Supplementary Information

The abundant marine bacterium *Pelagibacter* simultaneously catabolizes dimethylsulfoniopropionate to the gases dimethyl sulfide and methanethiol

Jing Sun[1], Jonathan D. Todd[4], J. Cameron Thrash[5], Yanping Qian[2], Michael C. Qian[2], Ben Temperton[6], Jiazhen Guo[7], Emily K. Fowler[4], Joshua T. Aldrich[8], Carrie D. Nicora[8], Mary S. Lipton[8], Richard D. Smith[8], Patrick De Leenheer[3], Samuel H. Payne[8], Andrew W. B. Johnston[4], Cleo L. Davie-Martin[1], Kimberly H. Halsey[1] and Stephen J. Giovannoni[1]*

[1]Department of Microbiology, [2]Department of Food Science, [3]Department of Mathematics, Oregon State University, Corvallis, Oregon 97331, USA; [4]School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK; [5]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, 70803, USA; [6]Department of Biosciences, University of Exeter, Exeter, UK [7]Qingdao Aquarium, Qingdao, Shandong 266003, China. [8]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, 99352, USA.

Contents:

5 Supplementary Notes

9 Supplementary Figures

4 Supplementary Tables

1 Supplementary Method

Supplementary Notes

Supplementary Note I: DMSP catabolic pathways in *Pelagibacterales*.

Supplementary Figure 1 shows DMSP catabolic pathways in diagrammatic form. Genes that are predicted to encode enzymes for the metabolism of acrylate were present in most *Pelagibacterales* strains. Acrylate can be metabolized to 3-hydroxypropionate (3-HP) by the action of AcuNK[1]. However, while AcuK is found in all strains, AcuN is not a core gene among the *Pelagibacterales*. 3-HP can be oxidized to an intermediate, malonate semialdehyde (mal-SA) and then acetyl-CoA, by an alcohol dehydrogenase (DddA) and mal-SA dehydrogenase (DddC), respectively. The predicted homologs for both *dddA* and *dddC* are found in all *Pelagibacterales* strains. *yhdH* is a homolog of *acuI* that recently was implicated in reductive 3-HP metabolism in *Rhodobacter sphaeroides* and *R. pomeroyi,* and was proposed as part of a novel pathway that converts acrylate to propionyl-CoA via acrylyl-CoA in those organisms [2,3]. This gene has homologs in all *Pelagibacterales* strains except HIMB59. The enzyme that converts acrylate to acrylyl-CoA has recently been identified as a propionate-CoA ligase (PrpE) in *R. pomeroyi*[4]. PrpE carries out multiple functions, and is involved in a third pathway for acrylate degradation via transformation to propionate and propionyl-CoA by acrylate reductase and PrpE, respectively. Acrylate reductase is missing from *Pelagibacterales,* but PrpE is present in all strains.

Supplementary Note II: Transcriptional and proteomic analysis of the DMSP metabolic pathways in *Pelagibacterales* strain HTCC1062.

To determine whether DMSP catabolic pathways are regulated in HTCC1062, we first examined changes in transcription in response to the addition of DMSP to the growth medium. Briefly, HTCC1062 cells were grown in autoclaved, filtered artificial seawater (ASW) media (1 mM NH_4Cl, 100 μM KH_2PO_4, 1 μM $FeCl_3$, 80 μM pyruvate, 40 μM oxaloacetate, 40 μM taurine, 50 μM glycine, 50 μM methionine and excess vitamins[5]) in the presence and absence of 1 μM DMSP. Changes in transcription were measured with Affymetrix GeneChip oligonucleotide microarrays (Microarray data was deposited in GEO (GSE65845)), as described in a previous publication from our research group[6]. Differences were deemed significant when genes exhibited either a 2-fold change or greater between treatments and controls, and when the fold change value indicating differential expression was supported by a Q-value of 0.05 or less (data not shown). No significant changes were observed in the expression of genes involved in DMSP metabolism (e.g., the genes in Supplementary Figure 6), including *dddK* and *dmdA* genes, indicating that these genes and pathways are not transcriptionally regulated.

Because transcriptional analysis reveals changes in transcription, but not whether genes are expressed and translated, we applied quantitative proteomics using the isobaric tag for relative and absolute quantitation (iTRAQ) method to compare DMSP catabolic pathway

proteins in cells grown in the presence and absence of DMSP. All proteins involved in DMSP metabolic pathways (AcuIK, DddAC, PreE, DmdABC) were detected, except DddK. The absence of DddK does not show that it is not present, but rather is likely an unfortunate consequence of DddK not producing detectable peptides in the iTRAQ experiments. The DddK peptide sequence exhibits an unusually small number of tryptic peptides that are predicted to produce MS/MS spectra. Studying data from many other experiments, we found that one peptide that likely originates from DddK was frequently detected in cells grown under a variety of conditions, but, since we don't score any protein as 'detected' unless two peptides are observed, DddK is marked as 'unobserved' in all of our work. In the iTRAQ experiments not even this single DddK peptide was observed; while unfortunate, this is not surprising because iTRAQ experiments have unique biases associated with peptide chemistry that can cause additional peptides to be missed.

In accord with the transcription data, the iTRAQ experiment revealed few significant changes in the abundance of proteins for DMSP metabolism (Supplementary Figure 7; Supplementary Table 2). Two proteins in the predicted pathways of DMSP cleavage and demethylation were among the significantly differentially abundant proteins, but the changes in protein abundance were small: DmdC was 25% more abundant in cultures amended with DMSP than in those amended with methionine, and DddC was 20% more abundant in cultures amended with methionine than those amended with DMSP.

To summarize, the microarray and iTRAQ data collectively provide compelling evidence that both pathways for DMSP catabolism are constitutively produced by HTCC1062 cells, whether DMSP is present in the medium or not.

Supplementary Note III: Enzymatic activities of DddKs in *Pelagibacterales* strains.
DddK homologs from strains HTCC9022 and HMIB5 were cloned and expressed, and as expected the *E. coli* transformants showed DMSP lyase activities similar to that observed in the original DddK from strain HTCC1062 (Supplementary Table 1). We also tested the most distant DddK-like proteins of strains HIMB114 (28% identity) and IMCC9063 (26% identity) from *Pelagibacterales* IIIa subclade[7]. However, the cloned genes from these strains had no DMSP lyase activity (were similar to *E. coli* with the empty vector).

Supplementary Note IV:

Twelve proteins were identified as having a differential expression of > 1.5-fold between DMSP and methionine treatments (Supplementary Figure 7; Supplementary Table 2). As MetF catalyzes the conversion of CH_3-THF to CH_2-THF, up-regulation of MetF in the presence of DMSP is consistent with increased concentrations of CH_3-THF resulting from

conversion of DMSP to MMPA by DmdA. Similarly, GcvT is required alongside FolD in the conversion of CH_2-THF to CHO-THF. SAR11_1724 is a protein of unknown function containing a YGGT domain conserved among integral membrane proteins of unknown function.

Quantitative proteomics provided evidence of up-regulation of PepQ (SAR11_0687) in the presence of DMSP. As structurally similar creatinases have previously been found to have DMSP lyase activity[8], SAR11_0687 was synthesized, cloned and overexpressed in *E. coli* as described previously. However, SAR11_0687 showed no evidence of DMSP lyase activity (data not shown), therefore it is unlikely that this protein is responsible for DMSP cleavage in HTCC1062.

