MtpB, a member of the MttB superfamily from the human intestinal acetogen *Eubacterium limosum*, catalyzes proline betaine demethylation

Jonathan W. Picking, Edward J. Behrman, Liwen Zhang, and Joseph A. Krzycki

Supporting Information

Supplemental Methods

Analysis of proline betaine and N-methyl proline in culture supernatants.

Thin layer chromatography was performed as described previously (1). Aliquots were removed from cultures during growth with proline betaine and centrifuged at 5,000 x g in a microcentrifuge for 5 minutes. Standards included *L*-proline betaine and *N*-methyl *L*-proline, synthesized as described above, and commercially available proline (Thermo-Fisher Scientific, Inc., Waltham MA). Samples (10 µL in total) were spotted onto Silica Gel 60 plates (Merck-Millipore) in 2.5 µL aliquots, which were dried before addition of the next aliquot. The mobile phase was 80% (vol/vol) phenol in H_2O. After development, the plates were dried and stained with bromocresol green that had been previously titrated with 0.1 M NaOH until dark green color began to appear. Plates were then scanned and the entire image adjusted using Adobe Photoshop to increase contrast between the background and the stained methylated amines.

For accurate mass analysis, unstained TLC spots were cut from the plate using stained lanes from the same plate as a reference. Cut segments were placed in a microcentrifuge tube with 1 mL H_2O, into which the silica substrate was scraped from the backing, and the backing was removed. The tubes were incubated for 1 hour at room temperature with periodic vortex mixing. The silica was removed by centrifugation at 16,100 x g for 2 minutes. The supernatant was vacuum concentrated to dryness, then dissolved in acetonitrile/water/acetic acid (1:1:0.001) followed by analysis using a Bruker MaXis ESI Q-Tof instrument (Bruker, Billerica, MA) using direct infusion. Sodium iodide was employed to tune and calibrate the instrument for accurate mass measurement which was operated in positive mode with a capillary spray voltage of 4 to

4.5 kV. The analysis temperature was 200 °C, and the scan range was set between 50-1500 m/z.

Proteomic sample preparations. Two sets of 4 replicate cultures (10 mL each) were grown on defined LS medium. One set was grown with 50 mM *L*-proline betaine, while the second set was grown on 50 mM *D,L*-lactate. Each culture was anaerobically harvested in mid-log phase ($OD_{600} \approx 0.45$) by centrifugation. The cells were resuspended in 22 mM potassium phosphate buffer, pH 7.2, pelleted again, then frozen in liquid nitrogen, and stored at -80 °C until needed. From each cell pellet, an aliquot was lyophilized then subjected to cyrogrinding and the subsequent addition of RIPA Buffer [50mM Tris-HCl 7.4, 150 mM NaCl, 1% NP-40, 0.25% Na-deoxycholate, 1 mM PMSF, 1x Roche Protease inhibitor cocktail, 1x HALT Phosphatase inhibitor cocktail (Thermo Fisher Scientific, Waltham MA), 25U/ml Benzonase nuclease (Novagen, Madison, WI)] at a ratio of 12 μl/mg cell pellet (original wet weight). The suspension was sonicated on ice twice for 10 seconds with a 10 second interval between pulses, then centrifuged at 13,000 RPM in a microfuge, the supernatant retained, and lysis buffer (7 M urea, 2M thiourea, 4% CHAPS in 30 mM Tris-HCl pH 8.5) added to the remaining pellet at a ratio of 2.5 μl/mg cell pellet (original wet weight). The resuspended pellet was subjected to two five-second sonication bursts on ice, then centrifuged again at 13,000 RPM. The two supernatants were combined and the protein concentration determined by Bradford assay using BSA as standard. The protein was then precipitated with TCA (final concentration 20% v/v) and centrifuged at 13,000 RPM for 20 min. The pellets were washed with ice-cold acetone to remove any remaining acid and resuspended in 50 μl of 0.1% Rapigest (Waters Corp., Milford, MA) dissolved in 50 mM ammonium

