SUPPLEMENTARY METHODS

Generating *fae* alleles and mutant strains

All mutant strains used in the study were created on the background of an *fae* deletion strain, CM2563 (variant of CM198.1 (Marx and Lidstrom 2002)). As described in the Methods section in the main text, we compared two *M. extorquens* strains (AM1 and DM4) to find conserved amino acid residues (those that are identical between the two species) and variable residues (those that differ between the two species). We then calculated, for each amino acid, which codons were significantly enriched in the conserved regions relative to the variable regions; these we designated as "frequent codons". Conversely, codons that were significantly relatively depleted at the conserved amino acid residues were called "rare codons". Using this list of most frequently and most rarely used codons in conserved residues of protein-coding genes in *Methylobacterium extorquens* AM1, we created six *fae* alleles. Alleles AF, AR and RN were straightforward to design using the list of rare and frequent codons. For strain AC, we used a published structure of FAE to determine active residues (Achary et al). For strains VA and CO, we used a multiple alignment of *fae* sequences across 26 species, including *M. extorquens* strains as well as species from other genera to determine the top 50% fae residues that were most conserved across species, and the bottom 50% residues that were most variable.

The *fae* alleles were synthesized by Integrated DNA Technologies (IDT, Coralville, IA) and delivered as an insert in a pSMARTHC-Kan vector. We PCR-amplified ~400 bp regions upstream and downstream from *fae* using CM501 genomic DNA using the primer pairs CM_syn_fae-uf and CM_syn_fae-ur or CM_syn_fae-df and CM_syn_fae-dr (Table S4). We inserted the upstream region into pCM433 after digestion with *Xba*I and *Apa*I (pCM433) and *Nhe*I and *Apa*I (upstream fragment), to make pDAFu. Next, we inserted the *fae* downstream region between the *Apa*I and *Age*I sites within this plasmid to create pDAFud. We then inserted *fae* alleles into the resulting plasmid (between the upstream and downstream regions) by digestion of pSMART::*fae* and pDAFud with *Pst*I and *Mlu*I, followed by ligation. We amplified the WT *fae* allele from CM501 with the primers CM_syn_fae0-f and CM_syn_fae0-r (Table S4). The resulting plasmids (pDAFu0d to pDAFu6d) carried each *fae* allele with both C and N terminal FLAG tags, along with the gene's flanking regions (Table S5). Finally, we digested the plasmids with *Nde*I to remove the N-terminal FLAG tag and self-ligated these, creating pDAF0 through pDAF6 (Table S5). The plasmids were conjugated into CM2563 to insert the synonymous alleles into the chromosomal location of *fae* via homologous recombination (Marx 2008), creating strains CM2556, CM2565, and CM2558 to CM2562 (Table S6). We ensured that the C-terminal FLAG tag did not impose a significant fitness cost (Figure S2). We sequenced the *fae* locus of each

mutant strain to confirm that the *fae* alleles were placed in the correct location and had the expected codon substitutions.

To create strains with plasmid-borne *fae* alleles with an inducible promoter, we amplified *fae* alleles from the pDAF0 to pDAF6 plasmids using the primer pair DAp51f and DAp43r. We digested the resulting PCR products and pHC115 (cumate-regulated expression vector) with *Bgl*II and *Eco*RI and ligated the resulting fragments, generating pDA115-F0 to pDA115-F6 (Table S5). We transformed CM2563 with each plasmid (selecting for kanamycin resistance to ensure plasmid carriage) to create CM2574 to CM2575 (Table S6).

