36%

Table 7 Properties and abundance of the *A. hydrogeniformans* OFOR-family proteins (PFOR family).

***A. hydrogeniformans* OS1 OFOR-family genes**							
Locus Tag	**Subunit**	**Predicted Size (AA)**	**Predicted MW (kDa)**	**Gene Order**	**Average Total Cell Protein**		
					Glucose	**Pyruvate**	**Glucose + Pyruvate**
Potential PFORs (based on subunit composition and homology)							
OFOR Homodimer							
OS1_0570	------	1184	130.55	porAGDB*	**5.494%**	**3.501%**	**3.689%**
OFOR Tetramer 1							
OS1_1194	δ	85	9.75	porDABG	0.003%	0.002%	0.003%
OS1_1195	α	379	41.45		0.027%	0.019%	0.021%
OS1_1196	β	280	30.94		0.027%	0.023%	0.029%
OS1_1197	γ	181	19.13		0.028%	0.030%	0.027%
OFOR Tetramer 2							
OS1_1680	δ	68	7.55	porDABG	0.002%	0.001%	0.002%
OS1_1681	α	377	41.95		0.061%	0.023%	0.042%
OS1_1682	β	274	29.73		0.096%	0.040%	0.064%
OS1_1683	γ	186	20.39		0.118%	0.046%	0.063%
OFOR Tetramer 3							
OS1_1765	δ	71	7.9	porDABG	0.037%	0.016%	0.033%
OS1_1766	α	353	38.72		0.386%	0.339%	0.343%
OS1_1767	β	247	27.04		0.150%	0.116%	0.129%
OS1_1768	γ	188	19.97		0.088%	0.073%	0.073%
Other OFORs							
KOR (2-oxoglutarate: ferredoxin oxidoreductase							
OS1_0358	α				0.027%	0.075%	0.076%
OS1_0359	β				0.030%	0.062%	0.068%
IOR (Indolepyruvate: Ferredoxin Oxidoreductase) 1							
OS1_0239	α				0.004%	0.004%	0.005%
OS1_0240	β				0.004%	0.003%	0.003%
IOR (Indolepyruvate: Ferredoxin Oxidoreductase) 2							
OS1_1209	α				0.333%	0.069%	0.086%
OS1_1210	β				0.147%	0.028%	0.038%

*fused into one protein

Table S1. Gene/protein orthologous candidates for the EMP pathway in *A. hydrogeniformans.* Please see Materials and Methods.

Gene	**Product**	**OS1 Locus Tag**	**Average Total Cellular Protein**		
			Glucose	**Pyruvate**	**Glucose & Pyruvate**
pts	phosphocarrier protein	OS1_0546	0.049%	0.039%	0.031%
	IIA component, nitrogen regulator	OS1_0755	0.014%	0.012%	0.030%
	IIC component, fructose	OS1_0756	0.012%	0.008%	0.021%
	IIC component, glucosamine	OS1_1851	0.145%	0.106%	0.145%
pgi	glucose-6-phosphate isomerase	OS1_0293	0.200%	0.525%	0.582%
pfk	Phosphofructokinase	OS1_0077	0.001%	0.000%	0.000%
		OS1_0754	0.003%	0.001%	0.005%
		OS1_1537	0.057%	0.104%	0.092%
		OS1_1613	0.009%	0.011%	0.011%
fba	Fructose bisphosphate aldolase	OS1_0322	0.069%	0.086%	0.147%
		OS1_0699	0.015%	0.016%	0.019%
		OS1_1515	0.225%	0.389%	0.366%
tpiA	triose-phosphate isomerase	OS1_1934	0.296%	0.293%	0.256%
gapA	Glyceraldehyde-3-phosphate dehydrogenase	OS1_1932	5.090%	5.149%	4.835%
pgk	Phosphoglycerate kinase	OS1_1933	4.008%	4.275%	4.123%
gor	Glyceralde-3-phosphate ferredoxin oxidoreductase	OS1_0269	0.274%	0.349%	0.279%
		OS1_1030	0.003%	0.032%	0.044%
		OS1_1044	0.000%	0.008%	0.011%
		OS1_1674	No protein detected		
pgm	Phosphoglycerate mutase	OS1_0109	0.205%	0.307%	0.338%
		OS1_0573	0.013%	0.031%	0.026%
		OS1_0578	0.001%	0.006%	0.004%
		OS1_0639	0.309%	0.386%	0.409%
		OS1_1168	No protein detected		
		OS1_2035	0.008%	0.004%	0.006%
		OS1_2095	0.001%	0.001%	0.001%
		OS1_2105	No protein detected		
eno	Enolase	OS1_1402	0.080%	0.096%	0.102%

pyk	Pyruvate kinase	OS1_1621	0.067%	0.144%	0.138%

Table S2. Gene/protein orthologous candidates for the pentose phosphate pathway by *A. hydrogeniformans.* The pathway was based on the MetaCyc database (2). Please see Materials and Methods.

Gene	**Product**	**OS1 Locus Tag**	**Average Total Cellular Protein**		
			Glucose	**Pyruvate**	**Glucose & Pyruvate**
zwf	NADP$^+$-dependent glucose-6-phosphate dehydrogenase	No ID	---	---	---
pgl	6-phosphogluconolactonase	No ID	---	---	---
gnd	6-phosphogluconate dehydrogenase	No ID	---	---	---
rpe	ribulose-phosphate 3-epimerase	OS1_0621	0.003%	0.004%	0.004%
rpiB	ribose-5-phosphate isomerase B	OS1_1736	0.045%	0.068%	0.073%
rpiA	ribose-5-phosphate isomerase A	No ID	---	---	---
tktA	transketolase 1	OS1_1374	0.068%	0.131%	0.131%
tktB	transketolase 2	OS1_1374	0.068%	0.131%	0.131%
talB	transaldolase B	OS1_0322	0.069%	0.086%	0.147%

Table S3. Gene/protein orthologous candidates for the Entner-Doudoroff pathway in *A. hydrogeniformans.* Please see Materials and Methods.

Gene	**Product**	**OS1 Locus Tag**	**Average Total Cellular Protein**		
			Glucose	**Pyruvate**	**Glucose & Pyruvate**
pts	PTS System	OS1_0546	0.049%	0.039%	0.031%
		OS1_0755	0.014%	0.012%	0.030%
		OS1_0756	0.012%	0.008%	0.021%
		OS1_1851	0.145%	0.106%	0.145%
glk	hexokinase	No ID	---	---	---
zwf	NADP$^+$-dependent glucose-6-phosphate dehydrogenase	No ID	---	---	---
edd	6-phosphogluconate dehydrataste	OS1_0422	0.058%	0.069%	0.083%
		OS1_2111	---	---	---
eda/kdgA	KDPG aldolase	OS1_0699	0.015%	0.016%	0.019%
gapA	Glyceraldehyde-3-phosphate dehydrogenase	OS1_1932	5.090%	5.149%	4.835%

garK, glxK	glycerate kinase	No ID	---	---	---
eno	Enolase	OS1_1402	0.080%	0.096%	0.102%
pyk	Pyruvate kinase	OS1_1621	0.067%	0.144%	0.138%

Table S4. Gene/protein orthologous candidates for the conversation of pyruvate to acetyl-CoA by *A. hydrogeniformans.* Includes candidates for replacement steps in those with no enzyme activity. The pathway was based on the MetaCyc database (2). Please see Materials and Methods.

Gene	**Product**	**OS1 Locus Tag**	**Average Total Cellular Protein**		
			Glucose	**Pyruvate**	**Glucose & Pyruvate**
pdhA	Pyruvate Dehydrogenase E1 Alpha	OS1_0089	0.000%	0.000%	0.000%
pdhB	Pyruvate Dehydrogenase E1 Beta	OS1_0088	No protein detected		
pdhC	Pyruvate Dehydrogenase E2	OS1_0084	0.001%	0.003%	0.003%
pdhD	Pyruvate Dehydrogenase E3	OS1_0776	0.013%	0.005%	0.006%
Por1	Pyruvate: ferredoxin oxidoreductase	OS1_0570	5.494%	3.501%	3.689%
porD2		OS1_1194	0.003%	0.002%	0.003%
porA2		OS1_1195	0.027%	0.019%	0.021%
porB2		OS1_1196	0.027%	0.023%	0.029%
porG2		OS1_1197	0.028%	0.030%	0.027%
porD3		OS1_1680	0.002%	0.001%	0.002%
porA3		OS1_1681	0.061%	0.023%	0.042%
porB3		OS1_1682	0.096%	0.040%	0.064%
porG3		OS1_1683	0.118%	0.046%	0.063%
porD4		OS1_1765	0.037%	0.016%	0.033%
porA4		OS1_1766	0.386%	0.339%	0.343%
porB4		OS1_1767	0.150%	0.116%	0.129%
porG4		OS1_1768	0.088%	0.073%	0.073%
pta	acetyl-CoA C-acetyltransferase	OS1_0712	0.623%	0.612%	0.613%
		OS1_0713	0.034%	0.015%	0.018%
		OS1_0863	0.019%	0.016%	0.023%
ack	Acetate kinase	OS1_0711	0.193%	0.144%	0.186%
		OS1_0714	0.060%	0.035%	0.043%
		OS1_1570	0.196%	0.174%	0.223%

Table S1. The protein sequences used to generate the phylogenetic tree in figure 4 are listed. (accession numbers):

> GAPOR: *Thermococcus kodakarensis* KOD1 YP_184576.1, *Pyrococcus furiosus* DSM 3638 NP_578193.1, *Pyrococcus abyssi* GE5 NP_127254.1, *Pyrococcus horikoshii* OT3 NP_142434.1, *Methanocaldococcus jannaschii* DSM 2661 NP_248179.1, *Methanococcus vannielii* SB ZP_01702330.1, *Methanococcus maripaludis* S2 NP_988065.1, *Staphylothermus marinus* F1 YP_001041574.1, *Caldivirga maquilingensis* IC-167 ZP_01711828.1, *Thermoproteus tenax*CAF18511.1, *Pyrobaculum islandicum* DSM 4184 YP_930734.1, *Pyrobaculum calidifontis* JCM 11548 YP_001055556.1, *Pyrobaculum aerophilum* str. IM2 NP_559028.1 WOR4: *Hyperthermus butylicus* A2BKT0, *Thermococcus kodakaraensis* Q5JI10, *Pyrococcus furiosus* O93736, *Pyrococcus horikoshii* O57750, *Pyrococcus abyssi* Q9V2P2, FOR: *Thermococcus litoralis* Q56303, *Pyrococcus furiosus* Q8U1K3, *Pyrococcus horikoshii* O74007, *Pyrococcus abyssi* Q9UZE9, *Thermoproteus tenax* Q704C0, *Pyrobaculum aerophilum* Q8ZT58, *Pyrobaculum islandicum* A1RT64, *Thermofilum pendens* A1RWC4, *Hyperthermus butylicus* DSM 5456 A2BJR2, AOR: *Thermococcus kodakarensis* KOD1 YP_182485.1, *Hyperthermus butylicus* DSM 5456 A2BLI7, *Pyrobaculum aerophilum* str. IM2 Q8ZYU2, *Pyrobaculum islandicum* DSM 4184 A1RSI3, *Thermoproteus tenax* Q703Z6, *Pyrococcus furiosus*Q51739, *Thermococcus kodakarensis* KOD1 Q5JE15, *Pyrococcus horikoshii* OT3 O58778, *Pyrococcus abyssi* GE5 Q9V035, WOR5: *Pyrococcus furiosus* AAL81604.1, *Thermofilum pendens* A1S030. (20)

8. References

1. **Tamura K, Stecher G, Peterson D, Filipski A, Kumar S.** 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **30:**2725-2729.

2. **Caspi R, Altman T, Dale J, Dreher K, Fulcher C, Gilham F, Kaipa P, Karthikeyan A, Kothari A, Krummenacker M, Latendresse M, Mueller L, Paley S, Popescu L, Pujar A, Shearer A, Zhang P, Karp P.** 2010. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res **37:**D473-479.

3. **Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics **25:**2271-2278.

4. **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23:**2947-2948.

5. **Maune MW, Tanner RS.** 2012. Description of Anaerobaculum hydrogeniformans sp. nov., an anaerobe that produces hydrogen from glucose, and emended description of the genus Anaerobaculum. Int J Syst Evol Microbiol **62:**832-838.

6. **Ben Hania W, Bouanane-Darenfed A, Cayol J-L, Ollivier B, Fardeau M-L.** 2016. Reclassification of Anaerobaculum mobile, Anaerobaculum thermoterrenum, Anaerobaculum hydrogeniformans as Acetomicrobium mobile comb. nov., Acetomicrobium thermoterrenum comb. nov. and Acetomicrobium hydrogeniformans comb. nov., respectively, and emendation of the genus Acetomicrobium. International Journal of Systematic and Evolutionary Microbiology **66:**1506-1509.

7. **Vartoukian SR, Palmer RM, Wade WG.** 2009. Diversity and Morphology of Members of the Phylum "Synergistetes" in Periodontal Health and Disease. Applied and Environmental Microbiology **75:**3777-3786.

8. **Jumas-Bilak E, Roudière L, Marchandin H.** 2009. Description of 'Synergistetes' phyl. nov. and emended description of the phylum 'Deferribacteres' and of the family Syntrophomonadaceae, phylum 'Firmicutes'. Int J Syst Evol Microbiol **59:**1028-1035

9. **Jumas-Bilak E, Marchandin H.** 2014. The Phylum Synergistetes doi:10.1007/978-3-642-38954-2_384.

10. **Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Mizrachi I, Ostell J, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Yaschenko E, Ye J.** 2008. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research **37:**D5-D15.

11. **Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW.** 2008. GenBank. Nucleic Acids Research **37:**D26-D31.

12. **Thauer RK, Jungermann K, Decker K.** 1977. Energy conservation in chemotrophic anaerobic bacteria. Bacteriol Rev **41:**100-180.

13. **Sieber JR, McInerney MJ, Gunsalus RP.** 2012. Genomic Insights into Syntrophy: The Paradigm for Anaerobic Metabolic Cooperation. Annual Review of Microbiology **66:**429-452.

14. **Ferry JF.** 2015. Acetate Metabolism in Anaerobes from the Domain Archaea. Life **5:**1454-1471.

.

15. **Cook LE, Gang SS, Ihlan A, Maune M, Tanner RS, McInerney MJ, Weinstock G, Lobos EA, Gunsalus RP.** 2018. Genome Sequence of Acetomicrobium hydrogeniformans OS1. Genome Announcements **6:**e00581-00518.

16. **Erde J, Loo RR, Loo JA.** 2014. Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. J Proteome Res **13:**1885-1895.

17. **Rappsilber J, Mann M, Ishihama Y.** 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nature Protocols **2:**1896.

18. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M.** 2014. Pfam: the protein families database. Nucleic Acids Research **42:**D222-D230.

19. **Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, Muniz-Rascado L, Bonavides-Martinez C, Paley S, Krummenacker M, Altman T, Kaipa P, Spaulding A, Pacheco J, Latendresse M, Fulcher C, Sarker M, Shearer AG, Mackie A, Paulsen I, Gunsalus RP, Karp PD.** 2011. EcoCyc: a comprehensive database of Escherichia coli biology. Nucleic Acids Res **39:**D583-590.

20. **Reher M, Gebhard S, Schönheit P.** 2007. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR) and nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN), key enzymes of the respective modified Embden-Meyerhof pathways in the hyperthermophilic crenarchaeota *Pyrobaculum aerophilum* and *Aeropyrum pernix*. FEMS Microbiol Lett **273:**196-205.

21. **Jones DT, Taylor WR, Thornton JM.** 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci **8:**275-282.

22. **Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martinez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, Latendresse M, Muniz-Rascado L, Ong Q, Paley S, Schroder I, Shearer AG, Subhraveti P, Travers M, Weerasinghe D, Weiss V, Collado-Vides J, Gunsalus RP, Paulsen I, Karp PD.** 2013. EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Res **41:**D605-612.

23. **James KL, Ríos-Hernández LA, Wofford NQ, Mouttaki H, Sieber JR, Sheik CS, Nguyen HH, Yang Y, Xie Y, Erde J, Rohlin L, Karr EA, Loo JA, Ogorzalek Loo RR, Hurst GB, Gunsalus RP, Szweda LI, McInerney MJ.** 2016. Pyrophosphate-Dependent ATP Formation from Acetyl Coenzyme A in Syntrophus aciditrophicus, a New Twist on ATP Formation. mBio **7:**e01208-01216.

24. **Mukund S, Adams MWW.** 1995. Glyceraldehyde-3-phosphate Ferredoxin Oxidoreductase, a Novel Tungsten-containing Enzyme with a Potential Glycolytic Role in the Hyperthermophilic Archaeon Pyrococcus furiosus. Journal of Biological Chemistry **270:**8389-8392.

25. **Mavromatis K, Stackebrandt E, Held B, Lapidus A, Nolan M, Lucas S, Mammon N, Deshpande S, Cheng J, Tapia R, Goodwin L, Pitluck S, Liolios K, Pagani I, Ivanova N, Mikhailova N, Huntemann M, Pati A, Chen A, Palaniappan K, Land ML, Rohde M, Spring S, Göker M, Woyke T, Detter J, Bristow J, Eisen J, Markowitz V, Hugenholtz P, Klenk H, Kyrpides N.** 2013. Complete genome sequence of the moderate thermophile Anaerobaculum mobile type strain (NGA(T)). Stand Genomic Sci **15:**47-57.

26. **Heider J, Ma K, Adams MW.** 1995. Purification, characterization, and metabolic function of tungsten-containing aldehyde ferredoxin oxidoreductase from the hyperthermophilic and proteolytic archaeon Thermococcus strain ES-1. Journal of Bacteriology **177:**4757-4764.

27. **Roy R, Menon AL, Adams MWW.** 2001. [11] Aldehyde Oxidoreductases from Pyrococcus furiosus, p 132-144, Methods in Enzymology, vol 331. Academic Press.

28. **Park M-O, Mizutani T, Jones PR.** 2007. Glyceraldehyde-3-Phosphate Ferredoxin Oxidoreductase from Methanococcus maripaludis. Journal of Bacteriology **189:**7281-7289.

