Mechanistic insights into the key marine dimethylsulfoniopropionate synthesis enzyme DsyB/DSYB

Chun-Yang Li[1,2,3]†, Jason C. Crack[4]†, Simone Newton-Payne[5]†, Andrew R. J. Murphy[6],

Xiu-Lan Chen[2,3], Benjamin J. Pinchbeck[5], Shun Zhou[1,5], Beth T. Williams[5], Ming Peng[2],

Xiao-Hua Zhang[1], Yin Chen[6], Nick E. Le Brun[4]*, Jonathan D. Todd[1,5]*, Yu-Zhong

Zhang[1,2,3]*

[1]Frontiers Science Center for Deep Ocean Multispheres and Earth System & College of Marine

Life Sciences, Ocean University of China, Qingdao, China

[2]State Key Laboratory of Microbial Technology, Marine Biotechnology Research Center,

Shandong University, Qingdao, China

[3]Laboratory for Marine Biology and Biotechnology, Pilot National Laboratory for Marine

Science and Technology, Qingdao, China

[4]Centre for Molecular and Structural Biochemistry, School of Chemistry, University of East

Anglia, Norwich Research Park, Norwich, UK.

[5]School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich,

UK.

[6]School of Life Sciences, University of Warwick, Coventry, UK.

† Chun-Yang Li, Jason C. Crack and Simone Newton-Payne contributed equally to this work.

* Corresponding author: Jonathan D. Todd, Jonathan.Todd@uea.ac.uk; Nick E. Le Brun,

N.Le-brun@uea.ac.uk; Yu-Zhong Zhang, zhangyz@sdu.edu.cn

Abstract

Marine algae and bacteria produce eight billion tonnes of the organosulfur molecule dimethylsulfoniopropionate (DMSP) in Earth's surface oceans annually. DMSP is an anti-stress compound and, once released into the environment, a major nutrient, signalling molecule and source of climate-active gases. The methionine transamination pathway for DMSP synthesis is used by most known DMSP-producing algae and bacteria. The *S*-directed *S*-adenosylmethionine-dependent methyltransferase (SAM-MT) 4-methylthio-2-hydroxybutyrate (MTHB) *S*-methyltransferase, encoded by the *dsyB/DSYB* gene, is the key enzyme of this pathway, generating *S*-adenosylhomocysteine (SAH) and 4-dimethylsulfonio-2-hydroxybutyrate (DMSHB). *dsyB/DSYB*, present in most haptophyte and dinoflagellate algae with the highest known intracellular DMSP concentrations, is shown to be far more abundant and transcribed in marine environments than any other known DMSP synthesis pathway *S*-methyltransferase gene. Furthermore, we demonstrate *in vitro* activity of the bacterial DsyB enzyme from *Nisaea denitrificans* and provide its crystal structure in complex with SAM and SAH-MTHB, which together provide the first important mechanistic insights into a DMSP synthesis enzyme. Structural and mutational analyses imply that DsyB adopts a proximity and desolvation mechanism for the methyl transfer reaction. Sequence analysis suggests that this mechanism is common to all bacterial DsyB enzymes and also, importantly, eukaryotic DSYB enzymes from e.g., algae that are the major DMSP producers in Earth's surface oceans.

Impact Statement

Dimethylsulfoniopropionate (DMSP) is one of Earth's most abundant organosulfur containing molecules which influences marine nutrient cycling, chemotaxis, atmospheric chemistry and potentially the climate. This study provides the first structural and mechanistic understanding of the key DMSP synthesis enzyme in marine bacteria (DsyB) and algae (DSYB) that are responsible for the annual production of > 8 billion tonnes of DMSP. DsyB is the first DMSP synthesis enzyme to be analyzed at the structural and mechanistic level. This study defines functional characteristics of the DsyB/DSYB enzyme family that has a central role in driving marine microbial cycling of organosulfur.

Key words: DMSP synthesis, marine sulfur cycle, *S*-methyltransferase, catalytic mechanism

INTRODUCTION

Approximately eight billion tonnes of the compatible solute dimethylsulfoniopropionate (DMSP) are produced annually in Earth's surface waters (1), constituting up to 10% of surface ocean organic carbon (2). Many marine algae, bacteria, corals and some plants produce DMSP (3) for its proposed functions as e.g., a compatible solute (4), grazing deterrent (5), antioxidant (6) and protectant against hydrostatic pressure (7). Furthermore, DMSP is a major nutrient for marine microorganisms, and a precursor for climate-active volatiles such as dimethyl sulfide (DMS) (3, 8, 9). DMSP was thought to be mainly produced by marine algae in Earth's surface oceans, but recent studies suggest that bacteria, particularly in marine sediment, are also important DMSP producers (3, 10-12).

Three pathways for DMSP synthesis have been proposed based on the identification of intermediates and enzyme activities in various model DMSP producers: a methylation pathway in some plants and bacteria, a transamination pathway in algae and bacteria, and a decarboxylation pathway in one dinoflagellate (10, 12-15) (Fig. 1). Of these (Fig. 1), the transamination pathway is thought to be the most important in marine environments as it functions in the majority of DMSP producing algae (spanning dinoflagellates, haptophytes and diatoms) and bacteria (10, 12, 14). The committed enzyme of the transamination pathway (Fig. 1) is an *S*-adenosylmethionine (SAM)-dependent 4-methylthio-2-hydroxybutyrate (MTHB) *S*-methyltransferase that yields 4-dimethylsulfonio-2-hydroxybutyrate (DMSHB) (13, 14, 16). Recently, the key MTHB *S*-methyltransferase enzyme 'DsyB' (Fig. 1), was identified in many DMSP-producing marine *Alphaproteobacteria* (10). Enzymes, termed DSYB, with ~33% amino acid identity to bacterial DsyB enzymes and that have SAM-dependent MTHB *S*-

81   methyltransferase activity are found in many eukaryotes including most DMSP-producing

82   dinoflagellates, haptophytes, corals and ~ 20% of diatoms (12). This is consistent with the

83   detection of DMSHB and its oxidative decarboxylation to DMSP in some prymnesiophytes,

84   diatoms and prasinophytes (14). *dsyB*/*DSYB* genes are robust indicators of an organism's

85   potential to produce DMSP (10, 12). The centric diatom *Thalassiosira pseudonana*, which lacks

86   DSYB, contains an isoform MTHB *S*-methyltransferase enzyme termed TpMMT, but this

87   enzyme has not been studied in any other organism (17). Published and new analysis here (see

88   below) shows that *DsyB*/*DSYB* genes are far more abundant in known DMSP-producing

89   microorganisms (phytoplankton and bacteria) and in marine environmental metagenome and

90   metatranscriptome datasets than other identified DMSP synthesis genes (10-12). Furthermore,

91   acquisition of *dsyB* is sufficient to enable some organisms to produce DMSP (10). Together

92   these data suggest that transamination pathway using DsyB/DSYB enzymes is the most

93   important marine DMSP synthesis pathway.

94   DSYB and DsyB belong to the SAM-dependent methyltransferase (SAM-MT) family (10,

95   12). SAM is the second most widely used enzyme substrate after ATP and is involved in many

96   important biological processes (18). SAM-MTs are categorized based on the methyl accepting

97   atom, usually *O*, *N*, *C* or *S* (19). The majority (54%) of known SAM-MTs are *O*-directed,

98   whereas only 3% are *S*-directed (19). SAM-MTs, which catalyze transmethylation via S_N2

99   nucleophilic substitution (20, 21), have evolved three distinct mechanisms: proximity and

100 desolvation (PD), general acid/base-mediated catalysis, and a metal-dependent mechanism (19).

101 There are no reported protein crystallographic studies on any DMSP synthesis enzyme and,

102 thus, the mechanism of *S*-directed SAM-MT in DMSP synthesis pathways, e.g., via

103   DsyB/DSYB or TpMMT, are unknown.

104   Here, we investigate *Nisaea denitrificans* DR41_21, a marine *Alphaproteobacterium* (22)

105   predicted to produce DMSP, and characterize its DsyB enzyme. X-ray crystallography and

106   mutational analyses were employed to establish the DsyB structure and predict its interaction

107   with SAM and MTHB substrates and reaction mechanism. Furthermore, sequence alignment

108   and structural analysis are used to infer mechanistic similarities between bacterial DsyB and

109   algal DSYB enzymes. We also probe marine microorganisms, metagenomes and

110   metatranscriptomes for the presence of DsyB/DSYB and other key SAM-MT in DMSP

111   synthesis pathways to investigate the importance of these proteins in the global oceans. Our

112   results provide the first insights into the mechanism of global DMSP production via the most

113   abundant known DMSP synthesis enzymes.

114

RESULTS AND DISCUSSION

115

Nisaea denitrificans DR41_21 is a DMSP-producing bacterium

116

117   Isolated from coastal Mediterranean Sea surface waters, *Nisaea denitrificans* DR41_21 (DSM

118   18348) is a marine *Alphaproteobacterium* of the *Rhodospirillaceae* family which was not

119   previously known to produce DMSP (22). *N. denitrificans* contains a DsyB enzyme, 337 amino

120   acid residues in length with 59% identity to *Labrenzia aggregata* DsyB and is thus predicted to

121   make DMSP (10). Indeed, cloned *N. denitrificans dsyB* conferred onto *Rhizobium*, a

122   heterologous host that lacks DsyB and makes no DMSP, MTHB *S*-methyltransferase activity.

123   Furthermore, *N. denitrificans dsyB* fully restored the production and accumulation of DMSP

124   (105 ± 3.4 pmol DMSP µg protein^{-1}) to an *L. aggregata dsyB*$^{-}$ deletion mutant, which produces

6

125   and accumulates no DMSP without it (wild type *L. aggregata* produces 99.8 ± 1.2 pmol DMSP

126   µg protein^{-1} (10)). *N. denitrificans* itself produced DMSP when grown in the absence of

127   methylated sulfur compounds, and production and *dsyB* transcription was enhanced by

128   increased salinity and by nitrogen starvation (Fig. 2A-B). This work further confirms that the

129   presence of *dsyB* and its transcription in a bacterium represent the ability of the strain to produce

130   DMSP and the levels it makes, respectively.

131   ***In vitro* characterization of *N. denitrificans* DsyB**

132   As shown above and in Curson et al., (10), DsyB has MTHB *S*-methyltransferase activity when

133   expressed in alphaproteobacterial hosts. However, *L. aggregata* DsyB and *Chrysochromulina*

134   DSYB enzymes (12) had no detectable MTHB *S*-methyltransferase activity when expressed in

135   *E. coli*. The same was generally true of the recombinant *N. denitrificans* DsyB enzyme purified

136   from *E. coli*, although variable MTHB *S*-methyltransferase activity was observed *in vitro* using

137   MTHB and SAM as substrates (see Materials and Methods). The reason for this lack of activity

138   upon isolation is unknown; one possibility is that the enzyme requires an essential co-factor or

139   modification that was provided by an algal or alphaproteobacterial host, but not by *E. coli* (12).

140   This hypothesis was initially supported by the fact that addition of heat denatured cell lysate

141   fractions (from a PD10 desalting column) liberated from the *L. aggregata dsyB* deletion mutant,

142   which produces no DMSP, recovered *N. denitrificans* DsyB MTHB *S*-methyltransferase activity.

143   Similar complementation was shown with the addition of heat-killed *Prymnesium parvum*

144   extract to DSYB in (12). The activated DsyB protein was shown to have K_m and V_{max} values of

145   0.14 mM and 365 nmol min^{-1} mg protein^{-1}, respectively, for MTHB (Fig. 2C), which were

146   similar to those previously established for *P. parvum* DSYB (0.09 mM and 294 nmol min^{-1} mg

147   protein^{-1}) in (12). The activated DsyB had a K_m of 0.16 mM and V_{max} 368.9 nmol min^{-1} mg

148   protein^{-1} for the co-substrate SAM (Fig. 2D), which were also similar to those obtained with *P.*

149   *parvum* DSYB (0.06 mM and 303 nmol min^{-1} mg protein^{-1}) in (12).

150   Liquid chromatography with mass spectrometry (LC-MS) and/or native mass

151   spectrometry was used in an attempt to identify the activation factor in the *L. aggregata dsyB*⁻

152   extract (Fig. S1). A prominent peak at 37,084 Da was observed in the LC-MS spectrum for both

153   the as-isolated and activated samples, which corresponds to DsyB with its N-terminal Met

154   residue cleaved (commonly observed for proteins over-expressed in *E. coli*) (23). A lower

155   intensity peak at +131 Da, corresponding to the full-length protein (37,215 Da), was also

156   observed in the as-isolated sample, indicating that the Met cleavage was not complete (Fig.

157   S1A). There was an additional minor peak at +269 Da of unknown origin in the activated

158   sample (Fig. S1B). Under non-denaturing conditions, both monomeric and dimeric forms of

159   DsyB were detected in the as-isolated sample, a feature commonly observed in non-denaturing

160   mass spectra of solution dimers (24-26). In the monomeric region, the main protein peaks (due

161   to cleaved and non-cleaved proteins) were again observed, but, in addition, a number of adduct

162   species were present in the spectrum. Two of these, at +36 and +98 Da, correspond to chloride

163   and (most likely) phosphate adducts. An additional adduct at +63 Da was also observed,

164   possibly due to metal ion binding. In general, the spectrum of the dimeric form of DsyB was

165   less well-resolved, but the main protein peak (at 74,168 Da) along with chloride and possible

166   metal ion adducts were all detected (Fig. S2). The non-denaturing mass spectrum of the

167   activated monomeric DsyB (Fig. S1B) revealed a number of adducts, including those most

168   likely due to chloride, phosphate and metal ion binding (all common within the as-isolated

169 DsyB), along with an additional adduct at +122 Da (and at +244 Da), which is likely due to

170 Tris buffer. Thus, we have no data to support there being a cofactor or modification of DsyB

171 caused by the addition of the heat-killed *L. aggregata dsyB*⁻ extract to as-isolated DsyB, and

172 further work is required to understand the variable nature of DsyB activity (see below).

173   The association of metal ions with DsyB was investigated further. Inductively coupled

174 plasma mass spectrometry (ICP-MS) analysis revealed variable metal ion content with some

175 preparations of as-isolated DsyB containing up to 0.85 Cu per protein, with other metals such

176 as Ni (up to 0.5 per protein), Zn (0.4) and Fe (0.14) also detected. However, there was no

177 correlation between metal ion content and activity of as-isolated samples.

178   Despite mostly lacking consistent *in vitro* MTHB *S*-methyltransferase activity (see below),

179 native MS showed that the as-isolated *N. denitrificans* DsyB enzyme binds to SAM. The

180 deconvoluted mass spectrum of a DsyB sample under non-denaturing conditions and containing

181 25 equivalents of SAM contained a peak in the DsyB dimer region at +870 Da (predicted mass

182 of a $(DsyB)_2$-$(SAM)_2$ is 74,966 Da), indicative of a $(DsyB)_2$-$(SAM-Cl)_2$ adduct, that was not

183 observed in the absence of SAM (Fig. S3A). Evidence for SAM binding was also apparent in

184 the monomer region, though the presence of chloride adducts spreads out the intensity in this

185 region (Fig. S3B). Evidence for an MTHB-bound form of DsyB was also observed (although,

186 again, the presence of chloride adducts spreads out intensity, Fig. S3B). Although care is needed

187 in interpreting intensities of peaks in the non-denaturing mass spectrum, the low intensity of

188 the SAM- and MTHB-bound forms of DsyB suggests relatively low affinities when these

189 substrates are present individually.

