SULFIDE ASSIMILATION IN ANAEROBES:

NOVEL PROTEINS FOR AN ANCIENT STRATEGY

A dissertation presented to the Department of Biochemistry & Molecular Biology and the Oregon Health & Science University School of Medicine in partial fulfillment of the requirements for the degree of Doctor of Philosophy

By

Benjamin Julius Rauch

June 2015

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the PhD dissertation of

Benjamin Julius Rauch

has been approved

John Perona, PhD (Advisor)

Kenneth Stedman, PhD

Buddy Ullman, PhD

Larry David, PhD

Mathew Thayer, PhD

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I wish to acknowledge my advisor, Dr. John Perona, for his unwavering support. When failed experiments left me discouraged, I could rely on John to be sensibly reassuring. I always felt relieved after consulting John.

I thank Jun Kai Zhang and Dr. William W. Metcalf for hosting me in their laboratory at the University of Illinois at Urbana-Champaign. Without those two weeks, the genetic manipulations performed in *Methanosarcina acetivorans,* which were indispensible to my thesis work, would not have been possible.

I am grateful to Drs. Buddy Ullman, Larry David and Ken Stedman for their guidance as members of my research advisory committee, and to the Department of Biochemistry and Molecular Biology of Oregon Health and Science University for allowing me to transfer into their graduate program. I also thank the Department of Chemistry of Portland State University for treating me as one of their own. Both departments were accommodating and generous in their support.

ABSTRACT

The metabolic network for sulfide assimilation and trafficking in methanogens is largely unknown. To discover novel proteins required for these processes, bioinformatic methods were used to identify genes co-occurring with the protein biosynthesis enzyme SepCysS, which converts phosphoseryl-tRNA^{Cys} to cysteinyl-tRNA^{Cys} in nearly all methanogens. The analyses revealed three conserved protein-coding genes, each containing molecular signatures predicting functions in sulfur metabolism. All three genes were also identified in more than 50 strictly anaerobic bacterial genera from nine distinct phyla. Genotype-dependent growth and metabolite labeling experiments conducted in *Methanosarcina acetivorans* demonstrated that two of the proteins (MA1821 and MA1822) are essential to a novel homocysteine biosynthesis reaction, consuming aspartate-4-semialdehyde as a precursor. Mutational analysis confirmed the importance of several structural elements, including a conserved cysteine residue present in MA1821 and a predicted 4Fe-4S cluster-binding domain present in MA1822. Additional genotype-dependent growth experiments determined that the third protein (MA1715) is essential for growth with sulfide when present as the lone sulfur source at concentrations below 800 µM, indicating that MA1715 could be involved in mobilizing sulfur for the biosynthesis of cysteine and homocysteine. Moreover, phylogenetic analyses indicate that all three novel protein families were inherited vertically from the ancestral euryarchaeote along with SepCysS, suggesting that these four proteins comprise an ancient metabolic strategy for the assimilation of sulfide.

INTRODUCTION

The past two decades have seen an exponential increase in the number of completely sequenced microbial genomes (**Fig. 0.1**). These data are being leveraged to identify and answer fundamental biological problems using novel approaches. One such problem took form when it was noticed that the gene encoding cysteinyl-tRNA synthetase (CysRS) is missing from the genomes of some methane-producing *Archaea* (methanogens) [1,2,3]. CysRS catalyzes the ATP-dependent attachment of cysteine to the 3' end of tRNA (**Fig. 0.2**). This yields cysteinyl-tRNA (Cys-tRNA), an essential substrate for ribosome-mediated protein synthesis. Thus, in the absence of CysRS, methanogens had to employ some novel, alternative process.

Eventually, it was revealed that methanogens employ an alternative two-enzyme pathway to synthesize Cys-tRNA [4], wherein tRNA is first charged with phosphoserine (Sep) by phosphoseryl-tRNA synthetase (SepRS) to yield Sep-tRNA (**Fig. 0.2**). Subsequently, Sep-tRNA is converted to Cys-tRNA by Sep-tRNA:Cys-tRNA synthase (SepCysS), requiring a sulfur donor whose source has yet to be identified.

Since both pathways for Cys-tRNA synthesis evolved prior to the time of the last universal common ancestor (LUCA) [5], it is remarkable that the SepRS-SepCysS pathway is present only in methanogens and the closely related *Archaeoglobales*. *Why did these Archaea retain the less efficient, two-enzyme pathway for Cys-tRNA synthesis? And, how might its elusive sulfur source be connected to other aspects of sulfur metabolism?* These were the guiding questions for the

research presented in this dissertation. The answers, it was surmised, could be found by searching the genomes.

Methanogens: contemporary organisms preserving ancient metabolic strategies

Biological methane production (methanogenesis) emerged during the Archaean eon, at least 2.7 billion years ago (Ga), and represents one of the oldest strategies for energy production [6,7]. Unlike today, the archaean Earth lacked molecular oxygen (O_2), which did not accumulate until 2.3 Ga [8]. In its absence, Earth had a reducing environment that was rich in sulfide (S^{2-}) and low in sulfate [9,10,11]. Methanogenesis evolved under these conditions and is consequently an oxygen-sensitive process.

Methanogenesis persists today in contemporary methanogens, which comprise seven phylogenetic orders (*Methano-bacteriales*, *-coccales*, *-pyrales*, *-plasmatales*, *-microbiales*, *-cellales* and *-sarcinales*) within the archaeal phylum, *Euryarchaeota*. These organisms have been isolated from a wide variety of anaerobic environments that range broadly in temperature, salinity and pH, and include hydrothermal vents, marine and fresh water sediments, rice paddies, anaerobic digesters, land fills, insect guts, cattle rumen and the human intestine [12].

The original methanogenesis reaction involves hydrogen-dependent reduction of carbon dioxide [13]. However, since this pathway is less efficient than other hydrogenotrophic forms of energy production, methanogens are often out-competed in hydrogen-containing environments. Thus, to facilitate infiltration of methanogens into other environments, the methanogenesis pathway expanded to include the fermentation of acetate and the disproportionation (dismutation) of methanol, methylamines and methylsulfides in the orders *Methanosarcinales* and *Methanoplasmatales*, and the use of formate as an additional electron source in *Methanomicrobiales*, *Methanocellales, Methanococcales* and *Methanobacteriales* [14]. Collectively, methanogenesis emits more than 500 teragrams of methane into the atmosphere, annually, which is of broad interest due to the prominent role of methane in the global carbon cycle and as a greenhouse gas [15].

The core enzymes of methanogenesis are oxygen-sensitive, and were inherited vertically by contemporary methanogens [13], suggesting that these organisms and their ancestors have continuously inhabited anaerobic environments for at least 2.7 billion years. It is therefore plausible that other ancient, oxygen-sensitive metabolic strategies may be preserved in contemporary methanogens as well. Sulfur metabolism is of particular interest, because methanogens exhibit several distinguishing features that appear related to their ancient origin: (i) Methanogens employ two unique thiol-containing cofactors—CoM and CoB—to facilitate the final methane-liberating step of methanogenesis [16,17]. (ii) The predicted open-reading frames within methanogen genomes encode cysteine with unusually high frequency [18]. (iii) Methanogens utilize sulfide as a sulfur source in place of sulfate, which is preferred by aerobic microbes. (iv) Methanogens lack canonical genes for the biosynthesis of cysteine, homocysteine and persulfide [9,19]. (v) The intracellular concentration of free cysteine is quite low in *Methanococcus maripaludis*, inconsistent with a role as a metabolic hub as is the case in aerobic microbes [20]. (vi) Methanogens employ the unique SepRS-SepCysS pathway for Cys-tRNA synthesis, which was inherited vertically from the ancestral euryarchaeote along with methanogenesis [4,5].

Sulfur assimilation in aerobic model organisms

Sulfur assimilation in aerobic microbes occurs by the reduction of sulfate, which is transported into cells and converted to sulfide through a highly conserved, ATP-dependent, four-step pathway [21]. Sulfide is then incorporated into homocysteine for methionine and S-adenosylmethionine biosynthesis, and into cysteine for dispersal into other metabolites. This trafficking proceeds by persulfide (S^{1-}) relay initiated by cysteine desulfurases (CD) [22], which remove sulfane (S^0) from free cysteine, yielding alanine and a persulfided cysteine residue on the enzyme ([CD]-S-SH; **Fig. 0.3A, C**). The terminal sulfur of the CD-bound persulfide group is then relayed as sulfane to conserved cysteine residues on other proteins through a series of nucleophilic substitution reactions. By this process sulfur is ultimately inserted into a broad variety of compounds, including enzyme cofactors and RNA modifications [23]. In many cases the mobile sulfane is reduced to sulfide prior to insertion (**Fig. 0.3C**).

Cysteine and homocysteine can be synthesized by direct sulfhydrylation or transsulfuration (**Fig. 0.3A**). Direct sulfhydrylation usually proceeds by way of acetylated serine and homoserine derivatives as intermediates. In *Escherichia coli*, O-acetylserine and sulfide are converted to cysteine and acetate by O-acetylserine sulfhydrylase (OASS) [21]. An analogous reaction in *Saccharomyces cerevisiae*, catalyzed by O-acetylhomoserine sulfhydrylase (OAHS), is employed for homocysteine biosynthesis [24]. In the hyperthermophilic, aerobic archaeon *Aeropyrum pernix* and other *Crenarchaeota*, O-phosphoserine is used in place of O-acetylserine as an intermediate of cysteine biosynthesis, as catalyzed by O-phosphoserine sulfhydrylase (OPSS) [25]. OPSS is closely related to OASS, and its occurrence may be limited to *Crenarchaeota*.

The thermodynamics of sulfide incorporation have been investigated for OASS from 25 different species and OAHS from 7 different species, that are catalogued in the BRENDA database [26] (http://www.brenda-enzymes.org/). For all 32 enzymes, the K_M parameter for sulfide incorporation ranges between 6 µM and 7 mM; however, the overall and middle 50% average K_M values are 700 µM and 400 µM, respectively, suggesting that millimolar levels of sulfide are required for efficient catalysis in most organisms.

Transsulfuration is the two-step interconversion of cysteine and homocysteine, where the sulfur atom from cysteine is incorporated into homocysteine, and vice versa. It proceeds by way of the transsulfuration-specific intermediate cystathionine, which is produced in the first step following the condensation of either homocysteine and serine, or cysteine and an activated homoserine derivative (**Figs. 0.3A** and **3.4B**). Cystathionine synthesis is catalyzed by cystathionine-β-synthase (CBS) in the homocysteine-to-cysteine direction, and by cystathionine-γ-synthase (CGS) in the opposite direction. In the second step, cystathionine is cleaved by cystathionine-γ-lyase (CGL) or cystathionine-β-lyase (CBL) to yield cysteine or homocysteine, respectively. Both O-succinyl- and O-phosphohomoserine are known substrates for CGS-dependent cystathionine biosynthesis.

All seven enzymes catalyzing sulfhydrylation and transsulfuration require pyridoxal phosphate (PLP) as a cofactor. There are five structurally distinct fold types of PLP-dependent enzymes whose members are collectively responsible for more than one hundred different

catalytic activities [27]. The CBS, OASS and OPSS enzymes are members of fold type II and have remarkably similar sequences. For example, *E. coli* OASS is 42% identical to OASS from *Methanosarcina acetivorans*, but is also 34% identical to OPSS from *A. pernix* and 35% identical to CBS from *S. cerevisiae* [28]. The OAHS, CBL, CGL and CGS enzymes belong to fold type I, and are also quite similar to one another. OAHS from *S. cerevisiae* is 50% identical to OAHS from *M. acetivorans*, 29% identical to CGS from *E. coli*, 25% identical to CBL from *E. coli* and 31% identical to CGL from *S. cerevisiae* [28]. This high level of sequence conservation makes it difficult to assess genomes for the presence and absence of individual proteins for sulfhydrylation and transsulfuration. Thus, to resolve ambiguities in protein classification, the occurrence of cysteine and homocysteine-activating enzymes should be taken into consideration.

The acetylated precursors to direct sulfhydrylation are synthesized by acetyl-CoA-dependent serine and homoserine acetyltransferases (SAT and HSAT, respectively). Since these intermediates function only in cysteine and homocysteine biosynthesis, the occurrence of SAT or HSAT in genomes is a useful indicator for the a presence or absence of direct sulfhydrylation (**Fig. 3.4B**). Likewise, the metabolic utility of O-succinylhomoserine is limited to transsulfuration, so its biosynthetic enzyme, homoserine O-succinyltransferase (HSST), may be indicative of cysteine-to-homocysteine transsulfuration. By contrast, phosphoserine is synthesized in all organisms as the precursor to serine, and its biosynthetic enzyme is therefore useless from the perspective of comparative genomics. Similarly, phosphohomoserine, which is synthesized by homoserine kinase (HK), is commonly used for the biosynthesis of threonine. So, the presence of HK is not necessarily indicative of cysteine-to-homocysteine transsulfuration.

The CD enzyme is also PLP-dependent, but is a member of fold-type V (which includes SepCysS as well). The sulfane sulfur it removes from cysteine is relayed as persulfide to biosynthetic enzymes in pathways that yield cofactors and nucleotide modifications. There are four named CD types, which are classified into two distinct sub-groups. Group I contains IscS and NifS, while group II contains SufS and CsdA [29]. Whereas sulfur mobilized by NifS, SufS and CsdA is used only for iron-sulfur cluster biosynthesis, IscS-derived persulfide is promiscuously relayed

for incorporation into the cofactors thiamine, biotin and molybdenum cofactor, and tRNA modifications s^2C32, ^{s}2U34, ms^2io^6A37 and ^{s}4U8, in addition to iron-sulfur clusters [30].

Downstream of IscS, sulfur relay can be mediated by a variety of intermediary sulfur transfer proteins. Of these, rhodanese domains (RHD; pfam00581) and ubiquitin-like proteins (Ubl) have been are best understood, and are worth mentioning to establish precedents for sulfur-transfer chemistry. RHD are approximately 120 aa in length and contains a single conserved cysteine residue, which accepts sulfane from CD. RHD may occur as a stand-alone protein or in fusions with other functional domains. Over two-hundred RHD-containing architectures are catalogued in the Pfam databse [31]; among these, the ThiI-RHD fusion from *E. coli* and related *Bacteria* has been studied most extensively [32–35].

The ThiI-RHD fusion protein is a bifunctional enzyme essential to the incorporation of sulfur into the s^4U8 tRNA modification and thiamine (for which ThiI is named). In s^4U8 synthesis, the conserved cysteine of RHD accepts sulfane from CD, which is then inserted directly into the adenylate-activated nucleotide as sulfide, enabled by the concomitant formation of a disulfide between the RHD cysteine and another conserved cysteine present in the ThiI domain. For subsequent rounds of catalysis, this disulfide needs to be reduced by an electron donor, which has yet to be identified. Similarly, in thiamine biosynthesis RHD accepts sulfane from CD, but does not directly insert it for thiazole synthesis. Instead, sulfane is relayed for insertion as sulfide into Ubl (termed ThiS in this instance), which serves as a sulfur carrier for thiazole biosynthesis catalyzed by the ThiGH protein complex [36,37].

Ubl is approximately 90 amino acids in length is best known for its role in protein degradation; however, it also serves as sulfur carrier for the biosynthesis of thiamine, molybdenum cofactor and s^2U34 [38,39]. In either role, Ubl need to be activated by E1-like proteins (E1l), which transfer an adenylate group to the C-terminus of Ubl. When used as a sulfur carrier, the adenylate of Ubl is substituted with sulfide liberated from a persulfided cysteine on E1l, which is enabled by concomitant disulfide formation. The resulting thiocarboxylated Ubl is a sulfur carrier that may release sulfide for insertion following water mediated hydrolysis.

The mechanisms of sulfur insertion into cofactors and nucleotide modifications are not clearly understood. However, it is typical for two or three cysteine residues to be essential to catalysis [40–42].

Novel sulfur metabolism enzymes await discovery in methanogens

In nearly all microbial genomes, known protein-codign genes for the biosynthesis of homocysteine can be found in combinations that are indicative of prototrophy. This is not the case for many methanogens, whose genomes are often devoid of known genes encoding enzymes catalyzing either direct sulfhydrylation or transsulfuration (**Fig. 0.4**) [20]. Since these organisms are capable of growth with sulfide as the lone sulfur source, they must employ some novel alternative process to incorporate sulfide into homocsyteine.

Similar logic led to the discovery of the SepRS-SepCysS pathway for Cys-tRNA synthesis. Since the canonical gene for Cys-tRNA synthesis, which encodes CysRS, was missing from methanogen genomes, there had to be an alternative enzymatic pathway in its place. SepRS and SepCysS were identified using a classical biochemical approach, in which the cellular lysate of *Methanocaldococcus jannaschii*, which lacks CysRS, was fractionated through a series of chromatographic separations. Fractions that catalyzed Cys-tRNA synthesis in the presence of exogenously added tRNA and a low-molecular-weight cell extract (< 5kDa), were found to contain two uncharacterized proteins that became known as SepRS and SepCysS [4].

The use of SepRS and SepCysS is coupled to a unique strategy for sulfur assimilation. From a genomic perspective, the SepRS-SepCysS pathway represents the lone route to cysteine biosynthesis in methanogens lacking the OASS- and CBS/CGL-dependent pathways. This notion is supported by the observation that a SepRS-deletion strain of *Methanococcus maripaludis*, which also encodes CysRS, requires exogenously added cysteine for growth [4]. Consistent with the absence of tRNA-independent cysteine biosynthesis pathways, the metabolic role of free cysteine is deemphasized in these organisms. In *M. maripaludis*, cysteine is present at lower concentrations, relative to *E. coli*, and does not serve as the sulfur source for methionine or iron-sulfur cluster biosynthesis [20]. Free cysteine, however, is not completely without metabolic utility,

since it likely serves as the sulfur source to CoA biosynthesis, based on the presence of phosphopantothenate-cysteine ligase homologs in all methanogens [28]. Presumably, free cysteine is generated and recycled following protein degradation for this purpose. Or perhaps some is generated by deacylation of Cys-tRNA.

More than half of genomes encoding the SepRS-SepCysS pathway, lack the OAHS- and CGS/CBL-dependent pathways (**Fig. 0.4**). Furthermore, while these genomes do encode all three enzymes for homoserine biosynthesis from aspartate, they do not share a common gene for homoserine activation, and frequently lack homoserine-activating genes in general (**Fig. 0.4**). Taken together, these correlations link SepRS and SepCysS to a novel pathway for homocysteine biosynthesis that awaits identification. The nature of this link could be the means through which both biosynthetic pathways acquire sulfur for catalysis (**Fig. 0.3B**).

The SepCysS reaction consumes a single unit of sulfide (S^{2-}). However, free sulfide does not serve as an efficient sulfur source for this enzyme *in vitro* [28,43]. Instead, it was proposed—and later supported—that persulfide relay is responsible for delivering sulfur to SepCysS. The IscS-type CD from *E. coli* productively relays persulfide to SepCysS *in vitro* and *in vivo* (when recombinantly expressed in *E. coli*) [44,45]. However, a CD is unlikely to be the ancestral sulfur source for SepCysS in methanogens because (i) known CD genes are often completely absent from the genomes of methanogens (**Fig. 0.4**); (ii) genes responsible for the synthesis of free cysteine are also often absent from methanogen genomes, consistent with the finding that the SepRS-SepCysS pathway provides the sole *de novo* cysteine biosynthesis pathway in *M. maripaludis* [4]; (iii) the intracellular pool of free cysteine in *M. maripaludis* is five to 10-fold lower than in aerobic bacteria [20]; (iv) cysteine is not the sulfur source for either iron-sulfur cluster or methionine biosynthesis in *M. maripaludis* [20]; (v) sulfur donation from CD to SepCysS would represent a fruitless cycle because it would require cysteine as a source for its own biosynthesis [23]. Unlike in aerobes, the low concentration of free cysteine in methanogens suggests that this metabolite does not play a central role as a metabolic hub for anabolic sulfur distribution. Nonetheless, the presence of SepCysS suggests that an alternative means for initiating persulfide relay exists.

In addition to SepCysS, persulfide groups have been identified in two *M. maripaludis* proteins involved in the biosynthesis of thiolated tRNA nucleosides, following recombinant expression in *E. coli* [29,30]. One of these proteins is a ThiI homolog, but lacks the RHD domain commonly fused with ThiI in aerobic microbes. In its place, *M. maripaludis* ThiI possesses a conserved CXXC motif, which adopts disulfide and persulfide configurations that are proposed to mediate catalysis. Remarkably, this ThiI homolog was capable of performing multiple turnovers of s^4U8 synthesis *in vitro* with sodium sulfide as a sulfur source, indicating that this enzyme does not require persulfide relay for catalysis. However, it remains unclear whether methanogen-type ThiI is capable of fixing sulfide *in vivo*. The other characterized tRNA thiolating enzyme of *M. maripaludis*, Ncs6, catalyzes the biosynthesis of s^2U34. Like *M. maripaludis* ThiI, Ncs6 possesses a conserved CXXC motif; however, sulfide-dependent synthesis of s^2U34 could not be reconstituted *in vitro* with Ncs6. Interestingly, Ncs6 was found to associate with Ubl (known as SAMP in *Archaea*) *in vivo*. This suggests *M. maripaludis* Ncs6 might require thiocarboxylated SAMP as a sulfur donor, which would be consistent with previous studies demonstrating that Ubl proteins function as sulfur carriers for s^2U34 biosynthesis catalyzed by Ncs6 homologs in yeast and the halophilic archaeon *Haloferax volcanii* [48,49].

The sulfide-dependent activity of ThiI and the association of Ncs6 with SAMP are contradictory in their implications. While the former observation suggests persulfide may have a diminished role in methanogens, which may instead use sulfide directly, the latter suggests persulfide-dependent processes known to occur in aerobes may also be employed in methanogens. All methanogen genomes encode homologs of SAMP, their E1l activating proteins and RHD [28]. Although none of these methanogen proteins have been examined at the biochemical or genetic levels, their presence suggests that sulfur transfer events taking place downstream from persulfide synthesis in methanogens might resemble those which occur in aerobic model organisms.

A model representing the ancestral-methanogen-type strategy for sulfide assimilation was constructed based on (i) the requirement of SepCysS for persulfide; (ii) the frequent absence of OAHS, CGS and CBL from methanogens; (iii) the absence of cysteine-to-homocysteine

transsulfuration from *M. maripaludis*; (iv) the assumption that the biosynthetic pathways for homocysteine and cysteine have a shared sulfur source; (v) the presence of RHD, SAMP and E1l proteins in methanogens (**Fig. 0.3B**). With this model as a guide, all archaeal genomes were searched to identify three novel proteins occurring with SepCysS (see Chapter 1). Subsequent genetic and biochemical experiments were used to confirm roles in sulfide assimilation for all three of these proteins (Chapters 2-4). Moreover, phylogenetic analyses indicated that the discovered proteins were inherited vertically from the ancestral euryarchaeote along with SepCysS, suggesting these proteins comprise an ancient strategy for sulfide assimilation that emerged in the Archaean eon.

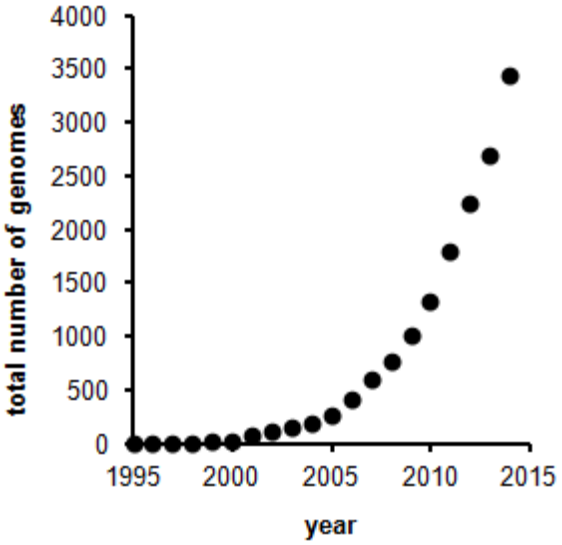


Figure 0.1 The total number of prokaryotic genomes by year. Genome Reports data were downloaded from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/GENOME_REPORTS/). The scatter plot reflects the total number of completely sequenced genomes from *Bacteria* and *Archaea* that were released by the end of each year. The research presented in this dissertation was conducted between 2009 and 2015. During this time, the number of genomes has increased by more than three-fold.

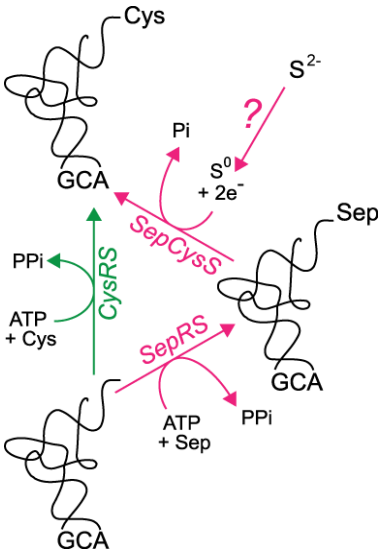


Figure 0.2 Dual pathways for cysteinyl-tRNA synthesis. The ancestral methanogen strategy (fuchsia) employs phosphoseryl-tRNA synthetase (SepRS), which catalyses the ATP-dependent aminoacylation of $tRNA^{Cys}$ with phosphoserine (Sep). The resulting Sep-$tRNA^{Cys}$ intermediate is converted to Cys-$tRNA^{Cys}$ by Sep-tRNA:Cys-tRNA synthase (SepCysS), which requires sulfane (S^0) for catalysis. Sulfane is thought to be delivered to SepCysS by persulfide relay; however, the ancestral methanogen enzyme for persulfide synthesis (?) has not been identified. For multiple turnovers, SepCysS requires the donation of two elections ($2e^-$), since sulfide (S^{2-}) is required is required to substitute for the leaving phosphate group (Pi). The conventional strategy (green) employs cysteinyl-tRNA synthetase (CysRS) for the ATP-dependent aminoacylation of $tRNA^{Cys}$ with cysteine.

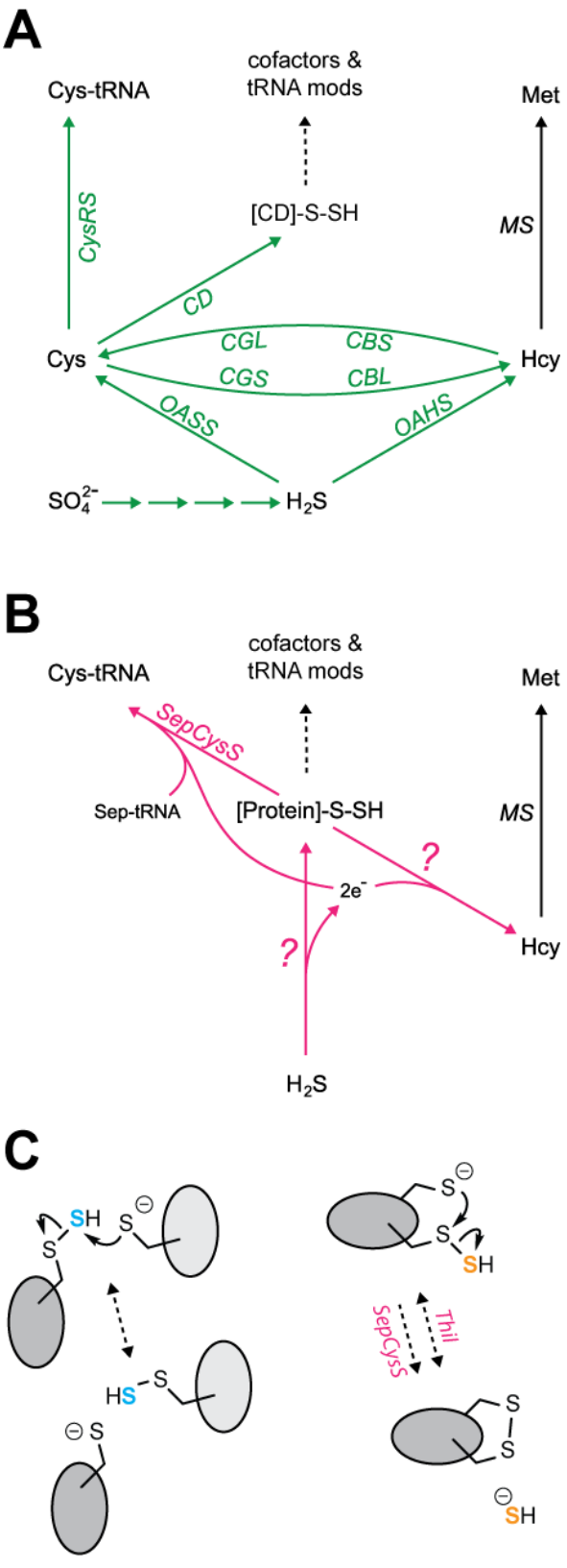


Figure 0.3 Sulfur assimilation and persulfide relay. **(A)** The pathways for sulfur assimilation typical of aerobic model organisms proceed from sulfate and include universally conserved enzymes and processes (black arrow) along with others largely absent from methanogens (green). These include methionine synthase (MS), RHD- and Ubl-dependent sulfur transfer (dashed arrows; see main text), cysteinyl-tRNA synthetase (CysRS), cysteine desulfurase (CD), O-acetylserine sulfhydrylase (OASS), O-acetylhomoserine sulfhydrylase (OAHS), cystathionine gamma-synthase (CGS), cystathionine beta-lyase (CBL), cystathionine beta-synthase (CBS), cystathionine gamma-lyase (CGL) and sulfate reduction. **(B)** Known and proposed reactions in the ancestral euryarchaeote (fuchsia) are presumed to occur in all genomes also encoding SepCysS. These include Sep-tRNA:Cys-tRNA synthase (SepCysS) and proposed sulfide-dependent persulfide and persulfide-dependent homocysteine biosynthesis enzymes (question marks indicate that the existence of the pathways is not established). **(C)** Transformations of persulfide (S^{1-}) include the relay of sulfane (blue; S^0) between cysteine residues of different proteins (left) as well as the release of sulfide (orange; S^{2-}) from persulfide and concomitant disulfide formation (right). SepCysS is thought to release sulfide irreversibly, since sulfide is a poor substrate, and therefore does not efficiently regenerate persulfide from disulfide. Conversely, the ThiI homolog of *Methanococcus maripaludis* may engage in reversible persulfide-disulfide exchange, since it is able to catalyze multiple turnovers of sulfide-dependent s^4U8 synthesis *in vitro* [47].

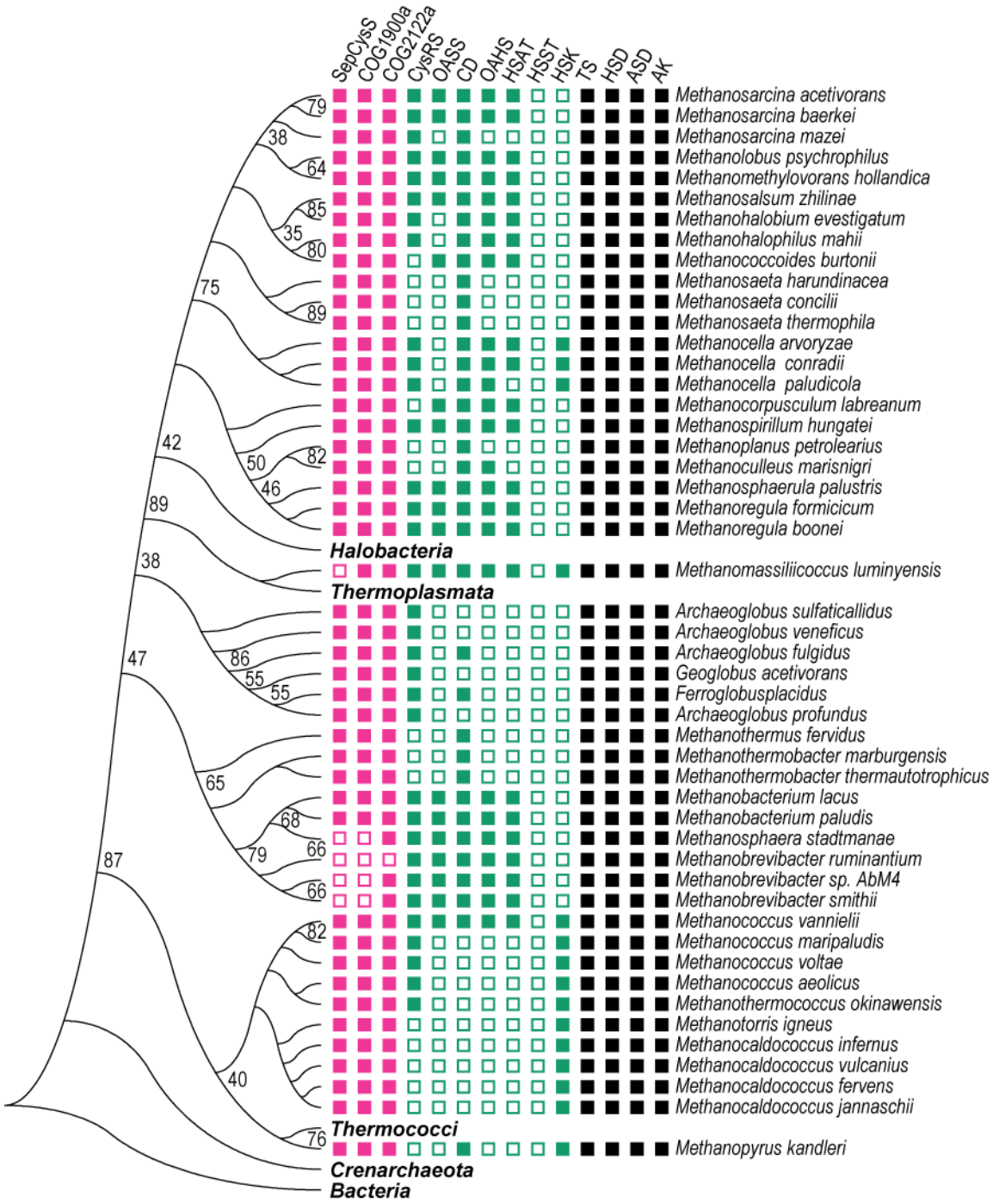


Figure 0.4 The occurrence profile of proteins for sulfide assimilation in methanogens and *Archaeoglobi*. The presence or absence of individual proteins is indicated with filled and open squares, respectively. Novel proteins COG1900a and COG2122b were found to co-occur with SepCysS and comprise an ancient strategy for sulfide assimilation (fuchsia; see subsequent chapters). Proteins for homoserine activation, sulfhydrylation and persulfide synthesis that are typically present in aerobic model organisms (green) are largely absent. These include homoserine acetyltransferase (HSAT), homoserine succinyltransferase (HSST), homoserine kinase (HK), O-acetylserine sulfhydrylase (OASS), O-acetylhomoserine sulfhydrylase (OAHS) and cysteine desulfurase (CD) whereas universally conserved proteins (black) for

the biosynthesis of threonine from aspartate are present. These include aspartate kinase (AK), aspartatesemialdehyde dehydrogenase (ASD), homoserine dehydrogenase (HSD) and threonine synthase (TS). Although HK was not detected in all genomes, it is likely present in a unique form since TS is O-phosphohomoserine-dependent and represents the lone biosynthetic route to threonine. The depicted tree is a phylogenetic reconstruction of 16S rRNA sequences that is intended to approximate organismal phylogeny. Bootstrapping numbers reflect a percentage of 1000 replicates, and are not shown when greater than or equal to 98%. The tree was constructed in MEGA 6 using the neighbor-joining method (p-distance) with complete deletion of gaps.

CHAPTER 1

IDENTIFICATION OF THREE NOVEL PROTEINS CO-CONSERVED WITH SEPCYSS

The logic of occurrence profiling

To identify novel genes involved in methanogen-type sulfide assimilation, SepCysS-dependent occurrence profiling was employed. This general bioinformatic method, also known as phylogenetic profiling [50], assumes that genes of interdependent function are likely to be present in, and absent from, the same genomes. Thus, the method seeks to infer a gene's function based on its phylogenetic distribution, relative to other genes with known function.

In accordance with the model for methanogen-type sulfide assimilation (see Introduction; **Fig. 0.3B**) the occurrence of the gene encoding SepCysS is predicted to correlate to hypothetical genes for the biosynthesis of persulfide and homocysteine. Because SepCysS requires persulfide for catalysis, the persulfide biosynthesis protein should be encoded in all genomes that have SepCysS. However, since the persulfide biosynthesis protein may also provide sulfur for methanogen-type homocysteine biosynthesis, which probably does not depend on SepCysS to function, it is plausible that the hypothetical proteins be encoded in genomes lacking SepCysS.

