Short Methionine Synthases

Morgan N. Price and Adam P. Arkin
Lawrence Berkeley National Lab

Abstract

Although most organisms synthesize methionine from homocysteine and 5-methyltetrahydrofolate, we propose that many bacteria and archaea use alternative enzymes that rely on other methyl donors. These methionine synthases are homologous to MetE, which uses 5-methyltetrahydrofolate, but they lack the N-terminal folate-binding domain of MetE. We describe four families of short MetE-like methionine synthases that are distantly related to each other (under 30% amino acid identity). The first family, MesA, is found only in methanogens, and was previously shown to be a methionine synthase *in vitro.* It is thought to obtain methyl groups from corrinoid (vitamin B12-binding) proteins. We describe two additional families, MesB and MesC, that are found only in anaerobic organisms that have the Wood-Ljungdahl pathway. We propose that most of these proteins obtain methyl groups from that pathway. The final family, MesD, is found only in aerobic bacteria and genetic data suggests that it does not use 5-methyltetrahydrofolate as a methyl donor. We propose that MesD obtains its methyl groups from an uncharacterized protein family (DUF1852).

Introduction

Methionine is required for protein synthesis and is also a precursor to S-adenosylmethionine, which is the methyl donor for most methyltransferases and is required for polyamine biosynthesis. Methionine is synthesized from aspartate by reduction and sulfhydrylation to homocysteine, which is followed by the transfer of a methyl group to give methionine (Figure 1). There are two well-studied forms of methionine synthase, both of which obtain the methyl group from 5-methyl-tetrahydrofolate (5-methyl-THF): MetH, which uses cobalamin (vitamin B12) as a cofactor, and MetE, which does not require cobalamin. MetH is about 40 times faster than MetE (Goulding, Postigo, and Matthews 1997; González et al. 1996). *Escherichia coli,* which cannot synthesize cobalamins, has both enzymes. There are also methyltransferases that convert homocysteine to methionine by using methylated nutrients such as glycine betaine or S-methylmethionine (enzyme commission numbers 2.1.1.5 or 2.1.1.10). Here we will focus on the synthesis of methionine without special nutrient requirements.

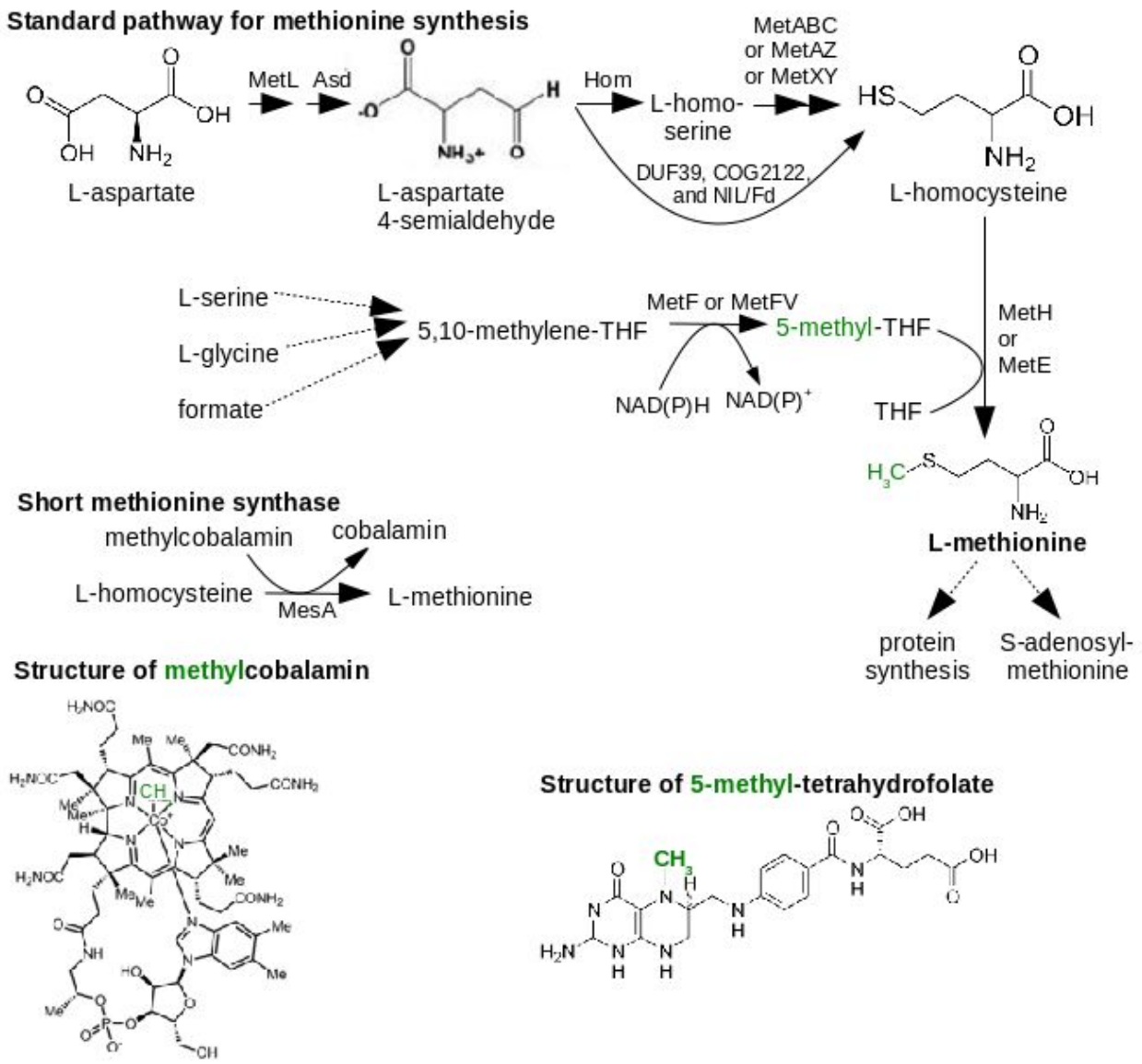

Figure 1: Overview of methionine synthesis. We show the standard pathway with 5-methyl-THF, the reaction catalyzed by the short methionine synthase MesA, and the structures of methylcobalamin and 5-methyl-THF. Although THF is shown with a single glutamyl residue, in the cell, THF is usually polyglutamylated. Similarly, cobalamin has 5,6-dimethylbenzamidazole as the lower ligand, but many organisms use cobinamides with other lower ligands. In *Methanobacterium* (the source of MesA for biochemical studies), the lower ligand is 5-hydroxybenzimidazole (Kräutler, Moll, and Thauer 1987), and the cobalamin is probably bound to a corrinoid protein.