190   Small molecule HILIC-MS analysis of reactions following addition of MTHB to DsyB-

191   SAM resulted in the detection of substrates SAM and MTHB, and products DMSHB and SAH

192   (Fig. 3). Non-denaturing mass spectrometry of similarly generated samples resulted in the loss

193   of SAM- and MTHB-bound forms of DsyB (Fig. S3B). Together, the data are consistent with

194   DsyB being a SAM-dependent MTHB *S*-methyltransferase.

195

196   **Overall structure of DsyB**

197   To analyze the catalytic mechanism of DsyB, we solved the crystal structures of complexes of

198   DsyB with SAM and with SAH-MTHB. The crystal structure of the DsyB-SAM complex was

199   determined by the single-wavelength anomalous dispersion (SAD) method using a

200   selenomethionine derivative (Se derivative) (Table S1).

201   Crystals of the DsyB-SAM complex belonged to the $P2_12_12_1$ space group, with four

202   molecules arranged as a tetramer in the asymmetric unit. Each DsyB molecule contains two

203   domains, an N-terminal domain (N-domain, Met1-Ala125) and a C-terminal domain (C-domain,

204   Thr126-Glu337), which can be seen binding to the SAM molecule (Fig. 4A). The DsyB C-

205   domain contains seven β-strands surrounded by six α-helices, which together adopt the typical

206   Rossmann-like α/β fold of Class I SAM-MTs (Fig. 4A). Structural analysis showed that two

207   DsyB monomers are tightly intertwined, mainly through interactions of residues from the N-

208   domains of two adjacent monomers (Fig. 4A). Analysis of DsyB using the PISA server

209   (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) predicted the DsyB dimer to be stable in

210   solution. Indeed, gel filtration analysis indicated that DsyB is likely a dimer in solution (Fig.

211   4B), consistent with the non-denaturing mass spectrometry data above (Figs. S2, S3A). These

212   results indicate that DsyB functions as a dimer in the same way as other SAM-MTs, whose N-

213 domains are also responsible for dimerization (19, 27, 28). ICP-MS and LC-MS analyses

214 showed that the as-isolated DsyB contained variable metals. However, in the crystal structure

215 of DsyB-SAM complex, no explicit electron density associated with metals was observed,

216 suggesting that the binding site of metals may be not specific in DsyB.

217   The crystals of the DsyB-SAH-MTHB complex belong to the $P2_1$ space group (Table S1)

218 and the resulting structure has a similar overall structure to that of the DsyB-SAM complex

219 (Fig. 4C). Interestingly, in this case the MTHB co-substrate molecule is located between the C-

220 domain and the N-domain of one DsyB monomer (Fig. 4C).

221

The conformational change of DsyB in binding SAM

223 During structural refinement of the DsyB-SAM complex, we found that three monomers (chain

224 A, B and C) of the asymmetric unit contained SAM molecules. The structures of these three

225 monomers are similar, with root mean square deviations (RMSDs) of ~0.5 Å between any two

226 monomers. The chain D of the DsyB-SAM complex is not bound to a SAM molecule. The

227 conformation of chain D is different to the other monomers bound to SAM, with a RMSD of

228 ~2.4 Å between chain D and chain A. Moreover, residues Asp123 to Tyr143 in chain D

229 exhibited weak electron density, suggesting that this region is highly flexible. By superposing

230 molecules of chain A and chain D, we observed that the N-domains of chain A and chain D are

231 almost completely aligned, whereas the C-domain rotates ~10° as a rigid body (Fig. 4D). These

232 structural differences indicate that DsyB possesses two conformations: an "open" form and a

233 "closed" form. Although DsyB can bind SAM and MTHB individually (Fig. S3B), the binding

234 of SAM triggers the conformational change of DsyB from the "open" form to the "closed" form,

235   shrinking the cavity between the N-domain and the C-domain of DsyB and possibly promoting

236   the subsequent binding of MTHB.

237   The structure of the DsyB-SAM complex is similar to that of the *Streptosporangium*

238   *sibiricum* SibL protein (PDB code: 4U1Q), a *C*-directed Class I SAM-MT, with an RMSD of

239   ~1.3 Å between these two structures. SAM binding also triggers the conformational change of

240   SibL from an "open" form to a "closed" form to complete the formation of a binding site for its

241   methyl acceptor 3-hydroxykynurenine (27). Similar conformational changes have also been

242   observed in other *C/O*-directed SAM-MTs, despite their low sequence identities (19, 28).

243

244   **Binding sites of SAM and MTHB**

245   The SAM molecule within the DsyB-SAM complex (Fig. 4A) is bound mainly by hydrogen

246   bonds with residues in the DsyB C-terminal domain. DsyB residues Asp223 and Ala224

247   participate in binding the adenine ring of SAM; Asp196 forms hydrogen bonds with the ribose

248   moiety of SAM; and Ser150, Gly173 and Ser239 interacts with the terminal amino acid moiety

249   of SAM (Fig. 5A). A similar binding mode was observed between the same DsyB C-terminal

250   residues for SAH in the DsyB-SAH-MTHB complex.

251   In the DsyB-SAH-MTHB complex, the electron density of the MTHB molecule is

252   relatively poor and the distance between SAH and MTHB is more than 6 Å, which is too far to

253   enable the methyl transfer reaction. We speculate that the position of MTHB observed in the

254   structure is not the exact location of MTHB when the reaction occurs under physiological

255   conditions, and that the observed structure represents a state where the MTHB molecule has

256   not completely entered into the active site. Nevertheless, the location of MTHB clearly implies

257   its initial binding site (Fig. 5B), and several residues likely involved in the binding of MTHB

258   were identified, including Tyr97 and Gln101 from the N-domain and Tyr129, Tyr142, Gln146

259   and His291 from the C-domain of DsyB (Fig. 5C).

260   To determine the importance of the Tyr97, Gln101, Tyr129, Tyr142, Gln146 and His291

261   in DsyB binding to MTHB, we performed site-directed substitutions of these residues and

262   assayed the *in vivo* MTHB *S*-methyltransferase activity of the resultant variant DsyB

263   derivatives in *R. leguminosarum*. *In vivo* assays were performed in *Rhizobium* because of the

264   sensitivity of *in vitro* assays, see above. Site-directed mutations of Tyr97, Gln101, Tyr129,

265   Tyr142, Gln146 or His291 severely decreased the enzymatic activity of DsyB (Fig. 5D),

266   indicating the potentially important roles of these residues in binding MTHB. In particular,

267   activity of the Tyr97Ala and Tyr142Ala variants was completely abolished (Fig. 5D). However,

268   Tyr97Phe and Tyr142Phe variants maintained > 30% residual activity (Fig. 5D), suggesting that

269   the elimination of activity of Tyr97Ala and Tyr142Ala substitutions may be caused by the

270   replacement of the aromatic side chain. This mutational analysis suggests that residues likely

271   involved in the binding of MTHB are not essential for catalysis.

272   Three distinct catalytic mechanisms have been reported for SAM-MTs, including the PD

273   mechanism, the general acid/base-mediated mechanism and the metal-dependent mechanism

274   (19). Structural and biochemical analyses indicate that the activity of DsyB is neither metal-

275   dependent nor catalytic residue-dependent, but is likely driven by the proximity effect. The

276   DsyB enzyme likely enables favourable orientations of MTHB and SAM molecules that bring

277   the sulfur atom of MTHB in close proximity to the methyl group of SAM.

278

13

The catalytic mechanism of DsyB

Based on our structural and biochemical results, we propose DsyB first binds a SAM molecule to generate a conformational change from "open" to "closed" state, which may promote the binding of MTHB (Fig. 6A). When an MTHB molecule enters the active site, DsyB brings the sulfur atom of MTHB close enough to the methyl group of SAM to allow nucleophilic attack on the methyl group of SAM (Fig. 6B). Subsequently, the generated DMSHB and SAH are released, and DsyB can re-bind a SAM molecule from the intracellular environment in preparation for the next reaction.

Ecm18, another bacterial *S*-directed SAM-MT, converts disulfide in triostin A to the thioacetal linkage in the peptide antibiotic echinomycin through two stages, the methylation of one sulfur atom of the disulfide and the rearrangement of the methylated disulfide to form the thioacetal (29). Thus, as we predict for DsyB, Ecm18 also uses the PD mechanism for its methyl transfer reaction (29). Thiopurine *S*-methyltransferase (TPMT) is a murine *S*-directed SAM-MT that methylates 6-mercaptopurine (30). Unlike DsyB, TPMT does not contain an N-terminal domain likely involved in dimerization, as its N-terminus only constitutes 40 residues (30). Although Arg147 and Arg221 are possible participants in 6-mercaptopurine deprotonation, the modest decrease in the enzymatic activities of the corresponding mutants suggests that TPMT may possess the PD strategy for catalysis (30). *Catharanthus roseus* CrSMT1 is another *S*-directed SAM-MT that methylates a broad range of substrates including benzene thiol and furfuryl thiol (31). Homology modelling suggests that CrSMT1 contains an N-domain for dimerization (31), which is similar to DsyB. However, CrSMT1 is thought to use a histidine residue as a general base to deprotonate the thiol group of the substrate (31). Thus, although the

301   *S*-directed SAM-MTs only constitute a small portion of the reported SAM-MTs (19), their

302   domain structures and catalytic mechanisms appear diverse.

303

Universality of the catalytic mechanism of DsyB

304

305   The majority of bacteria containing DsyB are *Rhodobacterales*, which are abundant in marine

306   environments, but this enzyme is also found in some *Rhizobiales* and *Rhodospirillales*

307   (including *N. dentrificans*) (10, 32). To investigate the ubiquity of the DsyB catalytic

308   mechanism, we performed sequence alignment of DsyB proteins from different

309   *Rhodobacterales*, *Rhizobiales* and *Rhodospirillales* bacteria (Fig. S4). Most residues involved

310   in initial MTHB binding (Tyr97, Gln101, Tyr129, Tyr142, Gln146 and His291) and SAM

311   binding (Ser150, Gly173, Asp196, Asp223, Ala224 and Ser239) are highly conserved in DsyB

312   proteins from these marine bacteria, indicating that mechanistic insight gained here for *N.*

313   *denitrificans* DsyB has universal significance in bacteria containing DsyB.

314   Eukaryotic DSYB, which may originate from bacterial DsyB, is a key enzyme for DMSP

315   synthesis in many phytoplankton, such as marine haptophytes, dinoflagellates and some

316   diatoms (12). DSYB shares ~33% sequence identity to DsyB, and we predicted the structure of

317   DSYB from *Chrysochromulina tobin* CCMP291 by homologous modelling using SWISS-

318   MODEL (https://swissmodel.expasy.org/) (33). Structural alignment of DSYB and DsyB

319   indicated that residues involved in binding MTHB are perfectly superposed (Fig. S5). Moreover,

320   sequence alignment of DsyB and DSYB from different marine algae showed that residues

321   which play important roles in DsyB are highly conserved in different DSYB proteins (Fig. S6),

322   suggesting that DSYB proteins adopt a similar catalytic mechanism to DsyB.

15

323

DsyB/DSYB are the most abundant and transcribed *S*-methyltransferase enzymes of known DMSP synthesis pathways in marine microorganisms and environments

Having the identity of the key *S*-methyltransferases in diverse DMSP synthesis pathways (DsyB/DSYB and TpMMT in the bacterial and algal transamination pathway, and MmtN and BurB in bacterial methylation pathways, Fig. 1), we carefully analyzed their presence in marine microorganisms and their abundance and transcript levels in published global metagenome and metatranscriptome datasets to quantify the potential environmental importance of these pathways.

Of the known DMSP synthesis enzymes, DsyB is by far the most abundant in sequenced and/or isolated bacteria (65.8 % of cultured DMSP-producing isolates) (Table S2)(7, 10, 11, 34, 35). DsyB is mostly found in alphaproteobacterial *Rhodobacterales*, *Rhizobiales* and *Rhodospirillales*, but is also sporadically found in e.g., an actinobacterial *Ponticoccus* isolate (7), and in some *Betaproteobacteria* and *Bacteroidetes* metagenome assembled genomes (36). MmtN has much fewer (14.4% of cultured DMSP-producing isolates), but equally diverse, host bacteria, being found in *Alphaproteobacteria*, *Gammaproteobacteria* and *Actinobacteria* (Table S2). Finally, BurB is confined to very closely related *Burkholderia* spp. that likely uses DMSP as an intermediate in toxin production (34).

This hierarchy of DMSP synthesis gene abundance in bacteria (DsyB > MmtN > BurB) was mirrored in marine environmental data. In the *Tara* Oceans prokaryotic database, both *dsyB* and *mmtN* were found throughout the water column (Fig. 7A), but no close homologs of *BurB* (e-value <1e-40) were detected. This is consistent with BurB-mediated DMSP production in

16

345 *Burkholderia* spp., possibly for toxin production, not being an important process in marine

346 systems. *DsyB* was significantly more abundant than *mmtN* in both the metagenomic (median

347 abundance 0.141% vs 0.00376%) (Kruskal-Wallis X^2 = 83.781, p < 0.001) and

348 metatranscriptomic (median abundance 0.2% vs 0.0364%) (Kruskal-Wallis X^2 = 33.64, p <

349 0.001) *Tara* Oceans datasets (Fig. 7A). Additionally, *dsyB* was found at 172 and 153 sampling

350 sites (treating each depth as a separate site) in the metagenomes and metatranscriptomes,

351 respectively, whereas *mmtN* was found at only 74 and 63 sites, respectively. Given this, our

352 analysis of median abundance overestimates the contribution of *mmtN* to DMSP production in

353 the global ocean. We therefore determined the relative abundance of *dsyB:mmtN* across depths

354 at each sampling site in both metagenomes and metatranscriptomes (Fig. S7). *DsyB* was more

355 abundant at almost all sites in the metagenomes, and was more highly expressed across most,

356 though there were a number of locations in the South Atlantic and South Pacific where *mmtN*

357 was predominant (Fig. S7). Taxonomic examination of both *dsyB* and *mmtN* sequences in the

358 *Tara* database (Fig. 7B) showed that both genes were exclusively from *Alphaproteobacteria*,

359 primarily within the Orders *Rhodobacterales* and *Rhizobiales* for *dsyB*, and the genus

360 *Thalassospira* for *mmtN* (Fig. 7B). These data highlight DsyB as the most abundant, transcribed

361 and, likely, important of the known bacterial DMSP synthesis enzymes in marine waters, which

362 likely plays a significant role in the global production of DMSP.

363   Moving to eukaryotic DMSP synthesis, we carefully analyzed available transcriptome data

364 from marine eukaryotes in the marine microbial eukaryote transcriptome sequencing project

365 (MMETSP) (37). The TpMMT MTHB *S*-methyltransferase has only been characterized in the

366 centric diatom *Thalassiosira pseudonana* CCMP1335, and close homologs (~70% protein

17

367 identity) with the same singular domain structure only exist in 17/82 diatom transcriptomes (7

368 of which also contain DSYB), and no other phytoplankton (Tables S3 and S4) (38-53). The next

369 most homologous TpMMT-like proteins, present in e.g., *Thalassiosira oceanica* (EJK59074)

370 and *Fistulifera solaris* (GAX25165) that are more diverse (the methyltransferase domain being

371 <50% identical to TpMMT), contain extra protein domains and, thus, are much larger proteins

372 whose function is unknown. These TpMMT-like proteins cannot be considered as functional

373 MTHB *S*-methyltransferase enzymes and are omitted from this study. In contrast, DSYB is

374 found in the transcriptomes of 47/61 dinoflagellates and 24/30 haptophytes, organisms known

375 to produce the highest levels of DMSP per cell (>50 mM) (39, 46). Furthermore, 15/82 diatom

376 transcriptomes, typically known to produce lower cellular DMSP levels (generally < 50 mM)

377 (39), and some *Ochrophyta*, *Cnidaria* and *Cilophora* transcriptomes also contained DSYB.