Still, the success of occurrence profiling depends on the designation of control genomes, which are predicted to lack the hypothetical homocysteine and persulfide biosynthesis proteins. Subtracting proteins occurring in control genomes serves to prevent the false identification of proteins involved in metabolic processes unrelated to sulfur metabolism. Since methanogenesis is thought to require more than 200 different proteins [51], and genomes encoding SepCysS belong to either methanogens or the closely related *Archaeoglobales*, which encode a near-complete set

of methanogenesis proteins [52,53], the ideal control genome would belong to a methanogen lacking SepCysS. At the time of the analysis, only three of the forty methanogens with completely sequenced genomes (*Methanobrevibacter smithii*, *Methanobrevibacter ruminantium*, *Methanosphaera stadtmanae*) were known to lack SepCysS (**Fig. 0.4**). Moreover, each of the SepCysS-lacking genomes encoded aerobe-type proteins for the biosynthesis of free cysteine, homocysteine and persulfide, suggesting that the analogous ancestral-methanogen-type proteins were no longer needed and thus could have been lost.

Occurrence profiling with COG and Pfam databases

A typical computational approach for occurrence profiling employs a relational database (a table) cataloguing the genome-dependent occurrence of all proteins as classified into function-specific families. Two prominent examples of databases used for occurrence profiling are Pfam [54,55] and COG [56,57] (clusters of orthologous groups). In terms of their utility, these databases are essentially the same, but they do differ in their systems for classifying proteins. Both are publicly available and can be downloaded for free from their respective websites.

In an initial, unsuccessful attempt to identify novel genes occurring in genomes with SepCysS, the COG database was downloaded and manipulated using SQL (structured query language). First, to eliminate false-positives, all data for COGs that were present in the SepCysS-lacking methanogens were removed from the database. Subsequently, to identify proteins co-occurring with SepCysS, data for COGs that were missing from individual SepCysS-encoding genomes were iteratively removed from the database until only data for COGs present in all SepCysS-encoding genomes remained. Since the only remaining COGs corresponded with SepRS and SepCysS, no new COGs were identified. Pfam-dependent occurrence profiling was not attempted.

It was reasoned that the failure of this approach had several possible causes. (i) Since microbes are capable of employing redundant metabolic pathways, the presence of the ancestral proteins for homocysteine and persulfide synthesis cannot be ruled out in the SepCysS-lacking methanogens, which encode OAHS, OASS and CD (**Fig. 0.4**). Thus, if the target genes were

indeed present in the genomes of SepCysS-lacking methanogens, then they would have been erroneously removed from the dataset. (ii) Several SepCysS-encoding methanogens also possess a full set of aerobe-type proteins for sulfide assimilation. Thus, in these organisms, the hypothetical ancestral proteins for homocysteine and persulfide biosynthesis would be dispensable. If the target genes were indeed lost from one of these genomes, then they would have been erroneously eliminated from the dataset. (iii) Since the grouping of proteins of unknown function into COGs is based solely on sequence conservation, misclassification might arise in cases where protein families exhibit relatively high or low sequence similarity. For example, a single group of orthologous proteins might be classified as two distinct groups if their sequences are sufficiently diverged. Or, alternatively, two paralogous groups of proteins might be classified as a single group if their sequences are sufficiently similar. Thus, if the target proteins were classified into COGs incorrectly, then they would have been erroneously eliminated from the dataset because their COG-based occurrence profiles would be incongruous with their actual occurrence profiles.

An alternative, manual approach for occurrence profiling

To avoid problems associated with the classical computational approach to occurrence profiling, an original, alternative method was employed (**Fig 1.1A**). The approach did not rely on static protein classifications like COG and Pfam, nor did it require the use of a strict occurrence profile. Instead, individual proteins were simultaneously classified into orthologous groups and assessed for their occurrence profile through the construction of phylogenetic trees. The disadvantage of this method is that it is relatively time-consuming, requiring up to one hour per protein analyzed. Nonetheless, it was favored since it was less likely to overlook potential proteins of interest.

Genomes with SepCysS encode between 1283 and 4540 total proteins. Thus, achieving complete genomic coverage with this method is not practicable, as it would require more than 1000 hours to search a relatively small genome. To accelerate the process, proteins were

preferentially selected for analysis if homologous to, or encoded adjacent to those with putative roles in sulfur assimilation.

This method led to the identification of three proteins of interest (**Fig. 1.1B**). The first protein was identified following a BLASTp search limited to all archaeal proteins of the non-redundant protein sequence database (NCBI) using the protein query MA2714 from *Methanosarcina acetivorans*, which is predicted to encode homoserine acetyl-transferase (HSAT). MA2714 was used as a query because of its proximity to MA2715, an OAHS homolog. Fused to the HSAT catalytic domain of MA2714 is an additional 150 aa domain annotated CBS. The BLASTp search identified a distinct set of orthologous proteins that contained a similar CBS domain, which was present in all SepCysS-encoding genomes and absent from all other archaeal genomes. The identified protein family contained an additional 350 aa domain, annotated as COG1900, which was of unknown function. The second (NIL-Fer) and third (COG2122) proteins of interest were identified because they were encoded in genomic neighborhoods with the COG1900a-CBS fusion protein (**Fig. 1.3**). Both were found to be present in all genomes encoding SepCysS (**Fig. 0.4**), and were of unknown function.

Description of the proteins of interest

The ~500 aa COG1900-CBS protein from *Archaea* contains two distinguishable regions. The Conserved Domain Database (CDD) identifies a ~350 aa N-terminal domain as COG1900 (DUF39; cl14897), followed by a ~150 aa C-terminal region assigned to COG0517 (CBS_pair superfamily; cl15354; **Fig. 1.1B**) [58,59]. A BLASTp search of the non-redundant protein sequence database (NCBI) using the N-terminal domain identified 373 unique protein sequences with significant alignments (e-value < 10^{-8})—all classified within COG1900. The COG1900 family was classified further into four subdivisions, COG1900a-d (**Fig. 1.2**; see below for further elaboration).

COG1900a contains all sequences from genomes encoding SepCysS, and 182 total unique protein sequences. Of these, 64 sequences belong to SepCysS-encoding *Euryarchaeota* and two sequences are from the *Methanoplasmatales*, a very recently described taxon that lacks SepCysS [60]. The remaining 118 sequences belong to 57 bacterial genera from nine different

phyla [19] (data not shown). All species found in the bacterial genera encoding COG1900a are anaerobic, suggesting that the function of the protein may be oxygen-sensitive. Further, COG1900a is often encoded in close genomic proximity with genes predicted to play a role in sulfur metabolism, especially those required for *de novo* biosynthesis of methionine and sulfur-containing cofactors (**Fig. 1.3**). The large N-terminal domain of COG1900a contains two highly conserved cysteine residues and lacks a conserved lysine residue, suggesting that the proteins do not bind pyridoxal phosphate (PLP; **Fig. 1.1B**).

In all SepCysS-encoding genomes, CBS domains (named for the protein cystathione-β-synthase) are located C-terminal to COG1900a. These CBS domains are absent from COG1900a proteins encoded in genomes lacking SepCysS, suggesting a nonessential function. CBS domains are thought to function in a regulatory capacity in response to binding of adenosine derivatives [61]. This is consistent with the recent finding that the CBS domains found in the *Methanocaldococcus jannaschii* COG1900a protein, MJ0100, undergo a conformational change in response to binding of *S*-adenosylmethionine (AdoMet) [62]. CBS domains are typically found in proteins involved in energy and sulfur metabolism [63]. Their presence exclusively in COG1900a proteins from genomes containing SepCysS suggests that the functions of COG1900a and SepCysS may be linked.

The ~130 amino acid NIL-Fer protein contains two distinct domains recognizable within the CDD. The ~60 aa N-terminal domain belongs to the NIL superfamily (cl09633; pfam09383; smart00930), a proposed substrate binding domain found in the C-terminal, intracellular region of the MetN subunit of the methionine ABC transporter [64]. The ~50 aa C-terminal portion of NIL-Fer is of the Fer4_7 superfamily (cl19102; pfam12838), which contains eight conserved cysteine residues and is predicted to bind two iron sulfur clusters [4Fe-4S]. All SepCysS-encoding genomes encode this NIL-Fer fusion protein. NIL-Fer is almost always encoded immediately downstream of COG1900a proteins, although a variation exists in *Thermotoga*, *Thermosipho*, *Fervidobacterium* and *Coprothermobacter* spp., in which the Fer4_7 iron-sulfur clusters from NIL-Fer are fused to the C-terminus of COG1900a and NIL is missing alltogether (**Fig. 1.3**). There are also several genomes that possess a presumed NIL-Fer ortholog lacking the NIL region (**Fig.**

1.3). Together, these observations suggest that NIL may be dispensable, that Fer4_7 is the essential functional element of NIL-Fer proteins, and that the functions of COG1900a and NIL-Fer proteins are linked.

Finally, the third ~250 amino acid protein of interest belongs to COG2122 (cl17892; PRK04334). COG2122 is uncharacterized but is classified within the ApbE superfamily (cl18387), which features FAD-binding enzymes implicated in thiamine biosynthesis and iron-sulfur cluster maintenance [65,66]. The protein contains a single conserved cysteine residue, and no conserved lysine residues. A BLASTp search of all genomes using the COG2122 representative from *Methanosarcina acetivorans* (MA1715) identifies 171 unique protein sequences (76 in *Archaea*) yielding significant alignments (<1e-3), all classified within COG2122. Two subdivisions were found and were designated COG2122a-b (**Fig. 1.2**), with proteins co-occurring with SepCysS falling in COG2122a. The occurrence of COG2122a in *Archaea* and *Bacteria* is nearly identical to that of COG1900a; the two families co-occur in over 98% of genomes. Furthermore, COG2122a is often encoded in close genomic proximity to COG1900a, and is also often found immediately downstream from NIL-Fer proteins (**Fig. 1.3**).

The evolutionary history of COG1900a, NIL-Fer and COG2122a proteins

Phylogenetic reconstructions of the COG1900a and COG2122a protein families show that each is congruent with respect to currently accepted organismal phylogenies (data not shown). To provide a further assessment, we repeated the analysis with a single sequence corresponding to the concatenated protein sequences for orthologs of COG1900a, NIL-Fer and COG2122a. This reconstruction again yielded a tree that is nearly identical to those observed for methanogen genome phylogeny and methanogenesis enzymes, and suggests all three of the identified proteins were present in the ancestral euryarchaeote (**Fig. 1.4A**) [13,67]. This is consistent with the notion that SepCysS is also an ancient gene [5], and that the mechanism for sulfide assimilation in methanogens is linked to the origins of methanogenesis.

To extend this analysis to the Bacteria, the evolutionary histories of COG1900a and COG2122a were examined further at the domain-domain level using ratios of evolutionary

distances (RED) analysis [68]. For both protein families, intra-phylum and -domain distances separating orthologs scaled proportionately to the corresponding control distances, suggesting widespread vertical inheritance within *Bacteria* and *Archaea* (**Fig. 1.4A-B**)**.** Interestingly, for both COG1900a and COG2122a, interdomain distances did not extrapolate linearly from the intradomain data**.** This nonlinearity likely reflects an ancient interdomain lateral gene transfer event that predated the divergence of bacterial phyla. Alternatively, nonlinearity might be the result of a change in the evolutionary rates for orthologs of both COG1900a and COG2122a. The analysis also detected the existence of lateral gene transfer among some species within the bacterial domain (**Fig. 1.4C**).

Classification of COG1900 subgroups

Phylogenetic reconstruction reveals that the 373 proteins recovered in BLASTp searches are classifiable into four subdivisions, which have been termed COG1900a-d (**Fig. 1.2**). COG1900a is the largest of the four. None of the proteins in the other three groups possess the CBS domain. COG1900b and COG1900c occur exclusively in *Cyanobacteria*, and consist of 108 and six sequences, respectively. Of the two cysteine residues conserved in all COG1900a proteins (C54 and C131 in *Methanosarcina acetivorans* protein MA1821), the proteins in COG1900b contain C131 only, while those in COG1900c and COG1900d possess neither. C54 is substituted as E, D or S in COG1900b, COG1900c and COG1900d, respectively (**Fig. 1.5**).

COG1900d proteins possess a Fer4_16 (pfam 13484) domain containing two 4Fe-4S clusters fused to their C-termini, and are found almost exclusively in methanogens. These proteins possess S54 instead of C54, and lack a consensus TDCYP motif that surrounds C131 of COG1900a (**Fig 1.5**). The lack of conservation of these cysteine residues in COG1900d, and the co-occurrence of COG1900d with COG1900a in methanogens suggests COG1900d and COG1900a are of diverged function, and that the function of COG1900d may be related to the methanogenesis lifestyle.

Discussion

Based on its requirement of persulfide as a sulfur source, and its frequent occurrence in genomes lacking known proteins for sulfur incorporation into homocysteine, the ancestral methanogen enzyme for cysteine biosynthesis, SepCysS, was considered to be a genomic marker for the presence of novel proteins for persulfide and homocysteine biosynthesis. An original approach was employed to perform SepCysS-dependent occurrence profiling, which identified the COG1900a, NIL-Fer and COG2122a proteins. The identified protein-coding genes were found to be present in all genomes encoding SepCysS as well as in over fifty genera of strictly anaerobic *Bacteria*. Phylogenetic reconstruction revealed that the COG1900a, NIL-Fer and COG2122a proteins were vertically inherited in methanogens, indicating a shared evolutionary history with SepCysS. Moreover, each of the identified proteins display molecular signatures consistent with roles in persulfide manipulation, and are found in conserved genomic neighborhoods with known sulfur assimilation proteins. Taken together, these observations suggest COG1900a, NIL-Fer and COG2122a may in fact be the missing proteins for persulfide and homocysteine biosynthesis.

In retrospect, it is now clear why the COG-dependent approach for occurrence profiling failed to identify COG1900a, NIL-Fer and COG2122a. The sequences of the COG1900 subgroups are conserved to the extent that they were misclassified as a single orthologous group. Although the COG1900a and COG1900d representatives within *Methanosarcina acetivorans* share 31% sequence identity, they are likely to be of diverged function since they are clearly distinguished by phylogenetic reconstruction (**Fig 1.2**) and are present together in nearly all methanogen genomes. Moreover, since COG1900d is present in methanogens lacking SepCysS and absent from *Archaeoglobales* possessing SepCysS, the classification of this group within COG1900 prevented the identification of COG1900 as a protein of interest.

The NIL-Fer protein was not identified by the COG-dependent approach because neither of its domains are catalogued within the COG database. Although NIL and Fer are both recognized within the Pfam, using Pfam instead of COG would not have identified NIL-Fer. This is because some of the NIL-corresponding regions of the identified NIL-Fer proteins are not classified within

NIL (pfam09383) or any other Pfam. For example, the NIL region of NIL-Fer from *Methanoregula boonei* shares 35% identity with that from *M. acetivorans*. Despite their clear homology, the *M. acetivorans* homolog is included within pfam09383, whereas the *M. boonei* homolog is not. Thus, according to the Pfam database, NIL-Fer is absent from the SepCysS-encoding genome of *M. boonei*.

The COG-dependent approach had overlooked COG2122a, not due to misclassification, but because it was present in several methanogens lacking SepCysS, and was thus eliminated from the dataset. The Pfam database misclassifies COG2122a proteins within the ApbE superfamily (pfam02424), which has representation in nearly all *Methanosarcinales*.

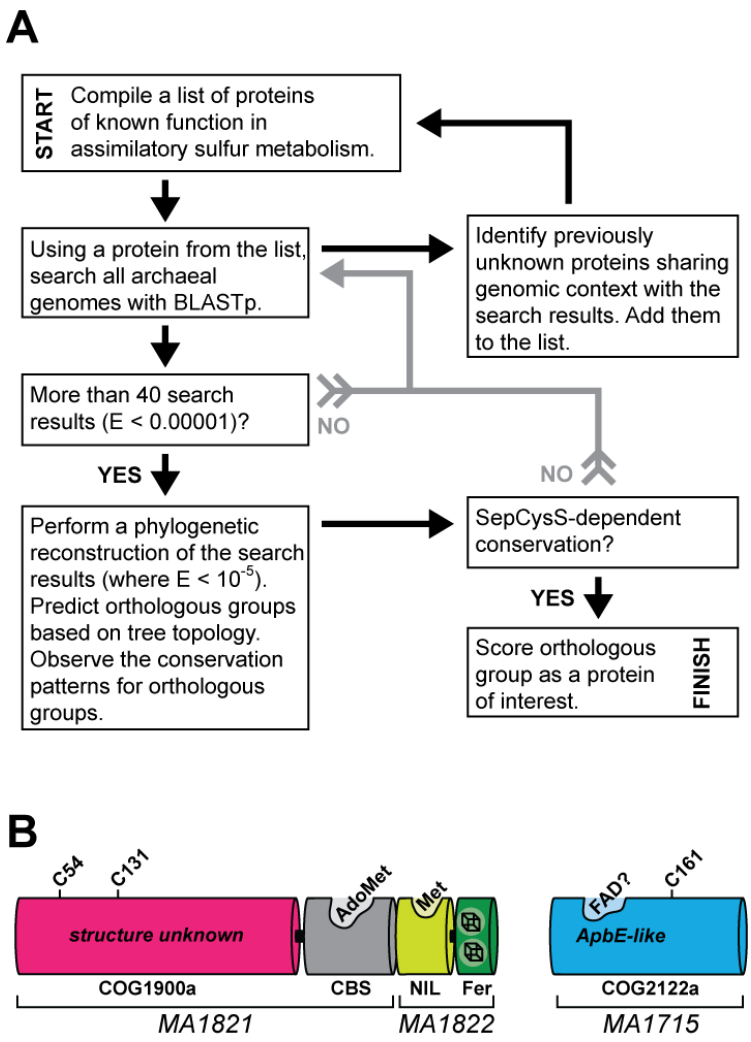


Figure 1.1 Novel proteins identified by SepCysS-dependent occurrence profiling. **(A)** A flow chart depicts the occurrence profiling strategy used to identify COG1900a-CBS, NIL-Fer and COG2122a. Over 100 archaeal genomes were used in this strategy at the time this study was completed. Of these, 41 genomes encoded SepCysS. **(B)** Cartoons for COG1900a-CBS, NIL-Fer and COG2122a, which are represented by *Methanosarcina acetivorans* proteins MA1821, MA1822 and MA1715 are drawn to scale. COG1900a (fuchsia) is of unknown structure and function, but contains two conserved cysteine residues. The C-terminal CBS domains (gray) of MA1821 are proposed to bind *S*-adenosylmethionine. The NIL domain (yellow) of MA1822 may bind methionine based on homology to the methionine transporter. The Fer_4_7 domain (Fer; green) of MA1822 has eight conserved cysteine residues thought to bind two [4Fe-4S] clusters. MA1715 (COG2122a; blue) and has a single conserved cysteine residue. Its bacterial paralog, ApbE, binds FAD and is involved in thiamine biosynthesis. It is not known if COG2122a binds FAD.

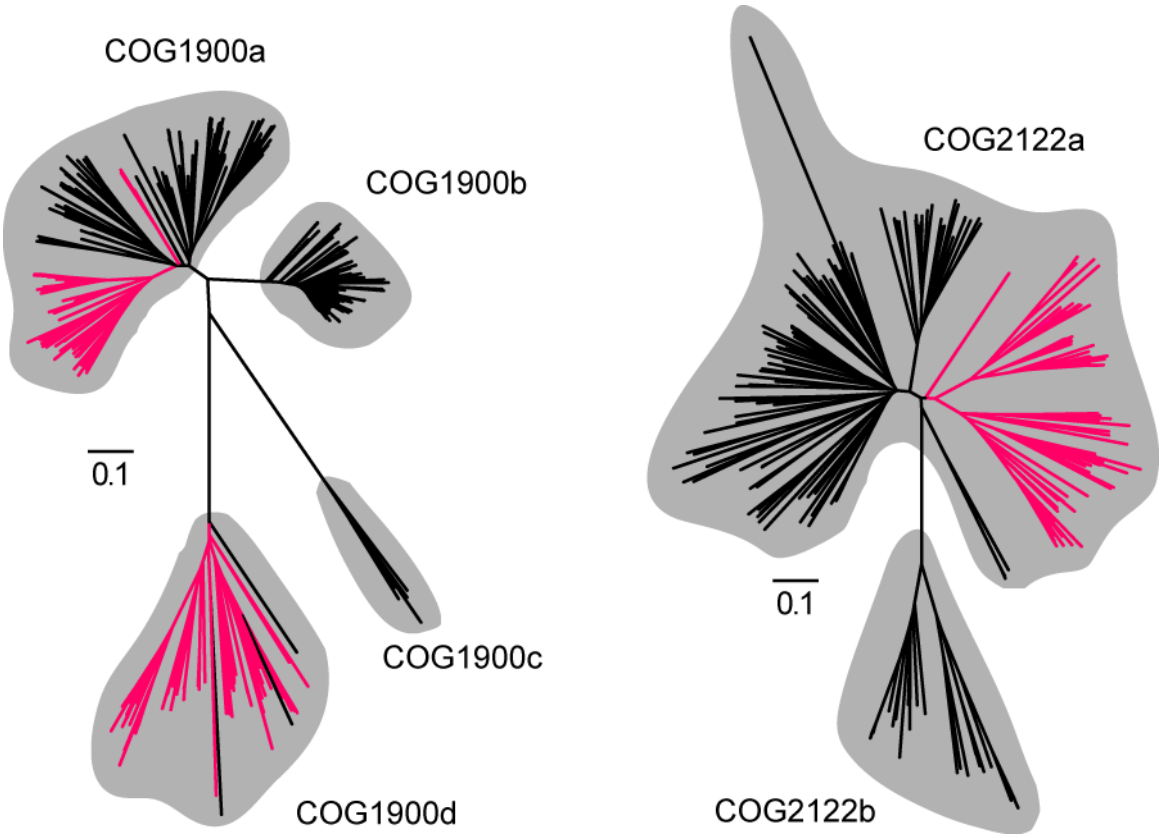


Figure 1.2 The subdivisions of COG1900 (PF01837; left) and COG2122 (PF02424**;** right) families. Protein sequences for all members of both families were identified using the BLASTp utility. The sequences were downloaded and aligned with Clustal W using default settings. Phylogenetic reconstructions were performed in MEGA5 using the neighbor-joining method with complete deletion of gaps. Both families are represented in *Archaea* (fuchsia) and *Bacteria* (black). Subdivisions COG1900a and COG2122a are the examined with reverse-genetics experiments in Chapters 3 and 4, respectively, and are exclusively present together in anaerobic microbes. COG1900d is present in nearly all methanogens and is thought to catalyze sulfur insertion into coenzyme M (see Chapter 3 Discussion). COG1900b-c occur in *Cyanobactiria*, which lack COG2122 proteins. COG2122b occurs in aerobic *Proteobacteria*.

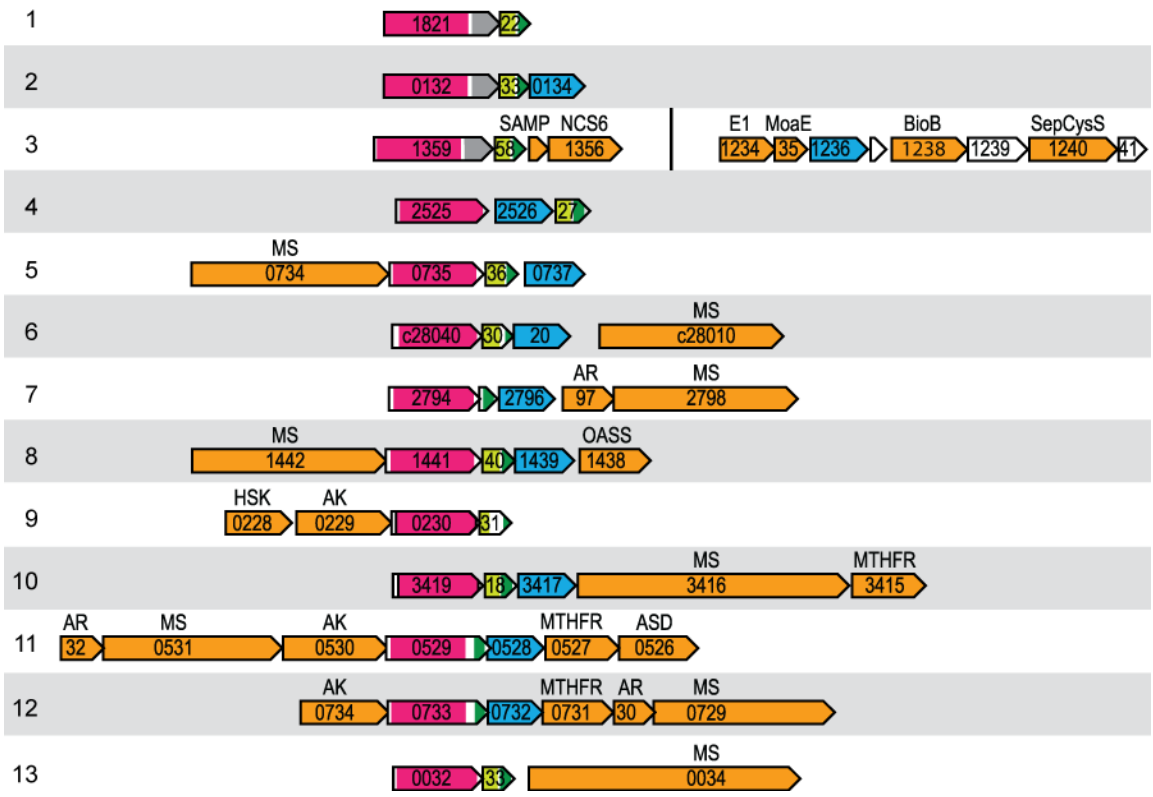


Figure 1.3 Genomic neighborhoods of the COG1900a, NIL-Fer and COG2122a proteins. All three proteins are observed to be encoded near other proteins with functions relevant to sulfur metabolism (orange). These include the small archaeal modifying protein and sulfur carrier (SAMP), the s^2U34 biosynthesis protein (Ncs6), the SAMP adenylating protein E1, (MoaE), (BioB), Sep-tRNA:Cys-tRNA synthase (SepCysS), methionine synthase (MS), the methionine synthase activating region (AR), O-acetylserine sulfhydrylase (OASS), homoserine kinase (HK), aspartate kinase (AK), methylenetetrahydrofolate reductase (MTHFR) and aspartatesemialdehyde dehydrogenase (ASD). Predicted open reading frames are labeled by NCBI locus number. COG1900a (pink) is fused to CBS domains (gray) in SepCysS-encoding genomes. Fusions of NIL (yellow-green) and Fer (green) domains are generally downstream from proteins containing COG1900a. However, in examples where NIL is absent, Fer is either fused to the C-terminus of COG1900a or occurs as a free-standing open reading frame downstream from COG1900a proteins. COG2122a proteins (blue) are often encoded downstream from NIL-Fer. Cartoons were constructed to scale with the assistance of the Microbial Genomic context Viewer (MGcV; http://mgcv.cmbi.ru.nl/) [69]. Data are shown for thirteen different genomes, which include *Methanosarcina acetivorans* (1), *Methanocella paludicola* (2), *Methanococcus maripaludis* S2 (3), *Desulfomonie tiedjei* (4), *Syntrophothermus lipocalidus* (5), *Acetobacterium woodii* (6), *Tepidanaerobacter acetatoxydans* (7), *Syntrophomonas wolfii* (8), *Thermosulfobium narugense* (9), *Odoribacter splanchnicus* (10), *Thermotoga thermarum* (11), *Thermosipho melanesiensis* (12) and *Sphaerochaeta globosa* (13).

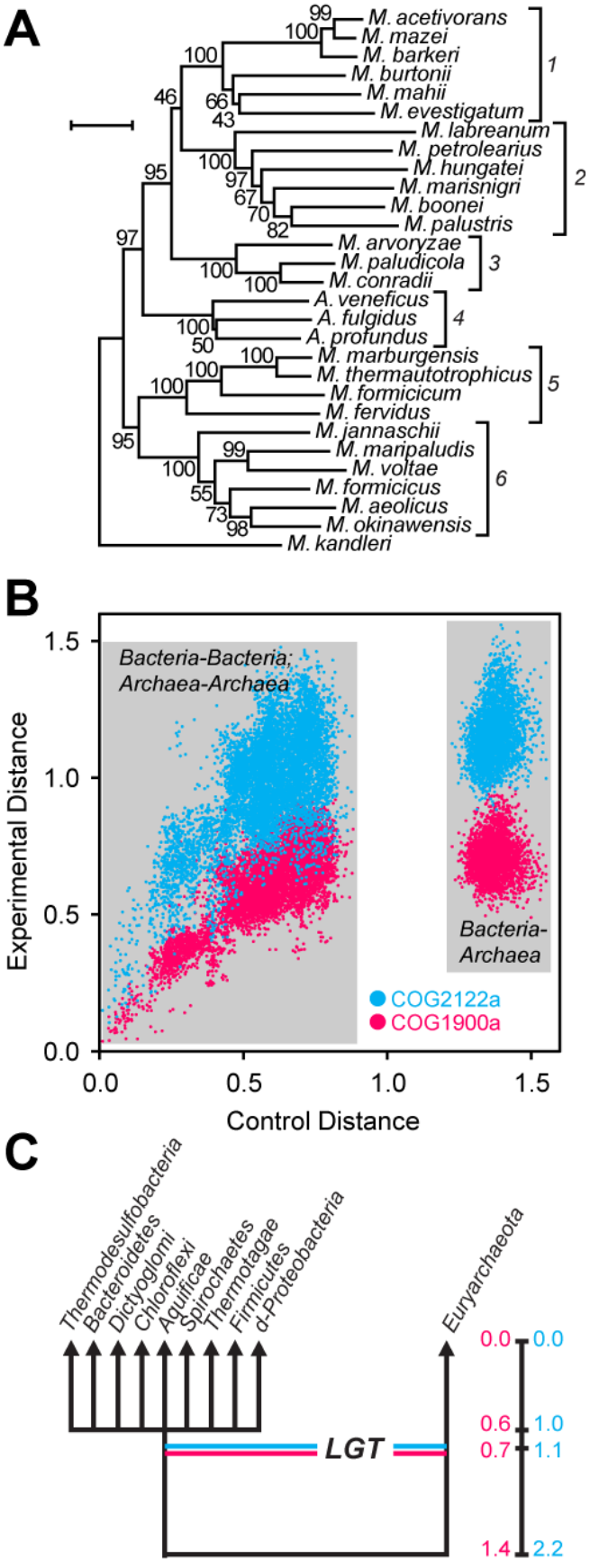
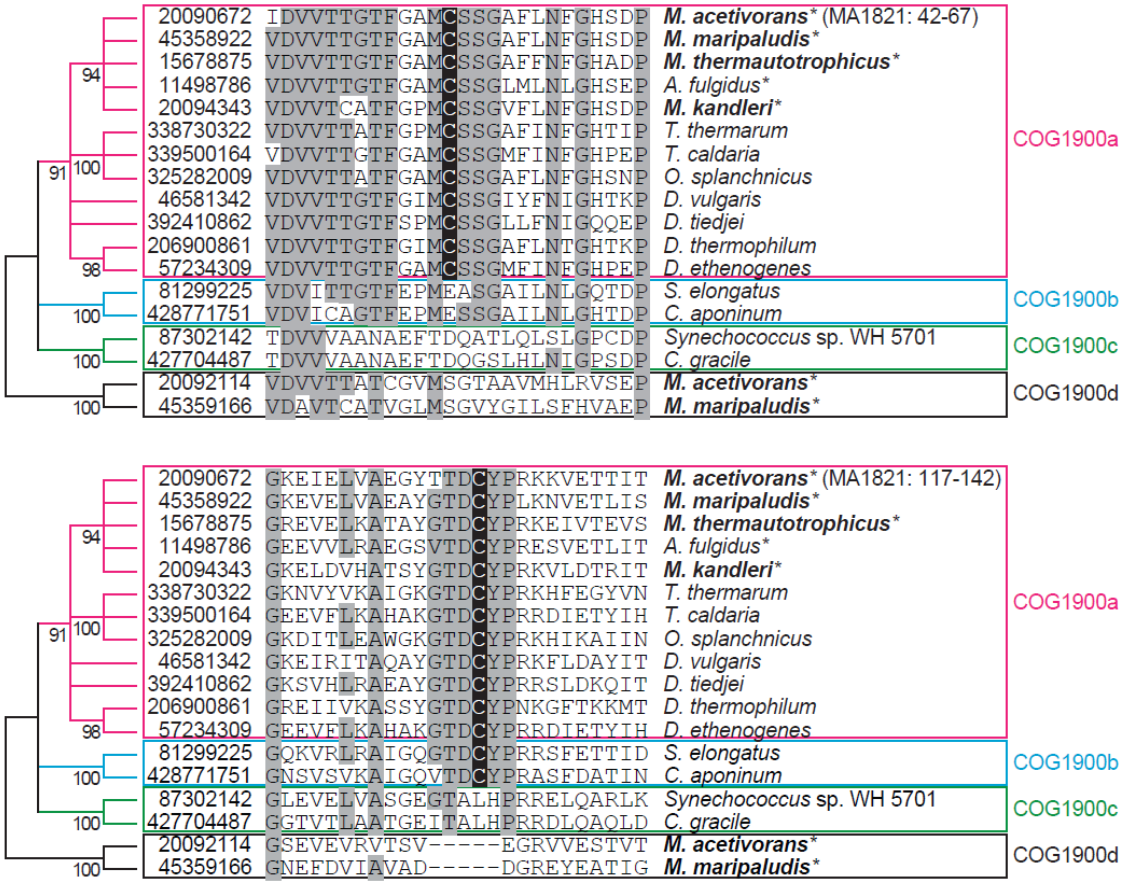


Figure 1.4 The evolutionary history of COG1900a, Nil-Fer and COG2122a. **(A)** Phylogenetic reconstruction of concatenated protein sequences for COG1900a, NIL-Fer and COG2122a proteins from selected SepCysS-encoding *Archaea*. Sequences were downloaded from GenBank and aligned with ClustalW, using default settings. A neighbor-joining tree was constructed in MEGA6 using the Poisson model, uniform rates and complete deletion of gaps. Individual nodes are labeled with the corresponding bootstrapping numbers as a percentage of 1000 replications. Taxonomic orders are grouped under brackets. These include *Methanosarcinales* (1), *Methano-microbiales* (2), *Methanocellales* (3), *Archaeoglobales* (4), *Methanobacteriales* (5) and *Methanococcales* (6). **(B)** Ratios of Evolutionary Distances (RED) analysis of COG1900a (pink) and COG2122a (blue). Evolutionary distances were calculated for inter- and intra-domain sequence pairs in MEGA6. The scatter plot relates experimental evolutionary distances for COG1900a and COG2122a to control evolutionary distances, which were calculated from an alignment of concatinated sequences for ribosomal proteins L2 and S5 and the protein translocase SecY. Sequences from 144 genomes encoding both COG1900a and COG2122a were identified and downloaded using BLASTp. All alignments were performed in MEGA6 using ClustalW with default settings. **(C)** The RED analysis detected an ancient lateral gene transfer event, which is depicted in the cartoon, and



possibly predates the divergence of bacterial phyla.

Figure 1.5 Sequence alignment of representatives from COG1900a, COG1900b, COG1900c and COG1900d revealing two conserved Cys residues (black) in COG1900a. Each conserved Cys is embedded in a highly conserved portion of the protein (gray). Sequences are identified by the corresponding NCBI GI number and the name of the host organism. Methanogens are shown in bold. *Archaea* are denoted with an asterisk. The phylogenetic tree representation at left is consistent with that depicted in Fig. 1.2.

CHAPTER 2

METHANOSARCINA ACETIVORANS: A MODEL ORGANISM FOR STUDYING SULFIDE ASSIMILATION

Metabolic diversity

The genome of *M. acetivorans* is the largest of all *Archaea*. At 5.8 Mb [70], it is more than twice the size of the average methanogen genome (2.5 Mb) and nearly five times larger than the smallest methanogen genome (1.2 Mb; *Methanothermus fervidus*). Genomic expansion has enabled *M. acetivorans* to use a variety of methanogenesis substrates, including methanol, acetate, carbon monoxide, methyl-amines and methyl-sulfides [71,72]. Moreover, it was this metabolic diversity which motivated the development of genetic technologies in *M. acetivorans*. Since there were multiple pathways for methanogenesis, individual pathways were dispensable, and deletion of the underlying genes was feasible.

Similarly, to study the genetic basis of sulfide assimilation, a metabolically diverse model organism is to be desired. *M. acetivorans* encodes what appears to be a full set of *E. coli*- and *S. cerevisiae*-type genes for the biosynthesis of cysteine (OASS: ma2720 and SAT: ma2721), cysteinyl-tRNA (CysRS: ma0749), homocysteine (OAHS: ma2715 and HSAT: ma2714) and persulfide (CD: ma0236, ma0808, ma1950, ma2718 and ma3264) **(Figs. 0.4 and 2.1)**. Thus, it was predicted that *M. acetivorans* would tolerate deletions of the genes encoding COG1900a-CBS (*ma1821*), NIL-Fer (*ma1822*), and COG2122a (*ma1715*) if they were indeed the hypothetical ancestral methanogen-type genes for sulfide assimilation. However, the presence of the *E. coli*- and *S. cerevisiae*-type genes also presented the possibility of the ancestral methanogen-type genes being displaced.