Besides MetH and MetE, two other types of methionine synthase have been reported. First, a short MetE-like protein from the methanogen *Methanobacterium thermoautotrophicum* has been studied biochemically (Schröder and Thauer 1999). We will call this protein MesA (methionine synthase short group A; Swiss-Prot identifier METE_METTM). *In vitro*, MesA transfers methyl

groups from methylcobalamin to homocysteine to form methionine (Figure 1), but it has no activity with 5-methyl-THF or 5-methyltetrahydromethanopterin as substrates (Schröder and Thauer 1999). (Tetrahydromethanopterin is a cofactor for methanogenesis that is similar to THF.) Because MesA has a very weak affinity for methylcobalamin (the Michaelis-Menten constant is above 20 mM), and because most of the cobalamin in methanogens is bound to corrinoid proteins, the physiological substrates of MesA are probably methyl corrinoid proteins (Schröder and Thauer 1999). MesA is homologous to the C-terminal catalytic domain of MetE and lacks the N-terminal domain of MetE, which is involved in binding folate (Pejchal and Ludwig 2005). It might seem surprising that MesA accepts methyl groups from cobalamin when it is homologous to the cobalamin-independent enzyme, but the catalytic mechanisms of MetE and MetH are similar: both MetE and MetH rely on a zinc cofactor to deprotonate the sulfur atom of homocysteine and activate it as a nucleophile (Matthews et al. 2003).

Second, a genetic study identified an unusual methionine synthase in *Acinetobacter baylyi* (de Berardinis et al. 2008). ACIAD3523 is required for methionine synthesis in the absence of cobalamin, and so is the adjacent gene ACIAD3524. Although ACIAD3523 was originally annotated as a MetE protein and is so described in the genetic study, ACIAD3523 lacks the N-terminal folate-binding domain, and it is distantly related to the C-terminal catalytic domain of MetE (under 30% identity). The associated protein ACIAD3524 belongs to the uncharacterized family DUF1852 (Pfam PF08908; (El-Gebali et al. 2019)). (DUF is short for domain of unknown function.) Homologs of ACIAD3523 are found adjacent to genes encoding DUF1852 in diverse bacteria, and we will call this family MesD (methionine synthase short with DUF1852). We will present genetic evidence that 5-methyl-THF is not the methyl donor for MesD.

We will also show that diverse anaerobic bacteria and archaea lack all of the known forms of methionine synthase, yet they grow in minimal defined media. We identified two additional families of short MetE-like proteins in these organisms, which we call MesB and MesC. Analysis of functional residues suggests that MesB and MesC are methionine synthases. Based on comparative genomics, we predict that most MesB and MesC proteins obtain methyl groups from the iron-sulfur corrinoid protein of the Wood-Ljungdahl pathway.

Results and Discussion

Anaerobic organisms that lack all known forms of methionine synthase

We previously ran the GapMind tool for reconstructing amino acid biosynthesis against 150 genomes of bacteria and archaea that grow in defined minimal media without any amino acids present (Price, Deutschbauer, and Arkin 2019). 16 of the 150 organisms had no candidates for any of the known varieties of methionine synthase (MetH, MetH split into two parts, three-part MetH (Price, Zane, et al. 2018), MetE, MesA, or MesD).

We searched for additional candidates for methionine synthase by using PSI-BLAST (Altschul et al. 1997) and the profile for protein family COG620 (Tatusov et al. 2003). COG620 matches both the N-terminal (folate-binding) and C-terminal (homocysteine-activating) domains of MetE, as well as MesA and MesD. In the hyperthermophilic archaea *Pyrolobus fumarii* 1A, we identified two hits to COG620, which correspond to the two domains of MetE. PYRFU_RS09465 (WP_048192068.1) contains the N-terminal domain (Meth_synt_1 in PFam; PF08267) and PYRFU_RS01495 (WP_014025837.1) contains the C-terminal domain (Meth_synt_2 in PFam; PF01717). Although the two genes are not near each other in the genome of *Pyrolobus fumarii*, close homologs are adjacent in other hyperthermophilic archaea (such as *Aeropyrum pernix* K1), which suggests that they comprise a split MetE.

The other 15 organisms with missing methionine synthases are strictly anaerobic. They include diverse sulfate-reducing bacteria (both δ-Proteobacteria and Firmicutes), a dehalogenating bacterium (*Dehalococcoides mccartyi*), methanogens from the genus *Methanosarcina*, and an iron-reducing archaea (*Ferroglobus placidus*).

In these anaerobes, we propose that short MetE-like proteins transfer methyl groups from corrinoid proteins to homocysteine to form methionine (as proposed for MesA). In 13 of the 15 anaerobes, we found a short MetE by using COG0620 and PSI-BLAST. Candidates in the other two organisms were identified by using protein BLAST with the original 13 candidates as queries. (The two additional candidates were 40%-41% identical to DET0516 from *Dehalococcoides mccartyi* 195.) The candidate proteins ranged in length from 317 to 361 amino acids. For comparison, MesA (SwissProt METE_METTM) is 309 amino acids and MetE from *E. coli* is 753 amino acids. By sequence similarity, the short MetE proteins cluster into two groups, a mostly-bacterial group which we will term MesB and an archaeal group which we will term MesC. MesB and MesC are distantly related to each other or to MetE, MesA, or MesD, with pairwise identities of under 30%.

MesB and MesC are probably methionine synthases

MesB and MesC have conserved the catalytic residues for methyl transfer that are known from studies of MetE. Specifically, MetE activates homocysteine via a zinc thiolate intermediate. The residues that coordinate the zinc atom in *E. coli*'s MetE are His641, Cys643, Glu665, and Cys726 (Taurog and Matthews 2006). These residues are conserved in MesA and MesD, which have experimental evidence supporting their roles as methionine synthases, and similar residues are present in MesB and MesC as well (Figure 2). MesB proteins have an aspartate instead of Glu665, and MesC proteins have a histidine instead of Cys643, but these substitutions are probably compatible with zinc binding. A few MesC proteins (from *Methanosarcina*) have a tyrosine aligning to Cys643, and we are not sure if this is compatible with zinc binding and methyltransferase activity. However, these genomes include other representatives of MesC, and those other proteins do have histidines aligning to Cys643.