Growth rate measurement

All growth rate and fitness assays were performed in 48-well cell culture plates (Costar, Corning, NY) containing 640 μL total liquid in each well. Plates were incubated with shaking on a 36–plate shaking tower (Liconic) run at 650 rpm (1mm orbit) in a room at 30 °C. We first inoculated 10 μL freezer stocks of each strain into 630 μL succinate medium (three biological replicates). After two days we transferred 10 μL to new plates containing 630 μL succinate medium, with three replicate wells per biological replicate. After two days of acclimation, we inoculated 10 μL culture from each well in 630 μL fresh methanol or succinate medium and measured OD_{600} at 1-hour intervals for 48 hours in an automated fashion, using a Victor 2.0 plate reader (Perkin Elmer, Waltham, MA). We calculated growth rate of the culture in each well as the slope from a linear regression fitted to the exponential phase of the growth curve (1% - 80% of maximum OD_{600}). We manually checked each regression and excluded data points that were clear outliers that probably arose from plate reading errors (< 2 data points were excluded out of a total of >10 total points for each curve). We followed a similar protocol to quantify the growth rate of regulated promoter strains, with the following modifications: (a) In the inoculation and acclimation stages, we added kanamycin (50 μg/mL) to ensure plasmid carriage (b) In the acclimation stage, we also added the inducer, cumate (4-isopropyl benzoic acid, Sigma, St Louis, MO), dissolved in methanol at the appropriate concentration for each treatment (since methanol is also used as a carbon source, we added pure methanol as necessary to ensure that all treatments received an equivalent total amount of methanol).

Competitive fitness measurements

We used the acclimation plates from growth rate measurements (above) to set up a separate set of competition assay plates in parallel, following a previously described protocol (Lee, Chou, and Marx 2009). We mixed cultures of each test strain with an equal volume of a reference strain that carries an mCherry fluorescent marker (CM2721). We stored 100 μL of this mixture (with 8 μL DMSO) at –80 °C to measure the initial ratio of the two competing strains. We inoculated 10 μL of the remaining mixture in

630 μL test medium in new 48 well plates. After 48 hours, we added 8 μL DMSO to 100 μL culture from these plates and froze it at –80 °C. After thawing the initial (t_0) and final (t_1) plates, we measured the initial and final ratio (R_0 and R_1) of fluorescent (reference) and non-fluorescent (test) strains using a flow cytometer (BD Biosciences LSR Fortessa). Assuming a net 64-fold population size expansion during growth, we estimated the fitness (W) of mutant strains relative to the wild type as:

$$W = \frac{\log\left(\frac{R_1 .64}{R_0}\right)}{\log\left(\frac{(1-R_1).64}{(1-R_0)}\right)}$$

mRNA and protein quantification

We inoculated –80 °C freezer stocks of each strain (3 biological replicates) in 5 mL succinate medium in 50 mL flasks. After 2 days, we transferred 310 μL culture at stationary phase to 50 mL flasks containing 20 mL fresh succinate medium. For strains bearing *fae* alleles on the regulated promoter (CM2574 through CM2680), we added kanamycin to ensure plasmid carriage. After 24 hours of growth in a shaking incubator (mid-log phase), we added 12 μL 100% methanol to each flask to induce *fae* expression (for regulated-promoter strains, we also added the appropriate amount of cumate inducer as described for growth rate measurements above). This approach of growth on permissive medium followed by induction with methanol was necessary because some strains could not grow on methanol alone (Figure 2a). One hour after induction, we added 2 mL Qiagen BacteriaProtect Reagent to 1 mL of each culture, collected cell pellets by centrifugation (4600 rpm at 4 °C for 5 minutes) and froze pellets at –20 °C for RNA extraction. We collected cells from the remaining culture volume by centrifugation and froze them at –80 °C for protein extraction.

We made cDNA from each RNA sample (Reverse Transcription kit, Qiagen, Valencia, CA) and set up quantitative PCR reactions (SYBRFast RTPCR kit, Qiagen, Valencia, CA) in 96 well plates (MX3000p machine, Stratagene, Santa Clara, CA). Due to the differences in the nucleotide sequences of each gene version, we used specific primers to quantify *fae* mRNA for each mutant, with the ribosomal protein gene (*rpsB*) as an endogenous reference (primer sequences in Table S7). In a separate experiment, we generated standard curves for each mutant (using six serial dilutions of purified PCR-amplified *fae* DNA) to confirm that all primer sets had similar amplification efficiency (90-110%). For each primer set, we used RNA isolated from the *fae* knockout strain (CM 2563) as a negative control, and purified PCR-amplified *fae* DNA as a positive control. Using three technical replicates for each biological replicate of a

strain, we calculated the average cycle threshold for detection of SYBR green dye (C_t). We then calculated *fae* mRNA expression relative to wild type as $2^{-\Delta\Delta}Ct$, where $\Delta\Delta C_{t\ (mutant)} = \Delta C_{t\ (mutant)} - \Delta C_{t\ (wild\ type)}$ and $\Delta C_t = C_{t\ (fae)} - C_{t\ (rpsB)}$.