Genetic technologies have also been developed in *Methanococcus maripaludis*, which has a relatively small genome of 1.7 Mb. This hydrogenotrophic methanogen grows with a doubling time of two hours, and is often favored as a model organism over *M. acetivorans*, which has a minimum doubling time of six hours when methanol is used as the methanogenesis substrate. Despite this advantage, a lack of metabolic diversity makes *M. maripaludis* undesirable to study sulfide assimilation at the genetic level. Unlike *M. acetivorans*, *M. maripaludis* lacks known genes for the biosynthesis of free cysteine, homocysteine and persulfide. Moreover, genome-wide transposon mutagenesis has revealed that the genes encoding COG1900a-CBS (*mmp1359*), NIL-Fer (*mmp1358*), and COG2122a (*mmp1236*) are probably essential [73], suggesting that *M. maripaludis* may not be a suitable model organism to study these proteins of interest at the genetic level.

Genetic manipulation of *M. acetivorans*

Genetic technologies for *Methanosarcina* were pioneered in the laboratory of William Metcalf at the University of Illinois, Urbana-Champaign, and have been reviewed extensively [74]. Relevant to this work are the pseudo-wild type parent strain for genetic manipulation (WWM75) [75], liposome-mediated transformation [76], the plasmid for markerless genetic exchange (pMP44) [77], the shuttle vector (pWM321) [76] and the promoter for tetracycline-dependent gene expression (P_{mcrB}(tetO1)) [75]. *Methanosarcina* are strictly anaerobic, and their handling requires specialized equipment and techniques. Thus, to complete the experiments described in this text, a Metcalf-style anaerobic facility was constructed in the laboratory (see Appendices).

The WWM75 strain (*Δhpt*::P_{mcrB}-*tetR*-φC31-*int*-attB) was used as a pseudo-wild type for the experiments described in this text. The *hpt* (*ma0717*) gene, which encodes the purine-salvage enzyme hypoxanthine phophoribosyltransferase, is replaced with a gene encoding the tetracycline-dependent transcriptional repressor (*tetR*) under the control of the constitutively active promoter (P_{mcrB}) from *Methanosarcina barkeri* heterodisulfide reductase subunit B (*mcrB*) . Also present at the same locus is a gene encoding *Streptomyces* phage recombinase φC31 and the corresponding site-specific recombination sequence φC31-*int*-attB, which efficiently

undergoes recombination with the complimentary φC31-*int*-attP sequence. The φC31 machinery is intended to enable the chromosomal integration of genes to be expressed under the transcriptional control of P_{mcrB}(tetO1).

Markerless gene-deletion strains from this study were engineered using pMP44, which takes advantage of the endogenous DNA repair machinery to facilitate homologous recombination. pMP44 does not replicate in *Methanosarcina*, and carries the *pac-hpt* cassette, which confers puromycin resistance and 8-aza-2,6-diaminopurine (8-ADP) sensitivity. To delete genes with pMP44, an engineered fusion of the up- and down-stream sequences of DNA that flank the gene of interest on the chromosome are cloned into the plasmid, adjacent to the *pac-hpt* cassette. The modified plasmid is transformed into the desired parent strain, which is then plated on puromycin-containing media to select for chromosomal integration. Selected puromycin-resistant (Pur^R) strains are mero-diploid, possessing both wild-type and mutant genotypes at the locus of plasmid integration. Mero-diploidy is resolved by counter-selection with 8-ADP-containing media, which is toxic to strains carrying the *hpt* gene. Those surviving the counter-selection have undergone a second homologous recombination event, and have either the knock-out or wild type genotype at the locus of interest.

Like other methanogens, *M. acetivorans* exhibits polyploidy [78]. Depending on growth phase, *M. acetivorans* carries between seven and twenty-three copies of its genome. This must be taken into consideration when performing genetic manipulations because selections for puromycin and 8-ADP resistance will almost certainly yield genotypically heterogeneous strains, where only some copies of the genome bear the desired mutation. To generate strains with genotypic homogeneity, selected colonies must undergo multiple rounds of replication on the same selective medium.

The *E. coli-Methanosarcina* shuttle vector, pWM321 [76], was engineered from the naturally occurring plasmid pC2A that was present in the *M. acetivorans* when it was isolated from a California marine sediment [71]. pWM321 confers puromycin resistance in *Methanosarcina* species with puromycin N-acetyltransferase (*pac*) under the control of the constitutively active promoter, P_{mcrB}, from *Methanococcus voltae*.

The promoter P_{mcrB}(tetO1) enables tetracycline-dependent gene expression in *Methanosarcina* [75]. Along with φC31-*int*-attP, P_{mcrB}(tetO1) is present in the plasmid pJK031A to enable knock-in experiments. Although this plasmid was not employed for this purpose, its promoter was cloned into pWM321 to yield pBR031, a plasmid for multi-copy protein expression, which was used successfully to facilitate gene add-back experiments as well as preparative protein purification **(Figs. 2.2, 2.5A, 3.2 and 3.4** and **Table 2.1)**.

Sulfur sources of *M. acetivorans*

Conventional growth media for *Methanosarcina* include cysteine (3 mM) and sodium sulfide (0.4 mM) as potential sulfur sources [77,79]. A clear rationalization for the inclusion of both reagents is lacking; however, this cocktail may have its roots in classical techniques for the cultivation of anaerobes, which use cysteine as a reducing agent and sulfide as the deliberate sulfur source [80]. Regardless, to proceed with genetic analysis in *M. acetivorans*, a basic understanding of sulfur assimilation had to be gained. In particular, two types of information were desired: (1) a list of viable sulfur sources and (2) the minimum concentration of sulfide required for growth.

A sulfur source-less growth medium [81] (HS_{DTT}) was developed based on those used previously [77,79,82], which includes dithiothreitol (DTT) as a reducing agent. Individual sulfur sources were supplemented into HS_{DTT} after autoclaving, prior to growth. *M. acetivorans* was tested for growth on HS_{DTT} supplemented with a variety of sulfur sources (**Fig. 2.3**; data not shown). Sulfide, cysteine, methionine were capable of supporting growth, whereas cystathionine, sulfite, sulfate and DTT were not. Thiosulfate, polysulfide and elemental sulfur were not tested because each would be susceptible to reduction, and release of sulfide, in the presence of DTT.

The viability of sulfide as a sulfur source comes without surprise since *M. acetivorans* encodes known genes for the sulfide-dependent biosynthesis of cysteine and homocysteine, and like other methanogens, inhabits sulfidic environments [83] (**Figs. 0.4** and **2.3A-D**). However, the ability to use cysteine and methionine (which can be recycled to homocysteine via S-adenosylmethionine) was not predicted since *M. acetivorans* does not encode canonical genes for transsulfuration, and implies either that classical transsulfuration does indeed occur, or that

novel processes enable growth under these conditions (**Fig. 2.3E**). Assuming it can be imported into *M. acetivorans*, as it is in *M. maripaludis* [82], the inability of cystathionine to support growth rules out classical transsulfuration (**Fig. 2.3F**).

To determine its minimum required sulfide concentration, *M. acetivorans* was tested for growth with sulfide as the lone sulfur source at four different concentrations (**Fig. 2.3A-D**). When sulfide was present between 0.2 and 3.2 mM, *M. acetivorans* consistently grew with a doubling time of 8 hours to an optical density (A_{600}) of 1.2. When the sulfide was lowered to 50 µM, doubling time increased to ~20 hours and growth yield decreased to an optical density (A_{600}) of ~0.75, indicating that these conditions are sulfide-limiting.

Growth with methionine as the lone sulfur source is a surprising phenomenon, implying that sulfur from methionine can be used for the synthesis of cysteine and all other sulfur-containing molecules in the cell. Methionine-dependent cysteine biosynthesis might proceed through methanethiol produced by methionine-gamma lyase (ma2532). Methanethiol, which is a substrate of methanogenesis in *Methanosarcina*, would then be converted to sulfide and used by the endogenous sulfide assimilation pathways. Alternatively, methionine could be recycled to homocysteine (via S-adenosylmethionine; **Fig. 3.4B**), from which sulfide might be liberated by the reverse reaction of OAHS.

The ability of cysteine to support growth can also be rationalized through plausible biochemical transitions. Cysteine can supply the cell's demand for the biosynthesis of nucleotide modifications and enzymatic cofactors through the activity of cysteine desulfurase (CD), which is typically responsible for mobilizing sulfur for this purpose in other microorganisms. However, the route of sulfur from cysteine to homocysteine or methionine is less clear in the absence of transsulfuration enzymes. One possibility is that CD-derived persulfide is relayed to a novel enzyme for the biosynthesis of homocysteine. Alternatively, if persulfide is not a direct substrate of homocysteine biosynthesis, CD-derived persulfide might be reduced to sulfide and used by OAHS or another enzyme for homocysteine biosynthesis. Independent of CD, sulfide might be liberated from cysteine by the reverse reaction of OASS.

Novel genes for sulfide assimilation await discovery in *M. acetivorans*

A complete set of *E. coli*- and *S. cerevisiae*-type genes for sulfide assimilation can be detected in the genome of *M. acetivorans.* While this is a desirable feature (as explained earlier in this chapter), it suggests that the analogous ancestral methanogen-type genes are not necessarily present. To prove their existence in *M. acetivorans*, the genes for OASS (*ma2720*) and OAHS (*ma2715*) were deleted individually and in combination (**Table 2.1** and **Fig. 2.4**). The resulting strains were capable of growth with sulfide as the lone sulfur source, indicating that other sulfide-dependent pathways for cysteine and homocysteine biosynthesis remained intact (**Fig. 2.3-D**). Furthermore, the growth rates and yields of these deletion strains were nearly identical to *wild type* at sulfide concentrations as low as 50 µM, suggesting that OAHS and OASS do not permit the assimilation of sulfide at concentrations lower than the alternative pathways.

Strains lacking OASS were incapable of growth with cysteine as the lone sulfur source, suggesting that its protein product is essential to cysteine utilization (**Fig. 2.3D**). One plausible explanation for the observed phenotype is that cysteine-dependent growth is enabled by OASS catalyzing its reverse reaction, which would liberate sulfide from cysteine. Sulfide could then be recycled for the biosynthesis of homocysteine and other necessary sulfur-containing compounds. The single-deletion strain lacking OAHS was not compromised in its ability to use cysteine as a sulfur source, indicating that OAHS is not essential to cysteine-dependent growth.

SepCysS from *M. acetivorans* consumes persulfide for Cys-tRNA synthesis

Previous studies examining the catalytic mechanism of SepCysS support the idea that the enzyme consumes sulfane from relayed persulfide for Cys-tRNA synthesis [84–86]. Although a persulfide modification was observed to be present on SepCysS using mass spectrometry, its validity should be questioned because the modification was synthesized in *E. coli*, where the protein was recombinantly expressed [85].

The notion that SepCysS prefers relayed sulfane over free sulfide was reaffirmed in a series of preliminary experiments (**Figs. 2.5** and **3.6** and **Table 3.2**). Following recombinant expression in its native host, *M. acetivorans*, the purified SepCysS enzyme was observed to synthesize Cys-

tRNA from Sep-tRNA without an exogenously added sulfur source under single-turnover conditions **(Fig. 2.5A-B)**. Activity under these conditions suggests the SepCysS sample contained a co-purifying sulfur source. Since the enzyme was washed and dialyzed extensively (see methods), it appears likely that the co-purifying sulfur source was covalently bound, and may have in fact been persulfide. Consistent with earlier reports [84,86], SepCysS was not observed to efficiently catalyze multiple turnovers of Cys-tRNA synthesis with sodium sulfide present as a sulfur source, further supporting the notion that relayed sulfane is preferred **(Fig. 2.5C**).

The cysteine residues corresponding to C51, C54 and C260 in *M. acetivorans* SepCysS, were found to be essential in the *M. jannaschii* homolog [87]. Although persulfide was only detected (as a trisulfide) between C51 and C54 following expression of SepCysS in *E. coli*, the presence of C260 was found to significantly influence the total amount of persulfide present on SepCysS [85]. Following expression in *M. acetivorans*, persulfide modifications were repeatedly detected only at C260 (**Fig. 3.6B** and **Table 3.2**; see Chapter 3 for further elaboration on the LC-MS/MS approach). However, the negative results at C51 and C54 should not be interpreted as conclusive, since peptides carrying these residues were rarely detected in general. Nonetheless the presence of persulfide at C260, along with the sulfide-independent activity of SepCysS, are consistent in their support of a catalytic mechanism requiring persulfide.

Discussion

M. acetivorans is a suitable model organism for determining the genetic basis of methanogen-type sulfide assimilation. Unlike *M. maripaludis*, it possesses a complete set of aerobe-type genes for the biosynthesis of cysteine, Cys-tRNA, homocysteine and persulfide. Thus, *M. acetivorans* deletion strains lacking the genes encoding COG1900a-CBS, NIL-Fer and COG2122a, which are thought to have roles in methanogen-type sulfide assimilation, are likely to be viable. Furthermore, the existence of tools for the subtraction and addition of specific genes make genetic studies practicable.

The ability of the OASS-OAHS (Δ*ma2720*Δ*ma2715*) double-deletion strain to grow comparably to the pseudo-wild type strain, with sulfide present as the lone sulfur source, clearly

demonstrates that alternative pathways for sulfide assimilation remain intact (**Fig. 2.3**). Thus, to confirm the roles predicted of COG1900a-CBS, NIL-Fer and COG2122a, their genes will be deleted in combination with those encoding OASS and OAHS, since they might be functionally redundant.

In the event of homocysteine auxotrophy, which could arise in double-deletion strains lacking *oahs*, the second deleted gene of interest (encoding COG1900a, NIL-Fer or COG2122a) may function in either a general process for sulfide assimilation or one limited to homocysteine biosynthesis. To determine which is the case, the same gene of interest will be deleted in combination with *oass*. An absence of cysteine auxotrophy in the resulting *oass*-double-deletion strain would indicate that the function of the gene of interest is specific to homocysteine biosynthesis.

In the event of cysteine auxotrophy, which might be apparent in double-deletion strains lacking *oass*, inferring the function of the second deleted gene of interest (encoding COG1900a, NIL-Fer or COG2122a) is less straightforward. For example, if deletion of the gene of interest were to result in homocysteine auxotrophy when deleted in combination with *oahs*, then that gene of interest would be implicated in a general process in sulfide assimilation, consistent with the proposed sulfide-dependent persulfide biosynthesis reaction of the model for ancestral methanogen-type sulfur metabolism (**Figs. 0.3B** and **2.1**). However, if homocysteine auxotrophy were not to be observed, then the function of the gene of interest may be limited to sulfur relay to SepCysS for Cys-tRNA biosynthesis (and cysteine biosynthesis in general). Alternatively, the gene's function might be specific to downstream processes in sulfur delivery for the biosynthesis of cofactors or ribonucleotide modifications. These processes are paralleled by CD-mediated sulfur relay, and thus would be restored upon the addition of cysteine to the growth medium (**Fig. 2.1**). Of course, this possibility rests on the assumption that cysteine recycled from SepCysS-dependent Cys-tRNA synthesis would not be sufficient to sustain normal levels of CD activity.

Finally, the persulfide-dependent activity of SepCysS purified from *M. acetivorans* appears to validate previous observations based on the study of sepCysS homologs that were recombinantly expressed in *E. coli* [84,85]. Since SepCysS was used as the focus of the occurrence profiling

approach (see Chapter 1) that resulted in the identification of COG1900a-CBS, NIL-Fer and COG2122a as candidates for the missing ancestral proteins for persulfide and homocysteine biosynthesis, it is of significant consequence that SepCysS requires persulfide for activity in methanogens. Moreover, this finding supports the notion that the direct incorporation of sulfide, in spite of its environmental availability, is not unanimously employed by sulfur inserting enzymes in methanogens. *M. maripaludis* ThiI, which catalyzes sulfide-dependent s^4U8 biosynthesis [47], may therefore be an outlier to the prevailing persulfide-dependent strategy (**Figs. 0.3B-C**).

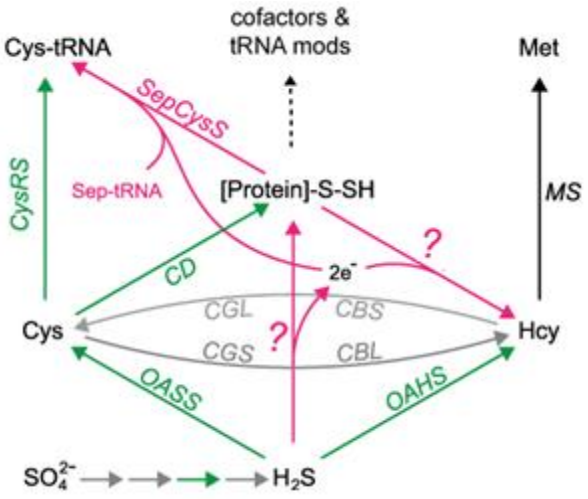


Figure 2.1 The model for sulfide assimilation in *Methanosarcina acetivorans*. Known and proposed reactions of the ancestral methanogen strategy (pink) are presumed to occur in all genomes also encoding SepCysS. These include Sep-tRNA:Cys-tRNA synthase (SepCysS) and proposed sulfide-dependent persulfide and homocysteine biosynthesis enzymes (question marks indicate that the existence of the pathways is not established). Known reactions for the *E. coli*- and *S. cerevisiae*-type strategy present in *M. acetivorans* (green), are often absent from other SepCysS-encoding genomes. These include cysteinyl-tRNA synthetase (CysRS), cysteine desulfurase (CD), *O*-acetylserine sulfhydrylase (OASS), and *O*-acetylhomoserine sulfhydrylase (OAHS). Known reactions for the *E. coli*- and *S. cerevisiae*-type strategy absent from *M. acetivorans* and some other SepCysS-encoding genomes, are indicated in gray. These include the transulfuration pathways for interconversion of Cys and Hcy, comprising the enzymes cystathionine gamma-synthase (CGS), cystathionine beta-lyase (CBL), cystathionine beta-synthase (CBS) and cystathionine gamma-lyase (CGL). Known enzymes common to both strategies are depicted in black (methionine synthase; MS). Protein-bound persulfide ([Protein]-SSH) likely provides sulfur for the biosyntheses of cofactors, tRNA modifications.

Table 2.1 *M. acetivorans* strains, plasmids and primers used in this study

strain	genotype	reference
WWM75	Wild type Δ*hpt*::(P*mcrB*tetR ΦC31 int attB)	75
BJR01	WWM75 Δ*ma1821-22*	19
BJR02	WWM75 Δ*ma1715*	0
BJR08	WWM75 Δ*ma2715*	19
BJR09	WWM75 Δ*ma2720*	19
BJR10	BJR01 Δ*ma2715*	19
BJR11	BJR09 Δ*ma1821-22*	19
BJR22	BJR09 Δ*ma2715*	19
BJR23	BJR22 Δ*ma1715*	0
BJR24	BJR09 Δ*ma1715*	0

plasmid	notes	reference
pMP44	Vector for markerless genetic exchange in *M. acetivorans*	77
pET22b	*E. coli* expression vector for C-term 6XHis fusion proteins	N
pWM321	Autonomously replicating *Methanosarcina-E. coli* shuttle vector	88
pJK031A	Carries the tetracycline-dependent promoter P*mcrB*(tetO1)	75
pBR004	*ma1821* cloned into pET22b via *NdeI*/*XhoI*.	19
pBR005	*ma1822* cloned into pET22b via *NdeI*/*XhoI*.	19
pBR006	*ma1715* cloned into pET22b via *NdeI*/*XhoI*	0
pBR016	Fused up- and down-stream *ma1821-1822* genomic flanking sequences, cloned into pMP44 via *SpeI*/*KpnI*	19
pBR023	Fused up- and down-stream *ma1715* genomic flanking sequences, cloned into pMP44 via *SpeI*/ *KpnI*	0
pBR031	P*mcrB*(tetO1) from pJK031A, cloned into pWM321 via *SpeI*/*SphI*.	19
pBR038	*ma1821-22* cloned into pET22b via *NdeI*/*XhoI*.	19
pBR041	Fused up- and down-stream *ma2720* genomic flanking sequences, cloned into pMP44 via *SpeI*/*KpnI*	19
pBR042	Fused up- and down-stream *ma2715* genomic flanking sequences, cloned into pMP44 via *SpeI*/*KpnI*	19
pBR050	*SphI* site removed by SDM (*ma1821*-S356S; AGC→AGT) of pBR004.	19
pBR051	*SphI* site removed by SDM (*ma1821*-H123H; CAT→CAC) of pBR050.	19
pBR052	*SphI* site removed by SDM (*ma1821*-S356S; AGC→AGT) of pBR038.	19
pBR053	*SphI* site removed by SDM (*ma1821*-H123H; CAT→CAC) of pBR052.	19
pBR054	*SphI*-site-free *ma1821* from pBR051 cloned into pBR031 via *SphI*/*ApaI*.	19
pBR055	*SphI*-site-free *ma1821-22* from pBR053 cloned into pBR031 via *SphI*/*ApaI*.	19
pBR056	*XhoI* site introduced into *ma1822* on pBR52 (H73L CAC→CTC; R74E AGG→GAG) by SDM	19
pBR058	MA1822-His6-coding region of pBR005 cloned into pBR031 via *SphI*/*ApaI*.	19
pBR059	*SphI*-site-free *ma1821* and $ma1822_{Fer}$ from pBR056 were fused by OEP and cloned into pBR031 via *SphI*/*ApaI*.	19
pBR060	MA1715-6XHis coding sequence from pBR06 inserted into pBR31 via SphI and BamHI	0
pBR062	$ma1822_{Fer}$ was removed from pBR056 via *XhoI*, yielding $ma1821\text{-}1822_{\Delta Fer}$.	19
pBR063	*SphI*-site-free $ma1821\text{-}22_{NIL}$ from pBR062 cloned into pBR031 via *SphI*/*ApaI*.	19
pBR070	OEP fusion of *SphI*-site-free *ma1821*from pBR051 and *ma1822* of pBR05, cloned into pBR031	19
pBR071	$ma1821_{C131A}$-*1822* was constructed from pBR55 via SDM (*ma1821*-C131A; TGC→GCG).	19
pBR073	$ma1821_{C54A}$-*1822* was constructed from pBR55 via SDM (*ma1821*-C54A; TGC→GCG).	19
pBR091	$ma1715_{C161A}$ was constructed from pBR60 via SDM (*ma1821*-C54A; TGT→GCT)	0
pBR998	An attempt to construct pBR053 by SDM resulted in the insertion of a 47 nt. sequence at the beginning of the CBS-coding region of *ma1821* in addition to the intended mutation. The insertion sequence shifted the reading frame such that an 11-aa sequence (364-LSTENSLSVCP-374) was inserted prior to a stop codon.	19
pBR999	$ma1821_{\Delta CBS}$-*22* from pBR998 was cloned into pBR998 via *SphI*/*ApaI*.	19

primer sequence	function
GAGGAAAACCATATGGTTGAAAAATCGGTTCATGAG	Forward primer; pBR004
GCTCAGTACCTCCTCTCGAGAAGCTTGC	Reverse primer; pBR004
GGAGGTACTGCATATGAAAATAAAGATCTGCATCCCC	Forward primer; pBR005
CTTTCATTCCTCGAGTTTTAAAGCGCCG	Reverse primer; pBR005
GACTCCGGGACTAGTTTTCGGAGCTTCAGGC	Upstream forward primer; OEP; pBR016
GCGCCGTGCGCCTCGGCTGTGACTACATTGACGC	Upstream reverse primer; OEP; pBR016
CACAGCCGAGGCGCACGGCGCTTTAAAACTCGGGG	Downstream forward primer; OEP; pBR016
CCACCCCGCCTGTTTGGTACCTCGGAATG	Downstream reverse primer; OEP; pBR016
GCGCTTTAAATACTAGTATGCTTCATTTATCGGAG	Forward primer; pBR031
CTACAGGACGTAGCATGCGAATTTCCTCC	Reverse primer; pBR031
GGGCCAGTCTTGCAATCGAACTAGTAAATAAAGG	Upstream forward primer; OEP; pBR041

CCTGAAGATATGCCCGGACAAGCGGTGTGTTTCCGATTGTCCATG	Upstream reverse primer; OEP; pBR041
CACACCGCTTGTCCGGGCATATCTTCAGGAGCGGCTACATGCGCAGC	Downstream forward primer; OEP; pBR041
GGGACAGGTACCCGTTGTCAATACAAGG	Downstream reverse primer; OEP; pBR041
GTTGGGGACTAGTTTTCCCTCTGCAGGGAAG	Upstream forward primer; OEP; pBR042
GAAAGCTGTTCATGGGCCGATCTTGCTCCCGTGACAGGGTCAGG	Upstream reverse primer; OEP; pBR042
CGGGAGCAAGATCGGCCCATGAACAGCTTTCAGCAGAAGAGCAG	Downstream forward primer; OEP; pBR042
GAGCCGATACGGTACCCTGGGACTGG	Downstream reverse primer; OEP; pBR042
CTCGACGTGCACGCAACTTCC	Forward SDM primer; pBR050 and 052
GGAAGTTGCGTGCACGTCGAG	Reverse SDM primer; pBR050 and 052
GGAAAATTCTTTGTCAGTATGCCCGTAGAAAAGC	Forward SDM primer; pBR051 and 053
GCTTTTCTACGGGCATACTGACAAAGAATTTTCC	Reverse SDM primer; pBR051 and 053
CTTTAAGAAGGAGATATACGCATGCTTGAAAAATCGG	Forward primer; pBR054, 055, 059 (OEP), 063, 070 (OEP), and 999
CCTTTCGGGCCCTGTTAGCAGC	Reverse primer; pBR055, 058, 059 (OEP), 063, 070 (OEP), and 999
CCGCCCCATTCTCGAGGACGAAGAAGAATGCGTGG	Forward primer for SDM to add an *XhoI* site from *ma1822* (H73L CAC→CTC; R74E AGG→GAG).
CCACGCATTCTTCTTCGTCCTCGAGAATGGGGCGG	Reverse primer for SDM to add an *XhoI* site from *ma1822* (H73L CAC→CTC; R74E AGG→GAG).
CTTTAAGAAGGAGATATACGCATGCAAATAAAGATCTGC	Forward primer; pBR058
GCAACTATTGCTCCCATGCTCAGTACCTCCTTGCGAGAAGCTTGC	Upstream reverse primer; OEP; pBR059
GGAGGTACTGAGCATGGGAGCAATAGTTGCAATTCTGGACCGCCCC	Downstream forward primer; OEP; pBR059
GGGACAGACGCGTATCCCAGGAAAG	Forward primer SDM: *pBR073*
CTTTCCTGGGATACGCGTCTGTCCC	Reverse primer SDM: *pBR073*
GGAGCCATGGCGTCTTCAGGTTTGATG	Forward primer SDM: *pBR071*
CATCAAACCTGAAGACGCCATGGCTCC	Reverse primer SDM: *pBR071*
GGATACAAGAGGAATACAGTATGGTGGAGCGCACG	Forward primer *ma1821* Southern blot probe
CCAGGCCGTCAGGATTCCTACCAGTTCC	Reverse primer *ma1821* Southern blot probe
CGGTTCCGATTTATCAGACCGCAGCTTATG	Forward primer *ma2715* Southern blot probe
CCTCATGAACTGCACCCTTGCTTTTGC	Reverse primer *ma2715* Southern blot probe
CCACATCACGAAGGGTCTTCATACGGA	Forward primer *ma2720* Southern blot probe
GCAAGAATGCCTTCTTCCCTTGCCAG	Reverse primer *ma2720* Southern blot probe
CGGGATCGAGATCTCAGTGGTGGTGGTGGTGGTGGTGCTCG	Upstream reverse primer; OEP; pBR070
CCACCACTGAGATCTCGATCCCGCGAAATTAATACGACTCAC	Downstream forward primer; OEP; pBR070
CCCGGAAACACTAGTATTGTGTCGCCAGAATAAGC	Upstream forward primer; OEP; pBR023
GTGATTTTCGGCACCTCCGCCAGCTTGTCTGGCGGCCCTTCAC	Upstream reverse primer; OEP; pBR023
CAAGCTGGCGGAGGTGCCGAAAATCACAAGAGCCGAGGTCAGGCATG	Downstream forward primer; OEP; pBR023
CATCGGGAGCTGGTACCGGGTTCTCCC	Downstream reverse primer; OEP; pBR023
GAGGCAAAGTACAGATTCTTCTGGACAGGTGG	Forward primer *ma1715* Southern blot probe
CCTATGTACTCCCCCTGAATCACAAGTGCCC	Reverse primer *ma1715* Southern blot probe
CCCCAAGTGAATCCATATGCCAGAACCC	Forward primer; pBR006
GCCGCCCCAACACTCGAGTTTTGTAATATACTCATGCC	Reverse primer; pBR006
GGAGATATACGCATGCCAGAACCCAC	Forward primer; pBR060
CCTTTCGGGCCCTGTTAGCAGC	Reverse primer; pBR060
CGATCACAGGAGTC~~GCT~~ACCTCGGCAGG	Forward SDM primer; pBR091
CCTGCCGAGGT~~AGC~~GACTCCTGTGATCG	Reverse SDM primer; pBR091

[0], This study
[N], Novagen
SDM, site-directed mutagenesis
OEP, overlap extension PCR

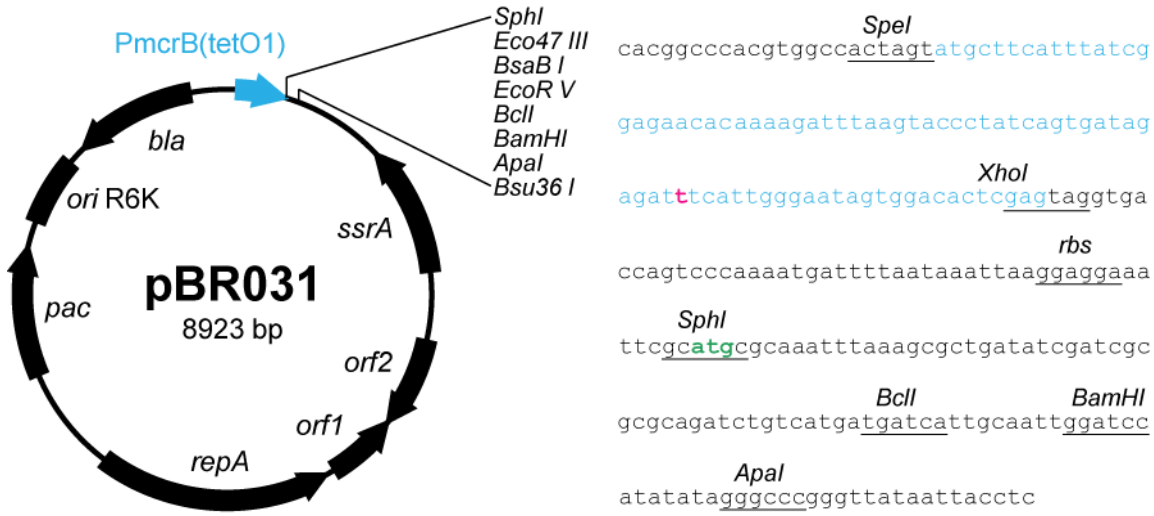


Fig. 2.2 Map of the pBR031 expression vector. The Tet-dependent *PmcrB*(tetO1) promoter (blue), transcriptional start site (pink) and translational start site (green) are indicated within the sequence of the regulatory region, shown below.

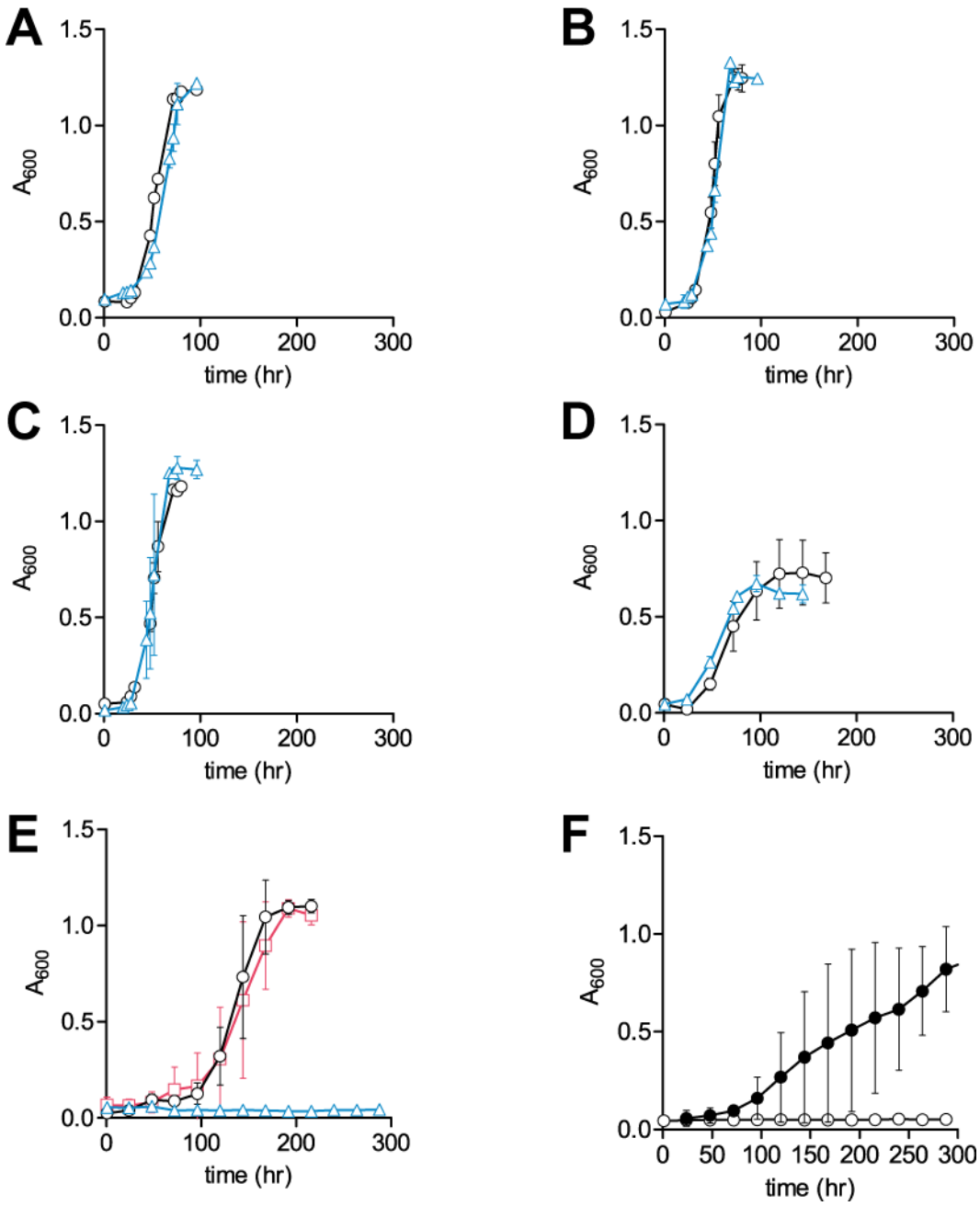


Figure 2.3 Growth of *Methanosarcina acetivorans* strains with various sulfur sources. WWM75 (wt; black circles), BJR08 (Δ*oahs*; pink squares) and BJR22 (Δ*oass*Δ*oahs*; blue triangles) were grown with HS_{DTT} media supplemented with 3.2 mM sodium sulfide **(A)**, 0.8 mM sodium sulfide **(B)**, 0.2 mM sodium sulfide **(C)**, 0.05 mM sodium sulfide **(D)**, 5 mM cysteine **(E)**, 5 mM methionine **(F**, closed circles**)** or 0.5 mM cystathionine **(F**, open circles**)**. Data points represent an average of 4-6 independent experiments. Error bars reflect standard deviations.