Supplementary Note V: DMSP transport
It is reasonable to propose that *Pelagibacter* cells can concentrate DMSP from the environment, where ambient concentrations are ~ 2 nM (Supplementary Table 3), to an intracellular DMSP concentration of greater ~ 180 mM (a concentration of 10^8 fold). Firstly, the transporter in question (OpuAC) was the sixth most highly detected *Pelagibacterales* protein in our study of the Sargasso Sea metaproteome[9], and one of the most highly detected proteins in cultured *Pelagibacter* proteomes. This transporter, which is often annotated as a glycine betaine transporter, is likely responsible for Kiene's observation that in seawater there is an abundant glycine betaine transporter that has a 5 nM half saturation constant and is competitively inhibited by DMSP[10]. Kiene wisely concluded that this is likely a multifunctional transporter that transports DMSP[10]. Note that *Pelagibacter* has been proven to transport both glycine betaine and DMSP, and has a single ABC transporter of the appropriate type. Thus, everything we report here is consistent with published knowledge on this topic.
In addition, it is also a reasonable prediction from thermodynamics. For active transport from 2 nM to 200 mM:
$$\Delta G = 2.303 \, RT \, log10 \, (10^{-9} \div 10^{-1})$$
= 10.9 kcal mole^{-1}
Since the transporter in question is an ABC transporter that relies on ATP hydrolysis, -12 kcal mol^{-1} is available.
The cytoplasmic volume of *Pelagibacter* cells (the same strain used in our paper about DMSP) was at ~0.01 μm^3 [11]. At an internal concentration of 200 mM, the amount of DMSP inside a cell would be (1 × 10^{-17} liters × 0.2 M) 2 × 10^{-18} moles DMSP/cell. This number is consistent with what is known about the biology of the smallest cells. Assuming a spherical cell, the estimated diameter of the cytoplasmic volume is ~0.12 μm.
To calculate flux, assume
$$Flux = 4\pi DRC_\infty$$
where R is the radius
D is the diffusion coefficient
C is the ambient concentration in the fluid
Assuming a D of 10^{-9} m^2 sec^{-1} (perhaps a slight overestimate) with an R of 0.06 μm

Then 1.5 x 10^{-21} moles DMSP per cell $\times$ sec^{-1}

Thus, without factoring in catabolism, it would take ~1300 seconds, or ~22 minutes, to accumulate DMSP to 200 mM, based on the laws of diffusion, active transport, and the assumption of 2 nM ambient DMSP.
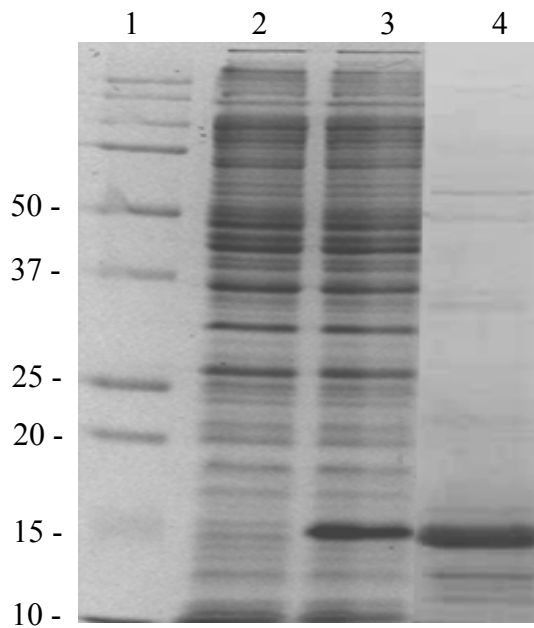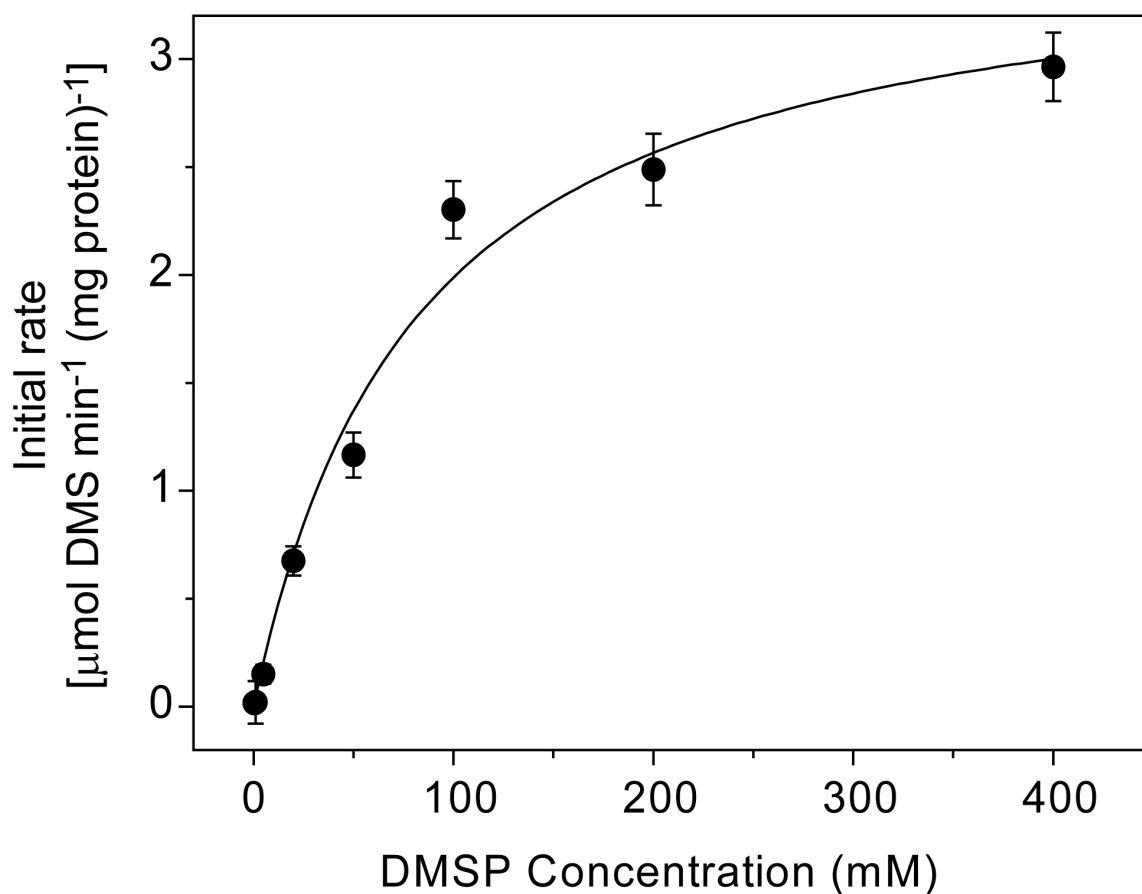
Supplementary Figures
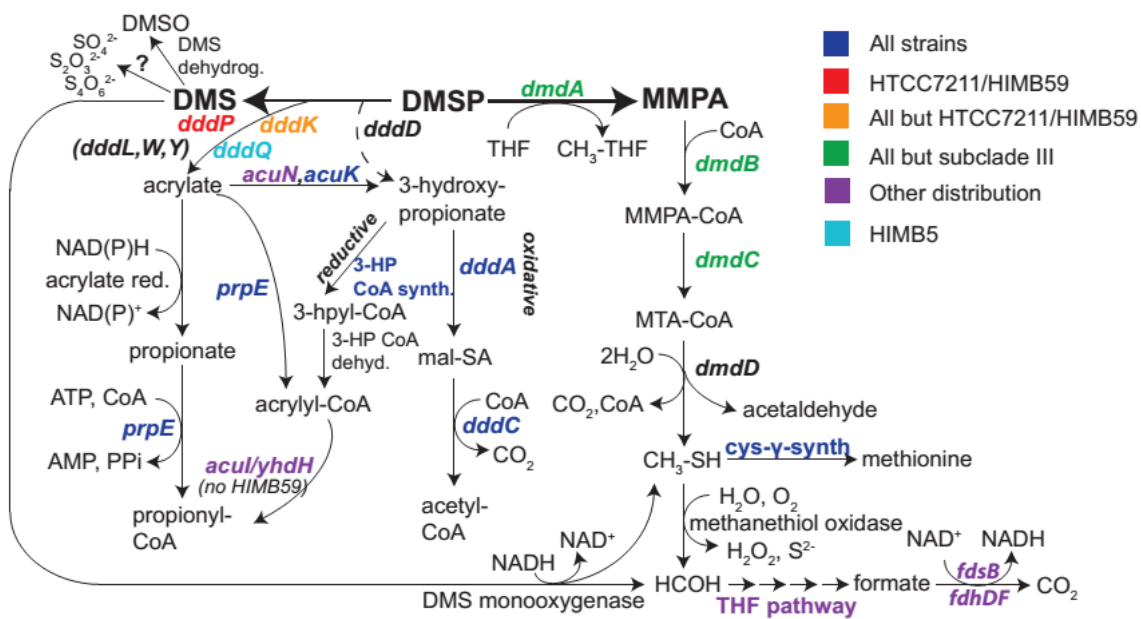
Supplementary Figure 1. DMSP catabolic pathways and homologs identified in *Pelagibacterales* genomes. The dashed line indicates a proposed pathway. Question marks indicate unknown enzymes. All strains belonging to the temperate ocean surface type Ia.1 have *dddK*, but most strains from subtropical ocean sites, type Ia.3, have *dddP*, or *dddQ* (HIMB5) (Supplementary Figure 5). Genes in black did not have identified homologs in *Pelagibacterales*. In the western Sargasso Sea, type IIIa is found in surface waters in the fall. HIMB59, a type V isolated from the subtropical Pacific, represents an early branch of *Pelagibacterales*[12].

```
DddW Rpom    GHQLRPHRHTPPEFYLGLEGSGIVTIDGVPHEIRAGVALYIPGDAEHGTVA
DddQ Rpom    GLYYPFHQHPAEEIYFILAGEAEFLMEGHPPPRRLGPGDHVFHPSGHPHATRT
DddQ HIMB5   NTFYTWHHHEAEEIYFVLSGKAKFESYGDKSEI-LGPNQARFHKSFQPHSLTT
DddK 1062    GGDLTLHYHSPAEIYVVTNGKGILNKSGKLETIKKGDVVYIAGNAEHALKNN
DddK HIMB5   GGNLTLHHHAPDEIYVVTNGSGTLNKSGELEEIKKGDVVYIAGNAKHALQNN
```
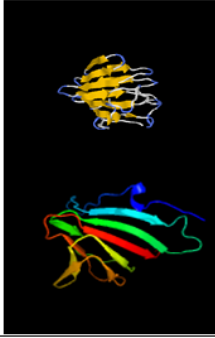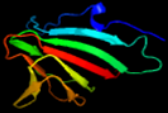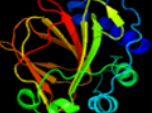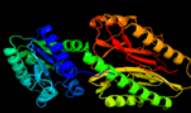
Supplementary Figure 2. Comparison of *Pelagibacterales* DddK and DddQ-like polypeptides with other cupin DMSP lyases. The amino acid sequences of the cupin domains of the known DddW and DddQ polypeptides in *R. pomeroyi* are lined up in comparison with the DddK polypeptides of *Pelagibacterales* strains HTCC1062 and HIMB5 (gene products SAR11_0394 and HIMB5_00004730, respectively) and the DddQ-like product of the HIMB5_00000220 gene of *Pelagibacterales* strain HIMB5. Residues conserved in all polypeptides are shown as red letters. Yellow shading denotes identical residues in the two DddK polypeptides and turquoise shading shows those residues in common in the DddQs of *Pelagibacterales* HIMB5 and of *R. pomeroyi*.