We extracted soluble and insoluble protein fractions for each sample, modifying the protocol described for yeast cells in (Geiler-Samerotte et al. 2011). We made the following buffers: (a) soluble protein buffer ("SPB", pH 6.8: 50mM Tris-HCl, 150 mM NaCl, 1:100 protease inhibitor cocktail [one EDTA-free complete Mini tablet (Roche, Branford, CT) dissolved in 500 μL protease-free water]) (b) insoluble protein buffer ("IPB", pH 6.8: SPB with 2% SDS and 2mM DTT) (c) 6x insoluble protein loading buffer ("IPLB": 50mM Tris-HCl, 0.05% bromophenol blue, 30% glycerol, 5% β-mercaptoethanol) (d) 6x soluble protein loading buffer ("SPLB": IPLB with 10% SDS). Throughout, we used a chilled Sorvall Legend RT centrifuge (4 °C, 4600 rpm) and ice-cold buffers, and kept samples on ice. We thawed frozen cell pellets, washed them in 500 μL Tris Buffered Saline (TBS), centrifuged suspensions and discarded the supernatant. We re-suspended washed pellets in 400μL SPB and lysed cells in screw-cap tubes with lysing matrix B (1 min at 6.5 m/s on a MPBio Fast-Prep 24 bead-beater). We centrifuged tubes for 1 min and transferred the supernatant to fresh tubes (soluble protein fraction). We washed pellets twice with 500μL SPB and extracted insoluble protein fraction in 500μL IPB (vortex 10 sec, centrifuge 1min and remove supernatant after each step). We stored soluble and insoluble fractions at –80 °C in 20μL aliquots. Using one aliquot, we first quantified the amount of protein in each sample using a Qubit Quant-iT Protein assay (Invitrogen, Grand Island, NY). We diluted samples as necessary (with SPB or IPB), added appropriate amount of 6x SPB or IPLB, denatured proteins at 100 °C for 5 mins, and loaded approximately equal amount of protein from each sample in 1.5 mm 10-20% Novex Tris-Glycine precast gels (Invitrogen). In each gel, we included a lane with 1.6 μg C-terminal FLAG-BAP fusion protein (Sigma, St Louis, MO) as a positive control, and a BenchMark pre-stained protein ladder (Invitrogen, Grand Island, NY) to estimate protein size. We used the X-cell sure lock system (Invitrogen, Grand Island, NY) for denaturing gel electrophoresis (100V for 80 minutes) and transferred proteins from gels to PVDF membranes with an iBLOT dry blotting system (Invitrogen). We cut each blot into two at the 49 kDa size standard. We stained the top half of the blot (larger proteins) with GelCode Blue stain (ThermoScientific, Waltham, MA) to determine protein loading in each lane. We blocked the other half of the blot (with low molecular weight proteins) with TBST (Tris Buffered Saline + 0.1% Tween-20) containing 5% nonfat milk for 1 hour at 30 °C. We incubated the blot at room temperature with murine monoclonal anti-FLAG M2 antibody (Sigma, St Louis, MO) diluted to 3.8 μg/ml in TBST with 3% nonfat milk, for 30 minutes. We then incubated the blot with the secondary antibody (Anti-mouse IgG-Peroxidase diluted 1:5000 in TBST with 5% milk) for 30 minutes, with TBST washes between each step. After multiple TBST washes (total 1 hour), we developed the blot with

SuperSignal West Femto Maximum Sensitivity Substrate (ThermoScientific, Waltham, MA) for 2 minutes. We imaged and analyzed band intensity from both GelCode stained loading control blots and chemiluminescent western blots using the AlphaInnotech ChemiImager. We quantified the amount of FAE produced by each strain in arbitrary units, after normalizing with the amount of reference protein loaded in each lane (~64 kDa, quantified from GelCode-stained blots). Results were qualitatively the same even when we used two other protein bands (115 or 82 kDa) as reference.