378 These data show DSYB to be the most abundant and widespread DMSP synthesis enzyme

379 known in eukaryotic DMSP-producing organisms.

380   Within the eukaryotic Marine Atlas of Tara Ocean Unigenes (MATOU), we found both

381 *DSYB* and *TpMMT* within epipelagic (surface, subsurface (SRF) and deep chlorophyll

382 maximum (DCM)) waters. Initial examination showed *DSYB* to be more abundant in $\leqslant$ 3 μm

383 fractions than in larger fractions (Fig. S8). Data from these smaller $\leqslant$ 3 μm fractions that

384 likely contain picoeukaryotes (CoP) were considered together. Likewise, fractions with a

385 minimum filter size of $\geqslant$ 3 μm that likely exclude picoeukaryotes (ExP) were also considered

386 together. Abundance was not significantly different between SRF and DCM sampling depths

387 for either CoP or ExP *DSYB* (Kruskal-Wallis $X^2 = 0.113$, $p = 0.74$, and Kruskal-Wallis $X^2 =$

388 0.004, $p = 0.95$, respectively), or for Cop or ExP *TpMMT* (Kruskal-Wallis $X^2 = 0.102$, $p = 0.75$,

389 and Kruskal-Wallis $X^2 = 0.194$, $p = 0.66$, respectively), and as such these sampling depths were

390 combined for the purposes of comparative analyses between *DSYB* and *TpMMT* abundance.

391   Metagenomes derived from the MATOU dataset revealed that *DSYB* was significantly

392 more abundant than *TpMMT* in both CoP and ExP fractions (Fig. 8A; *DSYB* vs *TpMMT* CoP

393 median abundance 4.99e-06 vs 5.01e-08 RPKM, post-hoc Dunn's test $z = 16.22$, $p < 0.001$, ExP

394 median abundance 2.2e-07 vs 1.99e-08 RPKM, post-hoc Dunn's test $z = 6.97$, $p < 0.001$).

395 Similarly, *DYSB* was significantly more abundant than *TpMMT* in MATOU derived

396 metatranscriptomes (Fig. 8A; *DSYB* vs *TpMMT* CoP median abundance 9.89e-06 vs 1.81e-07

397 RPKM, post-hoc Dunn's test $z = 15.16$, $p < 0.001$, ExP median abundance 7.21e-07 vs 1.36e-

398 07 RPKM, post-hoc Dunn's test $z = 7.33$). *DSYB* was also significantly more abundant within

399 the CoP fraction than the ExP fraction in the metagenome (Fig. 8A; median abundance 4.99e-

400 06 vs 2.2e-07, post-hoc Dunn's test $z = 11.30$, $p < 0.001$), and in the metatranscriptome (Fig.

401 8A; median abundance 9.89e-06 vs 7.21e-07, post-hoc Dunn's test $z = 12.20$, $p < 0.001$). In

402 contrast, *TpMMT* abundance was not significantly different between CoP and Exp fractions in

403 either the metagenome (Fig. 8A; median abundance 5.01e-08 vs 1.99e-08, post-hoc Dunn's test

404 $z = 1.61$, $p = 0.11$) or the metatranscriptome (Fig. 8A; median abundance 1.81e-07 vs 1.36e-07,

405 post-hoc Dunn's test $z = 0.92$, $p = 0.35$). Again, these analyses likely overestimated the

406 abundance of *DSYB* in the ExP fraction, because, in the metagenome, *DSYB* was detected at

407 138/140 CoP fraction sites, but was only found at 178/272 ExP fraction sites (X^2 (1, $N = 412$) =

408 56.77, $p < 0.001$). *TpMMT* abundance was also likely overestimated as *TpMMT* was detected

409 at 90/140 CoP fraction sites, and at 39/272 ExP fraction sites (X^2 (1, $N = 412$) = 107.21, $p <$

410 0.001). As such, *DSYB* was detected significantly more frequently in CoP (X^2 (1, $N = 280$) =

19

411 54.41, $p < 0.001$) and ExP (X^2 (1, $N = 544$) = 148.12, $p < 0.001$) fraction sites than *TpMMT*.

412 Similarly, within the metatranscriptome, *DSYB* was detected at 139/140 CoP fraction sites and

413 at 251/272 ExP fraction sites (X^2 (1, $N = 412$) = 8,98, $p < 0.01$). *TpMMT* was detected at 94/140

414 CoP fraction sites, and at 99/272 ExP fraction sites (X^2 (1, $N = 412$) = 35.09, $p < 0.001$). Again,

415 *DSYB* was detected at significantly more CoP (X^2 (1, $N = 280$) = 51.78, $p < 0.001$) and ExP (X^2

416 (1, $N = 544$) = 185.10, $p < 0.001$) fraction sites than *TpMMT* in the metatranscriptome data.

417 Given the greater abundance of *DSYB* over *TpMMT* in the environmental data, and that the

418 majority of environmental *DSYB* sequences are likely from dinoflagellates and/or haptophytes,

419 known to be high producers (39, 46) of DMSP compared to *TpMMT* in the generally low-

420 producing diatoms (39), DSYB is currently the most important known DMSP synthesis enzyme

421 (Fig. 8B).

422   To conclude, DMSP is an abundant and ecologically important organosulfur compound.

423 DsyB/DSYB enzymes catalyze the committed *S*-methylation of MTHB to generate DMSHB,

424 which is the key step of the transamination pathway for DMSP synthesis in most bacteria and

425 algae (10-12). Furthermore, DsyB/DSYB enzymes are present in the most prodigious DMSP-

426 producing haptophyte and dinoflagellate phytoplankton, and represent the most abundant and

427 transcribed *S*-methylase genes of known DMSP synthesis pathways in marine waters. In this

428 study, we solved the first crystal structures of bacterial DsyB-SAM and DsyB-SAH-MTHB

429 complexes and demonstrated the conversion of SAM and MTHB into SAH and DMSHB. Based

430 on structural and mutational analyses, the catalytic mechanism of DsyB is proposed, and has

431 universal significance in bacteria containing DsyB, and in marine algae containing DSYB. Our

432 results provide novel insights into DMSP synthesis, shedding light on the global sulfur cycling.

433

MATERIALS AND METHODS

Bacterial strains and growth conditions

E. coli BL21 (DE3) was grown in Lysogeny Broth (LB) medium at 37°C. *R. leguminosarum*

J391 was grown in TY (54) complete medium or Y (54) minimal medium (10 mM NH_4Cl as

nitrogen source) at 28 °C. *L. aggregata* J571 (*dsyB*⁻) was grown in YTSS (55) complete medium

or MBM (56) minimal medium (0.5 mM NH_4Cl as nitrogen source) at 30 °C. *N. denitrificans*

DR41_21 (DSM 18348), purchased from DSMZ, Germany, was cultured in the Difco 2216

medium at 30°C (http://www.dsmz.de/) or MBM medium under different salinity and nitrogen

levels for differential DMSP production experiments. Standard conditions were 10 mM NH_4Cl

and 35 practical salinity units (PSU) compared to 5 PSU (low salinity) or 50 PSU (high salinity).

Cultures were sampled for DMSP and RT-qPCR work in exponential-phase growth (after ~7

days). For nitrogen starvation experiments, exponential phase cells grown under standard

conditions were harvested and incubated overnight in standard MBM media with no added

NH_4Cl. Where necessary, 10 mM succinate was used as carbon source and antibiotics were

added at the following concentrations: gentamicin (20 µg ml⁻¹), streptomycin (400 µg ml⁻¹),

kanamycin (20 µg ml⁻¹), ampicillin (100 µg ml⁻¹), spectinomycin (200 µg ml⁻¹).

450

General *in vivo* and *in vitro* genetic manipulations

Plasmids (Table S5) were transferred to *E. coli* by transformation, and *R. leguminosarum* J391

or *L. aggregata* J571 by conjugation using the helper plasmid pRK2013 (57). Routine

restriction digestions and ligations for cloning were performed essentially as in Carrion et al.

455 (58). The oligonucleotide primers used in this study were synthesized by Eurofins Genomics

456 (Table S6). Sequencing of plasmids and PCR products was performed by Eurofins Genomics.

457   The *dsyB* gene was amplified from *N. denitrificans* genomic DNA and cloned into the

458 pET22b (Novagen, America) for the expression of DsyB with a C-terminal His-tag or into

459 pLMB509 for expression in *Rhizobium* and *Labrenzia* (59). Amino acid substitution mutations

460 in DsyB were generated using QuickChange® mutagenesis kit (Agilent Technologies) and the

461 primers in Table S6. All site directed mutant (SDM) variant plasmids were verified by DNA

462 sequencing.

463

464 **Reverse transcription quantitative polymerase chain reaction (RT-qPCR)**

465 RNA was isolated from 100 ml *N. denitrificans* cultures using RNeasy Mini Kit (Qiagen)

466 according to the manufacturer's protocol with some modifications. On-Column DNase

467 digestion was performed with RNase-Free DNase Set (Qiagen). Reverse transcription of 1 µg

468 of DNA-free RNA per sample was done using the QuantiTect Reverse Transcription Kit

469 (Qiagen). PCR on RNA and cDNA samples confirmed that RNA samples were DNA-free.

470   Primers for RT-qPCR for *N. dentrificans dsyB* and housekeeping genes *recA* and *gyrB*,

471 were designed using Primer5 (60) (Table S6). The qPCR experiments were performed on a

472 StepOnePlus instrument (Applied Biosystems). Quantification was performed using the

473 SYBR® Green JumpStart™ Taq ReadyMix™ (Sigma-Aldrich) following the manufacturer's

474 instructions. Reactions (20 µl) contained 2 µl cDNA and 0.8 µl primers (10 µM), with an

475 annealing temperature of 55 °C. For each condition and gene, the cycle threshold (Ct) values

476 of triplicate technical and biological replicates were averaged. Relative expression levels were

477 determined with efficiency correction (61). *dsyB* transcription was displayed as normalized fold

478 change to the standard condition.

479

480 **DsyB enzyme assays *in vivo***

481 To measure MTHB *S*-methyltransferase activity from pLMB509 clones (and SDM derivatives)

482 in *R. leguminosarum* J391, cultures were grown overnight in TY complete medium. Then 1 ml

483 of culture was centrifuged at 20,000 g for 2 min, resuspended in the same volume of Y medium

484 and then diluted 1:100 into 5 ml Y with 0.5 mM DL-MTHB (Sigma-Aldrich, 55875), 10 mM

485 taurine (to induce expression, Sigma-Aldrich, T0625) and gentamicin, and incubated for 60 h

486 at 28 °C before sampled for gas chromatography (GC) analysis to determine the amount of

487 DMSP product (see below).

488   To measure MTHB *S*-methyltransferase activity from pLMB509 clones expressing the

489 *dsyB* gene in the *L. aggregata dsyB⁻* mutant strain J571, cultures were grown overnight in YTSS

490 complete medium. Following incubation, 1 ml of culture was then centrifuged at 20,000 g for

491 2 min, resuspended in the same volume of MBM medium and then diluted 1:50 into 5 ml MBM

492 with 10 mM taurine (to induce expression; Sigma-Aldrich), gentamicin and rifampicin, and

493 incubated for 24 h at 30 °C. DL-MTHB (0.5 mM) was added as substrate to the cultures and

494 these were incubated for 4 h at 30 °C before sampled for GC and protein estimation by the

495 Bradford assay.

496   To measure DMSHB/DMSP in *Rhizobium* or *Labrenzia* assay mixtures, 200 µl of culture

497 was added to a 2 ml glass serum vial, then 100 µl 10 M NaOH was added and vials were

498 crimped with PTFE/rubber crimp caps immediately. Vials were incubated at 80°C for 10 min

499 (to capture DMSHB as well as DMSP) and then for 24 h at room temperature in the dark before

500 being monitored by GC assay. All GC assays involved measurement of headspace DMS using

501 a flame photometric detector (Agilent 7890 A GC fitted with a 7693 autosampler) and an HP-

502 INNOWax 30 m × 0.320 mm capillary column (Agilent Technologies J&W Scientific).

503 Calibration curves were produced by alkaline lysis of DMSP standards in water. The detection

504 limit for headspace DMS from DMSP was 0.015 nmol and from DMSHB was 0.3 nmol. DsyB

505 activity is expressed as pmol DMSHB/DMSP mg protein^{-1} min^{-1}. Protein concentrations were

506 determined using the Bradford method (BioRad). Control assays of *Rhizobium* or *Labrenzia*

507 J571 containing pLMB509 were carried out, as above, and gave no detectable DsyB activity.

508

509 **Protein expression and purification**

510 *E. coli* BL21 (DE3) containing pET22b::*dsyB* clones were cultured in LB medium containing

511 ampicillin at 37°C. At mid-exponential growth (OD$_{600}$ 0.5–0.7), 0.5 mM isopropyl β-D-1-

512 thiogalactopyranoside (IPTG) was added and the cells were incubated at 20°C for 16 h. Cells

513 were harvested by centrifugation (20 min, 7500 × g, 4°C), washed and resuspended in 25 mM

514 Tris-HCl, pH 8.0, 150 mM NaCl. Cell lysis was performed by three passages through a French

515 Press (16,000 psi), unbroken cells and cell debris were removed by centrifugation (30 min,

516 5,500 × g, 4°C) and the supernatant was recovered and subjected to centrifugation (60 min,

517 185,000 × g, 4°C) to pellet the membrane fraction. Soluble cell lysate was applied to a slurry

518 of Ni-NTA resin (Qiagen) at a 3:1 ratio for 90 min with shaking at 4°C. The lysate/slurry mix

519 was loaded into Econo-Pac polypropylene columns, washed with 50 mM Tris-HCl, 250 mM

520 NaCl, 20 mM imidazole, pH 8.0, and DsyB was eluted in 5 x 1 ml fractions using 50 mM Tris-

HCl, 250 mM NaCl, 250 mM imidazole, pH 8.0. Fractions containing DsyB were concentrated

and buffer exchanged into 2 ml of 50 mM Tris-HCl, 100 mM NaCl, pH 8.0 and applied to a

Superdex 200 10/300GL preparative grade gel filtration column (Cytiva). The purified protein

was flash-frozen in liquid nitrogen and stored at −80°C until required.

DsyB enzyme assays *in vitro*

Labrenzia aggregata dsyB$^-$ J571 (10) was grown to late exponential phase in MBM. Cell lysates

were prepared by centrifuging 100 ml of culture for 10 min at 2,500 *g*. The pellet was washed

and resuspended with 20 ml 20 mM HEPES, 150 mM NaCl, pH 7.5 before cell lysis via French

press (16,000 psi). The cell lysate was heat-treated at 80°C for 10 min to denature proteins, then

applied to a 10 ml PD10 column, eluted over 10 ml, and collected in 1ml fractions. DsyB

activity was monitored by performing *in vitro* enzyme assays with 50 µl of the separate heat-

killed extract fractions, 1 mM SAM (Sigma-Aldrich), 1 mM DL-MTHB and 1.97 µM DsyB or

no protein (control) in 400 µl total volume. Experiments were done as above with purified DsyB

without addition of heat-killed extracts, but these gave no activity (data not shown). Reactions

were incubated for 30 min at 28°C and then 800 µl of activated charcoal (38 mg ml^{-1} in 0.1 M

acetic acid) was added to the samples and mixed. Samples were centrifuged at 14,000 *g* for

15 mins and the supernatant was retained. For GC analysis, 200 µl of the supernatant was added

to a 2 ml vials with 10 M NaOH (100 µl) and vials were immediately crimped. Crimped vials

were then heated to 80 °C for 10 min and incubated in the dark at 22 °C for 16 h. These samples

were subsequently used for DMS quantification by GC analysis (as above). No DMS was

produced from the no DsyB protein controls.