Supplementary Figure 3: Stained SDS-PAGE image showing partial purification of histidine-tagged DddK (SAR11_0394). His-DddK has a predicted molecular mass of 15.8 kDa. Lane 1 = Precision Plus Protein Dual Colour Standards (Biorad); Lane 2 = soluble fraction of wild type *E. coli* BL21; Lane 3 = Soluble fraction of BL21 containing cloned *dddK* in pBIO2206; Lane 4 = DddK-containing sample used for kinetics determinations. 0.1% SDS-PAGE gels prepared with a 15% acrylamide resolving gel, topped with a 6% stacking gel. Loaded samples 1 and 2 are ~2.5 µg and the purified protein sample in Lane 4 was ~250 ng. Gels were run in vertical tanks (ATTO AE-6450) at 150 V for 2 hours in PAGE running buffer [25 mM Tris, 200 mM glycine, 0.1% SDS (w/v)]. Gels were stained with InstantBlue™ (Expedeon). Purity of DddK is 76%, which was determined by gel densitometry using *ImageJ*.

Supplementary Figure 4. Kinetic analysis of enzyme activity. Michaelis-Menten plot for the DMSP lyase activity of DddK (SAR11_0394). For kinetics analysis of DddK initial rates were fitted to the Michaelis-Menten equation using *Origin* software (version 8, Origin Labs). V_{max} was calculated as 3.61 ± 0.27 μmol DMS min^{-1}(mg protein)$^{-1}$, and K_m 81.9 ± 17.2 mM DMSP. DddK (1.2 μg) was in sodium phosphate buffer (pH 8.0.). The R^2 value for the fit is 0.982. Standards errors are indicated (n=3).