29. **Sieber JR, Sims DR, Han C, Kim E, Lykidis A, Lapidus AL, McDonnald E, Rohlin L, Culley DE, Gunsalus R, McInerney MJ.** 2010. The genome of Syntrophomonas wolfei: new insights into syntrophic metabolism and biohydrogen production. Environ Microbiol **12:**2289-2301.

Chapter 6. Conclusions and future directions

Contents

1. *A. hydrogeniformans* utilizes a modified Embden-Meyerhof pathway that resembles those found in thermophilic archaea

Synergistetes species have thus-far been most intensely studied in the context of their native habitats and the roles they play in those habitats. Most papers that mention *Synergistetes* are a metagenomic census of a given location of interest, and beyond a few studies that show their role in amino acid cycling in these communities (1), they are not studied in any detail. Regardless of their central metabolic role as amino acid metabolizers in many microbial communities, their non-amino-acid related metabolisms are also unstudied. Interest in *A. hydrogeniformans* glucose metabolism came of its utility to humans for hydrogen formation, independent of the contribution that pathway makes to an ecosystem. Given the correlation of *Synergistetes* OTU diversity to disease progression in periodontitis, a common dental disease, (2-4) research into the variety of metabolic pathways and the roles they play in communities may spike in the next few years. Metagenomics identification and census research only scratches the surface of mysteries still yet undiscovered in this unique and diverse phylum. Our study of the glycolysis pathway in *A. hydrogeniformans* reveals for the first time a unique variant previously unseen in bacteria. Preliminary scanning of the genomes of other *Acetomicrobium* species reveal the potential for similar metabolisms (table 1).

A. hydrogeniformans OS1's sugar fermentation pathway serves as an example of a variant metabolism not yet seen in *Synergistetes*, or, indeed, bacteria; in this case, the novelty is a modified pathway with 'archaeal-like' aspects. There is not yet information on the extent to which the *Synergistetes* gene pool represents horizontal gene transfer from archaea, but this is a potential source of diversity, and perhaps in some ecological niches *Synergistetes* species are performing activities that archaea would otherwise perform. Or, given the diversity of environments *Synergistetes* inhabit, there is the possibility during co-existence with archaea that horizontal gene transfer occurred, and there is evidence for horizontal gene transfer between archaea and bacteria (5-7). A hydrogen-producing bacterium would find a niche with hydrogenotrophic methanogens (archaea), who consume hydrogen and carbon dioxide to produce methane. In fact, co-culture of *A. hydrogeniformans* with a suitable methanogen strain resulted in enhanced growth and hydrogen formation (8-10).

We hypothesize that *A. hydrogeniformans*' modifications to the glycolysis pathway that result in the utilization of ferredoxin over NAD result in greater energy efficiency. This energy efficiency is reflected in the production of dihydrogen at the highest rate yet documented, a rate that approaches the theoretical maximum for the reaction calculated under the ideal circumstances of the realm of theory (9, 11).

Interest in the mechanics of an efficient energy production method extends beyond the purely academic. There is immense potential in discovering a method by which glucose, which can be produced literally from sunlight and CO_2, of which we have plenty, or a feedstock such as clarified sewage, of which we have plenty, can be converted into a non-carbon-containing combustible fuel. The data is unequivocal: slow climate change or destroy the biosphere as we know it. As there is no returning to the pre-industrial era the need for combustible fuels is unavoidable for the foreseeable future. This is how microbiologists nibble at the edges of this problem: monitor how the changing climate effects microbial populations, and discover/engineer microbes that can produce the fuels we require at a lower environmental cost for extraction and combustion. This is the sort of research our colleagues at the University of Oklahoma were conducting when they discovered a hydrogen-producing microbe from a phylum about which little is currently known.

These initial findings – the hydrogen-production potential of *A. hydrogeniformans* OS1, and the great potential of hydrogen as a clean biofuel—were the foundation of the research here catalogued. We wanted to know: what is the molecular mechanism for this production, and what is unique about this mechanism that allows for such efficient fermentation? *A. hydrogeniformans* was only recently discovered, as was its phylum *Synergistetes* only recently recognized as a unique taxonomic class, meaning there is much basic groundwork to be done to more fully understand and be able to manipulate this organism. But we have begun.

2. Future Questions

Why does *A. hydrogeniformans* express the GAPDH protein at high levels if it is not used?

It is unknown why *A. hydrogeniformans* synthesizes abundant protein that is not in use under the current growth conditions. Indeed, under the three conditions here tested (glucose only, pyruvate only, and glucose + pyruvate) the GAPDH protein constitutes a whopping 5% of total cellular protein (table S1, chapter 5)! The enzyme activity studies, however, were only performed under glycolytic conditions; we have no enzyme data for growth under pyruvate. Therefore, we do not know if GAPDH becomes active under these gluconeogenic conditions, or, indeed, whether GAPOR is still active under same. There is precedence for a GAPDH only required under gluconeogenesis in archaea *Pyrococcus furiosus* and *Thermococcus kodakarensis*, as outlined below:

In 2011 Matsubara and colleagues studied the altered Embden-Meyerhof (EM) pathway in the thermophilic archaeon *Thermococcus kodakarensis*, which resembles that previously studied in fellow thermophiles like *Pyrococcus furiosus* (12-16). Members of the genus *Thermococcus* have in common with *Pyrococcus* the same alterations to the EM pathway: the unique ADP-utilizing glucokinase and phosphofructokinase, and the utilization of GAPOR in place of GAPDH and PGK (12). Prior to this knockout study it had been established that *gor* (GAPOR) and *gapN* (a non-phosphorylating NAD(P)-utilizing GAP (a.k.a. G3P) dehydrogenase (GAPN)) gene expression are up-regulated by a glycolytic regulator (*Tgr*, or Thermococcales glycolytic regulator) during growth on malto-oligosaccharides (glycolytic growth conditions), while *gapDH* and *pgk* are downregulated (17). The knockout studies showed that GAPOR and GAPN are not necessary for growth under gluconeogenic growth conditions, but that GAPDH and PGK are both necessary under same. Conversely, the reverse was also shown: GAPOR and GAPN are necessary for growth under glycolytic conditions, but not GAPDH or PGK. (12) This mirrors the growth requirements in *Pyrococcus furiosus*. The precise role of GAPDH in gluconeogenesis is unknown, but a hint may lie in the fact that the enzyme activity of GAPDH and PGK is bidirectional, whereas the enzyme activity of GAPOR proceeds only in the glycolytic direction (12, 15, 18).

There is yet no expression or knockdown system in *A. hydrogeniformans*. It would be difficult to perform the sort of knockdown performed above on *Thermococcus*. But, the enzyme activity studies can be repeated on cells grown on pyruvate, which encourages gluconeogenesis.

If GAPDH and PGK activity are detected in these cells, and GAPOR and GAPN activity is not detected, the similarities between the C_6 utilization pathways in *A. hydrogeniformans* and in these thermophilic archaea becomes even stronger.

How might *A. hydrogeniformans* OS1 generate NADH without an oxidative pentose phosphate pathway?

As discussed in chapter 5 and in the preceding section, *A. hydrogeniformans* OS1 has a glycolytic metabolism that resembles that found in thermophilic archaea, in that it utilizes GAPOR but not GAPDH, but there is GAPDH protein present. As *A. hydrogeniformans*'s glycolysis pathway resembles these archaea's, perhaps its method of generating NADH without an oxidative pentose-phosphate pathway is similar. The 2011 Matsubara study mentioned above that explored the altered Embden-Meyerhof pathway in archaeon *Thermococcus kodakarensis* also elucidated the method by which *T. kodakarensis* is able to produce NADH in absence of a pentose phosphate pathway (12). At the time of publication it had already been determined *T. kodakarensis* utilizes a reversed ribulose monophosphate pathway for nucleotide biosynthesis, which accounts for the products of the second half of the PPP (19, 20). For NADH synthesis, the study showed activity by a non-phosphorylating GAP-dehydrogenase (GAPN, not to be confused with the phosphorylating version, GAPDH). This arrangement, in which GAPN contributes to the cell by generating NADH for biosynthesis in absence of a pentose phosphate pathway, is described in detail in Appendix I.

A survey of the *A. hydrogeniformans* OS1 genome reveals that it, like *T. kodakarensis*, is missing genes for the oxidative pentose phosphate pathway (chapter 4). Perhaps given its similar modifications to the Embden-Meyerhof pathway it also is utilizing GAPN to produce NADH. The GAPN reaction is classified as EC 1.2.1.9 and a survey of known GAPN from archaea reveals that the only common pfam is that for a NAD(P)-utilizing aldehyde oxidoreductase (pfam00171). The phosphorylating GAPDH is distinguished by pfam00044 and pfam02800, highlighting its sequence divergence from its non-phosphorylating cousin. We can therefore be reasonably sure that the so-labeled GAPDH in *A. hydrogeniformans* is in fact that phosphorylating GAPDH, or, at the very least, unlikely to be a GAPN. There is one gene in *A. hydrogeniformans* OS1 that contains pfam00171 and therefore is a good candidate for a GAPN: it is at locus tag OS1_1087, and its protein expression levels are as follows:

Gene Product Name	**Locus Tag**	**Average Total Cellular Protein**		
		Glucose	**Pyruvate**	**Glucose & Pyruvate**
propionaldehyde dehydrogenase	OS1_1087	0.008%	0.019%	0.017%

Table 1: Potential GAPN in *A. hydrogeniformans* OS1. Gene product is predicted to be a propionaldehyde dehydrogenase, but the protein contains the pfam for an NAD(P)-utilizing aldehyde oxidoreductase (pfam00171).

This shows a 2.4-fold increase in the abundance of this protein when grown on pyruvate as compared to glucose alone, and a similar 2.1-fold increase in expression when grown on glucose and pyruvate as compared to glucose alone. This contradicts our model: if we expected a GAPN to be utilized during glycolysis, as it is in thermophilic archaea, we would expect to see the opposite pattern. This protein may not be a GAPN, or if it is, GAPN has a function under gluconeogenic conditions we did not anticipate. But, first, before putting effort into all this protein identification, it would be more time-effective to redo the enzyme assays under conditions that would sense GAPN activity. And that is precisely what we plan to do.

To detect potential GAPN activity, the enzyme assay would probably have to be re-performed with the cells grown in pyruvate. The assay was only done on glucose-grown cells. In the original enzyme assay GAPDH activity was assayed in two ways: via the conversion of glyceraldeyde-3-phosphate using NAD(P), and by use of fructose-1,6-bisP and aldolase to create G3P. Neither of these assays would differentiate between phosphorylating and non-phosphorylating activity, in any case (personal communication, Michael J. McInerney). In addition, the concentration of G3P used in the assay would probably have to be increased significantly to trigger GAPN activity. The McInerney lab used 0.4 mM G3P in their assays. In one study of GAPN activity, the researchers had to use 4 mM G3P to detect GAPN activity, a tenfold difference (21). Dr. McInerney has said that they plan to do some more enzyme studies this summer, including one on pyruvate-grown cells. I will petition for those studies to include the 4mM G3P tests for GAPDH/GAPN activity.

Does the *A. hydrogeniformans* OS1 GAPOR use tungsten or molybdenum?

Thus-far characterized GAPORs all utilize tungsten as a metal cofactor, with one exception: the non-phosphorylating GAPOR from *Methanococcus maripaludis*. GAPOR_{Mm} utilizes molybdenum, another periodic table group 6 element (atomic number 42 to tungsten's 74). It was found that recombinant GAPOR_{Mm} from *E. coli* had no enzyme activity when isolated from or when grown in a medium containing only tungsten or no metal cofactor, indicating three things: first, tungsten can out-compete molybdenum for a spot in the enzyme; second, tungsten-containing GAPOR_{Mm} is inactive; and, third, GAPOR_{Mm} requires molybdenum as a cofactor. (18) This also indicates that attempts to determine whether an enzyme utilizes tungsten or molybdenum based on the size of the ion-binding area (as relative to the atomic radius of the metals) predicted in a structure may well be moot, as one is able to easily replace the other in a competition. Indeed, in the case of GAPOR_{Mm}, it is the larger metal, and the metal that does *not* work (W) that replaces the metal that *does* work (Mo) in a competition, indicating that in this case at least there is not a mechanism for the precise filtering of incorporated ion based on size;

not only this, but that the metal that does *not* work is preferentially assimilated or can force out the metal that does actually work, as activity drops to zero. Alternatively, the inhibitory effect of tungsten may not be exerted by it slotting into the ion binding site at all; it may interfere on the periphery of the enzyme or in some other way hinder its action.

In absence of structural studies that would reveal the identity of the ion used in the GAPOR in *A. hydrogeniformans*, biochemical assays can be performed to determine its identity. A repeat of the experiment with an *E. coli*-expressed *A. hydrogeniformans* GAPOR, similar to the one performed on *M. maripaludis*, may reveal metal cofactor exclusivity, or lack thereof. *A. hydrogeniformans* could also be grown on a tungsten-limited or molybdenum-limited medium to compare viability. Currently, *A. hydrogeniformans* is grown in a medium that contains both tungsten and molybdenum (personal communication, Farzaneh Sedighian).

This utilization of molybdenum over tungsten is further evidence that the *M. maripaludis* GAPOR is unique when compared to those previously studied, not only in that its organism of origin differs greatly from those of other GAPOR-holders metabolically (non-phosphorylating) and environmentally, but that GAPOR$_{\mathrm{Mm}}$ *itself* is unique of heretofore-studied GAPOR. The four AFOR proteins in *A. hydrogeniformans* have closer homology to the tungsten-containing GAPOR from *Pyrococcus furiosus* than the GAPOR from *M. maripaludis* (chapter 5). But, the physiology of *A. hydrogeniformans* is closer to that of *M. maripaludis* in some ways than that of *P. furiosus*. Most evident is their optimum growth temperatures: *P. furiosus* grows best at 100°C (22), while *M. maripaludis* grows best at 38°C (23) and *A. hydrogeniformans* grows best at 50°C (9). There is the potential here for endless speculation, but it is all just an intellectual exercise in absence of concrete data. The next step, and an experiment that may answer many of these questions, is to determine the structure of the *A. hydrogeniformans* GAPOR.

Table 1. Distribution of core metabolic genes/proteins of the genus *Acetomicrobium*. Strains are: *Acetomicrobium flavidum* DSM 20664, *Acetomicrobium thermoterrenum* DSM 13490, *Acetomicrobium mobile* DSM 13181, and *Acetomicrobium hydrogeniformans* ATCC BAA-1850 (OS1). Orthologs were identified by pinned region analysis tools at IMG JGI.

	A. flavidum	***A. thermoterrenum***	***A. mobile***	***A. hydrogeniformans***
Electron flow				
AFOR (GAPOR-like)	**5**	6	5	4
OFOR (PFOR-like)	7	8	7	7
Hyd1-like	1	1	1	1
Hyd2-like	1	1	1	1
EtfAB	1	1	1	1
Fdh*	2	2	2	2

*Genes with homology to the Fdh major subunit not contained in the Hyd1 (hydrogen formate lyase) or Hyd2 clusters.

References

1. **Tang Y-Q, Ji P, Hayashi J, Koike Y, Wu X-L, Kida K.** 2011. Tang Y-Q, Ji P, Hayashi J, Koike Y, Wu X-L, Kida K.. Characteristic microbial community of a dry thermophilic methanogenic digester: its long-term stability and change with feeding. Appl Microbiol Biotechnol 91: 1447-1461, vol 91.

2. **do Cabo Fernandes C, Rechenberg D-K, Zehnder M, Belibasakis GN.** 2014. Identification of Synergistetes in endodontic infections. Microbial Pathogenesis **73:**1-6.

3. **Horz H-P, Citron DM, Warren YA, Goldstein EJC, Conrads G.** 2006. Synergistes Group Organisms of Human Origin. Journal of Clinical Microbiology **44:**2914-2920.

4. **Koyanagi T, Sakamoto M, Takeuchi Y, Maruyama N, Ohkuma M, Izumi Y.** 2013. Comprehensive microbiological findings in peri-implantitis and periodontitis. Journal of Clinical Periodontology **40:**218-226.

5. **Garcia-Vallvé S, Romeu A, Palau J. 2000**. **Horizontal Gene Transfer in Bacterial and Archaeal Complete Genomes. Genome Research 10:1719-1725**.

6. **Nelson KE.** 1999. **Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. Nature 27:323-329**

7. **Deppenmeier U, Johann A, Hartsch T, Merkl R, Schmitz RA, Martinez-Arias R, Henne A, Wiezer A, Baumer S, Jacobi C, Bruggemann H, Lienard T, Christmann A, Bomeke M, Steckel S, Bhattacharyya A, Lykidis A, Overbeek R, Klenk HP, Gunsalus RP, Fritz HJ, Gottschalk G.** 2002. The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. J Mol Microbiol Biotechnol **4:**453-461.

8. **Menes RJ, Muxi L.** 2002. *Anaerobaculum mobile* sp. nov., a novel anaerobic, moderately thermophilic, peptide-fermenting bacterium that uses crotonate as an electron acceptor, and emended description of the genus *Anaerobaculum*. Int J Syst Evol Microbiol **52:**157-164.

9. **Maune MW, Tanner RS.** 2012. Description of Anaerobaculum hydrogeniformans sp. nov., an anaerobe that produces hydrogen from glucose, and emended description of the genus Anaerobaculum. Int J Syst Evol Microbiol **62:**832-838.

10. **Ben Hania W, Bouanane-Darenfed A, Cayol J-L, Ollivier B, Fardeau M-L.** 2016. Reclassification of Anaerobaculum mobile, Anaerobaculum thermoterrenum, Anaerobaculum hydrogeniformans as Acetomicrobium mobile comb. nov., Acetomicrobium thermoterrenum comb. nov. and Acetomicrobium hydrogeniformans comb. nov., respectively, and emendation of the genus Acetomicrobium. International Journal of Systematic and Evolutionary Microbiology **66:**1506-1509.

11. **Thauer RK, Jungermann K, Decker K.** 1977. Energy conservation in chemotrophic anaerobic bacteria. Bacteriol Rev **41:**100-180.

