THE EVOLUTION OF ACETYL-CoA SYNTHASE

PAUL A. LINDAHL[1*] and BELINDA CHANG[2,3]

[1] *Departments of Chemistry and of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843-3255, U.S.A.;* [2] *Department of Organismic and Evolutionary Biology, 22 Divinity Ave., Harvard University, Cambridge MA, 01238, U.S.A.;* [3] *University of Rome Tor Vergata, Rome, Italy*

Present address: The Rockefeller University, New York, NY 10021-6399, U.S.A., e-mail: changb@rockvax.rockefeller.edu

(Author for correspondence, e-mail: Lindahl@mail.chem.tamu.edu)*

(Received 1 August 2000; accepted in revised form 7 March 2001)

Abstract. Acetyl-coenzyme A synthases (ACS) are Ni–Fe–S containing enzymes found in archaea and bacteria. They are divisible into 4 classes. Class I ACS's catalyze the synthesis of acetyl-CoA from CO_2 + $2e^-$, CoA, and a methyl group, and contain 5 types of subunits (α, β, γ, δ, and ε). Class II enzymes catalyze essentially the reverse reaction and have similar subunit composition. Class III ACS's catalyze the same reaction as Class I enzymes, but use pyruvate as a source of CO_2 and $2e^-$, and are composed of 2 autonomous proteins, an $\alpha_2\beta_2$ tetramer and a $\gamma\delta$ heterodimer. Class IV enzymes catabolize CO to CO_2 and are α-subunit monomers. Phylogenetic analyses were performed on all five subunits. ACS α sequences divided into 2 major groups, including Class I/II sequences and Class III/IV-like sequences. Conserved residues that may function as ligands to the B- and C-clusters were identified. Other residues exclusively conserved in Class I/II sequences may be ligands to additional metal centers in Class I and II enzymes. ACS β sequences also separated into two groups, but they were less divergent than the α's, and the separation was not as distinct. Class III-like β sequences contained ~300 residues at their N-termini absent in Class I/II sequences. Conserved residues identified in β sequences may function as ligands to active site residues used for acetyl-CoA synthesis. ACS γ-sequences separated into 3 groups (Classes I, II, and III), while δ-sequences separated into 2 groups (Class I/II and III). These groups are less divergent than those of α sequences. ACS ε-sequence topology showed greater divergence and less consistency vis-à-vis the other subunits, possibly reflecting reduced evolutionary constraints due to the absence of metal centers. The α subunit phylogeny may best reflect the functional diversity of ACS enzymes. Scenarios of how ACS and ACS-containing organisms may have evolved are discussed.

Keywords: acetogens, anaerobic, archaea, carbon monoxide dehydrogenase, chemo-autotrophic, methanogens, origin of life, phylogenetic analysis

Abbreviations: ACS, acetyl-CoA synthase, identical to CODH, carbon monoxide dehydrogenase; PFOR, pyruvate ferredoxin oxidoreductase.

1. Introduction

Life apparently originated within a 0.5 billion year period between ~3.9 bya, when the Earth is thought to have cooled enough to allow life, and ~3.4 bya, when cyanobacteria-like photosynthetic organisms appeared (Mojzsis *et al*., 1996;

Origins of Life and Evolution of the Biosphere **31:** 403–434, 2001.

Schopf, 1996; Davis, 1999 and references therein). Scenarios describing the life-yielding events of this period can be categorized according to the raw materials assumed present (Maden, 1995; Edwards, 1998). Heterotrophic models assume that RNA, proteins and other complex molecules were present as the result of abiotic processes (Gilbert, 1986; Joyce, 1989; Campbell, 1991; Joyce *et al*., 1993). One such RNA strand is proposed to have catalytically synthesized copies of itself from available nucleotides. Proteins eventually replaced this function and RNA became a template for protein synthesis. Other enzymes degraded available carbohydrates and harnessed the energy released in those processes, while still others used that energy to drive nonspontaneous components of what had become a mutually catalytic (Kauffman, 1986, 1993) or hypercyclic (Eigen and Schuster, 1977, 1978, 1978) set of reactions. Available lipids spontaneously formed membranes that encapsulated these molecules, giving rise to the first living systems – anaerobic heterotrophs growing in an environment containing high levels of energy-rich nutrients.

In contrast, chemo-autotrophic models of how life began assume that organic molecules were not present and that the first set of autocatalytic reactions involved only simple inorganic compounds. Wächtershäuser has proposed a detailed model in which the first autocatalytic reaction was the reduction of CO_2 to activated acetyl groups at high temperatures on a positively-charged Fe–S (or Ni–Fe–S) surface (Wächtershäuser, 1988, 1990, 1992, 1997; Huber and Wächtershäuser, 1997). Additional reactions 'grafted on', leading to a mutually catalytic set of negatively charged surface-bound components called *surface metabolists*. These metabolists were eventually enclosed by hemispherical lipid membranes, which they generated, and were endowed with the capacity to synthesize proteins using that surface as a template. Certain proteins sequestered Ni–Fe–S ions from the surface as they lifted off into the cytoplasm, and self-associated to form the universal ancestor of acetyl-CoA synthase (ACS).

This ancient metalloenzyme catalyzed the synthesis of an analog of acetyl-CoA, from a methyl group, CO_2 and reducing equivalents (derived from H_2S). ACS served to increase the level and variety of enzymes in these metabolists. Some of these enzymes catalyzed the synthesis of primitive nucleic acids, which replaced the surface as a template for protein synthesis. At that point, a membrane-encapsulated organism abandoned the surface and began to grow and evolve.

The means by which this first living system evolved, and the group(s) of extant organisms into which it evolved are controversial (Achenbach-Richter, 1987; Woese, 1987, 1990, 1998; Olsen, 1994; Gogarten *et al*., 1996; Brown and Doolittle, 1997; Gupta, 1998). However, the Euryarchaeotes and the low G + C Gram-positive group of bacteria are located near (or at) the root of the phylogenetic tree, and there seems to be some consensus that the universal ancestor had properties of these organisms. This view is compatible with Wächtershäuser's model (1988), in that the Euryarchaota include anaerobic hyperthermophilic chemo-autotrophs such as *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, and

Methanopyrus kandleri. These archaea grow exclusively on simple inorganic compounds in the absence of O_2, at temperatures between 65–110 °C (Jones *et al*., 1983; Zeikus and Wolfe, 1972; Kurr *et al*., 1991; Burggraf, 1991). They reduce CO_2 to methane using H_2, and harness the free energy released by this process. During methanogenesis, CO_2 is reduced to a methyl group bound to tetrahydromethanopterin (H_4MPT), an archaeal analog of tetrahydrofolate (THF) (Thauer, 1998). The methyl group of CH_3-H_4MPT is transferred to coenzyme M, forming CH_3-S-CoM. This thioether reacts with Coenzyme B to generate CH_4 and the heterodisulfide CoM-SS-CoB. CoM-SS-CoB is the terminal electron acceptor used in oxidative phosphorylation.

The anabolic needs of these cells are met by synthesizing acetyl-coenzyme A using the Acetyl-CoA (or Wood-Ljungdahl) pathway (Thauer, 1998; Bhatnagar *et al*., 1991; Wood and Ljungdahl, 1991). During this process, CO_2 is reduced to CO in a reaction catalyzed by the modern-day ACS. ACS also catalyzes the synthesis of acetyl-CoA from CH_3-H_4MPT, CO and CoA. Acetyl-CoA is a central metabolite from which all other cellular components are ultimately derived. This aspect of their metabolism is schematically represented in Figure 1A.

Besides growing autotrophically as just described, most methanogens of the order *Methanosarcinales* (e.g. *Methanosarcina thermophila*, *Methanosarcina barkeri*, and *Methanosarcina frisia*) can grow heterotrophically by catabolizing acetate to CH_4 and CO_2 (Thauer, 1998; Krzycki *et al*., 1982; Weimer *et al*., 1978; Bhatnagar, 1991; Ferry, 1995, 1999). The only Methanosarcinale known to grow exclusively on acetate is *Methanosaeta soehngenii* (Bhatnagar, 1991). During *acetoclastic methanogenesis*, acetate is activated to acetyl-CoA, and then degraded to the methyl group of CH_3-H_4MPT, CoA, CO_2 and $2e^-$ in a reaction catalyzed by ACS. This degradation is essentially the reverse of that catalyzed by ACS in obligate autotrophs. Finally, the methyl group and $2e^-$ are converted into CH_4 (and increased ATP levels), using the lower portion of the methanogenesis pathway (Figure 1B).

The ACS's from *Methanosarcina* are composed of five types of subunits, including α, β, γ, δ, and ε (Grahame and DeMoll, 1996; Lu *et al*., 1994). The $\alpha\varepsilon$ complex from *M. thermophila* has been isolated by treating the enzyme with SDS, and was found to catalyze only the reversible oxidation of CO to CO_2 (Lu *et al*., 1994). The larger (72 kda) α subunit contains (at least) two types of metal-sulfur clusters called 'B' and 'C', while the 17 kda ε subunit is devoid of metals. Studies of homologous enzymes from *Clostridium thermoaceticum* (a.k.a. *Moorella thermoacetica*) and *Rhodospirillum rubrum* (see below) suggest that the active-site *C-cluster* consists of a Ni ion coordinated by 2–3 N/O and ∼2 S donor atoms (Tan *et al*., 1992) and bridged to an Fe_4S_4 cube through an unknown ligand (Hu *et al*., 1996). Three iron ions of the cube appear to be 4-coordinate and bound to the protein via ∼3 cysteinyl amino acid residues, while the remaining iron (the one most likely bridged to the Ni) is coordinated by 5–6 ligands including 1 histidine (DeRose *et al*., 1998). The *B-cluster* is a standard electron-transfer Fe_4S_4 cube probably coordinated by 4

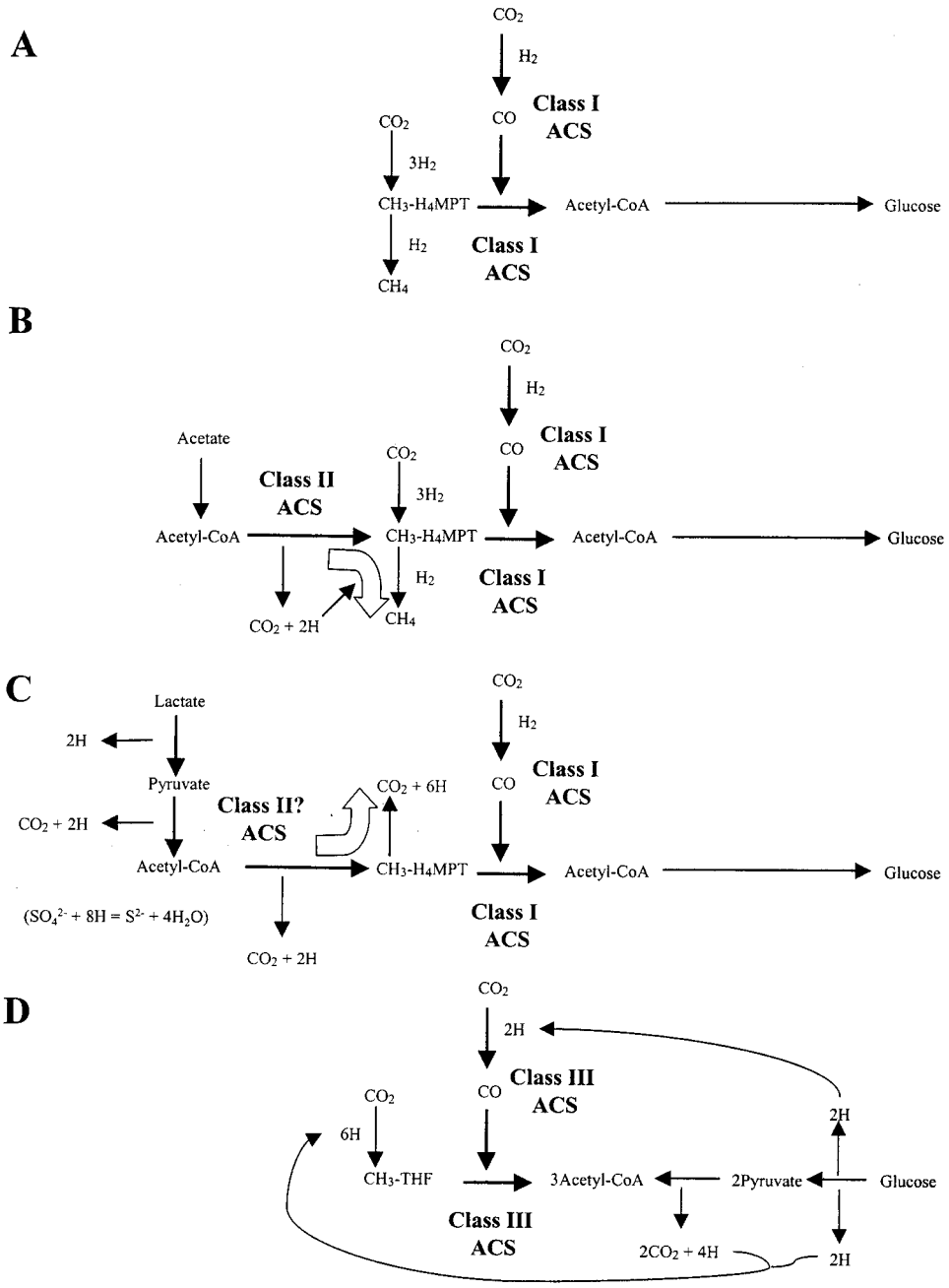

Figure 1. Simplified Metabolic Pathways Involving Acetyl-CoA Synthases. (A) Pathway used by obligate chemo-autotrophic methanogens, such as *M. jannaschii*, *M. thermoautotrophicum*, and *M. kandleri*. (B) Pathway used by facultative chemo-autotrophic methanogens able to catabolize acetate, such as *M. frisia* and *M. thermophila*. (C) Pathway used by sulfate-reducers such as *A. fulgidus*. (D) Pathway by homoacetogens such as *C. thermoaceticum*. The catabolic pathway involving ACS in purple non-sulfur bacteria is not shown.

cysteine residues. The α subunits from 3 different species have been reported to contain 9 conserved cysteines, proposed to be ligands to these clusters (Kerby *et al*., 1992; DeRose *et al*., 1998).