Supplementary Figure 5. Evolutionary relationships among *Pelagibacterales* DMSP lyases. An analysis of relationships among the SAR11 lyases we report on was done using BLASTP, protein structural prediction programs (I-TASSER, Phyre2) and SFams protein database[13]. **A)** Predicted protein structures of DddK from HTCC1062, DddQ from HIMB5 and DddP-like protein from HTCC7211 by I-TASSER and Phyre2. DddK, DddQ and DddP-like protein appear to belong to different protein families, although DddK and DddQ both have predicted cupin domains, and may belong to the same superfamily. Comparison of DddK, DddQ and DddP-like protein to the SFam database of hidden Markov models showed that each protein was recruited to a separate SFam model. Furthermore, the top-hitting family for each sequence was additionally found to be included in a separate 'clan' from the others (as defined by the 'precision 80' set of clans provided by SFam; families that reciprocally recruit at least 80% of each other's sequences are placed into a clan together). There is no evidence to show that these three proteins are evolutionarily related, which favors the explanation that these proteins are non-orthologous. **B)** The distribution of DMSP lyase families in *Pelagibacterales* strains. '**+**' means the homolog of DMSP lyase is present. '**-**' means the homolog of DMSP lyase is missing. 'OC' indicates strains that were isolated from the Oregon coast. 'SS' indicates strains that were isolated from the Sargasso Sea. The proteins that have been tested for DMSP lyase activity are squared in red.
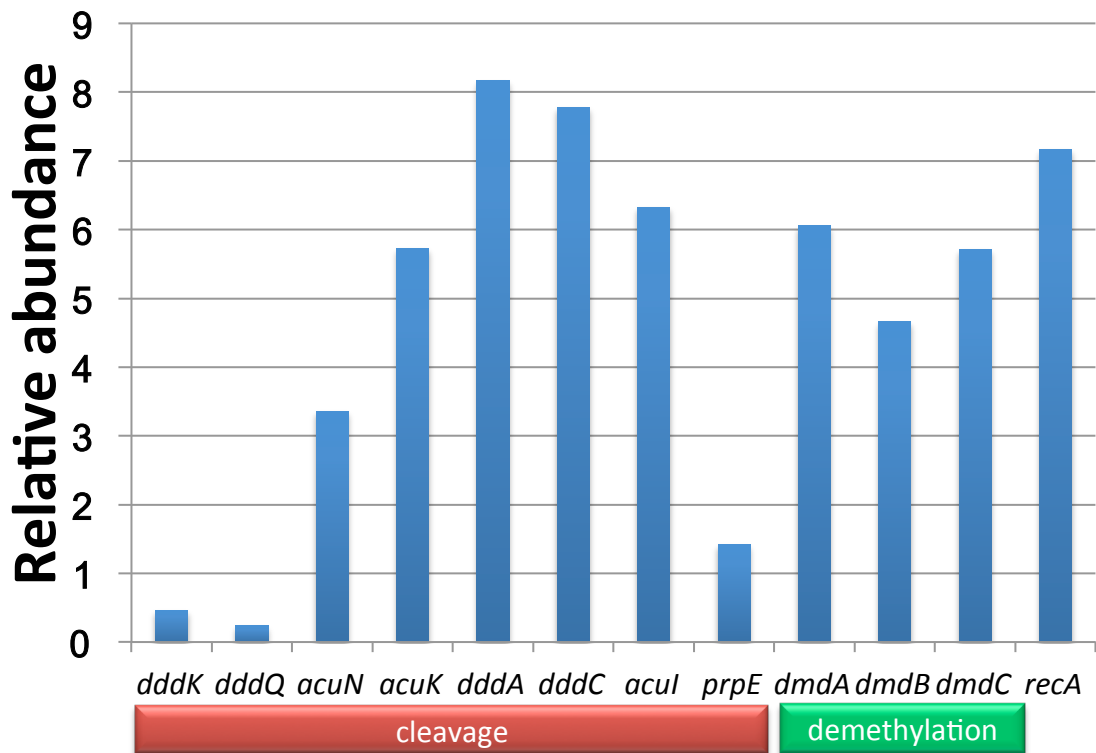
Supplementary Figure 6. The relative abundance of *Pelagibacterales* DMSP catabolism genes in the GOS dataset. Genes were identified as *Pelagibacterales* clades by a reciprocal best BLAST (RBB) approach. For each gene, the count of hits was normalized by gene length, and the normalized values were summed across species. The frequency of the single copy *recA* gene was used to evaluate the abundance of the DMSP metabolism genes.
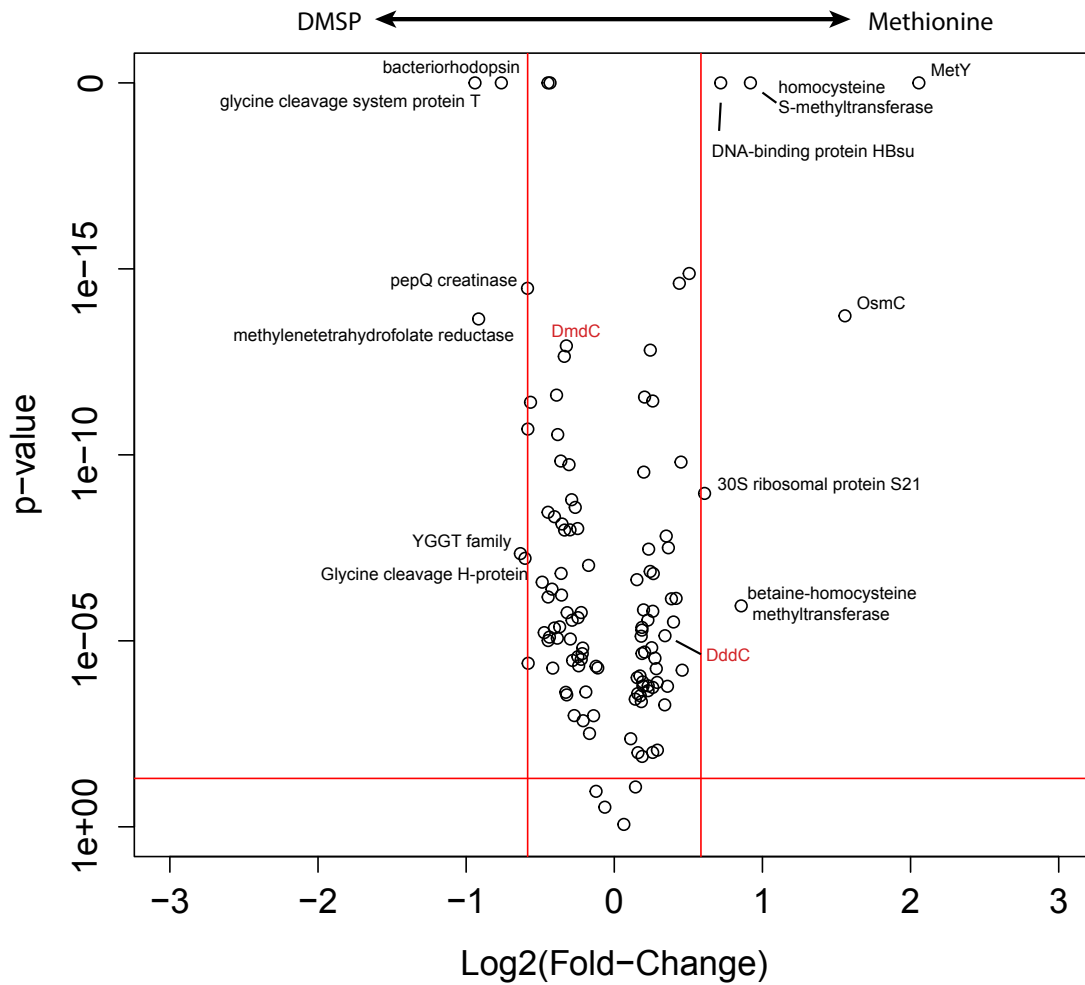
Supplementary Figure 7. Volcano Plot of differential protein expression between HTCC1062 cultures amended with DMSP (left) vs. cultures amended with methionine (right) as determined by quantitative iTRAQ proteomics. Horizontal red line indicates a p-value cut-off of 0.05; Vertical red lines indicate boundaries of 1.5-fold difference in expression.
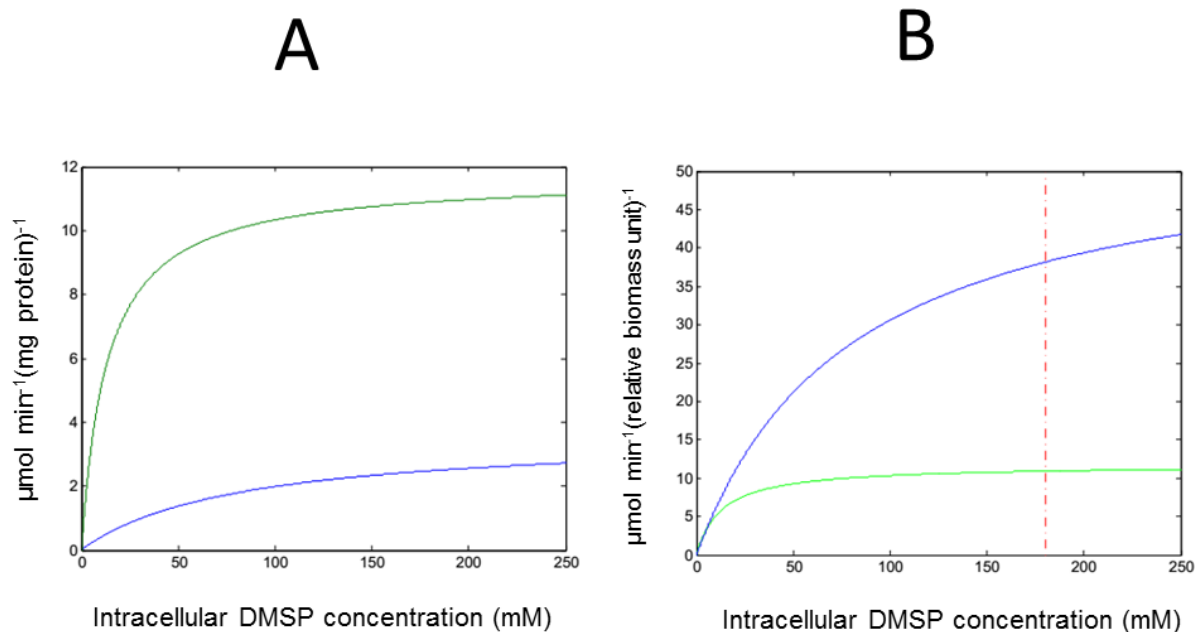
Supplementary Figure 8. Kinetic models for DMS and MeSH production as a function of intracellular DMSP concentration. A) Product formation rates modeled as $dP_1/dt = V_{m1}S/(S+K_{m1})$ for MeSH formation by DmdA (green) and modeled as $dP_2/dt = V_{m2}S/(S+K_{m2})$ for DMS formation by DddK (blue). Parameter values: $V_{m1} = 11.7$ µmol min^{-1} mg^{-1}, $K_{m1} = 13.2$ mM, $V_{m2} = 3.6$ µmol min^{-1} mg^{-1}, $K_{m2} = 81$ mM. Note that the model in Supplementary Figure 8A does not fit the observations shown in Figure 1A, but this is not a surprise because the model shown in Supplementary Figure 8A assumes equivalent amounts of the two enzymes, DddK and DmdA. **B)** To solve for rates of production of DMS and MeSH that match the observations shown in Figure 1A, we assumed an intracellular DMSP concentration of 180 mM, and adjusted activities such that the amount (by weight) of DddK (15.8 kDa) is X-fold (F=15) DmdA (~40kDa), yielding the model seen in Supplementary Figure 8B.
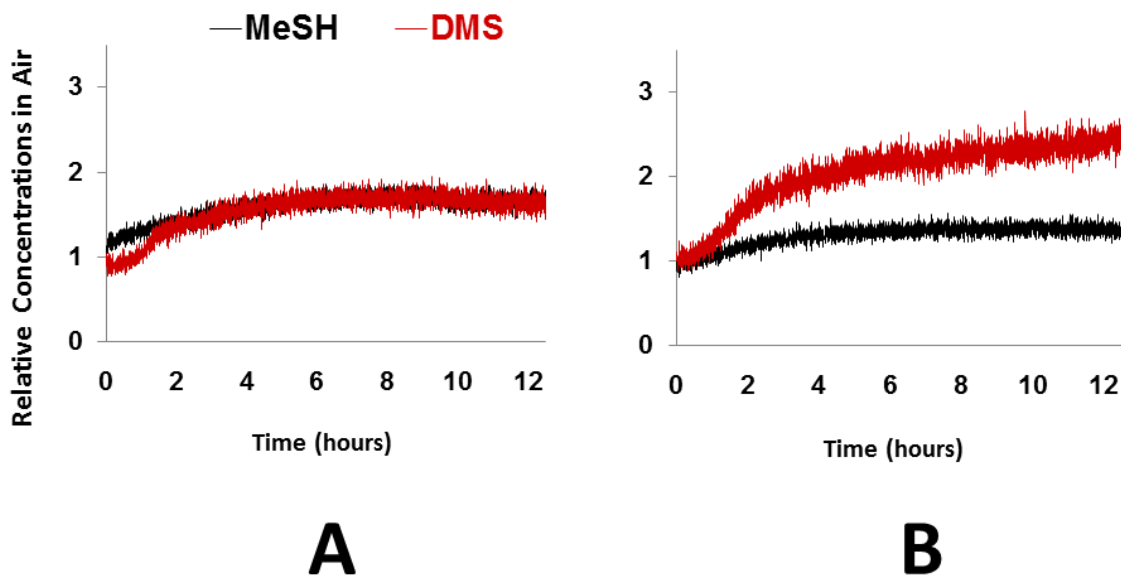
Supplementary Figure 9. Real-time gas-phase MeSH and DMS production measurements by PTR-TOF/MS. The data in panel A and B are from experiments similar to Figure 2A. This experiment was repeated three times. The second and the third repeats are shown here. HTCC1062 cell suspensions were incubated in ASW and subjected to a flow of fine bubbles. DMSP was added at T=0 to cells that had been grown in the absence of DMSP. Measurements are presented in relative concentration units and were normalized to the gas-phase concentrations of MeSH and DMS (m/z 49 and 63, respectively) at T=0.