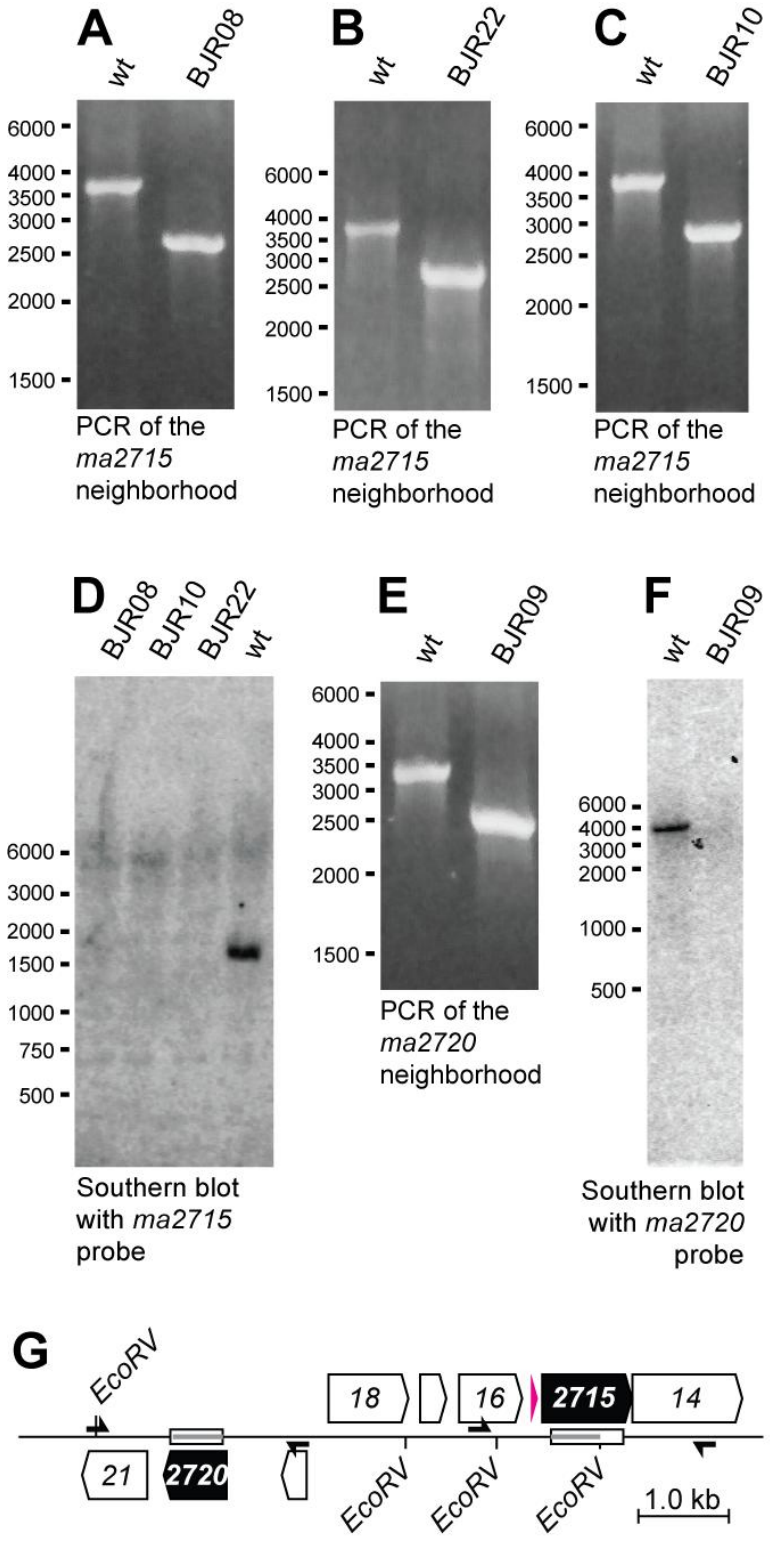


Figure 2.4 Markerless deletions of genes encoding OASS and OAHS. A knockout of *oass* (*ma2720*) was constructed in the pseudo-wild type strain (wt; WWM75) to yield BJR09 (Δ*oass*). Knockouts of *oahs* (*ma2715*) were constructed in WWM75, BJR01 (Δ*ma1821-22*) and BJR09 to yield BJR08 (Δ*oahs*), BJR10 (Δ*ma1821-22*Δ*oahs*) and BJR22 (Δ*oass*Δ*oahs*), respectively. PCR **(A, C** and **E)** and Southern blot **(B, D** and **F)** analyses are consistent with the desired manipulations **(G)**, which remove 771 and 1064 nt segments (white rectangles) from the chromosome for deletions of *ma2720* and *ma2715*, respectively. Primers for PCR analysis (black arrows) were used to amplify the genomic neighborhoods of *ma2720* and *ma2715*, which are predicted to measure 3132 and 3569 nt in the wild-type genotype, and 2361 and 2505 nt in the mutant genotypes, respectively. The probes used for Southern blot analysis (grey lines) measure 731 and 721 nt for *ma2720* and *ma2715*, respectively. They hybridize to the 4554 and 1522 nt fragments that lie between endogenous EcoRV restriction sites surrounding *ma2720* and *ma2715*, respectively.

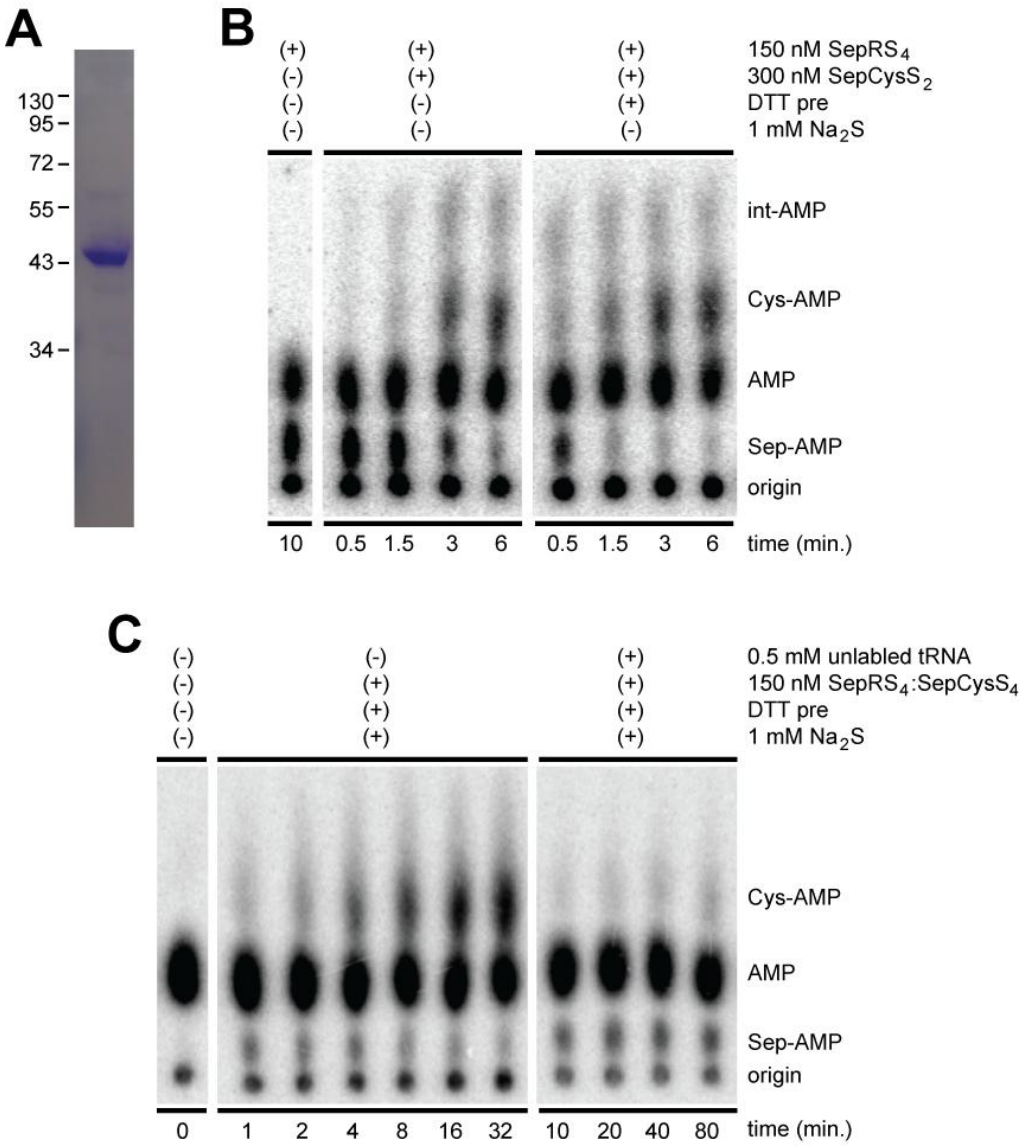


Figure 2.5 SepCysS catalyzes Cys-tRNA synthesis without an exogenously added sulfur source. **(A)** SDS PAGE of SepCysS-H_6 (43.5 kDa) that was purified anaerobically following expression with pBR031 in *M. acetivorans.* Approximately 0.5 mg of protein was recovered per liter culture. **(B-C)** SepCysS-dependent Cys-tRNA synthesis was monitored using thin layer chromatography (TLC), which resolved spots corresponding to uncharged tRNA (AMP), Sep-tRNA (Sep-AMP), an unknown intermediate (int-AMP) and Cys-tRNA (Cys-AMP). The intermediate was observed to decrease in concentration over time in other experiments (data not shown) as well as one published study [84]. In all experiments, < 25 nM radiolabeled [^{32}P-A76]-tRNA was present as a substrate. Prior to separation with TLC, quenched time points were digested with P1 nuclease to liberate 5'-[^{32}P]-AMP derivatives. Raw data TLC images were visualized by phosphorimaging. **(B)** Pre-incubation of SepCysS with 50 mM dithiothreitol (DTT pre) was observed to enhance activity under single turnover conditions in the absence of exogenously added sulfur sources, suggesting DTT may serve to reduce a disulfide, reported previously [85] that links two of the essential cysteine residues of SepCysS. **(C)** Supplementation with sodium sulfide could not support multiple turnovers of the SepCysS reaction.

CHAPTER 3

COG1900A AND ITS ASSOCIATED FERREDOXIN ARE ESSENTIAL TO A NOVEL ASPARTATE 4-SEMIALDEHYDE-DEPENDENT PATHWAY FOR HOMOCYSTEINE BIOSYNTHESIS

The COG1900a, NIL-Fer and COG2122a proteins are of unknown function, and were identified in all genomes also containing the tRNA-dependent pathway for cysteine biosynthesis. Their shared occurrence profile, evolutionary history and genomic context are each consistent with roles in sulfide assimilation (see Chapter 2).

The functions of COG1900a and and NIL-Fer were investigated using a series of gene-deletion and growth experiments conducted in *M. acetivorans*. The analysis demonstrated that both proteins are essential for homocysteine biosynthesis when a laterally inherited homocysteine biosynthesis gene, encoding O-acetylhomoserine sulfhydrylase (OAHS), was deleted. Metabolite labeling experiments revealed that COG1900a:NIL-Fer-dependent homocysteine biosynthesis uses aspartate-4-semialdehyde as a precursor.

Functional roles of MA1821 and MA1822

Within the genome of *M. acetivorans*, COG1900a and NIL-Fer are encoded by the adjacent genes *ma1821* and *ma1822*. Only two nucleotides separate the stop codon of *ma1821* and start codon of *ma1822*, suggesting that these genes could be expressed as a polycistronic message. To avoid potential issues stemming from interdependent gene expression, *ma1821* and *ma1822* were not deleted individually, but were instead excised together to yield the double-deletion strain (**Fig. 3.1A-B**, **E**). A triple-deletion strain was also constructed by further removing the *ma2715* gene, which encodes OAHS (**Fig. 2.3C-D**). The resulting strains—Δ*ma1821-22* and Δ*ma1821-*

*22*Δ*oahs*—were grown on a medium containing sulfide as the sole sulfur source. Under these conditions, Δ*ma1821-22* grew comparably to wild type, whereas the triple mutant Δ*ma1821-22*Δ*oahs* exhibited no growth (**Fig. 3.2A**). The robust growth of the Δ*ma1821-22* strain demonstrates that OAHS and OASS function efficiently in sulfide uptake.

Growth of the Δ*ma1821-22*Δ*oahs* triple mutant was recovered by addition of either methionine or homocysteine to the growth medium (**Fig. 3.2B**). This finding indicates that the strain is an homocysteine auxotroph, thereby implicating the *ma1821-22* in a novel, OAHS-independent biosynthetic route to homocysteine. To support this interpretation, the multi-copy, tetracycline-dependent expression plasmid pBR31 (see Chapter 2) was used to add back *ma1821-22* to the homocysteine auxotroph. The add-back strain grew comparably to wild type on sulfide-only medium supplemented with tetracycline (**Fig. 3.2C**), signifying the restoration of OAHS-independent homocysteine biosynthesis.

Unlike the ability of methionine or homocysteine to restore growth to the Δ*ma1821-22*Δ*oahs* strain, growth could not be recovered by adding cysteine to a growth medium containing sulfide (**Fig. 3.2B**). Therefore, it appears that the metabolic defect in the Δ*ma1821-22*Δ*oahs* strain is the inability of the cell to synthesize homocysteine from sulfide. The inability of cysteine to complement the deletions indicates that the transsulfuration pathway by which cysteine is converted to homocysteine is not present in *M. acetivorans*. This is consistent with the apparent absence of cystathione γ-synthase (CGS) or cystathionine β-lyase (CBL) in the genome, and with metabolite labeling experiments in *M. maripaludis* demonstrating that cysteine is not the sulfur source for methionine biosynthesis[20]. To further examine the possibility of transsulfuration despite the apparent absence of CGS and CBL, a medium containing sulfide and 3 mM cystathionine was tested for its ability to recover growth of the Δ*ma1821-22*Δ*oahs* strain (**Fig. 3.2B**). Like cysteine, cystathionine also fails to complement the deletions, providing further evidence that MA1821 and MA1822 are involved in direct sulfhydrylation of an amino acid precursor.

To examine whether *ma1821-22* may also play a role in cysteine biosynthesis, a triple deletion strain was constructed where *ma1821-22* was deleted in combination with *ma2720*, which encodes OASS (**Fig. 3.1C-E**). The resulting strain grows comparably to wild-type on

sulfide-only medium, and does not require the addition of exogenous cysteine (**Fig. 3.2A**). This finding suggests that MA1821 and MA1822 do not participate in the SepCysS-dependent pathway for cysteine biosynthesis.

It was considered that *ma1821-22* might confer the ability to assimilate sulfide at lower concentrations as an advantage over the alternative, laterally acquired pathways for sulfide assimilation. To determine if this were indeed the case, the Δ*ma1821-22* strain was grown with 3.2, 0.8, 0.2 or 0.05 mM sodium sulfide present as the lone sulfur source (**Fig. 3.3A-D**). The Δ*ma1821-22* strain grew comparably to the pseudo-wild type strain WWM75 under all conditions, suggesting that the *ma1821-22* genes do not function as an efficient sulfide sensor.

Also considered was the possibility that the *ma1821-22* genes may have a role in facilitating cysteine-dependent growth. Thus, the Δ*ma1821-22* strain was grown with 5 mM cysteine as the lone sulfur source. Under these conditions, the Δ*ma1821-22* strain grew comparably to wild type, and was thus not essential for cysteine-dependent growth, which is thought to proceed through OASS-and-cysteine-dependent sulfide biosynthesis (see Chapter 2).

Structure-function analysis of the MA1821 and MA1822 proteins

To identify protein structural elements essential to OAHS-independent homocysteine biosynthesis, a number of MA1821-22-coding region variants were constructed in pBR031 and transformed into the Δ*ma1821-22*Δ*oahs* strain (**Fig. 3.2C**). The resulting strains were then assayed for the capacity to biosynthesize homocysteine by monitoring growth in the presence of tetracycline, in a medium containing sodium sulfide as the sole sulfur source (**Fig. 3.2C**). As an initial experiment, the *ma1821* and *ma1822* genes were expressed in the homocysteine auxotroph triple chromosomal deletion strain. Neither gene alone was capable of restoring growth (doubling times (T_d) > 300 hr.), indicating that structural elements from both proteins are required for OAHS-independent homocysteine biosynthesis (**Fig. 3.2C**).

To explore the roles of particular subdomains, the following proteins were constructed (**Fig. 3.2C**): (i) full-length MA1821 together with the separately expressed NIL domain of MA1822, (ii) full-length MA1821 together with the separately expressed Fer domain of MA1822, (iii) the N-

terminal COG1900a portion of MA1821 together with separately expressed full-length MA1822. The truncated MA1822 protein consisting of the NIL domain alone was unable to restore a capacity for biosynthesis of homocysteine to the triple deletion mutant strain (**Fig. 3.2C**), indicating that the predicted iron-sulfur clusters in Fer are essential to catalytic function. However, the truncated MA1822 protein consisting of the Fer domain alone remained capable of restoring robust growth in the presence of sodium sulfide, suggesting that NIL does not confer an essential function in Homocysteine biosynthesis. Finally, the truncated MA1821 protein lacking both CBS domains also generated a strain capable of growth with sodium sulfide as the sole sulfur source, when coexpressed with full-length MA1822. However, in this case the doubling time of the strain was increased from 9 hr to 16 hr (**Fig. 3.2C**), indicating that these domains contribute to the efficiency of OAHS-independent Homocysteine biosynthesis.

The expendability of the CBS domains implies that the N-terminal COG1900a region of MA1821 has an essential role in homocysteine biosynthesis. To explore the roles of the conserved C54 and C131 residues in this domain, alanine mutations were separately introduced at each position. Surprisingly, the C131A mutant strain is capable of growth on sulfide-only medium, indicating that this cysteine residue is not functionally significant in this context. However, the strain bearing C54A was incapable of growth under these conditions (T_d > 300 hr), suggesting that the thiol of C54 has an essential role in OAHS-independent Homocysteine biosynthesis (**Fig. 3.2C**).

The precursor for *ma1821-22*-dependent homocysteine biosynthesis is aspartate-4-semialdehyde

Known pathways for *de novo* homocysteine biosynthesis proceed through acetyl-, succinyl- and phospho-homoserine (**Fig. 3.4**; see Introduction). However, since neither of the corresponding homoserine-activating enzymes are strictly conserved in genomes with COG1900a and NIL-Fer, alternative precursors were considered (**Fig. 0.4**). For its widespread use as an intermediate of threonine and lysine metabolism, aspartate-4-semialdehyde (Asa) was a particularly attractive candidate. Asa is synthesized from aspartate by aspartate kinase (AK)

and aspartate-semialdehyde dehydrogenase (ASD), which are both present in all genomes encoding COG1900a and NIL-Fer.

The possibility of Asa as a precursor to homocysteine was investigated in collaboration with the laboratory of Robert White at Virginia Polytechnic Institute. Cellular lysate of *Methanocaldococcus jannaschii*, a methanogen lacking OAHS, was assayed for its ability to synthesize homocysteine from exogenously added, isotopically labeled Asa, sulfide and water (**Table 3.1**). Homocysteine detection was carried out in the White lab by first labeling all thiol-containing molecules with monobromobimane and subsequently performing LC-ESI-MS for quantification in reference to a purified standard. Whereas only 40 µM homocysteine was present in lysate alone, upon the addition of Asa and sulfide, the homocysteine concentration increased five-fold. It was confirmed that the exogenously added compounds were responsible for the observed homocysteine production by repeating the experiment with isotopically labeled [3,3-D_2]Asa or $^{34}S^{2-}$. In both cases, homocysteine was almost completely enriched with the heavy isotopes. In a third labeling experiment, the lysate was supplemented with D_2O, which resulted in near-complete enrichment of deuterium. Collectively, these experiments indicate that homocysteine is synthesized from Asa, sulfide and water.

To determine if *ma1821-22* is responsible for mediating this reaction in *M. acetivorans*, an analogous set of experiments were performed with lysates from the Δ*oahs* and Δ*ma1821-22*Δ*oahs* strains. Whereas the Δ*oahs* lysate produced 28 µM homocysteine following the addition of Asa and sulfide, homocysteine was not detected in Δ*ma1821-22*Δ*oahs* lysate that was treated similarly. These data confirm that OAHS-independent homocysteine biosynthesis proceeds by way of Asa in *M. acetivorans*, and is mediated by the genes *ma1821-22*. However, they fall short of implicating the protein products (COG1900a and NIL-Fer) in catalysis (see Discussion).

Recombinant expression and purification of MA1821 and NIL-Fer

The *ma1821-22* genes are required for a novel, Asa-dependent homocysteine biosynthesis pathway. However, it is still uncertain whether the encoded proteins actually catalyze this reaction. Confirmation would require reconstitution of homocysteine biosynthesis *in vitro* with

purified enzyme. Attempts to express and purify recombinant MA1821, other COG1900 proteins and NIL-Fer in *E. coli* have been unsuccessful. Several IPTG-inducible expression plasmids were engineered; however, for each one, protein expression was either not detected or it yielded insoluble material (**Table A.7;** data not shown). Changes in IPTG concentration, temperature, ionic strength and detergents added during induced growth or cell lysis did not improve the results.

An alternative approach was pursued, in which recombinant expression was attempted in *M. acetivorans.* Successful protein expression was achieved using a pBR31-derived, tetracycline-dependent expression plasmid (pBR70) and the engineered homocysteine auxotroph (Δ*ma1821-22*Δ*oahs*), which was used as a host strain because it permitted higher levels of expression than the pseudo wild type strain WWM75 (**Fig. 3.2C** and **Table 2.1**). This plasmid, which restores homocysteine biosynthesis in the Δ*ma1821-22*Δ*oahs* strain, encodes the MA1821 (COG1900a-CBS) and MA1822 (NIL Fer) proteins—each modified with C-terminal poly-histidine tags. Approximately 200 µg of purified protein were recovered per liter of culture. Aerobic SDS PAGE analysis revealed that MA1821 migrates as a dimer when it is not treated with reducing agent, suggesting that pairs of MA1821 proteins form intermolecular disulfide bonds (**Fig. 3.5**).

Detection of persulfide modifications in MA1821

As a conserved strategy that spans all three domains of life, enzymes catalyzing sulfur incorporation for the biosynthesis of various metabolites are the recipients of the posttranslational modification, persulfide (**Fig. 0.3A-C**). The COG1900a protein family may catalyze sulfur insertion for homocysteine biosynthesis, and was therefore considered a candidate for persulfidation.

To identify persulfide modifications in COG1900a, the tryptic peptides of partially purified MA1821 (COG1900a-CBS) samples were subjected to LC-MS/MS analysis. Protein samples were processed in the presence or absence of dithiothreitol (DTT), which was administered to reduce persulfide back to unmodified cysteine. Thus, mass shifted cysteine residues could be positively identified as persulfide if they occurred primarily in the sample lacking DTT.

Persulfide was found to be present at C131 and C453, but not C54 (**Fig. 3.6** and **Table 3.2**). Whereas in the non-reduced sample 45-60% of all detected peptides containing C131 and C453 exhibited mass shifts indicative of persulfide, only 3-10% of the analogous peptides were similarly shifted in the reduced sample. Since the observed mass shifts are DTT-dependent, they are likely the result of persulfide, as opposed to some other modification of equivalent mass. Although persulfide was not detected at C54, the data are not persuasive in ruling this out as a possibility, since only two of the corresponding peptides were detected, and both occurred in the reduced sample. Nonetheless, it is of note that persulfide could be detected only at cysteine residues that are either nonconserved or nonessential to homocysteine biosynthesis, suggesting that these modifications may not be of functional significance.

Persulfide was also detected in twenty different contaminant proteins from the MA1821 sample and a similarly prepared SepCysS sample (**Table 3.2**; see Chapter 2). Of these, only one (MoaB) is predicted to function in a metabolic process requiring persulfide; however, even in this case, the persulfide occurs at a nonconserved cysteine residue. Thus, the lack of clear functional roles for the detected persulfides suggest they might be ornamental or artifactual.

Discussion

A series of genotype-dependent growth and metabolite labeling experiments have implicated the novel proteins MA1821 (COG1900a-CBS) and MA1822 (NIL-Fer) in a unique reductive condensation reaction of Asa and sulfide for the biosynthesis of homocysteine. Furthermore, since cysteine auxotrophy was not observed in the Δ*oass*Δ*ma1821-22* strain, MA1821 and MA1822 are probably not responsible for a more general process in sulfide assimilation, such as persulfide synthesis. Although the possibility of these proteins functioning in some regulatory capacity cannot be ruled out, this would be difficult to rationalize given that C54 and Fer are essential functional components of MA1821 and MA1822.

Structure-function analysis of MA1821 and MA1822 identified COG1900a and Fer as the essential domains for homocysteine biosynthesis. For a role in catalysis, the requirement of Fer is justified since the reaction consumes at least two electrons to reduce the aldehyde. Moreover, the

dispensability of the CBS and NIL domains is consistent with their absence from some bacterial homologs (**Fig. 1.3**). Thus, the functional unit, which is putatively involved in catalysis will be referred to as COG1900a:Fer (**Fig. 3.4A**).

Of the two conserved cysteine residues present in COG1900a, only C54 was found to be essential. This finding is consistent with the orthological assignments that distinguish the subgroups of COG1900, since COG1900b-d do not conserve C54, the notion that they are of diverged function is supported (**Fig. 1.5**). However, the identification of only one essential cysteine may indicate a departure from the only known strategy of ancestral-methanogen enzymes for sulfide insertion. Both SepCysS and ThiI (from *M. maripaludis*) have pairs of essential cysteine residues that engage in persulfide-disulfide exchange to release sulfide for catalysis [47,85] (**Fig. 0.3C**). Thus, the presence of only a single conserved cysteine suggests COG1900a may employ a different, disulfide-independent strategy for manipulating sulfur. Alternatively, COG1900a could still engage persulfide-disulfide exchange if C54 were to form an intermolecular disulfide. The DTT-dependent dimerization of MA1821 supports this idea (**Fig. 3.5**); however, it has yet to be determined if C54 is responsible for mediating the implicit disulfide bond.

The sulfur source for COG1900a:Fer-dependent homocysteine biosynthesis awaits identification. Although isotopically labeled sulfide was efficiently incorporated into homocysteine in whole-cell extracts of *M. acetivorans* and *M. jannaschii*, it should not be inferred that sulfide is used directly by COG1900a:Fer. Instead, sulfide might be oxidized to sulfane (within the context of a persulfide modification), relayed as such to COG1900a:Fer and subsequently reduced back to sulfide for catalysis. If these sulfide-sulfane transitions were to be mediated by small-molecule redox cofactors, such as nicotinamides, coenzyme M, coenzyme B, flavins or the methanogen-type flavin analog F_{420}, then they could be suppressed following dialysis, which would serve to remove the requisite electron donors and acceptors. Since the extracts used in metabolite labeling experiments were not dialyzed, and thus contained dialyzable redox cofactors, the exogenously added sulfide could have proceeded through sulfane prior to insertion into homocysteine. Unfortunately, additional dialysis-dependent experiments could not be performed

to rule out the use of sulfane as an intermediate. Since homocysteine synthesis from Asa also requires a two-electron transfer—altogether separate from persulfide relay—to reduce the aldehyde of Asa, dialysis might have the undesired effect of suppressing the aldehyde's reduction. Thus, any observed decrease in homocysteine formation that might follow dialysis would not necessarily signify that sulfane is used as an intermediate.

The possible use of persulfide by COG1900a:Fer was addressed further through LC-MS/MS analysis of tryptic peptides derived from MA1821. A similar approach has been used to identify persulfide on the methanogen proteins SepCysS, ThiI and Ncs6 [44,46,47]. In those cases, persulfide and disulfide were detected at conserved cysteine residues. While these findings are certainly provocative, their underlying methodologies may be flawed for two reasons. (i) The protein samples used for peptide analysis were obtained by recombinant expression in *E. coli*, and thus may not accurately reflect persulfidation as it occurs in methanogen cells. (ii) Measures were not taken to ensure that mass shifts interpreted as persulfide were actually due to the presence of an extra sulfur atom, as opposed to some other possible modification. The peptide analysis presented here avoid these issues, since the analyzed protein samples were obtained following expression in *M. acetivorans*, and additional negative-control experiments were conducted in which samples were treated with DTT to cleave persulfides.

Persulfide was not identified at C54. However, this negative finding is hardly conclusive, since peptides bearing C54 were detected only twice in general. The reduced detection of peptides with C54 may be the result of their size. Following digestion with trypsin, the expected C54 peptide is 30 amino acids in length. Future attempts to detect persulfide at C54 should invoke the use of pepsin in place of trypsin to decrease peptide length. The predicted peptic peptide would measure 8 amino acids in length, and thus might enable detection of persulfide, and possibly C54-C54 intermolecular disulfides.

Peptides containing C131 and C453 were observed 25 and 55 times, respectively. Furthermore, they were persulfided roughly half of the time in samples that were not treated with DTT, but were seldom persulfided in reduced controls. Thus, the occurrence of persulfide at these sites is strongly supported by the data. However, since C131 and C453 are not essential to

COG1900a:Fer-dependent homocysteine biosynthesis, their persulfides appear to be inconsequential for homocysteine biosynthesis. Similarly, persulfide was observed in twenty contaminating *M. acetivorans* proteins, all lacking clear roles in sulfur metabolism. This is inconsistent with the current paradigm for persulfide relay, which contends that persulfide modifications are relayed in a specific manner between conserved cysteines prior to sulfide incorporation [23]. Conversely, the repeated observation of function-less persulfide instead suggests that persulfide relay in methanogens may be governed by a stochastic process.

Finally, whereas all other known pathways for *de novo* homocysteine biosynthesis consume activated homoserine, the discovered pathway proceeds from Asa and consumes a pair of electrons (**Fig. 3.4A-B**). Although unique for homocysteine biosynthesis, this aldehyde-to-thiol transformation is analogous to the final step in coenzyme M biosynthesis [89] (**Fig. 3.4C**). The enzyme catalyzing sulfur insertion into coenzyme M is unknown, and may thus be accounted for by the COG1900a paralog, COG1900d, which occurs almost exclusively in methanogen genomes [28]. This role for COG1900d could be confirmed if a deletion strain lacking COG1900d were to exhibit coenzyme M auxotrophy (this has not been attempted). Alternatively, the COG1900d protein could be purified and used to catalyze coenzyme M biosynthesis *in vitro*. However, previous attempts in recombinant expression of COG1900d in *E. coli* have yielded insoluble protein [28] (**Table A.7**).

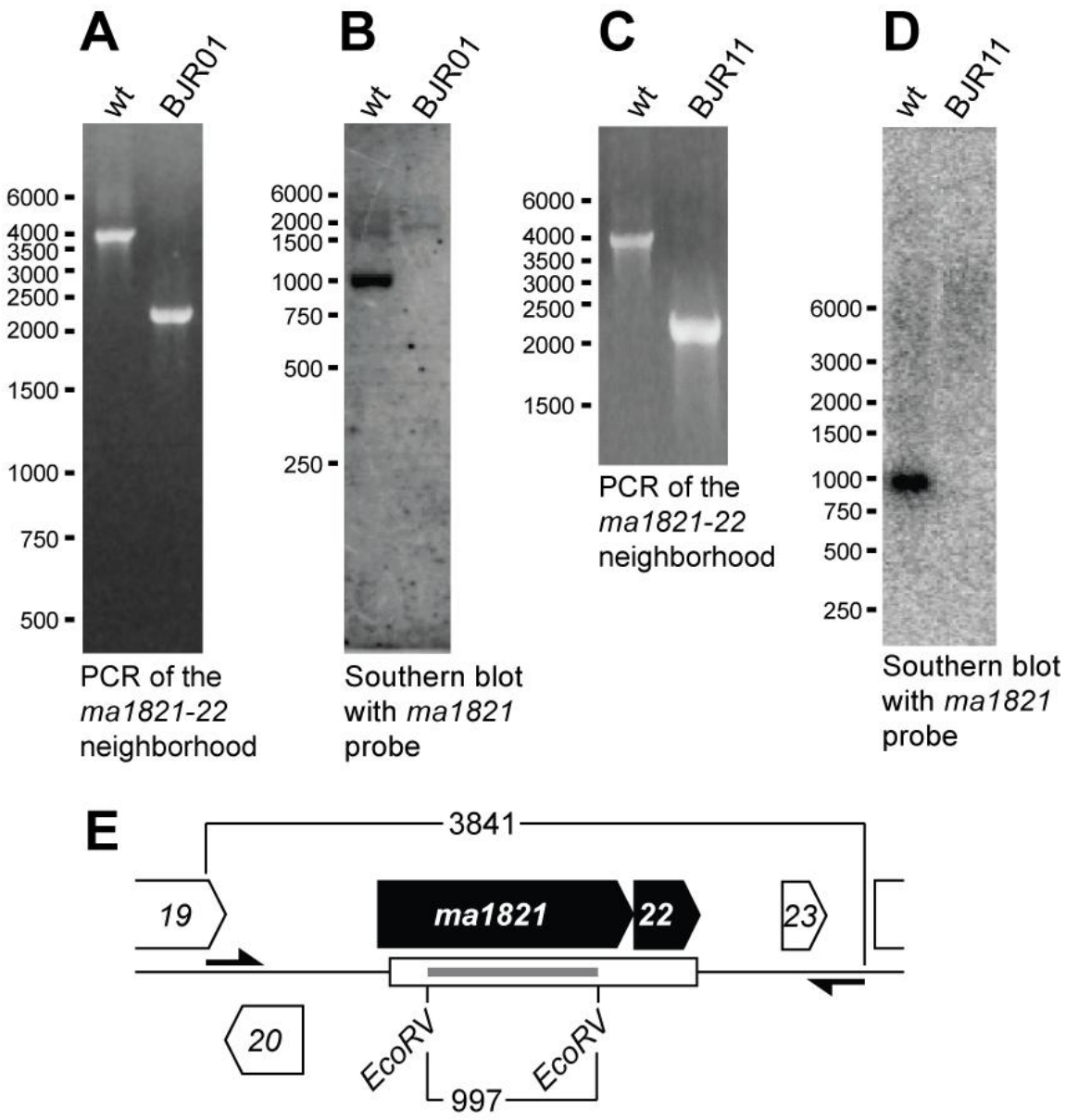


Figure 3.1 Markerless deletions of the genes encoding COG1900a-CBS and NIL-Fer (*ma1821-22*) were constructed in the WWM75 **(**wt; **A-B)** and BJR09 **(**Δ*oass*; **C-D)** strains, yielding BJR01 (Δ*ma1821-22*) and BJR11 (Δ*oass*Δ*ma1821-22*), respectively. PCR **(A, C)** and Southern blot (**B, D**) analyses are consistent with the desired manipulation **(E)**, which removes a 1790 nt segment (white rectangle) of *ma1821-22* from the chromosome. Primers for PCR analysis (black arrows) were used to amplify the genomic neighborhood of *ma1821-22*, which is predicted to measure 3841 nt in the wild-type strain and 2051 nt in the Δ*ma1821-22* strains. The probe used for Southern blot analysis (grey line) measures 990 nt and hybridizes to the 997 nt fragment that lies between two endogenous EcoRV restriction sites within *ma1821*.

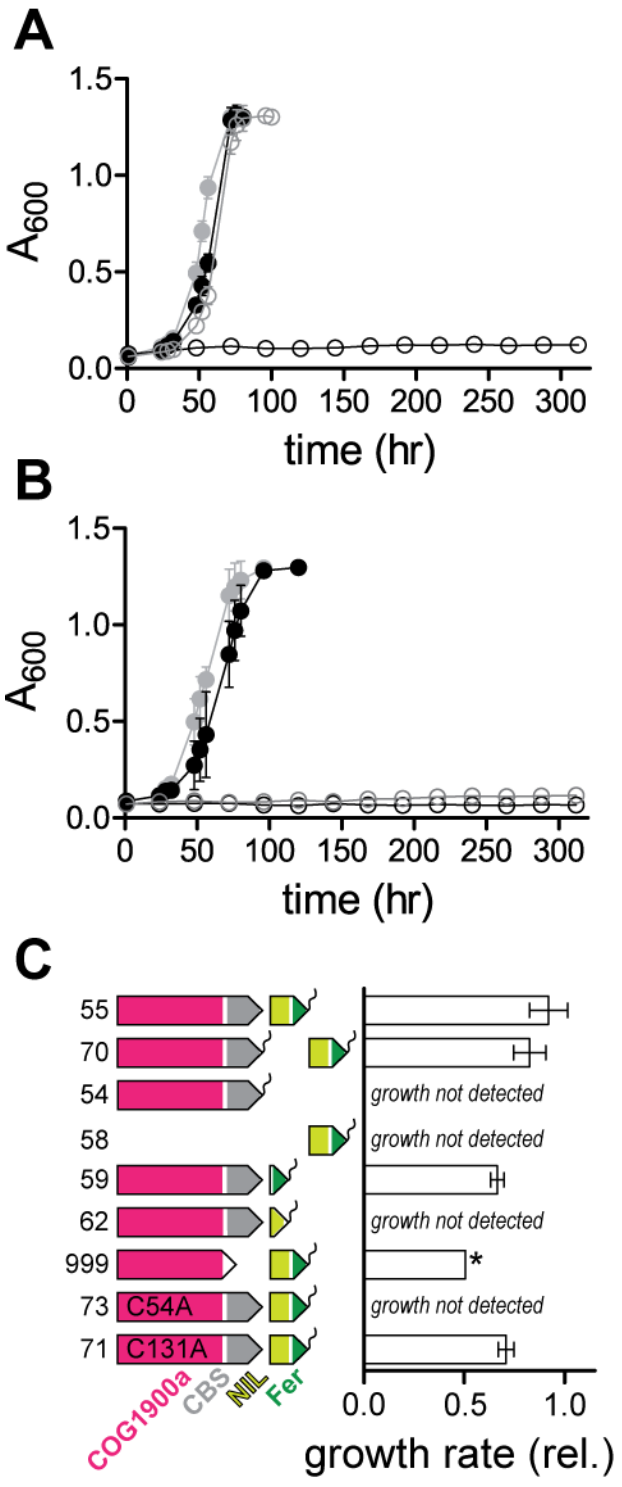


Figure 3.2 Genes encoding COG1900a-CBS and NIL-Fer are essential to OAHS-independent homocysteine biosynthesis **(A)** Strains BJR01 (Δ*ma1821-22*; gray filled circles), BJR10 (Δ*ma1821-22Δoahs*; open black circles), BJR10+pBR055 (Δ*ma1821-22* with *ma1821-22* added back; filled black circles; construct 55 from panel C), and BJR11 (Δ*oass*Δ*ma1821-22;* open gray circles) were grown on a medium containing sulfide as the sole sulfur source (HS_{DTT} medium supplemented with 0.80 mM sodium sulfide; see Methods). Only marginal growth of BJR10 with the *ma1821-22* genes added back in plasmid pBR031 is observed when the tetracycline inducer is omitted (data not shown). **(B)** Growth of BJR10 on a medium containing 0.8 mM sodium sulfide supplemented with cysteine (3 mM; black open circles), methionine (3 mM; filled gray circles) cystathionine (3 mM; gray open circles) or homocysteine (3 mM; filled black circles). **(C)** Add-back experiments for identification of the essential structural elements of *ma1821-22*. The identities of the color-coded domains are indicated at the bottom left. Variants were added back to the BJR10 strain using the shuttle vector pBR031 and subsequently grown on a medium containing 0.8 mM sodium sulfide as the sole sulfur source. Growths were conducted in the presence of tetracycline (50 µg/ml) and puromycin (2 µg/ml). The curved lines at the C-termini of the constructs represent poly-histidine tags. MA1821 and MA1822 are separately expressed from independent ribosomal binding sites. The intergenic spacing of two nucleotides in the *M. acetivorans* chromosome is preserved in constructs 55, 59, 62, 73 and 71. Construct 999 (CBS domain deletion) possesses a larger interdomain spacer, as does construct 70 containing a second His tag added to the C-terminus of MA1821. The corresponding plasmids are numbered at left as in Table 2.1. Statistical analysis was not performed for the strain indicated by the asterisk.