543   For kinetics analysis of *N. denitrificans* DsyB, the as-isolated protein was activated by

544   addition of 400 µl heat-killed cell lysate fractions liberated from the *L. aggregata dsyB* deletion

545   mutant. K_m and V_{max} values were determined by nonlinear analysis using 1.97 µM DsyB and 0-

546   2 mM SAM (fixed at 1 mM for DL-MTHB kinetic work), or 0-2 mM DL-MTHB (fixed at 1

547   mM for SAM kinetics work) (Fig. 2). The reaction mixture was incubated at 28°C for 30 min

548   before detection of DMSHB as above. Origin version 8.5 was used to calculate K_m.

549

550   **Mass spectrometry analysis**

551   Mass spectrometry coupled with liquid chromatography (LC-MS) was used to confirm the mass

552   of intact (but denatured) DsyB, and also for the analysis of small molecules. For analysis of

553   DsyB, denaturing LC-MS was conducted using a Bruker microQTof-QIII electrospray

554   ionisation time of flight mass spectrometer, operating in positive mode. The spectrometer was

555   calibrated with ESI-L Low concentration tuning mix (Agilent technologies). Samples were

556   prepared by ten-fold dilution of ~100 µM DsyB protein solution with 2% (v/v) acetonitrile and

557   0.1% (v/v) formic acid to 0.5 ml. Samples were chromatographically separated by an UltiMate

558   3000 HPLC system (Dionex) fitted with a ProSwift reversed phase RP-1S column (4.6 × 50

559   mm; Dionex). Hystar (Bruker Daltonics) was used to coordinate mass spectrometer and HPLC

560   operations. Bound proteins were eluted using an isocratic gradient (2 – 100% B) at a flow rate

561   of 0.2 ml min^{-1} using the following solvents: solvent A (water, 0.1% (v/v) formic acid); and

562   solvent B (acetonitrile, 0.1% (v/v) formic acid). The eluant was continuously infused into the

563   source of the mass spectrometer operating with the following parameters: dry gas flow 8 L min^{-}

564   ^{1}; nebuliser gas pressure 1.8 bar; dry gas 240 ˚C; capillary voltage 4,500 V; offset 500 V;

565   collision RF 650 Vpp.

566   Mass spectrometric analysis of small molecule substrates and products in DsyB assay

567   mixtures was performed using hydrophilic interaction liquid chromatography (HILIC) (62),

568   which is particularly useful for the separation of small polar compounds such as MTHB or

569   DMSHB. HILIC-MS (63) experiments were performed using the same mass spectrometer and

570   HPLC system as above, but with the latter fitted with a Luna NH$_2$ column (2 x 100 mm)

571   (Phenomenex). For HILIC chromatography, the following solvents were freshly prepared:

572   solvent A (95% (v/v) aqueous 5 mM ammonium formate pH 3.75, 5% (v/v) acetonitrile);

573   solvent B (95% (v/v) acetonitrile, 5% (v/v) aqueous 100 mM ammonium formate, pH 3.75).

574   Standard compounds (SAM, DL-MTHB, SAH (Sigma-Aldrich), DMSHB (10)) were used to

575   calibrate the elution profile of the HILIC column. Samples were brought to 92% (v/v)

576   acetonitrile and loaded onto a column pre-equilibrated with solvent B. An optimized HILIC

577   gradient was applied and compounds eluted (0.6 ml min^{-1}) using the HILIC gradient between

578   solvent A and solvent B, as previously described (10). The eluant was continuously infused into

579   the source of the mass spectrometer (optimized for $50 - 600$ *m/z*) with the following parameters:

580   dry gas flow 8.5 L/min; dry gas 200 °C; nebulizer pressure 1.2 bar; capillary voltage 4500 V;

581   offset 500 V; collision RF 400 Vpp. Each HILIC-MS run contained an internal sodium formate

582   calibration segment at the end of the run.

583   Non-denaturing MS (often referred to as native MS), in which non-covalently bound

584   protein-cofactor, simple protein-protein, or even multiprotein interactions are preserved (64,

585   65) , was used to investigate substrate binding and confirm the presence of dimeric DsyB. Prior

586   to analysis, protein samples were exchanged into 250 mM ammonium acetate pH 8.0 using

27

587 Zeba spin (Thermo Scientific) or PD mini trap (Cytiva) desalting columns and infused (0.3

588 ml/h) directly into the ESI source of the Bruker microQTof-QIII mass spectrometer with the

589 following parameters: dry gas flow 4 L min^{-1}; nebulizer gas pressure 0.8 bar; dry gas

590 temperature 190 °C; capillary voltage 3000 V; capillary offset 500 V; ion energy 8eV;

591 collisional RF 1500 Vpp; collision cell voltage 5 V; and, ion transmission range 1500 – 5500

592 *m/z*.

593 Processing, isotope pattern simulation and analysis of denaturing LC-MS, HILIC-MS and

594 non-denaturing MS data were carried out using Compass Data Analysis version 4.1. For

595 denaturing LC-MS and non-denaturing MS, neutral mass spectra were generated using ESI

596 compass Maximum Entropy deconvolution algorithm version 1.3. Proteins masses are reported

597 from peak centroids representing the isotope average neutral mass and compared to predicted

598 masses (Expasy) (66).

599

600 **DsyB activity detection using HILIC-MS**

601 Heat-killed J571 fractions that restored MTHB *S*-methyltransferase were added to pure DsyB

602 as above to yield activated samples for analysis here. Samples of as-isolated DsyB prepared in

603 HEPES buffer or activated DsyB, were immediately desalted (PD10, Cytiva) into 25 mM Tris,

604 100 mM NaCl, pH 8.0 prior to conducting HILIC or non-denaturing MS experiments. The

605 methyltransferase activity of DsyB was measured using DL-MTHB and SAM as substrates, as

606 previously described (17), with a slight modification. The reaction mixture (20 μl) consisted of

607 14 μl water, 2 μl of buffer (100 mM Tris-HCl, pH 7.5), 0.5 μl of 20 mM DL-MTHB, 1.5 μl of

608 32 mM SAM as co-substrate, and 2 μl of enzyme solution (DsyB in range 7 – 30 μM depending

609   on particular experiment). The reaction was incubated at 25 °C, over-night and quenched by

610   the addition of 230 μl acetonitrile. Samples were analyzed immediately by mass spectrometry.

611

612   **Crystallization and data collection**

613   The purified DsyB protein was concentrated to ~8 mg/ml in the buffer containing 100 mM

614   NaCl and 10 mM Tris-HCl (pH 8.0). Initial crystallization trials for DsyB were performed using

615   the sitting drop vapor diffusion method at 20°C. To obtain crystals of DsyB-SAM complex, the

616   purified DsyB protein was mixed with 1 mM SAM at 4°C for 30 min. Diffraction-quality

617   crystals of DsyB-SAM complex were obtained in hanging drops containing 0.1 M Hepes (pH

618   7.5), 0.2 M NaCl and 25% (wt/vol) polyethylene glycol (PEG) 3350 after 1-week incubation at

619   20°C. Crystals of the DsyB-SAM complex Se derivative were obtained in hanging drops

620   containing 0.1 M Bis-Tris propane (pH 7.5), 0.2 M sodium acetate and 20% (wt/vol) PEG 3350

621   after 1-week incubation at 20°C. To obtain crystals of DsyB-SAH-MTHB complex, the purified

622   DsyB protein was mixed with SAH (1 mM) and DL-MTHB (1 mM) at 4°C for 30 min. Crystals

623   of DsyB-SAH-MTHB complex were obtained in hanging drops containing 0.1 M Tris (pH 8.0),

624   0.2 M NaCl and 20% PEG 4000 after 1-week incubation at 20°C. X-ray diffraction data were

625   collected on the BL18U1&BL19U1 beamlines at the Shanghai Synchrotron Radiation Facility.

626   The initial diffraction data sets were processed by the HKL3000 program (67).

627

628   **Structure determination and refinement**

629   The crystals of DsyB-SAM complex belong to the $P2_12_12_1$ space group, while the crystals of

630   DsyB-SAH-MTHB complex belong to the $P2_1$ space group. The structure of DsyB-SAM

631   complex Se derivative was determined by SAD phasing. The crystal structures of DsyB-SAM

632   complex and DsyB-SAH-MTHB complex were determined by molecular replacement using

633   the CCP4 program Phaser (68) with the structure of the Se derivative as the search model. The

634   refinements of these structures were performed using Coot (69) and *Phenix* (70). All the

635   structure figures were produced with the PyMOL (http://www.pymol.org/).

636

637   **Analyses of DMSP synthesis genes in cultured microorganisms**

638   The presence or absence of DMSP synthesis genes in 111 cultured DMSP-producing bacteria

639   (published since the discovery of bacterial DMSP synthesis (7, 10, 11, 34, 35) was analyzed

640   (Table S2). This is based previously published work that analyzed their sequenced genomes

641   and/or used degenerate primers to detect the presence of *burB*, *dsyB* and/or *mmtN*. Percentage

642   abundances were calculated for all three DMSP synthesis genes within these cultured organisms,

643   as well as those containing both *dsyB* and *mmtN*.

644   Eukaryotic transcriptomes from the MMETSP (71) (Tables S3 and S4) were analyzed for

645   the presence of *DSYB* and/or *TpMMT* through tblastn searches against DSYB (12) and TpMMT

646   (17) sequences whose enzyme activity had been previously demonstrated. These were manually

647   curated to confirm identity (E value cutoff of $1e^{-30}$ for DSYB), although TpMMT has only been

648   shown to be functional in *T. pseudonana*, we did not assume any sequences below 70% identity

649   to *T. pseudonana* TpMMT to be functional. Strains confirmed to contain *DSYB* and/or *TpMMT*

650   are listed in Table S3 and summarised in Table S4, alongside literature reporting the presence

651   of DMSP synthesis in that particular strain (if tested).

652

Metagenome and metatranscriptome analyses.

Verified sequences (7, 10-12, 72) were aligned using ClustalOmega (73), and profile hidden Markov models (hmms) of *dysB*, *DSYB*, and *mmtN* were constructed using the hmmbuild function of hmmer 3.3 (74). Tara metagenome (OM-RGC_v2_metaG/MATOU_v1_metaG) (prokaryotic/eukaryotic, respectively), and metatranscriptome (OM-RGC_v2_metaT/ MATOU_v1_metaT) sequences together with their abundances and taxonomic assignations were downloaded from the Ocean Gene Atlas site (75) using an hmmsearch e-value threshold of 1e-70 (*dsyB*), 1e-80 (*DSYB*) or 1e-60 (*mmtN*). A blastp search (e-value threshold of 1e-80) was used for *TpMMT*, using the *Thalassiosira pseudonana TpMMT* sequence as query. Environmental *dsyB*/*DSYB* sequences were aligned with *Nisaea denitrificans dsyB* using ClustalOmega (73), and sequences that did not possess all six essential residues were excluded from further analysis. Environmental *TpMMT* sequences greater than 400 amino acids in length were also excluded from further analysis. Prokaryotic sequence abundances were normalized using the median abundance of 10 single copy marker genes/transcripts (76). This gave abundance as a percentage of single copy gene abundance (equivalent to the percentage of cells containing a copy) in the metagenome, and transcription as a percentage of single copy gene transcription in the metatranscriptome. These marker gene/transcript abundances were downloaded from the Ocean Gene Atlas using the hmm profiles developed by Milanese *et al.* (76) with an e-value threshold of 1e-80. The MATOU_v1_metaG (metagenomic) database featured few MIX and MES sampling sites (2 and 7, respectively), limiting the power of comparative analysis between sampling depths, thus, these sites were excluded from analysis. Statistical analysis was performed in R (version 4.02) using RStudio.

DATA AVAILABILITY

The structures of DsyB-SAM complex and DsyB-SAH-MTHB complex have been deposited

in the Protein Data Bank (PDB) under the accession codes 7WDQ and 7WDW, respectively.

CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

697

698 JDT and YZZ designed the research. NEL and YZZ directed the research. CYL, JCC, SNP,

699 ARJM, BJP and SZ performed the experiments. XLC, BTW, MP, and YC helped in data

700 analysis. CYL, JCC, BTW, ARJM and XLC wrote the manuscript. XHZ, YC and NEL edited

701 the manuscript.

ETHICS STATEMENT

702

703 This article does not contain any studies with human participants or animals performed by any

704 of the authors.

705

References

1. Gali M, Devred E, Levasseur M, Royer SJ, Babin M. A remote sensing algorithm for planktonic dimethylsulfoniopropionate (DMSP) and an analysis of global patterns. Remote Sens Environ. 2015;171:171-84.

2. Simo R, Archer SD, Pedros-Alio C, Gilpin L, Stelfox-Widdicombe CE. Coupled dynamics of dimethylsulfoniopropionate and dimethylsulfide cycling and the microbial food web in surface waters of the North Atlantic. Limnol Oceanogr. 2002;47(1):53-61.

3. Zhang XH, Liu J, Liu JL, Yang GP, Xue CX, Curson ARJ, et al. Biogenic production of DMSP and its degradation to DMS-their roles in the global sulfur cycle. Sci China Life Sci. 2019;62(10):1296-319.

4. Stefels J. Physiological aspects of the production and conversion of DMSP in marine algae and higher plants. J Sea Res. 2000;43(3):183-97.

5. Wolfe GV, Steinke M. Grazing-activated production of dimethyl sulfide (DMS) by two clones of Emiliania huxleyi. Limnol Oceanogr. 1996;41(6):1151-60.

6. Sunda W, Kieber DJ, Kiene RP, Huntsman S. An antioxidant function for DMSP and DMS in marine algae. Nature. 2002;418(6895):317-20.

7. Zheng Y, Wang J, Zhou S, Zhang Y, Liu J, Xue C-X, et al. Bacteria are important dimethylsulfoniopropionate producers in marine aphotic and high-pressure environments. Nat Commun. 2020;11(1):4658.

8. Vallina SM, Simó R. Strong relationship between DMS and the solar radiation dose over the global surface ocean. Science. 2007;315(5811):506-8.

9. Curson AR, Todd JD, Sullivan MJ, Johnston AW. Catabolism of dimethylsulphoniopropionate: microorganisms, enzymes and genes. Nat Rev Microbiol. 2011;9(12):849-59.

10. Curson AR, Liu J, Bermejo Martinez A, Green RT, Chan Y, Carrion O, et al. Dimethylsulfoniopropionate biosynthesis in marine bacteria and identification of the key gene in this process. Nat Microbiol. 2017;2:17009.

11. Williams BT, Cowles K, Bermejo Martinez A, Curson ARJ, Zheng Y, Liu J, et al. Bacteria are important dimethylsulfoniopropionate producers in coastal sediments. Nat Microbiol. 2019;4(11):1815-25.

12. Curson ARJ, Williams BT, Pinchbeck BJ, Sims LP, Martinez AB, Rivera PPL, et al. DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton. Nat Microbiol. 2018;3(4):430-9.

13. Summers PS, Nolte KD, Cooper AJL, Borgeas H, Leustek T, Rhodes D, et al. Identification and stereospecificity of the first three enzymes of 3-dimethylsulfoniopropionate biosynthesis in a chlorophyte alga. Plant Physiol. 1998;116(1):369-78.