12. **Matsubara K, Yokooji Y, Atomi H, Imanaka T.** 2011. Biochemical and genetic characterization of the three metabolic routes in Thermococcus kodakarensis linking glyceraldehyde 3-phosphate and 3-phosphoglycerate. Molecular Microbiology **81:**1300-1312.

13. **Heider J, Ma K, Adams MW.** 1995. Purification, characterization, and metabolic function of tungsten-containing aldehyde ferredoxin oxidoreductase from the

hyperthermophilic and proteolytic archaeon Thermococcus strain ES-1. Journal of Bacteriology **177:**4757-4764.

14. **Kengen SW, de Bok FA, van Loo ND, Dijkema C, Stams AJ, de Vos WM.** 1994. Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by Pyrococcus furiosus. J Biol Chem **269:**17537-17541.

15. **Mukund S, Adams MWW.** 1995. Glyceraldehyde-3-phosphate Ferredoxin Oxidoreductase, a Novel Tungsten-containing Enzyme with a Potential Glycolytic Role in the Hyperthermophilic Archaeon Pyrococcus furiosus. Journal of Biological Chemistry **270:**8389-8392.

16. **Kanai T, Imanaka H, Nakajima A, Uwamori K, Omori Y, Fukui T, Atomi H, Imanaka T.** 2005. Continuous hydrogen production by the hyperthermophilic archaeon, Thermococcus kodakaraensis KOD1. Journal of Biotechnology **116:**271-282.

17. **Kanai T, Akerboom J, Takedomi S, van de Werken HJG, Blombach F, van der Oost J, Murakami T, Atomi H, Imanaka T.** 2007. A Global Transcriptional Regulator in Thermococcus kodakaraensis Controls the Expression Levels of Both Glycolytic and Gluconeogenic Enzyme-encoding Genes. Journal of Biological Chemistry **282:**33659-33670.

18. **Park M-O, Mizutani T, Jones PR.** 2007. Glyceraldehyde-3-Phosphate Ferredoxin Oxidoreductase from Methanococcus maripaludis. Journal of Bacteriology **189:**7281-7289.

19. **Atomi H, Fukui T, Kanai T, Morikawa M, Imanaka T.** 2004. Description of Thermococcus kodakaraensis sp. nov., a well studied hyperthermophilic archaeon previously reported as Pyrococcus sp. KOD1. Archaea (Vancouver, BC) **1:**263-267.

20. **Orita I, Sato T, Yurimoto H, Kato N, Atomi H, Imanaka T, Sakai Y.** 2006. The Ribulose Monophosphate Pathway Substitutes for the Missing Pentose Phosphate Pathway in the Archaeon <em>Thermococcus kodakaraensis</em&gt. Journal of Bacteriology **188:**4698.

21. **Ettema TJG, Ahmed H, Geerling ACM, van der Oost J, Siebers B.** 2008. The non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN) of Sulfolobus solfataricus: a key-enzyme of the semi-phosphorylative branch of the Entner–Doudoroff pathway. Extremophiles **12:**75-88.

22. **Blamey JM, Adams MWW.** 1993. Purification and characterization of pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus.* Biochim Biophys Acta **1161:**19-27.

23. **Jones WJ, Paynter MJB, Gupta R.** 1983. Characterization of Methanococcus maripaludis sp. nov., a new methanogen isolated from salt marsh sediment. Archives of Microbiology **135:**91-97.

Appendix I: A Review of the AFOR and PFOR Enzyme Families

Contents

1. GAPOR (Glyceraldehyde-3-phosphate: ferredoxin oxidoreductase) and the Aldehyde:Ferredoxin Oxidoreductase (AFOR) Enzyme Family

NOTE: In this study, AOR will be used to refer to the broader family of aldehyde oxidoreductases, NAD(P)-utilizing variants included. If an AFOR is referenced from a paper that uses "AOR" to refer specifically to aldehyde:ferredoxin oxidoreductase, it will be in this study called "AFOR" for clarity.

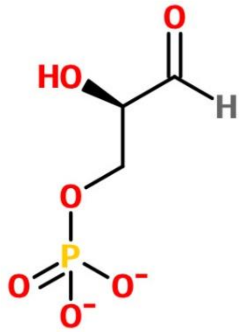

Figure 1 Glyceraldehyde-3-Phosphate (G3P or GAP). Drawn at emolecules.com.

GAPOR: Glyceraldehyde-3-phosphate: ferredoxin dehydrogenase

The first characterized GAPOR came from *Pyrococcus furiosus* (1), a hyperthermophilic archaeon that grows optimally at 100°C and ferments carbohydrates to produce acetate, H_2, and CO_2. In the presence of elemental sulfur, it also produces H_2S. It requires tungsten for growth, an element not commonly used in biological systems. (1-3)

At the time Mukund and colleagues began their study of *P. furiosus* GAPOR it had already been determined that its growth depended on tungsten, and two tungsten-containing, aldehyde-oxidizing, ferredoxin-utilizing enzymes, aldehyde:ferredoxin oxidoreductase (AFOR) and formaldehyde:ferredoxin oxidoreductase (FOR), had been purified. These were the first two discovered members of the AFOR family of enzymes. (Note that there is an AFOR *family* and the *enzyme* AFOR, which has a broad substrate range for aldehydes.) It had also been determined through a ^{13}C-glucose metabolism study that *P. furiosus* converted glucose via an Embden-Meyerhof pathway, and that substrate-level phosphorylation in this pathway utilized two kinases that required ADP, not ATP, as a substrate (4). Indeed, this was the first example of such kinases found, and lead the researchers to propose that they represented vestiges of an ancestral metabolism given *P. furiosus*'s great phylogenetic age, and contradicted earlier suggestions (5) that *P. furiosus* fermented sugars via a non-phosphorylating Entner-

Doudoroff pathway (pyroglycolysis). Discussion of these unique ADP-utilizing kinases moves beyond the scope of this particular study.

While the Kengen studies (4) indicated that that *P. furiosus* utilizes an Embden-Meyerhof pathway, enzyme activity studies of cell-free extracts from *P. furious* showed minimal glycolytic activity of two key enzymes in the common EM pathway: glyceraldehyde-3-phosphate dehydrogenase (GADPH) and phosphoglycerate kinase (PGK). However, under gluconeogenerative conditions (grown on pyruvate), the activities of these enzymes spike (6). Starting from this framework, Mukund et al found activity from a novel AFOR-family, tungsten-containing, ferredoxin-reducing enzyme proposed to act in place of GADPH *and* PGK in the *Pyrococcus* glycolytic pathway, and that enzyme was GAPOR.

The *P. furiosus* GAPOR is a monomeric, O_2-sensitive, 63kDa (on SDS-PAGE) enzyme which contains a pterin group and one tungsten and six iron atoms per molecule. It oxidized glyceraldehyde-3-phosophate (G3P) and reduced *P. furiosus* ferredoxin, and it did NOT oxidize formaldehyde, acetaldehyde, glyceraldehyde, benzaldehyde, glucose, glucose-6-phosphate, or glyoxylate, and it did NOT reduce NAD(P). (1) Unlike the previously-discovered AFOR and AFOR-family enzymes it had a narrow substrate range (G3P only). Though the reactions of GADPH and PGK are reversible, the conversion of G3P to 3-PGA by GAPOR is physiologically irreversible (1). An SDS-PAGE experiment that involved boiling the purified enzyme at 100°C for 10 minutes testified to the enzyme’s exceptional heat stability: there were two bands on the gel, one at 44kD (the folded protein) and one at 63kD (the linearized protein) (1). The fact that *any* folded enzyme remained after such harsh treatment is remarkable.

In differentiating GAPOR from other members of the AFOR family, and from other G3P-oxidizing enzymes, Mukund et al made this observation:

> "The presence of a single subunit was confirmed by N-terminal amino acid sequence analysis of a solution of GAPOR which gave rise to a single sequence (M K F S V L K L D V G K R E V E A Q E I E R E D I F G V V D Y G I M R H N E). This showed no homology to the N-terminal sequences of either A[F]OR or FOR from *P. furiosus* [(7)] or to GAPDH from *Pyrococcus woesei* [(8)]." (1)

As of April 17, 2018, a protein BLAST at NCBI of the aforementioned 38-residue sequence returned only GAPOR from three strains of *P. furiosus* as a 100% match, with e-values of 4e-18 to 4e-16. The next-best match represents a significant jump down in both e-value and percent identity: GAPOR from *Thermococcus chitonophagus*, at 74% identity and e-value of 1e-10. Of the top 45 BLAST hits, all are of the N-terminus of predicted GAPOR sequences from hyperthermophilic archaea. This N-terminal sequence, therefore, is probably not of use in identifying GAPORs in other organisms, especially in organisms other than hyperthermophilic archaea.

Twelve years after the *P. furiosus* GAPOR was characterized, in 2007, another GAPOR was characterized from the non-sugar-assimilating mesophilic archaeon *Methanococcus maripaludis* (9). Unlike the GAPOR from *P. furiosus* (henceforth GAPOR_{Pf}), which was isolated directly from its parent organism, the *M. maripaludis* GAPOR studied (henceforth GAPOR_{Mm}) was a recombinant from *E. coli*. Also unlike the GAPOR from *P. furiosus*, it is non-phosphorylating. *M. maripaludis* differs from the sugar-fermenting *P. furious* in that it has never been reported to assimilate any extracellular sugars, and also in that it is a mesophile, whereas *P. furiosus* is a thermophile. It is surprising that an organism not expected to make much use of the glycolysis pathway would have glycolytic enzymes. The reliance upon other sources of energy may explain why *M. maripaludis* can survive with a non-phosphorylating GAPOR.

Like *P. furiosus*, *M. maripaludis* also contains genes for NAD-dependent phosphorylating and nonphosphorylating glyceraldehyde-3-phosphate dehydrogenases (GAPDH and GAPN, respectively), which can catalyze part of the same reaction as GAPOR in the glycolytic pathway (9). Also like *P. furiosus*, GAPDH is only active during gluconeogenesis, though the roles of GAPOR and GAPDH in *M. maripaludis* have not yet been determined (9). The non-phosphorylating GAPOR in *M. maripaludis* may function in a manner analogous to the GAPN (non-phosphorylating GAPDH) in other hyperthermophiles in that it is involved in the production of some key metabolite for the cell, as it is not involved in substrate-level phosphorylation. This role for GAPN as modeled by *Thermococcus kodakarensis* is further outlined later in this section.

Park and colleagues note that as of 2007 *M. maripaludis* was the only non-hyperthermophile predicted to contain a GAPOR (9); *A. hydrogeniformans*, and many other *Synergistetes*, were not sequenced until after this observation, and of note Park et al did not specify the parameters of their search (archaea only, all prokaryotes, etc). A BLAST of the GAPOR-specific N-terminal sequence from *P. furiosus* (1) was performed on four *M. maripaludis* genomes (c5, c6, c7, x1) and returned no matches. This is not surprising as, noted above, the only BLAST hits returned of this Pyrococcus N-terminal sequence (of the entire NCBI Protein Database) are of GAPOR in hyperthermophlic archaea.

Also unlike other GAPORs from hyperthermophiles, in place of tungsten GAPOR_{Mm} utilizes molybdenum, another periodic table group 6 element (atomic number 42 to tungsten's

74). Indeed, it was found that GAPOR_{Mm} extracts had no enzyme activity when isolated from *E. coli* grown in a medium containing tungsten as well as molybdenum, or when grown in a medium containing only tungsten or no metal cofactor, indicating three things: first, tungsten can out-compete molybdenum for a spot in the enzyme; second, tungsten-containing GAPOR_{Mm} is inactive; and, third, GAPOR_{Mm} requires molybdenum as a cofactor. (9) This also indicates that attempts to determine whether an enzyme utilizes tungsten or molybdenum based on the size of the ion-binding area (as relative to the atomic radius of the metals) may well be moot, as one is able to easily replace the other in a competition. Indeed, in the case of GAPOR_{Mm}, it is the metal that does *not* work (W) that replaces the metal that *does* work (Mo) in a competition, indicating that in this case at least there is not a mechanism for the precise filtering of incorporated ion based on size; not only this, but that the metal that does *not* work is preferentially assimilated or can force out the metal that does actually work, as activity drops to zero. This utilization of molybdenum in place of tungsten, along with its non-phosphorylating activity, is further evidence that the *M. maripaludis* GAPOR itself is unique when compared to those previously studied, not only in that its organism of origin differs greatly from those of other GAPOR-holders metabolically and environmentally.

The first oxygen-tolerant GAPOR was discovered in microaerobic hyperthermophile *Pyrobaculum aerophilum*. GAPOR had thus far only been found in anaerobic organisms, and had been found to be oxygen-sensitive, but the GAPOR of *P. aerophilum* remained active at an atmosphere of O_2 1.5% (10).

In 2011 Matsubara and colleagues studied the altered Embden-Meyerhof (EM) pathway in the thermophilic archaeon *Thermococcus kodakarensis*, which resembles that previously studied in fellow thermophiles like *Pyrococcus furiosus* (1, 4, 11-13). Members of the genus *Thermococcus* have in common with *Pyrococcus* the same alterations to the EM pathway: the unique ADP-utilizing glucokinase and phosphofructokinase, and the utilization of GAPOR in place of GAPDH and PGK (11). Prior to this knockout study it had been established that *gor* (GAPOR) and *gapN* (a non-phosphorylating NAD(P)-utilizing GAP (a.k.a. G3P) dehydrogenase (GAPN)) gene expression are upregulated by a glycolytic regulator (*Tgr*, or Thermococcales glycolytic regulator) during growth on malto-oligosaccharides (glycolytic growth conditions), while *gapDH* and *pgk* are downregulated (14). The knockout studies showed that GAPOR and GAPN are not necessary for growth under gluconeogenic growth conditions, but that GAPDH and PGK are both necessary under same. Conversely, the reverse was also shown: GAPOR and GAPN are necessary for growth under glycolytic conditions, but not GAPDH or PGK. (11) This mirrors the growth requirements in *Pyrococcus furiosus*.

T. kodakarensis GAPN exhibits relatively low levels of activity unless intracellular G3P concentrations are high, and shows a strong allosteric response to glucose-1-phosphate (as opposed to glucose-6-phosphate), which, while not its substrate, acts as an activator and regulator (11). The researchers found it curious that *both* GAPOR and GAPN would be required for glycolytic growth, as they both catalyze the same reaction (G3P to 3-PGA). They hypothesize that the complementarity is found in their differing electron acceptors (ferredoxin for GAPOR and NAD+ for GAPN) (11).

In T. *kodakarensis* oxidized ferredoxin is readily re-generated by the reduced ferredoxin offloading its electron to a hydrogenase, whereas the regeneration of NAD from NADH is low, which leads the researchers to conclude that there is no fermentative metabolism utilizing NAD as an electron carrier in *T. kodakarensis*. This hypothesis is further bolstered by the observation that protein fermentation in *T. kodakarensis* requires the addition of an electron acceptor such as elemental sulfur to the growth medium. This necessity for addition of an electron acceptor suggests that the intrinsic capacity of *T. kodakarensis* to regenerate NAD is not sufficient to sustain growth, as a key step in amino acid fermentation requires use of an electron acceptor, and if there is not enough NAD available, the cell will not be able to harvest enough energy to divide without outside help in the form of adding another electron acceptor. Now, when applied to the glycolytic conditions tested (no amino acid fermentation), this same lack of an NAD-regeneration mechanism would hinder GAPN's ability to provide sufficient energy for the cell, and so it relies upon GAPOR to pick up the energetic slack with its easily-regenerated ferredoxin.

This, of course, begs the question: why would GAPN be necessary at all under glycolytic growth conditions if this is the case? GAPOR is the enzyme performing substrate-level phosphorylation at that step, so what role would a less efficient enzyme that cannot even generate ATP play? The answer may lie in the fact that in oxidizing G3P GAPN generates NADH, which is necessary for various forms of biosynthesis. Like many archaea, *T. kodakarensis* lacks the oxidative pentose phosphate pathway, which generates NADH, and so must generate NADH by another mechanism. If this hypothesis is correct, GAPN would be acting more as an NADH generator than an energy harvester. There is also the possibility that in the absence of GAPN too much G3P accumulates, and GAPOR has been reported to be "seriously inhibited" by high G3P concentrations. In that case, GAPN would also be acting to prevent GAPOR from being inhibited by keeping G3P concentrations at an ideal level. (1, 11, 13)

T. kodakarensis GAPOR was only studied in the context of determining the effect of a predicted *gor* gene knockout on growth; the protein was not purified and studied, and a metal cofactor was not determined (11). It is therefore listed as a putative GAPOR based on the action of homologous genes in *Pyrococcus* and the fact that its knockout inhibited glycolytic growth, and that GAPOR in *Pyrococcus* was shown to be necessary for glycolytic growth. These two circumstantial factors (high sequence homology and the same impact upon glycolysis) give reasonable confidence that the gene knocked out was in fact GAPOR.

AFOR: Aldehyde: ferredoxin oxidoreductase

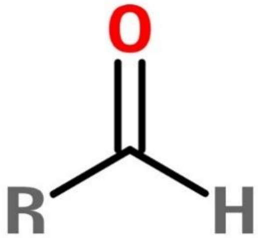
Figure 5: An aldehyde. Drawn at emolecules.com.

AFOR, aldehyde ferredoxin oxidoreductase, enzyme has the broadest substrate specificity of the AFOR-family enzymes, but it is most active with aldehydes derived from amino acids (3, 12, 15). Through a process-of-elimination study comparing oxidation activities, researchers also concluded that peptide-fermenter *T. litoralis* contains a ‘minor amount’ of an AFOR-like enzyme in addition to an FOR, but further study was not pursued (7). As of May 2, 2018 the only crystal structure for the AFOR enzyme in the RCSB Protein Databank comes from *Pyrococcus furiosus*(16).

FOR: Formaldahyde: ferredoxin oxidoreductase

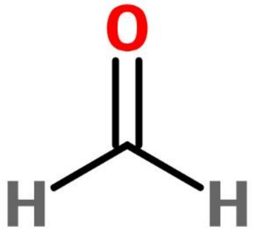
Figure 6: Formaldehyde. Drawn at emolecules.com.

FOR, formaldehyde ferredoxin oxidoreductase, has a narrower substrate specificity window than AFOR. Along with the eponymous formaldehyde, FOR has the most activity on C_4 to C_6 semi- and dialdehydes (17).