Assays of FAE activity

Purification of H_4MPT: We cultured the *M. extorquens* AM1 deletion mutant lacking *fae*, CM115K.1 (Vorholt et al. 2000) and carrying the plasmid pCM106 (Marx and Chistoserdova 2003) at 30 °C on a minimal salts medium (Attwood and Harder 1972) containing methanol (125 mM) and tetracycline (10 μg/ml). The presence of pCM106 allows for expression of an alternative formaldehyde-oxidation pathway that permits growth on methanol without FAE (Marx and Chistoserdova 2003). We harvested cells by centrifugation at an OD_{600} of 2.5. The cell paste (300 g) was introduced into an anaerobic chamber (Coy, Grass Lake, MI) containing 95% N_2 and 5% H_2. All experiments were performed in the dark. Cells were resuspended in 300 mL anoxic Buffer A (5 mM potassium phosphate buffer pH=4.8, 10 mM β-mercaptoethanol) and broken by boiling (15 min). Cell free extract was cleared by centrifugation in a sealed tube outside of the anaerobic glove box (28,000 x g, 45min, 4 °C), transferred back into the glove box, and 60 mL supernatant was applied to an OASIS weak anion exchange extraction cartridge (6 cc, 500 mg) (Waters, Milford, MA) previously activated with 1% formic acid and equilibrated with methanol. After loading, the column was washed with 1 column volume of H_2O. Partially purified H_4MPT was eluted with one column volume of Elution Solution 1 (5% NH_4OH, 80% methanol, 15% H_2O). The elution fraction was analyzed under UV-Visible light to confirm the characteristic maximal peaks of 247 nm and 302 nm. The fraction was also tested by monitoring NADP-dependent MtdB activity (Hagemeier et al. 2000). After corroborating activity, the active fractions were pooled (4 fractions of 3 mL each) and lyophilized under anoxic conditions. The lyophilized powder was transferred into the glove box and resuspended with anoxic Buffer A. The active fractions (6 mL each) were applied to an OASIS mixed-mode anion exchange cartridge (3 cc, 60 mg) (Waters, Milford, MA) previously activated with water and equilibrated with methanol. After loading, the column was washed with 1 column volume of water. While purified H_4MPT did not bind the sorbent, some contaminants did bind and were discarded. The flow trough fraction and washes were pooled (2 fractions of 4 mL each) and lyophilized under anoxic conditions. The powder was transferred into the glove box and resuspended with anoxic buffer A. The fraction was analyzed by UV-Visible analysis and activity as described above. The process was repeated

once more to further purify the cofactor, and the resulting lyophilized powder was transferred into the glove box and resuspended in anoxic Buffer B (100 mM Potassium phosphate buffer, pH= 7.8, 1 mL). Analysis by UV-Visible spectrum and NADP-dependent MtdB activity was corroborated in the final fraction. Analysis by MALDI-TOF (positive mode) corroborated the m/z typical of H_4MPT. The mass spectrometer (Quattro Micro API, Micromass ,Manchester, UK) was operated in the positive (3.5 kv) electrospray ionization (ESI) mode. Nitrogen was used as desolvation gas at 300 L/h and as cone gas at 50 L/h. The syringe pump was used to infuse the purified sample at a flow rate 5 µL/mL for MS and MS/MS analysis. The mass spectrometer was first operated in Q1MS mode to detect the interested (targeted) parent ions. It was then operated in MS/MS mode to look for the product ions for the selected parent. The collision energy was optimized to obtain good signal-to-noise ratio of the product ions.