14. Gage DA, Rhodes D, Nolte KD, Hicks WA, Leustek T, Cooper AJL, et al. A new route for synthesis of dimethylsulphoniopropionate in marine algae. Nature. 1997;387(6636):891-4.

15. Rousseau H, Rousseau-Gueutin M, Dauvergne X, Boutte J, Simon G, Marnet N, et al. Evolution of DMSP (dimethylsulfoniopropionate) biosynthesis pathway: Origin and phylogenetic distribution in polyploid Spartina (Poaceae, Chloridoideae). Mol Phylogenet Evol. 2017;114:401-14.

16. Ito T, Asano Y, Tanaka Y, Takabe T. Regulation of Biosynthesis of Dimethylsulfoniopropionate and Its Uptake in Sterile Mutant of Ulva Pertusa (Chlorophyta). J Phycol. 2011;47(3):517-23.

17. Kageyama H, Tanaka Y, Shibata A, Waditee-Sirisattha R, Takabe T. Dimethylsulfoniopropionate biosynthesis in a diatom Thalassiosira pseudonana: Identification of a gene encoding MTHB-

749 methyltransferase. Arch Biochem Biophys. 2018;645:100-6.

750 18. Cantoni GL. Biological methylation: selected aspects. Annu Rev Biochem. 1975;44:435-51.

751 19. Liscombe DK, Louie GV, Noel JP. Architectures, mechanisms and molecular evolution of natural
752 product methyltransferases. Nat Prod Rep. 2012;29(10):1238-50.

753 20. O'Hagan D, Schmidberger JW. Enzymes that catalyse S(N)2 reaction mechanisms. Natural Product
754 Reports. 2010;27(6):900-18.

755 21. Dunbar KL, Scharf DH, Litomska A, Hertweck C. Enzymatic Carbon-Sulfur Bond Formation in
756 Natural Product Biosynthesis. Chem Rev. 2017;117(8):5521-77.

757 22. Urios L, Michotey V, Intertaglia L, Lesongeur F, Lebaron P. Nisaea denitrificans gen. nov., sp nov
758 and Nisaea nitritireducens sp nov., two novel members of the class Alphaproteobacteria from the
759 Mediterranean Sea. Int J Syst Evol Micr. 2008;58:2336-41.

760 23. Meinnel T, Mechulam Y, Blanquet S. Methionine as Translation Start Signal - a Review of the
761 Enzymes of the Pathway in Escherichia-Coli. Biochimie. 1993;75(12):1061-75.

762 24. Pellicer Martinez MT, Crack JC, Stewart MY, Bradley JM, Svistunenko DA, Johnston AW, et al.
763 Mechanisms of iron- and O2-sensing by the [4Fe-4S] cluster of the global iron regulator RirA. eLife.
764 2019;8:e47804.

765 25. Volbeda A, Dodd EL, Darnault C, Crack JC, Renoux O, Hutchings MI, et al. Crystal structures of
766 the NO sensor NsrR reveal how its iron-sulfur cluster modulates DNA binding. Nat Commun. 2017;8:
767 15052.

768 26. Crack JC, Thomson AJ, Le Brun NE. Mass spectrometric identification of intermediates in the O2-
769 driven [4Fe-4S] to [2Fe-2S] cluster conversion in FNR. Proc Natl Acad Sci USA. 2017;114(16):E3215-
770 E23.

771 27. Chen SC, Huang CH, Lai SJ, Liu JS, Fu PK, Tseng ST, et al. Structure and mechanism of an
772 antibiotics-synthesizing 3-hydroxykynurenine C-methyltransferase. Sci Rep-Uk. 2015; 5:10100.

773 28. Martin JL, McMillan FM. SAM (dependent) I AM: the S-adenosylmethionine-dependent
774 methyltransferase fold. Curr Opin Struct Biol. 2002;12(6):783-93.

775 29. Hotta K, Keegan RM, Ranganathan S, Fang M, Bibby J, Winn MD, et al. Conversion of a Disulfide
776 Bond into a Thioacetal Group during Echinomycin Biosynthesis. Angew Chem Int Edit. 2014;53(3):824-
777 8.

778 30. Peng Y, Feng Q, Wilk D, Adjei AA, Salavaggione OE, Weinshilboum RM, et al. Structural basis of
779 substrate recognition in thiopurine s-methyltransferase. Biochemistry. 2008;47(23):6216-25.

780 31. Coiner H, Schroder G, Wehinger E, Liu CJ, Noel JP, Schwab W, et al. Methylation of sulfhydryl
781 groups: a new function for a family of small molecule plant O-methyltransferases. Plant J.
782 2006;46(2):193-205.

783 32. Dang HY, Li TG, Chen MN, Huang GQ. Cross-Ocean distribution of Rhodobacterales bacteria as
784 primary surface colonizers in temperate coastal marine waters. Appl Environ Microb. 2008;74(1):52-60.

785 33. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL:
786 homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296-
787 W303.

788 34. Trottmann F, Ishida K, Franke J, Stanisic A, Ishida-Ito M, Kries H, et al. Sulfonium Acids Loaded
789 onto an Unusual Thiotemplate Assembly Line Construct the Cyclopropanol Warhead of a Burkholderia
790 Virulence Factor. Angew Chem Int Edit. 2020;59(32):13511-5.

791 35. Liu J, Zhang Y, Liu J, Zhong H, Williams BT, Zheng Y, et al. Bacterial Dimethylsulfoniopropionate
792 Biosynthesis in the East China Sea. Microorganisms. 2021;9(3):657.

793   36. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled
794   genomes from the global oceans. Sci Data. 2018;5:170203.
795   37. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, et al. The Marine
796   Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional
797   diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol.
798   2014;12(6):e1001889.
799   38. Bucciarelli E, Ridame C, Sunda WG, Dimier-Hugueney C, Cheize M, Belviso S. Increased
800   intracellular concentrations of DMSP and DMSO in iron-limited oceanic phytoplankton Thalassiosira
801   oceanica and Trichodesmium erythraeum. Limnol Oceanogr. 2013;58(5):1667-79.
802   39. Keller MD, Bellows WK, Guillard RRL. Dimethyl Sulfide Production in Marine Phytoplankton.
803   Biogenic Sulfur in the Environment. ACS Symposium Series. 393: American Chemical Society; 1989.
804   p. 167-82.
805   40. Keller MD, Korjeff-Bellows W. Physiological Aspects of the Production of
806   Dimeyhtlsulfoniopropionate (DMSP) by Marine Phytoplankton. In: Kiene RP, Visscher PT, Keller MD,
807   Kirst GO, editors. Biological and Environmental Chemistry of DMSP and Related Sulfonium
808   Compounds. Boston, MA: Springer US; 1996. p. 131-42.
809   41. Meng A, Corre E, Probert I, Gutierrez-Rodriguez A, Siano R, Annamale A, et al. Analysis of the
810   genomic basis of functional diversity in dinoflagellates using a transcriptome-based sequence similarity
811   network. Mol Ecol. 2018;27(10):2365-80.
812   42. Caruana AMN, Le Gac M, Herve F, Rovillon GA, Geffroy S, Malo F, et al. Alexandrium pacificum
813   and Alexandrium minutum: Harmful or environmentally friendly? Mar Environ Res. 2020;160:105014.
814   43. Wolfe GV, Strom SL, Holmes JL, Radzio T, Olson MB. Dimethylsulfoniopropionate cleavage by
815   marine phytoplankton in response to mechanical, chemical, or dark stress. J Phycol. 2002;38(5):948-60.
816   44. Royer C, Borges AV, Lapeyra Martin J, Gypens N. Drivers of the variability of
817   dimethylsulfonioproprionate (DMSP) and dimethylsulfoxide (DMSO) in the Southern North Sea. Cont
818   Shelf Res. 2021;216:104360.
819   45. Caruana AMN, Steinke M, Turner SM, Malin G. Concentrations of dimethylsulphoniopropionate
820   and activities of dimethylsulphide-producing enzymes in batch cultures of nine dinoflagellate species.
821   Biogeochemistry. 2012;110(1-3):87-107.
822   46. Caruana AMN, Malin G. The variability in DMSP content and DMSP lyase activity in marine
823   dinoflagellates. Prog Oceanogr. 2014;120:410-24.
824   47. Gutierrez-Rodriguez A, Pillet L, Biard T, Said-Ahmad W, Amrani A, Simo R, et al. Dimethylated
825   sulfur compounds in symbiotic protists: A potentially significant source for marine DMS(P). Limnol
826   Oceanogr. 2017;62(3):1139-54.
827   48. Yost DM, Mitchelmore CL. Dimethylsulfoniopropionate (DMSP) lyase activity in different strains
828   of the symbiotic alga Symbiodinium microadriaticum. Mar Ecol Prog Ser. 2009;386:61-70.
829   49. Franklin DJ, Steinke M, Young J, Probert I, Malin G. Dimethylsulphoniopropionate (DMSP),
830   DMSP-lyase activity (DLA) and dimethylsulphide (DMS) in 10 species of coccolithophore. Mar Ecol
831   Prog Ser. 2010;410:13-23.
832   50. Steinke M, Wolfe GV, Kirst GO. Partial characterisation of dimethylsulfoniopropionate (DMSP)
833   lyase isozymes in 6 strains of Emiliania huxleyi. Mar Ecol Prog Ser. 1998;175:215-25.
834   51. Spielmeyer A, Gebser B, Pohnert G. Dimethylsulfide sources from microalgae: Improvement and
835   application of a derivatization-based method for the determination of dimethylsulfoniopropionate and
836   other zwitterionic osmolytes in phytoplankton. Mar Chem. 2011;124(1-4):48-56.

837   52.   Kinsey JD, Kieber DJ, Neale PJ. Effects of iron limitation and UV radiation on Phaeocystis
838   antarctica growth and dimethylsulfoniopropionate, dimethylsulfoxide and acrylate concentrations.
839   Environ Chem. 2016;13(2):195-211.

840   53.   Jean N, Bogé G, Jamet J-L, Richard S, Jamet D. Annual contribution of different plankton size
841   classes to particulate dimethylsulfoniopropionate in a marine perturbed ecosystem. J Mar Syst.
842   2005;53(1):235-47.

843   54.   Beringer JE. R Factor Transfer in Rhizobium-Leguminosarum. J Gen Microbiol. 1974;84(Sep):188-
844   98.

845   55.   Gonzalez JM, Whitman WB, Hodson RE, Moran MA. Identifying numerically abundant culturable
846   bacteria from complex communities: an example from a lignin enrichment culture. Appl Environ
847   Microbiol. 1996;62(12):4433-40.

848   56.   Baumann P, Baumann, L. The Prokaryotes: A Handbook on Habitats, Isolation and Identification
849   of Bacteria. 1st edn ed: Springer-Verlag Berlin Heidelberg; 1981.

850   57.   Figurski DH, Helinski DR. Replication of an origin-containing derivative of plasmid RK2
851   dependent on a plasmid function provided in trans. Proc Natl Acad Sci USA. 1979;76(4):1648-52.

852   58.   Carrion O, Curson ARJ, Kumaresan D, Fu Y, Lang AS, Mercade E, et al. A novel pathway producing
853   dimethylsulphide in bacteria is widespread in soil environments. Nat Commun. 2015;6:6579.

854   59.   Tett AJ, Rudder SJ, Bourdes A, Karunakaran R, Poole PS. Regulatable vectors for environmental
855   gene expression in Alphaproteobacteria. Appl Environ Microbiol. 2012;78(19):7137-40.

856   60.   Lalitha S. Primer Premier 5. Biotech Software & Internet Report. 2000;1(6):270-2.

857   61.   Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic
858   Acids Res. 2001;29(9):e45.

859   62.   Buszewski B, Noga S. Hydrophilic interaction liquid chromatography (HILIC)-a powerful
860   separation technique. Anal Bioanal Chem. 2012;402(1):231-47.

861   63.   Tang DQ, Zou L, Yin XX, Ong CN. HILIC-MS for metabolomics: An attractive and complementary
862   approach to RPLC-MS. Mass Spectrom Rev. 2016;35(5):574-600.

863   64.   Heck AJR. Native mass spectrometry: a bridge between interactomics and structural biology. Nat
864   Methods. 2008;5(11):927-33.

865   65.   Leney AC, Heck AJR. Native Mass Spectrometry: What is in the Name? J Am Soc Mass Spectr.
866   2017;28(1):5-13.

867   66.   Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein
868   identification and analysis tools in the ExPASy server. Methods Mol Biol. 1999;112:531-52.

869   67.   Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data
870   reduction and structure solution--from diffraction images to an initial model in minutes. Acta Crystallogr
871   D Biol Crystallogr. 2006;62(Pt 8):859-66.

872   68.   Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4
873   suite and current developments. Acta Crystallogr D Biol Crystallogr. 2011;67:235-42.

874   69.   Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr
875   D Biol Crystallogr. 2010;66:486-501.

876   70.   Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a
877   comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol
878   Crystallogr. 2010;66:213-21.

879   71.   Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, et al. The Marine
880   Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional

881   Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. PLoS Biol. 2014;12(6).

882   72.   Liao CS, Seebeck FP. In Vitro Reconstitution of Bacterial DMSP Biosynthesis. Angew Chem Int

883   Edit. 2019;58(11):3553-6.

884   73.   Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-

885   quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539-.

886   74.   Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching.

887   Nucleic Acids Res. 2011;39:W29-37.

888   75.   Villar E, Vannier T, Vernette C, Lescot M, Cuenca M, Alexandre A, et al. The Ocean Gene Atlas:

889   exploring the biogeography of plankton genes online. Nucleic Acids Res. 2018;46(W1):W289-W95.

890   76.   Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh HJ, Cuenca M, et al. Microbial abundance,

891   activity and population genomic profiling with mOTUs2. Nat Commun. 2019;10(1):1014.

892

Figure legends

Fig. 1. Predicted DMSP biosynthesis pathways (10). Different pathways are shown in different colours. Enzymes of interest in this study (DsyB/DSYB) are in red. *SMM is converted to DMSP-aldehyde directly in *Wollastonia*. Dotted lines represent unconfirmed steps of the decarboxylation pathway. Abbreviations: SMM, *S*-methylmethionine; MTOB, 4-methylthio-2-oxobutyrate; MTHB, 4-methylthio-2-hydroxybutyrate; DMSHB, 4-dimethylsulfonio-2-hydroxybutyrate; MTPA, 3-methylthiopropylamine; MMPA, methylmercaptopropionate.

Fig. 2. Characterisation of *N. dentrificans* DMSP production, *dsyB* transcription and the DsyB enzyme. *N. dentrificans* DMSP production (A) and *dsyB* transcription (B) observed under different conditions. Standard conditions were MBM medium at 35 PSU with 10 mM NH_4Cl compared to low salinity (5 PSU), high salinity (50 PSU) and nitrogen starvation conditions (where cells in standard MBM were resuspended in standard MBM with no added NH_4Cl). (C) A non-linear fit curve for MTHB methylation by DsyB. Initial rates were determined with 1.97 µM DsyB (50 mM Tris-HCl, 100 mM NaCl, pH 8.0) and 0 – 2 mM MTHB. K_m was 0.14 ± 0.02 mM. (D) A non-linear fit curve for SAM demethylation by DsyB. Initial rates were determined with 1.97 µM DsyB and 0-2 mM SAM in the same reaction buffer. K_m was 0.16 ± 0.002 mM. The error bar represents standard deviation of triplicate experiments.