Mukund et al first described the FOR from the peptide-fermenting archaeal hyperthermophile *Thermococcus litoralis* in 1993 (7). The enzyme is so-named because formaldehyde was used in the initial activity assays and was oxidized at an apparent K_m of 62 mM. However, it is hypothesized that formaldehyde is not the physiological substrate for FOR_{Tl}, but, rather, that it *in vivo* functions in some capacity in peptide fermentation. (7) The researchers were able to differentiate this 'new' enzyme from the previously-discovered AFOR enzyme in *P. furiosus* by FOR_{Tl}'s inability to also oxidize crotonaldehyde, whereas AFOR can oxidize both crotonaldehyde and formaldehyde (7). Later, purification and assays were performed on the *P. furiosus* FOR enzyme, and by comparison further refinements made to knowledge of the now-recognized-as-unique AOR enzyme (17). FOR_{Pf} oxidized formaldahyde well with a K_m of 15 mM, but, as is the case with *T. litoralis*, the physiological substrate is thought to be different: in this case, an aliphatic C_5 semi- or dialdehyde, such as glutaric dialdehyde, which was oxidized at a K_m of 1 mM. These FOR_{Pf} K_m values are for the purified FOR activated with HS^-; the activity of FOR was raised fivefold by treatment with HS^- under reducing conditions. Roy et al came upon this discovery by noticing that the initially-purified FOR had low enzymatic activity, and they hypothesized, and tested to be correct, that this was due to a loss of sulfide. (17) As *P. furiosus* is a carbohydrate fermenter, and not an obligate peptide-fermenter like *T. litoralis* (i.e. *T. litoralis* cannot grow on carbohydrates alone and peptides are an absolute growth requirement (7)), it is hypothesized that FOR serves a different metabolic role in *T. litoralis* than in *P. furiosus*. The only crystal structures for the FOR enzyme (glutarate-complexed and naked) come from *P. furiosus*(18).

WOR4: Tungsten-containing aldehyde:ferredoxin oxidoreductase 4

WOR4 protein was only found in *P. furiosus* grown with elemental sulfur (S^0). (Compare this with FOR, which experiences a fivefold increase in activity when treated with HS^- (17), although it is not known how sulfur compounds effect protein expression levels.) Although WOR4 has been placed in the AOR family due to protein sequence homology to other members, it was not found to oxidize any aliphatic or aromatic aldehydes or hydroxy acids, or to reduce any keto acids (2). Cold shock mRNA expression studies of *P. furiosus* grown at 72°C revealed that the wor4 transcript is upregulated in cells that have acclimated to the cold over many generations (data originally reported in (19); attention called to wor4 upregulation in (20)). Its precise role in cold shock is unknown.

WOR5: Tungsten-containing aldehyde:ferredoxin oxidoreductase 5

The fifth tungsten-containing aldehyde:ferredoxin oxidoreductase discovered in *P. furiosus* was found to have a broad substrate specificity, but on a different subset of molecules than that catalyzed by AFOR. While AFOR has a preference for aldehydes derived from amino acids (12, 21), WOR5 has a high affinity for substituted and nonsubstituted aliphatic and aromatic aldehydes, with the strongest affinity for hexanaldehyde ($V_{\max}$ = 15.6 U/mg, K_m = 0.18 mM at 60°C with methyl viologen). (20) In the same *P. furiosus* cold shock study mentioned in the section on WOR4, wor5 transcripts were shown to be upregulated fivefold in cells 1-5hrs after culture temperature was dropped to 72°C (19). Indeed, a temperature-dependent study of WOR5 activity on hexanaldehyde reveals that its activity drops precipitously at temperatures higher than 80°C, further supporting its suggested role as a cold shock-adaptation enzyme (20). As for WOR4, the role of WOR5 in cold adaptation is unknown.

Adjacent to wor5 on the genome was an ORF coding for a 19kDa protein predicted to have multiple iron-sulfur cluster binding sites that was upregulated to the same degree as wor5. The researchers who first isolated WOR5 suggest that ORF is co-regulated with WOR5 and that the protein resulting acts as an electron acceptor for WOR5, as ferredoxin does for other members of the AFOR family. A cyclic voltammogram of *P. furiosus* ferredoxin showed a reversible electron transfer from the ferredoxin when it was placed with WOR5 and hexanal at 60°C, indicating that ferredoxin, also, may be a physiological redox partner for WOR5. (20)

Native and SDS-PAGEs suggest that WOR5 is a homodimer of a 67kDa subunit. Its sequence indicates a loss of one of four highly conserved cysteine residues in its 4Fe-4S iron-sulfur cluster binding sites. One would expect this to mean that it binds instead a 3Fe-4S cluster instead, but the electron paramagnetic resonance spectrum of the enzyme unambiguously indicates a 4Fe-4S cluster. An aspartate residue in place of the 'lost' cysteine is predicted to act as a ligand in this case. (20)

'False' positives for finding bacterial GAPORs and AFORs in papers; or, the world of difference contained in one letter

The nomenclature and classification of enzymes considered "aldehyde oxidoreductases" has been refined since many of the initial studies were done in the early 90s, and this can lead to confusion in classifying enzymes in old papers by current definitions. A 1992 paper discussed a "tungsten-containing aldehyde oxidoreductase" from bacterium *Clostridium thermoaceticum* (now known as ***Moorella thermoacetica*****), a trimeric protein that "**reduce[s] reversibly non-activated carboxylic acids to the corresponding aldehydes" with NADP (22), instead of ferredoxin. It is classified as an "AOR" in the original and in some subsequent citing papers. This is the same acronym used for the enzyme family of aldehyde oxidoreductases that utilizes ferredoxin, and it has been cited in papers since as an example of a member of that ferredoxin-utilizing enzyme family. When citations of this first instance are chained away from the original paper, merely saying that ***M. thermoacetica*** contains an "AOR", the reader comes away with the misconception that there is a ferredoxin-utilizing AFOR in this bacterium. This would be a gross error if one is evaluating the phylogenetic spread of this enzyme family. Indeed, it contains a tungsten cofactor, which is unusual in biological systems, and it acts upon the same substrate, but the electron acceptor is different. Later papers that mark the difference omit this bacterial NAD(P)-utilizing AOR from AFOR lists.

A 2014 version of the conserved domain database for NCBI for the domain "glyceraldehyde-3-phosphate ferredoxin oxidoreductase, GAPOR N-terminal [Pyrococcus furiosus, DSM 3638, Peptide Partial, 38 aa]" (GenBank: AAB34742.1), the aforementioned N-terminal GAPOR tag for hyperthermophile GAPORs, cited the existence of a "hydroxycarboxylate viologen oxidoreductase from *Proteus vulgaris*, the sole member of the AOR family containing molybdenum" (citing source (23)). This seems a very exciting note, as it claims that not only is there is an "AOR" enzyme (implied, by its presence on the GAPOR page, that this AOR family is the ferredoxin-utilizing aldehyde oxidoreductase variety) in a bacterium, but also one that utilizes molybdenum. However, closer inspection of the cited 1994 paper reveals that this *Proteus* enzyme reduces 2-oxocarboxylates, not aldehydes. It *is* similar to the ferredoxin-utilizing AOR family in that it does not utilize NAD(P); rather, in *in vitro* experiments in was able to use a range of electron mediators. (23) It is also worth noting that

> The amino acid sequence of the N-terminal of HVOR… shows significant similarity from position 4 to 15 with that of the tungsten containing [NADPH-utilizing] aldehyde oxidoreductase from *Clostridium thermoaceticum* [***Moorella thermoacetica*]** (22) from position 3-14 (8 of 12 are identical and 4 of 12 are conserved substitutions). Comparing the same range of sequence of HVOR with positions 3 - 14 of tungsten-containing [ferredoxin-utilizing] aldehyde oxidoreductase from *Thermococcus litoralis* (7) there are 8 of 12 identical and 2 of 12 conservatively substituted amino acid residues. (23) *brackets my clarifications; citations updated to reflect citations list for this paper*

This is an interesting similarity between these three similar enzymes of differing electron acceptors and may indicate common descent. But, if the researcher is to stick to the definition of

"AFOR" family in that it must utilize ferredoxin AND reduce aldehydes, this HVOR from *Proteus vulgaris* does not qualify.

One also finds statements such as this in a 2003 paper on a "tungsten-containing aldehyde oxidoreductase of *Eubacterium acidaminophilum*":

> The tungsten-containing AOR family is subdivided into aldehyde oxidoreductase/aldehyde dehydrogenase (AOR) (3, 12, 16, 22, 24, 25), formaldehyde oxidoreductase (FOR) (7, 18), glyceraldehyde-3-phosphate oxidoreductase (GAPOR) (1) and carboxylic acid reductase (CAR) (26, 27). These enzymes have mainly been isolated from hyperthermophilic archaea like *Pyrococcus furiosus* or *Thermococcus litoralis* and from acetogenic bacteria like *Moorella thermoacetica* and *Clostridium formicoaceticum*. (28) *citations updated to reflect citations list for this paper*

In this case, the "tungsten-containing AOR family" incorporates both NAD(P)- (*Moorella thermoacetica* (22)) and ferredoxin-utilizing species. The *Desulfovibrio gigas* study did not determine the physiological electron acceptor, but successfully used benzylviologen as an artificial electron acceptor while catalyzing a variety of aldehydes—acetaldehyde, propionaldehyde, and benzaldehyde being the three strongest—and found enzyme samples utilizing NAD and NADP to be inactive (25). Its N-terminal sequence showed 'no similarity' with other W-AORs and for this point the paper cited the *T. litoralis* AFOR study (7, 28) from 1993. The 1995 *D. gigas* study (25) also referenced a 1993 study of an earlier-discovered DCPIP-dependent molybdenum-containing aldehyde oxidoreductase from *D. gigas* (29), the latter of which was not directly cited in the 2003 *E. acidaminophilum* study (28). It described a DCPIP-dependent, non-NAD(P)-utilizing molybdenum aldehyde oxidoreductase (also called in study a "molybdenum hydroxylase") (29). Not cited under this study as an example of GAPOR (as it had not yet been characterized) is the molybdenum-containing *M. marpialudis* GAPOR (9).

Furthermore, this is a 'non-phosphorylating GAPOR', making it functionally equivalent to a GAPN, but with a ferredoxin electron acceptor.

2. PFOR (Pyruvate:Ferredoxin Oxidoreductase) and the OFOR (2-oxoacid:ferredoxin oxidoreductase) Enzyme Family

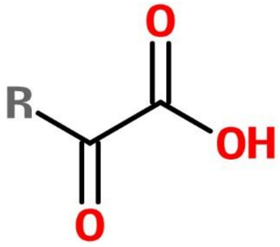

Figure 7: A 2-oxoacid (aka 2-ketoacid or α-ketoacid). Drawn at emolecules.com.

In most aerobic bacteria and mitochondria, the breakdown of 2-oxoacids (also known as α-ketoacids) such as pyruvate is handled by large multi-enzyme complexes called 2-oxoacid dehydrogenases. In anaerobic bacteria and archaea, as well as some amitochondrial eukaryotes, these reactions are catalyzed by the smaller 2-oxoacid: ferredoxin oxidoreductases. The dehydrogenases utilize NAD(P), while, as name would imply, the OFORs utilize ferredoxin. These electrons sequestered by the ferredoxins are low-potential (typically less than -500mV) and can be utilized in the reduction of difficult-to-reduce compounds such as H^+ (to H_2), CO_2, N_2, and sulfur and aromatic compounds (30). A few aerobic organisms, as well, utilize the OFORs: of note, nitrogen-fixing cyanobacteria (aerobes) of the genus *Anabaena* utilize OFORs for their 2-oxoacid metabolism needs (31), along with a few specific examples to be discussed in greater detail in subsequent sections.

The OFOR enzymes utilize thiamine pyrophosphate (TPP) to cleave carbon-carbon bonds; specifically, this enzyme family cleaves the bond at the α carbon and integrates coenzyme A with the R-keto group, while liberating CO_2 and two electrons (figure 8).

Figure 8: The general reaction schema for members of the OFOR enzyme family.

OFORs utilize [4Fe-4S] clusters for electron capture. These centers allow the OFORs to capture lower-potential electrons; but, these Fe-S centers can only process one electron at a time (as opposed to FAD, which processes two). Gibson et al neatly summarize how TPP is an ideal cofactor for these enzymes (internal sources updated for this bibliography):

> The one-electron nature of [4Fe-4S] cluster redox chemistry, as opposed to FAD or NAD(P)$^+$, which undergo two-electron redox cycles, means that OFORs must make use of a catalytic cofactor that is both capable of activating the substrate 2-oxoacid, as well as undergoing one-electron radical transfer. This cofactor is thiamine pyrophosphate (TPP), an organic cofactor that in its active form contains a nucleophilic carbanion [(32)] and is capable of stabilizing a 1-electron oxidized radical species. These properties make TPP well suited for converting the two-electron currency of bond oxidation into the one-electron currency recognized by [4Fe-4S] clusters. (33)

The 'active' form of TPP above mentioned is a yilde, or a zwitterion with two adjacent atoms with opposite charges and with a net charge of zero. This yilde form contains the carbanion in a thiazole ring (figure 9-B) and is in equilibrium (figure 9-C) with the cationic form (figure 9-A).

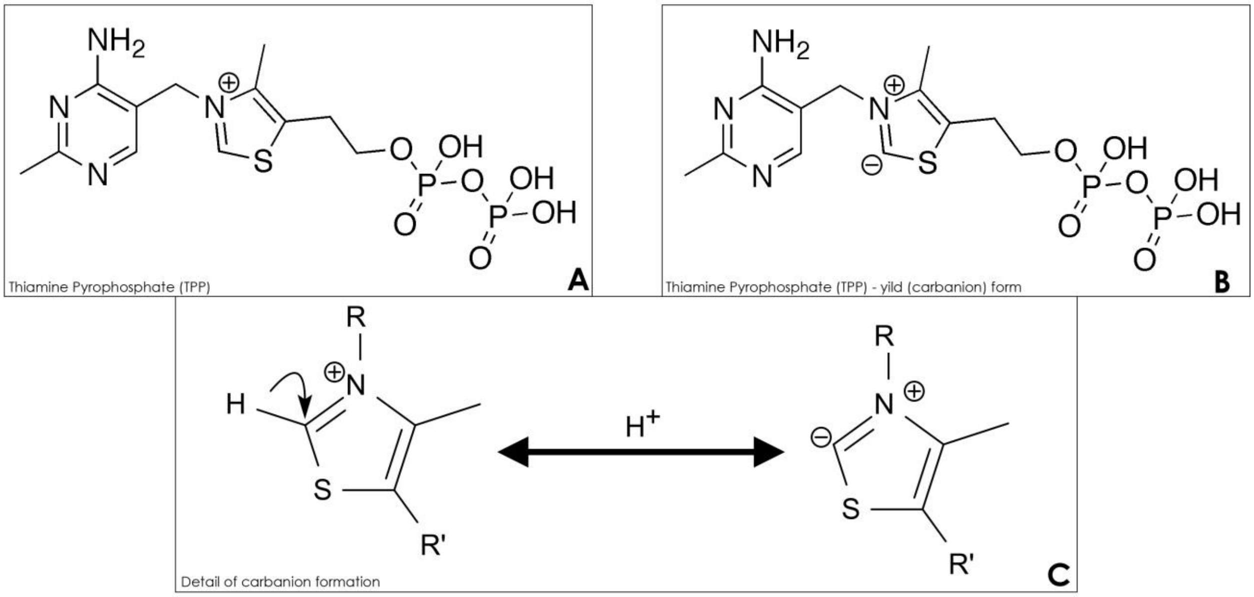

Figure 9: Thiamine pyrophosphate (TPP) (A), with detail of carbanion form (B) and abbreviated equilibrium diagram of carbanion center (C). Figures 6A and 6B modified from an original diagram by Wikipedia user Edgar181 (legal name undisclosed), released into public domain under a creative commons license. Figure 6C modified from a figure by Wikipedia user Marshall Strother (username Mcstrother), released into public domain under a creative commons license.

Unlike the 2-oxoacid dehydrogenases, the 2-oxoacid ferredoxin oxidoreductases tend to catalyze reversible reactions (i.e. they can also fix carbon dioxide), with some exceptions, such as the oxalate ferredoxin oxidoreductase (OOR) and the pyruvate decarboxylase (PDC) function of the archaeal dual-function PFOR/PDC enzyme. Indeed, the OOR and PDC-function also catalyze a completely different reaction than the one featured in figure 8, as will be elaborated in their respective sections. But, figure 8 outlines the most common reaction schema for OFOR-family enzymes, including pyruvate: ferredoxin oxidoreductase (PFOR). This ability to fix carbon dioxide, a greenhouse gas, has made OFORs an attractive target of study for researchers concerned with environmental microbiology and other methods of atmospheric clean-up (30); this is in addition to the already-discussed interest in OFORs as liberators of low-potential electrons for efficient metabolisms that can be co-opted for efficient biofuel production.

The mechanism of action of the OFORs involves the formation and breakdown of a stable free-radical intermediate (34). TPP is necessary for the stabilization of this radical intermediate, given its unique chemistry (figure 9).

The group that did the pioneering studies on the OOR (Gibson et al, as quoted above regarding TPP), which will be further described below, wrote an excellent review article in 2016 on this family of enzymes (35). It served as a much-needed update to the until-then definitive review by Stephen Ragsdale published in 2003 (36). Interest in these enzymes spans decades, but there has been a burst of recent discoveries that greatly expand our understanding of the family. Oxalate oxidoreductase (OOR) is the newest member of this ferredoxin-utilizing 2-oxoacid oxidoreductase family and was defined in 2010 (37) and crystalized in 2015 (33, 38), both from bacterium *Moorella thermoacetica*. 2-oxoacid: ferredoxin oxidoreductase (OFOR) was crystalized in 2016 from archaeon *Sulfolobus tokodaii* (34), twenty years after it was first described in same (then *Sulfolobus* sp. Strain 7) (39) and thirty-five years after it was described in archaeon *Halobacterium salinarum* (then being called a PFOR, though it works on many 2-oxoacids) (31).