The β subunits of ACS contain a different Ni–X–Fe_4S_4 cluster called 'A' (Xia *et al*., 1997; Xia and Lindahl, 1996). In some enzymes, the *A-cluster* is responsible for converting the acetyl group of acetyl-CoA to CO and a methyl group, while in others it promotes the reverse reaction. Sequences of homologous β subunits contain 6 conserved cysteines (Maupin-Furlow and Ferry, 1996), 4 of which probably coordinate the cube (Barondeau, 1997). The Ni ion is coordinated by $\sim$2 S and 2 N/O donors (Russell *et al*., 1998). Circumstantial evidence suggests that two mechanistically important low-potential redox-active cysteine ligands may be coordinated to (or located near) the Ni ion (Barondeau, 1997). During catalysis, a proteinacious base near the A-cluster, suggested to be a histidine, deprotonates/protonates the substrate/product CoA (Wilson and Lindahl, 1999).

A $\gamma\delta$ subunit complex of the *M. thermophila* ACS can also be isolated (Jablonski *et al*., 1993). The γ subunit contains an Fe_4S_4 cube while the δ subunit noncovalently binds a cobalamin. The cube is an electron transfer agent that reductively activates Co^{2+} to Co^{1+}, while the cobalamin is used to transfer a methyl cation group between CH_3-H_4MPT (or CH_3-THF) and the A-cluster (Menon and Ragsdale, 1998; Ragsdale and Kumar, 1996). Four conserved cysteine residues in the γ subunit are probably used to coordinate the Fe_4S_4 cluster (Menon and Ragsdale, 1998).

Archaeoglobus fulgidus is a hyperthermophilic archaeon closely related to *Methanosarcinales*, but it is not able to generate methane (Thauer, 1998; Möller-Zinkhan *et al*., 1990). Rather, it catabolizes lactate completely to $3CO_2$ while reducing sulfate to sulfide (Figure 1C). Lactate is oxidized to pyruvate, which is then converted to acetyl-CoA and CO_2. In an ACS-catalyzed reaction similar to that occurring in Methanosarcinales, acetyl-CoA is degraded to CO_2 and the methyl group of CH_3-H_4MPT. However, the methyl group is oxidized to CO_2 rather than being reduced to CH_4. The ability to use sulfate as the terminal electron acceptor obviates the need for *A. fulgidus* to generate and use CoM-SS-CoB for that purpose, and the lower portion of the methanogenesis pathway is not needed. The quaternary structure of the ACS from *A. fulgidus* is similar to that from *Methanosarcina* (Dai, 1998). *A. fulgidus* is metabolically flexible, in that it can also grow autotrophically on CO_2, H_2, and thiosulfate, using ACS to synthesize acetyl-CoA (Vorholt, 1995) By contrast, *Archaeoglobus lithotrophicus* only grows autotrophically. The related *Ferroglobus placidus* grows autotrophically on ferrous ion (or H_2) and nitrate (Vorholt, 1995, 1997). Both organisms also contain ACS and employ the Acetyl-CoA pathway.

Numerous sulfate-reducing *bacteria* also contain ACS, including *Desulfobacterium autotrophicum*, *Desulfobacterium vacuolatum*, *Desulfosarcina variabilis*, *Desulfococcus multivorans*, *Desulfotomaculum orientis*, *Desulfotomaculum acetoxidans*, *Desulfovibrio desulfuricans* and *Desulfovibrio vulgaris* (Klemps, 1985;

Schauder, 1986, 1989; Länge *et al*., 1989; Meyer, 1984, 1985). Some use the Acetyl-CoA pathway to grow autotrophically (on CO_2/H_2/sulfate) while others use it to ferment pyruvate or lactate completely to CO_2 (rather than stopping at the acetate level). *Desulfotomaculum acetoxidans* has been reported to synthesize two types of ACS enzymes, depending on whether cells are grown autotrophically or heterotrophically (Schauder *et al*., 1989). ACS enzymes in sulfate-reducing bacteria are not well characterized.

The anaerobic Ni-requiring homoacetogenic bacterium *C. thermoaceticum* catabolizes glucose, forming three acetates (Diekert and Thauer, 1978; Diekert *et al*., 1979; Wood and Ljungdahl, 1991; Drake, 1994). Glucose is converted into two pyruvates and $4e^-$ via glycolytic enzymes (Figure 1D). Pyruvate-ferredoxin oxidoreductase (PFOR) converts these pyruvates to 2 acetyl-CoA, 2 CO_2, and $4e^-$. The ACS in *C. thermoaceticum* uses $2e^-$ to reduce a CO_2 to CO, and a series of other enzymes use the other $6e^-$ to reduce another CO_2 to the methyl group of CH_3-THF. Finally, ACS catalyzes the reaction of CO, coenzyme A, and CH_3-THF, forming acetyl-CoA. Some acetyl-CoA is hydrolzyed to acetate, and free energy released in the process is used to drive ATP synthesis. The remaining acetyl-CoA is used in anabolism.

ACS from *C. thermoaceticum* and *Acetobacterium woodii* are $\alpha_2\beta_2$ tetramers (Xia *et al*., 1996; Balch *et al*., 1977; Ragsdale *et al*., 1983). Their α and β subunits are homologous to the β and α subunits, respectively, of the 5-subunit ACS complexes. *C. thermoaceticum* contains an autonomous corrinoid-iron-sulfur protein (CoFeSP) that transfers the methyl group from CH_3-THF to the A-cluster of this ACS (Ragsdale *et al*., 1987; Ragsdale and Kumar, 1996). CoFeSP is an $\alpha\beta$ dimer whose subunits are homologous to the γ and δ subunits of 5-subunit ACS's*. Acetogens are not known to contain homologs to the ε subunits of the 5-subunit enzymes.

ACS-related enzymes are found in carboxydotrophic bacteria such as the purple non-sulfur photosynthetic *R. rubrum* (Bonam and Ludden, 1987). This organism can grow in the dark by oxidizing CO to CO_2, reducing $2H^+$ to H_2, and coupling these processes to synthesize ATP (Kerby *et al*., 1995). The monomeric enzyme that catalyzes the oxidation of CO to CO_2 is homologous to the α subunit of the 5-component ACS's (Kerby *et al*., 1992, 1995)†.

We have classified known ACS homologs into four classes, based on catalytic activity, metabolic function, and protein characteristics. *Class I* ACS's are anabolic

* To allow comparison with methanogenic 5-subunit $\alpha\beta\gamma\delta\varepsilon$ ACS enzymes, the β and α subunits of the ACS and the β and α subunits of the CoFeSP from *C. thermoaceticum* are referred to as the α, β, γ and δ subunits, respectively.

† This enzyme from *Rhodospirillium rubrum* is commonly called carbon monoxide dehydrogenase (CODH). Due to its homology with the α subunit of ACS's, we refer to it as a type of ACS even though it does not catalyze the synthesis of acetyl-CoA. Likewise, the enzymes from Methanosarcina, commonly called acetyl-CoA decarbonylase/synthases (ACDS), will be referred to as ACS's. These designations allow us to use a single nomenclature for ACS-related homologous subunits from all organisms.

enzymes that catalyze the synthesis of acetyl-CoA, using CO_2 as one of its substrates and H_2 (via hydrogenase) as electron donor. *Class II* ACS's are catabolic enzymes that degrade the acetyl group of acetyl-CoA to a methyl group, CO_2 and $2e^-$. Both Class I and II enzymes are 5-subunit ($\alpha\beta\gamma\delta\varepsilon$) complexes. *Class III* enzymes catalyze the synthesis of acetyl-CoA, using pyruvate (via PFOR) as an electron donor and source of CO_2. They consist of two proteins, an $\alpha_2\beta_2$ tetramer that reduces CO_2 to CO and then combines CO with the methyl group and CoA to form acetyl-CoA, and a $\gamma\delta$ heterodimer that transfers the methyl group to and from the tetramer. Class III enzymes are used in both catabolism (facilitating reduction of the terminal electron acceptor in anaerobic respiration) and anabolism (promoting the synthesis of the metabolite acetyl-CoA). *Class IV* enzymes catalyze only the oxidation of CO to CO_2 and $2e^-$, and are used in catabolism. They are monomers that consist of only an α-subunit.

ACS's are found in four groups of organisms, including Euryarchaeotes, low G + C Gram-positive bacteria, α-proteobacteria, and δ-proteobacteria. All ACS-containing organisms except α-proteobacteria are strict anaerobes. The ACS-containing α-proteobacterium *R. rubrum* is a facultative anaerobe. The orders within the Euryarchaeotes that contain ACS include the Methanobacteriales, Methanococcales, Methanopyrales, Methanosarcinales and Archaeoglobales. Members of the first three orders are hyperthermophilic methanogens that are *obligate* chemo-autotrophs, while many members of the latter two orders are methanogens that grow at less extreme temperatures and are *facultative* chemo-autotrophs. Low G + C Gram-positive bacteria that contain ACS include homoacetogens and other Clostridia, as well as Gram-positive sulfate-reducers. δ-proteobacteria include Gram-negative sulfate-reducers.

The goal of this study was to determine how ACS evolved and evaluate the roles it has served during evolution. Our approach was to examine the phylogenies of the subunits, and correlate these to the phylogenies, metabolism and genetic composition of extant organisms that contain ACS.

2. Methods

ACS subunit sequences were identified in BLAST searches of GenBank, and aligned using a combination of the GCG PILEUP routine, NCBI Entrez comparisons, the ClustalW package, and visual inspection. Alignments for each of the subunits are available online at http://sakmarsgi.rockefeller.edu/~changb/Public/ACS.

Phylogenetic analyses of protein sequences using maximum parsimony as well as distance methods (Swofford *et al.*, 1996; Swofford, 1999), including neighbor-joining (Saitou and Nei, 1987) and minimum evolution (Rzhetsky and Nei, 1992), were performed using a beta test version of the PAUP* v.4.0b2 program (Sinauer Associates, Sunderland, MA). Bootstrap methods were used to assess the degree of confidence of nodes in the phylogeny (Felsenstein, 1985).

Maximum likelihood phylogenetic methods (Felsenstein, 1981; Kishino *et al*., 1990) using models for protein sequences (JTT; Jones *et al*., 1992) implemented in the program MOLPHY (Adachi and Hasegawa, 1994) were used to compare the likelihood of candidate tree topologies generated by the parsimony and distance methods. A search strategy implemented in MOLPHY called 'star decomposition' was also used. Although this does not guarantee the maximum likelihood topology, it is similar to the neighbor-joining algorithm (Saitou and Nei, 1987), in that it starts from a star phylogeny, and sequentially groups pairs of taxa that give the highest likelihood score for the tree at each step (Adachi and Hasegawa, 1994). Finally, maximum likelihood distances between pairs of taxa were calculated using MOLPHY, and this distance matrix was also used to reconstruct a phylogeny using neighbor-joining methods. All computer analyses were performed on Macintoshes G3 and G4, and a Sun UltraSPARC 1 model 140 workstation.

3. Results and Discussion

3.1. Phylogeny of α subunits

Seventeen α-subunit sequences, including 3 from *A. fulgidus*, 2 each from *M. jannaschii*, *M. frisia*, *C. difficile*, and *C. acetobutylicum*, and 1 each from *M. kandleri*, *C. thermoaceticum*, *R. rubrum*, *M. thermophila*, *M. soehngenii*, and *M. thermoautotrophicum* were obtained from GenBank, aligned (Figure 2) and subjected to phylogenetic analysis using a variety of algorithms. For this data set, the same topology (Figure 3) was found using maximum likelihood methods (with the star decomposition algorithm), and neighbor-joining methods with ML-estimated distances, both calculated under the JTT substitution model (Jones *et al*., 1992). Table I shows the pairwise distances estimated by maximum likelihood under the JTT model. Bootstrap analysis, under maximum parsimony and NJ distance methods, show well-supported divergences into two groups, including Class I/II sequences and Class III/IV-like sequences, as well as for several nodes within each of these groups. All Class I/II sequences A–H originate from archaea while most Class III/IV-like sequences (I–L, N, O) originate from bacteria. Three Class III/IV-like sequences (M, P, Q) originate from archaea. Thus, this tree deviates from the 16S rRNA tree (Woese, 1987), possibly the result of lateral gene transfer.