Fig. 3. Identification of DsyB substrates and products by LC-MS. (A) The chemical equation for MTHB *S*-methylation into DMSHB. (B) Extracted ion chromatograms for MTHB, SAM, DMSHB and SAH. Here, mass spectrometry data were analyzed to extract ion counts as a function of elution volume for the *m/z* ions indicated, which correspond to the substrates and

915 products of the DsyB-catalyzed reaction. The red broken lines indicate elution volumes of the

916 molecules when run as standards. (C) – (F) Mass spectra recorded for the eluted species, as

917 indicated. Inset are spectra over a narrower *m/z* range (black lines) along with simulated spectra

918 (showing the isotope distribution, red line) for each molecule, providing clear confirmation of

919 identity.

920 **Fig. 4. Overall structure analysis of DsyB**. (A) Ribbon representation of DsyB dimer. Each

921 DsyB monomer contains an N-domain and a C-domain. SAM molecules are shown as sticks

922 coloured in cyan. (B) Analysis of the form of DsyB in solution by gel filtration. Conalbumin

923 (molecular mass = 75,000 Da; GE Healthcare) and ovalbumin (molecular mass = 43,000 Da;

924 GE Healthcare) were used as markers. The predicted molecular mass of DsyB monomer is

925 37,215 Da. (C) The overall structure of DsyB-SAH-MTHB complex. The SAH molecule

926 (coloured in purple) and the MTHB molecule (coloured in green) are represented as sticks. (D)

927 Superimposed structures of DsyB with (coloured in purple) and without (coloured in yellow)

928 binding the SAM molecule.

929 **Fig. 5. Analyses of residues of DsyB involved in binding SAM and MTHB.** (A) Interactions

930 between DsyB residues and SAM. SAM is coloured in purple. The possible hydrogen bonds

931 are represented by dashed lines. The $2F_o$–F_c density for SAM is contoured in blue at 1.5σ. (B)

932 Electrostatic surface of the crystal structure of DsyB. The SAM binding site and the MTHB

933 binding site can be clearly identified. (C) The binding site of MTHB. Residues of DsyB that

934 may participate in binding MTHB are shown in yellow. The $2F_o$–F_c density for SAH (coloured

935 in purple) is contoured in grey at 1.5σ. The $2F_o$–F_c density for MTHB (coloured in green) is

936 contoured in grey at 1.0σ. (D) Enzymatic activities of WT DsyB and site-directed mutants. The

937   error bar represents standard deviation of triplicate experiments.

938   **Fig. 6. A proposed catalytic mechanism of DsyB**. (A) The schematic diagrams of the DsyB

939   conformational change triggered by the binding of SAM. (B) The sulfur atom of MTHB attacks

940   on the methyl group of SAM to generate DMSHB and SAH. The MTHB molecule and DMSHB

941   molecule are shown in black. And the SAM molecule and the SAH molecule are shown in red.

942   **Fig. 7. Analysis of MATOU for *dsyB* and *mmtN*.** (**A**) Normalized abundance of *dsyB* and

943   *mmtN* in *Tara* metagenomes and metatranscriptomes, by sampling depth. DCM = deep

944   chlorophyll maximum layer, MES = mesopelagic layer, MIX = epipelagic wind mixed layer,

945   SRF = surface water layer. Abundances are normalized as a percentage of the median gene or

946   transcript abundance of 10 single copy marker genes. Box plots show median values (central

947   black line), and lower and upper hinges correspond to the 1st and 3rd quartiles of the data.

948   Kruskal-Wallis X^2 values for comparisons of median abundances between *dsyB* and *mmtN*

949   (across all depths combined) are shown. (**B**) Taxonomic assignment and relative abundance (as

950   a percentage) of *dsyB* and *mmtN* sequences in the *Tara* metagenomes and metatranscriptomes.

951   Taxa designated *Alphaproteobacteria* lack further taxonomic resolution.

952   **Fig. 8. Analysis of MATOU for *DSYB* and *TpMMT*.** (**A**) Normalized abundance of *DSYB* and

953   *TpMMT* in MATOU metagenome and metatranscriptomes, by fractions containing

954   picoeukaryotes (i.e., with a minimum filter size of < 3 μm) (blue) and fractions excluding

955   eukaryotes (i.e., with a minimum filter size of ≥ 3 μm) (red). DCM and SRF depths are

956   combined for the purposes of this analysis. Abundances are normalized as reads per kilobase

957   per million mapped reads (RPKM). Box plots show median values (central black line), and

958   lower and upper hinges correspond to the 1st and 3rd quartiles of the data. Kruskal-Wallis X^2

959 values for comparisons between *DSYB*/*TpMMT* and fraction abundance are shown. Letters

960 denote gene or transcript/fraction combinations that are significantly different ($p < 0.05$) by

961 post-hoc Dunn's test, using Holm's correction. (**B**) Taxonomic assignment and relative

962 abundance (as a percentage) of *DSYB* and *TpMMT* sequences in the MATOU metagenome and

963 metatranscriptome. Taxonomy is reported as Phylum (if available) for *DSYB* and as Class (if

964 available) for *TpMMT*. Taxa designated *Eukarya* lack further taxonomic resolution.

Fig. 1

Fig. 2

(A) DMSP (pmol (μg protein)$^{-1}$ d^{-1}) vs conditions: Standard, Low salinity, High salinity, Nitrogen starvation

(B) Normalized fold change (*dsyB*) vs conditions: Standard, Low salinity, High salinity, Nitrogen starvation

(C) V (nmol mg^{-1} min^{-1}) vs MTHB (mM); *Km*: 0.14 ± 0.02 mM

(D) V (nmol mg^{-1} min^{-1}) vs SAM (mM); *Km*: 0.16 ± 0.002 mM

Fig. 3

(A)

MTHB → DsyB (SAM SAH) → DMSHB

(B)

(C)

SAM

399.1606 (1+)

Simulation [C₁₅H₂₂N₆O₅S] + H⁺

399.1606 (1+)

(D)

MTHB

193.0136 (1+)

Simulation [C₅H₁₀O₃S] + Na⁺

173.0304 (1+)

151.0445 (1+) +Na 173.0304 (1+) +Na

(E)

SAH

385.1415 (1+)

Simulation [C₁₄H₂₀N₆O₅S] + H⁺

385.1415 (1+)

(F)

DMSHB

165.0645 (1+)

Simulation [C₆H₁₂O₃S] + H⁺

165.0645 (1+)

45

Fig. 4

(A)

(B)

(C)

(D)

Fig. 5

(A)

Ser150
Ser239
Asp223
Gly173
Asp196
Ala224

(B)

MTHB binding site

SAM binding site

(C)

His291
Tyr129
Tyr142
Gln101
Gln146
Tyr97

(D)

nmol DMSP/mg protein

WT Y97A Y97F Q101A Q101E Y129A Y129F Y142A Y142F Q146A Q146E H291A

47

Fig. 6

(A)

(B)

Fig. 7

(A)

Metagenome

dsyB   *mmtN*

Abundance (%)

Kruskal−Wallis X^2 = 83.781, p < 0.001

DCM MES MIX SRF DCM MES MIX SRF

Metatranscriptome

dsyB   *mmtN*

Abundance (%)

Kruskal−Wallis X^2 = 33.64, p < 0.001

DCM MES MIX SRF DCM MES MIX SRF

(B)

Metagenome

Relative Abundance (%)

dsyB   *mmtN*

DCM MES MIX SRF DCM MES MIX SRF

Metatranscriptome

Relative Abundance (%)

dsyB   *mmtN*

DCM MES MIX SRF DCM MES MIX SRF

Taxonomy

Alphaproteobacteria
Rhodobacteraceae
Rhizobiales bacterium NRL2
Amorphus coralli
Labrenzia
Nisaea denitrificans
Thalassospira
Thalassospira lucentensis

Fig. 8

(A)

Metagenome

DSYB   *TpMMT*

Kruskal–Wallis X^2 = 355.86, $p < 0.001$

Metatranscriptome

DSYB   *TpMMT*

Kruskal–Wallis X^2 = 356.23, $p < 0.001$

(B)

DSYB   *TpMMT*

Taxonomy

- Eukaryota
- Dinophyceae
- Haptophyceae
- Coccolithales
- Coccosphaerales
- Isochrysidales
- Phaeocystales
- Prymnesiales
- Stramenopiles
- Bacillariophyta
- Dictyochophyceae
- Coscinodiscophyceae
- Mediophyceae

Mechanistic insights into the key marine dimethylsulfoniopropionate

synthesis enzyme DsyB/DSYB

Supplementary materials:
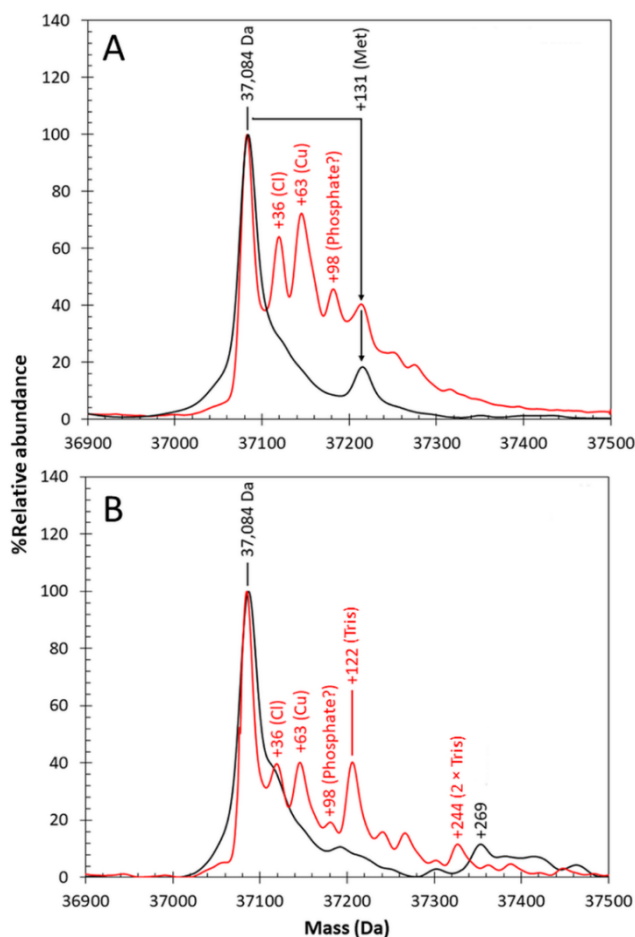
Fig. S1. Mass spectrometric analyses of DsyB. (A) Deconvoluted LC-MS spectrum of as-isolated DsyB (~10 μM) in 25 mM Tris, 100 mM NaCl pH 8 (black line), and deconvoluted non-denaturing spectrum of as-isolated DsyB (~20 μM) in 250 mM ammonium acetate pH8.0 buffer (red line). The partial cleavage of the N-terminal Met residue is indicated by the observation of two protein peaks separated by the mass of a single Met residue (131 Da). (B) As in (A) but DsyB was pre-activated by addition of lysate from *L. aggregata dsyB* deletion strain. The mass of the main protein peak is indicated and adduct species are labelled with the additional mass and origin (if known). The DsyB sample used for experiments reported in (B)

977   was different to that of (A), with a more extensive degree of N-terminal Met cleavage; hence,

978   the +131 Da peak is less well resolved.

979
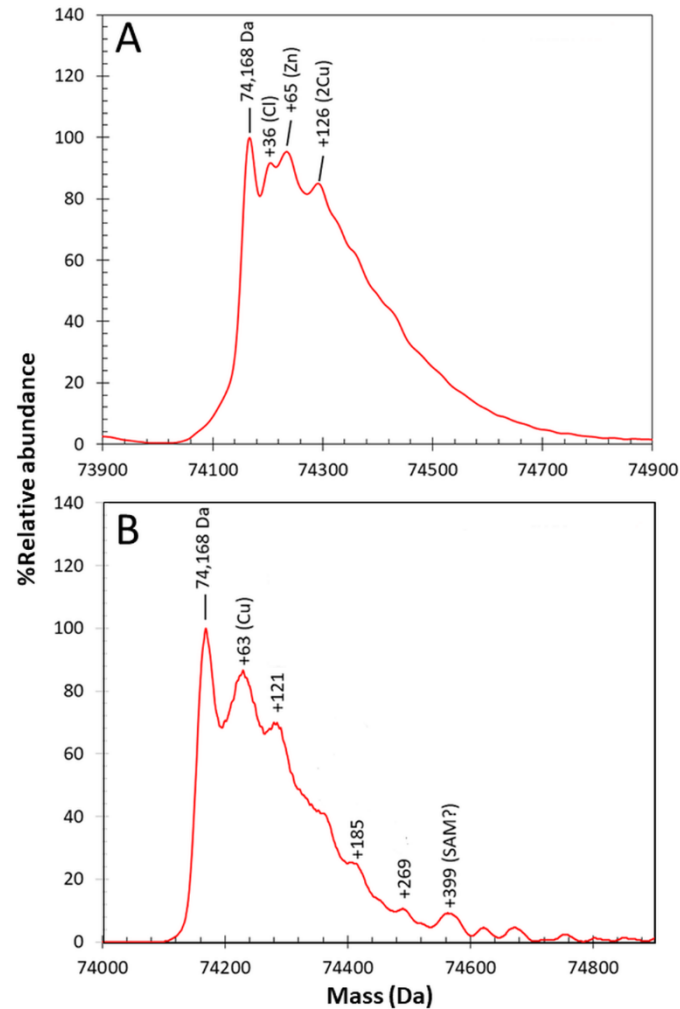
980   **Fig. S2. Mass spectrometric analyses of dimeric DsyB under non-denaturing conditions**.

981   (A) Deconvoluted non-denaturing mass spectrum of as-isolated DsyB (~20 μM) in 250 mM

982   ammonium acetate pH 8.0 buffer.   (B) As in (A) but DsyB was pre-activated by addition of

983   lysate from *L. aggregata dsyB* deletion strain. The mass of the main DsyB dimer peak is

984   indicated and adduct species are labelled with the additional mass and origin (if known).

985

986

987 **Fig. S3. Substrate binding to DsyB probed by mass spectrometry**. Deconvoluted mass

988 spectra of DsyB under non-denaturing conditions in the presence of SAM and MTHB. (A)

989 Peaks due to a DsyB dimer, metal (+65 Da) and $(SAM-Cl)_2$ adducts are indicated. (B) The

990 DsyB monomer and adducts due to SAM and MTHB are as indicated. Presence of both SAM

991 and MTHB leads to reduction in substrate adduct peaks.