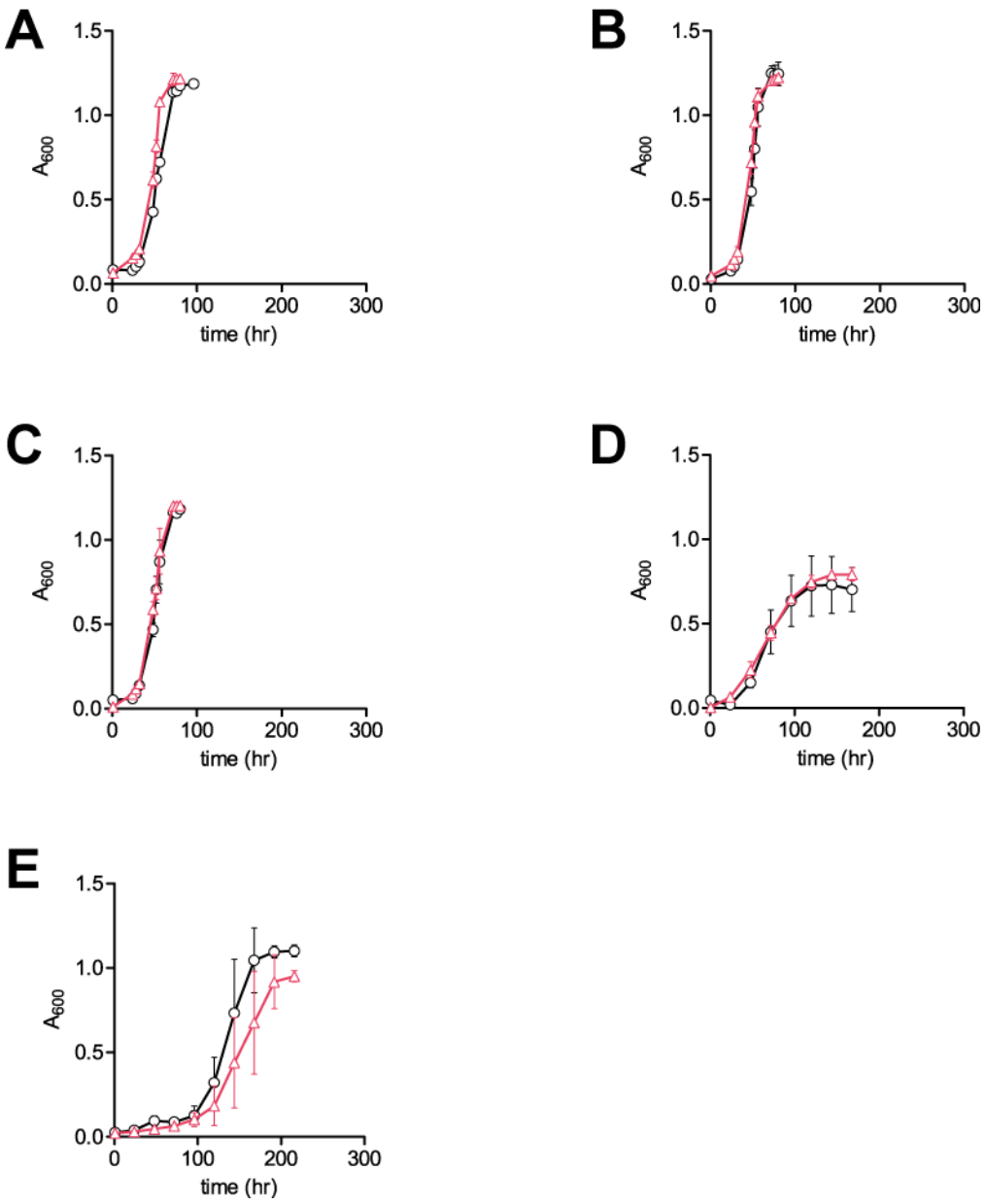


Figure 3.3 Genes encoding COG1900a and NIL-Fer are not essential to growth at lower at lower sulfide concentrations or with cysteine. WWM75 (wt; black circles) and BJR01 (Δ*ma1821-22*; pink triangles) were grown with HS_{DTT} media supplemented with **(A)** 3.2 mM sodium sulfide **(B)**, 0.8 mM sodium sulfide **(C)**, 0.2 mM sodium sulfide **(D)**, 0.05 mM sodium sulfide or **(E)** 5 mM cysteine. Data points represent an average of 4-6 independent experiments. Error bars reflect standard deviations.

Table 3.1 Incorporation of precursors into homocysteine in *M. jannaschii* and *M. acetivorans* cell extracts

experiment	cell extract	HS^-	Asa	HA	isotopic label	isotopic incorporation	homocysteine (μM)
1	Mj (wt)	-	-	-	-	na	40
2	-	+	+	-	-	na	< 0.7
3	Mj (wt)	+	+	-	18% ^{2}H	15% ^{2}H	200
4	Mj (wt)	+	+	-	$H^{34}S^-$	89% ^{34}S	340
5	Mj (wt)	+	+	-	[3,3-$^{2}H_2$]Asa	68% ^{2}H	130
6	Ma (wt)	+	+	-	$H^{34}S^-$	92% ^{34}S	22
7	Ma (Δ*ma1821-22*Δ*oahs*)	+	+	-	$H^{34}S^-$	na	nd
8	Ma (Δ*oahs*)	+	+	-	$H^{34}S^-$	67% ^{34}S	28
9	Ma (Δ*ma1821-22*)	+	-	+	$H^{34}S^-$	99% ^{34}S	1000
10	Ma (Δ*oahs*)	+	-	+	$H^{34}S^-$	48% ^{34}S	12
11	Ma (Δ*ma1821-22*)	+	+	-	$H^{34}S^-$	74% ^{34}S	13

na, not applicable.
nd, too low to measure accurately.
Mj, *Methanocaldococcus jannaschii*
Ma, *Methanosarcina acetivorans*

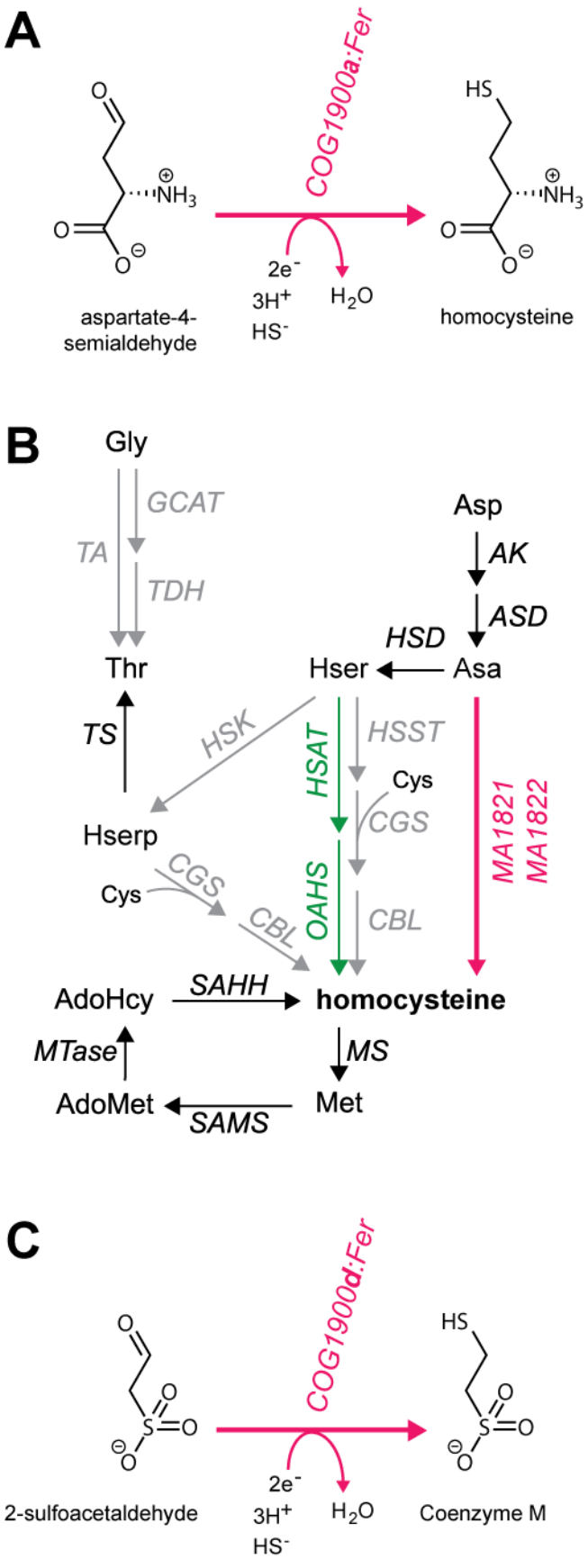


Figure 3.4 Homocysteine and coenzyme M biosynthesis from the perspective of carbon. **(A)** The COG1900a family of proteins, and its associated ferredoxin, were found to synthesize homocysteine (Hcy) from aspartate-4-semialdehyde (Asa). The ferredoxin is proposed to mediate the electron transfer required to reduce the aldehyde. **(B)** The pathways for *de novo* homocysteine biosynthesis and recycling include the enzymes aspartate kinase (AK), aspartate semialdehyde dehydrogenase (ASD), homoserine dehydrogenase (HSD), homoserine succinyltransferase (HSST), homoserine kinase (HSK), homoserine acetyltransferase (HSAT), *S*-adenosylmethionine synthetase (SAMS), AdoMet-dependent methyltransferases (MTase), S-adenosylhomocysteine hydrolase (SAHH), threonine aldolase (TA), glycine C-acetyltransferase (GCAT), threonine dehydrogenase (TDH) and threonine synthase (TS), and the intermediates glycine (Gly), threonine (Thr), O-phosphohomoserine (Hserp), cysteine (Cys), homoserine (Hser), aspartate-4-semialdehyde (Asa), aspartate (Asp), methionine (Met), S-adenosylmethionine (AdoMet), and S-adenosylhomocysteine (AdoHcy). An HSK homolog cannot be detected in the *M. acetivorans* genome, yet the functioning of TS with a phosphohomoserine substrate may represent the only route to threonine in this organism. Therefore, a homoserine kinase may be present in some undetected form. Enzymes that are universally conserved (black), present typically in aerobes but absent from *M. acetivorans* (grey), present

typically in aerobes but present in *M. acetivorans* (green) and of the ancestral methanogen strategy (fuschia) are distinguished by color. **(C)** The enzyme synthesizing methanogenesis-specific coenzyme M from 2-sulfoacetaldehyde has not been identified. In analogy with COG1900a:Fer-catalyzed homocysteine biosynthesis, COG1900d:Fer, which is conserved in methanogens, and extremely rare outside of methanogens, is proposed to catalyze coenzyme M biosynthesis from 2-sulfoacetaldehyde.

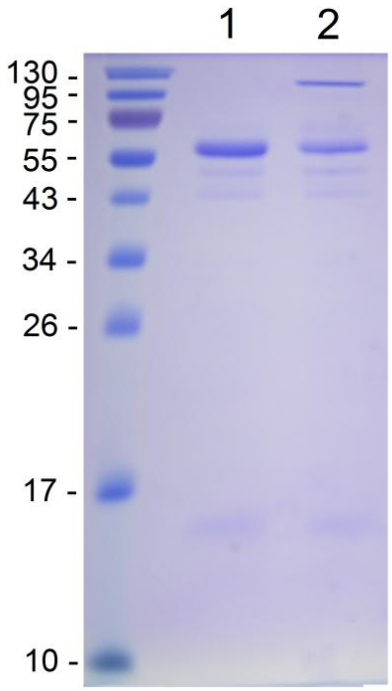


Figure 3.5 COG1900a-CBS exhibits DTT-dependent dimerization. SDS PAGE of MA1821-H_8 and MA1822-H_6 purified by recombinant expression in *M. acetivorans*. The engineered homocysteine auxotroph (Δ*ma1821-22*Δ*oahs*) was transformed with pBR070 (Fig. 3.2 and Table 2.1). The resulting transformant was grown in HS_{DTT} supplemented with 50 µg/ml tetracycline and 800 µM sodium sulfide. The proteins were purified anaerobically by affinity chromatography using an immobilized-nickel resin (see Methods). Approximately 200 µg protein was recovered per liter of culture. Protein samples were analyzed by aerobically run SDS PAGE with **(1)** or without **(2)** dithiothreitol present as a reducing agent in the loading dye, which also contained SDS as a denaturant. The predicted molecular weights of MA1821-H_8 and MA1822-H_6 are 55.4 kDa and 15.3 kDa, respectively.

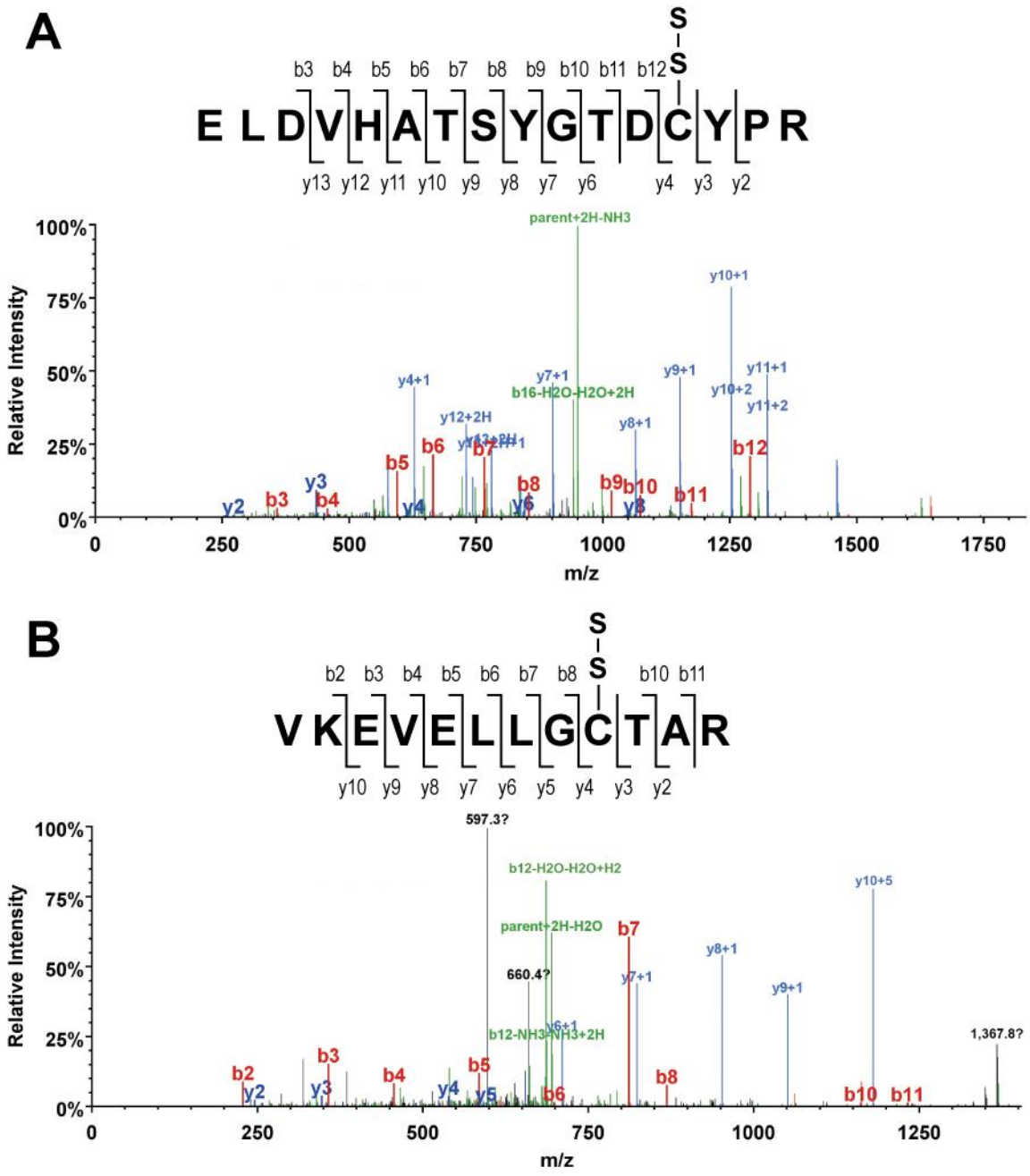


Figure 3.6 Example MS/MS fragmentation spectra detecting persulfide on MA1821 and SepCysS. Spectra for the tryptic peptides $_{119}$ELDVHATSTGTDCTPR$_{134}$ **(**MA1821; **A)** and $_{252}$VKEVELLGCTAR$_{263}$ **(**SepCysS; **B)** are both consistent with +89 mass shifts at cysteine, which is to be expected for persulfide (+32) that is alkylated with iodoacetate (+57). The detected ions are labeled on the spectrum. The parent ions were detected with errors of -190 ppm **(A)** and -220 ppm **(B)**.

Table 3.2 Detected peptides with persulfide from *Methanosarcina acetivorans*

protein	annotation	residue	peptides (persulfided/total)	
			+ DTT	- DTT
MA0022[1]	iron-dependent repressor	C155	0	1
MA0022[1,2]	iron-dependent repressor	C53	0	3
MA0086[1]	Hsp60 chaperonin	C289	0	1
MA0187[2]	MoaB	C153	0	1
MA0503[1]	unknown function	C59[?]	0	1
MA0532[2]	dimethylamine methyltransferase	C99	1	2
MA0606[2]	Trans-Isoprenyl Diphosphate Synthase	C306	1	0
MA0722[nd]	SepCysS	C51*	0/3	0/0
MA0722[nd]	SepCysS	C54*	0/3	0/0
MA0722[2]	SepCysS	C80	0/10	3/8
MA0722[2]	SepCysS	C260*	0/13	11/20
MA0933[2]	dimethylamine methyltransferase	C99	0	1
MA1173[1]	GDP-mannose 4,6 dehydratase	C183	0	1
MA1276[2]	N-ethylammeline chlorohydrolase	C189*	0	2
MA1276[2]	N-ethylammeline chlorohydrolase	C375	0	2
MA1524[2]	Nucleoside diphosphate kinase	C146	0	1
MA1652[1]	dihydrolipoamide dehydrogenase	C42*	0	1
MA1821[nd]	COG1900a	C54*	0/2	0/0
MA1821[1]	COG1900a	C131*	1/10	9/15
MA1821[1]	COG1900a	C453	1/33	10/22
MA2425[2]	dimethylamine methyltransferase	C99	1	2
MA3025[2]	glucosamine-1-phosphate N-acetyltransferase	C127	0	1
MA3025[2]	glucosamine-1-phosphate N-acetyltransferase	C21	0	1
MA3025[2]	glucosamine-1-phosphate N-acetyltransferase	C237	0	1
MA3631[1]	Sirohydrochlorin ferrochelatase	C94	0	1
MA4099[1]	COG1844 (unknown function)	C58*	0	1
MA4243[1]	iron-molybdenum cluster-binding protein	C117	0	1
MA4271[2]	cell division protein FtsZ	C55	0	1
MA4410[2]	GlcD-GlpC fusion	C642*	2	0
MA4590[1]	GMP synthase subunit A	C148*	0	1
MA4615[1]	2-isopropylmalate synthase	C147	0	1
MA4615[1]	2-isopropylmalate synthase	C86*	0	1

[1], persulfided peptide identified in a partially purified sample of MA1821
[2], persulfided peptide identified in a partially purified sample of SepCysS
[nd], persulfided peptide was not detected
*, cysteine is conserved
[?], cysteine conservation could not be determined
DTT, dithiothreitol
+, protein sample was prepared in the presence of DTT
-, protein sample was prepared in the absence of DTT

CHAPTER 4

COG2122A IS REQUIRED FOR SULFIDE UTILIZATION AT LOWER CONCENTRATIONS IN METHANOSARCINA ACETIVORANS

In the previous chapter, COG1900a:Fer were found to participate in a novel pathway for aspartate-4-semialdehyde-dependent homocysteine biosynthesis. The data are consistent with the notion that COG1900a:Fer catalyze this aldehyde-to-thiol transformation. However, it remains unknown whether the COG1900a:Fer reaction uses free sulfide, or an alternative form of sulfur, as a substrate.

The occurrence profile of COG2122a is nearly identical (98% similar) to that of COG1900a [19], and both proteins, along with SepCysS, appear to have been inherited vertically by contemporary methanogens from the ancestral euryarchaeote. This suggests that these proteins could have interdependent functions. However, considering that the enzymes for aspartate-4-semialdehyde biosynthesis are universally conserved, only two plausible roles for COG2122a remain: (1) COG2122a facilitates sulfur donation to COG1900a:Fer and, possibly, SepCysS; (2) COG2122a is involved in regulating the function of COG1900a:Fer.

The function of COG2122a was investigated using a series of gene-deletion and growth experiments conducted in *M. acetivorans*. The analysis could not determine the specific role of COG2122a. However, this protein was essential to sulfide utilization at lower concentrations. Deletion of the gene encoding COG2122a from the pseudo-wild type genetic background led to a 100-fold increase in the amount of sulfide required to support growth. However, when deleted from a strain lacking OASS, the effect was less pronounced.

Construction of *ma1715* deletion strains

In *M. acetivorans*, the lone representative of COG2122a is encoded by the gene *ma1715*. Consistent with the experimental strategy described in Chapter 2, and successfully implemented in Chapter 3, markerless deletions of *ma1715* were constructed in pseudo-wild type (WWM75), Δ*oass* (BJR09) and Δ*oass*Δ*oahs* (BJR22) genetic backgrounds, which yielded the BJR02 (Δ*ma1715*), BJR24 (Δ*oass*Δ*ma1715*) and BJR23 (Δ*oass*Δ*oahs*Δ*ma1715*) strains, respectively (**Fig. 4.1** and **Table 2.1**). The Δ*oahs*Δ*ma1715* genotype could not be constructed. Instead the strain bearing the Δ*oass*Δ*oahs*Δ*ma1715* triple deletion was used to determine if *ma1715* is required for OAHS-independent homocysteine biosynthesis.

Single deletion of *ma1715* results in a loss of growth with lower sulfide concentrations

The single deletion strain lacking *ma1715* was severely impaired for growth with sulfide as the lone sulfur source. Whereas the pseudo-wild type strain grew with maximal efficiency in the presence of 0.2-3.2 mM sulfide, and with reduced efficiency at 0.05 mM sulfide, the Δ*ma1715* strain did not grow at sulfide concentrations below 3.2 mM. Even at this elevated concentration, growth of the Δ*ma1715* strain was still impaired relative to the pseudo-wild type strain (**Fig. 4.2A-D**). Thus, the COG2122a protein may be involved in a general process for sulfide utilization. Consistent with this notion, the Δ*ma1715* strain was also incapable of growth with cysteine present as the lone sulfur source (**Fig. 4.2E**). Since cysteine utilization is thought to proceed through OASS-dependent sulfide synthesis (see Chapter 2), COG2122a may be required to assimilate the presumed small amount of sulfide originating from cysteine by this route.

To reaffirm the notion that *ma1715* is indeed responsible for the observed phenotypes, *ma1715* was added back to the Δ*ma1715* strain using pBR031 (pBR060; **Table 2.1**). The plasmid restored growth with sulfide- and cysteine-only media (**Fig. 4.2A-E**). However, growth did not proceed with the rates typical of the pseudo wild type strain under all conditions. This might be due to the presence of the poly-histidine tag added to the C-terminus of the *ma1715*-conding sequence. Alternatively, the relatively slow growth may be due to under- or over-expression of

ma1715 arising from unfavorable promoter strength. Nonetheless, the recovered growth appears significant, (**Fig. 4.2A**).

The conserved cysteine of COG2122a is not essential

To assess the importance of C161, the lone conserved cysteine residue of COG2122a, a pBR060 derivative bearing the C161A mutation (pBR090; **Table 2.1**) was added back to the Δ*ma1715* strain. In the presence of 50 µM sulfide or 5 mM cysteine the resulting transformant grew comparably to the same strain when transformed with pBR060. This indicates that C161 is not essential to the function of COG2122a (**Fig. 4.5A**).

COG2122a is not required for OASS-independent cysteine biosynthesis at higher sulfide concentrations

The Δ*oass*Δ*ma1715* double deletion strain was capable of growth with sulfide as the lone sulfur source at higher concentrations (**Fig. 4.3A-B**). Since OASS and SepCysS are thought to be the only two enzymes catalyzing *de novo* cysteine biosynthesis in *M. acetivorans* (**Fig. 2.1**), this suggests COG2122a is not required for tRNA-dependent cysteine biosynthesis under these conditions. However, involvement of COG2122a in the SepCysS reaction cannot yet be ruled out in general (see below).

Relative to the Δ*ma1715* strain, the Δ*oass*Δ*ma1715* strain exhibited enhanced growth with higher concentrations (0.8-3.2 mM) of sulfide. The double-deletion strain grew comparably to wild type under these conditions, whereas the single deletion strain did not grow with 0.8 mM sulfide, and grew modestly with 3.2 mM sulfide (**Figs. 4.2A-B** and **4.3A-B**). Since a similar OASS-dependent growth phenotype was not observed in the Δ*oass*Δ*oahs* strain (see Chapter 2; **Fig. 2.3A-D**), these data suggest *M. acetivorans* is impaired for proper sulfide management when COG2122a is absent (see Discussion).

Like the Δ*ma1715* single-deletion strain, the Δ*oass*Δ*ma1715* double-deletion strain was incapable of growth when cysteine or lower concentrations (0.05-0.2 mM) of sulfide were present

as the lone sulfur source (**Figs. 4.2C-E** and **4.3C-E**). This suggests COG2122a may be required for SepCysS activity at lower sulfide concentrations.

MA1715 produced by recombinant expression in *E. coli* is a monomeric protein

The *ma1715* gene was recombinantly expressed under aerobic conditions in *E. coli*. Although the protein that was produced was largely insoluble, approximately 1 mg soluble protein per liter culture was recovered following affinity purification with an immobilized nickel resin. Following purification, the partially purified protein sample was analyzed by size-exclusion chromatography (**Fig. 4.4A-B**). The most prominent 280 nm-absorbing peak eluted with a retention volume of 16.5 ml. Since two size-standard proteins of 44 kDa and 17 kDa eluted with retention volumes of 15 ml and 17 ml, respectively, the MA1715 peak likely corresponds to a single monomeric unit of 28 kDa.

Structural comparison of COG2122a and ApbE

The ApbE family of proteins are structurally homologous to COG2122a. The function of ApbE has been examined for several bacterial homologs, which are thought to catalyze the Mg^{2+}-dependent cleavage of flavin-adenine dinucleotide (FAD) to yield flavin mononucleotide (FMN) and adenosine monophosphate (AMP) [90]. Several x-ray crystal structures are available for ApbE homologs [90–92]. Of these, the structure of *Trepomona pallidum* ApbE with FAD and two catalytic Mg^{2+} ions is particularly informative because it depicts a productive complex.

To assess the likelihood of whether COG2122a binds FAD, protein sequences representative of the ApbE and COG2122a protein families were compared in a multiple sequence alignment (**Fig. 4.5A**). Consensus identities of the ApbE proteins were largely nonconserved in COG2122a. Also, three extended gaps were observed in the COG2122a sequences (77-81, 117-183 and 270-286 in *T. pallidum* ApbE). Since the corresponding regions in ApbE comprise roughly half of the FAD-binding cleft, it seems unlikely that COG2122a is capable of binding FAD similarly (**Fig. 4.5B**).

The conserved ApbE residues D306 and T310 (*T. pallidum* numbering), which together coordinate a magnesium ion, are conserved in COG2122a proteins, suggesting COG2122a may bind a metal ion (**Fig. 4.5**). Consistent with this notion, an unpublished xray crystal structure of COG2122a from *Desulfovibrio vulgaris* (pdb: 2O34) features a sodium ion bound by the residues analogous to D306 and T310. However, it remains unclear whether sodium is in fact the cognate ligand of COG2122a, or whether any metal ion is.

Discussion

A series of genotype-dependent growth experiments have implicated the novel protein COG2122a in mediating efficient sulfide uptake. Although the experiments were not conclusive in determining the exact function of COG2122a, the data suggest this protein could be responsible for mobilizing sulfur for insertion into cysteine and homocysteine. Both deletion strains lacking COG2122a (Δ*ma1715* and Δ*oass*Δ*ma1715*) were deficient for growth with sulfide as the lone sulfur source when present at lower concentrations (0.05-0.2 mM) (**Fig. 4.2-4.3**).

When the gene encoding OASS is present, the single deletion strain lacking COG2122a is impaired for growth at higher sulfide concentrations (0.8-3.2 mM). However, when deleted from the strain lacking OASS, the resulting double-deletion strain is capable of efficient growth under the same conditions. This suggests that OASS, which exhibits a K_M value of 0.5 mM for sulfide [93], may sequester sulfide from other sulfide-dependent biosynthetic pathways, since the OASS-dependent effect is manifested at concentrations above, but not below, the determined K_M value.

The fact that the both COG2122a-deletion strains are not capable of growth with lower concentrations of sulfide (0.05-0.2 mM) suggests COG2122a may be necessary for SepCysS-dependent cysteine biosynthesis when sulfide is limiting. A possible functional link between COG2122a and SepCysS is supported by phylogenetic analyses, which indicate that COG2122a is present in all SepCysS-encoding organisms and that both proteins were present in the ancestral euryarchaeote (see Introduction and Chapter 2).

The functional link between SepCysS and COG2122a is confounded in part by the observation that the lone conserved cysteine residue of COG2122a is not essential to growth with

0.05 mM sulfide (**Fig. 4.2D**). Since SepCysS is thought to require relayed sulfane for efficient activity, and cysteine residues are the only known mediators of sulfane delivery, COG2122a does not conform to the paradigm of sulfur relay (see Introduction and Chapter 2; **Figs. 2.5** and **3.5** and **Table 3.4**; [84,85]). Thus, if COG2122a were to mediate sulfane delivery to SepCysS, it would employ a novel, persulfide-independent mechanism to do so. Furthermore, because COG2122a conserves the metal-chelating residues of ApbE (D306 and T310; **Fig 4.5**), and the yield of soluble COG2122a was significantly increased for recombinant expression in *E. coli* under anaerobic conditions in the presence of iron (**Table A.7**), it is suspected that oxygen sensitive ferrous iron may be involved in COG2122a function.

Alternatively, COG2122a may function in some regulatory capacity. Since COG2122a is essential for growth at lower sulfide concentrations, it may therefore be responsible for activating pathways capable of sulfide assimilation at lower sulfide concentrations.

Finally, a triple-deletion strain lacking COG2122a, OASS and OAHS is under investigation in the laboratory. Although a COG2122a-OAHS double-deletion strain was desired, attempts to engineer this strain were unsuccessful. In its place the triple mutant will be used to assess whether COG2122a is required for COG1900a:Fer-dependent homocysteine biosynthesis. If the strain in question were to exhibit homocysteine auxotrophy, then it would appear likely that COG2122a is essential to mobilize sulfur for insertion into homocysteine.

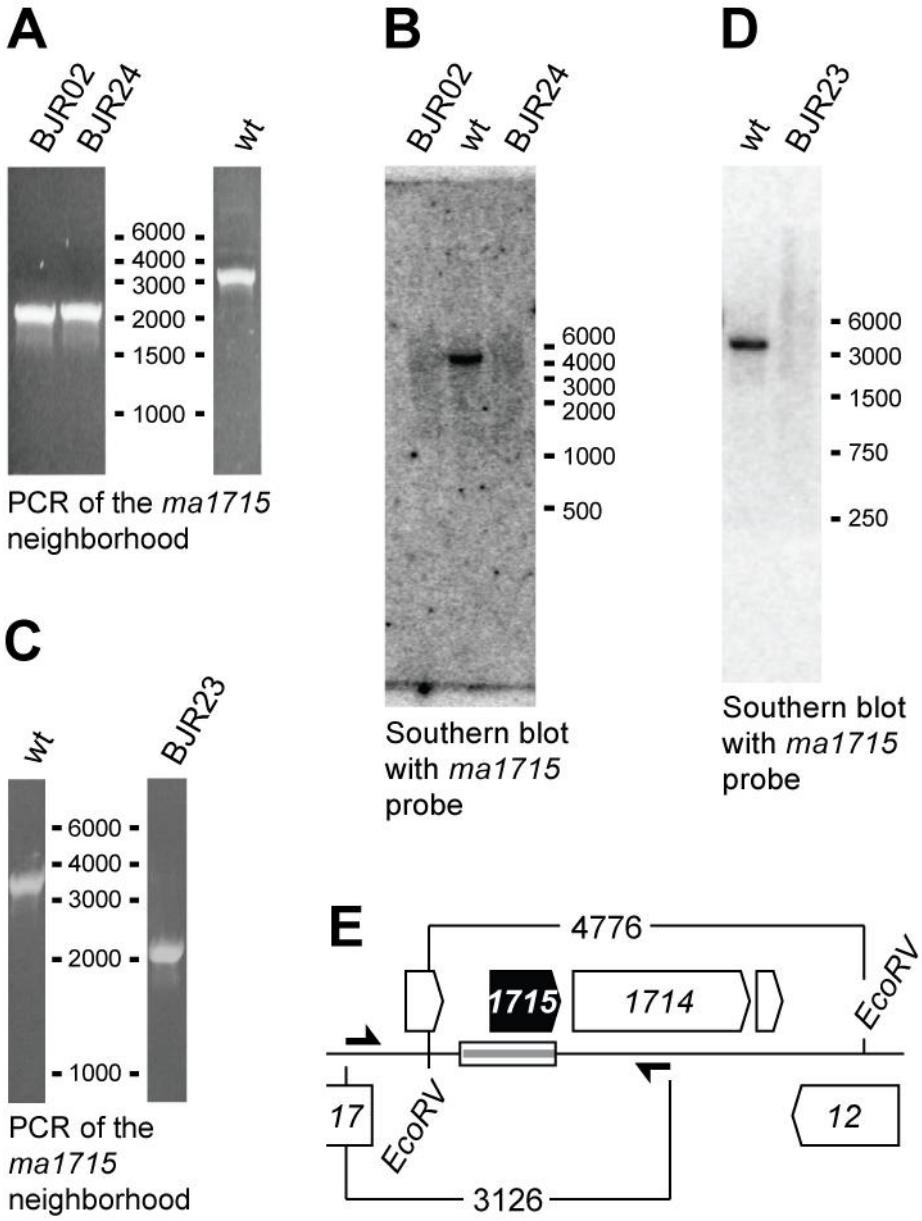


Figure 4.1 Markerless deletions of the gene encoding COG2122a (*ma1715*). Knockouts were constructed in the WWM75 (pseudo-wt; **A-B**), BJR09 (Δ*oass*; **A-B**) and BJR22 (Δ*oass*Δ*oahs*; **C-D**) strains to yield BJR02, BJR24 and BJR23, respectively (see Table 2.1). PCR (**A, C**) and Southern blot (**B, D**) analyses are consistent with the desired manipulation (**E**), which removes a 1049 nt segment (white rectangle) overlapping with *ma1715* in the chromosome. Primers for PCR analysis (black arrows) were used to amplify the genomic neighborhood of *ma1715*, which is predicted to measure 3126 nt in the wild type genotype and 2077 nt in mutant genotypes. The probe used for Southern blot analysis (grey line) measures 1049 nt and hybridizes to *ma1715* and the adjacent upstream non-coding region, which lies between two endogenous EcoRV restriction sites separated by 4776 nt in the wild type genotype.