Most nodes within the group of Class III/IV-like are well supported. A subgroup within Class I/II, consisting of *M. frisia* sequences D and E, and *M. thermophila* sequence F, also shows high bootstrap support, as does the basal node separating the *M. thermoautotrophicum* sequence A and *M. jannaschii* sequence B from the others. Moreover, these results are robust to exclusions of the N- and C-termini from the phylogenetic analysis. However, other basal nodes within the group of Class I/II show very weak bootstrap support, with the various methods supporting conflicting resolutions of these nodes, especially in the placement of the *A. fulgidus*

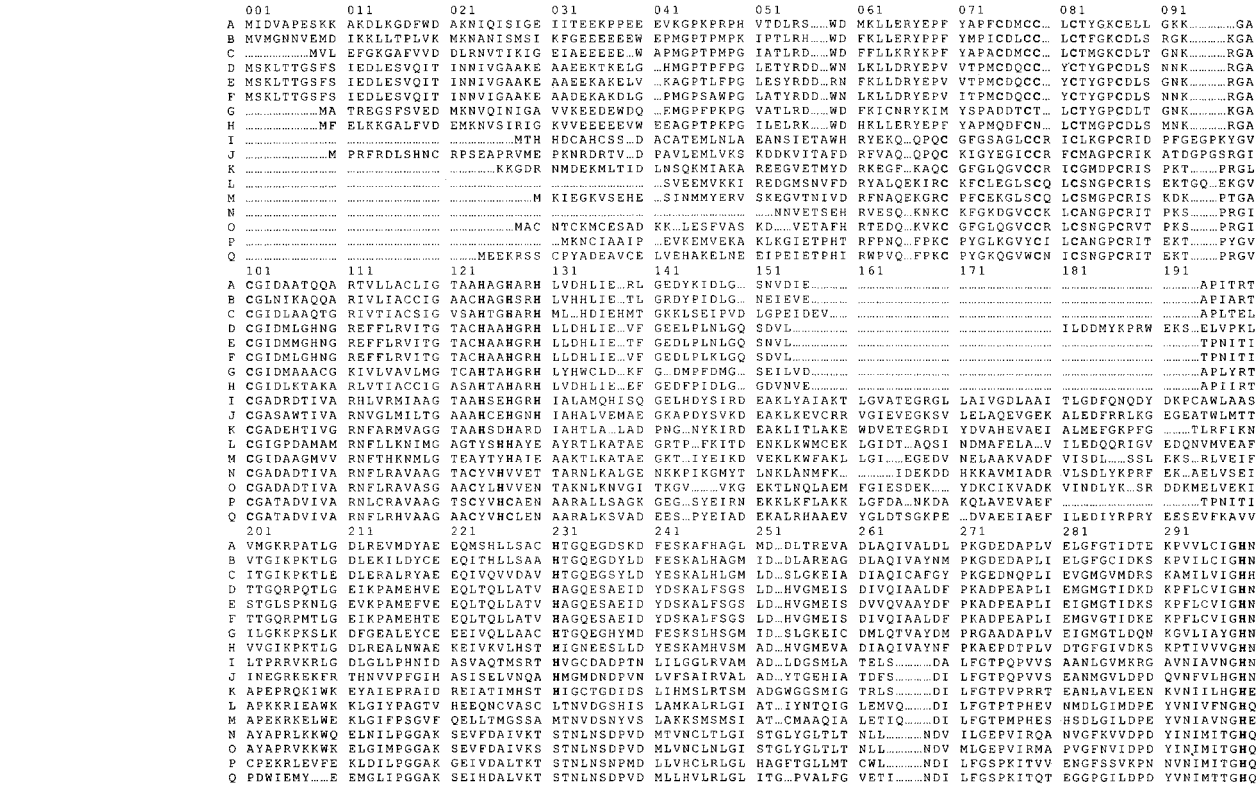

Figure 2. Alignment of ACS α Subunit Sequences. (A) *M. thermoautotrophicum* (AAB86180); (B) *M. jannaschii* (Q57617); (C) *A. fulgidus* (AAB91266); (D) *M. frisia IA* (Q49161); (E) *M. frisia* IIA (Q49163); (F) *M. thermophila* (AAC44650); (G) *M. soehngenii* (A39764); (H) *A. fulgidus* (AAB90136); (I) *R. rubrum* (P31896); (J) *C. thermoaceticum* (P27989); (K) *C. difficile* (Sanger Center); (L) *C. acetobutylicum* (Genome Therapeutics); (M) *A. fulgidus* (AAB89403); (N) *C. acetobutylicum* (Genome Therapeutics), unfinished genome); (O) *C. Difficile* (Sanger Centre, unfinished genome); (P) *M. jannaschii* (AAB98724); (Q) *M. kandleri* (Shin *et al*., 1999). Positions 901 and 902 of sequence H is PR (not shown). Selected conserved sequences have been highlighted.

Figure 2. (continued)

Figure 2. (continued)

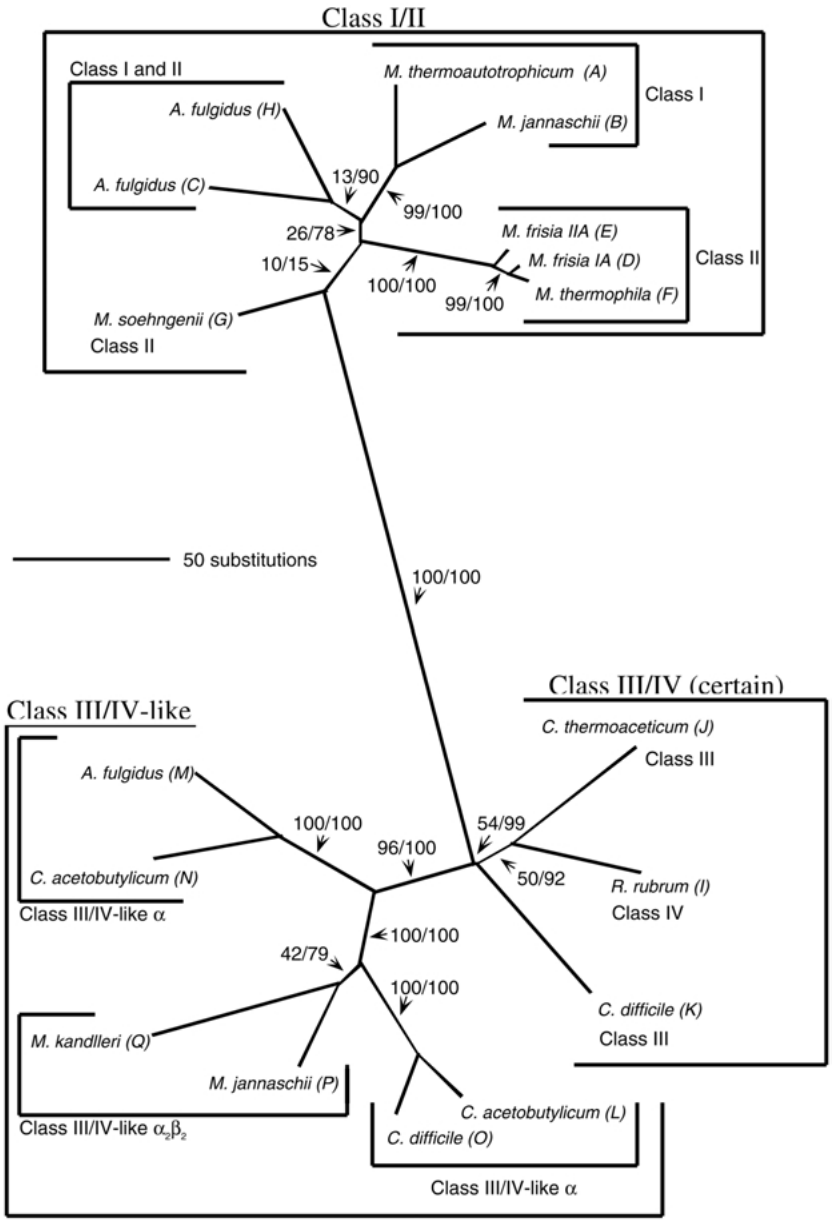

Figure 3. Phylogeny of ACS α-Subunit Amino Acid Sequences, based on a data set of 17 aligned sequences, and 902 sites (see Figure 2). This tree represents the maximum likelihood topology found using the star decomposition method under the JTT model (Jones *et al*., 1992). This topology was also found using neighbor-joining methods with maximum likelihood-estimated distances (Table I). Bootstrap percentage values for 100 replications are indicated above each node: the first value calculated using maximum parsimony with equally-weighted characters, and the second value calculated using neighbor-joining distance methods. Branch lengths were estimated by maximum likelihood under the JTT model.

TABLE I

ACS α subunit pairwise distances estimated by maximum likelihood under the JTT model (MOLPHY). See Figure 2 for accession numbers

C. difficile (O)	–	1.15	1.50	1.63	0.37	1.31	1.46	1.43	0.89	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
M. kandleri (Q)	1.15	–	1.85	1.82	1.09	1.56	1.67	1.46	0.90	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
C. difficile (K)	1.50	1.85	–	1.10	1.57	1.52	1.03	1.44	1.44	2.92	2.99	2.91	2.69	2.93	3.00	2.95	2.98
C. thermoaceticum (J)	1.63	1.82	1.10	–	1.65	1.55	0.91	1.49	1.47	2.77	2.85	2.81	2.63	2.68	2.90	2.76	2.95
C. acetobutylicum (L)	0.37	1.09	1.57	1.65	–	1.35	1.51	1.36	0.86	2.94	3.00	3.00	3.00	3.00	3.00	3.00	3.00
C. acetobutylicum (N)	1.31	1.56	1.52	1.55	1.35	–	1.54	0.74	1.27	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
R. rubrum (I)	1.46	1.67	1.03	0.91	1.51	1.54	–	1.48	1.38	2.76	2.82	2.79	2.61	2.59	2.68	2.71	2.67
A. fulgidus (M)	1.43	1.46	1.44	1.49	1.36	0.74	1.48	–	1.21	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
M. jannaschii (P)	0.89	0.90	1.44	1.47	0.86	1.27	1.38	1.21	–	3.00	3.00	3.00	2.97	3.00	3.00	3.00	3.00
M. frisia IA (D)	3.00	3.00	2.92	2.77	2.94	3.00	2.76	3.00	3.00	–	0.10	0.15	0.97	0.92	1.04	0.99	0.93
M. thermophila (F)	3.00	3.00	2.99	2.85	3.00	3.00	2.82	3.00	3.00	0.10	–	0.18	0.98	0.91	1.02	0.97	0.92
M. frisia IIA (E)	3.00	3.00	2.91	2.81	3.00	3.00	2.79	3.00	3.00	0.15	0.18	–	0.95	0.92	1.03	0.96	0.94
M. soehngenii (G)	3.00	3.00	2.69	2.63	3.00	3.00	2.61	3.00	2.97	0.97	0.98	0.95	–	0.93	0.98	1.00	0.83
M. thermoautotrophicum (A)	3.00	3.00	2.93	2.68	3.00	3.00	2.59	3.00	3.00	0.92	0.91	0.92	0.93	–	0.58	0.88	0.89
M. jannaschii (B)	3.00	3.00	3.00	2.90	3.00	3.00	2.68	3.00	3.00	1.04	1.02	1.03	0.98	0.58	–	0.87	0.89
A. fulgidus (H)	3.00	3.00	2.95	2.76	3.00	3.00	2.71	3.00	3.00	0.99	0.97	0.96	1.00	0.88	0.87	–	0.74
A. fulgidus (C)	3.00	3.00	2.98	2.95	3.00	3.00	2.67	3.00	3.00	0.93	0.92	0.94	0.83	0.89	0.89	0.74	–

sequences H and C, and *M. soehngenii* sequence G. The lack of resolution in these basal nodes of the Class I/II group of sequences is also reflected in the extremely short branch lengths of these internodes (Figure 3). The longer branches within the group of Class III/IV-like sequences, relative to those of Class I/II, indicates greater overall divergence among the Class III/IV-like sequences (see Table I), consistent with the greater metabolic diversity of the member organisms.

Class I/II and III/IV-like sequences contain several strictly conserved residues including 7 cysteines (C082, C087, C101, C387, C594, C625, and C675) and 1 histidine (H299). Two additional cysteines (C070, C079) and 2 aspartate residues are conserved within Class III/IV-like sequences. Twelve additional cysteines (C078, C334, C335, C485, C488, C491, C495, C526, C529, C532, C536, C801), 5 additional histidines (H124, H127, H130, H231, H680), 3 tyrosines, 1 tryptophan, and ~25 asparate/glutamate residues are conserved within Class I/II sequences, reflecting their greater homogeneity vis-a-vis Class III/IV-like sequences. The unusually large number of negatively charged conserved sequences (also evident in the β-sequences, see below) is consistent with Wächtershäuser's postulate (1988) that primordial proteins were negatively charged.

The two most studied α subunits from the Class III/IV-like sequences (from the Class III enzyme in *C. thermoaceticum* and the Class IV enzyme in *R. rubrum*) each contain one B- and C-cluster. These clusters are probably coordinated by ~9 cysteines, 1 histidine, and ~3 N/O donors (discussed above), using the 9 cysteines, 1 histidine, and aspartate residues conserved in Class III/IV-like sequences. H299 and C675 have been mutated in the enzyme from *R. rubrum*, the effects of which are consistent with them serving as ligands to the C-cluster (Spangler *et al*., 1997; Staples, 1999).