992

993

994
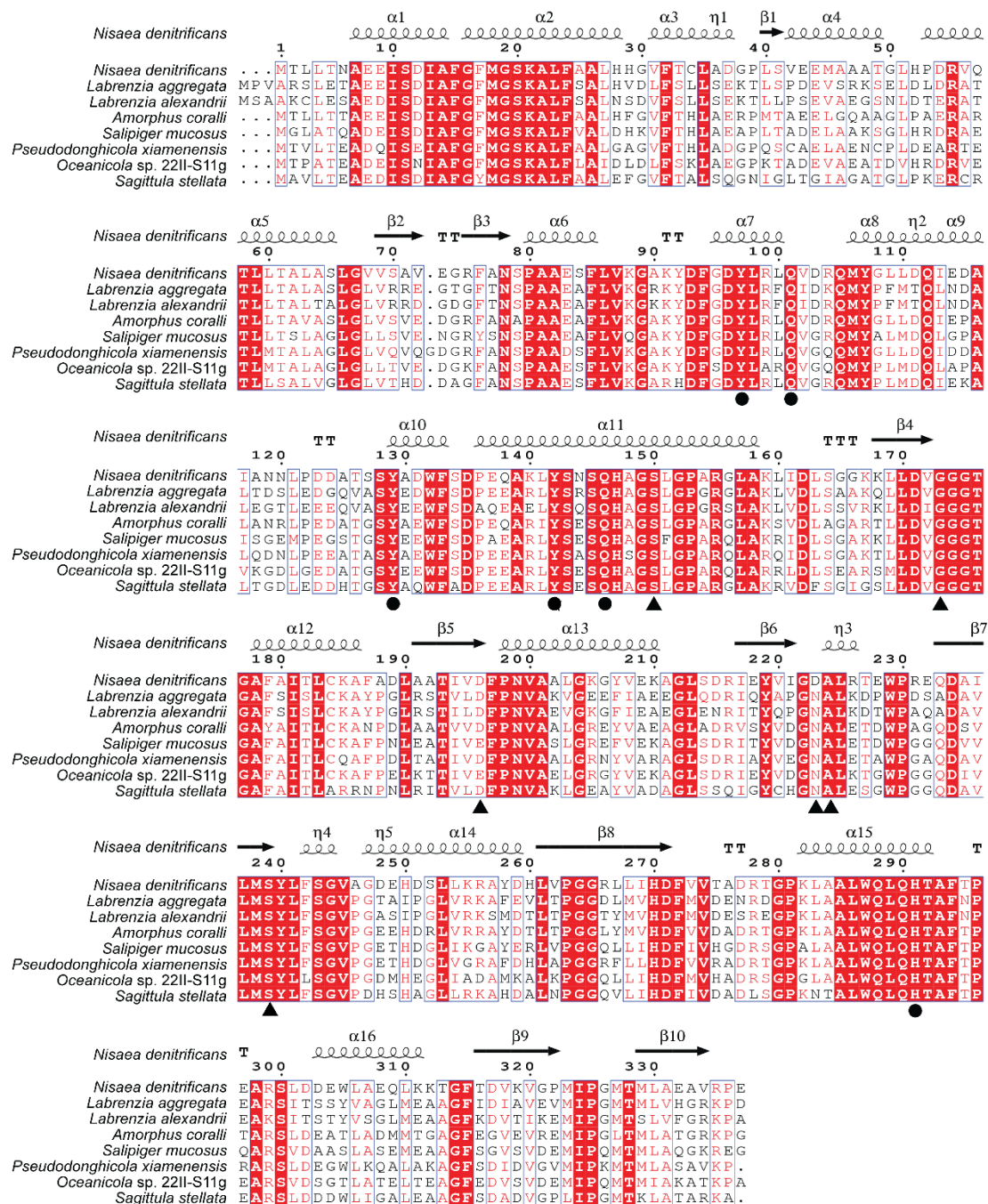
Fig. S4. Sequence alignment of bacterial DsyB proteins. *Nisaea denitrificans* belongs to the order Rhodospirillales, *Amorphus coralli* belongs to Rhizobiales, and the other strains are Rhodobacterales. Black dots indicate residues involved in MTHB binding, and black triangles indicate residues involved in SAM binding.

999

1000

1001
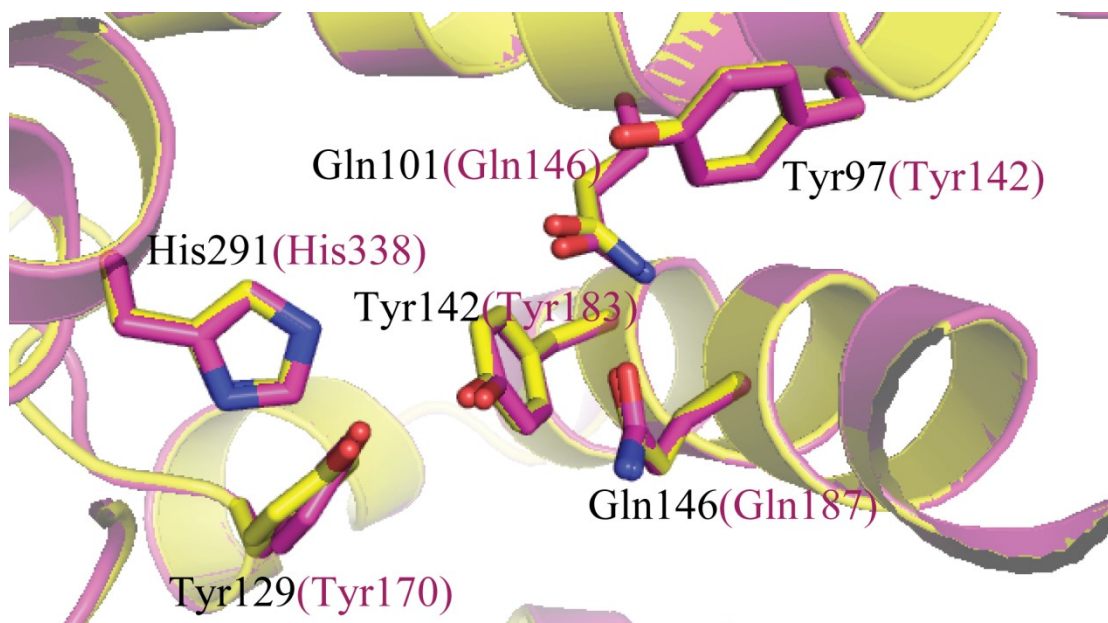
1002 **Fig. S5. Structural alignment of DsyB from *N. denitrificans* (yellow) and DSYB from**

1003 ***Chrysochromulina tobin* CCMP291 (purple).** The structure of DSYB was modelled using

1004 SWISS-MODEL (https://swissmodel.expasy.org/). Residues involved in binding MTHB from

1005 DsyB and from DSYB are labelled in black and purple, respectively.

1006

1007

1008

1009

1010

1011

1012

α1   α2   α3 η1   β1 α4

Nisaea denitrificans

```
                                                                              1
Nisaea denitrificans          ..............|.....|..................MT LLT....................
Acropora cervicornis          .MAAAFAAARGLTRAISRCASKRVPEQALFSRLSAPAPPALAALP..GVRAFSRTSW..ES
Chrysochromulina tobin CCMP291 MLLVTARRGLFQSRAVLRATP..........ALATAAPALIGAR....RTLCNSLF..EL
Fragilariopsis cylindrus CCMP1102 ..........MTLSTPTTVG............SGEDEVSKKN................
Lingulodinium polyedrum CCMP1936  .MAFAPRTSPLLARAVTRAWPR..AQRSAAAALRSWARPALAAKPPLASRAFARTSW..DE
Symbiodinium microadriaticum CCMP2467 .........MSRPIRRLKV..........VTGHTTALRTSA...........CG
Prymnesium parvum CCAP946/1B  MLRLAPR...LPTRALVRHALR........AHALPLARPALPSGT....RLFAS.....AP
Emiliania huxleyi CCMP1516    ...MLLRLGSMLARRGLHAAS...........ARFAAAPALRLGR...SRALSTAPVDDEM
```

```
                              α1          α2              α3   η1          β1    α4
Nisaea denitrificans          .NAEEISDIAFGFMGSKALFAALHHGVFTCLADGP.........LSVEEMAAATGLHPDR
Acropora cervicornis          GDDVDLDSVAYGFMASQALFSALELGIFDKVAAAGE.......KGCAAKDVQQACGVEGPR
Chrysochromulina tobin CCMP291 SDDIDVDSVSYGFMASQAIFCALEVGVYEAVAGAGE......GGLTADALKTACSINAPR
Fragilariopsis cylindrus CCMP1102 DDQPAIDDIAYGFMGSKALSVALKIGVFDAIDVVSKGEDNAVEGGATLNQILTHCKVPKER
Lingulodinium polyedrum CCMP1936 GEDVDLDSVAYGFMASQALFTGLELGIFDHIAAAGA.......GGLSAAGIGKACGIEAPR
Symbiodinium microadriaticum CCMP2467 SEDVDLDSLAYGFMASQAFAALELGVFDYLAKGP.........QSVSDLAEGCGVPCNR
Prymnesium parvum CCAP946/1B  ADDIDVDNVAYGFMASQALFTGLEMGLFDAIAAGPE.......AGLNIDELKAAANCSAPR
Emiliania huxleyi CCMP1516    EDDIDVDGVAYGFMASQALFAGLELGIFDAIAAAG........GSVDLPSLQRSSGVTAPR
```

```
                              α5          β2    β3    α6        α7      α8  η2
Nisaea denitrificans          VQTLLTALASLGVVSA....VEGRFANSPAAESFLVKGAKYDFGDYLRLQVDRQMYGLLDQ
Acropora cervicornis          LTTLLTALTAVKCLRR...SDEGLYTLSPNTAQYMVSSSRHYYGDYLQYQIGRQFYHRMGA
Chrysochromulina tobin CCMP291 LQTLLTLSTLAIKSLKRN..AATGAYTLSPNSARYLVQSSRQYYGDYLRYQMGRQFYHRMGA
Fragilariopsis cylindrus CCMP1102 LRTLLSACVALKLINRRVLRGEDVFSLPEASAEQLVKSSKRYWGDYIVGQVDAQFYVRMKD
Lingulodinium polyedrum CCMP1936 VQTLLTSLVAVKCLKR...DASAMYTLSPNTAQYMVTSSRHFYGDYLRYQIGRQFYHRMGA
Symbiodinium microadriaticum CCMP2467 LQTLLTALVAAKCLRR...DESQLYSNSPNVQKFMVSTSKAYYGDYFKHQVGGLFYARMGQ
Prymnesium parvum CCAP946/1B  LQTLVTSLVAIKSLKR...TPDGRYTLSPNTARFLVQSSKQYYGDYLKYQMGRQFYHRMGA
Emiliania huxleyi CCMP1516    LQTLLTALVASKALRR...SRDGAYDLSPNAARFLVSSSRHFYGDYLKLQIGRQFYHRMGA
                                                                  ●     ●
```

```
                              α9   TT   α10        α11                      TT...
Nisaea denitrificans          IEDAIANNLPDDATSSYADWFSDPEQAKLYSNSQHAGSLGPARGLAK....LIDLSG...
Acropora cervicornis          LPEVMTT..GKAPS..YASWFSDPEVAKTYTQAQHNGSVATAKYLVR...KKLDLGG...
Chrysochromulina tobin CCMP291 LPDVMTT..GEAPS..YASWFSDPATAAMYTQAQHNGSVATAKYLVK...KKLQLGS...
Fragilariopsis cylindrus CCMP1102 LDTTILT..GDTSSDGYEAWFDKDPDAAKRYTQAQHNGSLATGYGLLKR.LPELSQGEKYP
Lingulodinium polyedrum CCMP1936 LPEVMTS..GKAPS..YASWFSDPEVARTYTQAQHNGSVATAKYLIK...KKLQLGG...
Symbiodinium microadriaticum CCMP2467 LAKVLRG..EEVLD..YSQWFSDPTVASLYTQAQHNGSLATAKSLFR....KVSLAT...
Prymnesium parvum CCAP946/1B  LPDVMTS..GEAPS..YASWFSDPETAATYTQAQHNGSVATAKYLIK...KKLQLGD...
Emiliania huxleyi CCMP1516    LTDVMKT..GEAPS..YASWFSDPEVAATYTRAQHNGSVATAKALLKRMAGSVSLDG...
                                              ●              ●      ●   ▲
```

```
                              β4    α12        β5    α13              β6
Nisaea denitrificans          GKKLLDVGGGTGAFAITLCKAFADLAATIVDFPNVAALGKGYVEKAG..LSDRTEYVIGDA
Acropora cervicornis          ISSMLDVGGGSGAFSYVFTEATPGLKSTVLELPEVCRTGEGIKAQQPQDIQDRVSFVELDA
Chrysochromulina tobin CCMP291 VTSMLDVGGGSGAFSYVFVESTPGLKSTVLELPEVCKTGNGIKAKQSPDVQARVAFIELDA
Fragilariopsis cylindrus CCMP1102 NLRMLDIGGGSGAFSIATARKVKDADCVILDLPNVIKVAEEIISKEDEQVRNRLSTVALSA
Lingulodinium polyedrum CCMP1936 ISAMLDVGGGSGAFSYVFTQATPGLHSKVLELPEVCRTGEGIRAKQPEDVRSRVSFVELDA
Symbiodinium microadriaticum CCMP2467 TGRMLDVGGGSGAFSLQAAQMNPELRATVLDLPEVCRVGRSLMEQAASPEKSRVGFKELDA
Prymnesium parvum CCAP946/1B  AATMLDVGGGSGAFSYVFTEATPGLSSTVLELPEVCRTGEAIKAKQPVSVQERVKFVELDA
Emiliania huxleyi CCMP1516    ASKMLDVGGGSGAFSYVFTGKHPQLSSQILELPEVCATGNTIRAEQPADVQQRVSFSELDA
                                 ▲                     ▲                   ▲▲
```

```
                              η3  β7    η4  η5  α14          β8       TT
Nisaea denitrificans          LRT.EWPREQ...DAILMSYLFSGVAGDEHDSLLKRAYDHLVPGGRLLIHDFVVTADRTGP
Acropora cervicornis          TSP.DWPVSDSNYDIVLMSYISGSVPESVILPLYKNAFKALRPGGRLLVHDFVVNDSLDGP
Chrysochromulina tobin CCMP291 TKP.DWPVKEADYELVLMSYISGSVPEAIIGKLYENAFKALKPGGRLLVHDFMVNDSLDGP
Fragilariopsis cylindrus CCMP1102 ADPGNWMIEDESFDVVLMSYVSGSISSDALSGLYQNAFRALRPGGMVVIHDFFVDNNGEGP
Lingulodinium polyedrum CCMP1936 SSP.TWPVDDSAFDVVLMSYISGSVPEPIIGSLYANAMKALRPGGRLLVHDFVVNDSLDGP
Symbiodinium microadriaticum CCMP2467 TSP.AWPVKAESQDLALMSYLCGSVPEDVIAQLFQNAFRVLRSGGRLVVHDFVDDTRDGP
Prymnesium parvum CCAP946/1B  TSP.DWPVNDAAYDVVLMSYISGSVPESIIGALYANAYKALKPGGRLLVHDFVVDNSLDGP
Emiliania huxleyi CCMP1516    TSP.DWPVPDGEYDLVLMSYISGSVPEPVVGALYANAYKALKPGGRLLVHDFMVDDSLDGP
                                                ▲
```
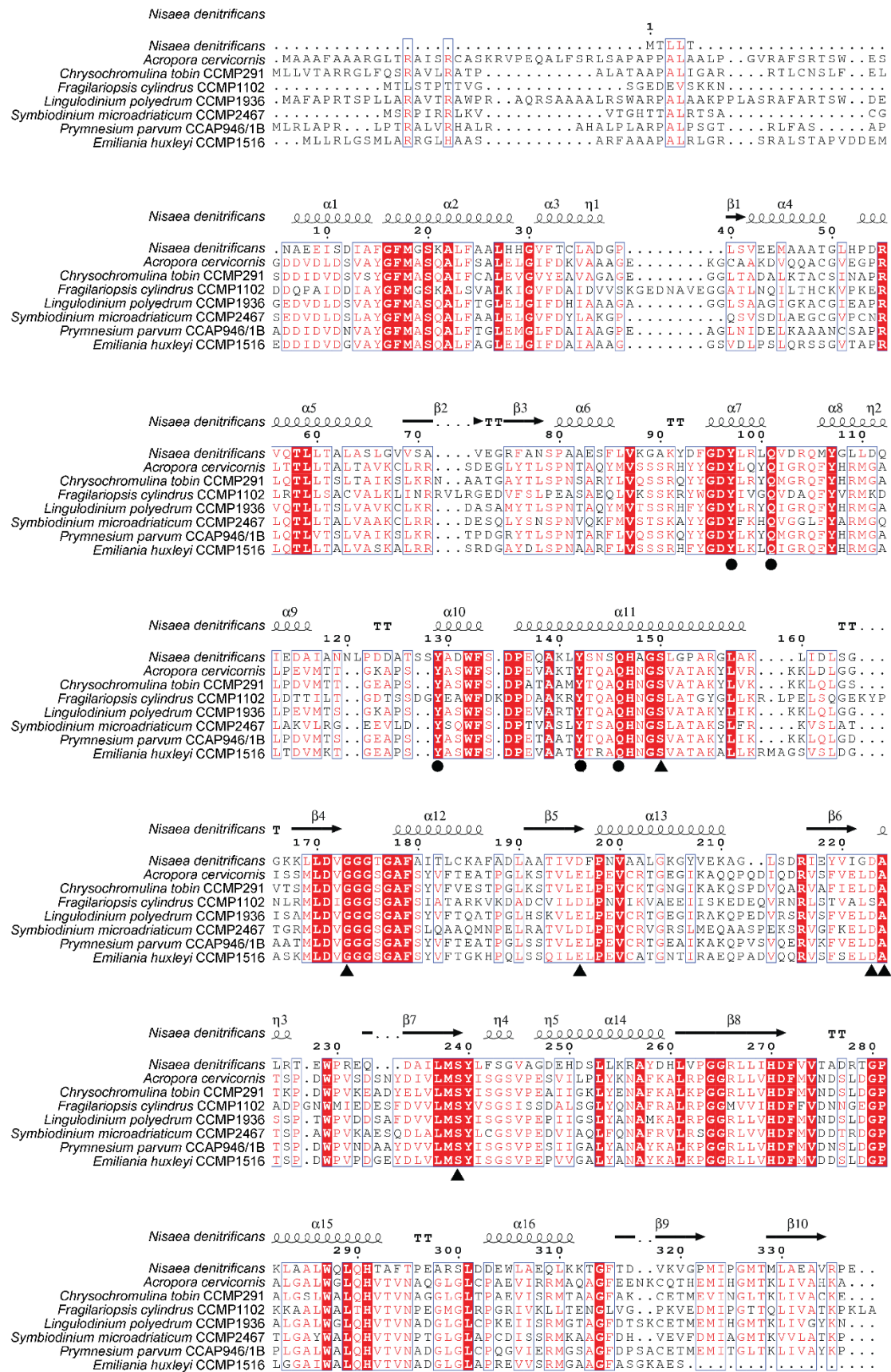
```
                              α15        TT   α16            β9       β10
Nisaea denitrificans          KLAALWQLQHTAFTPEARSLDDEWLAEQLKKTGFTD..VKVGPMIPGMTMLAEAVRPE..
Acropora cervicornis          ALGALWGLQHVTVNAQGLGLCPAEVIRRMAQACFEENKCQTHEMIHGMTKLIVAHKA...
Chrysochromulina tobin CCMP291 ALGSLWALQHVTVNAGGLGLTPAEVISRMTAACFAK..CETMEVINGLTKLIVACKF...
Fragilariopsis cylindrus CCMP1102 KKAALWALTHVTVNPEGMGLRPGRIVKLLTENCLVG..PKVEDMIPGTTQLIVATKPKLA
Lingulodinium polyedrum CCMP1936 ALGALWGLQHVTVNADGLGLCPKEIISRMGTACFDTSKCETMEMIHGMTKLIVGYKN...
Symbiodinium microadriaticum CCMP2467 TLGAYWALQHVTVNPTGLGLAPCDISSRMKAACFDH..VEVFDMIAGMTKVVLATKP...
Prymnesium parvum CCAP946/1B  PLGALWALQHVTVNADGLGLCPQGVIERMGSACFDPSACETMEMITGLTKLIVAYKP...
Emiliania huxleyi CCMP1516    LGGAIWALQHVTVNADGLGLAPREVVSRMGAACFASGKAES...............
                                     ●
```