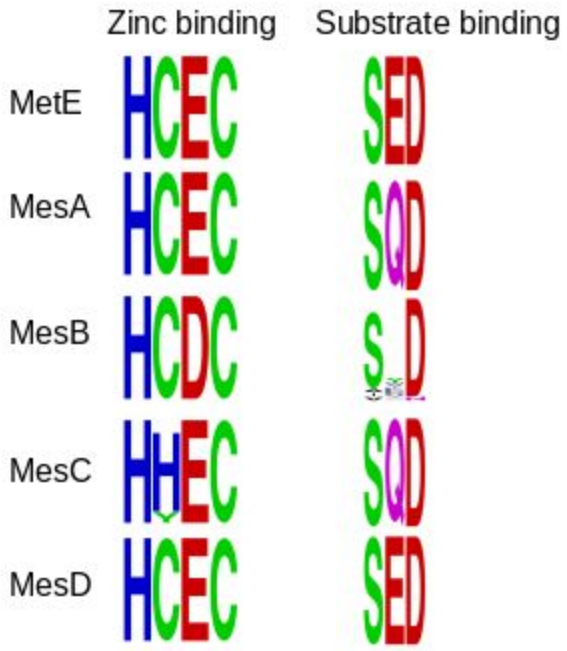

Figure 2: Functional residues of short methionine synthases. We show sequence logos (Crooks et al. 2004) for the zinc-coordinating and substrate-binding residues of each family of methionine synthases. The height of each position shows its conservation within the family, as measured by information content or bits. In MetE from *E. coli*, the zinc-coordinating residues are H641, C643, E665, and C726, and the substrate-binding residues are S433, E484, and D600.

We then asked if MesB and MesC were likely to bind homocysteine. Structural data suggests that several side chains in MetE form hydrogen bonds with the amino or carboxyl groups of homocysteine or methionine ((Ferrer et al. 2004); PDB:1U1J). (In the *E. coli* residue numbering, Ser433 binds the carboxyl group and Glu484 and Asp600 bind the amino group.) Similarly, in a crystal structure for a MesD protein bound to selenomethionine, which is an analog of methionine (PDB:3RPD), the corresponding side chains are in proximity to the amino and carboxyl groups of selenomethionine (Ser22, Glu73, and Asp188 in 3RPD). As shown in Figure 2, MesA and MesC have similar residues, but with a glutamine instead of Glu484. Glutamine could also form a hydrogen bond with the amino group of homocysteine, so we propose that MesA and MesC bind homocysteine (or methionine) in the same manner that MetE and MesD do. Since MesA is active with homocysteine as a substrate (Schröder and Thauer 1999), the identity of these residues in MesC suggests that MesC is also a methionine synthase.

In MesB, the serine and aspartate residues are mostly conserved, but the glutamate is not. The region corresponding to Glu484 (*E. coli* numbering) is quite variable among MesB proteins (i.e., around Val68 in DET0516). Nevertheless, residues that bind both the amino groups and carboxy groups of homocysteine are conserved, so we expect that MesB proteins are methionine synthases.

Furthermore, a conserved gene cluster links MesB to methionine synthesis. For example, MesB from *Dehalococcoides mccartyi* 195 (DET0516) is in a cluster of genes related to S-adenosylmethionine (SAM) synthesis and salvage (Figure 3). These include SAM synthetase (*metK*); adenosylhomocysteinase (*ahcY*), which breaks down the byproduct of SAM-dependent methylases; and two genes involved in converting *S*-methyl-5'-thioadenosine, which is a byproduct of other SAM-dependent enzymes, back to methionine. This cluster is conserved in other species of *Dehalococcoides* and in a representative of a related genus (*Dehalogenimonas lykanthroporepellens* BL-DC-9). Because methionine is the precursor to SAM, the conserved clustering of MesB with SAM-related genes suggests that MesB proteins are involved in methionine synthesis.

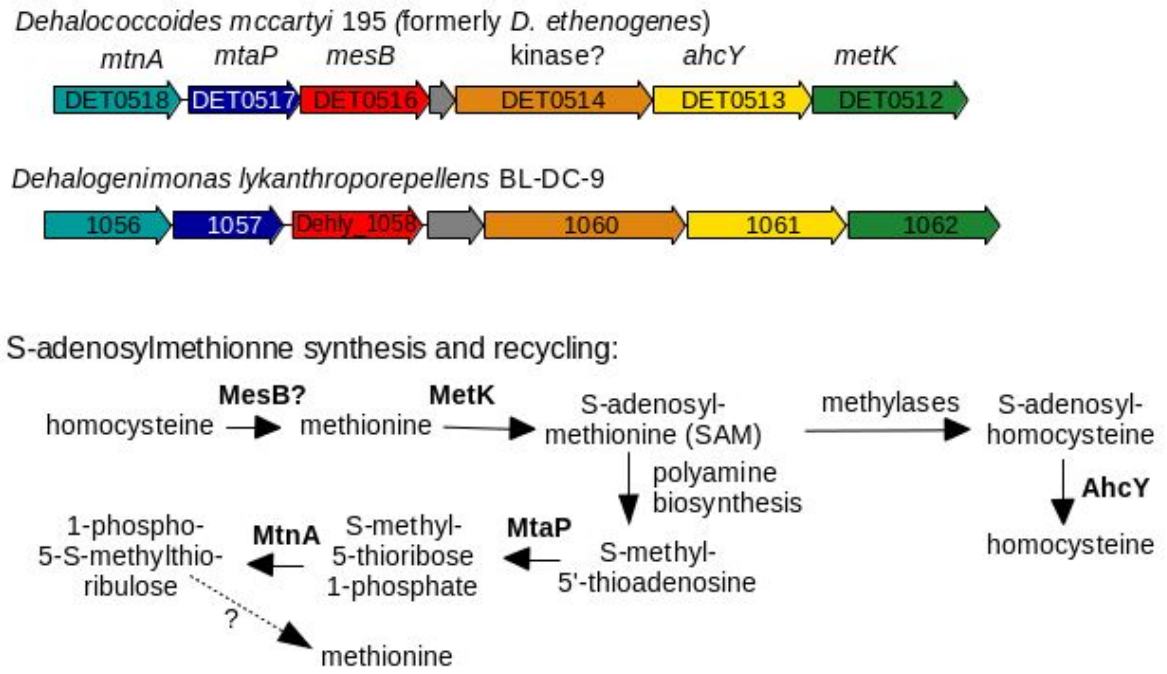

Figure 3: Conserved clustering of *mesB* with genes for the synthesis and recycling of S-adenosylmethionine. The fate of the 1-phospho-5-S-methylthioribulose in *Dehalococcoides mccartyi* is uncertain. *D. mccartyi* does not contain either of the two known anaerobic pathways for recycling this compound, which begin with 5-(methylthio)ribulose-1-phosphate isomerase (North et al. 2016) or 5-(methylthio)ribulose-1-phosphate aldolase (North et al. 2017). Gene drawings were modified from PATRIC (North et al. 2017).

These results suggest that MesB and MesC are zinc-dependent enzymes, and that they are likely to bind homocysteine and act as methionine synthases. MesB and MesC are also the only candidates we identified for the missing methionine synthases in these organisms. But since these proteins lack the folate-binding domain of MetE, where do the methyl groups come from?