There are a total of 19 cysteine, 6 histidine and ~25 glutamate/aspartate residues conserved in Class I/II sequences. Eight of the cysteines are located in two short regions (483–498 and 512–570), and are spaced at intervals typical of those used to coordinate Fe_4S_4 clusters. Neither region exists in Class III/IV-like sequences. Using the three sequences available at that time (from *R. rubrum, C. thermoaceticum*, and *M. soehngenii*), Kerby *et al*. (1992) noted that the *M. soehngenii* sequence contained these extra cysteines. They suggested that the subunit from this organism has 2 Fe_4S_4 clusters not present in those from *R. rubrum* and *C. thermoaceticum*. Maupin-Furlow and Ferry (1996) also noticed the same motifs in the α sequence from *M. thermophila*. Our more extensive analysis confirms these earlier observations, and we now conclude that α subunits from Class I and II enzymes contain 2 extra Fe_4S_4 clusters that are not present in Class III- or IV-like enzymes. These extra clusters are probably used to transfer electrons to and from donor and acceptor proteins. A similar situation seems to exist between the Fe hydrogenases from *Desulfovibrio desulfuricans* and *Clostridium pasteurianum* (Peters *et al*., 1998; Nicolet *et al*., 1999). Both contain the active-site H-cluster and have homologous sequences, but the latter enzyme contains 2 extra Fe_4S_4

clusters. These clusters redirect electrons to locations required exclusively for the *C. pasteurianum* enzyme.

Two other regions in the Class I/II and III/IV-like sequences are also interesting from a comparative context. In Class I/II sequences, residues 124–130 inclusive contain 3 conserved histidine residues. Two Class III/IV-like sequences (from *C. thermoaceticum*, and *R. rubrum*) also have histidine in the conserved positions (*C. difficile* sequence K has histidine in two of the three positions), but the other Class III/IV-like sequences have 1–2 histidine residues in different patterns. This is consistent with the three Class III/IV-like sequences being more closely related to the Class I/II sequences, as reflected in the topology. These patterns suggest that Class I and II enzymes are more ancestral and that these conserved histidines may have gradually lost or modified their function in Class III- and IV-like enzymes.

A similar situation occurs with residues 332–335 inclusive. G332 and D(E)339 are conserved among both groups, but C334 and C335 are only strictly conserved in Class I/II sequences. These residues are conserved in 3 Class III/IV-like sequences (*C. thermoaceticum*, *R. rubrum*, and *C. difficile* sequence K) while the other Class III/IV-like sequences have 0–2 cysteine residues in other patterns around the conserved positions. This again suggests that these 2 cysteine residues serve a function in Class I and II enzymes that was lost or modified as Class III/IV-like enzymes evolved from Class I/II.

These His_3Cys_2 residues may be ligands that coordinate an additional (as of yet hypothetical) metal center in α subunits from Class I and II enzymes or they may serve another function. In summary, α subunits from Class I and II enzymes may contain the B- and C-clusters common to all α subunits, 2 extra Fe_4S_4 clusters, and possibly another center or function involving the His_3Cys_2 conserved residues. All α-subunits from Class III- and IV-like enzymes lack these two extra Fe_4S_4 clusters, and all except *C. thermoaceticum*, *R. rubrum* and *C. difficile* lack the same His_3Cys_2 functionality.

3.2. PHYLOGENY OF β SUBUNITS

Seven sequences of β subunits, including 2 from *M. jannaschii*, and 1 each from *C. difficile*, *C. thermoaceticum*, *M. thermophila*, *A. fulgidus*, and *M. thermoautotrophicum* were obtained from GenBank, and aligned (Alignments for the β, γ, δ, and ε sequences are available online at http://sakmarsgi.rockefeller.edu/∼changb/Public/ACS). The most obvious difference among the β subunit sequences is the presence of a ∼300 residue region in the N-terminus of the *C. difficile* sequence A, *C. thermoaceticum* sequence B, and *M. jannaschii* sequence C, and the absence of this region in the other sequences D–G. Maupin-Furlow and Ferry (1996) also observed this using a more limited data set. This region appears to be found in sequences from Class III enzymes but not those of Class I or II. The *M. jannaschii* enzyme affording sequence C, which is not known to be from a Class III enzyme but which has this region nevertheless, will be called Class III-like.

TABLE II

ACS β subunit pairwise distances estimated by maximum likelihood under the JTT model

C. difficile (A)	–	0.69	0.92	0.89	1.18	1.06	1.18
C. thermoaceticum (B)	0.69	–	0.93	0.94	1.08	1.03	1.12
M. jannaschii (C)	0.92	0.93	–	0.10	0.79	0.61	0.80
M. jannaschii (D)	0.89	0.94	0.10	–	0.74	0.56	0.80
M. thermophila (E)	1.18	1.08	0.79	0.74	–	0.74	0.56
M. thermoautotrophicum (F)	1.06	1.03	0.61	0.56	0.74	–	0.79
A. fulgidus (G)	1.18	1.12	0.80	0.80	0.56	0.79	–

Excluding this extra N-terminal region from the phylogenetic analysis results in a topology that separates Class I/II from Class III sequences, except for the Class III-like *M. jannaschii* sequence C (Figure 4). This sequence, with its N-terminal region excluded, is most closely related to the Class I/II *M. jannaschii* sequence D (which lacks such an N-terminal region) (Table II). Including the N-terminal region in the analysis does not result in cleanly separated Class I/II vs. Class III, in either the parsimony or distance analyses unless gapmode is set to 'newstate' in the parsimony analysis. This indicates that, although the presence of the N-terminal region distinguishes Class III-like from Class I/II sequences, the rest of the sequence does not support this distinction.

All β sequences have 6 conserved cysteines (C510, C513, C522, C532, C599, and C601), 3 conserved histidines (H411, H422, and H725), 1 conserved methionine (C339) and 16 conserved aspartate and glutamate residues. Four of the cysteine residues are probably used to coordinate the Fe_4S_4 component of the A-cluster (assuming that the bridge is a cysteine), and the Ni is probably coordinated by 1–2 additional cysteines and 2 N/O (likely histidine) ligands. The conserved methionine is a potential ligand to the Ni, as thioether coordination helps stabilize Ni in the observed 1+ valence (Farmer *et al.*, 1993; Russell *et al.*, 1998). The 2 cysteine ligands to the Ni may constitute the D-site (Barondeau and Lindahl, 1997), and another histidine residue may serve to deprotonate/protonate Coenzyme A during catalysis (Wilson and Lindahl, 1999). Thus, essentially all conserved cysteine, methionine, and histidine residues can be assigned using previously proposed ligands and functional groups, and it appears unlikely that these subunits contain additional metal centers.

3.3. PHYLOGENY OF γ, δ, AND ε SUBUNITS

Six γ and δ subunit sequences, and 7 ε subunits were obtained from GenBank and aligned. These include 1 γ and δ sequence each from *M. jannaschii*, *C. difficile*,

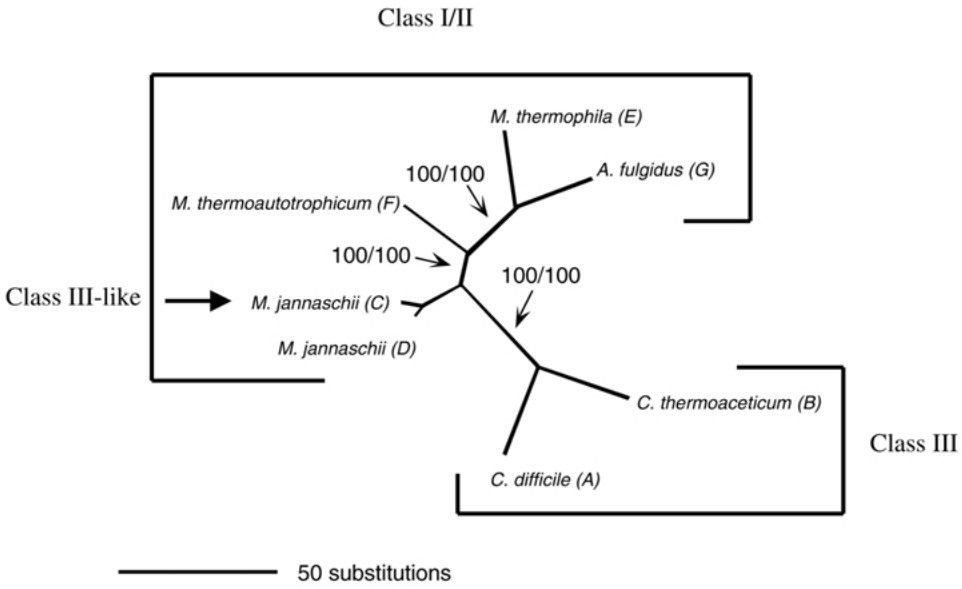

Figure 4. Phylogeny of ACS β-Subunit Amino Acid Sequences, based on a dataset of 7 aligned sequences, and 831 sites. Letters next to sequences correspond to those in the alignment, available at http://sakmarsgi.rockefeller.edu/∼changb/Public/ACS. This tree topology was found in 100 of 100 heuristic searches with random addition using maximum parsimony methods (with equally weighted characters), as well as in neighbor-joining analyses. Bootstrap percentage values for 1000 replications are indicated above each node, the first value calculated using parsimony methods, and the second value using NJ distance methods. This tree represents the maximum likelihood topology found using the star decomposition method (under the JTT model, excluding all gaps), as well as the topology found by neighbor-joining with maximum likelihood-estimated pairwise distances (Table II). The numbers of substitutions along each branch are proportional to the branch length, as indicated by the legend. These inferred branch lengths were estimated by maximum likelihood under the JTT model. Genbank accession numbers for sequences include *C. difficile* (Sanger Centre), *C. thermoaceticum* (P27988), *M. jannaschii* (AAB98134), *M. jannaschii* (E64319), *M. thermophila* (AAC44652), *M. thermoautotrophicum* (AAB86182), *A. fulgidus* (AAB90857).

TABLE III

ACS γ subunit pairwise distances estimated by maximum likelihood under the JTT model

C. difficile (A)	–	1.20	1.64	1.64	1.56	1.75
C. thermoaceticum (B)	1.20	–	1.23	1.53	1.48	1.40
M. thermoautotrophicum (C)	1.64	1.23	–	1.07	0.79	1.06
A. fulgidus (D)	1.64	1.53	1.07	–	1.21	0.72
M. jannaschii (E)	1.56	1.48	0.79	1.21	–	1.25
M. thermophila (F)	1.75	1.40	1.06	0.72	1.25	–

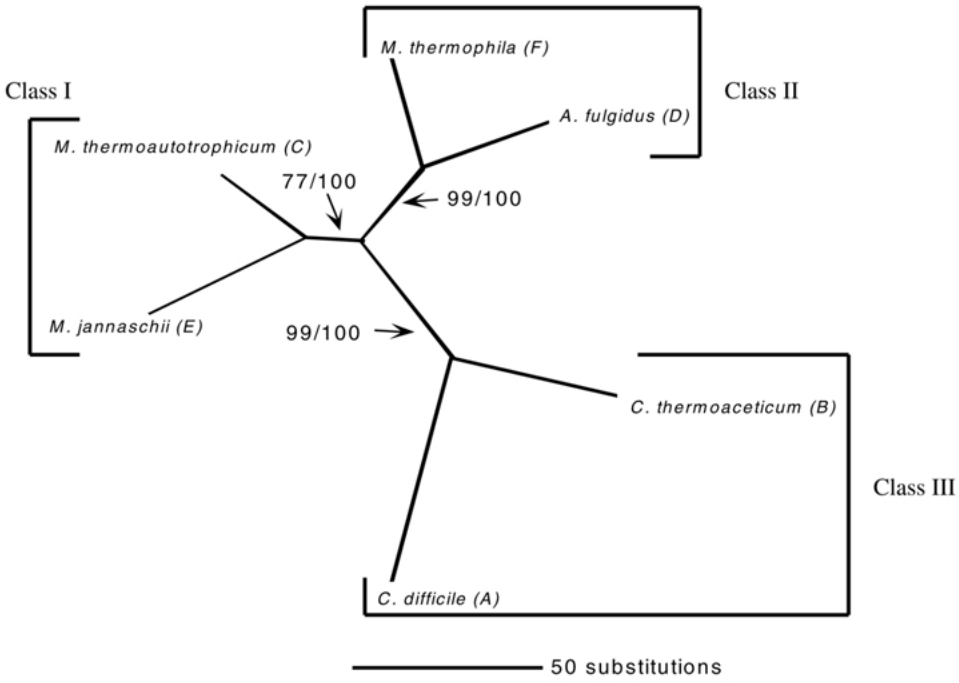

Figure 5. Phylogeny of ACS γ-Subunit Amino Acid Sequences, based on a dataset of 6 aligned sequences, and 503 sites. Letters next to sequences correspond to those in the alignment, available at http://sakmarsgi.rockefeller.edu/~changb/Public/ACS. This tree represents the maximum likelihood topology found using the star decomposition method under the JTT model (Table III), with branch lengths proportional to those estimated under this model. Bootstrap percentage values for 1000 replications are indicated above each node: the first value calculated using maximum parsimony with equally-weighted characters, and the second value calculated using neighbor-joining distance methods. The numbers of substitutions along each branch are proportional to the branch length, as indicated by the legend. GenBank accession numbers include *C. difficile* (Sanger Centre), *C. thermoaceticum* (Q07340), *M. thermoautotrophicum* (AAB86185), *A. fulgidus* (AAB90860), *M. jannaschii* (AAB98093), *M. thermophila* (Q50539).