Supplementary Tables

Supplementary Table 1. Enzymatic activities of DddK proteins from a variety of *Pelagibacterales* strains

Pelagibacterales strain	DMS production (μmol min^{-1} mg^{-1})
803 pUC57 (*E. coli*: empty vector)	0.013 ± 0.000
HTCC1062 (YP_265818)	1.233 ± 0.087
HIMB5 (WP_014953073)	0.588 ± 0.114
HTCC9022 (WP_028037226)	0.975 ± 0.041
HIMB114 (WP_009359929)	0.013 ± 0.003
IMCC9063 (WP_013695448)	0.011 ± 0.000

Supplementary Table 2. Proteins with differential expression > 1.5-fold between HTCC1062 cultures amended with methionine vs. cultures amended with DMSP as determined by quantitative iTRAQ proteomics. Fold change was calculated using the LIBRA module of the Trans-Proteomic Pipeline and by linear mixed-effects model encompassing a fixed treatment effect and random effect for each peptide associated with the protein. Bold-text indicates proteins enriched in DMSP amended cultures.

Locus	Protein	Coverage* (%±1 s.d.)	LIBRA fold change (Met/DMSP)	Lmer fold change (Met/DMSP)
SAR11_1030	MetY	60.6 ± 0.8	3.46	4.12
SAR11_0750	homocysteine S-methyltransferase	30.9 ± 2.5	1.88	1.89
SAR11_0817	non-specific DNA-binding protein HBsu	76.5 ± 0.6	1.06	1.65
SAR11_1172	OsmC	50.6 ± 12.7	2.74	2.94
SAR11_1173	betaine-homocysteine methyltransferase	36.3 ± 1.4	1.82	1.81
SAR11_0578	30S ribosomal protein S21	55.7 ± 6.3	1.15	1.52
SAR11_0625	**proteorhodopsin**	**17.0 ± 3.9**	**0.59**	**0.59**
SAR11_0687	**pepQ creatinase**	**23.8 ± 4.2**	**0.66**	**0.67**
SAR11_0667	**GcvH glycine cleavage H-protein**	**26.5 ± 3.2**	**NA**	**0.66**
SAR11_1264	**MetF methylenetetrahydrofolate reductase**	**28.7 ± 6.8**	**0.57**	**0.53**
SAR11_1265	**GcvT glycine cleavage system protein T**	**34.7 ± 5.2**	**0.43**	**0.52**
SAR11_1724	**YGGT family**	**13.6 ± 0.0**	**0.63**	**0.64**

*Coverage of the total protein length by peptides with a PeptideProphet probability > 0.95

Supplementary Table 3: Concentrations of dissolved DMSP (DMSPd) in the Oceans as reported in the literature.

DMSPd concentration	location	References
10-100 nM	Antarctic coastal waters	14
1-34 nM	Wadden Sea	15
1.1 - 15 nM	Mauritanian upwelling regions	16
Up to 6 nM	Monterey Bay, CA	17
0.1 nM to 11 nM	Northern Gulf of Mexico	18*
1-10 nM	Gulf of Mexico mesotrophic shelf	19*
0.2 – 2.6 nM	Gulf of Mexico oligotrophic oceanic	
5.6 -198.8 nM (180-6360 ng S/L)	The North Sea and English Channel	20*
65 nM	The Bay of Villefrance-sur-mer	21*
2 - 9 nM	Sargasso Sea	22*
2.5 – 11.4 nM	Vineyard Sound, Massachuset	
Up to 30 nM	Delaware Bay	23*
3nM	Western Mediterranean waters	24*
4 - 150 nM	The North Sea	25*
1 - 1.6 nM	Mediterranean	
1.1 – 24 nM	The North Atlantic	

*These measurements were made before Kiene et al.[26] reported that DMSPd concentration measurements can be influenced by filtration artifacts.

Supplementary Table 4. Accession numbers used in Figure 3 and Supplementary Figure 1. This table is provided as a separate file.

Supplementary Methods:

Synthesis and cloning of *Pelagibacterales ddd* genes that encode DMSP lyases

The intact *dddK* genes from SAR11 strains HTCC1062 (SAR11_0394); HTCC9022 (no gene tag available), HIMB5 (HIMB5_00004730); and *dddK*-like genes from *Pelagibacterales* strains HIMB114 (no gene tag) and IMCC9063 (SAR11G3_00808); and *dddQ* from strain HIMB5 (HIMB5_00000220) were each synthesized with codon usage being optimised for expression in *E. coli*. The genes were provided cloned into pUC57 containing the engineered ribosome binding site sequences "TCTAGAAATAATTTTGTTTAACTTTAAG**AAGGAG**ATATACAT**ATG**" (from pET21) incorporated directly upstream of their ATG start codons. These recombinant plasmids were each transformed into *E. coli* 803 on LB media containing 100 μg mL^{-1} ampicillin and assayed for DMSP lyase activity, as described below.

The *dddK* and *dddQ* genes of strains HTCC1062 and HIMB5 respectively were then sub-cloned into the expression vector pET16b using *Nde*I and *Bam*HI, and the resulting plasmids were each transformed into competent *E. coli* BL21 on LB media containing ampicillin. Transformants were used for protein purification, taking advantage of the His tag, which is incorporated into pET16a-based recombinant plasmids.

Assays of DMSP lyase

E. coli 803 or BL21 strains containing cloned *ddd* genes cloned in pUC57 or pET16b, respectively or with the vectors alone, were grown at 37°C in 5 mL of LB broth containing ampicillin (100 μg mL^{-1}) to an OD$_{600}$ of 0.8. The cells were diluted 10-fold into 300 μL M9 media containing 1 or 5mM DMSP and 100 nM IPTG for pET16b clones in 2 mL vials (Alltech Associates). Vials were incubated at 28°C for 18 hours and the concentrations of DMS in the headspace were measured by gas chromatography, using a flame photometric detector (Agilent 7890A GC fitted with a 7693 autosampler) and HP-INNOWax 30 m x 0.320 mm column (Agilent Technologies J&W Scientific) capillary column. The assayed cells were pelleted, re-suspended and washed three times in 1 mL of phosphate buffered saline pH 7.4, then lysed by sonication (6 × 10 s, full power) and the protein concentrations were estimated as described by Bradford[27].

Purification and characterization of DddK

A 50 mL culture of *E. coli* containing the recombinant plasmid in which *dddK* of strain HTCC1062 cloned in pET16b was grown in LB at 28ºC in the presence of 100 nM IPTG. The cells were harvested, pelleted and re-suspended in 1.4 mL NPI-10 buffer (50 mM sodium phosphate, 300 mM sodium chloride, 10 mM imidazole), then lysed by sonication (lane 3, Supplementary Figure 3). The lysate was centrifuged at 13,000 RPM, and the soluble fraction was applied, in two loads of 0.7 mL, to a *Qiagen* Ni-NTA spin column. The column was washed three times with NPI-30 buffer (50 mM sodium phosphate, 300 mM sodium chloride, 30 mM imidazole). Then, the bound His-DddK was eluted at pH 8.0 using NPI-300 buffer (50 mM sodium phosphate, 300 mM NaCl, 300 mM imidazole), an aliquot of which is shown in Supplementary Figure 3.