MesB is linked to the iron-sulfur corrinoid protein of the Wood-Ljungdahl pathway

In *Dehalococcoides mccartyi*, which appears to use MesB as its sole methionine synthase, labeling experiments demonstrated that the methyl group of methionine is derived from the methyl group of acetate (Zhuang et al. 2014). The acetate is probably broken down by acetyl-CoA decarbonylase (also known as acetyl-CoA synthase). Most organisms with acetyl-CoA decarbonylase/synthase use it as part of the Wood-Ljungdahl pathway to split acetate to 5-methyl-THF and carbon dioxide, or (in reverse) to convert CO_2 to acetyl-CoA, but *D. mccartyi* lacks the CO dehydrogenase (*acsA*) or the methyltransferase that would transfer methyl groups to THF (*acsE*). Labeling experiments showed that the carboxyl carbon of acetate is released as CO, which confirms that CO dehydrogenase is truly absent (Zhuang et al. 2014). Acetyl-CoA decarbonylase (*acsB*) is expected to transfer the methyl group from acetyl-CoA to a corrinoid iron-sulfur protein (CoFeSP, encoded by *acsCD*). Growing *D. mccartyi* with external methionine leads to a 50% reduction in the release of CO (Zhuang et al. 2014), which suggests that providing methyl groups to methionine is the major role of the incomplete Wood-Ljungdahl pathway (Figure 4).

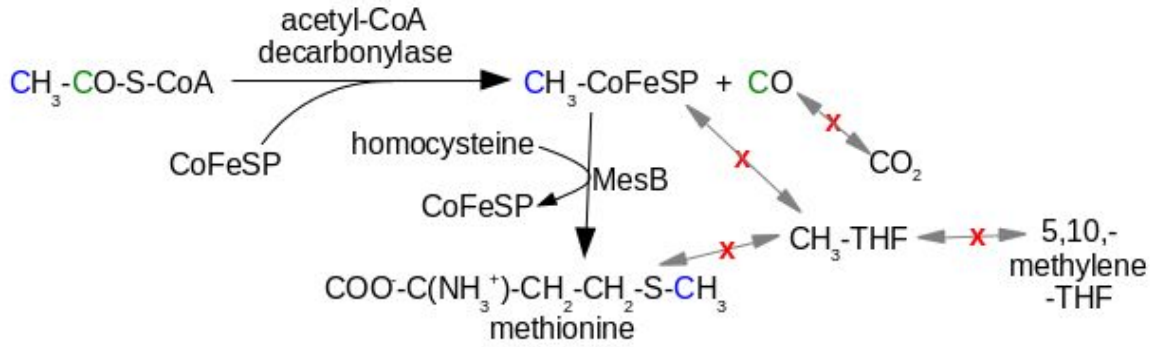

Figure 4: The proposed pathway for methionine synthesis in *Dehalococcoides mccartyi*. Steps that are usually present in organisms with the Wood-Ljungdahl pathway, but appear to be absent from *D. mccartyi*, are shown with a red x.

We propose that the MesB in *D. mccartyi* obtains methyl groups directly from the CoFeSP protein. This would explain why *D. mccartyi* lacks all of the known enzymes that transfer methyl groups to or from THF. We already mentioned that *D. mccartyi* lacks the methyltransferase from the Wood-Ljungdahl pathway (*acsE*), and *D. mccartyi* also seems to lack 5,10-methylene-THF reductase (MetF), which interconverts 5-methyl-THF and 5,10-methylene-THF. In *E. coli* and in most other bacteria, MetF is required for methionine synthesis because it is the only pathway from central metabolism to methyl-THF. If methyl-THF is not the methyl donor for methionine synthesis, then this explains why MetF is not necessary. In fact, we predict that cells of *D. mccartyi* do not contain methyl-THF.