C. thermoaceticum, *M. thermophila*, *A. fulgidus*, and *M. thermoautotrophicum*, 2 ε sequences from *A. fulgidus*, and 1 ε sequence each from *M. jannaschii*, *M. thermophila*, *M. thermoautotrophicum*, *M. frisia* and *M. soehngenii*. The phylogeny of the γ subunit shows well-supported divergences between sequences from Class I (*M. jannaschii*, *M. thermoautotrophicum*), Class II (*M. thermophila*, *A. fulgidus*), and Class III enzymes (*C. difficile*, *C. thermoaceticum*). However, the divergence between these three groups of sequences is not as great as that of α sequences (Figure 5 and Table III). This is also reflected in the shorter branch lengths dividing the groups in the γ subunit tree, and the smaller pairwise distances between members of different groups. As previously noted, 4 cysteine residues are conserved in the γ sequences (C019, C022, C027, and C044), and these serve to coordinate the Fe_4S_4 cluster in this subunit. Phylogenetic analysis of the δ subunit

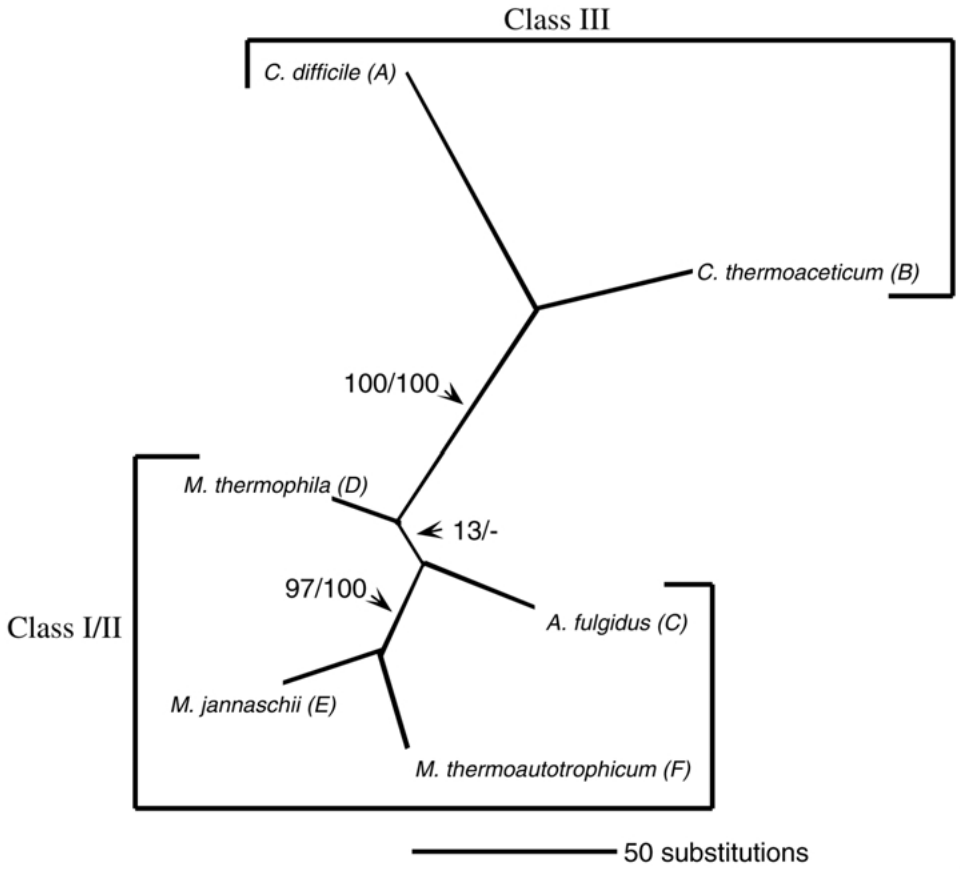

Figure 6. Phylogeny of ACS δ-Subunit Amino Acid Sequences, based on a dataset of 6 aligned sequences, and 476 sites. Letters next to sequences correspond to those in the alignment, available at http: //sakmarsgi.rockefeller.edu/∼changb/Public/ACS. This tree represents the maximum likelihood topology found using the star decomposition method under the JTT model (Table IV), with branch lengths proportional to those estimated under this model. Bootstrap percentage values are indicated above each node: maximum parsimony/neighbor-joining (for 100, and 1000 replications, respectively). The numbers of substitutions along each branch are proportional to the branch length, as indicated by the legend. GenBank accession numbers include *C. difficile* (Sanger Centre), *C. thermoaceticum* (Q07341), *A. fulgidus* (AAB90859), *M. thermophila* (Q50538), *M. jannaschii* (AAB98094), *M. thermoautotrophicum* (AAB86184).

shows a well-supported separation into Class I/II and III sequences (Figure 6 and Table IV).

Phylogenetic analysis of the ε sequences shows a well-supported monophyletic group of 3 sequences from Class II enzymes (*M. thermophila* and the two *M. frisia* sequences), but no defined groupings of any of other sequences (Figure 7 and Table V). The greater pairwise divergences among ε sequences, and the less

TABLE IV

ACS δ subunit pairwise distances estimated by maximum likelihood under the JTT model

A. difficile (A)	–	1.03	1.53	1.41	1.62	1.64
C. thermoauceticum (B)	1.03	–	1.28	1.24	1.47	1.43
A. fulgidus (C)	1.53	1.28	–	0.71	0.96	0.94
M. thermophila (D)	1.41	1.24	0.71	–	0.97	0.93
M. jannaschii (E)	1.62	1.47	0.96	0.97	–	0.74
M. thermoautotrophicum (F)	1.64	1.43	0.94	0.93	0.74	–

TABLE V

ACS ε subunit pairwise distances estimated by maximum likelihood under the JTT model

A. fulgidus (A)	–	1.07	1.30	1.48	1.45	1.73	1.78	1.49
A. fulgidus (B)	1.07	–	1.36	1.11	1.48	1.24	2.05	1.51
M. frisia IA (C)	1.30	1.36	–	1.34	0.08	1.78	1.94	0.28
M. soehngenii (D)	1.48	1.11	1.34	–	1.45	1.61	2.51	1.50
M. thermophila (E)	1.45	1.48	0.08	1.45	–	1.74	1.96	0.29
M. thermoautotrophicum (F)	1.73	1.24	1.78	1.61	1.74	–	1.90	1.78
M. jannaschii (G)	1.78	2.05	1.94	2.51	1.06	1.90	–	1.97
M. frisia IIA (H)	1.49	1.51	0.28	1.50	0.29	1.78	1.97	–

conservative nature of its phylogeny may be due to the absence of metal centers in the ε subunit that could have constrained protein structure and function.

3.4. Distribution of ACS genes in archaea and bacteria

M. thermoautotrophicum contains genes encoding one Class I enzyme, including α, β, γ, and ε subunits, all of which are located in the same region of the genome (Smith *et al.*, 1997). *M. jannaschii* contains genes encoding 2α, 2β, 1γ, 2 δ and 1ε, subunits (Bult *et al.*, 1996). The known metabolism of this organism as well as the presence of an ε subunit indicates that some subunits are organized into a Class I enzyme. The γ and δ genes are located adjacent to each other, about 40 genes before the 2β genes, 1 of the α genes, and the ε gene. This α ('B' in Figures 2 and 3) and one of the β sequences within the gene cluster ('D' in Figure 4) are of the Class I variety, and are probably associated with the γ, δ, and ε genes found in the same region. The remaining α and β genes, 'P' in Figure 3 and 'C' in Figure 4, respectively, are both Class III-like. If the product of these sequences form a Class III enzyme, the γ and δ subunits would need to associate

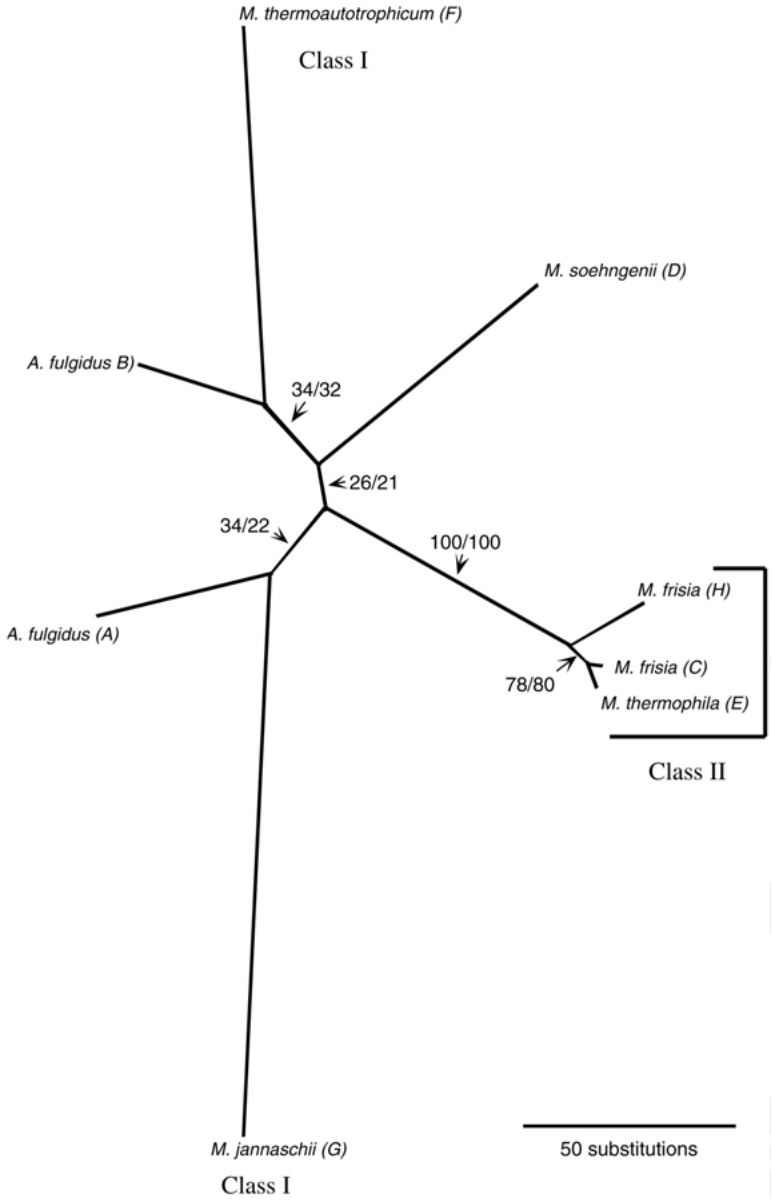

Figure 7. Phylogeny of ACS ε-Subunit Amino Acid Sequences, based on a dataset of 8 aligned sequences, and 188 sites. Letters next to sequences correspond to those in the alignment, available at http://sakmarsgi.rockefeller.edu/~changb/Public/ACS. This tree represents the maximum likelihood topology found using the star decomposition method under the JTT model (Table V). Bootstrap percentage values are indicated above each node: maximum parsimony/neighbor-joining (for 100, and 1000 replications, respectively). The numbers of substitutions along each branch are proportional to the branch length, as indicated by the legend. These inferred branch lengths were estimated by maximum likelihood under the JTT model. GenBank accession numbers include *A. fulgidus* (AAB90135), *A. fulgidus* (AAB91265), *M. frisia* (Q49162), *M. soehngenii* (P26693), *M. thermophila* (AAC44651), *M. thermoautotrophicum* (AAB86181), *M. jannaschii* (AAB98136), *M. frisia* (Q49164).

with both Class I (as part of a complex) and Class III enzymes (as an autonomous protein), since there is only one copy of each in the genome. The presence of a Class III-like enzyme in *M. jannaschii* is surprising, as this organism does not exhibit the metabolism of Class III-containing acetogens (Selkov, 1997). Thus, the Class III-like enzyme may have a unique activity.

Conclusions regarding *M. kandleri* are less certain because its genome has not been sequenced. Only a single α subunit sequence has been reported (Shin *et al.*, 1999) and our analysis indicates that it is of the Class III/IV-like variety. Given the metabolism of this organism, *M. kandleri* probably contains one Class I enzyme and one Class III-like enzyme, similar to the constitution of *M. jannaschii*.

A. fulgidus contains genes for 3α, 1β, 1γ, 1δ, and 2ε subunits (Klenk *et al.*, 1997). Genes for the β, γ and δ subunits are located in a group. An α gene (Figures 2 and 3, sequence H) and ε gene are located in another group $\sim$700 genes from the $\beta\gamma\delta$ group. The products of these two groups of genes probably function together as a Class I or II enzyme. The second α gene (Figures 2 and 3, sequence C) and the remaining ε gene are located together $\sim$1300 genes away from the first $\alpha\varepsilon$ group, suggesting that their products function together as another Class I or II enzyme. The products of one $\alpha\varepsilon$ gene-pair may be used for the Class I enzyme while those of the other may be part of the Class II enzyme. The same $\beta\gamma\delta$ complex must associate with both $\alpha\varepsilon$ complexes, as the organism contains only one sequence of each of these genes. The last α gene (Figures 2 and 3, sequence M) is located in a region lacking any other ACS sequence, and our analysis indicates that it is of the Class III/IV-like variety. If this α subunit were part of a Class III-like enzyme, it would use the same $\beta\gamma\delta$ subunits as used by the Class I and II enzymes. This would be novel in that other Class III α subunits function with β subunits that have an additional 300-residue region at their N-terminus, while the β subunit in *A. fulgidus* lacks this region. Alternatively, the isolated α subunit may function as a Class IV-like monomer, but this organism is not known to be carboxydotrophic. In summary, *A. fulgidus* contains three ACS enzymes, including a Class I, Class II and either a Class III-like or Class IV-like enzyme.