bicarbonate. Five µl of DTT (5µg/µL in 50 mM ammonium bicarbonate) was added and the sample was incubated at 56°C for 15 min. Five µL of iodoacetamide (15 mg/ml in 50mM ammonium bicarbonate) was then added and the sample was kept in the dark at room temperature for 30 min. Sequencing grade-modified trypsin (Promega, Madison WI) in 50 mM ammonium bicarbonate was added to the sample at an estimated ratio of 1:50 trypsin-substrate and incubated at 37 °C overnight. After digestion, a final concentration of 0.1% TFA was added prior to incubation at 37 °C for 30 min to precipitate Rapigest. The sample was centrifuged and the supernatant dried and respsupended in 50 mM acetic acid for the measurement of peptide concentration using a Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham MA).

Peptide analysis with Orbitrap Fusion Mass Spectrometer. Prior to MS/MS, samples were subjected to 2-D LC separation using either an Acquity UPLC M-Class LC (Water's Corp., Milford, MA) or a Thermo Scientific 2D RSLC HPLC system. Peptides were separated using the same basic protocol on either instrument. Samples (12 µg) were first separated on a 5 mm x 300 µm BEH C18 column with 3 µm particle size and 130 Å pore size. Solvent A was composed of 20 mM ammonium formate, pH 10 and solvent B was 100% acetonitrile. Peptides were eluted from column in 8 successive fractions using 9.5, 12.4, 14.3, 16.0, 17.8, 19.7, 22.6 and 50% solvent B. Each eluted fraction was then trapped, diluted, neutralized, and desalted on a µ-Precolumn Cartridge (Thermo Fisher Scientific) for the second dimensional separations performed with a 15 cm x 75 µm PepMap C18 column (ThermoFisher Scientific, Waltham, MA) with 3 µm particle size and 100 Å pore size. For M-Class LC, the flow rate for the analytical column was 500 µL/min. The gradient was: for 0 to 3 min, 3%

solvent B; for 3 to 45min, 55% solvent B; for 45 to 47 min, 85% solvent B; and finally solvent B was kept at 85% for another 2 min before bought back to 3% for 10 min to equilibrate the column for the next separation. For the Thermo Scientific 2D RSLC HPLC system, the flow rate for the analytical column was 500 µl/min. The gradient was: from 0 to 5 min, 2% solvent B; from 5 to 38 min, 35% solvent B; from 38 to 46 min, 35-55% solvent B; and for 46 to 47 min, 55-90% solvent B. Solvent B was then kept at 90% for 1 min before being returned to 2% solvent B. The system was equilibrated for 11 min before the next separation. MS/MS data was acquired with a spray voltage of 1.7 KV and a capillary temperature of 275 °C. The scan sequence of the mass spectrometer was based on the preview mode data dependent TopSpeed™ method: the analysis was programmed for a full scan recorded between *m/z* 400 – 1600 and a MS/MS scan to generate product ion spectra to determine amino acid sequence in consecutive scans starting from the most abundant peaks in the spectrum in the next 3 seconds. To achieve high mass accuracy MS determination, the full scan was performed at FT mode and the resolution was set at 120,000. The AGC Target ion number for FT full scan was set at 2×10^5 ions, maximum ion injection time was set at 50 ms and micro scan number was set at 1. MS/MS was performed using ion trap mode to ensure the highest signal intensity of MS^2 spectra using both CID (for 2+ and 3+ charges) and ETD (for 4+ to 7+ charges) methods. The AGC Target ion number for ion trap MS^2 scan was set at 1000 ions, maximum ion injection time was set at 100 ms and micro scan number was set at 1. The CID fragmentation energy was set to 35%. Dynamic exclusion is enabled with a repeat count of 1 within 60s and a low mass width and high mass width of 10 ppm.