Purification of MtdB: For high-level expression of MtdB, the *mtdB* gene was amplified and cloned into pET28b as described by Rasche (Rasche, Havemann, and Rosenzvaig 2004) with the exception that the DNA template used for the PCR amplification was chromosomal DNA from wild type *M. extorquens* AM1. The construct was transformed into BL21-AI. This strain was grown at 37 °C in Superbroth media with kanamycin (50 µg/mL). IPTG (1 mM) and arabinose (0.2%) were added to induce expression of MtdB when the culture reached OD_{600}=0.5. Cultures were grown after induction for 3 hours at 28 °C and cells were harvested by centrifugation (4729 x g; 10 min, 4 °C). The cell paste (30 g) was resuspended in 10 mL of buffer A (50 mM Tris-HCl, pH=8.0, 5 mM imidazole, and 15% (v/v) glycerol) and cells were broken by French Press. Cell-free extracts were cleared by centrifugation (28,078 x g, 45min, 4 °C), and the supernatant was applied onto a Ni-charged chelating Sepharose column (8 mL) previously equilibrated with buffer A. After loading, the column was washed with 5 column volumes of buffer A. MtdB-His6 protein was eluted off the column by using an imidazole gradient 0 to 500 mM over 50 mL in buffer A. Fractions (4 fractions of 3 mL each) were pooled and desalted using a PD-10 gel filtration column (8.3 mL bed volume, 5 cm bed height) equilibrated with buffer B (100 mM Potassium Phosphate buffer pH=7.5). The protein was concentrated using centrifugal filter devices (Amicon-Ultra, 10K, 4000 x g, 15 min, 4 °C). The concentrated protein sample (100 µM) was used in the experiments as indicated. Protein concentration was determined by the bicinchoninic acid method (Pierce).

***FAE activity*:** In *M. extorquens* AM1, FAE catalyzes the conversion of formaldehyde to methylene H_4MPT. The resulting intermediate is then used as a substrate by the enzyme H_4MPT dehydrogenase (MtdB), catalyzing its dehydrogentaion with NADP as a cosubstrate to generate methenyl H_4MPT and NADPH (Vorholt et al. 1998). Taking advantage of the physiological sequence of the reactions, we used an indirect assay (monitoring production of NADPH) to measure the activity of FAE by adding purified MtdB to the assay. Although the spontaneous coupling of H_4MPT and formaldehyde occurs, using

slightly alkaline pH and the presence of high amounts of magnesium increases the relative rate of the enzymatic reaction considerably. The contribution of the spontaneous reaction is shown in Figure 2e (*fae* deletion strain) (Vorholt et al. 2000). The activity assay was performed as described by Vorholt (Vorholt et al. 2000) with the following differences. We used 50 mM Potassium Phosphate buffer pH=7.8, purified methylene-H_4MPT dehydrogenase MtdB (100 µM, 50 µL) and co-substrate NADP (125 µM). The total volume of the reaction was 1 mL. Activity was monitored at 340 nm (i.e., NADPH production) after addition of formaldehyde (2 mM).

tAI, mRNA folding and other predictions

We tested various hypotheses that predict the effect of transcript properties and codon usage on translation. We used the codonR program to calculate the overall tRNA adaptation index (tAI) for each *fae* allele as described by dos Reis (Reis 2003), with tRNA gene copy numbers extracted from the *M. extorquens* AM1 genome. We also calculated local tAI for each codon sequentially to estimate local translation speed specific to each codon.

Using the Vienna RNAfold program (Lorenz et al. 2011) at 30° C (the growth temperature for *M. extorquens*) and with RNA folding parameters described by Andronescu et al (Andronescu et al. 2007), we estimated the minimum free energy of folding of each allele, for (a) the entire mRNA (b) the first 170 bp, and (c) across windows of 50 nt starting from 100 nucleotides upstream of the start codon. We used the TransTerm database (http://uther.otago.ac.nz/; (Jacobs et al. 2009)) to identify translation termination sites and other RNA regulatory elements within each *fae* allele. We used the RBSHidesign webserver (https://salis.psu.edu/software/; (Salis, Mirsky, and Voigt 2009)) to predict the translation initiation rate for each allele at 30 °C. However, the predicted initiation rate for the wild type allele was very low, which does not agree with the high rate of protein production we observe. Hence, we suspect that the design algorithm does not generate accurate predictions for *M. extorquens*, and we did not use these results to make inferences about translation initiation.