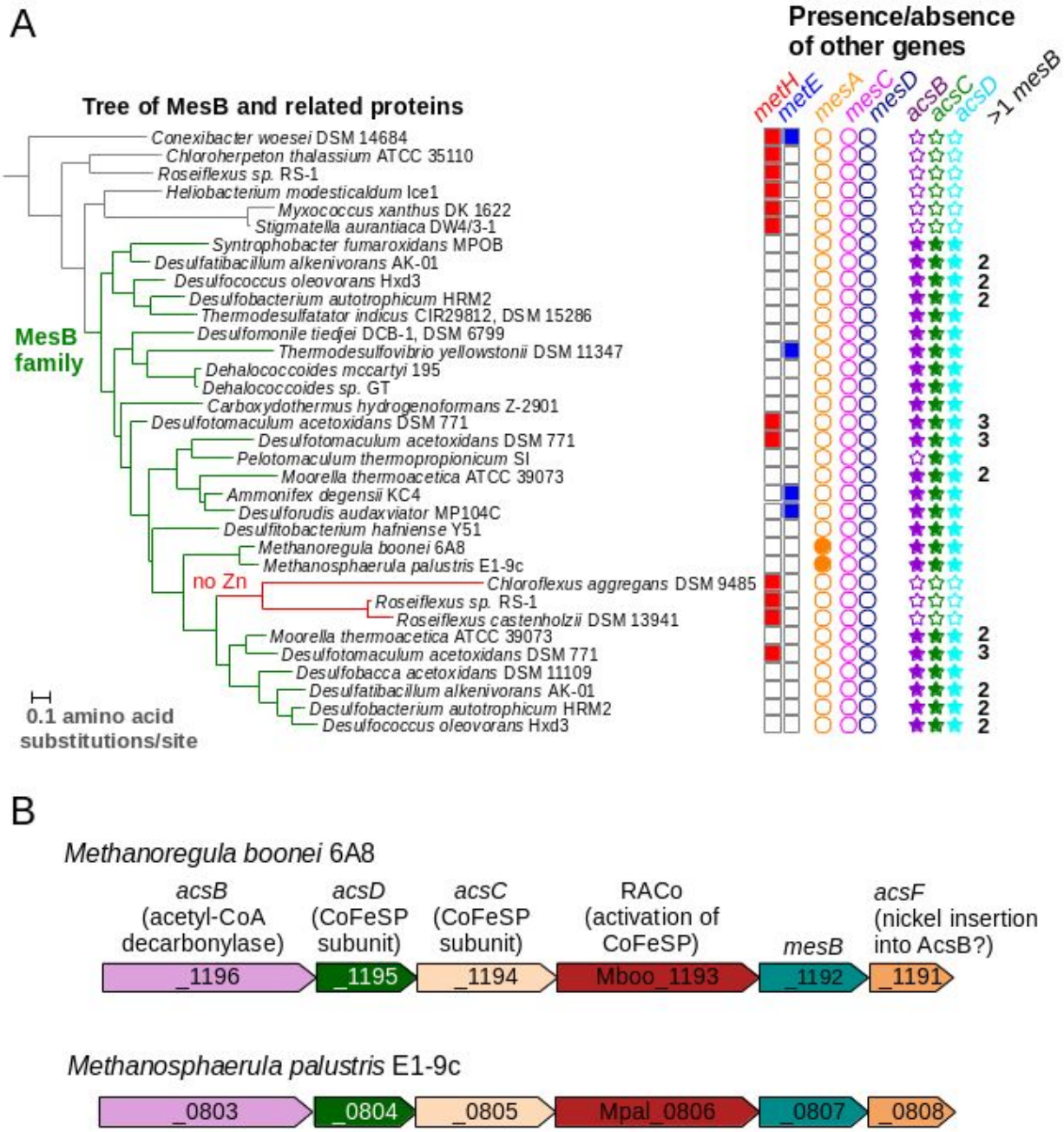

Figure 5: Comparative genomics links MesB to the Wood-Ljungdahl pathway. (A) A phylogenetic tree of MesB and related proteins from MicrobesOnline (Dehal et al. 2010). The MesB family is highlighted in green and a subfamily that lacks Zn-coordinating residues is highlighted in red. On the right, filled symbols indicate the presence in that genome of other methionine synthases or of the Wood-Ljungdahl pathway (*acsBCD*, also known as *cdhCED*). If the genome contains more than one *mesB* gene, we show the number. (B) Conserved clustering of *mesB* with genes from the Wood-Ljungdahl pathway. Gene drawings were modified from MicrobesOnline.

Two additional pieces of evidence link MesB to the Wood-Ljungdahl pathway. First, all of the genomes with MesB also contain the CoFeSP protein. There are homologs of MesB in some anoxygenic phototrophs (Chloroflexales) that lack the Wood-Ljungdahl pathway (i.e., Chloroflexus and Roseiflexus in Figure 5A), but the MesB-like proteins of Chloroflexales lack the zinc-coordinating residues, so they must have a different function.

Second, although MesB is primarily found in bacteria, it is also found in a few methanogens. In methanogens, the *mesB* genes are in an apparent operon with *acsCD* (Figure 5B). The conserved operon suggests a direct functional relationship.

Overall, we predict that most of the MesB proteins use the CoFeSP protein from the Wood-Ljungdahl pathway as the methyl donor. Some of the bacterial genomes with MesB have 2-3 copies (Figure 5A), and some of these paralogs might bind another corrinoid protein.

Potential methyl donors for MesC

MesC is found only in anaerobic archaea that encode the CoFeSP protein of the Wood-Ljungdahl pathway. This includes methanogens from the orders Methanosarcina and Methanotrichales, iron-reducing *Ferroglobus* and *Geoglobus*, and sulfate-reducing *Archaeoglobus*. The CoFeSP protein is the most obvious candidate for the methyl donor.

Although most of these organisms have one short methionine synthase, representatives of the genus *Methanosarcina* have 2-3 copies of MesC (and no other methionine synthases). The multiple MesC proteins within *Methanosarcina* genomes seem to have arisen by lineage-specific duplications: they cluster together in a phylogenetic tree, and two of the three paralogs are near each other in the genome. *Methanosarcina* can grow on various methyl donors via an array of specialized corrinoid proteins (Fu and Metcalf 2015), so we speculate that the paralogs might accept methyl groups from other corrinoid proteins besides CoFeSP.

MesA is linked to methanogenesis

The MesA family is found solely in methanogens. The closest homologs from other organisms are the catalytic domains of split MetE proteins. For example, the characterized MesA protein (METE_METTM) is 38% identical to the catalytic part of the split MetE from *Pyrolobus fumarii* that we mentioned previously

Although most of the methanogens that have MesA encode the Wood-Ljungdahl pathway, some do not, for instance, *Methanobrevibacter smithii.* This suggests that the CoFeSP is not the physiological methyl donor for MesA. Also, as shown in Figure 5A, some methanogens contain both MesA and MesB, which suggests that the two methionine synthases might use different methyl donors. Because MesA is linked to methanogenesis, and because methyl-tetrahydromethanopterin:coenzyme M methyltransferase is the key corrinoid protein in methanogenesis, we speculate that this is the methyl donor.

MesD does not require 5-methyl-THF as a methyl donor

In contrast to MesA, MesB, and MesC, the short methionine synthase MesD and the accompanying DUF1852 are found only in aerobic bacteria. These include both strict aerobes and facultative anaerobes. Close homologs of MesD, with over 70% identity to ACIAD3523, are present in six different divisions of bacteria (α-Proteobacteria, β-Proteobacteria, γ-Proteobacteria, Actinobacteria, Bacteroidetes, and Verrucomicrobia).

If MesD accepts methyl groups from 5-methyl-THF, then the methylene-THF reductase (MetF) would be required for its activity. We have evidence from three different genera that MetF is not required for MesD's activity.

First, in *Acinetobacter baylyi*, MesD and DUF1852 are required for growth if neither methionine nor vitamin B12 are available (de Berardinis et al. 2008). (*A. baylyi* also has a cobalamin-dependent methionine synthase (MetH), but MetH is probably not active under these conditions because *A. baylyi* cannot synthesize cobalamin.) In a constraint-based metabolic model of *A. baylyi* in which 5-methyl-THF is a precursor to methionine, MetF is predicted to be essential for growth in these conditions (Durot et al. 2008), which illustrates that MetF is the only known path to 5-methyl-THF. Nevertheless, MetF from *A. baylyi* is not essential for growth in a defined minimal medium with no vitamins (de Berardinis et al. 2008).

Second, although most of the organisms with MesD also contain MetH or MetE, we found a few genomes in which MesD seems to be the sole methionine synthase. One of these is *Arthrobacter aurescens* TC1, which can grow in a defined minimal medium with the herbicide atrazine as the sole source of carbon (Strong et al. 2002). The genome of *A. aurescens*, which is complete, does not contain MetF, and the degradation pathway for atrazine does not involve 5-methyl-THF or other folate derivatives ((Sajjaphan et al. 2004); pathway PWY-5724 in MetaCyc, (Caspi et al. 2018)). Again, this implies that 5-methyl-THF is not required for MesD's activity.

Third, we used large-scale genetic data from Sphingomonas koreensis DSMZ 15582 (Price, Wetmore, et al. 2018) to test the roles of MesD, DUF1852, and MetF. *S. koreensis* also contains MetH (split into two parts), and these experiments used defined media that contains vitamin B12, so both methionine synthases could be functional. Nevertheless, MesD and DUF1852 were important for growth in most of the defined media experiments across a variety of carbon sources (Figure 6). In some of these conditions, neither MetH and MetF were required for growth (left side of Figure 6), which suggests that MesD can be the primary methionine synthase and that it does not require 5-methyl-THF.