A comparative analysis of the members of the OFOR family reveals significant structural similarity, despite the variety of holoenzyme structures. The domain labeling originates with that given to the homodimeric *Desulfovibrio africanus* PFOR crystalized in 1999. As is the case with the AFOR family of enzymes, the OFOR family takes its name from the broad-spectrum substrate-catalyzing member of its family.

PFOR: Pyruvate: Ferredoxin Oxidoreductase

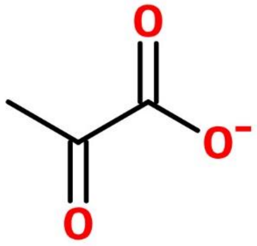

Figure 10: Pyruvate. Drawn at emolecules.com.

The first PFOR was crystalized from the bacterium *Desulfovibrio africanus* in 1999 (40, 41). Prior to this, researchers had discovered that PFOR, unlike pyruvate dehydrogenase (PDH), utilized ferredoxin as an electron acceptor, as opposed to NAD(P), and, again unlike PDH, it catalyzed a reversible reaction: it could also act as a pyruvate *synthase*. Whereas PDH is part of a massive multi-enzyme complex (the PDC, or pyruvate dehydrogenase complex), PFOR is a single enzyme. It is oxygen-sensitive and thus-far found only in anaerobes, whereas PDH is not oxygen-sensitive and is found in aerobic organisms.

PFOR has attracted attention as a potential antimicrobial drug target. As humans (and indeed all aerobic organisms) use PDH for their own pyruvate-metabolizing needs, anaerobic microbes that use PFOR can in theory be targeted via this enzyme with little harm to the host. Many of the studies on *Helicobacter pylori*'s PFOR were driven by this interest, for example. PFOR is also of interest to the pharmaceutical industry as a means to generate acetyl-CoA, a critical component in many drug-manufacturing processes. Indeed, the study of the structure and function of *Citrobacter sp. 77* PFOR was influenced by this need, wherein researchers tested the feasibility of *in vitro* industrial-scale acetyl-CoA manufacture by an array of PFORs (42). There is clearly ample economic and humanitarian impetus to further study this enzyme.

Bacterial PFORs

The *Desulfovibrio africanus*, *Moorella thermoactica,* and *Citrobacter sp. S-77* PFORs are homodimers, a configuration thus far only found in bacteria, whereas the PFORs in *Helicobacter pylori* and *Thermotoga maritima* are a dimer of heterotetramers, resembling those found in archaea. The *D. africanus* PFOR was the first crystalized and from this study was derived the domain labeling scheme still in use in discussing the structure of OFORs.

Archaeal PFOR/PDC Dual-Function Enzyme and Composite *por/vor* Transcription Unit

The archaeal ‘PFOR’ as studied in hyperthermophilic archaea *Pyrococcus furiosus* and *Thermococcus guaymasensis* is a dual-function enzyme: it combines the activity of PFOR, which catalyzes the reversible conversion of pyruvate to acetyl-CoA, and the activity of pyruvate decarboxylase (PDC), which irreversibly converts pyruvate to acetaldehyde. The holoenzyme is a heterodimer of heterotetramers, resembling the PFORs found in bacteria *Helicobacter pylori* and *Thermotoga maritima*. The gene order resembles same, with an important distinction, as will be outlined below. Both modes of activity for this enzyme are oxygen-sensitive. In *T. guaymasensis*, the time required for a 50% loss of PFOR activity (relative to activity at time zero when no air was introduced, at which point the sample was uncorked and stirred) was about 40 minutes at 4°C, and the time required for a 50% loss of PDC activity was about 30 minutes at same (43). (The researchers note, though, that due to technical difficulties these tests were not performed on the same batch of enzyme, so the difference in time-to-deactivation may merely reflect experimental technicalities, and their candor and honesty is appreciated.)

The bifunctional PDC/PFOR was first isolated in *P. furiosus* (15) and then from *T. guaymasensis* (43), the latter during a survey of the ethanol generation pathway. Acetaldehyde is a key component of the ethanol-production pathway in *T. guaymasensis*. Prior to this study the alcohol dehydrogenase converting acetaldehyde to ethanol in *T. guaymasensis* had been identified, but researchers had found no evidence of then-known enzymes for catalyzing the *formation* of acetaldehyde, leaving a hole in this metabolic pathway.

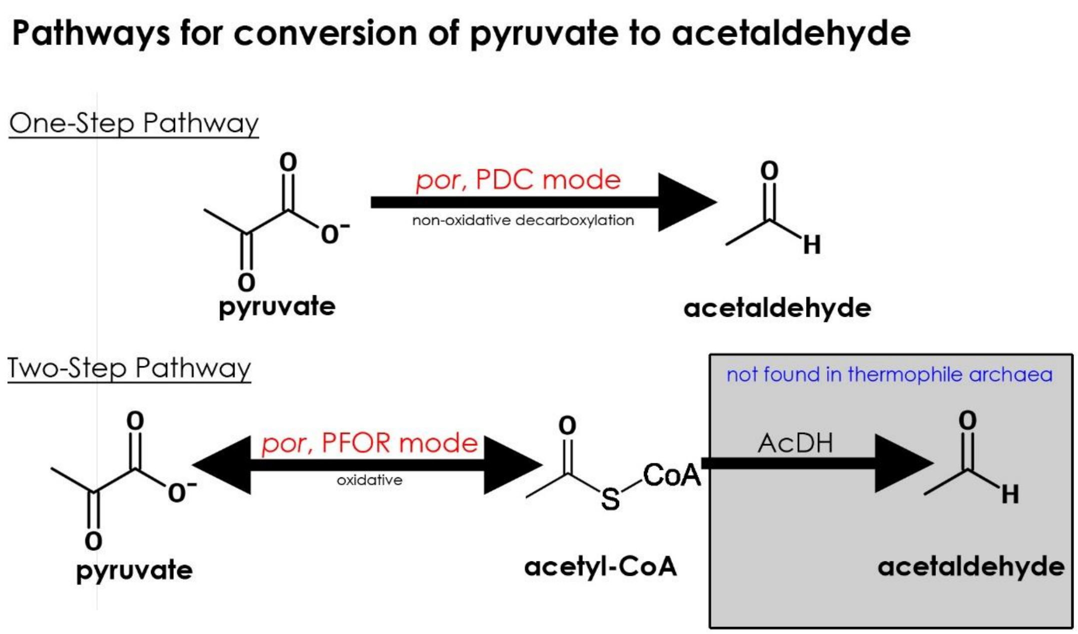

Figure 11: A comparison of the two methods available for converting pyruvate to acetaldehyde.

There are two known pathways for the generation of acetaldehyde from pyruvate: the one-step pathway utilizes 1) a pyruvate decarboxylace (PDC), and the two-step pathway involves 1) pyruvate being converted to acetyl-CoA either by PFOR or an NADH-utilizing pyruvate dehydrogenase (PDH), and then 2) acetyl-CoA being converted to acetaldehyde by a CoA-acetylating acetaldehyde dehydrogenase (AcDH). In thermophilic archaea and some bacteria the pyruvate-to-acetyl-CoA step is catalyzed by a PFOR in place of PDH. A survey of thermophilic archaeal genomes did not reveal any genes with homology to known autonomous PDC or AcDH. Furthermore, in a study of enzyme activity of cell free extracts of *P. furiosus* and *T. guaymasensis*, no AcDH activity was detected (43). The researchers looked to the dual-function PDC/PFOR as described in *P. furiosus* as a potential answer, as in PDC mode the *por* enzyme catalyzes the reaction all the way through to acetaldehyde, integrating into its own function the function served by AcDH wherein acetyl-CoA is converted to acetaldehyde. If the *T.*

guaymasensis ‘PFOR’ were to function as a PDC/PFOR as in *P. furiosus*, the PDC-mode of the enzyme would plug the metabolic hole in being able to produce acetaldehyde. This is exactly what the researchers found.

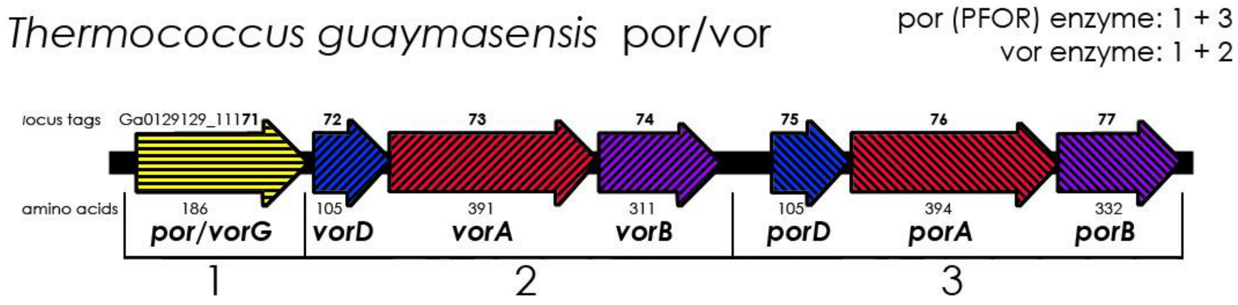

Figure 12: Gene organization of the por/vor *genes in* Thermococcus guaymasensis*. This modular setup is also found in the PFOR/PDC in* Pyrococcus furiosus *and other hyperthermophilic archaea. Bacterial tetrameric PFORs lack the insertion of the* vor *gene cassette. There are three ‘cassettes’ or transcription units that are arranged into two different combinations: the* vor *enzyme is formed by gene clusters 1+2, and the* por*/PFOR enzyme is formed by gene clusters 1+3. The gamma subunit is shared between both enzymes. Each enzyme is expressed independently (i.e. they are not always simultaneously expressed). In the original 2014 paper describing these genes there had not yet been a genome; locus tags added to this diagram based on 2016 genome. Diagram adapted from figure 4 in Eram et al 2014.*

The PFORs in thermophilic archaea *Pyrococcus furiosus* and *Thermococcus guaymasensis* are tetramers that resemble the tetrameric form found in bacteria, with a striking exception in their gene organization: between the gamma subunit (*porG*) and the delta, alpha, and beta subunits (*porD*, *porA*, *porB*) of the PFOR/PDC is inserted a three-gene cassette for the delta, alpha, and beta subunits of a branched-chain keto-acid oxioreductase (*vorD*, *vorA*, *vorB*). The VOR enzyme itself will be discussed in a later section. This parcel of genes is arranged into three ‘transcription units’, labeled in figure 11 as 1, 2, and 3. TU1, consisting solely of *por/vorG*, codes for a gamma subunit shared by both enzymes. TU2 codes for the *vorDAB* series of genes. To generate the VOR enzyme, TU1 and TU2 are transcribed. TU3 contains the *porDAB* genes. To generate the PDC/PFOR enzyme, TU1 and TU3 are transcribed. (43) *porGDAB* codes for a dual-function enzyme that exhibits *both* pyruvate decarboxylase (PDC) and pyruvate: ferredoxin oxidoreductase (PFOR) activity. Note that this activity is relegated *to the same enzyme*, that generated by *porGDAB*; *vorGDAB* is a separate enzyme complex. So, this one parcel of genes that codes for two separate enzymes could also be said to code for *three* separate functions. VOR, like PFOR, can also catalyze the oxidative decarboxylation of pyruvate, though it works also on other substrates, whereas PFOR/PDC has a strong preference for pyruvate. Despite sharing a gamma subunit, and despite the *vorDAB* genes being inserted between the gamma subunit and the *porDAB* genes, VOR and PFOR/PDC are not necessarily co-expressed in *P. furiosus* (43), although there is co-expression (along with other β-ketoacid oxidoreductase, or KOR, genes) of the VOR and PFOR/PDC in *P. furiosus* when grown on peptides (44). The distribution of archaeal consensus transcription start and termination sites follows the pattern one would expect for the three transcription units in *P. furiosus* (45), further evidence for their existence as such.

The only parameters thus far found to determine whether the PDC/PFOR enzyme from *T. guaymasensis* prefers to act as a PDC or a PFOR are temperature and pH (table 3). Other parameters that may determine the enzyme's preferred activity are not yet known. (43) The optimum pH range for the PDC mode in both *P. furiosus* and *T. guaymasensis* PDC/PFOR was higher than those found for PDCs in plants and fungi (usually between 5.0 and 7.5) (43, 46). The PFOR mode of activity in the *T. guaymansensis* PDC/PFOR had an optimum activity at pH 8.4, which is close to the optimum pH for PFOR mode in PDC/PFOR from *P. furiosus*, and for the overall operating of the four-subunit PFOR from the thermophilic bacterium *Thermotoga maritima* (43, 47, 48).

Activity	Optimum Temperature	Temperature notes	Optimum pH	pH range (at least 75% max activity)
PDC	85°C	Activity increases to 85°C, then drops sharply.	9.5	9-10
PFOR	95°C	Activity increases to/peaks at 95°C (assay limit).	8.4	7.5-10

Table 3: Parameters determining preference for PDC or PFOR activity in the combined PDC/PFOR enzyme in *Thermococcus guaymasensis*. Other parameters that may determine the enzyme's preferred activity are not yet known. Temperature measured from 25-95°C. pH measured from range 6-12. Data taken from Elam et al 2014 (43).

Both the PDC and PFOR modes of activity require coenzyme A. However, only in the PFOR mode of activity does it serve as a *substrate*. *In vitro* studies of the *T. guaymasensis* PDC/PFOR showed that acetyl-CoA, specifically, was necessary for PFOR enzyme activity, and that desulfo-CoA (an analogue of acetyl-CoA lacking the reactive -SH group) was inhibitory. For PDC activity, desulfo-CoA was an acceptable substitute, in which the enzyme exhibited 75% (*T. guaymasensis*) or 80% (*P. furiosus*) of the activities exhibited when incubated with acetyl-CoA. This indicates that when the enzyme is functioning in its PDC capacity, acetyl-CoA is serving a structural or stabilizing role, rather than a catalytic role, as it serves in PFOR mode. (43)

Archaeoglobus fulgidus and *Methanothermobacter thermautotrophicus*: tetrameric PFORs in archaea, but without the inserted *vorDAB* genes. PDC function unknown

A four-subunit PFOR was also discovered in the thermophilic archaeon *Archaeoglobus fulgidus*. It was found to oxidize pyruvate, but did *not* oxidize other tested 2-oxoacids (in this case: 2-oxoglutarate, indolepyruvate, phenylpyruvate, glyoxylate, and hydroxypyruvate) (49). The initial study did not discuss the existence or lack thereof of the intervening *vorDAB* genes, as found inserted in the PFOR genes in the other studied thermophile archaea. Examination of three *A. fulgidus* genomes (DSM 8774, 7324, and VC-16/DSM4304) on IMG reveals that in this particular archaeon the tetrameric PFOR genes are arranged as they are in bacteria: *without* the *vorDAB* genes, neither inserted in the PFOR genes nor anywhere in the genome at all. Also, unlike the PFOR in *T. guaymasensis* and *P. furiosus*, both native PAGE and size-exclusion gel chromatography indicated that the holoenzyme is a single heterotetramer (only αβγδ, not $(\alpha\beta\gamma\delta)_2$) with a mass around 120-125kDa.

Methanothermobacter thermautotrophicus (formerly known as *Methanobacterium thermoautotrophicum*), a methanogenic archaeon, also has a four-subunit PFOR (50). Examination of two *M. thermoautotrophicus* genomes on IMG (strains CaT2 and Delta H) reveals that the PFOR genes do not have inserted within them the VOR genes, either. However, unlike *A. fulgidus*, *M. thermoautotrophicus* does have a VOR (50), merely isolated in the genome and not inserted into the PFOR gene complex. Also unlike the one in *Archaeoglobus fulgidus*, *M. thermoautotrophicus*'s PFOR was found to oxidize *only* pyruvate and not other tested 2-oxoacids (2-oxoglutarate, indolepyruvate, phenylpyruvate, glyoxylate, and hydroxypyruvate). The *M. thermoautotrophicus* PFOR was able to oxidize 2-oxobutyrate at roughly half the rate of its main substrate, pyruvate (50). It also weakly catalyzed the oxidative decarboxylation of hydroxypyruvate and glyoxylate, but there was no detectable activity on 2-oxoglutarate, 2-oxoisovalerate, or indolepyruvate. One may therefore be inclined to do as was done with the *Halobacterium salinarum* "PFOR" and re-classify it as an OFOR, the broad-spectrum member of the enzyme family. However, its differing substrate specificity (in the case of the *H. salinarum* OFOR: "2-oxobutarate, pyruvate and 2-oxoglutarate in decreasing order of efficiency" (31), whereas the *M. thermoautotrophicus* PFOR has a preference for pyruvate) and different structure (in the case of the *H. salinarum* OFOR: a two-subunit complex lacking a self-contained Fe-S domain), suggest that it is more like the other PFORs than the OFOR and should be for now classified as such.

What is *not* known, however, about the four-subunit *Archaeoglobus fulgidus* and *Methanothermobacter thermautotrophicus* PFORs, is whether they also exhibit PDC activity like the four-subunit PFORs in fellow archaea *Pyrococcus* and *Thermococcus*. It is necessary to know this for proper categorization of the *A. fulgidus* and *M. thermoautotrophicus* PFORs according to the current scheme. Therefore, given the lack of further study on these enzymes, I must tentatively place these PFORs in its own category, for the time being, and it would be unwise to rely too heavily upon them to make phylogenetic comparisons. In the lack of the *vorDAB* genes is demonstrated a potential branch-point for evaluating at which point the heterotetrameric PFOR’s *por* genes gained or lost the intervening *vor* genes as seen in *Pyrococcus* and *Thermococcus*, and from that, it would be possible to construct a rudimentary sketch for a potential horizontal gene transfer point with bacteria (i.e. *H. pylori* and *T. maritima*).

VOR: 2-Ketoisovalarate (branched-chain α-ketoacid) Ferredoxin: Oxidoreductase

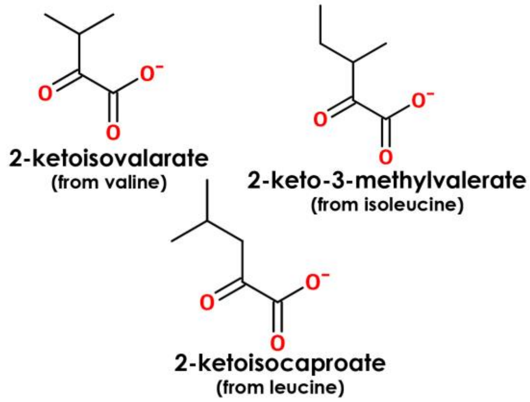

Figure 13 The branched-chain amino acid-derived alpha-ketoacids. Drawn at emolecules.com.