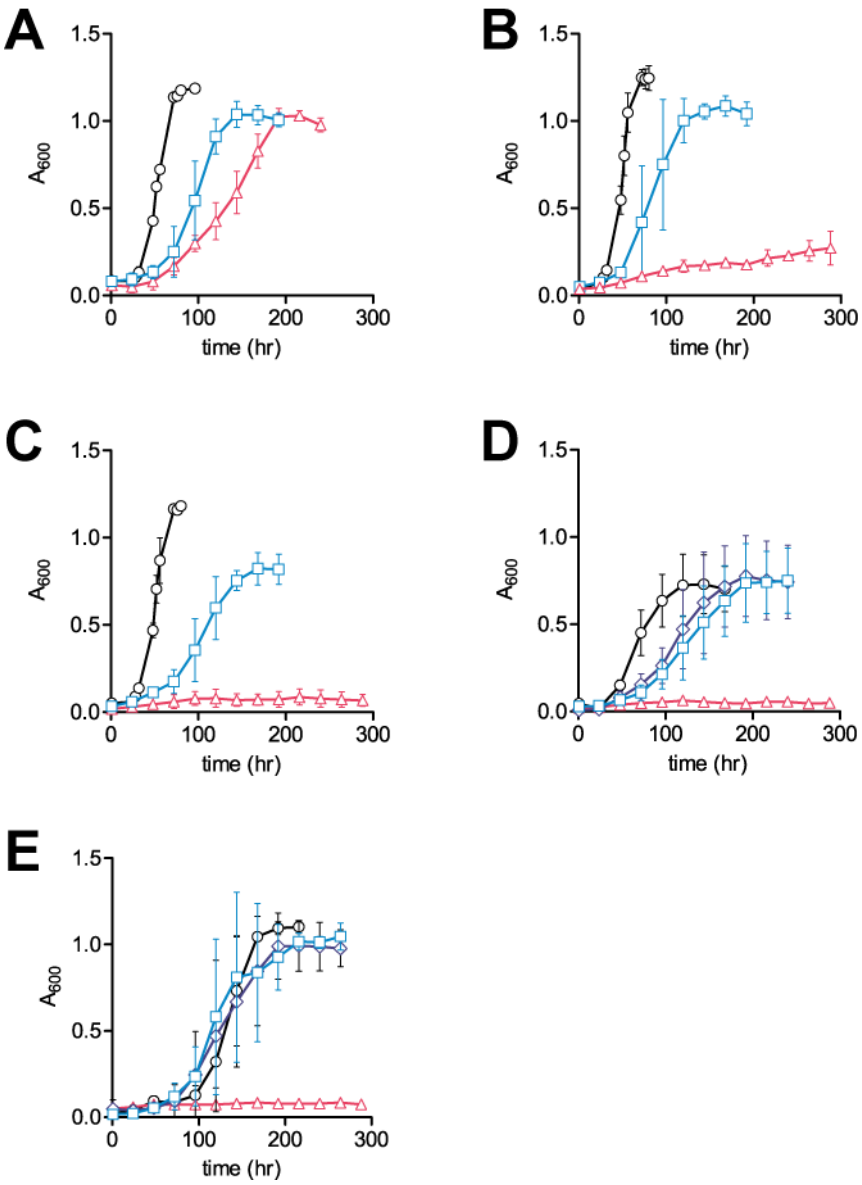


Figure 4.2 Single deletion of the gene encoding COG2122a results in an increased requirement of sulfide for growth. WWM75 (wt; black circles), BJR02 (Δ*ma1715*; pink triangles), BJR02+pBR60 (Δ*ma1715*+*ma1715*; light blue squares) and BJR02+pBR91 (Δ*ma1715*+*ma1715*$_{C161A}$; dark blue diamonds) were grown with HS_{DTT} media supplemented with 3.2 mM sodium sulfide **(A)**, 0.8 mM sodium sulfide **(B)**, 0.2 mM sodium sulfide **(C)**, 0.05 mM sodium sulfide **(D)** or 5 mM cysteine **(E)**. Data points represent an average of 4-6 independent experiments. Error bars reflect standard deviations.

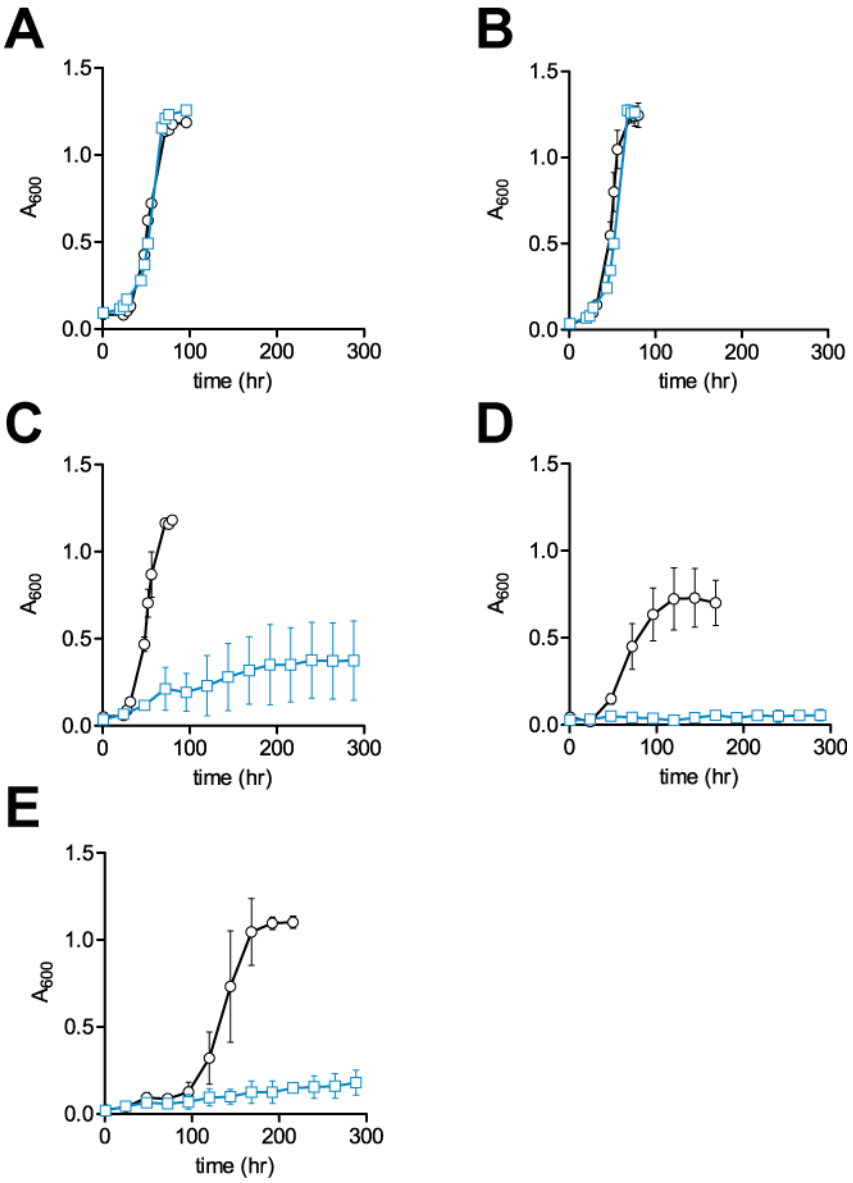


Figure 4.3 Further deletion of the gene encoding OASS from the COG2122a-lacking strain partially restores growth at lower sulfide concentrations (see Fig. 4.2). WWM75 (wt; black circles) and BJR24 (Δ*ma2720*Δ*ma1715*; light blue squares) were grown with HS_{DTT} media supplemented with 3.2 mM sodium sulfide **(A)**, 0.8 mM sodium sulfide **(B)**, 0.2 mM sodium sulfide **(C)**, 0.05 mM sodium sulfide **(D)** or 5 mM cysteine **(E)**. Data points represent an average of 4-6 independent experiments. Error bars reflect standard deviations.

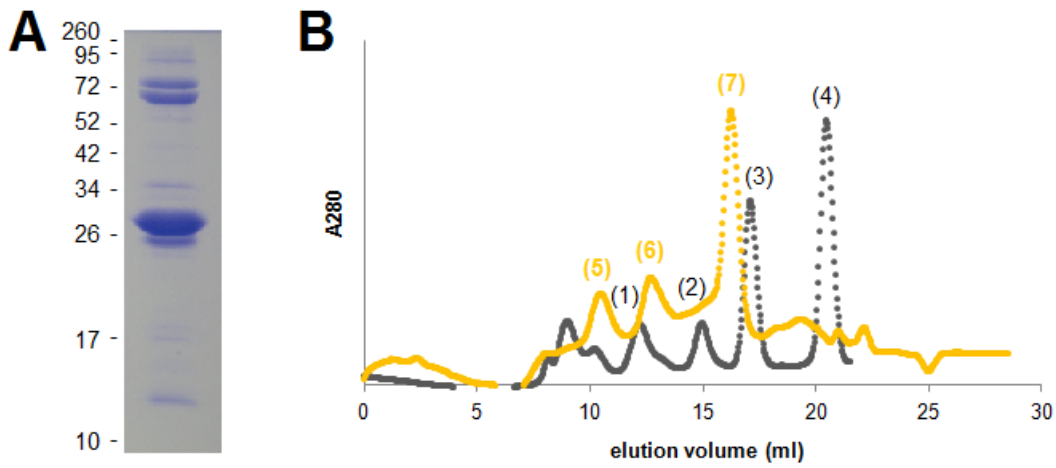


Figure 4.4 Recombinant expression of COG2122a in *Escherichia coli* yields a monomeric protein. **(A)** SDS PAGE of MA1715 purified by Ni-NTA chromatography following recombinant expression in *E. coli*. The predicted monomeric molecular weight of MA1715 is 27.8 kDa. **(B)** Size-exclusion chromatography of the partially purified MA1715 protein sample featured in the gel (yellow orange). Proteins of known particle size were run as controls (gray; 1 = 158 kDa; 2 = 44 kDa; 3 = 17 kDa; 4 = 1.4 kDa). Peak 7 corresponds to a monomer of MA1715, whereas peaks 5 and 6 may be accounted for by the other two prominent peaks featured in the gel.

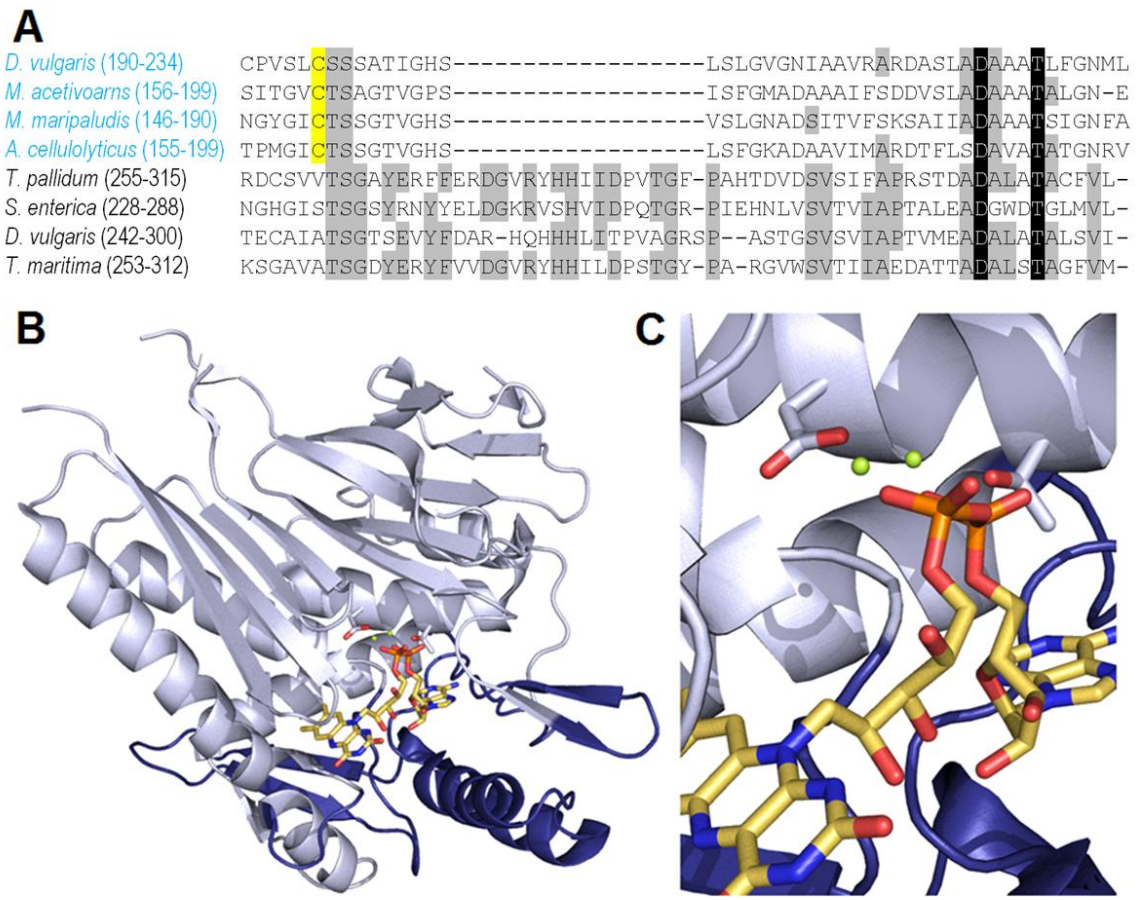


Figure 4.5 Structural comparison of COG2122a and ApbE. **(A)** A sequence alignment of representatives of ApbE (black) and COG2122a (blue) is shown. Consensus identities of ApbE are shaded grey, and are rarely conserved in COG2122a. Residues D306 and T310 (shaded black) of *T pallidum* ApbE may coordinate the magnesium ions required for FAD cleavage. The highly conserved cysteine residue of COG2122a (yellow highlight) is not conserved in ApbE. **(B-C)** The x-ray crystal structure of *T. pallidum* ApbE (grey and blue cartoon) is shown with FAD (yellow sticks) and two Mg^{2+} ions (green spheres). Regions of ApbE that are not conserved in COG2122a (gaps > 4 amino acids; blue cartoon) form approximately half of the FAD-binding interface, which contact the adenylate and flavin moieties directly. **(C)** D306 and T310 are depicted as sticks.

CONCLUDING REMARKS

Sulfur is an essential element to all known cellular life. It is present in proteins as the genetically encoded amino acids cysteine and methionine, and also occurs in numerous enzymatic cofactors and ribonucleotide modifications. Thus, the metabolic pathways for the assimilation of sulfur—and the enzymes that facilitate them—are of fundamental biological importance. The preceding chapters recount the discovery of three novel proteins, which are present in a large group of anaerobic *Bacteria* and *Archaea* and have different roles in sulfide assimilation.

The COG1900a and NIL-Fer proteins together catalyze a novel reductive condensation reaction of aspartate-4-semialdehyde and sulfide to synthesize the universal precursor to methionine, homocysteine. The COG2122a protein may also be involved in homocysteine biosynthesis, as it is essential for growth in *M. acetivorans* under sulfide-limiting conditions. Moreover, all three proteins were inherited vertically by contemporary methanogens, along with the tRNA-dependent cysteine biosynthesis proteins SepRS and SepCysS, from the ancestral euryarchaeote, suggesting these five proteins comprise an ancient strategy for sulfide assimilation that emerged in the Archaean eon.

The ancestral-methanogen-type pathways for cysteine and homocysteine biosynthesis bear one striking similarity. Both consume amino acid intermediates that are synthesized upstream from the analogous intermediates employed in *E. coli*- and *S. cerevisiae*-type pathways. Whereas the pathways prevailing in aerobic systems consume acetyl-serine and acetyl- or succinyl-homoserine for cysteine and homocysteine biosynthesis, respectively, the corresponding

methanogen-type pathways consume phosphoserine and aspartate-4-semialdehye, which both occur two steps upstream from their aerobic counterparts (**Fig. 5.1**). Given the early origin of the methanogen-type pathways, their preference for upstream metabolites may be a reflection of their emergence from a more primitive and condensed metabolic network. Conversely, the use of specialized downstream metabolites in *E. coli*- and *S. cerevisiae*-like systems may indicate that a more expansive and controllable metabolic network is advantageous in these organisms.

From the perspective of functional genomics, the characterization of the COG1900a and COG2122a genes is a significant achievement. Both genes encode proteins that are annotated within the Pfam database as domains of unknown function (DUF). DUFs comprise approximately 20% of all predicted proteins from completely sequenced genomes [94,95]. Without functional information, DUFs represent a prevalent obstacle in understanding the relationship between genotype and phenotype, and thus limit the utility of the rapidly expanding pool of publicly available genomic data. This is particularly true for pursuits in metabolic engineering, which seek to manipulate extant biochemical pathways in their native host organisms or to introduce new pathways into a more manageable systems. Several pathways of interest—including those for heavy-metal remediation (*Desulfovibrio* spp.), cellulose breakdown (*Acetivibrio cellulolyticus*), halocarbon remediation (*Dehalococcoides* spp.) and methanogenesis—originate from anaerobes encoding COG1900a and COG2122a. Since both proteins have roles in sulfide assimilation, and thus influence local sulfide pools, it is not difficult to imagine how COG1900a and COG2122a might interact with these pathways. For example, since heavy-metal remediation proceeds by precipitation of metals following non-enzymatic reaction with sulfide produced from sulfate reduction [96], it is plausible that the activities of COG1900a and COG2122a may influence the efficiency of this process. Similarly, since COG2122a appears to have a general role in sulfide assimilation, and may thus be involved in iron-sulfur cluster biosynthesis, processes like methanogenesis, which rely on multiple iron-sulfur cluster-binding proteins, might be influenced by the activity COG2122a.

The means through which COG1900a, NIL-Fer and COG2122a were identified are of note as a template for future pursuits in gene discovery. All three proteins were identified based on their

presence in genomes encoding the SepCysS protein. Since SepCysS was thought to require a sulfane for activity, it was reasoned that a novel sulfide-dependent sulfane biosynthesis protein would be present in genomes with SepCysS. Moreover, since many of the genomes with SepCysS lacked known proteins for sulfur insertion into homocysteine, it seemed plausible that SepCysS and novel proteins for sulfane and homocysteine biosynthesis were linked as components of the ancestral-methanogen-type strategy for sulfide assimilation. Additional, *in silico* experiments to assess the evolutionary history and identify the conserved genomic neighborhoods of COG1900a, NIL-Fer and COG2122a supported the notion that these uncharacterized proteins might be involved in an ancestral pathway for sulfide assimilation. Thus, prior to the genotype-dependent growth experiments, which eventually demonstrated roles for each protein in sulfide assimilation, a variety of complementary information from the literature and publicly available genomes were used to identify and approximate the function of three novel proteins. Considering that the employed analyses required minimal investment, they might serve as a model for identifying other elusive proteins.

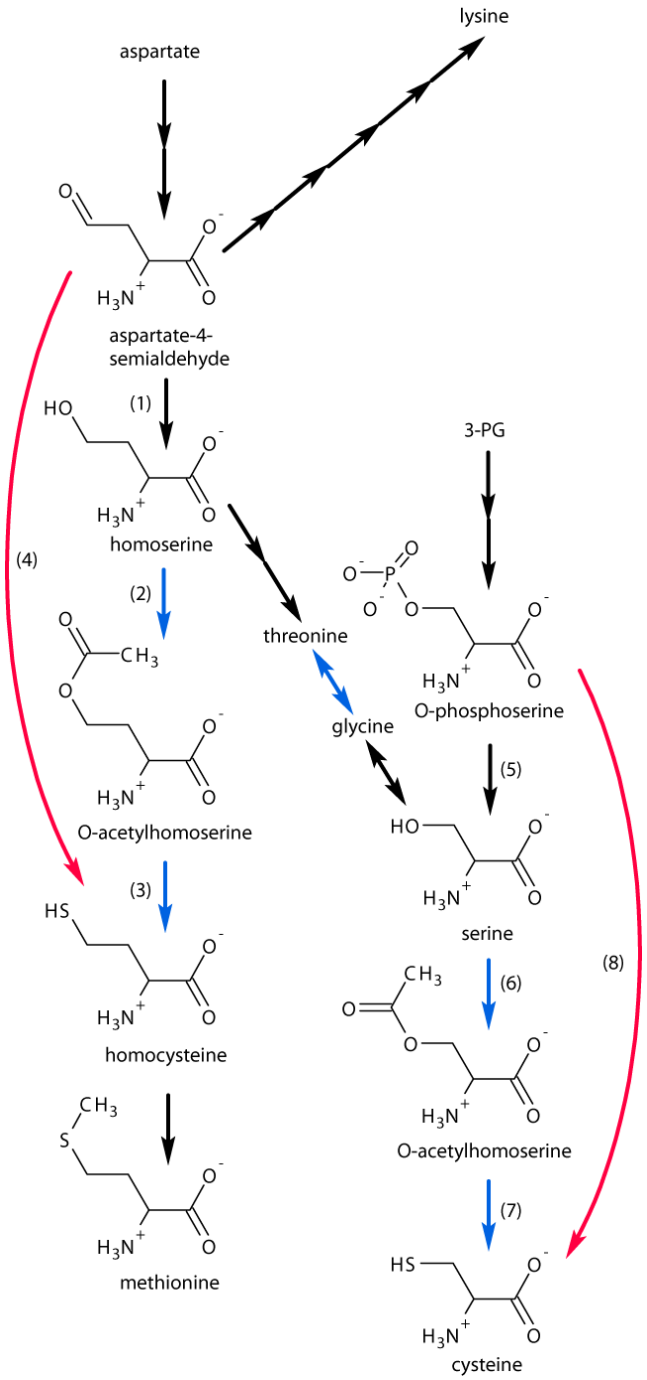


Figure 5.1 The strategy for cysteine and homocysteine biosynthesis in methanogens reflects its ancient emergence. Methanogen-type reactions (pink) do not use specialized metabolic intermediates for cysteine and homocysteine biosynthesis, which are preferred in the aerobe-type reactions (blue). Instead precursors for the methanogen-type reactions serve as intermediates in near-universal processes (black). Transsulfuration reactions, which proceed through the specialized intermediates O-succinylhomoserine and cystathionine, are excluded for simplicity. Labeled reactions are catalyzed by (1) homoserine dehydrogenase, (2) homoserine acetyltransferase, (3) O-acetylhomoserine sulfhydrylase, (4) COG1900a:Fer, (5) phosphoserine phosphatase, (6) serine acetyltransferase, (7) O-acetylserine sulfhydrylase and (8) SepCysS. Note that the SepCysS reaction is tRNA-dependent, and that this is not reflected in the figure. Metabolites are labeled by name. Only 3-phosphoglycerate is abbreviated (3-PG).

METHODS

Strains and media

Escherichia coli strains were grown aerobically in LB medium under the appropriate antibiotic selection. DH5α was used as the host for pET-22b(+), DH5α/λ-pir was used for the *oriR6K*-dependent plasmids pMP44 and pWM321 [97], and WM3118 for the *oriV*-dependent pJK031A [75]. 10 mM rhamnose was supplemented into the WM3118 cultures prior to plasmid purification, to permit plasmid replication at high copy number.

M. acetivorans strain WWM75 and derived strains were grown in high-salt (HS) medium to promote single-cell morphology [75,98,99]. All growths were carried out at 37^{o} C, under 2 psi N_2:CO_2 (80:20), with methanol (125 mM) as the methanogenesis substrate. HS medium supplemented with 3 mM methionine (HS_{Met}) was used for routine strain propagation and genetic manipulations. Growth-monitoring experiments were conducted in HS medium lacking cysteine, sodium sulfide, thiamine, biotin and thioctic acid, with 1.5 mM dithiothreitol added as a reducing agent (HS_{DTT}). Sulfur sources were added to HS_{DTT} after autoclaving, immediately prior to use. Dithiothreitol alone does not support growth of the *M. acetivorans* strains used in this study (data not shown).

Construction of markerless deletions in M. acetivorans.

Upstream and downstream flanking sequences, extending 1000-1250 bp from the gene to be deleted, were amplified separately by PCR from WWM75 genomic DNA (**Table 2.1**). Opposing

flanking sequences were then fused by overlap extension PCR, yielding mutant genotypes suitable for insertion into pMP44 using *SpeI* and *KpnI* sites (**Table 2.1**).

Liposome-mediated transfection was used to construct mero-diploid intermediate strains as described [100]. Purified intermediate colonies were grown in 10 ml liquid medium without selection. After reaching quiescence, the strain was propagated for 15-20 additional generations. Counter-selective plating was performed by resuspending mature cultures (A600 ~ 1.0) in 0.1 ml HS_{Met} medium, and spreading 30 µl of the cell suspension at dilutions of 1:1, 1:10 and 1:100 on agar-solidified media containing 200 µg/ml 8-aza-2,6-diaminopurine sulfate (8ADP; Santa Cruz Biotechnology). Under these conditions, 8ADP-resistant colonies emerged after 20 days. Typically eight colonies were picked and grown in 10 ml liquid medium without selection. After assessing genotype by PCR, confirmed knockout strains were restreaked for purification on agar-solidified medium containing 200 µg/ml 8ADP. Purified strains were grown to quiescence without selection in 10 ml liquid medium and analyzed further by DNA hybridization via Southern blotting anlysis, which was performed as described [101,102].

Complementation of mutants with pBR031.

To construct pBR031 (**Fig. 2.2**), P_{mcrB}(tetO1) was first amplified from pJK031A and subsequently cloned into pWM321 using the *SpeI* and *SphI* sites. Prior to insertion into pBR031, the coding regions for *ma1821, ma1821-1822*, *ma1822* and *ma1715* were first cloned into pET22b(+) (Novagen), via *NdeI* and *XhoI* restriction sites, yielding pBR004, pBR038, pBR05 and pBR60, respectively. Using site-directed mutagenesis, two endogenous *SphI* restriction sites occurring in *ma1821* were removed from pBR004 and pBR038, resulting in pBR051 and pBR053. Coding regions from pBR051, pBR053 and pBR05 were amplified and cloned into pBR031 using *Sph* and *ApaI* sites, resulting in pBR054, pBR055 and pBR058. Cysteine mutants were constructed in *ma1821* and *ma1715* by site-directed mutagenesis on template plasmids pBR055 and pBR060.

The Δ*ma1821-22*Δ*Fer* mutant was constructed from pBR055 in two steps. First, using site-directed mutagenesis, an *XhoI* restriction site was introduced immediately downstream from the

NIL-coding region of *ma1822*, yielding pBR056. Second, the Fer-coding region was removed from pBR056 by digestion with *XhoI*, followed by intramolecular ligation, yielding pBR062.

The *ma1821-22*$_{\Delta NIL}$ mutant was constructed by amplification of *ma1821* and *ma1822*$_{Fer}$ from pBR053. *Ma1821* and *ma1822*$_{Fer}$ were then fused by overlap extension and cloned into pBR031 using *SphI* and *ApaI*, yielding pBR059.

Plasmids derived from pBR031 were introduced into the *Δma1821-22Δoahs* strain by liposome-mediated transfection, as described with minor variation [100] (see Appendix III). Puromycin-resistant colonies were used in subsequent growth experiments without further purification.

Growth experiments.

All growth experiments were conducted at 37^{o} C in Balch tubes containing 10 ml HS_{DTT} medium supplemented with various sulfur sources. Starter cultures were grown to late exponential phase ($0.50 < A_{600} < 0.70$) and diluted into the experimental growth medium such that $A_{600(initial)} = 0.010$. To minimize carry-over of sulfur sources from HS_{Met} medium, starter cultures were grown in the experimental growth medium for at least seven generations when possible, such that carry-over concentrations for cysteine, methionine and sulfide were less than 1.2 µM, 1.2 µM and 0.3 µM, respectively. Of necessity, in the case of strains auxotrophic for homocysteine, starter cultures were carried out with HS_{Met} medium.

Measurement of A_{600} was performed using a Turner 330 spectrophotometer. The instrument was modified to accommodate 18 mm Balch tubes, so that culture tubes could be inserted directly in the light path.

Doubling times were calculated by fitting data points ($A_{600} < 1.1$) to the exponential growth equation:

$$y = ae^{kx}$$

where y is time-dependent A_{600}, a is the initial A_{600}, k is the rate constant for growth and x is time. Prior to fitting, individual time courses were normalized by subtracting a single value from all time points such that A_{600} of the 1-hr time point = 0.01. Data were fit in GraphPad Prism without fixing

a to 0.01. Reported growth rates and growth curves reflect the mean of five independent experiments. Error bars represent standard deviations.

Anaerobic technique

A facility suitable for genetic analyses in *Methanosarcina* was constructed by standard methods [74,103–106]. To remove trace levels of oxygen, N_2:CO_2:H_2 and N_2:CO_2 mixes were purified through a glass column (Chemglass, AF-0200-01), packed with 200 g of copper turnings. The column was heated externally with heating tape at a rate of 1050 W. The gas purifier is integrated into copper/stainless steel manifolds via PVC tubing. Anaerobic media was prepared using the boiling method under a stream of N_2:CO_2 (80:20). Boiling was carried out in 2000 ml borosilicate bottles with GL-45 screw-thread openings (Kimble NO. 14396) over a 10" x 10" hot plate (Fisher Scientific™ Isotemp™). Bottles were sealed with butyl rubber stoppers (OCHS, 444704) and GL-45 screw caps with apertures (Chemglass, CG-195-05). After boiling, media were supplemented with methanol, ammonium sulfate, potassium phosphate, cysteine, methionine, sodium sulfide and DTT inside a type-B vinyl anaerobic chamber (Coy Laboratories). The anaerobic chamber was maintained with N_2:CO_2:H_2 (70:20:10), with hydrogen content held near 3%. Note that hydrogen is consumed to remove oxygen by the palladium catalyst (See appendix II). Media were distributed into the appropriate glassware inside the anaerobic chamber. Balch tubes (Chemglass, CLS-4209-10) and serum bottles with natural-gum rubber stoppers (OCHS: 191000, 191111, 444702) were used for liquid and agar-solidified media, respectively. The gas contents of Balch tubes containing media were adjusted to 2 psi N_2:CO_2 (80:20) and subsequently autoclaved. Growths on agar-solidified media (1.5% w/v) were carried out on 10-cm Petri dishes containing 35 ml media. Live plates were stored under 4 psi N_2:CO_2:H_2S (80:20:0.1) in anaerobic jars (Almore International, Inc.: 15000), residing in a 37^o C incubator within the anaerobic chamber.

Purification of genomic DNA from M. acetivorans strains.

The procedure employed was adapted from the published description [107]. Cells from saturated cultures (1-4 ml) were harvested by centrifugation (30 sec. at 16000 rcf). Cells were resuspended in 350 µl lysis buffer (10 mM Tris·HCl, 50 mM EDTA, 1% SDS (pH 8.0)) by pipetting, and were then incubated at 80^{o} C for 5 minutes with intermittent vortexing. After cooling to room temperature, 750 µl isopropanol was added, and the contents were mixed by inversion. The resulting precipitate was harvested by centrifugation (5 min. at 16000 rcf.). The recovered DNA pellet was resuspended in 200 µl TE buffer (10 mM Tris·HCl, 1 mM EDTA (pH 7.5) and treated with 100 µg/ml RNaseA for 30 min. at 65^{o} C. The reaction was cooled to 37^{o} C, 100 µg/ml proteinase K was added, and the reaction was allowed to proceed for 15 minutes. Proteins were then extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1). Genomic DNA was recovered by ethanol precipitation of the aqueous phase.

Southern Blot

Methods for DNA hybridization were performed as described, with minor variation [101,102]. Genomic DNA (gDNA) was purified from 4 ml saturated cell culture, and a 5-10 µg aliquot was digested with FastDigest *EcoRV* (Thermo) in accordance with the manufacturer's recommendations. Digested gDNA samples, including a wild-type control, were separated by agarose gel electrophoresis (85 mm x 60 mm). Separated DNA was denatured in the gel by shaking for 30 min. at ambient temperature with denaturation buffer (1.5 M NaCl, 0.5 M NaOH). After neutralization by further shaking for 30 min. at room temperature with neutralization buffer (1 M ammonium acetate, 20 mM NaOH), the fragmented, separated, denatured gDNA was transferred to a 0.45 µm-NYTRAN®N filter (Whatman™) with neutralization buffer by capillary action for 12-16 hrs. at ambient temperature. Following transfer, the DNA was baked into the membrane by incubating in an oven set to 80^{o} C for 1 hr. DNA probes specifically targeting *ma1821* (991 bp), *oahs* (*ma2715;* 722 bp), *oass* (*ma2720*; 732 bp) and *ma1715* (997 bp) were synthesized by PCR (**Table 2.1**). After purification using a GeneJet PCR cleanup kit (Thermo), probes (10 pmol dsDNA) were end-labeled with ^{32}P at the 5'-phosphate positions using T4

polynucleotide kinase (Thermo) in accordance with the manufacturer's recommendations. Labeled probes were denatured by boiling at 105^{o} C for 5 minutes and subsequent flash-freezing in dry ice-ethanol. Probes were stored at -20^{o} C. The baked membrane was incubated in hybridization buffer (7% SDS, 0.5 M monobasic sodium phosphate, 1 mM EDTA, with pH adjusted to 7.2 using phosphoric acid) for 1 hr. at 65^{o} C, rotating in a rotisserie oven. Denatured salmon sperm DNA was added to a final concentration of 100 µg/ml. After another hour rotating in the oven at 65^{o} C, the labeled, denatured probe was added and allowed to hybridize for >10 hr. The membrane was washed twice with 0.2X SSC/0.1% SDS at RT, and thrice more—each time rotating at 65^{o} C for 20 minutes. Finally, the membrane was visualized by phosphorimaging.

Bioinformatic methods.

Homologous proteins were identified using the Basic Local Alignment Search Tool (BLAST) [108]. Protein sequence queries were used to search the non-redundant protein sequence database with default algorithm parameters. Protein sequences producing significant alignments were downloaded as a single fasta file. Alignments were performed with ClustalW [109] using default settings. Phylogenies were reconstructed from multiple-protein sequence alignments in Molecular Evolutionary Genetics Analysis (MEGA5) [110] using the neighbor-joining method. The Poisson model for amino acid substitutions was employed with uniform rates and pair-wise deletion of gaps. Phylogenies were tested by bootstrapping with 2000 replicates. The genomic contexts of genes of interest were analyzed using the NCBI Sequence Viewer (http://www.ncbi.nlm.nih.gov/tools/sviewer/) and the Microbial Genomic context Viewer (MGcV; http://mgcv.cmbi.ru.nl/) [69].

Expression and purification of SepCysS from *M. acetivorans*

All manipulations were performed anaerobically with the assistance of the anaerobic chamber. Buffers were chilled to 4^{o} C prior to use. Low-temperature conditions were maintained inside of the anaerobic chamber by using freezer packs.

The WWM75 strain was transformed with pBR049, which encodes SepCysS from *M. acetivorans* with a C-terminal polyhistidine tag under the transcriptional control of P_{mcrB}tet(O1).

The resulting transformant was cultured at 37^{o} C with HS_{Met} in the presence of 2 µg/ml puromycin. Upon reaching an optical density of 0.15, 100 mg tetracycline was added to the culture, which was allowed to grow at 37^{o} for an additional 24 hours.

After reaching an optical density of ~1.0, cells were harvested by centrifugation for 20 minutes at 7,000 rcf. After resuspension in buffer A (10 mM tris HCl pH = 8.0, 1 M KCl, 10 mM imidazole, 15% glycerol) cells were lysed by bead beating. This was performed by votexing the cell suspension with 4 ml of 0.1 mm glass beads in a 50 ml falcon tube for approximately 2 minutes. 20 µM PLP and 10 µg/ml DNase were added to the crude lysate, which was then cleared by centrifugation for 40 minutes at 17,000 rcf. 16 ml of supernatant was recovered and cleared further by filtration through a 0.45 µm membrane.

SepCysS was purified by affinity chromatography with an immobilized nickel resin. The cleared lysate was incubated for 30 minutes with 4 ml of nickel resin that was pre-equilibrated in buffer A. The lysate-nickel resin mixture was poured over a column (diameter = 1 cm). The resin was then washed with 40 ml buffer A, 40 ml buffer B (10 mM tris HCl pH = 8.0, 100 mM KCl, 10 mM imidazole, 15% glycerol) and 16 ml buffer C (10 mM tris HCl pH = 8.0, 100 mM KCl, 200 mM imidazole, 15% glycerol). The protein eluting from the column with buffer C was concentrated with a centrifugal filtration device (50 kDa MWCO) to a concentration of ~0.3 mg/ml, and transferred to storage buffer (50 mM HEPES, pH = 8.0 via NaOH, 100 mM NaCl and 50% glycerol) by dialysis. The concentration of SepCysS was determined by UV absorbance assuming ε_{280} = 49195 $M^{-1}cm^{-1}$. Approximately 0.4 mg of SepCysS were recovered per liter of culture.