Sequence information from the MS/MS data was processed by converting raw files into a mgf files using MS convert (ProteoWizard) and then mgf files from each of the fractions was merged into a single merged file (mgf) using an in-house program, RAW2MZXML_n_MGF_batch (merge.pl, a Perl script). Isotope distributions for the precursor ions of the MS/MS spectra were deconvoluted to obtain the charge states and monoisotopic *m/z* values of the precursor ions during the data conversion. The resulting mgf files were searched using Mascot Daemon by Matrix Science version 2.5.1 (Boston, MA) and searched against databases of the protein sets encoded in genomes of *E. limosum* ATCC 8486 (2) and *E. limosum* KIST612 (3) maintained at the National Center for Biotechnology Information. The mass accuracy of the precursor ions was set to 10 ppm and an allowance for selection of one ^{13}C peak for each identified peptide was also included into the search. The fragment mass tolerance was set to 0.5 Da. Considered variable modifications were oxidation (Met), deamidation (Asn and Gln) and carbamidomethylation (Cys). Four missed cleavages for the enzyme were permitted. A decoy database was also searched to determine the false discovery rate (FDR) and peptides were filtered according to the FDR. The significance threshold for peptide identification was set at $p<0.05$, Only proteins identified with <1% FDR as well as a minimal of 2 peptides are accepted for quantitation. Label free quantitation was performed using the spectral count approach, in which the relative protein quantitation is measured by comparing the number of MS/MS spectra identified from the same protein in each of the multiple LC/MSMS datasets. Scaffold (Proteomic Software, Inc., Portland OR) was used to compile data and assign emPAI values to estimate the mol % of each identified protein within the total set of identified proteins (4). Student-t test was

performed using Scaffold to evaluate if the fold change for certain proteins between growth conditions is significant ($p<0.05$).

Cloning, expression, and purification of recombinant proteins. All primer pairs used for gene amplification and cloning are listed in Table S1. The genes encoding MtpB, MtqC, MtqA, and RamQ were amplified from *E. limosum* genomic DNA and inserted into pSpeed using the polymerase incomplete primer extension (PIPE) method (5). The pSpeed-based *mtqC* expression vector was modified by the removal of the N-terminal 6x His tag and the insertion of a C-terminal Strep II tag using the PIPE method.

MtqA was produced in *E. coli* BL21 (DE3) grown in LB media containing 50 mg/L kanamycin at 37 °C with shaking until an OD_{600nm} of 0.4 to 0.6 was reached. Induction was by addition of 1 mM IPTG and incubation at 25 °C with shaking for three additional hours. Cells were pelleted, frozen in liquid N_2, and stored at -80 °C until use. Pellet from 2 L of culture were resuspended in 30 mL buffer A (20 mM sodium phosphate, 500 mM sodium chloride, and 20 mM imidazole, pH 7.2) and lysed using a French pressure cell at 20,000 psi. The lysate was centrifuged at 41,400 x g for 30 minutes at 4 °C. Clarified extract was loaded onto a 1 mL HisTrap FF crude column (GE Life Sciences) and eluted with a 40 mL linear gradient of 100% buffer A to 100% buffer B (buffer A with 500 mM imidazole). The eluted peak containing MtqA was exchanged into storage buffer (50 mM sodium phosphate, 150 mM sodium chloride, and 10 % glycerol, pH 7.2) and kept at -80 °C until use.

MtpB was produced as described above, with the exception that 1 mM arabinose was used in place of IPTG, and all steps from lysis on were performed

anaerobically using cell pellet from 4 L of culture. Resuspension of cell pellet, assembly of the French pressure cell, and chromatography took place in an anaerobic chamber (Coy Lab Products, Grass Lakes MI) with the addition of 2 mM DTT and 10% glycerol to buffers A and B. Following anaerobic HisTrap chromatography, MtpB was further purified using two Superose 12 10/300 GL columns (GE Life Sciences) connected in tandem. The MtpB-containing peak from the HisTrap eluate was loaded in a volume of 250 µL and eluted by isocratic flow of storage buffer. MtpB was stored in an anaerobic vial on ice in a 4 °C cold-room in storage buffer supplemented with 2 mM DTT and 10% glycerol. Activity was retained over 2 weeks under these conditions.