To determine the enzyme kinetics of DddK, 1.2 µg of the protein (76% pure) was added to 30 µL NPI-10 buffer (pH 8.0) containing varying DMSP concentrations, in sealed vials. Initial reaction rates were measured by assaying DMS in the headspace over a 30 minute incubation period at 22°C.

Nuclear magnetic resonance (NMR)

From a 5 mL culture of *E. coli* BL21, a 2 mL aliquot was re-suspended in 1 mL of 20 mM Tris:D_2O (pH 6.45). Cells were sonicated, the debris removed by centrifugation and the soluble fraction was incubated at 22°C for 1 h in the presence of 3 mM ^{13}C-DMSP[28]. 15 µL of 70% perchloric acid was added per mL^{-1} then incubated on ice for 15 min. NMR analysis of the sample was done as described in Todd *et al*[28].

Bioinformatics analysis and proposed DMSP metabolic pathways

We expanded on the knowledge obtained in Grote *et al*[12] by doing additional homology searches for DMSP metabolism genes using profile hidden markov models (HMMs)[29]. Because they are constructed with a range of probabilistic values for a given site in a protein, profile HMMs are superior to BLAST for finding distantly related homologs[30]. In this workflow, representative genes for the reactions in Figure 3 and Supplementary Figure 1 were obtained from the original publications [3,31,32], searches of NCBI, and E.C. number searches based on figures from [1,33], and references therein (all starting sequence data is included in Table S4). These representative sequences were then searched against a database of profile HMMs created for over 436,000 protein families built with Markov clustering[7] using hmmscan from the HMMER3 package[34] on default settings. SFam HMMs with lowest expect values to the representative sequences were then searched against our *Pelagibacterales* genomes using hmmsearch on default settings. Homologs were classified based on HMMs having comparative expect values to both the representative sequence and a *Pelagibacterales* gene sequence.

Metagenomic analysis

To identify the relative abundance of *Pelagibacterales* genes involved in DMSP metabolism in surface water metagenomes, predicted proteins encoded by homologs of *acuIKN, dddACKPQ, dmdABC* and *prpE* were identified in all 14 genomes from the *Pelagibacter* clade (HTCC1002, HTCC1013, HTCC1062, HTCC7211, HIMB5, HIMB59, HIMB058, HIMB083, HIMB114, HIMB140, HTCC8051, HTCC9022, HTCC9565, IMCC9063) currently in the complete Integrated Microbial Genomes (IMG, http://img.jgi.doe.gov/) database (v. 400). Homologs were determined using a previously described comparative genomics analysis pipeline[12]. Genes within each cluster were used as queries in a TBLASTN (v. 2.2.22+) search against the Global Ocean Survey (GOS) nucleotide database available from CAMERA (http://camera.calit2.net/) with query filtering disabled and default *e*-value cutoff (-seg no –max_target_seqs 10000000). Nucleotide sequences returned from this search were used in a reciprocal best-BLAST (RBB)[35] filtering step against the amino acid sequences in the complete IMG database, returning the best hit to

each nucleotide query (BLASTX, -max_target_seqs 1 –seg no). If the best hit for a nucleotide sequence in the RBB analysis was a protein sequence from the original gene cluster, the nucleotide query was recorded as a successful hit; otherwise it was rejected. For each gene, the count of hits was normalized by gene length, and the normalized values were summed across species. The frequency of the single copy *recA* gene was used to evaluate the abundance of the DMSP metabolism genes.

Quantitative proteomics

HTCC1062 was grown in ASW amended with 100 μM NH_4Cl, 10 μM KH_2PO_4, 0.1 μM $FeCl_3$, 1 mM pyruvate, 500 μM glycine and excess vitamins[5]. Biological triplicate Samples were amended with 1 μM DMSP, or 1 μM methionine and samples with both 1 μM DMSP and 1 μM methionine are treated as positive controls. Cells were all harvested by centrifugation at the same time point in the exponential phase. Prior to harvesting, cultures were treated with chloramphenicol (0.01 g L^{-1}) and protease inhibitor cocktail Set II (0.1 mL L^{-1}, CalBiochem #539132). Cell pellets were immediately stored in - 80℃ prior to iTRAQ analysis at the Pacific Northwest National Laboratory (PNNL).

Each cell pellet was brought up to 100 μL with 8M urea (Sigma-Aldrich, St. Louis, MO) and sonicated in a water bath with ice until the pellet went into solution. The samples were briefly spun and transferred to PCT MicroTube barocycler pulse tubes with 150 μL caps (Pressure Biosciences Inc., South Easton, MA). The MicroTubes were placed in a MicroTube cartridge and barocycled for 10 cycles (20 s at 35,000 psi back down to ambient pressure for 10 s). All of the material was removed from the MicroTubes and transferred to 1.5 mL micro-centrifuge tubes. A Bicinchoninic acid (BCA) (ThermoScientic, Rockford, IL) assay was used to determine protein concentration. Dithiothreitol (DTT) was added to each sample at a concentration of 5 mM (Sigma-Aldrich, St. Louis, MO) and incubated at 60℃ for 1 h. The samples were then diluted 10-fold with 100 mM NH_4HCO_3, and tryptic digestion (Promega, Madison, WI) was performed at a 1:50 (w/w) ratio with the addition of 1 mM $CaCl_2$ to stabilize the trypsin and reduce autolysis. The sample was incubated for 3 h and cleaned via C-18 solid phase extraction (SPE) (Supelco, Bellefonte, PA) on a Gilson GX-274 ASPEC automated SPE system (Gilson Inc., Middleton, WI). The samples were dried to 50 μL and assayed with a Direct Detect IR Spectrometer (EMD Millipore, Billerica, MA) to determine the final peptide concentration.

Each sample set of 3 along with the pooled sample was dried in a speed-vac until near dryness and brought up to 30 μL with 1M Triethylammonium bicarbonate buffer (TEAB). Each sample was isobarically labeled using iTRAQ Multiplex (4-plex) Kits (ABsciex, Framingham, MA) according to the manufactures instructions. Briefly, 50 μL of isopropanol was added to each reagent (iTRAQ 114-117), vortexed and allowed to dissolve for 5 min with occasional vortexing. Reagents were then added to the samples, vortexed and incubated for 1 h at room temperature. The reaction was quenched by adding 100 μL of water to the sample with incubation for 15 min at room temperature. The samples within each set were then combined and dried in the speed vac to remove the organics. Each set was cleaned using Discovery C18 50 mg/1 mL solid phase extraction tubes as described above and once again assayed with BCA to determine the final peptide concentration. There were three technical replicates per sample.

Samples were diluted to a volume of 900 μL with 10 mm ammonium formate buffer (pH

10.0), and resolved on a XBridge C18, 250x4.6 mm, 5 μM with 4.6x20 mm guard column (Waters, Milford, MA). Separations were performed at 0.5 mL/min using an Agilent 1100 series HPLC system (Agilent Technologies, Santa Clara, CA) with mobile phases (A) 10 mM ammonium formate, pH 10.0 and (B) 10 mM ammonium formate, pH 10.0/acetonitrile (10:90). The gradient was adjusted from at 100% A to 95% A over the first 10 min, 95% A to 65% A over minutes 10 to 70, 65% A to 30% A over minutes 70 to 85, maintained at 30% A over minutes 85 to 95, re-equilibrated with 100% A over minutes 95 to 105, and held at 100% A until minute 120. Fractions were collected every 1.25 min after the first 15 min (96 fractions). Every 12th fraction was then combined for a total of 12 samples (each with n=8 fractions pooled) for each of the 3 sets. All fractions were dried under vacuum and 20 μL of 25 mM ammonium bicarbonate was added to each fraction for storage at -20℃ until LC-MS/MS analysis.

Each iTRAQ run generated 152,8902345 spectra, identifying 1196 out of 1324 proteins in HTCC1062. Of these, 112 showed significantly different expression between DMSP and methionine treatments (methionine·DMSP^{-1} fold-change median = 0.913, 1st quartile=0.783, 3rd quartile=1.19).