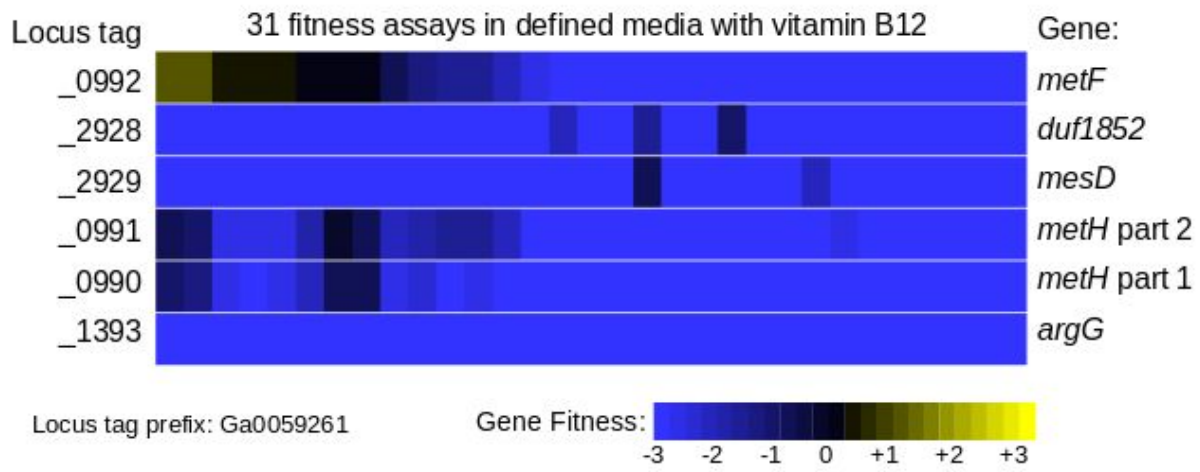

Figure 6: *Sphingomonas koreensis* can grow in minimal media by using MesD and not MetF. Each cell in the heatmap shows a fitness value from a different experiment. Only fitness assays with a defined media and a single carbon source are shown; all of these media included vitamin B12. Each fitness value is the log2 change in the abundance of mutants in that gene during that experiment (3-7 generations). Experiments are ordered by the fitness value of *metF*.

Since MesD does not obtain methyl groups from 5-methyl-THF, and since MesD and DUF1852 are often accompanied by a flavin reductase (i.e., ACIAD3522), we propose that DUF1852 acquires a methyl group by a reductive process and MesD then transfers the methyl group to methionine. Because MesD and DUF1852 are present in diverse bacteria but are restricted to aerobes, we also speculate that oxygen or hydrogen peroxide are involved in DUF1852's function. Again, this suggests that DUF1852 is an oxidoreductase, but we do not have a specific proposal for the substrate of DUF1852.

Materials and Methods

Alignments of the MetE superfamily

Literature on MetE and related proteins was retrieved using PaperBLAST (Price, Wetmore, et al. 2018; Price and Arkin 2017) and using PaperBLAST family search (for PF01717, the catalytic domain).

To align MetE with MesD, we used the structure-guided aligner MAFFT-DASH (Rozewicki et al. 2019).

To align MetE to MesA, MesB, and MesC, we obtained diverse sequences for each of these families, along with related sequences, from MicrobesOnline (Dehal et al. 2010) , and aligned them using MUSCLE (Edgar 2004). Although these sequences are quite divergent, the conservation of functional residues (Figure 2) shows that the alignment is accurate.

Substrate-binding residues

Substrate-binding residues were determined from PDB:1U1J (MetE from *Arabidopsis thaliana* in complex with zinc and methionine) and PDB:3RPD (MesD from *Shewanella sp.* W3-18-1 in complex with zinc and selenomethionine) using the ligand interaction viewer at rcsb.org and the ligplot tool from PDBsum (Laskowski and Swindells 2011). Compared to 1U1J, 3RPD has an additional hydrogen bond involving the carboxylate group of selenomethionine and the side chain of Tyr226. Tyr226 is in a MesD-specific insertion and does not align with MetE or other short methionine synthases, but Tyr226 is conserved within the MesD family.

Phylogenetic tree of MesB and relatives

To infer a phylogenetic tree of MesB and related proteins (Figure 5A), we selected the 88 closest homologs of DET0516 in MicrobesOnline. We removed two truncated proteins and a highly diverged protein (VIMSS11031200 from *Mahella australiensis* DSM 15567). The remaining 85 proteins were aligned using MUSCLE, and the alignment was trimmed to relatively-confident columns with Gblocks (Castresana 2000). We used a minimum block length of 2 and allowed at most half gaps at any position. A phylogenetic tree was inferred from the trimmed alignment with FastTree and the JTT+CAT model (Price, Dehal, and Arkin 2010). Figure 5A shows the proteins that are expected to be MesB (given the presence of the Wood-Ljungdahl pathway and functional residues) and their closest neighbors in the tree.

Phylogenetic profiling

Phylogenetic profiling of methionine synthases (as shown in Figure 5A) was conducted using MicrobesOnline (Dehal et al. 2010). To define the presence and absence of MetE, we used TIGR01371; for MetH, we used COG1410; for MesA, we used hits of 180 bits or higher to MTH775 (VIMSS 20772), which is over 90% identical to the characterized protein (METE_METTM); and for MesC, we used hits of 190 bits or higher to MA0053 (VIMSS 233378) from *Methanosarcina acetivorans* C2A. For AcsB, we used COG1614 (also known as CdhC). For AcsC, we used COG1456 (also known as CdhE). For AcsD, we used COG2069 (also known as CdhD).

Because MicrobesOnline includes only 1,654 prokaryotic genomes, we used larger databases for additional tests of which organisms contain MesA, MesB, or MesC.

To test if MesA is present in any organisms besides methanogens, we used the jackhmmer tool at the European Bioinformatics Institute's website (Potter et al. 2018) to find homologs of METE_METTM in UniProt reference proteomes. (Jackhmmer searches were run between November 22-25, 2019.) We used just one iteration of jackhmmer for all of our analyses. We found 44 hits with $E < 10^{-47}$ (over 40% identity), which were all from methanogens. These included representatives of 18 genera from the orders Methanobacteriales, Methanococcales, Methanocellales, Methanococcales, and Methanopyrales. In general, methanogens have MesA

except that representatives of the order Methanosarcinia have MesC instead. (MetH or MetE are not found in methanogens.)

To test if MesB is present in any organisms that lack the Wood-Ljungdahl pathway, we compared the jackhmmer results for MesB, CdhD, and CdhE. For MesB, we used DET0516 from *Dehalococcoides mccartyi*; for CdhD and CdhE, we used representatives from *Methanosarcina acetivorans* C2A as queries (ACDD1_METAC and ACDG_METAC, respectively). We considered hits to DET0516 with $E < 10^{-34}$ to be MesB (corresponding to roughly 29% identity or above; the weakest hit we kept was to *Candidatus Frackibacter sp.* WG12). Of the 151 hits for MesB, 148 were from species that contain CdhD and CdhE. The exceptions were *Roseiflexus sp.* RS-1, which as discussed above contains a non-enzymatic homolog of MesB; the related bacterium *Oscillochloris trichoides*, which contains a similar protein as *Roseiflexus* does; and the anaerobic bacterium *Desulfocucumis palustris,* which contains CdhD and the decarbonylase CdhC but seems to lack CdhE.