M. thermophila contains genes (1α, 1β, 1γ, 1δ, and 1ε) encoding one Class II enzyme (Maupin-Furlow and Ferry, 1996). However, its genome has not been sequenced, and additional ACS-related genes may be present. *M. frisia* contains two sets of $\alpha\varepsilon$ genes, reminiscent of the situation in *A. fulgidus* (Eggen *et al.*, 1996). Given the similar metabolism of *M. thermophila* and *M. frisia*, and their close phyologenetic relationship to *A. fulgidus* (Woese, 1991), it seems likely that each organism may contains 2 $\alpha\varepsilon$ groups and a single group of $\beta\gamma\delta$ genes. *M. soehngenii* has been reported to contain one set of $\alpha\varepsilon$ genes (Eggen *et al.*, 1991). Given the metabolism of this organism, the products of these genes are probably part of a Class II enzyme.

C. thermoacetium contains 1α, 1β, 1γ, and 1δ genes, which encodes the best-studied example of a Class III enzyme (Roberts *et al.*, 1989; Morton *et al.*, 1991; Lu *et al.*, 1993). A gene encoding an ε subunit has not been identified, and no

such subunit is required for catalytic activity. Whether *C. thermoacetium* contains other ACS-related genes is unknown, as the genome of this organism has not been sequenced.

The genome of the closely related organism, *C. difficile*, has been sequenced (Sanger Center). It contains a similar assortment of ACS genes, including 2α, 1β, 1γ, and 1δ sequences. *C. difficile* is an obligate anaerobe that grows in human intestines and causes antibiotic-associated colitis, antibiotic-related diarrhea, and nosocomial infections (Sednaoui *et al.*, 1999; Freeman and Wilcox, 1999; Taege and Adal, 1999; Wilcox and Modi, 1999). To the best of our knowledge, the role of ACS in this organism has not been investigated. However, our analysis suggests that one $\alpha\beta\gamma\delta$ set probably encode a Class III enzyme, similar to that in *C. thermoaceticum*, and that the remaining α subunit is of the Class III-like or Class IV-like variety.

The genome for *C. acetobutylicum* strain ATCC 824 has also been sequenced, and it contains 2α genes (Genome Therapeutics Corp., Waltham, MA). Both α subunits are of the Class III-like or Class IV-like variety. *C. acetobutylicum* is an obligate anaerobe that produces large quantities of acids (acetate and butyrate) and solvents (acetone, butanol and ethanol) (Peguin, 1994). The role of ACS in this organism has not been investigated, and the presence of these genes was unexpected, in that cell extracts of this organism lacked CO oxidation activity (Kim *et al.*, 1984). This activity is (or had been, up until now) a property of all characterized ACS-related enzymes, and it argues against a Class IV function for the α subunits. It may be significant that the proportion of acids vs. solvents produced by *C. acetobutylicum* depends on the presence or absence of CO (CO increases solventogenesis and decreases acidogenesis). The mechanism of this CO-dependent effect is largely unknown, though it was attributed to the ability of CO to inhibit hydrogenase activity. The ACS-related α subunits may be involved in this process, but possibly not as enzymes that oxidize CO to CO_2.

The carboxydotrophic bacterium *R. rubrum* contains one α gene, which encodes the best-characterized Class IV enzyme (Kerby *et al.*, 1992). Its genomic sequence has not been determined, so the presence of additional ACS-related genes is uncertain.

3.5. Model for the evolution of ACS

Constructing the evolutionary history of an enzyme from phylogenetic analyses is difficult and inevitably fraught with uncertainties, especially when phylogenetic trees are unrooted. In such cases, evolutionary models, including the one presented below, should be viewed as nothing more than constructions that make analyses comprehensible and provide interesting hypotheses for future testing.

The simplest model that is compatible with the subunit phylogenies assumes that all classes of ACS evolved from a common ancestor consisting of α, β, γ, δ and ε subunits having characteristics of extant Class I enzymes. The ancestral α

subunit had residues that coordinated the B- and C-clusters, two unnamed Fe_4S_4 clusters, and an unidentified His_3Cys_2 functionality. The ancestral β subunit contained residues that coordinated the A-cluster, while the ancestral γ and δ subunits had residues that bound an Fe_4S_4 cluster and a cobalamin, respectively.

The most dramatic divergence to occur resulted in a new enzyme form, called Class III/IV-like. The α subunit of this form lacked the two unnamed Fe_4S_4 clusters, and the β subunit contained an extra ~300 amino acid residues at its N-terminus. The quaternary structure of the enzyme was $\alpha_2\beta_2$ with an autonomous $\gamma\delta$ dimer.

Other events were less dramatic. The Class I/II enzyme diverged into the Class I and II enzymes found in extant obligate and facultative autotrophic archaea, respectively, but this did not require any major changes. The Class III/IV-like enzyme diverged into Class III-like and Class III/IV enzymes, and was associated with the loss or modification of the His_3Cys_2 functionality in the Class III-like α subunit. The Class III-like enzyme diverged into an $\alpha_2\beta_2 + \gamma\delta$ form (as in *M. jannaschii* and possibly *M. kandleri*) and two monomeric forms (as in solventogenic/acidogenic Clostridia and possibly *A. fulgidus*). The Class III/IV enzyme diverged into two group, including an $\alpha_2\beta_2 + \gamma\delta$ Class III enzyme (as in *C. thermoaceticum* and *C. difficile*) and an α monomeric Class IV enzyme (as in *R. rubrum*).

3.6. Model for the evolution of ACS-containing organisms

To discuss how ACS-containing organisms may have evolved, something about the composition and level of nutrients available in the environment must be assumed. There are two extreme views. The autotrophic-to-heterotrophic ($a \rightarrow h$) view assumes that the prebiotic environment was devoid of nutrients except for simple inorganic molecules such as CO_2 and H_2. Increasingly energy-rich nutrients, including acetate, pyruvate, lactate, sulfate, glucose, CO and O_2, would have appeared gradually, either as waste products from living organisms or as decomposition products of dead organisms. In contrast, the heterotrophic-to-autotrophic ($h \rightarrow a$) view assumes that the prebiotic environment was abundant in energy-rich nutrients. Then, in certain environments, these nutrients gradually depleted as bio-consumption surpassed abiotic production.

Asuming the $a \rightarrow h$ view, the earliest organisms would have been obligate autotrophs with properties similar to the Methanobacteriales, Methanococcales, and Methanopyrales. They would have used a Class I ACS and the methanogenesis pathway to grow on CO_2/H_2 (Wood, 1991; Koch, and Schmidt, 1991). It is simplest to assume that these organisms contained a single Class I ACS enzyme, similar to *M. thermoautotrophicum*.

The first major event in the evolution of the original ACS was also the most dramatic and least understood. These organisms duplicated their α and β genes and the copies evolved into a Class III-like enzyme. The $\gamma\delta$ gene-products used to 'service' the Class I enzyme was probably also used by the Class III-like enzyme. The closest descendents of the resulting organisms would be *M. jannaschii* and

M. kandleri. However, given that the metabolism of these organisms and *M. thermoautotrophicum* are so similar, the function for the Class III-like enzyme remains obscure.

Once acetate became available, some ACS-containing organisms evolved into facultative autotrophs with properties of Methanosarcina, including *M. thermophila*, *M. frisia* and *M. soehngenii*. This process would have included a gene-duplicated Class I ACS evolving into a Class II enzyme. These events would have allowed more flexible growth on either acetate or CO_2/H_2.

When sulfate became available (Wagner *et al*., 1998), populations of facultative autotrophs evolved to use this ion as a terminal electron acceptor. When lactate became available, some sulfate-reducers used their Class II ACS's in the process of catabolizing lactate to $3CO_2$. However, they maintained Class I enzymes and thus the flexibility to grow autotrophically. Other ACS-containing organisms employed nitrate in a similar capacity. As sulfate and nitrate replaced CoM-SS-CoB as electron acceptors, the terminal portion of the methanogenesis pathway became irrelevant. The closest descendents of these organisms would be the Archaeoglobales. Low G + G Gram-positive sulfate-reducing bacteria (*Desulfotomaculum*) have properties suggesting a close relationship to the Archaeoglobales. However these bacteria used THF rather than H_4MPT as a methyl transfer coenzyme.

Some low G + C Gram-positive ACS-containing bacteria were not sulfate or nitrate reducers. When glucose became available, some of these organisms evolved to operate gluconeogenesis in reverse and catabolize glucose *anaerobically* (Fothergill-Gilmore and Michels, 1993). Glucose was oxidized first to two pyruvates, $4e^-$ and $2CO_2$'s. Then the two pyruvates were oxidized further (by PFOR) to 2 acetyl-CoA, 2 CO_2, and $4e^-$. The top part of the remnant methanogenesis pathway used 6 of the 8 electrons to reduce CO_2 to the methyl group of CH_3-THF. In a reaction catalyzed by a Class III ACS, the remaining 2 electrons were used to reduce CO_2 to CO. Finally, the Class III enzyme combined CO with the methyl group, forming a third molecule of acetyl-CoA. Some of the acetyl-CoA hydrolyzed to acetate as a means of harnessing free-energy, while the remainder was used in anabolism. At this point, Class I and II enzymes became superfluous, and variants lacking genes encoding them flourished. The closest living relatives of these organisms would be the acetogens, including *C. thermoaceticum*, as well as *C. difficile*.

Other variants developed the ability to catabolize glucose via a dismutation process in which some glucose molecules were oxidized to CO_2 while others were reduced to organic acids and solvents. These organisms no longer required Class III ACS, and so variants lacking $\beta\gamma\delta$ genes were selected. The gene for the α subunit was retained, however, and used for other (unknown) purposes. The closest living relatives of these organisms would be solventogenic clostridia such as *C. acetobutylicum*.

Gram-negative organisms are thought to have evolved as a response to environments containing antibiotics (Gupta, 1998). Relatives of Gram-negative sulfate-reducing ACS-containing variants would be the δ-proteobacteria (including Des-

ulfobacterium, Desulfococcus, Desulfovibrio, and Desulfosarcina). Upon exposure to environments containing CO, some Gram-negative organisms evolved the ability to grow on this molecule. They coupled the oxidation of CO to CO_2, using the α subunit (Class IV ACS), to the reduction of protons via a membrane-bound hydrogenase, and harnessed the free energy released through oxidative phosphorylation. Whether this preceded or followed the evolution of photosynthesis in these organisms is unknown. In any event, the closest relatives of the Gram-negative carboxydotrophic photosynthesizing organisms would be the α-proteobacteria *R. rubrum*.

With the proliferation of photosynthesis, populations of ACS-containing organisms were exposed to increasing levels of O_2. Variants able to utilize O_2 as their terminal electron acceptor flourished. By catabolizing sugars *aerobically*, these heterotrophs extracted so much energy from their environment that their survival was no longer constrained in this manner, and they began to evolve in new and more varied directions. ACS was irreversibly damaged by O_2 and the reactions it had catalyzed became superfluous. As a consequence, genes encoding ACS were abandoned by organisms whose closest living relatives are aerobic bacteria.

According to this model, the survival and prosperity of early organisms was severely limited by their ability to extract energy from their environments, due to an inadequate supply of nutrients. Early organisms used ACS as a tool to extract increasing amounts of energy from their environment as new energy-rich molecules appeared. Thus, the enzyme facilitated the conversion from autotrophic to heterotrophic metabolism. Once populations harnessed energy from the aerobic respiration of sugars, ACS was abandoned. The abundant energy obtained by aerobic respiration allowed organisms to proliferate and evolve into more advanced life forms. ACS was retained only by organisms that adhered to their original and increasingly obscure (anaerobic, hot, and nutritionally impoverished) environments.

Assuming the h $\rightarrow$ a perspective suggests the reverse role of ACS, namely as a tool for allowing organisms to survive in increasingly inhospitable and nutritionally impoverished environments. We prefer the a $\rightarrow$ h assumption, as it is simpler, but our analysis is strictly neutral on this issue. For this reason, the phylogenetic trees presented are unrooted. However, in either case, the evolution of ACS appears to have allowed organisms to optimize the extraction of energy from environments that were changing in terms of nutrient composition and level.

Acknowledgments

We would like to thank Professor Rolf K. Thauer for critically reading the manuscript prior to publication. This work was supported initially by an NSF/Sloan Postdoctoral Fellowship in Molecular Evolution, followed by an NSF Internal Fel-

lowship (both to B.S.W.C.), as well as the National Institutes of Health (GM46441, to P.A.L.).

Note Added in Proof

Another α sequence has been recently reported (González, J. M. and Robb, F. T.: 2000, 'Genetic analysis of *Carboxydothermus hydrogenoformans* carbon monoxide dehydrogenase genes cooF and coos', *FEMS Microbiol. Lett.* **191**, 243–247). With the exception of His 231, it has the same conserved His and Cys residues as sequence J in Figure 2.