MS/MS datasets were searched against predicted proteins from *Ca.* P. ubique HTCC1062 using MSGF+ (http://proteomics.ucsd.edu/Software/MSGFPlus.html) with the following search parameters: dynamic methionine oxidation; partial trypsin digest; 20 ppm tolerance. Reporter ion intensities were collected using MASIC[36] and processed through the MAC (Multiple Analysis Chain) pipeline to aggregate, filter and generate cross-tabulated results for processing. Redundant peptide identifications had reporter ion intensities summed for a unique peptide result. Proteins were tested for significantly different expression between cultures grown on DMSP and those grown on methionine using a linear mixed-effects model below encompassing a fixed treatment effect and random effect for each peptide associated with the protein using the lme4 package in R:

$$y_{ij} = \mu + b_i + p_j + \epsilon_{ijk}$$
$$p_j \sim \mathcal{N}(0, \sigma_p^2)$$
$$\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$$
$$i = 1, ..., m \; j = 1, ..., n$$

where μ is the intercept, b is the treatment effect, p is the random intercept associated with each peptide and ϵ is the per observation variation. The resulting linear model was tested for significance of the treatment fixed effect using ANOVA generating a p-value. The p-value was then adjusted for multiple comparisons using Benjamini-Hochberg p-value correction:

$$p.adj = \frac{n}{Rank(pval)} pval$$
$$Rank(pval) = \text{Rank of the p-value in the result set}$$
$$n = \text{number of comparisons}$$

Proteins with significantly different expression between DMSP and methionine treatments were verified with a complimentary analysis using the Trans-Proteomic Pipeline[37]. MS/MS spectra were searched against predicted proteins from HTCC1062 using X!Tandem with identical parameters as before. Spectral matches were filtered using PeptideProphet (p > 0.95) and ProteinProphet (p > 0.90). Relative protein abundances were calculated using LIBRA using a default conditions file for 4-channel iTRAQ.

Linear regression showed that estimated fold-change size between lmer4 and LIBRA analyses for the 12 most differentially expressed proteins was highly correlated (coefficient=1.133 0.04 (s.e.), R^2 = 0.99, F = 795.2, d.f. = 10, p = 7.31×10^{-11}) (SAR11_0667, the glycine cleavage H-protein was removed from the analysis as LIBRA failed to estimate relative abundances for this protein).

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium[38] via the PRIDE partner repository with the dataset identifier PXD001717.

DMSP model
Substrate competition model
Consider the following reaction scheme:

$$S + E_1 \longleftrightarrow C_1 \longrightarrow P_1 + E_1$$
$$S + E_2 \longleftrightarrow C_2 \longrightarrow P_2 + E_1$$
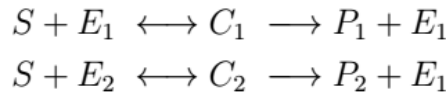
where S; E_1; E_2 and P_1; P_2 represent the concentrations of DMSP, demethylase, DMSP lyase, and MeSH, DMS respectively. The intermediate complexes are C_1 and C_2.

Assuming mass action kinetics, we can write the differential equations for the concentrations of the various compounds (the dots on the left hand side are shorthand for $d=dt$):

$$\begin{aligned}
\dot{S} &= -k_1 S E_1 - k_3 S E_2 + k_{-1} C_1 + k_{-3} C_2 \\
\dot{C_1} &= k_1 S E_1 - (k_{-1} + k_2) C_1 \\
\dot{C_2} &= k_3 S E_2 - (k_{-3} + k_4) C_2 \\
\dot{E_1} &= -k_1 S E_1 + (k_{-1} + k_2) C_1 \\
\dot{E_2} &= -k_3 S E_2 + (k_{-3} + k_4) C_2 \\
\dot{P_1} &= k_2 C_1 \\
\dot{P_2} &= k_4 C_2
\end{aligned}$$

The parameters k_i and k_{-i} are the rate constants of the reactions: k_1 and k_{-1} for the first reversible reaction (k_1 for the forward reaction and k_{-1} for the backward reaction), k_2 for the formation step of the first product P_1, k_3 and k_{-3} for the second reversible reaction, and finally k_4 for the formation step of the second product P_2.

The total concentration of the enzymes E_1 and E_2, both in free and in bound form, is constant: and equal to, say E_1^0 and E_2^0 respectively:

$$E_1(t) + C_1(t) = E_1^0, \text{ and } E_2(t) + C_2(t) = E_2^0$$

and thus we can eliminate E_1 and E_2 from the equations to get:

$$
\begin{aligned}
\dot{S} &= -k_1 S(E_1^0 - C_1) - k_3 S(E_2^0 - C_2) + k_{-1}C_1 + k_{-3}C_2 \\
\dot{C_1} &= k_1 S(E_1^0 - C_1) - (k_{-1} + k_2)C_1 \\
\dot{C_2} &= k_3 S(E_2^0 - C_2) - (k_{-3} + k_4)C_2 \\
\dot{E_1} &= -k_1 S(E_1^0 - C_1) + (k_{-1} + k_2)C_1 \\
\dot{E_2} &= -k_3 S(E_2^0 - C_2) + (k_{-3} + k_4)C_2 \\
\dot{P_1} &= k_2 C_1 \\
\dot{P_2} &= k_4 C_2
\end{aligned}
$$

Quasi-steady state

We make the following quasi-steady state assumptions:

$$
\dot{C_1} = 0 = \dot{C_2}.
$$

In other words, the two complexes SE_1 and SE_2 are assumed to be in steady state. Setting the corresponding equations in the model above equal to zero, yields:

$$
\begin{aligned}
0 &= k_1 S(E_1^0 - C_1) - (k_{-1} + k_2)C_1 \\
0 &= k_3 S(E_2^0 - C_2) - (k_{-3} + k_4)C_2
\end{aligned}
$$

We can solve these to express C_1 and C_2 in terms of S:

$$
\begin{aligned}
C_1 &= \frac{E_1^0 S}{S + \frac{k_{-1} + k_2}{k_1}} \\
C_2 &= \frac{E_2^0 S}{S + \frac{k_{-3} + k_4}{k_3}}
\end{aligned}
$$

The rates of product formation of P_1 and P_2 are respectively $dP_1/dt = k_2 C_1$ and $dP_2/dt = k_4 C_2$, and they take the usual Michaelis-Menten form:

$$
\frac{dP_1}{dt} = \frac{V_{m1}S}{S + K_{m1}} \tag{1}
$$

$$
\frac{dP_2}{dt} = \frac{V_{m2}S}{S + K_{m2}} \tag{2}
$$

with

$$
V_{m1} = k_2 E_1^0 \text{ and } K_{m1} = \frac{k_{-1} + k_2}{k_1} \tag{3}
$$

$$
V_{m2} = k_4 E_2^0 \text{ and } K_{m2} = \frac{k_{-3} + k_4}{k_3} \tag{4}
$$

Remark: Suppose we would consider only one of the reactions to take place, so that there is no competition for the substrate:

$$
S + E_1 \longleftrightarrow C_1 \longrightarrow P_1 + E_1
$$

In this case, making a similar quasi-steady state assumption, it can be shown that the rate of formation of P_1 is still be given by the Michaelis-Menten form (1) with the same expressions (3) for the maximal formation rate V_{m1} and half-saturation constant K_{m1}. A similar conclusion holds for the rate of formation of P_2.

In other words, whether one assumes the competition model or the single enzyme model

(no competition), the rate of formation of both products in terms of the substrate S, remains the same. This implies that when determining the values of V_m and K_m of both products experimentally, it does not matter whether this is done for the natural organism (which satisfies the competition model), or for the cloned system (which satisfies the single enzyme model).

Supplementary References

1   Curson, A. R. J., Todd, J. D., Sullivan, M. J. & Johnston, A. W. B. Catabolism of dimethylsulphoniopropionate: microorganisms, enzymes and genes. *Nature Reviews Microbiology*, 1-11, doi:10.1038/nrmicro2653 (2011).

2   Schneider, K., Asao, M., Carter, M. S. & Alber, B. E. *Rhodobacter sphaeroides* uses a reductive route via propionyl coenzyme A to assimilate 3-hydroxypropionate. *Journal of Bacteriology* **194**, 225-232, doi:10.1128/JB.05959-11 (2012).