Expression and purification of and MA1821 and MA1822 from *M. acetivorans*

All manipulations were performed anaerobically with the assistance of the anaerobic chamber. Buffers were chilled to 4^{o} C prior to use. Low-temperature conditions were maintained inside of the anaerobic chamber by using freezer packs.

The homocysteine auxotrophic strain, BJR10, was transformed with pBR070 which encodes MA1821 and MA1822 from *M. acetivorans* with C-terminal polyhistidine tags under the transcriptional control of P_{mcrB}tet(O1). The resulting transformant was cultured at 37^{o} C to an

optical density of 1.0 with HS_{DTT} supplemented with 2 µg/ml puromycin, 75 µg/ml and 1.5 mM sodium sulfide.

Techniques for cell harvest, cell lysis and protein purification were carried out like those for SepCysS, but with minor variation. All buffers used to purify MA1821 and MA1822 lacked glycerol and contained 2 mM B-mercaptoethanol. Cell lysis was performed in the presence of an EDTA-free protease inhibitor cocktail tablet (Roche) and 10 ml glass beads. Pyridoxal phosphate (PLP) was not administered. Approximately 1.2 mg of MA1821 were recovered pre liter culture (assuming ε_{280} = 49000 $M^{-1}cm^{-1}$). Note that peptides for MA1822 were not observed in this sample following trypsin LC-MS/MS analysis.

Preparation of protein samples for trypsin LC-MS/MS

SepCysS and MA1821-MA1822 protein samples were dialyzed anaerobically at 4^{o} C into a buffer lacking glycerol and reducing agents (50 mM tris-HCl, pH = 8.0, 100 mM KCl) with a dilution factor of 1:10,000. The dialyzed samples were re-assessed for their protein concentrations by UV absorption. Following transfer to 1.5 ml microcentrifuge tubes, solvent was removed from the protein samples using a SpeedVac™ that was run for 16 hours at low heat. Protein precipitates were then resuspended in 4X urea buffer (8 M urea, 1 M trisHCl pH = 8.5, 8 mM $CaCl_2$ and 200 mM methylamine) so that the final protein concentration was 0.5 mg/ml. 10 µl aliquots (5 µg) were then distributed into 1.5 ml microcentrifuge for alkylation under reduced or non-reducing conditions.

Samples to be alkylated under reducing conditions were supplemented with 1 µl of 200 mM DTT and incubated for 15 minutes at 50^{o} C. Following an additional 5 minute incubation at room temperature, the reduced protein was supplemented with 1 µl 500 mM iodoacetic acid. Alkylation was allowed to proceed for 30 minutes in the dark at room temperature. Reactions were then quenched following the addition of 2 µl of 200 mM DTT and an additional 15 minute incubation in the dark at room temperature. Samples were stored for several days at -20^{o} C prior to digestion with trypsin.

Immediately following resuspension in 4X urea buffer, samples to be alkylated in the absence of reducing agents were supplemented with 1 µl of 50 mM iodoacetic acid and allowed to incubate for 30 minutes in the dark at room temperature. Samples were stored for several days at -20^{o} C prior to digestion with trypsin.

Alkylated protein samples were supplemented with water and trypsin so that 40 µl mixtures resulted, in which substrate proteins were in excess of trypsin by 25-fold ($^{mass}/_{mass}$). These mixtures were incubated at 37^{o} C overnight and subsequently quenched by the addition of 2 µl of 88% formic acid. Digested samples were analyzed by LC-MS/MS, which was conducted in the laboratory of Larry David at OHSU.

Expression and purification of MA1715 from *E. coli*

BL21(DE3)pLysS was transformed with pBR006, which expresses *ma1715* with an engineered C-terminal polyhistidine tag under the control of the *lac* promoter (**Table 2.1**). The resulting transformant was cultured in 2,000 ml LB medium in the presence of 50 µg/ml ampicillin and 35 µg/ml chloramphenocol, shaking 200 rpm at 37^{o} C. Upon reaching an optical density of 0.9, the culture was induced for expression by adding 1 mM IPTG. The culture was allowed to continue growth for an additional 5 hours under the same conditions. The optical density eventually reached 2.2. Cells were harvested by centrifugation at 5,000 rcf for 20 minutes at 4^{o} C.

Cells were resuspended in buffer D (20 mM trisHCl, pH = 8.0, 1 M NaCl, 10 mM imidazole, 10% glycerol and 2 mM BME) by vortexing. An additional 2 ml of buffer D containing 25 mg lysozyme and 4 mg PMSF were added to the suspended cells, which were then mixed by inversion. Cell lysis was assisted by sonication, which was conducted for 1 minute at an intensity of 20 W. The crude lysate was then supplemented with 10 µg/ml DNaseI and 10 µg/ml RNaseA and allowed to incubate on ice for 30 minutes. Insoluble debris were cleared by centrifugation at 20,000 rcf for 40 minutes at 4^{o} C. The supernatant was dialyzed against 1,000 ml buffer D for 16 hours at 4^{o} C. The dialyzed lysate was then filtered through a 0.45 µm pore membrane.

Cleared lysate was run through a 2 ml bed of immobilized-nickel resin that was equilibrated in buffer D. After washing the column with 30 ml buffer D, MA1715 was eluted with 10 ml buffer E

(20 mM trisHCl, pH = 8.0, 1 M NaCl, 300 mM imidazole, 10% glycerol and 2 mM BME). The elution sample was concentrated four-fold, down to 2.5 ml using a centrifugal filtration device (MWCO = 10 kDa). The protein was then dialyzed into an anaerobic storage buffer (20 mM trisHCl, pH = 8.0, 100 mM NaCl, 40% glycerol) and stored at -20^{o} C.

Size exclusion chromatography of MA1715

Size exclusion chromatography was conducted aerobically with a supredex 200 column (GE Healthcare) in line with an FPLC, using MA1715 that was purified from *E. coli* by nickel-affinity chromatography. Prior to analysis, proteins were dialyzed into mobile phase (20 mM HEPES-KOH, pH = 7.5, 2 mM BME and 100-1000 mM NaCl). Samples were injected from a 0.1 ml loop and were separated by maintaining a flow rate of 0.75 ml/min. Data for the time-resolved absorbance of 280 nm light were collected with UNICORN software and were manipulated in Microsoft Excel for visualization.

Purification of *M. acetivorans* SepRS following recombinant expression in *E. coli*

BL21(DE3)pLysS was transformed with pBR007, which encodes *M. acetivorans* SepRS with an N-terminal polyhistidine tag under the transcriptional control of the *lac* promoter. The transformant was cultured at 37^{o} C in 1000 ml LB medium shaking at 200 rpm. After reaching an optical density of 0.6, 0.6 mM IPTG was added to induce protein production for an additional 5 hours of culture under the same conditions. Cells were harvested by centrifugation at 5,000 rcf for 20 minutes and stored at -20^{o} C after discarding the culture medium.

All manipulations were carried out at 0-4^{o} C. Cells were resuspended in 20 ml buffer E (50 mM $Na_xH_{2-x}PO_4$, pH = 7.8, 150 mM NaCl and1 mM BME) supplemented with 1 mg/ml lysozyme and one EDTA-free protease inhibitor cocktail tablet (Roche). After sonication for a combined 6 minutes (5 sec on, 10 sec off), 4 µg/ml DNaseI was added to the crude cell lysate and allowed to react for one hour. Insoluble debris were then cleared by centrifugation at 17,000 rcf for 40 minutes and subsequent filtration through a 0.45 µm membrane.

To purify SepRS, the cleared lysate was passed through a 4 ml bed of immobilized-nickel resin that was equilibrated with buffer E. After washing the resin with 80 ml of buffer E, supplemented with 20 mM imidazole, SepRS was eluted by washing the resin with 12 ml of buffer D supplemented with 200 mM imidazole. The purified protein was dialyzed into buffer F (50 mM HEPES-NaOH, pH = 8.0, 50 mM NaCl, 1 mM BME and 50% glycerol) for storage at -20^{o} C. Protein concentration was then assessed by UV absorbance. 6.6 mg purified protein was recovered.

Activity assay for SepCysS

The tRNA^{Cys} substrate for SepRS and SepCysS was prepared in accordance with established methods for DNA template synthesis and *in vitro* transcription with T7 RNA polymerase [111]. Its sequence, GCCAAGAUGGCGGAGCGGCUACGCAAUCGCCUGCAGAGC GAUUCCAUUCCGGUUCGAAUCCGGAUCUUGGCUCCA, is encoded within the genome of *M. acetivorans* at the chromosomal locus 1,740,455-526. The template for transcription was synthesized with the Klenow fragment of DNA polymerase from two DNA primers, AATTCCTGCAGTAATACGACTCACTATAGCCAAGATGGCGGAGCGGCTACGCAATCGCCTG CAGAG and [2-OMe]U[2-OMe]GGAGCCAAGATCCGGATTCGAACCGGAATGGAATCGCT CTGCAGGCGAT, which hybridize to form a 13 bp overlapping region. Each 2 ml transcription reaction contained 400-1000 pmol template DNA along with 40 mM tris▪HCl (pH = 8.0), 25 mM $McCl_2$ 2 µM spermidine, 0.01% tritonX-100, 40 mM DTT, 4 mM ATP, 4 mM UTP, 4 mM GTP, 4mM CTP, 2 units pyrophosphatase and 0.08 mg T7 RNA polymerase, and was incubated 16-24 hours at 37^{o} C. After phenol-chloroform extraction and ethanol precipitation, tRNA-sized transcripts were purified from unincorporated NTPs and DNA template on a 15% polyacrylamide:bisacrylamide (29:1) gel with 8 M urea. The tRNA band was excised and extracted with TE6 (10 mM bistrisHCl, 1 mM EDTA, pH = 6.0). Using a centrifugal filtration device (Amicon; 10 kDa MWCO), the tRNA sample was washed such that the final urea concentration was less than 1 µM.

Prior to use in SepCysS activity assays, $tRNA^{Cys}$ was labeled with ^{32}P at the 3'-internucleotide linkage. Incorporation of ^{32}P was achieved with the assistance of the CCA adding enzyme (tRNA nucleotidyltransferase) and [γ-^{32}P]ATP as described [112]. After incorporation, labeled $tRNA^{Cys}$ was purified away from unincorporated ATP after separation on a 15% polyacrylamide:bisacrylamide (29:1) gel lacking denaturants by extracting the RNA-containing gel slice passively overnight at room temperature with anaerobic TE6 buffer.

SepRS and SepCysS reactions were performed at 37^{o} C or at room temperature, anaerobically in a reaction buffer containing 50 mM NaOAc (pH = 6.0), 5 mM $Mg(OAc)_2$ (pH = 6.0), 2 mM ATP, 0.2 mM Sep, 20 mM KCl, 5 mM DTT. Note that SepRS activity is enhanced significantly at pH = 6.0 [28,113] as opposed to pH = 7.5 [114]. However the preferred pH for SepCysS activity was not determined. Time points were quenched by diluting 1:3 into solution, containing 200 mM NaOAc, 0.2% SDS, pH = 5.2. Subsequently, quenched time points were digested with *Penicillium citrinum* P1 nuclease (Sigma) at a concentration of 0.0125 u/ul. Free nucleotides were separated by thin layer chromatography across 10 cm PEI cellulose sheets (Sigma) with 1 M acetic acid, pH = 3.5 using NH_4OH as the source of base equivalents.

REFERENCES

1. Hamann, C. S., Sowers, K. R., Lipman, R. S. & Hou, Y. M. An archaeal aminoacyl-tRNA synthetase missing from genomic analysis. *J. Bacteriol.* **181,** 5880–5884 (1999).
2. Tumbula, D. *et al.* Archaeal aminoacyl-tRNA synthesis: diversity replaces dogma. *Genetics* **152,** 1269–1276 (1999).
3. Bult, C. J. *et al.* Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science* **273,** 1058–1073 (1996).
4. Sauerwald, A. *et al.* RNA-dependent cysteine biosynthesis in archaea. *Science* **307,** 1969–1972 (2005).
5. O'Donoghue, P., Sethi, A., Woese, C. R. & Luthey-Schulten, Z. A. The evolutionary history of Cys-tRNACys formation. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 19003–19008 (2005).
6. Blank, C. E. Not so old Archaea – the antiquity of biogeochemical processes in the archaeal domain of life. *Geobiology* **7,** 495–514 (2009).
7. Hayes, J. in *Early life on Earth* 220–236 (Columbia University Press, 1994).
8. Bekker, A. *et al.* Dating the rise of atmospheric oxygen. *Nature* **427,** 117–120 (2004).
9. Liu, Y., Beer, L. L. & Whitman, W. B. Methanogens: a window into ancient sulfur metabolism. *Trends Microbiol.* **20,** 251–258 (2012).
10. Kojima, S., Hanamuro, T., Hayashi, K., Haruna, M. & Ohmoto, H. Sulphide minerals in early Archean chemical sedimentary rocks of the eastern Pilbara district, Western Australia. *Mineral. Petrol.* **64,** 219–235 (1998).

11. Habicht, K. S., Gade, M., Thamdrup, B., Berg, P. & Canfield, D. E. Calibration of sulfate levels in the archean ocean. *Science* **298,** 2372–2374 (2002).

12. Kletzin Arnulf. in *Arcahaea: Molecular and Cellular Biology* 71 (2007).

13. Blank, C. E. Phylogenomic dating--the relative antiquity of archaeal metabolic and physiological traits. *Astrobiology* **9,** 193–219 (2009).

14. Ferry, J. G. & Kastead, Kyle A. in *Arcahaea: Molecular and Cellular Biology* 288–304 (2007).

15. Reeburgh, W. S. in *Microbial Growth on C1 Compounds* (eds. Lidstrom, M. E. & Tabita, F. R.) 334–342 (Springer Netherlands, 1996).

16. Noll, K. M., Rinehart, K. L., Tanner, R. S. & Wolfe, R. S. Structure of component B (7-mercaptoheptanoylthreonine phosphate) of the methylcoenzyme M methylreductase system of Methanobacterium thermoautotrophicum. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 4238–4242 (1986).

17. McBride, B. C. & Wolfe, R. S. A new coenzyme of methyl transfer, coenzyme M. *Biochemistry (Mosc.)* **10,** 2317–2324 (1971).

18. Klipcan, L., Frenkel-Morgenstern, M. & Safro, M. G. Presence of tRNA-dependent pathways correlates with high cysteine content in methanogenic Archaea. *Trends Genet.* **24,** 59–63 (2008).

19. Rauch, B. J., Gustafson, A. & Perona, J. J. Novel proteins for homocysteine biosynthesis in anaerobic microorganisms. *Mol. Microbiol.* **94,** 1330–1342 (2014).

20. Liu, Y., Sieprawska-Lupa, M., Whitman, W. B. & White, R. H. Cysteine is not the sulfur source for iron-sulfur cluster and methionine biosynthesis in the methanogenic archaeon Methanococcus maripaludis. *J. Biol. Chem.* **285,** 31923–31929 (2010).

21. Sekowska, A., Kung, H. F. & Danchin, A. Sulfur metabolism in Escherichia coli and related bacteria: facts and fiction. *J. Mol. Microbiol. Biotechnol.* **2,** 145–177 (2000).

22. Hidese, R., Mihara, H. & Esaki, N. Bacterial cysteine desulfurases: versatile key players in biosynthetic pathways of sulfur-containing biofactors. *Appl. Microbiol. Biotechnol.* **91,** 47–61 (2011).

23. Mueller, E. G. Trafficking in persulfides: delivering sulfur in biosynthetic pathways. *Nat. Chem. Biol.* **2,** 185–194 (2006).

24. Thomas, D. & Surdin-Kerjan, Y. Metabolism of sulfur amino acids in Saccharomyces cerevisiae. *Microbiol. Mol. Biol. Rev.* **61,** 503–532 (1997).

25. Mino, K. & Ishikawa, K. A novel O-phospho-L-serine sulfhydrylation reaction catalyzed by O-acetylserine sulfhydrylase from Aeropyrum pernix K1. *FEBS Lett.* **551,** 133–138 (2003).

26. Scheer, M. *et al.* BRENDA, the enzyme information system in 2011. *Nucleic Acids Res.* **39,** D670–676 (2011).

27. Aitken, S. M., Lodha, P. H. & Morneau, D. J. K. The enzymes of the transsulfuration pathways: Active-site characterizations. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1814,** 1511–1517 (2011).

28. Rauch BJ. Unpublished observations. (2015).

29. Ayala-Castro, C., Saini, A. & Outten, F. W. Fe-S cluster assembly pathways in bacteria. *Microbiol. Mol. Biol. Rev. MMBR* **72,** 110–125, (2008).

30. Mihara, H. & Esaki, N. Bacterial cysteine desulfurases: their function and mechanisms. *Appl. Microbiol. Biotechnol.* **60,** 12–23 (2002).

31. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42,** D222–D230 (2014).

32. Martinez-Gomez, N. C., Palmer, L. D., Vivas, E., Roach, P. L. & Downs, D. M. The rhodanese domain of thiI is both necessary and sufficient for synthesis of the thiazole moiety of thiamine in salmonella enterica. *J. Bacteriol.* **193,** 4582–4587 (2011).

33. Kambampati, R. & Lauhon, C. T. Evidence for the transfer of sulfane sulfur from IscS to ThiI during the in vitro biosynthesis of 4-thiouridine in Escherichia coli tRNA. *J. Biol. Chem.* **275,** 10727–10730 (2000).

34. Palenchar, P. M., Buck, C. J., Cheng, H., Larson, T. J. & Mueller, E. G. Evidence that ThiI, an enzyme shared between thiamin and 4-thiouridine biosynthesis, may be a sulfurtransferase that proceeds through a persulfide intermediate. *J. Biol. Chem.* **275,** 8283–8286 (2000).

35. You, D., Xu, T., Yao, F., Zhou, X. & Deng, Z. Direct evidence that ThiI is an ATP pyrophosphatase for the adenylation of uridine in 4-thiouridine biosynthesis. *Chembiochem Eur. J. Chem. Biol.* **9,** 1879–1882 (2008).

36. Kriek, M. *et al.* Thiazole synthase from Escherichia coli: an investigation of the substrates and purified proteins required for activity in vitro. *J. Biol. Chem.* **282,** 17413–17423 (2007).

37. Leonardi, R., Fairhurst, S. A., Kriek, M., Lowe, D. J. & Roach, P. L. Thiamine biosynthesis in Escherichia coli: isolation and initial characterisation of the ThiGH complex. *FEBS Lett.* **539,** 95–99 (2003).

38. Maupin-Furlow, J. A. Prokaryotic Ubiquitin-Like Protein Modification. *Annu. Rev. Microbiol.* **68,** 155–175 (2014).

39. Shigi, N., Sakaguchi, Y., Asai, S.-I., Suzuki, T. & Watanabe, K. Common thiolation mechanism in the biosynthesis of tRNA thiouridine and sulphur-containing cofactors. *EMBO J.* **27,** 3267–3278 (2008).

40. Bouvier, D. *et al.* TtcA a new tRNA-thioltransferase with an Fe-S cluster. *Nucleic Acids Res.* gku508 (2014).

41. Nakagawa, H. *et al.* Crystallographic and mutational studies on the tRNA thiouridine synthetase TtuA. *Proteins* **81,** 1232–1244 (2013).

42. Neumann, P. *et al.* Crystal structure of a 4-thiouridine synthetase-RNA complex reveals specificity of tRNA U8 modification. *Nucleic Acids Res.* **42,** 6673–6685 (2014).

43. Hauenstein, S. I. & Perona, J. J. Redundant synthesis of cysteinyl-tRNACys in Methanosarcina mazei. *J. Biol. Chem.* **283,** 22007–22017 (2008).

44. Liu, Y. *et al.* Catalytic mechanism of Sep-tRNA:Cys-tRNA synthase: sulfur transfer is mediated by disulfide and persulfide. *J. Biol. Chem.* **287,** 5426–5433 (2012).

45. Yuan, J. *et al.* A tRNA-dependent cysteine biosynthesis enzyme recognizes the selenocysteine-specific tRNA in Escherichia coli. *FEBS Lett.* **584,** 2857–2861 (2010).

46. Liu, Y., Long, F., Wang, L., Söll, D. & Whitman, W. B. The putative tRNA 2-thiouridine synthetase Ncs6 is an essential sulfur carrier in Methanococcus maripaludis. *FEBS Lett.* **588,** 873–877 (2014).

47. Liu, Y. *et al.* Biosynthesis of 4-thiouridine in tRNA in the methanogenic archaeon Methanococcus maripaludis. *J. Biol. Chem.* **287,** 36683–36692 (2012).

48. Miranda, H. V. *et al.* E1- and ubiquitin-like proteins provide a direct link between protein conjugation and sulfur transfer in archaea. *Proc. Natl. Acad. Sci.* **108,** 4417–4422 (2011).

49. Leidel, S. *et al.* Ubiquitin-related modifier Urm1 acts as a sulphur carrier in thiolation of eukaryotic transfer RNA. *Nature* **458,** 228–232 (2009).

50. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 4285–4288 (1999).

51. Kaster, A.-K. *et al.* More Than 200 Genes Required for Methane Formation from H_2 and CO_2 and Energy Conservation Are Present in *Methanothermobacter marburgensis* and *Methanothermobacter thermautotrophicus*. *Archaea* **2011,** e973848 (2011).

52. Gao, B. & Gupta, R. S. Phylogenomic analysis of proteins that are distinctive of Archaea and its main subgroups and the origin of methanogenesis. *BMC Genomics* **8,** 86 (2007).

53. Klenk, H. P. *et al.* The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus. *Nature* **390,** 364–370 (1997).

54. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42,** D222–D230 (2014).

55. Sonnhammer, E. L., Eddy, S. R. & Durbin, R. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* **28,** 405–420 (1997).

56. Tatusov, R. L. *et al.* The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4,** 41 (2003).

57. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. A genomic perspective on protein families. *Science* **278,** 631–637 (1997).

58. Marchler-Bauer, A. & Bryant, S. H. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* **32,** W327–331 (2004).

59. Marchler-Bauer, A. *et al.* CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39,** D225–229 (2011).

60. Borrel, G. *et al.* Phylogenomic data support a seventh order of Methylotrophic methanogens and provide insights into the evolution of Methanogenesis. *Genome Biol. Evol.* **5,** 1769–1780 (2013).

61. Baykov, A. A., Tuominen, H. K. & Lahti, R. The CBS domain: a protein module with an emerging prominent role in regulation. *ACS Chem. Biol.* **6,** 1156–1163 (2011).

62. Lucas, M. *et al.* Binding of S-methyl-5'-thioadenosine and S-adenosyl-L-methionine to protein MJ0100 triggers an open-to-closed conformational change in its CBS motif pair. *J. Mol. Biol.* **396,** 800–820 (2010).

63. Baykov, A. A., Tuominen, H. K. & Lahti, R. The CBS domain: a protein module with an emerging prominent role in regulation. *ACS Chem. Biol.* **6,** 1156–1163 (2011).

64. Kadaba, N. S., Kaiser, J. T., Johnson, E., Lee, A. & Rees, D. C. The high-affinity E. coli methionine ABC transporter: structure and allosteric regulation. *Science* **321,** 250–253 (2008).

65. Beck, B. J. & Downs, D. M. The apbE gene encodes a lipoprotein involved in thiamine synthesis in Salmonella typhimurium. *J. Bacteriol.* **180,** 885–891 (1998).

66. Skovran, E. & Downs, D. M. Lack of the ApbC or ApbE protein results in a defect in Fe-S cluster metabolism in Salmonella enterica serovar Typhimurium. *J. Bacteriol.* **185,** 98–106 (2003).

67. Bapteste, E., Brochier, C. & Boucher, Y. Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. *Archaea Vanc. BC* **1,** 353–363 (2005).

68. Farahi, K., Pusch, G. D., Overbeek, R. & Whitman, W. B. Detection of lateral gene transfer events in the prokaryotic tRNA synthetases by the ratios of evolutionary distances method. *J. Mol. Evol.* **58,** 615–631 (2004).

69. Overmars, L., Kerkhoven, R., Siezen, R. J. & Francke, C. MGcV: the microbial genomic context viewer for comparative genome analysis. *BMC Genomics* **14,** 209 (2013).

70. Galagan, J. E. *et al.* The genome of M. acetivorans reveals extensive metabolic and physiological diversity. *Genome Res.* **12,** 532–542 (2002).

71. Sowers, K. R., Baron, S. F. & Ferry, J. G. Methanosarcina acetivorans sp. nov., an acetotrophic methane-producing bacterium isolated from marine sediments. *Appl. Environ. Microbiol.* **47,** 971–978 (1984).

72. Rother, M. & Metcalf, W. W. Anaerobic growth of Methanosarcina acetivorans C2A on carbon monoxide: An unusual way of life for a methanogenic archaeon. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 16929–16934 (2004).

73. Sarmiento, F., Mrázek, J. & Whitman, W. B. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon Methanococcus maripaludis. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 4726–4731 (2013).

74. Buan, N., Kulkarni, G. & Metcalf, W. Genetic methods for methanosarcina species. *Methods Enzymol.* **494,** 23–42 (2011).

75. Guss, A. M., Rother, M., Zhang, J. K., Kulkarni, G. & Metcalf, W. W. New methods for tightly regulated gene expression and highly efficient chromosomal integration of cloned genes for Methanosarcina species. *Archaea Vanc. BC* **2,** 193–203 (2008).

76. Metcalf, W. W., Zhang, J. K., Apolinario, E., Sowers, K. R. & Wolfe, R. S. A genetic system for Archaea of the genus Methanosarcina: Liposome-mediated transformation and construction of shuttle vectors. *Proc. Natl. Acad. Sci.* **94,** 2626–2631 (1997).

77. Pritchett, M. A., Zhang, J. K. & Metcalf, W. W. Development of a markerless genetic exchange method for Methanosarcina acetivorans C2A and its use in construction of new genetic tools for methanogenic archaea. *Appl. Environ. Microbiol.* **70,** 1425–1433 (2004).

78. Hildenbrand, C., Stock, T., Lange, C., Rother, M. & Soppa, J. genome copy numbers and gene conversion in Methanogenic Archaea. *J. Bacteriol.* **193,** 734–743 (2011).

79. Sowers, K. R., Boone, J. E. & Gunsalus, R. P. Disaggregation of Methanosarcina spp. and growth as single cells at elevated osmolarity. *Appl. Environ. Microbiol.* **59,** 3832–3839 (1993).

80. Hungate, R. E. The anaerobic mesophilic cellulolytic bacteria. *Bacteriol. Rev.* **14,** 1–49 (1950).

81. Rauch, B. J., Gustafson, A. & Perona, J. J. Novel proteins for homocysteine biosynthesis in anaerobic microorganisms. *Mol. Microbiol.* **94,** 1330–1342 (2014).

82. Liu, Y., Sieprawska-Lupa, M., Whitman, W. B. & White, R. H. Cysteine Is not the sulfur source for iron-sulfur cluster and methionine biosynthesis in the Methanogenic Archaeon Methanococcus maripaludis. *J. Biol. Chem.* **285,** 31923–31929 (2010).

83. Liu, Y., Beer, L. L. & Whitman, W. B. Methanogens: a window into ancient sulfur metabolism. *Trends Microbiol.* **20,** 251–258 (2012).

84. Hauenstein, S. I. & Perona, J. J. Redundant Synthesis of Cysteinyl-tRNACys in Methanosarcina mazei. *J. Biol. Chem.* **283,** 22007–22017 (2008).

85. Liu, Y. *et al.* Catalytic mechanism of Sep-tRNA:Cys-tRNA synthase: sulfur transfer is mediated by disulfide and persulfide. *J. Biol. Chem.* **287,** 5426–5433 (2012).

86. Liu, Y. *et al.* Ancient translation factor is essential for tRNA-dependent cysteine biosynthesis in methanogenic archaea. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 10520–10525 (2014).

87. Helgadóttir, S., Sinapah, S., Söll, D. & Ling, J. Mutational analysis of Sep-tRNA:Cys-tRNA synthase reveals critical residues for tRNA-dependent cysteine formation. *FEBS Lett.* **586,** 60–63 (2012).

88. Metcalf, W. W., Zhang, J. K., Apolinario, E., Sowers, K. R. & Wolfe, R. S. A genetic system for Archaea of the genus Methanosarcina: liposome-mediated transformation and construction of shuttle vectors. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 2626–2631 (1997).

89. white, R. H. Biosynthesis of coenzyme M (2-mercaptoethanesulfonicacid). *Biochemistry (Mosc.)* **24,** 6487–6493 (1985).

90. Deka, R. K., Brautigam, C. A., Liu, W. Z., Tomchick, D. R. & Norgard, M. V. The TP0796 lipoprotein of Treponema pallidum is a bimetal-dependent FAD pyrophosphatase with a potential role in flavin homeostasis. *J. Biol. Chem.* **288,** 11106–11121 (2013).

91. Boyd, J. M. *et al.* FAD binding by ApbE protein from Salmonella enterica: a new class of FAD-binding proteins. *J. Bacteriol.* **193,** 887–895 (2011).

92. Han, G. W. *et al.* Crystal structure of the ApbE protein (TM1553) from Thermotoga maritima at 1.58 A resolution. *Proteins* **64,** 1083–1090 (2006).

93. Borup, B. & Ferry, J. G. O-Acetylserine sulfhydrylase from Methanosarcina thermophila. *J. Bacteriol.* **182,** 45–50 (2000).

94. Goodacre, N. F., Gerloff, D. L. & Uetz, P. Protein Domains of Unknown Function Are Essential in Bacteria. *mBio* **5,** e00744–13 (2014).

95. Bateman, A., Coggill, P. & Finn, R. D. DUFs: families in search of function. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **66,** 1148–1152 (2010).

96. Lovley, D. R. & Phillips, E. J. Reduction of uranium by Desulfovibrio desulfuricans. *Appl. Environ. Microbiol.* **58,** 850–856 (1992).

97. Kolter, R., Inuzuka, M. & Helinski, D. R. Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K. *Cell* **15,** 1199–1208 (1978).

98. Sowers, K. R., Boone, J. E. & Gunsalus, R. P. Disaggregation of Methanosarcina spp. and Growth as Single Cells at Elevated Osmolarity. *Appl. Environ. Microbiol.* **59,** 3832–3839 (1993).

99. Metcalf, W. W., Zhang, J. K., Shi, X. & Wolfe, R. S. Molecular, genetic, and biochemical characterization of the serC gene of Methanosarcina barkeri Fusaro. *J. Bacteriol.* **178,** 5797–5802 (1996).

100. Buan, N., Kulkarni, G. & Metcalf, W. Genetic methods for methanosarcina species. *Methods Enzymol.* **494,** 23–42 (2011).

101. Alphey, L. & Parry, H. D. in *DNA cloning 1* (eds. Glover, D. M. & Hames, B. D.) 121–123, 129–135 (Oxford University Press, 1995).

102. Parry, H. D. & Alphey, L. in *DNA cloning 1* (eds. Glover, D. M. & Hames, B. D.) 143–158 (Oxford University Press, 1995).

103. Balch, W. E., Fox, G. E., Magrum, L. J., Woese, C. R. & Wolfe, R. S. Methanogens: reevaluation of a unique biological group. *Microbiol. Rev.* **43,** 260–296 (1979).

104. Sowers, K. R. & Noll, K. M. in *Methanogens* (eds. Sowers, K. R. & Schreier, H. J.) 15–46 (Cold Spring Harbor Laboratory Press, 1995).

105. Wolfe, R. S. Techniques for cultivating methanogens. *Methods Enzymol.* **494,** 1–22 (2011).

106. Kohler, P. R. A. & Metcalf, W. W. Genetic manipulation of Methanosarcina spp. *Front. Microbiol.* **3,** 259 (2012).

107. Sowers, K. R. in *Methanogens* (eds. Sowers, K. R. & Schreier, H. J.) 369–378 (Cold Spring Harbor Laboratory Press, 1995).

108. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).

109. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinforma. Oxf. Engl.* **23,** 2947–2948 (2007).

110. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28,** 2731–2739 (2011).

111. Sherlin, L. D. *et al.* Chemical and enzymatic synthesis of tRNAs for high-throughput crystallization. *RNA N. Y. N* **7,** 1671–1678 (2001).

112. Wolfson, A. D. & Uhlenbeck, O. C. Modulation of tRNAAla identity by inorganic pyrophosphatase. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 5965–5970 (2002).

113. Zhang, C.-M., Liu, C., Slater, S. & Hou, Y.-M. Aminoacylation of tRNA with phosphoserine for synthesis of cysteinyl-tRNA(Cys). *Nat. Struct. Mol. Biol.* **15,** 507–514 (2008).

114. Hauenstein, S. I., Hou, Y.-M. & Perona, J. J. The Homotetrameric Phosphoseryl-tRNA Synthetase from Methanosarcina mazei Exhibits Half-of-the-sites Activity. *J. Biol. Chem.* **283,** 21997–22006 (2008).

115. Mossessova, E. & Lima, C. D. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell* **5,** 865–876 (2000).

APPENDIX I

PREPARATION OF HIGH-SALT MEDIA FOR METHANOSARCINA CULTURE

Two types of high-salt (HS) media are used in the lab: HS_{Met} and HS_{DTT}. The HS_{Met} medium is used for routine strains propagation and performing genetic manipulations. It includes multiple sulfur sources (cysteine, methionine and sodium sulfide) and sulfur-containing cofactors (thiamine, biotin and thioctic acid). The HS_{DTT} medium is used for growth experiments. It lacks usable sulfur sources and sulfur-containing cofactors. Dithiothreitol (DTT) is used as a reducing agent in place of cysteine. For both media, methanol is used as the methanogenesis substrate, carbon dioxide as the carbon source and ammonium as the nitrogen source.

Step 1: The day before. Media receptacles, such as Balch tubes, should be moved to the anaerobic chamber at least one day in advance. It takes approximately two and 24 hours to passively remove oxygen from glass and plastic surfaces, respectively.

Step 2: Preparing solutions A and B. Solutions A and B are to be prepared separately prior to heat-mediated gas exchange in Pyrex bottles with GL-45 threads. Their contents are listed below (Tables A1-2). Typically, two-liter bottles are used for A and B when preparing greater than, or equal to, one liter. Occasionally it is necessary to scale down the prep. When this is the case, I use smaller-sized bottles (1 liter or 0.5 liters). Each bottle should be fitted with a butyl stopper and GL-45 aperture cap (Fig. A1).

Step 2: Heat-mediated gas exchange. Bring solutions A and B to a boil using the large hot plate that lives on the bench that also houses the gas manifold.

- Place the bottles on the hot plate while it is turned off.
- Remove the red GL-45 aperture caps and stoppers.
- Place the stoppers back on the mouths of the bottles. Do not make a seal; gas needs to be able to escape the bottles.
- Set the temperature to 400^{o} C. It will take at least 15 minutes for the solutions to boil.
- Allow both solutions to boil hard for 10-20 minutes.
- Turn off the hot plate.
- Using the gas probes apply streams of anaerobic gas (20% CO_2, 80% N_2) to solutions A and B. Do this while the bottles are still on the hot plate, immediately after turning off the heat. Adjust the flow rate of the gas through the probes so that the gas is audible and tactile, but not noisy and turbulent
- After two minutes of gas exchange on the hot plate, move the bottles from the hotplate to the bench. Allow it to sit there for an additional two minutes. Continue to apply anaerobic gas.
- Transfer the bottles to an ice-water bath. The bottles will cool to room temperature after 15-30 minutes on ice. When this happens, seal the bottles with their stoppers while removing their probes. Then, turn off the gas, remove the bottles from the ice and wipe the bottles dry.

Step 3a: Additional components for HS_{Met}. Weigh out and package the media components listed below in weigh-paper envelopes secured, but not sealed, by aluminum foil (Table A3, Fig. A2). Transfer these into the anaerobic chamber along with anaerobic solutions A and B. Special care must be given when manipulating and weighing sodium sulfide. This reagent is stored as large crystals under oil in an amber glass bottle in the fume hood. It is difficult to obtain sodium sulfide crystals of a precise size. Typically, a larger-than-needed crystal is brought into the anaerobic chamber and used to make a stock solution with anaerobic water.

To obtain a sodium sulfide crystal of suitable mass, a larger crystal is removed from the bottle and crushed with a hammer in the fume hood. A smaller crystal fragment can then be washed with de-ionized water, dried with a Kimwipes paper and weighed.

Step 3b: Additional components for HS_{DTT}. Weigh out and package the media components listed below in weigh-paper envelopes secured, but not sealed, by aluminum foil (Table A4, Fig. A2). Transfer these into the anaerobic chamber along with anaerobic solutions A and B.

Step 4: Mixing the media components. This step is performed inside of the anaerobic chamber and calls for the use of three anaerobic reagents, water, methanol and phosphate buffer. All three are stored inside of the anaerobic chamber.