To test if MesC is present in any organisms that lack the Wood-Ljungdahl pathway, we compared jackhmmer results for MesC, CdhE, and CdhD. To represent MesC, we used a protein from *Methanosarcina acetivorans* C2A (Q8TUL3_METAC). We found a weak hit of MesC to protein A0A166AZ17 from *Methanobrevibacter curvatus*, which does not contain CdhD or CdhE. A0A166AZ17 is more similar to MesA (METE_METTM; 47% identity instead of 22% identity), so we disregarded this hit. The other hits were all to anaerobic archaea that contain CdhD and CdhE, except for a few hits to metagenome-assembled draft genomes from uncultured archaeal lineage MSBL-1. MesC was found in nine draft genomes of MSBL-1, and CdhD and/or CdhE were found in five of these nine. The Wood-Ljungdahl pathway is reported to be present in many MSBL-1 genomes (Mwirichia et al. 2016). Because these are fragmented assemblies from metagenomics, we are not sure if the Wood-Ljungdahl pathway is truly absent from some of the MSBL-1 genomes that encode MesC-like proteins.

To test if MesD and DUF1852 are present only in aerobic bacteria, we first used MicrobesOnline to identify homologs of ACIAD3523 (MesD) that are likely to have the same function in methionine synthesis. Specifically, we used the MicrobesOnline tree-browser to check if they were adjacent to DUF1852. We chose a bit score threshold of 390 bits, as homologs above this threshold were almost always adjacent to DUF1852. We also excluded a homolog from the oomycete *Phytophthora capsici*, which could be contamination or might indicate the acquisition of DNA from a *Stenotrophomonas* bacterium. This left 106 genomes from 44 genera that contain MesD. We used FAPROTAX (Louca, Parfrey, and Doebeli 2016) along with web searches to confirm that all of these genera are aerobic. To check more broadly that MesD/DUF1852 are found only in aerobic bacteria, we used AnnoTree (Mendler et al. 2019) to obtain a list of 235 genera that contain DUF1852. After removing suffixes (i.e., converting “Erythrobacter_B” to “Erythrobacter”), we had 206 named genera, of which 170 were distinct from the genus names in MicrobesOnline. We examined a random sample of 50 of these genera and all were aerobic.

We also ran jackhmmer against UniProt reference proteomes with the sequence of ACIAD3523 or PFam's HMM model for DUF1852 (PF08908.11) as queries. 418 of the top 422 hits to ACIAD3523 (corresponding to $E < 10^{-173}$ or roughly 75% identity and above) were from organisms that contain DUF1852. One of these organisms, *Rubritalea profundi*, was from the Verrucomicrobia. *R. profundi* is a strict aerobe (Song et al. 2018).

Overall, we confirmed that MesA is found only in methanogens; MesB and MesC are found only in organisms with the Wood-Ljungdahl pathway; and MesD is found only in aerobic bacteria that also contain DUF1852.

Acknowledgements

ENIGMA is a Scientific Focus Area Program supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTLFoundational Science through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

References

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research* 25 (17): 3389–3402.

Berardinis, Véronique de, David Vallenet, Vanina Castelli, Marielle Besnard, Agnès Pinet, Corinne Cruaud, Sumitta Samair, et al. 2008. "A Complete Collection of Single-Gene Deletion Mutants of Acinetobacter Baylyi ADP1." *Molecular Systems Biology* 4 (March): 174.

Caspi, Ron, Richard Billington, Carol A. Fulcher, Ingrid M. Keseler, Anamika Kothari, Markus Krummenacker, Mario Latendresse, et al. 2018. "The MetaCyc Database of Metabolic Pathways and Enzymes." *Nucleic Acids Research* 46 (D1): D633–39.

Castresana, J. 2000. "Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis." *Molecular Biology and Evolution* 17 (4): 540–52.

Crooks, Gavin E., Gary Hon, John-Marc Chandonia, and Steven E. Brenner. 2004. "WebLogo: A Sequence Logo Generator." *Genome Research* 14 (6): 1188–90.

Dehal, Paramvir S., Marcin P. Joachimiak, Morgan N. Price, John T. Bates, Jason K. Baumohl, Dylan Chivian, Greg D. Friedland, et al. 2010. "MicrobesOnline: An Integrated Portal for Comparative and Functional Genomics." *Nucleic Acids Research* 38 (Database issue): D396–400.

Durot, Maxime, François Le Fèvre, Véronique de Berardinis, Annett Kreimeyer, David Vallenet, Cyril Combe, Serge Smidtas, Marcel Salanoubat, Jean Weissenbach, and Vincent Schachter. 2008. "Iterative Reconstruction of a Global Metabolic Model of Acinetobacter Baylyi ADP1 Using High-Throughput Growth Phenotype and Gene Essentiality Data." *BMC Systems Biology* 2 (October): 85.

Edgar, Robert C. 2004. "MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput." *Nucleic Acids Research* 32 (5): 1792–97.

El-Gebali, Sara, Jaina Mistry, Alex Bateman, Sean R. Eddy, Aurélien Luciani, Simon C. Potter,

Matloob Qureshi, et al. 2019. “The Pfam Protein Families Database in 2019.” *Nucleic Acids Research* 47 (D1): D427–32.

Ferrer, Jean-Luc, Stéphane Ravanel, Mylène Robert, and Renaud Dumas. 2004. “Crystal Structures of Cobalamin-Independent Methionine Synthase Complexed with Zinc, Homocysteine, and Methyltetrahydrofolate.” *The Journal of Biological Chemistry* 279 (43): 44235–38.

Fu, He, and William W. Metcalf. 2015. “Genetic Basis for Metabolism of Methylated Sulfur Compounds in Methanosarcina Species.” *Journal of Bacteriology* 197 (8): 1515–24.

González, J. C., K. Peariso, J. E. Penner-Hahn, and R. G. Matthews. 1996. “Cobalamin-Independent Methionine Synthase from Escherichia Coli: A Zinc Metalloenzyme.” *Biochemistry* 35 (38): 12228–34.

Goulding, C. W., D. Postigo, and R. G. Matthews. 1997. “Cobalamin-Dependent Methionine Synthase Is a Modular Protein with Distinct Regions for Binding Homocysteine, Methyltetrahydrofolate, Cobalamin, and Adenosylmethionine.” *Biochemistry* 36 (26): 8082–91.

Kräutler, B., J. Moll, and R. K. Thauer. 1987. “The Corrinoid from Methanobacterium Thermoautotrophicum (Marburg Strain). Spectroscopic Structure Analysis and Identification as Co Beta-Cyano-5’-Hydroxybenzimidazolyl-Cobamide (factor III).” *European Journal of Biochemistry / FEBS* 162 (2): 275–78.