We also tested whether the frequency of ribosomal pausing due to Shine-Dalgarno (SD)-like sequences explains protein production and fitness of our mutants. We first calculated the frequency of hexamers in each allele, and tested whether it was correlated with its binding energy to the SD sequence. A significant negative correlation would indicate that SD-like sequences internal to the mRNA are avoided in *fae.* Because the anti- Shine-Dalgarno sequence of *M. extorquens* is the same as that of *E. coli*, we used the binding affinities calculated by Li et al (Li, Oh, and Weissman 2012). Individual hexamers were not present at very high frequencies in our alleles, since *fae* is a relatively small gene (510 bp). Therefore, to test for a correlation between protein production and frequency of hexamers with high binding affinity to anti-SD, we pooled hexamers with high affinities (less than -6 or -4 kcal/mol).

SUPPLEMENTARY REFERENCES

Andronescu M, Condon A, Hoos HH, Mathews DH, Murphy KP. 2007. Efficient parameter estimation for RNA secondary structure prediction. Bioinformatics 23:i19–28.

Attwood MM, Harder W. 1972. A rapid and specific enrichment procedure for Hyphomicrobium spp. Antonie Van Leeuwenhoek 38:369–377.

Geiler-Samerotte KA, Dion MF, Budnik BA, Wang SM, Hartl DL, Drummond DA. 2011. Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. Proceedings of the National Academy of Sciences 108:680–685.

Hagemeier CH, Chistoserdova L, Lidstrom ME, Thauer RK, Vorholt JA. 2000. Characterization of a second methylene tetrahydromethanopterin dehydrogenase from Methylobacterium extorquens AM1. Eur. J. Biochem. 267:3762–3769.

Jacobs GH, Chen A, Stevens SG, Stockwell PA, Black MA, Tate WP, Brown CM. 2009. Transterm: a database to aid the analysis of regulatory sequences in mRNAs. Nucleic Acids Research 37:D72–6.

Lee M-C, Chou H-H, Marx CJ. 2009. ASYMMETRIC, BIMODAL TRADE-OFFS DURING ADAPTATION OF METHYLOBACTERIUMTO DISTINCT GROWTH SUBSTRATES. Evolution 63:2816–2830.

Li G-W, Oh E, Weissman JS. 2012. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature 484:538–541.

Lorenz R, Bernhart SH, Hoener Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. Algorithms for Molecular Biology 6:26.

Marx C, Chistoserdova L. 2003. Formaldehyde-detoxifying role of the tetrahydromethanopterin-linked pathway in Methylobacterium extorquens AM1. Journal of Bacteriology 185:7160–7168.

Marx CJ, Lidstrom ME. 2002. Broad-host-range cre-lox system for antibiotic marker recycling in gram-negative bacteria. BioTechniques 33:1062–1067.

Marx CJ. 2008. Development of a broad-host-range sacB-based vector for unmarked allelic exchange. BMC Res Notes 1:1.

Rasche ME, Havemann SA, Rosenzvaig M. 2004. Characterization of two methanopterin biosynthesis mutants of Methylobacterium extorquens AM1 by use of a tetrahydromethanopterin bioassay. Journal of Bacteriology 186:1565–1570.

Reis dos M. 2003. Unexpected correlations between gene expression and codon usage bias from microarray data for the whole Escherichia coli K-12 genome. Nucleic Acids Research 31:6976–6985.

Salis HM, Mirsky EA, Voigt CA. 2009. Automated design of synthetic ribosome binding sites to control protein expression. Nature Biotechnology 27:946–950.

Vorholt JA, Chistoserdova L, Lidstrom ME, Thauer RK. 1998. The NADP-dependent methylene tetrahydromethanopterin dehydrogenase in Methylobacterium extorquens AM1. Journal of Bacteriology 180:5351–5356.

Vorholt JA, Marx CJ, Lidstrom ME, Thauer RK. 2000. Novel formaldehyde-activating enzyme in Methylobacterium extorquens AM1 required for growth on methanol. Journal of Bacteriology 182:6645–6650.