E. coli lysate containing MtqC apoprotein was produced as described for MtqA, except that pellet from 4 L of culture was resuspended in 30 mL ST binding buffer (100 mM Tris, 300 mM sodium chloride, and 1 mM EDTA, pH 8.0). Clarified extract was loaded onto a Strep-Tactin XT high capacity column (5 ml total volume, IBA Lifesciences, Gottingen, Germany), followed by a 25 mL of ST binding buffer. Apo-MtqC was eluted with 25 mL ST binding buffer containing 50 mM biotin.

The gene encoding RamQ was expressed anaerobically in *E. coli* SG13009 in LB media containing 50 mg/L kanamycin, supplemented with 80 mM glucose, 80 mM sodium fumarate, and 44 mM potassium phosphate buffer, pH 7.2. Cultures were grown at 32 °C and induction was by anaerobic addition of 1 mM IPTG once OD_{600nm} reached 0.3-0.4. Anaerobic cysteine HCl (1 mM) and ferrous ammonium sulfate (0.1 mM) were added at the time of induction and again after two more hours of incubation at 32 °C. An OD_{600nm} of approximately 1.0 was reached four hours post-induction at which time the

cells were pelleted anaerobically, resuspended in 50 mM MOPS, pH 7.0, pelleted again and frozen in liquid N_2 and stored at -80 °C until used. Anaerobic lysis and HisTrap chromatography were performed as described for MtpB. The RamQ-containing peak from the HisTrap eluate was exchanged into storage buffer and stored in an anaerobic vial at -80 °C until use.

Reconstitution of MtqC holoprotein with cobalamin hydrochloride.

Recombinant MtqC apoprotein was concentrated to 5 mL using Amicon Ultra-0.5 10K centrifugal filter devices in a tabletop centrifuge, then made anaerobic by repeated flush/evacuation using N_2 on a vacuum manifold. The entire volume was then added to 20 mL of anaerobic 3.5 M glycine betaine, 1 mM hydroxocobalamin hydrochloride, 10 mM dithiothreitol in 50 mM Tris, pH 7.2, and the solution was gently stirred at 4 °C in darkness overnight (6). Subsequent steps were performed in the anaerobic chamber. MtqC holoprotein was concentrated to 500 µL prior to purification by Superose 12 as described for MtpB, but with two runs, each with an injection of 250 µL. MtqC holoprotein was stored in an anaerobic vial at -80 °C until use. Incorporation of cobalamin was measured using the absorbance change following dicyano derivatization using a $\Delta\varepsilon_{368nm}$ of 1.7×10^4 $M^{-1}cm^{-1}$ (7). Preparations varied from 0.94 to 1.02 moles cobalamin:mole MtqC.