VORs play an important role in the degradation of aliphatic amino acids (51). Indeed, until its discovery in methanogenic archaeon *Methanothermobacter thermautotrophicus* (formerly known as *Methanobacterium thermoautotrophicum*) they had only been found in peptide-fermenting, hyperthermophlic archaea, such as *Pyrococcus* and *Thermococcus*. In *Pyrococcus furiosus*, *Thermococcus litoralis*, and *Thermococcus paralvinellae* (formerly known as *Thermococcus* sp. ES-1)**,** the *vorGDAB* genes are in that composite *por/vor* gene cluster with the shared gamma subunit as outlined above in the PFOR section. No complete genome for *Pyrococcus endeavori* (formerly known as *Pyrococcus* sp. ES-4) has yet been sequenced, nor were the genes isolated in the study, so the presence or absence of the adjacent *por* genes cannot be confirmed.

In the Heider study (51), VOR activity was detected in extracts from four hyperthermophilic, proteolytic archaea. *Pyrococcus furiosus* and *Thermococcus litoralis* do not rely upon the presence of S^0 for growth, whereas *Pyrococcus endeavori* (in that paper *Pyrococcus* sp. ES-4), and *Thermococcus paralvinellae* (in that paper *Thermococcus* sp. ES-1) do require S^0 for growth**.** However, of these four enzymes, the group only reported protein size for *T. litoralis* VOR, other than to note that "[the other three VORs] gave rise to SDS gel patterns very similar to that of *T. litoralis* VOR". The group only reported enzyme kinetics for the two Thermococcus species' VORs (*T. litoralis* and *T. paralvinellae*). The justification given was that given N-terminal sequence similarity and gel banding similarity, "[i]t is assumed that the biochemical and spectroscopic properties of the other enzymes are analogous [to those of the *T. litoralis* VOR]" (51). I find this a dangerous assumption, but the additional data from these experiments is lost to all but the original researchers, and so make the best of what they did report. Gel banding indicated that the four VORs associated as a dimer of heterotetramers, $(\alpha\beta\gamma\delta)_2$, with *T. litoralis* holoenzyme size around 230 kDa and subunit sizes as follows: α - 47 kDa, β - 34 kDa, γ - 23.5 kDa, and δ -13.1 kDa (appendix 2 table 2).

Initial enzyme activity tests measured *T. litoralis* and *T. paralvinellae* VOR activity upon a group of 2-oxoacids with *T. litoralis* ferredoxin (or benzyl / methyl viologen) and CoA. The *T.*

litoralis VOR oxidized most efficiently 2-ketoacid derivatives of the branched-chain amino acids valine, leucine, and isoleucine, as well as methionine. Pyruvate and aryl-pyruvates (2-ketobutyrate, pyruvate, phenylpyruvate, and phenylglyoxylate) were poorly oxidized (K_m > 200 µM), and there was no activity detected upon 2-ketoglutarate. No activity was detected upon any substrate when NAD or NADP were substituted as electron acceptors. *T. paralvinellae* VOR showed the same preferences, but with even less activity upon pyruvate and no activity on any aromatic amino acid-derived 2-ketoacids. The researchers therefore conclude that the physiological substrates of VOR are the branched-chain amino acid and methionine keto-derivatives, ferredoxin, and CoA. (51)

T. litoralis VOR was found to act also in the reverse direction, as a 2-ketoisovalarate synthase. It produced 2-ketoisovalarate and CoA from isobutyryl-coenzyme A, reduced benzyl viologen, and CO_2 at 5% the oxidation (forward reaction) rate at 85°C (51). This indicates potential reverse activity for the other branched chain amino acid keto-derivatives as well. The *T. litoralis* VOR acted with optimum efficacy at pH 7.0 for both reaction directions, with a sharp drop in efficacy one pH point in either direction (< 50% activity at pH 6.0 and < 25% at pH 8.0), and had an optimum 2-ketoisovalarate oxidation temperature of 90°C. The *T. paralvinellae* VOR had an optimum 2-ketoisovalarate oxidation temperature exceeding 98°C. The optimum pH for the *T. paralvinellae* VOR was not reported.

IOR: Indolepyruvate: Ferredoxin Oxidoreductase

IORs play an important role in the degradation of aromatic amino acids. Once the 2-ketoacid derivative of the aromatic amino acid is created, it is converted to an aryl-CoA via oxidative decarboxylation by the IOR (51). Indeed, until its discovery in methanogenic archaeon *Methanothermobacter thermautotrophicus* (formerly known as *Methanobacterium* *thermoautotrophicum*) it had only been found in peptide-utilizing, hyperthermophlic archaea (50).

IORs have two unique subunits, α and β.

OFOR: 2-Oxoacid (α-ketoacid) Ferredoxin: Oxidoreductase

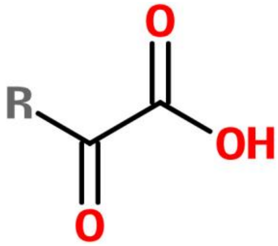

Figure 9: A 2-oxoacid (aka 2-ketoacid or α-ketoacid). Drawn at emolecules.com.

The hyperthermophilic archaeon *Sulfolobus tokodaii* contains two copies of the OFOR enzyme, designated StOFOR1 and StOFOR2. StOFOR1 was determined to be the most active in cell extracts, and was active on a variety of 2-oxoacids, but in the *E. coli*-based recombinant studies it crystalized poorly. StOFOR2, activity for which is considerably lower than StOFOR1 and which exhibits a preference for pyruvate, crystalized well, and the StOFOR1 structure was only resolved by a StOFOR2-modeled molecular replacement. It was also of StOFOR2 that ligand-free and pyruvate-complexed structures were determined. (34) It is unfortunate that the enzyme that, by all evidences, is most physiologically active and exhibits the greatest flexibility of substrate unique of OFORs did not crystalize well enough to provide a structure, but research on the OFOR enzyme is in its infancy and more resolved structures are sure to follow, perhaps leading to knowledge of what differentiates the promiscuous OFOR from the pyruvate-preferring PFOR.

The *Sulfolobus* OFOR organizes as a dimer of heterodimers ($(\alpha\beta)_2$ configuration, gene order αβ). It lacks the iron-sulfur domain found in other OFOR-family enzymes and necessary for its function, so researchers hypothesize that it integrates a free ferredoxin from another gene. (34) In a mutagenesis study of StOFOR1 D468aA and K49bI mutants lost 2-oxoglutarate activity but retained pyruvate activity, indicating these residues are critical in OFOR's ability to work on multiple substrates (34).

Halobacterium salinarum (formerly known as *Halobacterium halobium*) is an archaeon. It is well-known for its ability to survive in high-salt environments and for its utilization of bacteriorhodopsin, a light-driven proton pump. In 1981 and 1992 there were studies on its two-subunit "pyruvate: ferredoxin oxidoreductase", and a 1996 PFOR/VOR study that used it to construct a PFOR phylogeny had some difficulty fitting it into an alignment with other PFORs before they thought to check for gene re-arrangements (45), an early indicator of its uniqueness from the other PFORs. However, the OFOR studies on *Sulfolobus tokodaii* were performed in 2016, and given that *H. salinarum* is also an archaea and also has a member of the OFOR family composed of two subunits, and lacking an internal Fd domain, it is likely the "PFOR" found in *H. salinarum* is actually what we would now call an "OFOR". Indeed, the original 1981 paper that described the "PFOR" in *H. salinarum* characterized it by its ability to catalyze "2-oxobutarate, pyruvate and 2-oxoglutarate in decreasing order of efficiency" (31). The researchers decided to name it "PFOR" given that they thought pyruvate was the most likely physiological substrate. This was reasonable at the time, but we now know more about this enzyme family and the variety contained therein. By current nomenclature, this "PFOR" would be classified as an "OFOR" for its broad substrate specificity, and is referenced as such in the chart and the paper henceforth.

For further evidence that this *H. salinarum* "PFOR" should be classified as an OFOR, consider the differing substrate specificities of the two OFORs found in *S. tokodaii,* and compare the amino acid sequences for the *H. salinarium* OFOR (henceforth HsOFOR) and the amino acid sequences of the more promiscuous StOFOR1 and the more pyruvate-preferring StOFOR2. In a mutagenesis study of StOFOR1, D468aA and K49bI mutants lost 2-oxoglutarate activity but retained pyruvate activity, indicating these residues are critical in retaining the dual functionality (34). The *H. salinarum* OFOR alpha subunit has an asparagine (N) aligned with the critical aspartic acid residue, D468a, with no D residues 10 places up or downstream of the aligned spot. Both StOFOR1 and StOFOR2 have this aspartic acid residue. However, the *H. salinarum* beta subunit *does* have a lysine residue aligned with K49b, along with both StOFOR1 and StOFOR2. However, the structural data for the pyruvate-bound version of StOFOR2 does not show either of these residues to be directly involved in binding pyruvate (PDB structure 5B47). A 2-oxoglutarate-complexed protein was not crystalized. Clearly, the change of a hydroxl group to an amino group does not change substrate specificity in the *H. salinarium* OFOR.

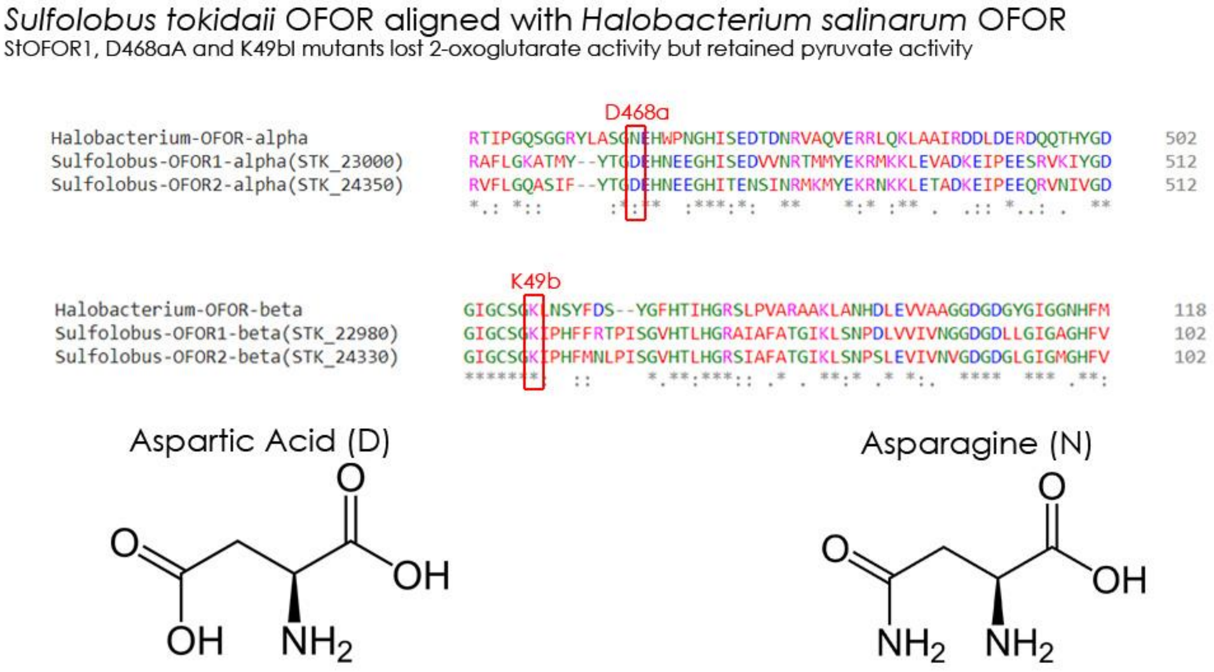

Figure 14: Comparison of substrate-specificity-determining residues in the Sulfolobus tokodaii OFORs and the Halobacterium salinarium OFOR, as determined by mutagenesis studies.

A. fulgidus genomes also note a two-subunit gene complex with OFOR homology, but to date no actual biochemical studies have been performed.

OGOR (aka KOR): 2-oxoglutarate (Ferredoxin) Oxidoreductase

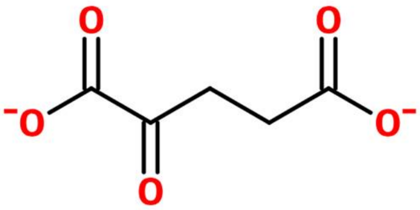

Figure 15: 2-oxoglutarate (aka α -ketoglutarate). Figure drawn at emolecules.com

(NOTE: This enzyme has also been called KOR (50).) Also found in archaeon *Halobacterium salinarum* (along with the OFOR we above discussed; again, note, in the paper cited here the OFOR was at the time called a PFOR despite its broad substrate specificity) was an enzyme closely resembling its OFOR, but only with measured activity on 2-oxoglutarate (not 2-oxobutarate or pyruvate, also tested) (31). This was therefore an OGOR, or a 2-oxoglutarate:ferredoxin oxidoreductase. The molecular weight of the OGOR complex was determined to be around 245kDa, which, given that it consists of two subunits of molecular weights 88kDa (α) and 36kDa (β), would indicate the holoenzyme structure is that of a dimer of heterodimers: $(\alpha\beta)_2$. (31)

In 1996, an OGOR was also found in *Hydrogenobacter thermophilus* TK-6, a thermophilic, aerobic, autotrophic, hydrogen-oxidizing bacterium (52, 53). Of note, *H. thermophilus* was found to be able to fix carbon dioxide under aerobic conditions utilizing a reductive TCA cycle—notable because this method of CO_2 fixation is mostly found in anaerobes. In radiolabeled carbon studies of CO_2 fixation, this research group identified key enzymes involved in this pathway, and those enzymes included PFOR and OGOR (53). OGOR had a high specificity for 2-oxoglutarate over other 2-oxoacids, and was determined to be a 105kDa heterodimer consisting of two subunits of molecular weights 70kDa (α) and 35kDa (β). Specifically, its activity was as noted: it "did not react with oxalacetate, oxomalonate, 2-oxoisocaproate, or phosphoenolpyruvate. [The OFOR] reacted with 2-oxobutyrate, pyruvate, and 2-oxoisovalerate at an activity level of less than 0.4 to 0.7% relative to that with 2-oxoglutarate" (53). This contrasts with the reported activity of the *H. salinarum* OFOR, above, which was reported not to exhibit *any* activity on 2-oxobutarate or pyruvate (31), in binary ability (zero activity/any activity) if not in magnitude. Here we note other differences that jump out upon reading appendix 2 table 2: *H. salinarum*, an archaeon, and *H. thermophilus*, a bacterium; in *H. salinarum*, a predicted dimer of heterodimers for the holoenzyme, and in *H. thermophilus*, a predicted simple heterodimer. For similarities, *H. salinarum* and *H. thermophilus* are both aerobes, and both use a ferredoxin-utilizing analogue of an enzyme (OGOR), whereas other aerobes usually utilize the NAD(P)-utilizing, multienzyme-complex analogue (2-oxoglutarate dehydrogenase).

Yoon et al note that PFOR seems to be involved in many metabolic pathways in many organisms, whereas in *H. thermophilus*, at least, OGOR is specific for the reductive TCA (53).

Methanothermobacter thermautotrophicus (formerly known as *Methanobacterium thermoautotrophicum*) is a methanogenic, thermophilic, anaerobic archaeon capable of utilizing H_2 and CO_2 as a sole energy and carbon source, respectively. CO_2 fixation occurs via an acetyl-CoA and carbon monoxide dehydrogenase-containing pathway, of which acetyl-CoA and pyruvate are key intermediates, and 2-oxoglutarate is synthesized from pyruvate. (50) *M. thermautotrophicus*'s OGOR was found to contain four subunits. Compare this to the two-subunit OGORs of *H. salinarum* and *H. thermophilus*. It was also found to have a strong preference for 2-oxoglutarate and exhibited no activity on the other 2-oxoacids tested, with the exception of slight activity on 2-oxoadipinate (50). As only an SDS-PAGE was run on the protein and not a native PAGE, holoenzyme arrangement cannot be predicted.

OOR: Oxalate (Ferredoxin) Oxidoreductase

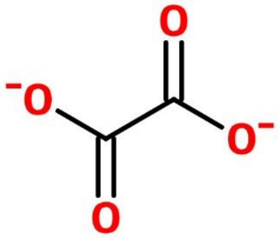

Figure 16: Oxalate. Drawn at emolecules.com.

Oxalate oxidoreductase (OOR) was first characterized in bacterium *Moorella thermoacetica* and is unique of OFORs in that it does not require CoA for catalysis (37, 38). In this regard it differs also from the until-then only-known anaerobic oxalate degrading enzyme, oxalyl-CoA decarboxylase (54). The carbon dioxide and low-potential electrons (bound to ferredoxin) from the breakdown of oxalate are fed into the Wood−Ljungdahl pathway for acetogenesis (33, 37, 38). This allows *M. thermoacetica* to live on oxalate, a molecule that is a toxic metabolic waste product in many organisms and indeed is even secreted by some plants into soil as a defense mechanism against predators (35, 55). While OOR was shown to act upon a variety of 2-oxoacids, its activity was strongest on oxalate. Its cellular expression in *M. thermoacetica* is induced by oxalate, but not by other 2-oxoacids. These two facts indicate that the physiological substrate is likely oxalate. (37) Gibson et al, who did the most recent OOR structure study, note that OOR is unique of OFOR enzymes, with "its primary function being the assimilation of an exogenous nutrient, rather than conversion and recycling of cellular metabolites, as is the case for the other OFORs" (35).