3   Todd, J. D., Curson, A. R. J., Sullivan, M. J., Kirkwood, M. & Johnston, A. W. B. The *Ruegeria pomeroyi acuI* gene has a role in DMSP catabolism and resembles *yhdH* of *E. coli* and other bacteria in conferring resistance to acrylate. *PLoS One* **7**, e35947, doi:10.1371/journal.pone.0035947 (2012).

4   Reisch, C. R. *et al.* Metabolism of dimethylsulphoniopropionate by *Ruegeria pomeroyi* DSS-3. *Molecular Microbiology* **89**, 774-791, doi:10.1111/mmi.12314 (2013).

5   Carini, P., Steindler, L., Beszteri, S. & Giovannoni, S. J. Nutrient requirements for growth of the SAR11 isolate 'Candidatus Pelagibacter ubique' HTCC1062 on a defined medium. *ISME J* **7**, 592-602, doi:doi:10.1038/ismej.2012.122 (2013).

6   Smith, D. P. *et al.* Transcriptional and translational regulatory responses to iron limitation in the globally distributed marine bacterium *Candidatus* pelagibacter ubique. *PLoS One* **5**, e10487, doi:10.1371/journal.pone.0010487 (2010).

7   Kevin, L. V. *et al.* High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. *The ISME Journal* **7**, 1322-1332, doi:10.1038/ismej.2013.32 (2013).

8   Kirkwood, M., Le Brun, N. E., Todd, J. D. & Johnston, A. W. B. The *dddP* gene of *Roseovarius nubinhibens* encodes a novel lyase that cleaves dimethylsulfoniopropionate into acrylate plus dimethyl sulfide. *Microbiology* **156**, 1900-1906, doi:10.1099/mic.0.038927-0 (2010).

9   Sowell, S. M. *et al.* Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME J* **3**, 93-105, doi:10.1038/ismej.2008.83 (2009).

10   Kiene, R. P., Williams, L. P. H. & Walker, J. E. Seawater microorganisms have a high affinity glycine betaine uptake system which also recognizes dimethylsulfoniopropionate *Aquatic Microbial Ecology* **15**, 39-51 (1998).

11   Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**, 630-633, doi:10.1038/nature00917 (2002).

12   Grote, J. *et al.* Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *mBio* **3**, doi:10.1128/mBio.00252-12 (2012).

13   Sharpton, T. J. *et al.* Sifting through genomes with iterative-sequence clustering produces a large, phylogenetically diverse protein-family resource. *BMC Bioinformatics* **13**, 264, doi:10.1186/1471-2105-13-264 (2012).

14   Gibson, J. A. E., Garrick, R. C., Burton, H. R. & McTaggart, A. R. Dimethylsulfide and the algaPhaeocystis pouchetii in antarctic coastal waters. *Marine Biology* **104**, 339-346, doi:10.1007/bf01313276 (1990).

15   van Duyl, F. C., Gieskes, W. W. C., Kop, A. J. & Lewis, W. E. Biological control of short-term variations in the concentration of DMSP and DMS during a Phaeocystis spring bloom. *Journal of Sea Research* **40**, 221-231, doi:http://dx.doi.org/10.1016/S1385-1101(98)00024-0 (1998).

16   Zindler, C., Peeken, I., Marandino, C. A. & Bange, H. W. Environmental control on the variability of DMS and DMSP in the Mauritanian upwelling region. *Biogeosciences* **9**, 1041-1051, doi:10.5194/bg-9-1041-2012 (2012).

17   Varaljay, V. A. *et al.* Single-taxon field measurements of bacterial gene regulation controlling DMSP fate. *ISME J* **9**, 1677-1686, doi:10.1038/ismej.2015.23 (2015).

18   Kiene, R. in *Biological and Environmental Chemistry of DMSP and Related Sulfonium Compounds* (eds RonaldP Kiene, PieterT Visscher, MaureenD Keller, & GunterO Kirst) Ch. 29, 337-349 (Springer US, 1996).

19   Kiene, R. P. & Linn, L. J. Distribution and turnover of dissolved DMSP and its relationship with bacterial production and dimethylsulfide in the Gulf of Mexico. *Limnology and Oceanography* **45**, 849-861, doi:10.4319/lo.2000.45.4.0849 (2000).

20   Turner, S. M., Malin, G., Liss, P. S., Harbour, D. S. & Holligan, P. M. The seasonal variation of dimethyl sulfide and dimethylsulfoniopropionate concentrations in nearshore waters1. *Limnology and Oceanography* **33**, 364-375, doi:10.4319/lo.1988.33.3.0364 (1988).

21   Belviso, S. *et al.* Production of dimethylsulfonium propionate (DMSP) and dimethylsulfide (DMS) by a microbial food web. *Limnology and Oceanography* **35**, 1810-1821, doi:10.4319/lo.1990.35.8.1810 (1990).

22   Ledyard, K. M. & Dacey, J. W. H. Microbial cycling of DMSP and DMS in coastalandoligotrophicseawater. *Limnology and Oceanography* **41**, 33-40, doi:10.4319/lo.1996.41.1.0033 (1996).

23   Iverson, R. L., Nearhoof, F. L. & Andreae, M. O. Production of dimethylsulfonium propionate and dimethylsulfide by phytoplankton in estuarine and coastal waters. *Limnology and Oceanography* **34**, 53-67, doi:10.4319/lo.1989.34.1.0053 (1989).

24   Simó, R., Grimalt, J. O. & Albaigés, J. Dissolved dimethylsulphide, dimethylsulphoniopropionate and dimethylsulphoxide in western Mediterranean waters. *Deep Sea Research Part II: Topical Studies in Oceanography* **44**, 929-950, doi:http://dx.doi.org/10.1016/S0967-0645(96)00099-9 (1997).

25   Simó, R., Pedrós-Alió, C., Malin, G. & Grimalt, J. O. Biological turnover of DMS, DMSP and DMSO in contrasting open-sea waters. *Marine Ecology Progress Series* **203**, 1-11, doi:10.3354/meps203001 (2000).

26   Kiene, R. P. & Slezak, D. Low dissolved DMSP concentrations in seawater revealed by small-volume gravity filtration and dialysis sampling. *Limnology and Oceanography: Methods* **4**, 80-95, doi:10.4319/lom.2006.4.80 (2006).

27   Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248-254, doi:S0003269776699996 [pii] (1976).

28   Todd, J. D. *et al.* Molecular dissection of bacterial acrylate catabolism--unexpected links with dimethylsulfoniopropionate catabolism and dimethyl sulfide production. *Environmental microbiology* **12**, 327-343, doi:10.1111/j.1462-2920.2009.02071.x (2010).

29   Eddy, S. R. What is a hidden Markov model? *Nature Biotechnology* **22**, 1315-1316 (2004).

30   Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755 (1998).

31   Boden, R. *et al.* Purification and characterization of dimethylsulfide monooxygenase from *Hyphomicrobium sulfonivorans*. *Journal of Bacteriology* **193**, 1250-1258, doi:10.1128/JB.00977-10 (2011).

32   McDevitt, C. A., Hugenholtz, P., Hanson, G. R. & McEwan, A. G. Molecular analysis of dimethyl sulphide dehydrogenase from *Rhodovulum sulfidophilum*: its place in the dimethyl sulphoxide reductase family of microbial molybdopterin‐containing enzymes. *Molecular Microbiology* **44**, 1575-1587 (2002).

33   Reisch, C. R., Moran, M. A. & Whitman, W. B. Bacterial catabolism of dimethylsulfoniopropionate (DMSP). *Frontiers in Microbiology* **2**, 172, doi:10.3389/fmicb.2011.00172 (2011).

34   Eddy, S. R. Accelerated profile HMM searches. *PLoS Computational Biology* **7**, e1002195 (2011).

35   Wilhelm, L. J., Tripp, H. J., Givan, S. A., Smith, D. P. & Giovannoni, S. J. Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biology Direct* **2**, 27, doi:10.1186/1745-6150-2-27 (2007).

36   Monroe, M. E., Shaw, J. L., Daly, D. S., Adkins, J. N. & Smith, R. D. MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. *Computational Biology Chemistry* **32**, 215-217, doi:10.1016/j.compbiolchem.2008.02.006 (2008).

37   Keller, A. & Shteynberg, D. Software pipeline and data analysis for MS/MS proteomics: the trans-proteomic pipeline. *Methods in Molecular Biology* **694**, 169-189, doi:10.1007/978-1-60761-977-2_12 (2011).

38   Vizcaino, J. A. *et al.* ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature Biotechnology* **32**, 223-226, doi:10.1038/nbt.2839 (2014).