References

Achenbach-Richter, L., Gupta, R., Stetter, K. O. and Woese, C. R.: 1987, 'Were the original eubacteria thermophiles?', *Syst. Appl. Microbiol.* **9**, 34–39.

Adachi, J. and Hasegawa, M.: 1994, 'MOLPHY', Ver. Institute of Statistical Mechanics, Tokyo, Japan.

Balch, W. E., Schoberth, S., Tanner, R. S. and Wolfe, R. S.: 1977, '*Acetobacterium*, a new genus of hydrogen-oxidizing, carbon dioxide-reducing, anaerobic bacteria', *Int. J. Syst. Bacteriol.* **27**, 355–361.

Barondeau, D. P. and Lindahl, P. A.: 1997, 'Methylation of carbon monoxide dehydrogenase from *Clostridium thermoaceticum* and the mechanism of acetyl-CoA synthesis', *J. Am. Chem. Soc.* **119**, 3959–3970.

Bhatnagar, L., Jain, M. K. and Zeikus, J. G.: 1991, 'Methanogenic Bacteria', in J. M. Shively and L. L. Barton (eds), *Variation in Autotrophic Life*, Academic Press, N.Y. Chapter 9, pp. 251–270.

Bonam, D. and Ludden, P. W.: 1987, 'Purification and characterization of CO dehydrogenase, a nickel, zinc, iron-sulfur protein, from *Rhodospirillum rubrum*', *J. Biol. Chem.* **262**, 2980–2987.

Brown, J. R. and Doolittle, W. F.: 1997, 'Archaea and the prokaryote-to-eukaryote transition' *Microbiol. Rev.* **61**, 456–502.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J.-F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., Nguyen, D., Utterback, T. R., Kelley, J. M., Peterson, J. D., Sadow, P. W., Hanna, M. C., Cotton, M. D., Roberts, K. M., Hurst, M. A., Kaine, B. P., Borodovsky, M., Klenk, H.-P., Fraser, C. M., Smith, H. O., Woese, C. R. and Venter, J. C.: 1996, 'Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*', *Science* **273**, 1058–1073.

Burggraf, S., Stetter, K. O., Rouviere, P. and Woese, C. R.: 1991, '*Methanopyrus kandleri*: An archaeal methanogen unrelated to all other known methanogens', *Syst. Appl. Microbiol.* **14**, 346–351.

Campbell, J. H.: 1991, 'An RNA replisome as the ancestor of the ribosome', *J. Molec. Evol.* **32**, 3–5.

Dai, Y.-R., Reed, D. W., Millstein, J. H., Hartzell, P. L., Grahame, D. A. and DeMoll, E.: 1998, 'Acetyl-CoA decarbonylase/synthase complex from *Archaeoglobus fulgidus*', *Arch. Microbiol.* **169**, 525–529.

Davis, B. K.: 1999, 'Evolution of the genetic code', *Prog. Biophys. Mol. Biol.* **72**, 157–243.

DeRose, V. J., Anderson, M. E., Lindahl, P. A. and Hoffman, B. M.: 1998, 'ENDOR of the C-cluster of CO dehydrogenase from *Clostridium thermoaceticum*: Evidence for histidine, cysteine, and hydroxyl coordination to the Fe_4S_4 moiety', *J. Am. Chem. Soc.* **120**, 8767–8776.

Diekert, G. B. and Thauer, R. K.: 1978, 'Carbon monoxide oxidation by *Clostridium thermoaceticum* and *Clostridium formicoaceticum*', *J. Bacteriol.* **136**, 597–606.

Diekert, G. B., Graf, E. G. and Thauer, R. K.: 1979, 'Nickel requirement for carbon monoxide dehydrogenase formation in *Clostridium pasteurianum*', *Arch. Microbiol.* **122**, 117–120.

Drake, H. L.: 1994, *Acetogenesis*, Chapman and Hall, New York.

Edwards, M. R.: 1998, 'From a soup or a seed? Pyritic metabolic complexes in the origin of life', *Trends in Ecology and Evolution* **13**, 178–181.

Eggen, R. I. L., Gerrling, A. C. M., Jetten, M. S. M. and de Vos, W. M.: 1991, 'Cloning, expression, and sequence analysis of the genes for carbon monoxide dehydrogenase of *Methanothrix soehngenii*', *J. Biol. Chem.* **266**, 6883–6887.

Eggen, R. I. L., van Kranenburg, R., Vriesema, A. J. M., Gerrling, A. C. M., Verhagen, M. F. J. M., Hagen, W. R. and de Vos, W. M.: 1996, 'Carbon monoxide dehydrogenase from *Methanosarcina frisia* Göl', *J. Biol. Chem.* **271**, 14256–14263.

Eigen, M. and Schuster, P.: 1977, 'The hypercycle, a principle of natural self-organization: Part A; emergence of a hypercycle', *Die Naturwissenschaften* **64**, 541–565.

Eigen, M. and Schuster, P.: 1978, 'The hypercycle, a principle of natural self-organization: Part B; The abstract hypercycle', *Die Naturwissenschaften* **65**, 7–41.

Eigen, M. and Schuster, P.: 1978, 'The hypercycle, a principle of natural self-organization: Part C; The realistic hypercycle', *Die Naturwissenschaften* **65**, 341–369.

Farmer, P. J., Reibenspies, J. H., Lindahl, P. A. and Darensbourg, M. Y.: 1993, 'Effects of sulfur site modification on the redox potentials of derivatives of [N,N′-Bis(2-mercaptoethyl)-1,5,-diazacyclooctanato]nickel(II)' *J. Am. Chem. Soc.* **115**, 4665–4674.

Felsenstein, J.: 1981, 'Evolutionary trees from DNA sequences: A maximum likelihood approach', *J. Mol. Evol.* **17**, 368–376.

Felsenstein, J.: 1985, 'Confidence limits on phylogenies: An approach using the bootstrap', *Evol.* **39**, 783–791.

Ferry, J. G.: 1995, 'CO Dehydrogenase', *Ann. Rev. Microbiol.* **49**, 305–333.

Ferry, J. G.: 1999, 'Enzymology of one-carbon metabolism in methanogenic pathways', *FEMS Microbiol. Rev.* **23**, 13–38.

Fothergill-Gilmore, L. A. and Michels, P. A. M.: 1993, 'Evolution of glycolysis', *Progress in Biophysics and Molecular Biology* **59**, 105–236.

Freeman, J. and Wilcox, M. H.: 1999, 'Antibiotics and *Clostridium difficile*', *Microbes and Infection* **1**, 377–384.

Gilbert, W.: 1986, 'Origin of Life – The RNA World', *Nature* **319**, 618–618.

Gogarten, J. P., Hilario, E. and Olendzenski, L.: 1996, 'Gene Duplication and Horizontal Gene Transfer During Early Evolution', in D. McL. Roberts, P. Sharp, G. Alderson and M. A. Collins (eds), *Evolution of Microbial Life*, Cambridge University Press, Cambridge, U.K., pp. 267–292.

Grahame, D. A. and DeMoll, E.: 1996, 'Partial reactions catalyzed by protein components of the acetyl-CoA decarbonylase synthase enzyme complex from *Methanosarcina barkeri*', *J. Biol. Chem.* **271**, 8352–8358.

Gupta, R. S.: 1998, 'Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among Archaebacteria, Eubacteria, and Eukaryotes', *Microbiol. and Mol. Biol. Rev.* **62**, 1435–1491.

Henikoff, S. and Henikoff, J. G.: 1992, 'Amino acid substitution matrices from protein blocks', *Proc. Natl. Acad. Sci. U.S.A.* **39**, 10915–10919.

Hu, Z., Spangler, N. J., Anderson, M. E., Xia, J. Q., Ludden, P. W., Lindahl, P. A. and Münck, E.: 1996, 'Nature of the C-cluster in Ni-containing carbon monoxide dehydrogenases', *J. Am. Chem. Soc.* **118**, 830–845.

Huber, C. and Wächtershäuser, G.: 1997, 'Activated acetic acid by carbon fixation on (Fe, Ni) S under primordial conditions', *Science* **276**, 245–247.

Jablonski, P. E., Lu, W. P., Ragsdale, S. W. and Ferry, J. G.: 1993, 'Characterization of the metal centers in the corrinoid/iron-sulfur component of the CO dehydrogenase enzyme complex from *Methanosarcina thermophila* by EPR and spectroelectrochemistry', *J. Biol. Chem.* **268**, 325–329.

Jones, W. J., Leigh, J. A., Mayer, F., Woese, C. R. and Wolfe, R. S.: 1983, '*Methanococcus jannaschii* sp. Nov., an extremely thermophilic methanogen from a submarine hydrothermal vent', *Arch. Microbiol.* **136**, 254–261.

Jones, D. T., Taylor, W. R. and Thornton, J. M.: 1992, 'The rapid generation of mutation data matrices from protein sequences', *Comp. Appl. Biosci.* **8**, 275–282.

Joyce, G. F.: 1989, 'RNA evolution and the origins of life', *Nature* **338**, 217–224.

Kauffman, S. A.: 1986, 'Autocatalytic sets of proteins', *J. Theor. Biol.* **119**, 1–24.

Kauffman, S. A.: 1993, *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, N.Y.

Kerby, R. L., Hong, S. S., Ensign, S. A., Coppoc, L. J., Ludden, P. W. and Roberts, G. P.: 1992, 'Genetic and physiological characterization of the *Rhodospirillum rubrum* carbon monoxide dehydrogenase system', *J. Bacteriol.* **174**, 5284–5294.

Kerby, R. L., Ludden, P. W. and Roberts, G. P.: 1995, 'Carbon monoxide dependent growth of *Rhodospirillum rubrum*' *J. Bacteriol.* **177**, 2241–2244.

Kim, B. H., Bellows, P., Datta, R. and Zeikus, J. G.: 1984, 'Control of carbon and electron flow in *Clostridium acetobutylicum* fermenation: Utilization of carbon monoxide to inhibit hydrogen production and to enhance butanol yields', *Appl. Environ. Microbiol.* **48**, 764–770.

Kishino, H., Miyata, T. and Hasegawa, M.: 1990, 'Maximum likelihood inference of protein phylogeny and the origin of chloroplasts', *J. Mol. Evol.* **31**, 151–160.

Klemps, R., Cypionka, H., Widdel, F. and Pfennig, N.: 1985, 'Growth with hydrogen, and further physiological characteristics of Desulfotomaculum species', *Arch. Microbiol.* **143**, 203–208.

Klenk, H.-P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Graham, D. E., Kyrpides, N. C., Fleischmann, R. D., Quackenbush, J., Lee, N. H., Sutton, G. G., Gill, S., Kirkness, E. F., Dougherty, B. A., McKenney, K., Adams, M. D., Loftus, B., Peterson, S., Reich, C. I., McNeil, L. K., Badger, J. H., Glodek, A., Zhou, L. X., Overbeek, R., Gocayne, J. D., Weidman, J. F., McDonald, L., Utterback, T., Cotton, M. D., Spriggs, T., Artiach, P., Kaine, B. P., Sykes, S. M., Sadow, P. W., DAndrea, K. P., Bowman, C., Fujii, C., Garland, S. A., Mason, T. M., Olsen, G. J., Fraser, C. M., Smith, H. O., Woese, C. R. and Venter, J. C.: 1997, 'The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*', *Nature* **390**, 364–370.

Koch, A. L. and Schmidt, T. N.: 1991, 'The first cellular bioenergetic process; primitive generation of a proton-motive force', *J. Mol. Evol.* **33**, 297–304.

Krzycki, J. A., Wolkin, R. H. and Zeikus, J. G.: 1982, 'Comparison of unitrophic and mixotrophic substrate metabolism by an acetate-adapted strain of *Methanosarcina barkeri*', *J. Bacteriol.* **149**, 247–254.

Kurr, M., Huber, R., König, H., Jannasch, H. W., Fricke, H., Trincone, A., Kristjansson, J. K. and Stetter, K. O.: 1991, '*Methanopyrus kandleri*, gen. and sp. Nov. represents a novel group of hyperthermophilic methanogens, growing at 110 °C', *Arch. Microbiol.* **156**, 239–247.

Länge, S., Scholtz, R. and Fuchs, G.: 1989, 'Oxidative and reductive acetyl CoA/carbon monoxide dehydrogenase pathway in *Desulfobacterium autotrophicum*', *Arch. Microbiol.* **151**, 77–83.

Lu, W. P., Schiau, I., Cunningham, J. R. and Ragsdale, S. W.: 1993, 'Sequence and expression of the gene encoding the corrinoid/iron-sulfur protein from *Clostridium thermoaceticum* and reconstitution of the recombinant protein to full activity', *J. Biol. Chem.* **268**, 5605–5614.

Lu, W. P., Jablonski, P. E., Rasche, M., Ferry, J. G. and Ragsdale, S. W.: 1994, 'Characterization of the metal centers of the Ni/Fe–S component of the carbon-monoxide dehydrogenase enzyme complex from *Methanosarcina thermophila*', *J. Biol. Chem.* **269**, 9736–9742.

Maden, B. E. H.: 1995, 'No soup for starters? Autotrophy and the origins of metabolism', *Trends in Biochem. Sci.* **20**, 337–341.