- Combine solutions A and B. Mix by shaking.
- Add ammonium chloride. mix by shaking.
- Add anaerobic phosphate buffer. Use 5 ml of 1 M KH_2PO_4 (pH 6.8) per liter media. Mix by shaking.
- Add anaerobic methanol. Use 5 ml 100% methanol per liter media. Mix by shaking.
- Add de-ionized water to the media so that the desired volume is obtained.
- Reduce the media. For HS_{Met} add cysteine. For HS_{DTT} add dithiothreitol. Mix by shaking. The oxygen indicator, resazurin, will turn from blue to pink to clear as it becomes reduced. This takes approximately 30 minutes.
- For HS_{Met}, add methionine and sodium sulfide after the media have turned clear. Mix my shaking
- Distribute the media into receptacles.

Step 5a: Balch tube preparation. Balch tubes are the gold standard for culturing anaerobes at volumes of 10 ml or less. They have an outer diameter of 18 mm and can be sealed with 20 mm stoppers and aluminum crimps (Fig. A3A). They can be stored in standard Fisherbrand 20 mm diameter tube racks. Generally, 40-tube racks are used.

- Stopper the tubes inside the anaerobic chamber. Stoppers should be pre-greased
- Remove stopped tubes from the anaerobic chamber.
- Crimp the stopped tubes on the manifold bench using a crimping tool (Fig. A3). After crimping. HS_{Met} tubes are ready for autoclaving.
- Using the gas manifold needles, adjust the headspaces of ten sealed HS_{DTT} Balch tubes in parallel to 2 psi of 20% CO_2 and 80% N_2. Perform three evacuation-fill cycles. For the first two cycles evacuate to -20 inHg and fill to 8 psi. For the third cycle, evacuate to -20 inHg and fill to 2 psi.

Step 5b: Solid media prep. Solid media contain 1.5% agar and are autoclaved twice—once after the initial prep and a second time immediately prior to pouring plates. The bottles used to contain solid media are without a name. They are manufactured by the German company OCHS and are sealed by rubber stoppers, which are held in place by screw caps (Fig. A3B). In theory, GL45-thread borosilicate bottles could be used instead, but this has not been tried. Agar is stored as 4.5 g aliquots inside of the anaerobic chamber. This amount of agar is sufficient to prepare 300 ml of solid media, which can be used to pour 10-12 plates.

Step 6: Autoclaving. Autoclave for 1 hour at 121^{o} C, using liquid cycle. Since the media are sealed in bottles and tubes, they may explode. Protect yourself by using a secondary container. Place solid media bottles in a 4-liter Nalgene bucket with an inverted 2-liter Nalgene bucket as cover. Place Balch tubes in an autoclaving tray and cover them with a second tray.

Notes

- After autoclaving liquid media, a whiter precipitate will form. After the media cool, continually mix the media using an orbital shaker.
- After autoclaving solid media**,** make sure that all agar is homogeneously dissolved. Mix by inversion.
- The oxygen scrubber should be turned on at least 15 minutes prior to the administration of gas. Turn it off when it is no longer in use.
- After exposure to HS media, glassware need to be washed in 0.1 M HCl for 24 hr to remove residual salt.
- Vitamin and trace elements solutions are prepared as 1000 ml aqueous (Tables A5-6). They are stored aerobically at room temperature. Vitamin solution should be filter sterilized and protected from exposure to light.

APPENDIX II

THE ANAEROBIC FACILITY

To perform the genetic manipulations described in this text, a specialized anaerobic facility was established based on similar facilities present in the laboratory of William Metcalf at the University of Illinois at Urbana-Champaign. Due to limited funding and space, the Metcalf-style facility could not be emulated. Thus, out of necessity, the facility described here was designed with efficiency in mind.

The anaerobic chamber.

Anaerobic manipulations were conducted within a type-B vinyl anaerobic chamber manufactured by Coy Laboratories (Grass Lake, MI). The anaerobic atmosphere (0-5 ppm oxygen) is maintained with hydrogen (3-5%) which reduces oxygen at room temperature in the presence of palladium catalyst. The chamber is equipped with an automatic airlock, through which objects can be transferred. Upon activation, the airlock is twice evacuated (-20 in Hg) and filled with nitrogen and is evacuated and filled once more with a gas mixture containing 10% hydrogen, 20% carbon dioxide an 70% nitrogen. Although not required for the removal of oxygen, carbon dioxide is included in the gas mix to maintain the bicarbonate buffer system of Methanosarcina growth media

The anaerobic chamber requires little routine maintenance. During periods of heavy use (more than one airlock cycle per day), catalyst cartridges should be replaced at least once per week. However, during periods of minimal use (less than one airlock cycle per day), catalyst

cartridges need only be replaced monthly. Between uses, catalyst cartridges must be regenerated by incubation at 100-150^{o} C for more than one hour. This serves to remove water. Eventually, the catalyst pellets in each cartridge should be replaced. This should be performed once per year.

Gas manifolds.

To purify and distribute gas mixtures for anaerobic manipulations, a custom manifold system was constructed a mixture of stainless steel and copper tubing and stainless steel and brass Swagelok fixtures (Fig. A4).

Trace amounts of oxygen were removed from gas mixtures by directing gas through an oxygen scrubber--a glass column (Chemglass, AF-0200-01) packed with 200 g of copper turnings. When in use, the oxygen scrubber was heated externally with heating tape at a rate of 1050 W, which allows for both oxidation and reduction of the copper catalyst. The oxygen scrubber tolerates up to 15 psi of pressure. Exceeding 15 psi, may result in leaks or explosions. Accordingly, gas pressures are typically limited to 12 psi.

Preparation of anaerobic materials.

Equipment can be made anaerobic simply by extended incubation in the anaerobic chamber. However, it should be noted that different material require different incubation times to become free of oxygen. Generally, borosilicate glass items require one-hour incubations prior to use, whereas plastic items require 24 hours. Equipment that are passed through the airlock are subjected to considerable convection. This is particularly problematic for sterile items which may become compromised when subjected to extreme fluctuations in pressure and the associated convection. When this is the case, items are sealed in sterilization pouches, which are permeable to air but not potentially contaminating particulates (Fig. A5).

Most anaerobic solutions are prepared by transferring the solute into the anaerobic chamber and dissolving it in anaerobic solvent (see solvent prep below).

Solid reagents require shielding from convection in the air lock. To achieve this, solids are transferred through the air lock in a protective envelope comprised of weigh paper an aluminum foil (Fig. A2).

For liquids, not including growth media, that need to be made anaerobic prior to transfer into the chamber, oxygen is simply displaced using the gas diffuser (Fig. A4). Typically, the N_2:CO_2 gas mixture is used for oxygen displacement. However, N_2 alone is used on occasion, when the solution pH is of concern. Oxygen displacement may be conducted in any GL-45-threaded borosilicate bottle fitted with a butyl rubber stopper (OCHS, Germany) and aperture cap (Schott, wherever; Fig. A3). For liquids of relatively low viscosity, such as water, methanol and potassium phosphate buffer, gas exchange is administered at room temperature for 30 minutes. Alternatively, for viscous liquids, such as those containing sucrose and glycerol, gas exchange is administered for 60 minutes over a hot plate set to 100^{o} C.

APPENDIX III

TRANSFORMATION OF METHANOSARCINA ACETIVORANS

Transformation of Methanosarcina acetivorans is mediated by the cationic detergent DOTAP (N-[1-(2,3-Dioleoyloxy)propyl]-N,N,N-trimethylammonium methyl-sulfate). The procedure for transformation invokes the use of specialized tools and techniques and is described here in detail.

Materials and reagents. Ensure that the following materials and reagents are prepared at least one day prior to transformation

- One Balch tube containing 10 ml sterile HS_{Met} medium per transformation
- Live cultures. Each transformation requires a Balch tube containing a 10 ml culture of the strain of interest. Transformation efficiency will be greatest if the culture is in late-exponential phase ($0.3 < A_{600} < 0.8$). For most strains, a late-exponential culture can be obtained by inoculating 10 ml HS_{Met} with 0.75 ml saturated culture, and incubating the resulting culture at 37^{o} C for 24 hours.
- Sterile, capped, 16 mm borosilicate tubes; three per transformation (Fig. A5D).
- Petri dishes; two per transformation. This needs to be inside the chamber for at least 24 hours.
- Anaerobic buffered sucrose solution (see notes).
- Sterilized single-use culture spreaders; two per transformation (Fig. A5). This needs to be inside the chamber for at least 24 hours.

- Plasmid DNA. Each transformation requires, 20 µl of DNA at a concentration of 100 ng/µl. This quantity can be obtained from a miniprep.

Transformation protocol. This is performed inside of the anaerobic chamber. The entire procedure takes approximately 1 hour of work and 20 hours of wait time.

- Transfer DNA and DOTAP solutions into the anaerobic chamber. The DOTAP vial's septum must be secured in place with a piece of tape so that the vacuum or the airlock does not remove the septum. Unopened vials do not need this, as they are crimped. Similarly, tubes containing DNA must have holes punched through their caps using a syringe needle.
- For each transformation, add 100 µl of buffered sucrose solution to a sterile 16 mm borosilicate tube.
- Add 30 µl of DOTAP to the sucrose solution. Mix by swirling.
- Add 20 µl of DNA to the sucrose-DOTAP solution. Mix by swirling. Incubate the mixture at room temperature for 15-30 minutes. To improve transformation efficiency, use the same quantity of DNA, but at a smaller volume and supplement the missing volume with sucrose solution.
- While incubating the sucrose-DOTAP-DNA mixture, harvest the strains to be transformed by transferring them to 16 mm borosilicate tubes and subjecting them to centrifugation. Discard the supernatants.
- Re-suspend the pelleted cells in 900 µl sucrose solution. Do so by pipetting gently
- After incubating the sucrose-DOTAP-DNA mixture for 15-30 minutes, add 850 µl of re-suspended cells to the sucrose-DOTAP-DNA mixture. Mix by swirling. Allow the resulting transformation reaction to incubate at room temperature for 2-4 hours.
- Recover the transformed strains by injecting entire transformation reactions into Balch tubes with 10 ml HS_{Met}. Incubate for 12-16 hours at 37^{o} C.

Plating protocol. Cultures of *Methanosarcina acetivorans* on solid media are stored under 4 psi of N_2:CO_2:H_2S (80:20:0.001%) inside of anaerobic jars (Fig. A5). Hydrogen sulfide functions to maintain the reducing environment necessary for methanogenic growth. Due to its propensity to react with palladium, hydrogen sulfide must be held within a secondary container. The gas manifold within the anaerobic chamber is used to facilitate gas exchange of anaerobic jars.

- Autoclave a bottle of solid media for one hour at 121^o C, liquid cycle
- Incubate the bottle in a 60^o C water bath for at least 20 minutes.
- Inject the bottle with the desired drug or additive.
- Transfer the bottle into the anaerobic chamber
- Pour 25-30 ml of media per plate into 10 cm Petri dishes.
- Transfer entire recovered transformation reactions to a 16 mm borosilicate tube.
- Using a pipette, transfer 100 µl of recovered transformation to a plate of solidified media. Spread the cells using a turntable and plastic cell spreader (Fig. A5)
- Pellet the recovered transformation reactions by centrifugation in a 16 mm borosilicate tube (Fig. A5).
- Pour off all the supernatant except for 50-100 µl
- Using a pipette, gently re-suspend the cell pellet in the residual supernatant and transfer 30-100 µl to a plate of solidified media. Spread the cells using a turntable and plastic cell spreader (Fig. A5)
- Place the plates inside of an anaerobic jar.
- Perform gas exchange on the jar by subjecting it to six cycles of evacuation (-10 inHg) and filling (4 psi) with N_2:CO_2:H_2S (80:20:0.001%).
- Store the sealed jar inside the 37^o C incubator within the anaerobic chamber. Colonies will emerge after 7-12 days.
- Prior to reopening the jar, perform gas exchange by subjecting it to six cycles of evacuation (-10 inHg) and filling (4 psi) with N_2:CO_2 (80:20%).

Notes

- Buffered sucrose solution contains 850 mM sucrose and 50 mM sodium bicarbonate and 30 µM cysteine (pH = 6.8). To prepare the solution, sucrose and sodium bicarbonate are dissolved in anaerobic water in a Pyrex bottle with GL-45 threads. This requires a bit of stirring. The solution is then transferred to a hot plate set to 100^{o} C. After heating the solution for one hour, the gas diffuser is used to purge the solution with N_2:CO_2 (80%:20%) for an additional hour with heat. The bottle is then stopped and transferred into the anaerobic chamber, where its solution is subjected to filter sterilization.
- Sterile technique relies on the judicious use of the foot-operated heating coil (Fig. F5). Objects including stopped Balch tubes, unstopped Balch tubes, unstopped solid media bottles, and 16 mm borosilicate tubes are routinely sterilized by holding them within 5 mm of the red-hot heating coil for approximately 10 seconds.
- Syringes are used frequently to transfer small volumes (< 3 ml) of anaerobic cell cultures. Accordingly, the anaerobic chamber is stocked with disposable, sterile, needled, 21-guage, 1 ml and 3 ml syringes.
- To open crimped DOTAP vials and Balch tubes inside of the anaerobic chamber, there are two sets of de-crimper tools. They look like pliers and are not depicted in this section. Generally, when transferring an entire culture from a Balch tube to a 16 mm borosilicate tube, the Balch tube is opened using the de-crimping tool and the culture is simply poured into the 16 mm borosilicate tube.

APPENDIX IV

RECOMBINANT EXPRESSION OF PROTEINS IN ESCHERICHIA COLI

Fourteen different plasmids were cloned for recombinant expression of various proteins of interest in *E. coli*. Although not all of these plasmids were employed successfully, the outcomes of different expression experiments are worth summarizing since they might be of value to the laboratory, which continues to work with these proteins (**Table A.7**).

Three plasmids were engineered to express SepCysS from *M. acetivorans* (pBR001, pBR002 and pBBR003). In each, SepCysS is modified to contain a polyhistidine tag. However, these constructs differ in terms of where the tag is located. Whether expressed in Rosetta2(DE3)pLysS (at 16^{o} C or 37^{o} C) or Arctic Express cells (at 10^{o} C), SepCysS was consistently observed to form inclusion bodies. The same result (insoluble protein) was observed for the *Methanosarcina mazei* SepCysS that was used previously for Cys-tRNA synthesis experiments conducted *in vitro* [84]. It should be noted, however, that recombinant expression of soluble *M. jannaschii* SepCysS has been reported, but expression of *M. jannaschii* homolog was not attempted here [85,86].

Several expression plasmids were engineered to express COG1900a (pBR004, pBR038, pBR096, pBR100 and pBR102). After initial results demonstrated that COG1900a-CBS from *M. acetivorans* (MA1821) is insoluble following expression alone, or together with NIL-Fer, the gene encoding the COG1900a-Fer fusion protein of *Thermotoga lettingae*, which lacks NIL, was cloned for expression. In each expression experiment the *Thermotoga lettingae* COG1900a-Fer protein exhibited prohibitively low levels of expression. Recently, all of the COG1900a-bearing plasmids were assayed for expression under anaerobic conditions. In each case, the outcomes of

anaerobic expression experiments resembled those observed for aerobic experiments—yielding either high levels of insoluble protein, or undetectable levels of protein in general.

Insoluble MA1821 from a pBR004-Rosetta2(BL21)pLysS expression could be dissolved in 8 M urea, 6 M guanidinium or 10% N-lauroylsarcosine, but not 1 mM CHAPS or 2% tritonX 100. Since urea and guanidinium are chaotropic agents, MA1821 is probably denatured under these conditions. However, when dissolved with N-lauroylsarcosine, a detergent which does not necessarily denature protein, it is unclear whether the native fold of MA1821 is preserved. When analyzed by SEC, after dialyzing N-lauroylsarcosine way from the protein, MA1821 migrated beyond the exclusion limit of the column (~600 kDa), suggesting that MA1821 had aggregated under these conditions. Furthermore, the fact that TritonX100 and CHAPS could not dissolve MA1821 following expression in *E. coli*, and that detergents were not needed to dissolve MA1821 following expression in *M. acetivorans*, suggests that MA1821 is not a membrane-associated protein. Attempts to refold urea-dissolved MA1821 were probably unsuccessful, yielding large particles (> 600 kDa).

The gene encoding COG1900d-Fer fusion from *M. jannaschii* was also cloned for expression (pBR113). However, this protein could only be expressed insolubly.

An original attempt to express MA1715 from pBR006 aerobically in *E. coli* resulted in mostly insoluble protein. This protein could be refolded as a monomer (judging by SEC elution time). However, a recent attempt to express MA1715 from the same plasmid under anaerobic conditions yielded significantly more soluble protein.

Andrew Gustafson, James Stanek and Camden Driggers performed a portion of the expression experiments. Andrew demonstrated that insoluble MA1821 resulting from expression in Rosetta2(DE3)pLysS with pBR004 could be dissolved with N-lauroylsarcosine. He was also the first to purify MA1715 following expression in Rosetta2(DE3)pLysS with pBR006. James performed refolding of insoluble MA1821 that resulted from expression in Rosetta2(DE3)pLysS with pBR004.Camden performed all expression experiments with pBR096, pBR100, pBR102 and pBR089 that were conducted at 10^{o}, 16^{o} and 37^{o} C. He also assisted with all anaerobic expression experiments.

Figure A.1. Stopper and cap for GL45 thread bottles.

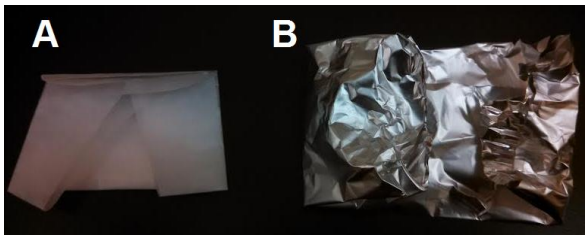


Figure A.2. Folded weigh-paper envelope alone (A) and supported with aluminum foil (B).

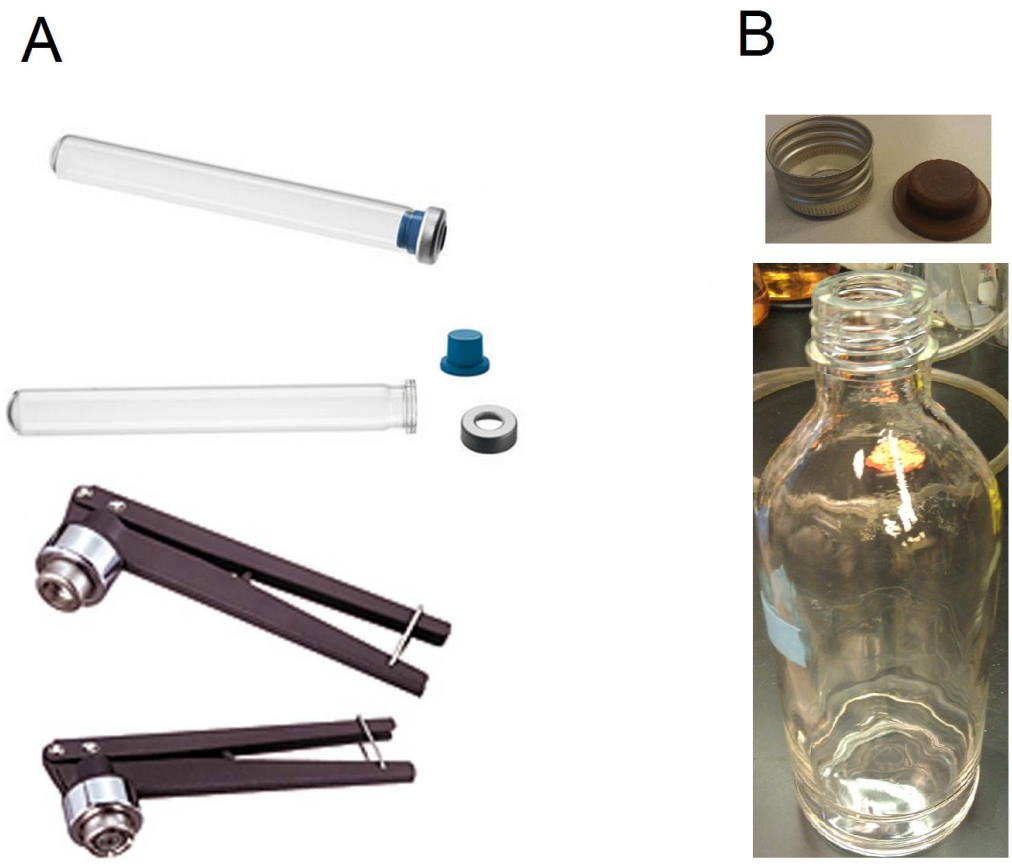


Figure A.3 Anaerobic media containers and tools. (A) Balch tubes are sealed using a rubber stopper secured by an aluminum crimp. The crimp is fastened to the tube using a crimping tool. An additional de-crimping tool is required to remove the crimp. (B) Solid media bottles are sealed with a rubber stopper secured in place with an aluminum screw cap. The Balch tube photograph was obtained from www.chemglass.com. The crimping and de-crimping photograph was obtained form www.medicalsuppliers.com

Table A.1 HS media solution A

	Mass (0.5-liter prep)	Mass (1-liter prep)	Mass (1.5-liter prep)	Mass (2-liter prep)	MW	[final]
NaCl	11.7 g	23.4 g	35.1 g	46.8 g	58.45	0.4 M
$NaHCO_3$	1.9 g	3.8 g	5.7 g	7.6 g	84.01	45 mM
KCl	0.5 g	1.0 g	1.5 g	2.0 g	74.56	13 mM
0.1% resazurin	0.5 ml	1.0 ml	1.5 ml	2.0 ml	251.2	4 µM
Trace elements*	5.0 ml	10.0 ml	15.0 ml	20.0 ml		
Vitamin solution*#	5.0 ml	10.0 ml	15.0 ml	20.0 ml		
De-ionized Water	500 ml	750 ml	1100 ml	1300 ml		

*Contents are described in the notes section below.
#For HS_{DTT} medium, use vitamin solution lacking thiamine, biotin and thioctic acid

Table A.2 HS media solution B

		Mass (0.5-liter prep)	Mass (1-liter prep)	Mass (1.5-liter prep)	Mass (2-liter prep)	MW	[final]
$MgCl_2 \cdot 6H_2O$		5.5 g	11.0 g	16.5 g	22.0 g	203.3	54 mM
$CaCl_2 \cdot 2H_2O$		0.15 g	0.3 g	0.45 g	0.6 g	147.0	2 mM
De-ionized Water	200 ml	350 ml	600 ml	800 ml			

Table A.3. Additional components for HS_{Met}

	Mass (0.5-liter prep)	Mass (1-liter prep)	Mass (1.5-liter prep)	Mass (2-liter prep)	MW	[final]
Methionine	0.20 g	0.40 g	0.60 g	0.80 g	149.2	2.7 mM
Cysteine-HCl	0.25 g	0.50 g	0.75 g	1.00 g	175.6	2.8 mM
$Na_2S \cdot 9H_2O$	0.05 g	0.10 g	0.15 g	0.20 g	240.2	0.4 mM
NH_4Cl	0.50 g	1.00 g	1.50 g	2.00 g	53.5	19 mM

Table A.4. Additional components for HS_{DTT}

	Mass (0.5-liter prep)	Mass (1.0-liter prep)	Mass (1.5-liter prep)	Mass (2.0-liter prep)	MW	[final]
Dithiothreitol (DTT)	0.11 g	0.23 g	0.34 g	0.46 g	154	1.5 mM
NH_4Cl	0.5 g	1.0 g	1.5 g	2.0 g	53.5	19 mM

Table A.5. Vitamin solution contents per liter aqueous stock

	Mass (mg)	**MW (g/mol)**	**[stock] (μM)**	**[final] (nM)**
p-aminobenzoic acid	10	137.1	72.9	729
Nicotinic acid	10	123.1	81.2	812
Ca pantothenate	10	238.3	41.9	419
Pyridoxine	10	205.6	48.6	486
Riboflavin	10	376.4	26.6	266
Thiamine HCl	10	337.3	29.6	296
a-Biotin	5	244.3	20.4	204
Folic acid	5	441.1	11.3	113
a-Liopic acid	5	206.3	24.2	242
Vitamin B_{12}	5	1355.4	3.7	37

Table A.6. Trace element solution contents per liter aqueous stock

	Mass (g)	**MW (g/mol)**	**[stock] (mM)**	**[final] (μM)**
Nitriloacetic acid (trisodium salt)	1.5	257.1	5.8	58
$Fe(NH_4)_2(SO_4)_2$	0.8	392.1	2	20
Na_2SeO_3	0.2	172.9	1.1	11
$CoCl_2 \cdot 6H_2O$	0.1	237.9	0.4	4
$MnSO_4 \cdot H_2O$	0.1	169	0.6	6
$Na_2MoO_4 \cdot 2H_2O$	0.1	241.9	0.4	4
$Na_2WO_4 \cdot 2H_2O$	0.1	329.9	0.3	3
$ZnSO_4 \cdot 7H_2O$	0.1	287.5	0.3	3
$NiCl_2 \cdot 6H_2O$	0.1	237.7	0.4	4
H_3BO_3	0.01	61.83	0.16	1.6
$CuSO_4 \cdot 5H_2O$	0.01	249.7	0.04	0.4

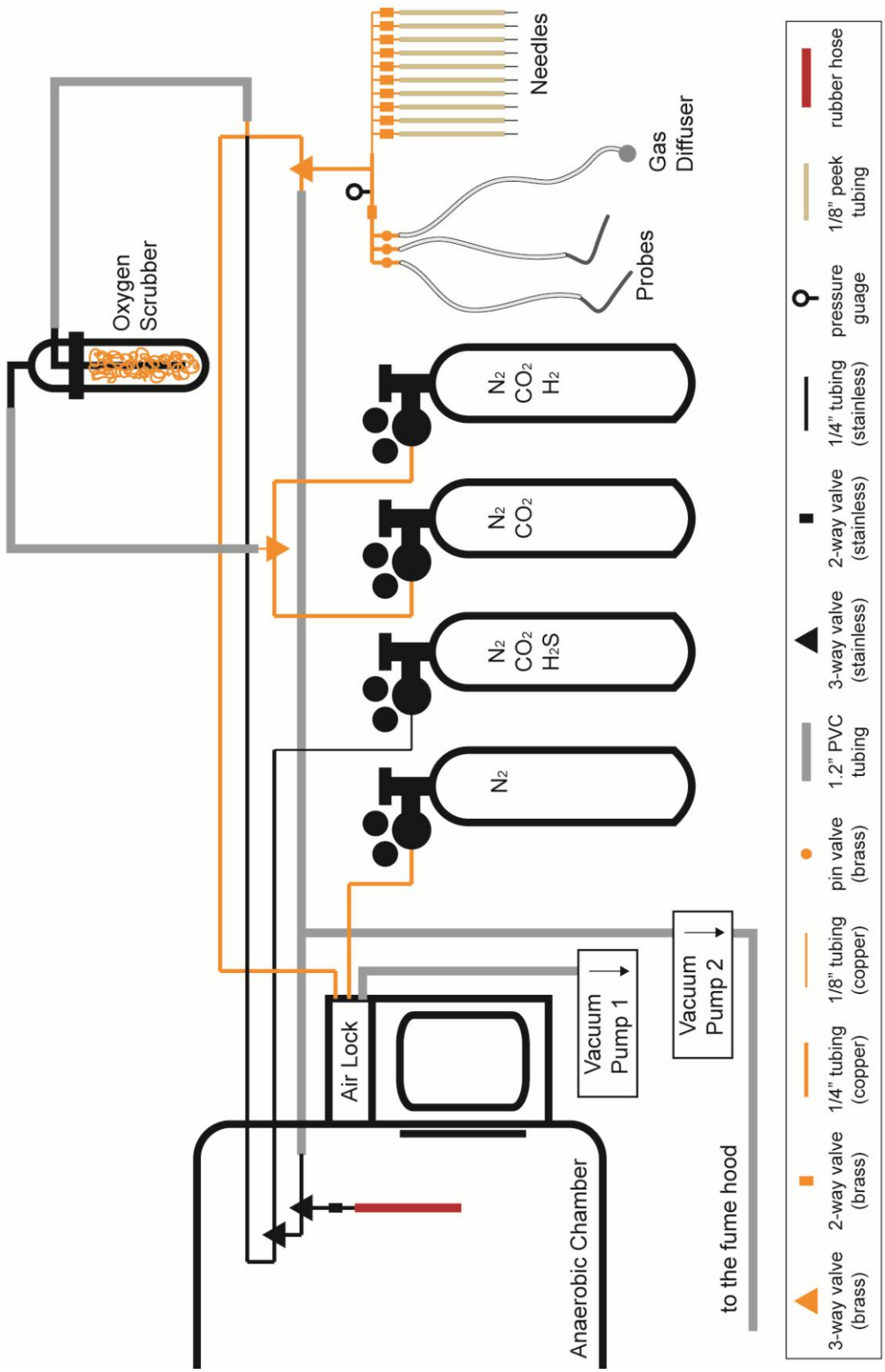


Figure A.4 The gas manifold system

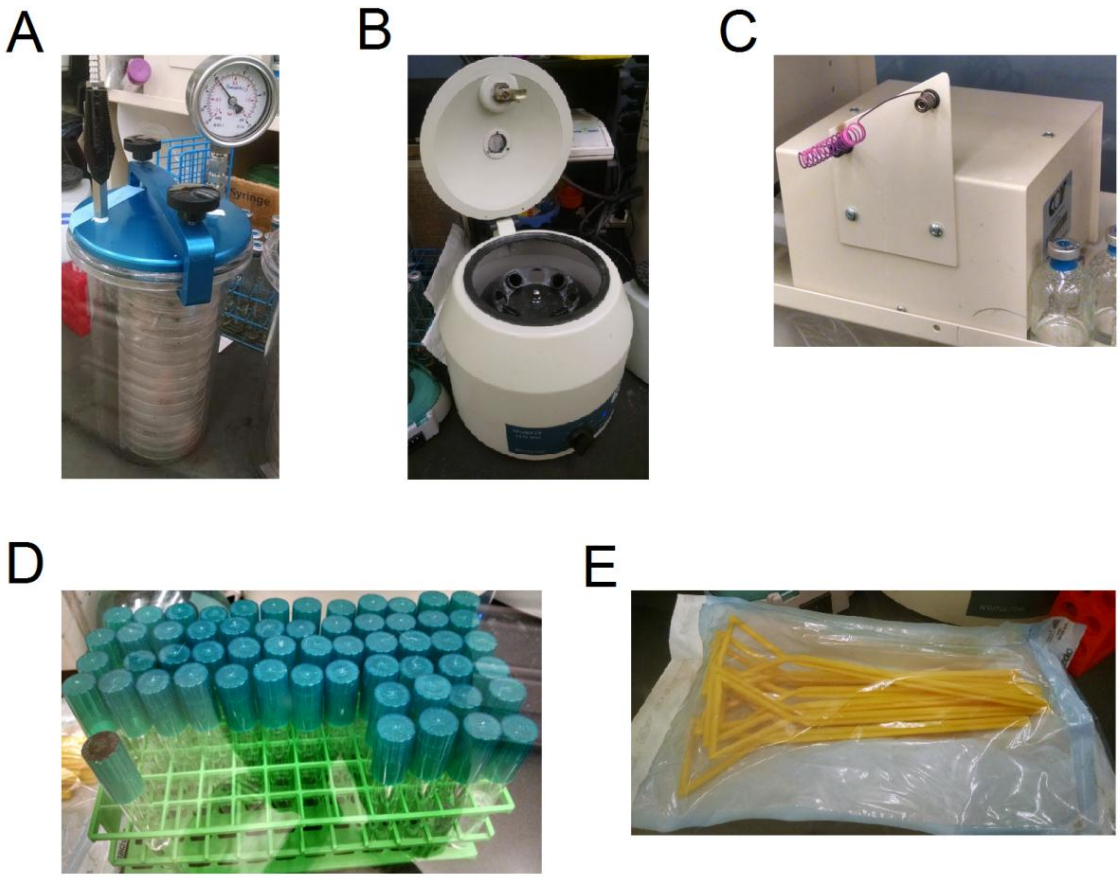


Figure A.5. Special materials for transformation of M. acetivorans. Transformation reactions are spread onto agar solidified media with plastic culture spreaders (E) and a turntable. The culture spreaders can be reused after autoclaving and are shown inside of a sterilization pouch, which keeps the spreaders sterile when passed through the air lock. Cultures on solid media are stored under 4 psi of N_2:CO_2:H_2S (80:20:0.001%) inside of an anaerobic jar (A). Cells present in liquid culture are harvested by centrifugation using a clinical centrifuge (B), which accommodates four capped 16X130 mm borosilicate tubes (D). In the absence of a Bunsen burner, sterile manipulations are carried out with the assistance of a heating coil (C), which is operated by a foot-pedal control.

Table A.7 Summary of *E. coli* protein expression constructs

plasmid	proteins encoded	source	parent plasmid	host strain	conditions: result
pBR001	SepCysS-X_{14}-H_6	*ma0722*	pET-22b(+)	Rosetta2(DE3)pLysS	● 24 hr at 16° C: >50 mg insol. ● 5 hr at 37° C: >50 mg insol.
pBR002	H_{10}-SepCysS	*ma0722*	pET-16b	Rosetta2(DE3)pLysS Arctic Express	● 24 hr at 16° C: >50 mg insol. ● 5 hr at 37° C: >50 mg insol. ● 24 hr at 10° C: >50 mg insol
pBR003	SepCysS-H_6	*ma0722*	pET-22b(+)	Rosetta2(DE3)pLysS Arctic Express	● 24 hr at 16° C: >50 mg insol. ● 5 hr at 37° C: >50 mg insol. ● 24 hr at 10° C: >50 mg insol
pBR007	H_{10}-SepRS	*ma0090*	pET-16b	Rosetta2(DE3)pLysS	● 5 hr 37° C: 8-50 mg sol.
pBR008	H_{10}-CysRS	*ma0749*	pET-16b	Rosetta2(DE3)pLysS	● 5 hr 37° C: 16 mg, sol.
pBR004	COG1900a-CBS-H_6	*ma1821*	pET-22b(+)	Rosetta2(DE3)pLysS Arctic Express	● 24 hr at 16° C: >50 mg insol. ● 5 hr at 37° C: >50 mg insol., which could be dissolved with 10% N-lauroylsarcosine, 8 M urea or 6 M guanadinium, but could not be dissolved with CHAPS or TritonX 100. ● refolding yields soluble complex > 600 kDa ● 24 hr at RT, anaerobic: >50 mg, insol. ● 24 hr at 10° C: >50 mg insol
pBR005	NIL-Fer-H_6	*ma1822*	pET-22b(+)	Rosetta2(DE3)pLysS	● 24 hr at 16° C: >50 mg insol. ● 5 hr at 37° C: >50 mg insol.
pBR038	COG1900a-CBS; NIL-Fer-H_6	*ma1821-22*	pET-22b(+)	Rosetta2(DE3)pLysS	● 5 hr at 37° C: >50 mg, insol. ● 24 hr at RT, anaerobic: > 5 mg insol.
pBR096	COG1900a-Fer-H_6	*tlet1363*	pET-22b(+)	BL21(DE3)pLysS	● 24 hr at 16° C: nd ● 5 hr at 37° C: nd ● 24 hr at RT, anaerobic: nd
pBR100	H_6-SUMO-COG1900a-Fer	*tlet1363*	pET-28b(+)-SUMO*	BL21(DE3)pLysS Arctic Express	● 24 hr at 16° C: nd ● 5 hr at 37° C: nd ● 24 hr at RT, anaerobic: nd ● 24 hr at 10° C: < 1 mg, sol., which could not be separated from co-purifying protein
pBR102	H_6-SUMO-COG1900a-Fer; COG2122a-H_6	*tlet1363-62*	pET-28b(+)-SUMO*	BL21(DE3)pLysS Arctic Express	● 24 hr at 16° C: nd ● 5 hr at 37° C: nd ● 24 hr at RT, anaerobic: nd ● 24 hr at 10° C: < 1 mg, sol., which could not be separated from co-purifying protein
pBR113	H_6-SUMO-COG1900d-Fer	*mj1681*	pET-28b(+)-SUMO*	BL21(DE3)	● 5 hr at 37° C: >50 mg, insol. ● 24 hr at RT, anaerobic: >50 mg, insol.
pBR006	COG2122a-H_6	*ma1715*	pET-22b(+)	Rosetta2(DE3)pLysS	● 5 hr at 37° C: 1 mg sol., >50 mg insol. ● Refolding of insol. yields sol. monomer, verified by SEC ● 24 hr at RT, anaerobic: > 5 mg sol.
pBR089	H_6-SUMO-COG2122a	*dvu1097*	pSGX4	BL21(DE3)pLysS	● 24 hr at 16° C: ~3 mg, sol. ● 24 hr at RT, anaerobic: nd

mg, milligrams protein per liter culture
nd, expression not detected
ma, *Methanosarcina acetivorans*
tlet, *Thermotoga lettingae*
mj, *Methanocaldococcus jannaschii*
dvu, *Desulfovibrio vulgaris*
X_{14}, 14-amino acid linker
H_n, polyhistidine tag
SEC, size exclusion chromatography
sol., soluble protein
insol., insoluble protein
*, see reference [115]