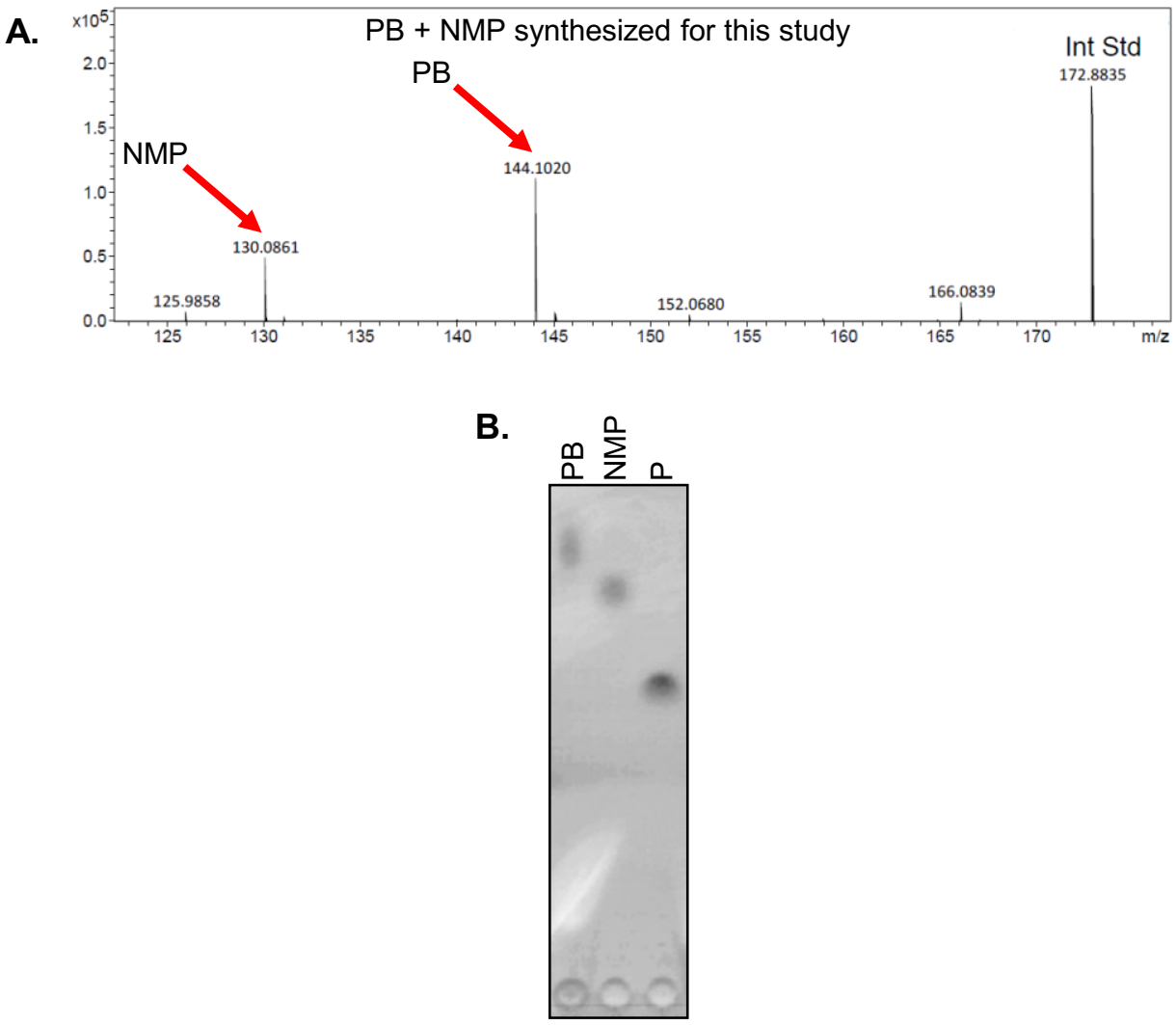

FIG S1 (A) Mass spectrometry of *L*-proline betaine (PB) and *N*-methyl *L*-proline (NMP) synthesized for this study. For comparison, the calculated theoretical m/z for *L*-proline betaine and *N*-methyl-*L*-proline are 144.1019 and 130.0863, respectively. Int Std; internal standard. (B) Thin layer chromatography of PB and NMP synthesized for this study. Commercially available *L*-proline (P) was also included for comparison.