Maupin-Furlow, J. A. and Ferry, J. G.: 1996, 'Characterization of the cdhD and cdhE genes encoding subunits of the corrinoid/iron-sulfur enzyme of the CO dehydrogenase complex from *Methanosarcina thermophila*', *J. Bacteriol.* **178**, 340–346.

Menon, S. and Ragsdale, S. W.: 1998, 'Role of the [4Fe-4S] cluster in reductive activation of the cobalt center of the corrinoid iron-sulfur protein from *Clostridium thermoaceticum* during acetate biosynthesis', *Biochemistry* **37**, 5689–5698.

Meyer, O. and Rhode, M.: 1984, 'Microbial Growth on C1 Compounds', in R. L. Crawford and R. S. Hanson (eds), *Proc. 4th Int. Symp. Am. Soc. Microbiology*, Washington DC, pp. 26–33.

Meyer, O. and Fiebig, K.: 1985, in H. Degn, R. P. Cox and H. Toftlund (eds), *Gas Enzymology*, D. Reidel, Dordrecht, pp. 147–168.

Mojzsis, S. J., Arrhenius, G., McKeegan, K. D., Harrison, T. M., Nutman, A. P. and Friend, C. R. L.: 1996, 'Evidence for life on earth before 3800 million years ago', *Nature* **384**, 55–59.

Möller-Zinkhan and Thauer, R. K.: 1990, 'Anaerobic lactate oxidation to $3CO_2$ by *Archaeoglobus fulgidus* via the CODH pathway: Demonstration of the acetyl-CoA carbon-carbon cleavage reaction in cell extracts', *Arch. Microbiol.* **153**, 215–218.

Morton, T. A., Runquist, J. A., Ragsdale, S. W., Shanumgasundaram, T., Wood, H. G. and Ljungdahl, L. G.: 1991, 'The primary structure of the subunits of carbon monoxide dehydrogenase/acetyl-CoA synthase from *Clostridium thermoaceticum*', *J. Biol. Chem.* **266**, 23824–23828.

Nicolet, Y., Piras, C., Legrand, P., Hatchikian, C. E. and Fontecilla-Camps, J. C.: 1999, '*Desulfovibrio desulfuricans* iron hydrogenase: The structure shows unusual coordination to an active site Fe binuclear center', *Structure with Folding and Design* **7**, 13–23.

Olsen, G. J., Woese, C. R. and Overbeek, R.: 1994, 'The winds of (evolutionary) change: Breathing new life into microbiology', *J. Bacteriol.* **176**, 1–6.

Peguin, S., Goma, G., Delorme, P. and Soucaille, P.: 1994, 'Metabolic flexibility of *Clostridium acetobutylicum* in response to methyl viologen addition', *Appl. Microbiol. Biotechnol.* **42**, 611–616.

Peters, J. W., Lanzilotta, W. N., Lemon, B. J. and Seefeldt, L. C.: 1998, 'X-ray crystal structure of the Fe-only hydrogenase (Cpl) from *Clostridium pasteurianum* to 1.8 angstrom resolution', *Science* **282**, 1853–1858.

Ragsdale, S. W., Ljungdahl, L. G. and DerVartanian, D. V.: 1983, 'Isolation of carbon monoxide dehydrogenase from *Acetobacterium woodii* and comparison of its properties with those of the *Clostridium thermoaceticum* enzyme', *J. Bacteriol.* **155**, 1224–1237.

Ragsdale, S. W., Lindahl, P. A. and Münck, E.: 1987, 'Mössbauer, EPR, and optical studies of the corrinoid/iron-sulfur protein involved in the synthesis of acetyl coenzyme A by *Clostridium thermoaceticum*', *J. Biol. Chem.* **262**, 14289–14297.

Ragsdale, S. W. and Kumar, M.: 1996, 'Nickel-containing carbon monoxide dehydrogenase/acetyl-CoA synthase', *Chem. Rev.* **96**, 2515–2539.

Roberts, D. L., James-Hagstrom, J. E., Garvin, D. K., Gorst, C. M., Runquist, J. A., Baur, J. R., Haase, F. C. and Ragsdale, S. W.: 1989, 'Cloning and expression of the gene cluster encoding key proteins involved in acetyl-CoA synthesis in *Clostridium thermoaceticum*: CO dehydrogenase, the corrinoid/Fe–S protein, and methyltransferase', *Proc. Natl. Acad. Sci. U.S.A.* **86**, 32–36.

Russell, W. K., Stålhandske, C. M. V., Xia, J., Scott, R. A. and Lindahl, P. A.: 1998, 'Spectroscopic, redox and structural characterization of the Ni-labile and nonlabile forms of the acetyl-CoA synthase active site of carbon monoxide dehydrogenase', *J. Am. Chem. Soc.* **120**, 7502–7510.

Rzhetsky, A. and Nei, M.: 1992, 'A simple method for estimating and testing minimum-evolution trees', *Mol. Biol. Evol.* **9**, 945–967.

Saitou, N. and Nei, M.: 1987, 'The neighbor-joining method: A new method for reconstructing phylogenetic trees', *Mol. Biol. Evol.* **4**, 406–425.

Schauder, R., Eikmanns, B., Thauer, R. K., Widdel, F. and Fuchs, G.: 1986, 'Acetate oxidation to CO_2 in anaerobic bacteria via a novel pathway not involving reactions of the citric acid cycle', *Arch. Microbiol.* **145**, 162–172.

Schauder, R., Preuss, A., Jetter, M. and Fuchs, G.: 1989, 'Oxidative and reductive acetyl-CoA carbon monoxide dehydrogenase pathway in *Desulfobacterium Autotrophicum*. 2. Demonstration of the enzymes of the pathway and comparison of CO dehydrogenase', *Arch. Microbiol.* **151**, 84–89.

Schopf, J. W.: 1996, 'Are the Oldest Fossils Cyanobacteria?' in D. McL. Roberts, P. Sharp, G. Alderson and M. A. Collins (eds), *Evolution of Microbial Life*, Cambridge University Press, Cambridge, U.K., pp. 23–61.

Sednaoui, P. E., Mantih, B. and Cauwell, M.: 1999, ' 'Second look' cytotoxicity assay for the diagnosis of *Clostridium difficile* antibiotic-associated colitis', *Pathol. Biol.* **47**, 415–421.

Selkov, E., Maltsev, N., Olsen, G. J., Overbeek, R. and Whitman, W. B.: 1997, 'A Reconstruction of the metabolism of *Methanococcus jannaschii* from sequence data', *Gene* **197**, GC11–26.

Shin, H.-S., Ryu, J.-R., Han, Y.-S., Choi, Y.-J. and Yu, Y. G.: 1999, 'Random Sequence analysis of the genomic DNA of *Methanopyrus kandleri* and molecular cloning of the gene encoding a homologue of the catalytic subunit of CODH', *J. Microbiol. Biotechnol.* **9**, 404–413.

Smith, D. R., DoucetteStamm, L. A., Deloughery, C., Lee, H. M., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D. Y., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrokovski, S., Church, G. M., Daniels, C. J., Mao, J. I., Rice, P., Nolling, J. and Reeve, J. N.: 1997, 'Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: Functional analysis and comparative genomics', *J. Bacteriol.* **179**, 7135–7155.

Spangler, N. J., Meyers, M. R., Gierke, K. L., Kerby, R. L., Roberts, G. P. and Ludden, P. W.: 1998, 'Substitution of valine for histidine 265 in carbon monoxide dehydrogenase from *Rhodospirillum rubrum* affects activity and spectroscopic states', *J. Biol. Chem.* **273**, 4059–4064.

Staples, C. R., Heo, J., Spangler, N. J., Kerby, R. L., Roberts, G. P. and Ludden, P. W.: 1999, '*Rhodospirillum rubrum* CO-dehydrogenase. Part 1. Spectroscopic studies of CODH variant C531A indicate the presence of a binuclear [FeNi] cluster', *J. Am. Chem. Soc.* **121**, 11034–11044.

Swofford, D. L., Olsen, G. J., Waddell, P. J. and Hillis, D. M.: 1996, 'Phylogenetic Inference', in D. M. Hillis, C. Moritz, and B. K. Mable (eds), *Molecular Systematics*, Sinauer, Sunderland, Massachusetts, pp. 407–514.

Swofford, D. L.: 1999, 'PAUP*, Phylogenetic Analysis Using Parsimony (* and Other Methods)', Ver. 4.0., Sinauer, Sunderland, Massachusetts.

Taege, A. J. and Adal, K. A.: 1999, '*Clostridium difficile* diarrhea and colitis: A clinical overview' *Cleveland Clinic Journal of Medicine* **66**, 503–507.

Tan, G. O., Ensign, S. A., Ciurli, S., Scott, M. J., Hedman, B., Holm, R. H., Ludden, P. W., Korszun, Z. R., Stephens, P. J. and Hodgson, K. O.: 1992, 'On the structure of the nickel/iron/sulfur center of the carbon monoxide dehydrogenase from *Rhodospirillum rubrum*: An X-ray absorption spectroscopy study' *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4427–4431.

Thauer, R. K.: 1998, 'Biochemistry of methanogenesis: A tribute to Marjory Stephenson', *Microbiology* **144**, 2377–2406.

Vorholt, J., Kunow, J., Stetter, K. O. and Thauer, R. K.: 1995, 'Enzymes and coenzymes of the CODH pathway for autotrophic CO_2 fixation in *Archaeoglobus lithotrophicus* and the lack of CODH in the heterotrophic *A. profundus*', *Arch. Microbiol.* **163**, 112–118.

Vorholt, J. A., Hafenbradl, D., Stetter, K. O. and Thauer, R. K.: 1997, 'Pathways of autotrophic CO_2 fixation and dissimilitory nitrate reduction to N_2O in *Ferroglobus placidus*', *Arch. Microbiol.* **167**, 19–23.

Wächtershäuser, G.: 1988, 'Before enzymes and templates: Theory of surface metabolism', *Microbiol. Rev.* **52**, 452–484.

Wächtershäuser, G.: 1990, 'Evolution of the first metabolic cycles', *Proc. Natl. Acad. Sci. U.S.A.* **87**, 200–204.

Wächtershäuser, G.: 1992, 'Groundworks for an evolutionary biochemistry: The iron-sulfur world', *Prog. Biophys. Mol. Biol.* **58**, 85–201.

Wächtershäuser, G.: 1997, 'The origin of life and its methodological challenge', *J. Theor. Biol.* **187**, 483–494.

Wagner, M., Roger, A. J., Flax, J. L., Brusseau, G. A. and Stahl, D. A.: 1998, 'Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration', *J. Bacteriol.* **180**, 2975–2982.

Weimer, P. J. and Zeikus, J. G.: 1978, 'Acetate metabolism in *Methanosarcina barkeri*', *Arch. Microbiol.* **119**, 175–182.

Wilcox, M. H. and Modi, N.: 1999, 'Nosocomial Diarrhoea', *Current Opinion in Infectious Diseases* **12**, 341–345.

Wilson, B. E. and Lindahl, P. A.: 1999, 'Equilibrium dialysis study and mechanistic implications of coenzyme A binding to acetyl-CoA synthase/carbon monoxide dehydrogenase from *Clostridium thermoaceticum*', *J. Biol. Inorg. Chem.* **4**, 742–748.

Woese, C. R.: 1987, 'Bacterial Evolution', *Microbiol. Rev.* **51**, 221–271.

Woese, C. R., Kandler, O. and Wheelis, M. L.: 1990, 'Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya', *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576–4579.

Woese, C. R., Achenbach, L., Rouviere, P. and Mandelco, L.: 1991, 'Archaeal phylogeny: Reexamination of the phylogenetic position of *Archaeoglobus fulgidus* in light of certain composition-induced artifacts', *System. Appl. Microbiol.* **14**, 364–371.

Woese, C.: 1998, 'The Universal Ancestor', *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6854–6859.

Wood, H. G.: 1991, 'Life with CO or CO_2 and H_2 as a source of carbon and energy', *FASEB J.* **5**, 156–163.

Wood, H. G. and Ljungdahl, L. G.: 1991, 'Autotrophic Character of Acetogenic Bacteria', in J. M. Shively and L. L. Barton (eds), *Variation in Autotrophic Life*, Academic Press, N.Y., Chapter 8, pp. 201–250.

Xia, J. and Lindahl, P. A.: 1996, 'Assembly of an exchange-coupled [Ni:Fe_4S_4] cluster in the α metallosubunit of CO dehydrogenase from *Clostridium thermoaceticum* with spectroscopic properties and CO-binding ability mimicking the acetyl-CoA synthase active site', *J. Am. Chem. Soc.* **118**, 483–484.

Xia, J., Sinclair, J. F., Baldwin, T. O. and Lindahl, P. A.: 1996, 'CO Dehydrogenase from *Clostridium thermoaceticum*: Quaternary structure and stoichiometry of its SDS-induced dissociation', *Biochemistry* **35**, 1965–1971.

Xia, J., Hu, Z., Popescu, C., Lindahl, P. A. and Münck, E.: 1997, 'Mössbauer and EPR study of the Ni-activated α subunit of carbon monoxide dehydrogenase from *Clostridium thermoaceticum*', *J. Am. Chem. Soc.* **119**, 8301–8312.

Zeikus, J. G. and Wolfe, R. S.: 1972, '*Methanobacterium thermoautotrophicus* sp., an anaerobic, autotrophic, extreme thermophile', *J. Bacteriol.* **109**, 707–713.