A battery of structural studies have been performed on OOR_{Mt}, starting with a resting-state crystal structure (38) followed by a study of the enzyme co-crystalized with oxalate (33). The first study established that it arranges as a dimer of heterotrimers $(\alpha\gamma\beta)_2$ with a gene order of αγβ. When compared to the functional domains as established by *D. africanus* (40), it lacks domains IV (a short domain of unknown function found in the homodimeric PFORs, annotated as an EKR) and VII (the C-terminal domain of unknown function thus far found only in homodimeric PFORs and which is predicted to aid in enzyme stabilization). Note that these domains are missing, also, from all OFORs that are not homodimeric PFORs from bacteria. It is also missing a four-helix segment of domain VI (contains part of the TPP-binding domain (40)), and this four-helix segment is of unknown function. Although it lacks a separate subunit δ, which in the tetrameric OFORs contains the iron-sulfur center, an analogous sequence is found on the end of subunit γ, indicating a possible fusion. (38) In the second study, the researchers determined that the active site, which houses the TPP cofactor and one of three [4Fe-4S] sites, is more polar than the analogous active site structure in PFORs. This is in congruence with oxalate's more-charged structure as compared to that of pyruvate. (33)

Oxalate metabolism has been of interest to the medical field given its ability to cause kidney stones, renal failure, and crystalline arthritis. It is difficult to metabolize and humans and most animals lack the endogenous ability to break it down. Indeed, in mammals it is a terminal metabolite primarily secreted via the kidneys, and it is there that it can begin to cause trouble when it accumulates. (56) There has therefore been interest in any methods of inducing oxalate breakdown in the body. Future research on OOR can reasonably be anticipated, and with research on this member of the OFOR family will come a greater comparative understanding of other members, including PFOR. The ligand-bound structure study of $\mathrm{OOR}_{\mathrm{Mt}}$ has already illuminated much about PFOR by comparison, and lays a foundation for more accurate prediction of the substrate of a computationally-predicted member of the OFOR family.

References

1. **Mukund S, Adams MWW.** 1995. Glyceraldehyde-3-phosphate Ferredoxin Oxidoreductase, a Novel Tungsten-containing Enzyme with a Potential Glycolytic Role in the Hyperthermophilic Archaeon Pyrococcus furiosus. Journal of Biological Chemistry **270:**8389-8392.

2. **Roy R, Adams MWW.** 2002. Characterization of a Fourth Tungsten-Containing Enzyme from the Hyperthermophilic Archaeon Pyrococcus furiosus. Journal of Bacteriology **184:**6952-6956.

3. **Mukund S, Adams MW.** 1991. The novel tungsten-iron-sulfur protein of the hyperthermophilic archaebacterium, Pyrococcus furiosus, is an aldehyde ferredoxin oxidoreductase. Evidence for its participation in a unique glycolytic pathway. Journal of Biological Chemistry **266:**14208-14216.

4. **Kengen SW, de Bok FA, van Loo ND, Dijkema C, Stams AJ, de Vos WM.** 1994. Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by Pyrococcus furiosus. J Biol Chem **269:**17537-17541.

5. **Schäfer T, Schönheit P.** 1992. Maltose fermentation to acetate, CO2 and H2 in the anaerobic hyperthermophilic archaeon Pyrococcus furiosus: evidence for the operation of a novel sugar fermentation pathway. Archives of Microbiology **158:**188-202.

6. **Schäfer T, Schönheit P.** 1993. Gluconeogenesis from pyruvate in the hyperthermophilic archaeon Pyrococcus furiosus: involvement of reactions of the Embden-Meyerhof pathway. Archives of Microbiology **159:**354-363.

7. **Mukund S, Adams MW.** 1993. Characterization of a novel tungsten-containing formaldehyde ferredoxin oxidoreductase from the hyperthermophilic archaeon, Thermococcus litoralis. A role for tungsten in peptide catabolism. Journal of Biological Chemistry **268:**13592-13600.

8. **Zwickl P, Fabry S, Bogedain C, Haas A, Hensel R.** 1990. Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaebacterium Pyrococcus woesei: characterization of the enzyme, cloning and sequencing of the gene, and expression in Escherichia coli. Journal of Bacteriology **172:**4329-4338.

9. **Park M-O, Mizutani T, Jones PR.** 2007. Glyceraldehyde-3-Phosphate Ferredoxin Oxidoreductase from Methanococcus maripaludis. Journal of Bacteriology **189:**7281-7289.

10. **Reher M, Gebhard S, Schönheit P.** 2007. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR) and nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN), key enzymes of the respective modified Embden-Meyerhof pathways in the hyperthermophilic crenarchaeota *Pyrobaculum aerophilum* and *Aeropyrum pernix*. FEMS Microbiol Lett **273:**196-205.

11. **Matsubara K, Yokooji Y, Atomi H, Imanaka T.** 2011. Biochemical and genetic characterization of the three metabolic routes in Thermococcus kodakarensis linking glyceraldehyde 3-phosphate and 3-phosphoglycerate. Molecular Microbiology **81:**1300-1312.

12. **Heider J, Ma K, Adams MW.** 1995. Purification, characterization, and metabolic function of tungsten-containing aldehyde ferredoxin oxidoreductase from the hyperthermophilic and proteolytic archaeon Thermococcus strain ES-1. Journal of Bacteriology **177:**4757-4764.

13. **Kanai T, Imanaka H, Nakajima A, Uwamori K, Omori Y, Fukui T, Atomi H, Imanaka T.** 2005. Continuous hydrogen production by the hyperthermophilic archaeon, Thermococcus kodakaraensis KOD1. Journal of Biotechnology **116:**271-282.

14. **Kanai T, Akerboom J, Takedomi S, van de Werken HJG, Blombach F, van der Oost J, Murakami T, Atomi H, Imanaka T.** 2007. A Global Transcriptional Regulator in Thermococcus kodakaraensis Controls the Expression Levels of Both Glycolytic and Gluconeogenic Enzyme-encoding Genes. Journal of Biological Chemistry **282:**33659-33670.

15. **Ma K, Hutchins A, Sung S-JS, Adams MWW.** 1997. Pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon, *Pyrococcus furiosus*, functions as a CoA-dependent pyruvate decarboxylase. Proc Natl Acad Sci USA **94:**9608-9613.

16. **Chan MK, Mukund S, Kletzin A, Adams MW, Rees DC.** 1995. Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. Science **267:**1463.

17. **Roy R, Mukund S, Schut G, M. Dunn D, Weiss R, W. W. Adams M.** 1999. Purification and Molecular Characterization of the Tungsten-Containing Formaldehyde Ferredoxin Oxidoreductase from the Hyperthermophilic Archaeon Pyrococcus furiosus: the Third of a Putative Five-Member Tungstoenzyme Family, vol 181.

18. **Hu Y, Faham S, Roy R, Adams MWW, Rees DC.** 1999. Formaldehyde ferredoxin oxidoreductase from Pyrococcus furiosus: the 1.85 Å resolution crystal structure and its mechanistic implications11Edited by I. A. Wilson. Journal of Molecular Biology **286:**899-914.

19. **Weinberg MV, Schut GJ, Brehm S, Datta S, Adams MWW.** 2005. Cold Shock of a Hyperthermophilic Archaeon: Pyrococcus furiosus Exhibits Multiple Responses to a Suboptimal Growth Temperature with a Key Role for Membrane-Bound Glycoproteins. Journal of Bacteriology **187:**336.

20. **Bevers LE, Bol E, Hagedoorn P-L, Hagen WR.** 2005. WOR5, a Novel Tungsten-Containing Aldehyde Oxidoreductase from Pyrococcus furiosus with a Broad Substrate Specificity. Journal of Bacteriology **187:**7056-7061.

21. **Roy R, Menon AL, Adams MWW.** 2001. [11] Aldehyde Oxidoreductases from Pyrococcus furiosus, p 132-144, Methods in Enzymology, vol 331. Academic Press.

22. **Strobl G, Feicht R, White H, Lottspeich F, Simon H.** 1992. The tungsten-containing aldehyde oxidoreductase from Clostridium thermoaceticum and its complex with a viologen-accepting NADPH oxidoreductase. Biological chemistry Hoppe-Seyler **373:**123-132.

23. **Trautwein T, Krauss F, Lottspeich F, Simon H.** 1994. The (2R)-hydroxycarboxylate-viologen-oxidoreductase from Proteus vulgaris is a molybdenum-containing iron-sulphur protein. European Journal of Biochemistry **222:**1025-1032.

24. **Mukund S, Adams MW.** 1990. Characterization of a tungsten-iron-sulfur protein exhibiting novel spectroscopic and redox properties from the hyperthermophilic archaebacterium Pyrococcus furiosus. Journal of Biological Chemistry **265:**11508-11516.

25. **Hensgens CM, Hagen WR, Hansen TA.** 1995. Purification and characterization of a benzylviologen-linked, tungsten-containing aldehyde oxidoreductase from Desulfovibrio gigas. Journal of Bacteriology **177:**6195-6200.

26. **Hiltrud W, Gerd S, Richard F, Helmut S.** 1989. Carboxylic acid reductase: a new tungsten enzyme catalyses the reduction of non-activated carboxylic acids to aldehydes. European Journal of Biochemistry **184:**89-96.

27. **Huber C, Skopan H, Feicht R, White H, Simon H.** 1995. Pterin cofactor, substrate specificity, and observations on the kinetics of the reversible tungsten-containing aldehyde oxidoreductase fromClostridium thermoaceticum. Archives of Microbiology **164:**110-118.

28. **Rauh D, Graentzdoerffer A, Granderath K, Andreesen Jan R, Pich A.** 2003. Tungsten-containing aldehyde oxidoreductase of Eubacterium acidaminophilum. European Journal of Biochemistry **271:**212-219.

29. **Barata BA, LeGall J, Moura JJ.** 1993. Aldehyde oxidoreductase activity in Desulfovibrio gigas: in vitro reconstitution of an electron-transfer chain from aldehydes to the production of molecular hydrogen. Biochemistry **32:**11559-11568.

30. **Chen PY-T, Aman H, Can M, Ragsdale SW, Drennan CL.** 2018. Binding site for coenzyme A revealed in the structure of pyruvate:ferredoxin oxidoreductase from Moorella thermoacetica. Proceedings of the National Academy of Sciences **115:**3846-3851.

31. **Lorenz K, Dieter O.** 1981. Purification and Properties of Two 2-Oxoacid:Ferredoxin Oxidoreductases from Halobacterium halobium. European Journal of Biochemistry **116:**587-594.

32. **Breslow R.** 1958. On the Mechanism of Thiamine Action. IV.1 Evidence from Studies on Model Systems. Journal of the American Chemical Society **80:**3719-3726.

33. **Gibson MI, Chen PY-T, Johnson AC, Pierce E, Can M, Ragsdale SW, Drennan CL.** 2016. One-carbon chemistry of oxalate oxidoreductase captured by X-ray crystallography. Proceedings of the National Academy of Sciences **113:**320.

34. **Yan Z, Maruyama A, Arakawa T, Fushinobu S, Wakagi T.** 2016. Crystal structures of archaeal 2-oxoacid:ferredoxin oxidoreductases from Sulfolobus tokodaii. Scientific Reports **6:**33061.

35. **Gibson MI, Chen PY-T, Drennan CL.** 2016. A structural phylogeny for understanding 2-oxoacid oxidoreductase function. Current opinion in structural biology **41:**54-61.

36. **Ragsdale SW.** 2003. Pyruvate Ferredoxin Oxidoreductase and Its Radical Intermediate. Chemical Reviews **103:**2333-2346.

37. **Pierce E, Becker DF, Ragsdale SW.** 2010. Identification and Characterization of Oxalate Oxidoreductase, a Novel Thiamine Pyrophosphate-dependent 2-Oxoacid Oxidoreductase That Enables Anaerobic Growth on Oxalate. Journal of Biological Chemistry **285:**40515-40524.

38. **Gibson MI, Brignole EJ, Pierce E, Can M, Ragsdale SW, Drennan CL.** 2015. The Structure of an Oxalate Oxidoreductase Provides Insight into Microbial 2-Oxoacid Metabolism. Biochemistry **54:**4112-4120.

39. **Zhang Q, Iwasaki T, Wakagi T, Oshima T.** 1996. 2-Oxoacid:Ferredoxin Oxidoreductase from the Thermoacidophilic Archaeon, <i>Sulfolobus</i> sp. Strain 7. The Journal of Biochemistry **120:**587-599.

40. **Charon M-H, Volbeda A, Chabriere E, Pieulle L, Fontecilla-Camps JC.** 1999. Structure and electron transfer mechanism of pyruvate:ferredoxin oxidoreductase. Current Opinion in Structural Biology **9:**663-669.

41. **Chabrière E, Charon MH, Volbeda A, Pieulle L, Hatchikian EC, Fontecilla–Camps JC.** 1999. Crystal structures of the key anaerobic enzyme pyruvate:ferredoxin oxidoreductase, free and in complex with pyruvate. Nature Structural Biology **6:**182.

42. **Takenaka M, Yoon K-S, Matsumoto T, Ogo S.** 2017. Acetyl-CoA production by encapsulated pyruvate ferredoxin oxidoreductase in alginate hydrogels. Bioresource Technology **227:**279-285.

43. **Eram MS, Oduaran E, Ma K.** 2014. The Bifunctional Pyruvate Decarboxylase/Pyruvate Ferredoxin Oxidoreductase from Thermococcus guaymasensis. Archaea **2014:**13.

44. **Schut GJ, Brehm SD, Datta S, Adams MWW.** 2003. Whole-Genome DNA Microarray Analysis of a Hyperthermophile and an Archaeon: Pyrococcus furiosus Grown on Carbohydrates or Peptides. Journal of Bacteriology **185:**3935-3947.

45. **Kletzin A, Adams MW.** 1996. Molecular and phylogenetic characterization of pyruvate and 2-ketoisovalerate ferredoxin oxidoreductases from Pyrococcus furiosus and pyruvate ferredoxin oxidoreductase from Thermotoga maritima. Journal of Bacteriology **178:**248-257.

46. **König S.** 1998. Subunit structure, function and organisation of pyruvate decarboxylases from various organisms. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology **1385:**271-286.

47. **Blamey JM, Adams MWW.** 1993. Purification and characterization of pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus.* Biochim Biophys Acta **1161:**19-27.

48. **Blamey JM, Adams MWW.** 1994. Characterization of an ancestral type of pyruvate ferredoxin oxidoreductase from the hyperthermophilic bacterium, *Thermotoga maritima.* Biochemistry **44:**1000-1007.

49. **Kunow J, Linder D, Thauer RK.** 1995. Pyruvate: ferredoxin oxidoreductase from the sulfate-reducing Archaeoglobus fulgidus: molecular composition, catalytic properties, and sequence alignments. Archives of Microbiology **163:**21-28.

50. **Tersteegen A, Linder D, Thauer Rudolf K, Hedderich R.** 2004. Structures and Functions of Four Anabolic 2-Oxoacid Oxidoreductases in Methanobacterium Thermoautotrophicum. European Journal of Biochemistry **244:**862-868.

51. **Heider J, Mai X, Adams MW.** 1996. Characterization of 2-ketoisovalerate ferredoxin oxidoreductase, a new and reversible coenzyme A-dependent enzyme involved in peptide fermentation by hyperthermophilic archaea. Journal of Bacteriology **178:**780-787.

52. **Yun N-R, Arai H, Ishii M, Igarashi Y.** 2001. The Genes for Anabolic 2-Oxoglutarate: Ferredoxin Oxidoreductase from Hydrogenobacter thermophilus TK-6. Biochemical and Biophysical Research Communications **282:**589-594.

53. **Yoon KS, Ishii M, Igarashi Y, Kodama T.** 1996. Purification and characterization of 2-oxoglutarate:ferredoxin oxidoreductase from a thermophilic, obligately chemolithoautotrophic bacterium, Hydrogenobacter thermophilus TK-6. Journal of Bacteriology **178:**3365-3368.

54. **Quayle JR.** 1963. CARBON ASSIMILATION BY PSEUDOMONAS OXALATICUS (OX 1). 7. DECARBOXYLATION OF OXALYL-COENZYME A TO FORMYL-COENZYME A. Biochemical Journal **89:**492.

55. **Franceschi VR, Nakata PA.** 2005. CALCIUM OXALATE IN PLANTS: Formation and Function. Annual Review of Plant Biology **56:**41-71.

56. **Marengo SR, Romani AMP.** 2008. Oxalate in renal stone disease: the terminal metabolite that just won't go away. Nature Clinical Practice Nephrology **4:**368.

Appendix II: Tables for Review of the AFOR and PFOR Enzyme Families

AFOR family proteins	Substrate	occurs as	size SDS	size native	organism	cofactor	Domain
AFOR	Various aldehydes (including formaldehyde(1)); wide-range, but preference for amino acid-derived aldehydes(2, 3)	homodimer(2)	65kDa (2)	Cannot access original paper	*P. furiosus*(2)	W(2)	A
		homodimer(3)	75kDa(3)	105kDa(3)	*Thermococcus* strain ES-1(3)	W(3)	A
FOR	Formaldehyde (non-physiological), and other C4-C6 di- and semialdehydes and C1-C3 aldehydes(1, 2, 4, 5)	homotetramer (2, 4, 5)	68kDa(4)	275kDa(4)	*P. furiosus*(2-5)	W(2, 4, 5)	A
		homotetramer (1)	70kDa(1)	280kDa(1)	*T. litoralis*(1)	W(1)	A
WOR4	Unknown; may be involved in S^0 metabolism(6)	homodimer(6)	69kDa(6)	129kDa(6)	*P. furiosus* (grown w/ S^0)(6)	W(6)	A
WOR5	Various aldehydes; greatest activity on hexanaldehyde(7)	homodimer(7)	67kDa(7)	135kDa(7)	*P. furiosus*(7)	W(7)	A
GAPOR	glyceraldehyde-3-phosphate(2, 8, 9)	monomer(8)	63kDa(8)	44kDa(8)	*P. furiosus*(8)	W(8)	A
		monomer(9)	70kDa(9) (size-exclusion chromatography purification)		*M. maripaludis* recomb. *E. coli*(9)	Mo(9)	A
		monomer(10)	67kDa(10)	60kDa(10)	*P. aerophilum*(10)	ND	A
		Not determined	Not measured		*T. kodakarensis*(11)*	ND	A

Table 1: The aldehyde:ferredoxin oxidoreductase (AFOR) family of enzymes. AFOR is both a specific enzyme in the family (broad specificity for aldehydes) and is frequently used as the name of the enzyme family as a whole, of which members have

varying specificities for specific aldehydes. Although WOR4 has been placed in the AFOR family due to homology to other members, it was not found to oxidize any aliphatic or aromatic aldehydes or hydroxy acids, or to reduce any keto acids (6). All enzymes are predicted to utilize ferredoxin as an electron receptor. All known members of the AFOR family consist of one unique subunit, in differing arrangements for holoenzymes. WOR4 protein was only found in *P. furiosus* grown with elemental sulfur (S^0). As of May 16, 2019 the only crystal structure for the AFOR enzyme in the RCSB Protein Databank comes from *Pyrococcus furiosus*(12). The only crystal structures for the FOR enzyme (glutarate-complexed and naked) also come from *P. furiosus*(5). ND = not determined

**T. kodakarensis* GAPOR was only studied in the context of determining the effect of a predicted *gor* gene knockout on growth; the protein was not purified and studied (11). It is therefore listed as a putative GAPOR based on the action of homologous genes in *Pyrococcus* and the fact that its knockout inhibited glycolytic growth, and that GAPOR in *Pyrococcus* was shown to be necessary for glycolytic growth. These two circumstantial factors (high sequence homology and the same impact upon glycolysis) give reasonable confidence that the gene knocked out was in fact GAPOR.