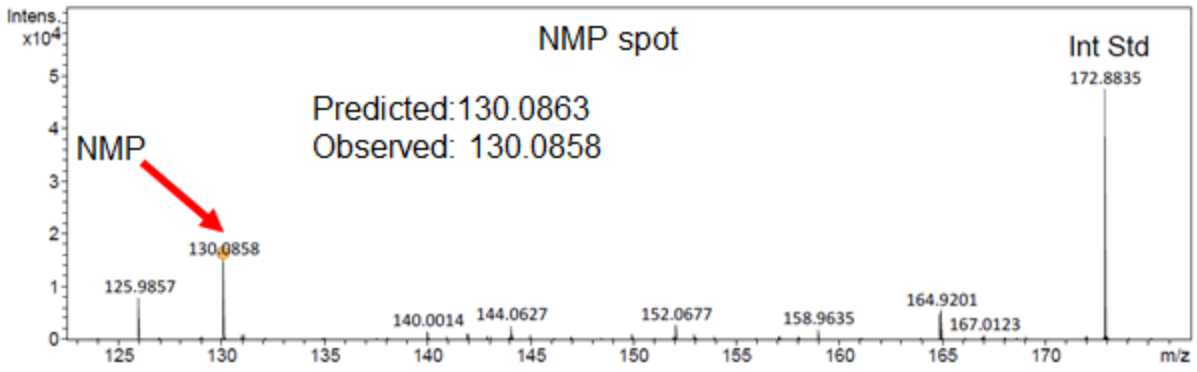

FIG S2 Mass spectrometric analysis confirms *N*-methyl proline (NMP) is a product of proline betaine catabolism by *E. limosum*. The area with an Rf corresponding to the putative *N*-methyl proline spot was extracted from an unstained lane loaded with supernatant from log phase. The extract was then subjected to mass spectrometry as described in Experimental Procedures. The predicted m/z value corresponds to $C_6H_{12}NO_2$ for the [M+H]+ ion of *N*-methyl proline.