Laskowski, Roman A., and Mark B. Swindells. 2011. “LigPlot+: Multiple Ligand-Protein Interaction Diagrams for Drug Discovery.” *Journal of Chemical Information and Modeling* 51 (10): 2778–86.

Louca, Stilianos, Laura Wegener Parfrey, and Michael Doebeli. 2016. “Decoupling Function and Taxonomy in the Global Ocean Microbiome.” *Science* 353 (6305): 1272–77.

Matthews, Rowena G., April E. Smith, Zhaohui S. Zhou, Rebecca E. Taurog, Vahe Bandarian, John C. Evans, and Martha Ludwig. 2003. “Cobalamin-Dependent and Cobalamin-Independent Methionine Synthases: Are There Two Solutions to the Same Chemical Problem?” *Health Care Analysis: HCA: Journal of Health Philosophy and Policy* 86 (12): 3939–54.

Mendler, Kerrin, Han Chen, Donovan H. Parks, Briallen Lobb, Laura A. Hug, and Andrew C. Doxey. 2019. “AnnoTree: Visualization and Exploration of a Functionally Annotated Microbial Tree of Life.” *Nucleic Acids Research* 47 (9): 4442–48.

Mwirichia, Romano, Intikhab Alam, Mamoon Rashid, Manikandan Vinu, Wail Ba-Alawi, Allan Anthony Kamau, David Kamanda Ngugi, et al. 2016. “Metabolic Traits of an Uncultured Archaeal Lineage--MSBL1--from Brine Pools of the Red Sea.” *Scientific Reports* 6 (January): 19181.

North, Justin A., Anthony R. Miller, John A. Wildenthal, Sarah J. Young, and F. Robert Tabita. 2017. “Microbial Pathway for Anaerobic 5’-Methylthioadenosine Metabolism Coupled to Ethylene Formation.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (48): E10455–64.

North, Justin A., Jaya Sriram, Karuna Chourey, Christopher D. Ecker, Ritin Sharma, John A. Wildenthal, Robert L. Hettich, and F. Robert Tabita. 2016. “Metabolic Regulation as a Consequence of Anaerobic 5-Methylthioadenosine Recycling in Rhodospirillum Rubrum.” *mBio* 7 (4). https://doi.org/10.1128/mBio.00855-16.

Pejchal, Robert, and Martha L. Ludwig. 2005. “Cobalamin-Independent Methionine Synthase (MetE): A Face-to-Face Double Barrel That Evolved by Gene Duplication.” *PLoS Biology* 3 (2): e31.

Potter, Simon C., Aurélien Luciani, Sean R. Eddy, Youngmi Park, Rodrigo Lopez, and Robert D.

Finn. 2018. “HMMER Web Server: 2018 Update.” *Nucleic Acids Research* 46 (W1): W200–204.
Price, Morgan N., and Adam P. Arkin. 2017. “PaperBLAST: Text Mining Papers for Information about Homologs.” *mSystems* 2 (4). https://doi.org/10.1128/mSystems.00039-17.
Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin. 2010. “FastTree 2--Approximately Maximum-Likelihood Trees for Large Alignments.” *PloS One* 5 (3): e9490.
Price, Morgan N., Adam M. Deutschbauer, and Adam P. Arkin. 2019. “GapMind: Automated Annotation of Amino Acid Biosynthesis.” *bioRxiv*. https://doi.org/10.1101/741918.
Price, Morgan N., Kelly M. Wetmore, R. Jordan Waters, Mark Callaghan, Jayashree Ray, Hualan Liu, Jennifer V. Kuehl, et al. 2018. “Mutant Phenotypes for Thousands of Bacterial Genes of Unknown Function.” *Nature* 557 (7706): 503–9.
Price, Morgan N., Grant M. Zane, Jennifer V. Kuehl, Ryan A. Melnyk, Judy D. Wall, Adam M. Deutschbauer, and Adam P. Arkin. 2018. “Filling Gaps in Bacterial Amino Acid Biosynthesis Pathways with High-Throughput Genetics.” *PLoS Genetics* 14 (1): e1007147.
Rozewicki, John, Songling Li, Karlou Mar Amada, Daron M. Standley, and Kazutaka Katoh. 2019. “MAFFT-DASH: Integrated Protein Sequence and Structural Alignment.” *Nucleic Acids Research* 47 (W1): W5–10.
Sajjaphan, Kannika, Nir Shapir, Lawrence P. Wackett, Michael Palmer, Barbara Blackmon, Jeff Tomkins, and Michael J. Sadowsky. 2004. “Arthrobacter Aurescens TC1 Atrazine Catabolism Genes trzN, atzB, and atzC Are Linked on a 160-Kilobase Region and Are Functional in Escherichia Coli.” *Applied and Environmental Microbiology* 70 (7): 4402–7.
Schröder, I., and R. K. Thauer. 1999. “Methylcobalamin:homocysteine Methyltransferase from Methanobacterium Thermoautotrophicum. Identification as the metE Gene Product.” *European Journal of Biochemistry / FEBS* 263 (3): 789–96.
Song, Jaeho, Yeonjung Lim, Yochan Joung, Jang-Cheon Cho, and Kazuhiro Kogure. 2018. “Sp. Nov., Isolated from Deep-Seawater and Emended Description of the Genus in the Phylum.” *International Journal of Systematic and Evolutionary Microbiology* 68 (4): 1384–89.
Strong, Lisa C., Charlotte Rosendahl, Gilbert Johnson, Michael J. Sadowsky, and Lawrence P. Wackett. 2002. “Arthrobacter Aurescens TC1 Metabolizes Diverse S-Triazine Ring Compounds.” *Applied and Environmental Microbiology* 68 (12): 5973–80.
Tatusov, Roman L., Natalie D. Fedorova, John D. Jackson, Aviva R. Jacobs, Boris Kiryutin, Eugene V. Koonin, Dmitri M. Krylov, et al. 2003. “The COG Database: An Updated Version Includes Eukaryotes.” *BMC Bioinformatics* 4 (September): 41.
Taurog, Rebecca E., and Rowena G. Matthews. 2006. “Activation of Methyltetrahydrofolate by Cobalamin-Independent Methionine Synthase.” *Biochemistry* 45 (16): 5092–5102.
Zhuang, Wei-Qin, Shan Yi, Markus Bill, Vanessa L. Brisson, Xueyang Feng, Yujie Men, Mark E. Conrad, Yinjie J. Tang, and Lisa Alvarez-Cohen. 2014. “Incomplete Wood-Ljungdahl Pathway Facilitates One-Carbon Metabolism in Organohalide-Respiring Dehalococcoides Mccartyi.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (17): 6419–24.