OFOR family proteins	substrate	product	rever-sible?	holoenzyme organization	gene order	sub. size SDS	holoenzyme size native	AA length	organism	domain
OFOR	Pyruvate, 2-oxoglutarate, 2-oxobutyrate + $^-$SCoA (13, 14)	(R)CoA + CO$_2$ + 2e$^-$	Y	$(\alpha\beta)_2$(13, 14)	αβ(13, 14)	α 70kDa(13) β 34kDa(13)	Not reported	α 627‡ aa(13) β 305 aa(13)	*Sulfolobus tokodaii*(13, 15)	A
						α 86kDa(14) β 42kDa(14)	256kDa(14)	α 628 aa(14) β 312 aa(14)	*Halobacterium salinarum*¤(14)	A
PFOR	Pyruvate + $^-$SCoA(16-19)	Acetyl-CoA + CO$_2$ + 2e$^-$	Y	α_2(16-19)	-------	133kDa(19)	256kDa(17-19)	1232 aa	*Desulfovibrio africanus*(17-19)	B
						120kDa(20)	Not reported	1711 aa	*Moorella thermoactica*(20)	B
						120kDa ± 5kDa(16)	240kDa ± 20kDa(16)	1174 aa	*Citrobacter sp. S-77*(16)	B
				$(\alpha\beta\gamma\delta)_2$	γδαβ	α 48kDa(21) β 36kDa(21) γ 24kDa(21) δ 14kDa(21)	Not reported	α 407 aa β 314 aa γ 186 aa δ 130 aa	*Helicobacter pylori*(21)	B
						α 44.3kDa(22) β 36.4kDa(22) γ 21.3kDa(22) δ 11.3kDa(22)	Not reported	α 392 aa β 324 aa γ 192 aa δ 99 aa	*Thermotoga maritima*(22)	B

OFOR family proteins	substrate	product	rever-sible?	holoenzyme organization	gene order	sub. size SDS	holoenzyme size native	AA length	organism	do-main
PFOR/PDC	Pyruvate + $^-$SCoA†(23) (dual function on same substrate)	PFOR mode: Acetyl-CoA + CO$_2$ + 2e$^-$(23) PDC mode: acetaldehyde(23)	PFOR Y(23) PDC N(23)	(αβγδ)$_2$(23)	γ*δαβ (23)	α 43.8kDa(23) β 36.1kDa(23) γ 19.9kDa(23) δ 11.9kDa(23)	258kDa ± 56kDa(23)	α 394 aa(23) β 332 aa(23) γ 186 aa(23) δ 105 aa(23)	*Thermococcus guaymasensis*(23)	A
						α 44.2kDa(22) β 36.3 kDa(22) γ 20.0kDa(22) δ 12.0 kDa(22)	Not reported	α 396 aa β 331 aa γ 185 aa δ 105 aa	*Pyrococcus furiosus*(22)	A
PFOR (PDC?)	Pyruvate + $^-$SCoA(24)	Acetyl-CoA + CO$_2$ + 2e$^-$	Y(23)	αβγδ(24)	γδαβ(24)	α 45kDa(24) β 33kDa(24) γ 25kDa(24) δ 13kDa(24)	120kDa(24)	α 391 aa β 294 aa γ 191 aa δ 97 aa	*Archaeoglobus fulgidus*(24)	A
	Pyruvate, 2-oxobutyrate, hydroxypyruvate, glyoxylate + $^-$SCoA(25)	(R)CoA + CO$_2$ + 2e$^-$		Not reported		α 43kDa(25) β 31kDa(25) γ 22kDa(25) δ 15kDa(25)	Not reported	α 383 aa β 288 aa γ 177 aa δ 81 aa	*Methanothermobacter thermautotrophicus* ※ (25)	A

OFOR family proteins	substrate	product	reversible?	holoenzyme organization	gene order	sub. size SDS	holoenzyme size native	AA length	organism	domain
OGOR (KOR)	2-oxoglutarate + $^-$SCoA (14, 26)	Succinyl-CoA + CO$_2$ + 2e$^-$ (25)	Y(26)	(αβ)$_2$(14)	αβ(14, 26)	α 88kDa(14) β 36kDa(14)	245kDa(14)	α 582 aa β 289 aa	*Halobacterium salinarum*¤ (14)	A
				αβ(27)		α 70kDa(26, 27) β 35kDa(26, 27)	105kDa(27)	α 607 aa(26) β 295 aa(26)	*Hydrogenobacter thermophilus* *(26, 27)*	B
				Not reported	δαβγ	α 40kDa(25) β 31kDa(25) γ 24kDa(25) δ 12kDa(25)	Not reported	α 378 aa β 286 aa γ 189 aa δ 67 aa	*Methanothermobacter thermautotrophicus* ※ (25)	A
VOR	2-ketoisovalerate (keto-valine), 2-ketoisocaproate (keto-leucine), 2-keto-3-methylvalerate (keto-isoleucine), 2-keto-methylthiobutyrate (keto-methionine) + $^-$SCoA(28)	(Acyl)-CoA + CO$_2$ + 2e$^-$	Y(28)	(αβγδ)$_2$(28)	γδαβ§	α 47kDa(28) β 34kDa(28) γ 23kDa(28) δ 13kDa(28)	230kDa(28)	α 390 aa β 311 aa γ 186 aa δ 105 aa	*Thermococcus litoralis*(28)	A
						α 44.0kDa(22) β 34.8kDa(22) γ 20.0kDa(22) δ 11.9kDa(22)	Not reported	α 394 aa β 311 aa γ 185 aa δ 105 aa	*Pyrococcus furiosus*(22, 28)	A
					δα(βγ)	α 37kDa(25) βγ 55kDa(25) δ 15kDa(25)	Not reported	α 352 aa βγ 477 aa δ 79 aa	*Methanothermobacter thermautotrophicus* ※ (25)	A

OFOR family proteins	substrate	product	reversible?	holoenzyme organization	gene order	sub. size SDS	holoenzyme size native	AA length	organism	domain
IOR	Indole-pyruvate + $^-$SCoA	S-2-(indol-3-yl) acetyl-CoA + CO_2	Y	$(\alpha\beta)_2$(29)	αβ	α 66kDa(22) β 23kDa(22)	180kDa(29)	α 646 aa β 214 aa	*Pyrococcus furiosus*(22)	A
						α 67kDa(25) β 23kDa(25)	Not reported	α 618 aa β 196 aa	*Methanothermobacter thermautotrophicus* ※ (25)	A
OOR	Oxalate and slow activity on variety of 2-oxoacids(30, 31)	$2CO_2 + 2e^-$ (30, 31)	N(30)	$(\alpha\beta\gamma)_2$(30)	αγβ(30)	α 42 kDa(31) β 39 kDa(31) γ 35 kDa(31)	243kDa(31)	α 395 aa(31) β 314 aa(31) γ 315 aa(31)	*Moorella thermoacetica*(30, 31)	B

Table 2: The OFOR (2-oxoacid: ferredoxin oxidoreductase) family of enzymes, with several representative examples of each enzyme.

*The *por* genes in *Thermococcus guaymasensis* and *Pyrococcus furiosus* are integrated with the 2-ketoisovalerate ferredoxin oxidoreductase (*vor)* genes. The two enzymes share the same γ subunit. In place of the asterisk are the *vor* δαβ subunits.

†Both the PDC and PFOR modes of activity *require* coenzyme A. However, only in the PFOR mode of activity does it serve as a *substrate*. In PDC mode the *por* enzyme catalyzes the reaction all the way through to acetaldehyde, integrating into its own function the function served by ACSI and ACSII wherein acetyl-CoA is converted to acetaldehyde. The substrates and products shown in the chart reflect the enzyme when it is functioning as a PFOR.

‡ *Sulfolobus tokodaii* has two copies of the OFOR; in the first, the alpha subunit is 627aa in length, and the second alpha subunit is 628aa in length. Both beta subunits are 305aa.

¤*Halobacterium salinarum* (formerly known as *Halobacterium halobium*) was originally noted to have a "PFOR" in the 1981 purification paper, but it actually has a broad substrate specificity on 2-oxoacids (14). Therefore, it is here referred to as an "OFOR".

§ These VORs are the same as those integrated into the PFOR/PDC genes. They share a gamma subunit and the *porDAG* genes follow. Please see figure 13 for clarification.

※*Methanothermobacter thermautotrophicus* was formerly known as *Methanobacterium thermoautotrophicum* and is named as such in the original isolation papers for its OFORs.

References

1. **Mukund S, Adams MW.** 1993. Characterization of a novel tungsten-containing formaldehyde ferredoxin oxidoreductase from the hyperthermophilic archaeon, Thermococcus litoralis. A role for tungsten in peptide catabolism. Journal of Biological Chemistry **268:**13592-13600.

2. **Roy R, Menon AL, Adams MWW.** 2001. [11] Aldehyde Oxidoreductases from Pyrococcus furiosus, p 132-144, Methods in Enzymology, vol 331. Academic Press.

3. **Heider J, Ma K, Adams MW.** 1995. Purification, characterization, and metabolic function of tungsten-containing aldehyde ferredoxin oxidoreductase from the hyperthermophilic and proteolytic archaeon Thermococcus strain ES-1. Journal of Bacteriology **177:**4757-4764.

4. **Roy R, Mukund S, Schut G, M. Dunn D, Weiss R, W. W. Adams M.** 1999. Purification and Molecular Characterization of the Tungsten-Containing Formaldehyde Ferredoxin Oxidoreductase from the Hyperthermophilic Archaeon Pyrococcus furiosus: the Third of a Putative Five-Member Tungstoenzyme Family, vol 181.

5. **Hu Y, Faham S, Roy R, Adams MWW, Rees DC.** 1999. Formaldehyde ferredoxin oxidoreductase from Pyrococcus furiosus: the 1.85 Å resolution crystal structure and its mechanistic implications11Edited by I. A. Wilson. Journal of Molecular Biology **286:**899-914.

6. **Roy R, Adams MWW.** 2002. Characterization of a Fourth Tungsten-Containing Enzyme from the Hyperthermophilic Archaeon Pyrococcus furiosus. Journal of Bacteriology **184:**6952-6956.

7. **Bevers LE, Bol E, Hagedoorn P-L, Hagen WR.** 2005. WOR5, a Novel Tungsten-Containing Aldehyde Oxidoreductase from Pyrococcus furiosus with a Broad Substrate Specificity. Journal of Bacteriology **187:**7056-7061.

8. **Mukund S, Adams MWW.** 1995. Glyceraldehyde-3-phosphate Ferredoxin Oxidoreductase, a Novel Tungsten-containing Enzyme with a Potential Glycolytic Role in the Hyperthermophilic Archaeon Pyrococcus furiosus. Journal of Biological Chemistry **270:**8389-8392.

9. **Park M-O, Mizutani T, Jones PR.** 2007. Glyceraldehyde-3-Phosphate Ferredoxin Oxidoreductase from Methanococcus maripaludis. Journal of Bacteriology **189:**7281-7289.

10. **Reher M, Gebhard S, Schönheit P.** 2007. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR) and nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN), key enzymes of the respective modified Embden-Meyerhof pathways in the hyperthermophilic crenarchaeota *Pyrobaculum aerophilum* and *Aeropyrum pernix*. FEMS Microbiol Lett **273:**196-205.

11. **Matsubara K, Yokooji Y, Atomi H, Imanaka T.** 2011. Biochemical and genetic characterization of the three metabolic routes in Thermococcus kodakarensis linking glyceraldehyde 3-phosphate and 3-phosphoglycerate. Molecular Microbiology **81:**1300-1312.

12. **Chan MK, Mukund S, Kletzin A, Adams MW, Rees DC.** 1995. Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. Science **267:**1463.

13. **Zhang Q, Iwasaki T, Wakagi T, Oshima T.** 1996. 2-Oxoacid:Ferredoxin Oxidoreductase from the Thermoacidophilic Archaeon, <i>Sulfolobus</i> sp. Strain 7. The Journal of Biochemistry **120:**587-599.

14. **Lorenz K, Dieter O.** 1981. Purification and Properties of Two 2-Oxoacid:Ferredoxin Oxidoreductases from Halobacterium halobium. European Journal of Biochemistry **116:**587-594.

15. **Yan Z, Maruyama A, Arakawa T, Fushinobu S, Wakagi T.** 2016. Crystal structures of archaeal 2-oxoacid:ferredoxin oxidoreductases from Sulfolobus tokodaii. Scientific Reports **6:**33061.

16. **Takenaka M, Yoon K-S, Matsumoto T, Ogo S.** 2017. Acetyl-CoA production by encapsulated pyruvate ferredoxin oxidoreductase in alginate hydrogels. Bioresource Technology **227:**279-285.

17. **Chabrière E, Charon MH, Volbeda A, Pieulle L, Hatchikian EC, Fontecilla–Camps JC.** 1999. Crystal structures of the key anaerobic enzyme pyruvate:ferredoxin oxidoreductase, free and in complex with pyruvate. Nature Structural Biology **6:**182.

18. **Charon M-H, Volbeda A, Chabriere E, Pieulle L, Fontecilla-Camps JC.** 1999. Structure and electron transfer mechanism of pyruvate:ferredoxin oxidoreductase. Current Opinion in Structural Biology **9:**663-669.

19. **Pieulle L, Magro V, Hatchikian EC.** 1997. Isolation and analysis of the gene encoding the pyruvate-ferredoxin oxidoreductase of Desulfovibrio africanus, production of the recombinant enzyme in Escherichia coli, and effect of carboxy-terminal deletions on its stability. Journal of Bacteriology **179:**5684-5692.

20. **Chen PY-T, Aman H, Can M, Ragsdale SW, Drennan CL.** 2018. Binding site for coenzyme A revealed in the structure of pyruvate:ferredoxin oxidoreductase from Moorella thermoacetica. Proceedings of the National Academy of Sciences **115:**3846-3851.

21. **Hughes NJ, Clayton CL, Chalk PA, Kelly DJ.** 1998. Helicobacter pylori porCDAB and oorDABCGenes Encode Distinct Pyruvate:Flavodoxin and 2-Oxoglutarate:Acceptor Oxidoreductases Which Mediate Electron Transport to NADP. Journal of Bacteriology **180:**1119.

22. **Kletzin A, Adams MW.** 1996. Molecular and phylogenetic characterization of pyruvate and 2-ketoisovalerate ferredoxin oxidoreductases from Pyrococcus furiosus and pyruvate ferredoxin oxidoreductase from Thermotoga maritima. Journal of Bacteriology **178:**248-257.

23. **Eram MS, Oduaran E, Ma K.** 2014. The Bifunctional Pyruvate Decarboxylase/Pyruvate Ferredoxin Oxidoreductase from Thermococcus guaymasensis. Archaea **2014:**13.

24. **Kunow J, Linder D, Thauer RK.** 1995. Pyruvate: ferredoxin oxidoreductase from the sulfate-reducing Archaeoglobus fulgidus: molecular composition, catalytic properties, and sequence alignments. Archives of Microbiology **163:**21-28.

25. **Tersteegen A, Linder D, Thauer Rudolf K, Hedderich R.** 2004. Structures and Functions of Four Anabolic 2-Oxoacid Oxidoreductases in Methanobacterium Thermoautotrophicum. European Journal of Biochemistry **244:**862-868.

26. **Yun N-R, Arai H, Ishii M, Igarashi Y.** 2001. The Genes for Anabolic 2-Oxoglutarate: Ferredoxin Oxidoreductase from Hydrogenobacter thermophilus TK-6. Biochemical and Biophysical Research Communications **282:**589-594.

27. **Yoon KS, Ishii M, Igarashi Y, Kodama T.** 1996. Purification and characterization of 2-oxoglutarate:ferredoxin oxidoreductase from a thermophilic, obligately chemolithoautotrophic bacterium, Hydrogenobacter thermophilus TK-6. Journal of Bacteriology **178:**3365-3368.

28. **Heider J, Mai X, Adams MW.** 1996. Characterization of 2-ketoisovalerate ferredoxin oxidoreductase, a new and reversible coenzyme A-dependent enzyme involved in peptide fermentation by hyperthermophilic archaea. Journal of Bacteriology **178:**780-787.

29. **Mai X, Adams MW.** 1994. Indolepyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon Pyrococcus furiosus. A new enzyme involved in peptide fermentation. Journal of Biological Chemistry **269:**16726-16732.

30. **Gibson MI, Brignole EJ, Pierce E, Can M, Ragsdale SW, Drennan CL.** 2015. The Structure of an Oxalate Oxidoreductase Provides Insight into Microbial 2-Oxoacid Metabolism. Biochemistry **54:**4112-4120.

31. **Pierce E, Becker DF, Ragsdale SW.** 2010. Identification and Characterization of Oxalate Oxidoreductase, a Novel Thiamine Pyrophosphate-dependent 2-Oxoacid Oxidoreductase That Enables Anaerobic Growth on Oxalate. Journal of Biological Chemistry **285:**40515-40524.