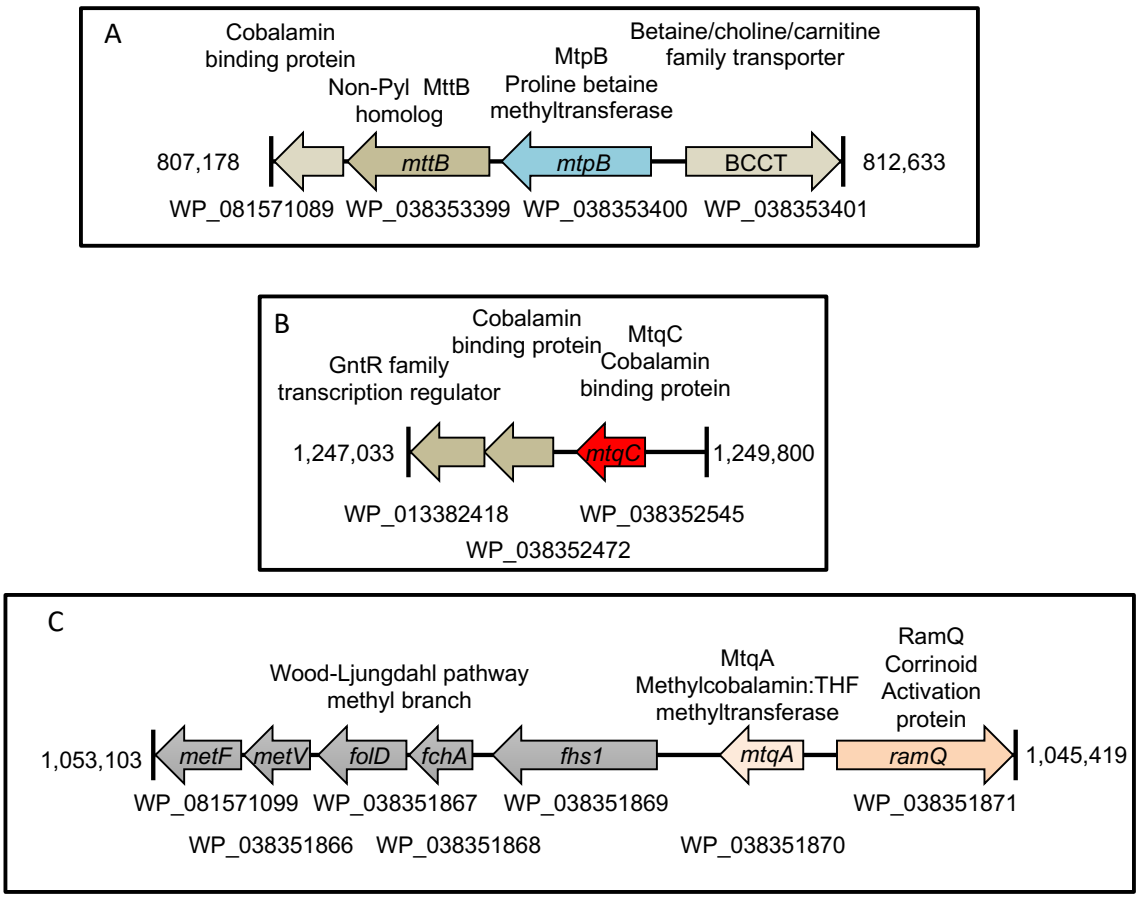

FIG S3 Genomic context of the genes encoding (A) MtpB (blue), (B) MtqC (red), and (C) MtqA and RamQ (pink). The numbers on the left and right of each genomic segment are from the nucleotide numbering used for the complete *E. limosum* ATCC 8486 genome maintained at the National Center for Biotechnology Information under accession number NZ_CP019962.1.

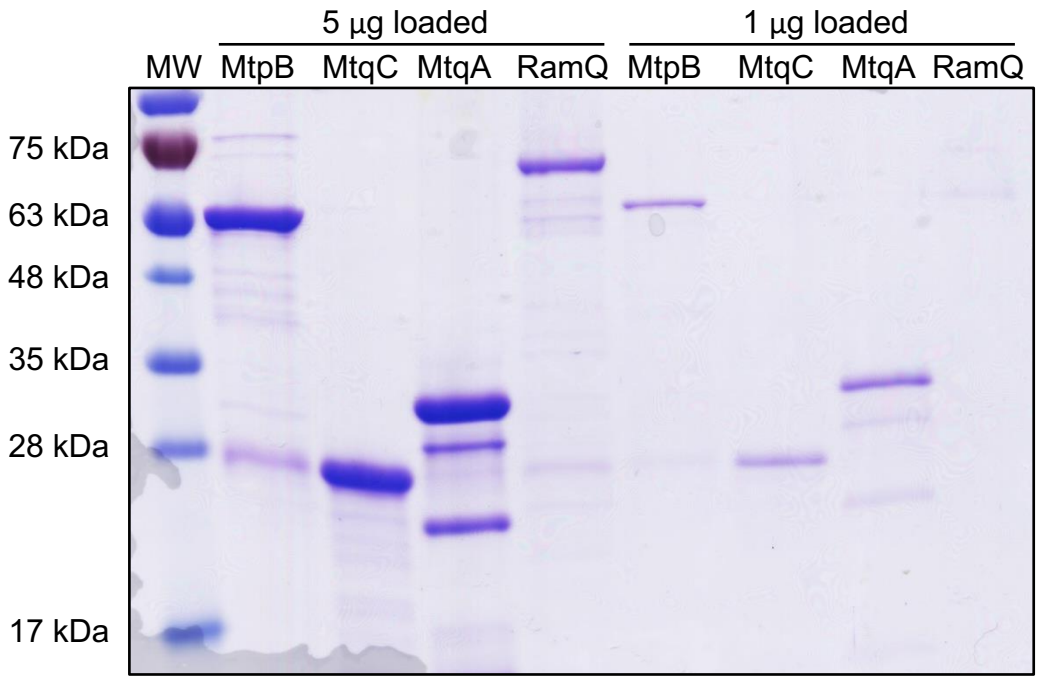

FIG S4 SDS-PAGE of recombinant proteins used in this study. Expected molecular masses are as follows: RamQ, 66.7 kDa; MtpB, 56.7 kDa; MtqC, 23.4 kDa; MtqA, 32.1 kDa. The masses of the molecular weight standards (MW) are indicated on the left side of the gel.

SUPPLEMENTARY TABLE S1 Primers used in this study

Name	Nucleotide sequence, 5' to 3'	Source
EI-MtpB-IF	CTGTACTTCCAGGGC|ATGTACGTTAACAG AAGATTTTATGAC	This study
EI-MtpB-IR	AATTAAGTCGCGTTA|AAACTTTTCGTCTT CAGGCAG	This study
DUNCAN-MtqC-F	CTGTACTTCCAGGGC|ATGGCAGATTGGAA AAATTTAACACAGG	Kountz, Zhang, and Krzycki, manuscript submitted
DUNCAN-MtqC-R	AATTAAGTCGCGTTA|TCAGCCAACCCATG CCTGGCAGATCTTAACG	Kountz, Zhang, and Krzycki, manuscript submitted
MtqC-RemoveNTag-F	acataccc**ATG**GCAGATTGGAAAAATTTAA CAC	This study
MtqC-RemoveNTag-R	TCTGCCATgggtatgtatatctccttctta aag	This study
MtqC-CStrep-F	GGAGTCACCCACAATTTGAAAAG**TGA**CGCG ACTTAATTAACTCG	This study
MtqC-CStrep-R	CAAATTGTGGGTGACTCCAtcctccGCCAA CCCATGCCTG	This study
DUNCAN-MtqA-F	CTGTACTTCCAGGGC|ATGATTATTATTGG AGAAAAACTGAACGG	Kountz, Zhang, and Krzycki, manuscript submitted
DUNCAN-MtqA-R	AATTAAGTCGCGTTA|TTATTTCGCCTGAA CTGGTCCAATCTTGCC	Kountz, Zhang, and Krzycki, manuscript submitted
DUNCAN-RamQ-F	CTGTACTTCCAGGGC|ATGCGCGTATTGTT CCCATTACTTGAGG	Kountz, Zhang, and Krzycki, manuscript submitted
DUNCAN-RamQ-R	AATTAAGTCGCGTTA|TTAAATCATCACAT TATCCCCGAACATGG	Kountz, Zhang, and Krzycki, manuscript submitted
pSpeed-F	taacgcgacttaattAAACGGTCTCCAGCT TGGCTGTTTTGGC	(5)
pSpeed-R	gccctggaagtacagGTTTTCGTGATGATG ATGATGATG	(5)

References

1. Ticak, T., Kountz, D. J., Girosky, K. E., Krzycki, J. A., and Ferguson, D. J. (2014) A nonpyrrolysine member of the widely distributed trimethylamine methyltransferase family is a glycine betaine methyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4668-4676
2. Song, Y., Shin, J., Jeong, Y., Jin, S., Lee, J. K., Kim, D. R., Kim, S. C., Cho, S., and Cho, B. K. (2017) Determination of the genome and primary transcriptome of syngas fermenting *Eubacterium limosum* ATCC 8486. *Sci. Rep.* 10.1038/s41598-017-14123-3
3. Roh, H., Ko, H. J., Kim, D., Choi, D. G., Park, S., Kim, S., Chang, I. S., and Choi, I. G. (2011) Complete genome sequence of a carbon monoxide-utilizing acetogen, *Eubacterium limosum* KIST612. *J. Bacteriol.* **193**, 307-308
4. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **4**, 1265-1272
5. Klock, H. E., Koesema, E. J., Knuth, M. W., and Lesley, S. A. (2008) Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. *Proteins* **71**, 982-994
6. Schilhabel, A., Studenik, S., Vodisch, M., Kreher, S., Schlott, B., Pierik, A. J., and Diekert, G. (2009) The ether-cleaving methyltransferase system of the strict anaerobe A*cetobacterium dehalogenans*: analysis and expression of the encoding genes. *J. Bacteriol.* **191**, 588-599
7. Grahame, D. A. (1991) Catalysis of acetyl-CoA cleavage and tetrahydrosarcinapterin methylation by a carbon monoxide dehydrogenase-corrinoid enzyme complex. *J. Biol. Chem.* **266**, 22227-22233