Fig. S6. Sequence alignment of DsyB and eukaryotic DSYB proteins. Black dots indicate

1015   residues involved in MTHB binding, and black triangles indicate residues involved in SAM
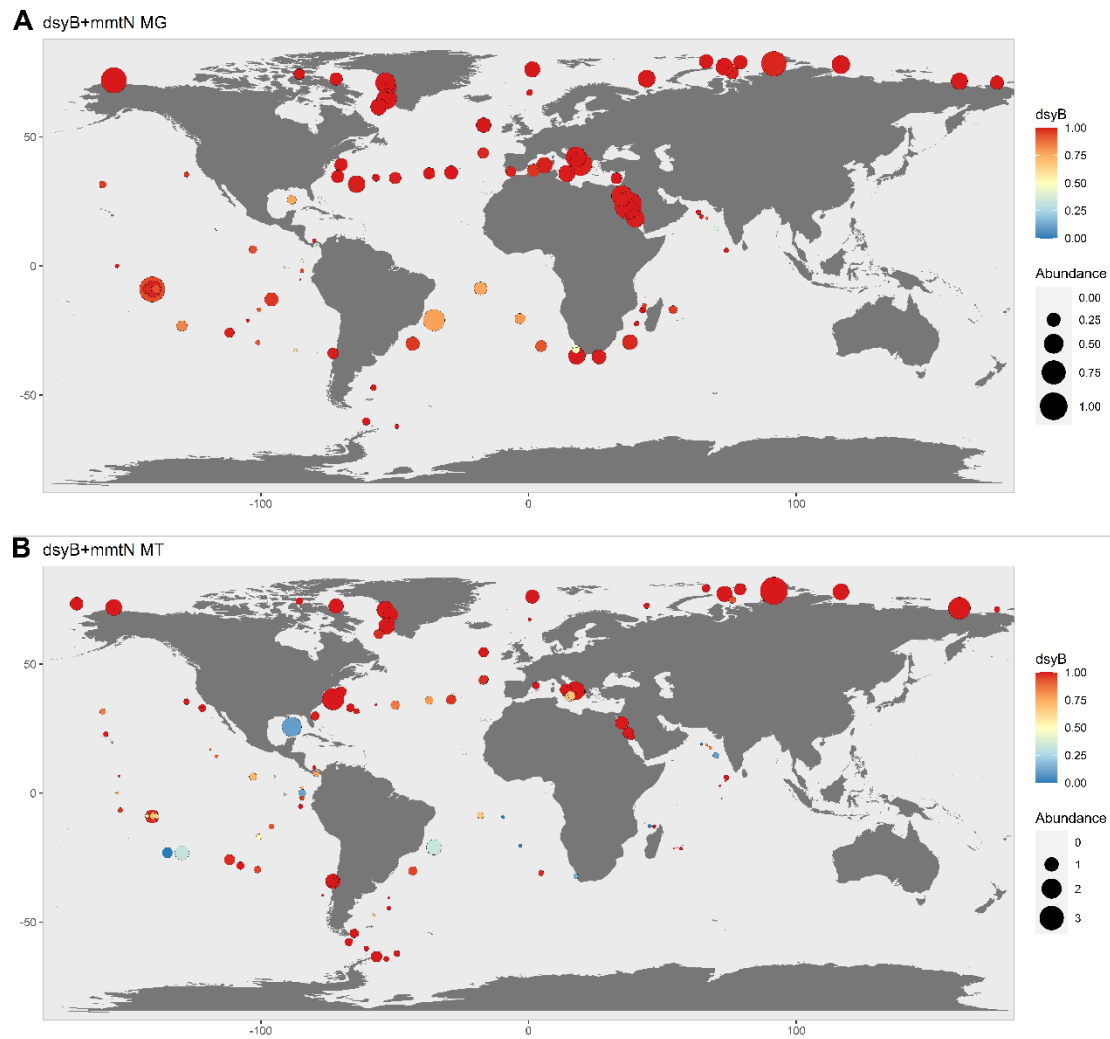
1016   binding.

1017

1018
1019
1020
1021 **Fig. S7. Bubble plot of global distribution of *dsyB/mmtN* in ocean metagenomes (A) and**
1022 **metatranscriptomes (B).** The sum of the average abundances of *dsyB* and *mmtN* across all
1023 depths as a percentage of the median of 10 single copy marker genes are shown by the bubbles.
1024 The colour bar indicates the relative abundance of *dsyB:mmtN* with red indicating 100% *dsyB*
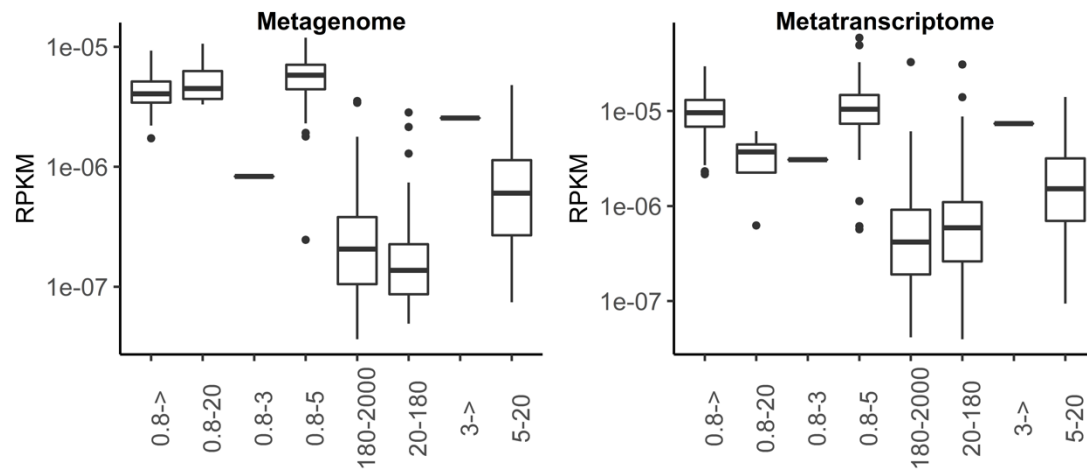1025 and blue indicating 100% *mmtN*.
1026

1027
1028 **Fig. S8. *DSYB* abundance (reads per kilobase per million mapped reads) by filter fraction**
1029 **size in the metagenome and metatranscriptome databases.**

1031
Table S1. Crystallographic data collection and refinement

Parameters	DsyB-SAM complex Se derivative	DsyB-SAM complex	DsyB-SAH-MTHB complex
Diffraction data			
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_1$
Unit cell			
a, b, c (Å)	75.6, 114.9, 152.6	76.5, 115.9, 153.1	88.1, 69.3, 104.3
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90. 0	90.0, 92.2, 90.0
Resolution range (Å)	50.0-2.7 (2.8-2.7) *	50.0-2.4 (2.49-2.40)	50.0-2.4 (2.49-2.40)
Redundancy	26.2 (25.3)	6.8 (7.0)	3.7 (3.7)
Completeness (%)	100.0 (100.0)	99.9 (100.0)	99.7 (99.7)
R_{merge}**	0.2 (0.4)	0.1 (0.5)	0.1 (0.5)
$I/\sigma I$	37.8 (9.9)	42.5 (6.9)	24.0 (4.1)
Refinement statistics			
R-factor		0.19	0.19
Free R-factor		0.25	0.27
RMSD from ideal geometry			
Bond lengths (Å)		0.008	0.008
Bond angles (°)		1.16	1.13
Ramachandran plot (%)			
Favoured		93.7	95.9
Allowed		6.2	4.1
Outliers		0.1	0
Overall B-factors (Å^2)		46.9	43.0

1032 *Numbers in parentheses refer to data in the highest resolution shell.

1033 ** $R_{merge}=\sum_{hkl}\sum_i|I(hkl)_i - <I(hkl)>|/\sum_{hkl}\sum_i I(hkl)_i$, where I is the observed intensity, $<I(hkl)>$ represents

1034 the average intensity, and $I(hkl)_i$ represents the observed intensity of each unique reflection.

1035

1036 **Table S5. Strains and Plasmids used in this study.**

1037

Strain/Plasmid	Description	Reference
Escherichia coli BL21 (DE3)	Strain used for DsyB expression for protein purification	New England BioLabs.
Rhizobium leguminosarium J391	Streptomycin resistant derivative of wild type strain 3841 used for the expression of genes cloned in plasmid pLMB509	Young et al. (2006)
Labrenzia aggregata J571	*Labrenzia aggregata* LZB033 with mutation in *dsyB* gene.	Curson et al. (2017)
Nisea denitrificans DR41_21	Wild type strain DSM 18348	DSMZ, Leibniz Institute, Germany.
pLMB509	Plasmid vector for taurine inducible expression of cloned genes in J391 and J571.	Tett et al. (2012)
pRK2013	helper plasmid used in triparental mating	Figuski and Helinski. (1979)
SNP-1304	*Nisaea denitrificans* DR41_21 *dsyB* gene coned into pLMB509	this study
SNP-1305	derivative of SNP-1304 with Y97A mutation	this study
SNP-1306	derivative of SNP-1304 with Q101A mutation	this study
SNP-1307	derivative of SNP-1304 with Y129A mutation	this study
SNP-1308	derivative of SNP-1304 with Y142A mutation	this study
SNP-1309	derivative of SNP-1304 with Q146A mutation	this study
SNP-1310	derivative of SNP-1304 with H291A mutation	this study

1038
1039

Table S6. Primers used in this study.

Primers	Sequence (5'-3')	Purpose
DsyB-F	GGAATTCCATATGACGTTGCTGACAAACGCC	Amplification of the
DsyB-R	CCGCTCGAGCTCCGGCCGCACGGCCTCG	genomic *dsyB* gene
Y97A-F	AATACGATTTCGGCGACGCTCTGCGTCTGCAGGTGG	Construction of the
Y97A-R	CCACCTGCAGACGCAGAGCGTCGCCGAAATCGTATT	mutant Tyr97Ala
Y97F-F	CGAAATACGATTTCGGCGACTTTCTGCGTCTGC	Construction of the
Y97F-R	GCAGACGCAGAAAGTCGCCGAAATCGTATTTCG	mutant Tyr97Phe
Q101A-F	GACTATCTGCGTCTGGCGGTGGACCGGCAGAT	Construction of the
Q101A-R	ATCTGCCGGTCCACCGCCAGACGCAGATAGTC	mutant Gln101Ala
Q101A-F	CTATCTGCGTCTGGAGGTGGACCGGCA	Construction of the
Q101A-R	TGCCGGTCCACCTCCAGACGCAGATAG	mutant Gln101Glu
Y129A-F	GATGCCACCAGCTCCGCCGCCGACTGGTTCTC	Construction of the
Y129A-R	GAGAACCAGTCGGCGGCGGAGCTGGTGGCATC	mutant Tyr129Ala
Y129F-F	GATGCCACCAGCTCCTTCGCCGACTG	Construction of the
Y129F-R	CAGTCGGCGAAGGAGCTGGTGGCATC	mutant Tyr129Phe
Y142A-F	CGGAACAGGCAAAGCTCGCTTCCAACAGCCAGCATG	Construction of the
Y142A-R	CATGCTGGCTGTTGGAAGCGAGCTTTGCCTGTTCCG	mutant Tyr142Ala
Y142F-F	GGAACAGGCAAAGCTCTTTTCCAACAGCCAGCATG	Construction of the
Y142F-R	CATGCTGGCTGTTGGAAAAGAGCTTTGCCTGTTCC	mutant Tyr142Phe
Q146A-F	CTCTATTCCAACAGCGCGCATGCCGGCTCCCT	Construction of the
Q146A-R	AGGGAGCCGGCATGCGCGCTGTTGGAATAGAG	mutant Gln146Ala
Q146E-F	TCTATTCCAACAGCGAGCATGCCGGCTCC	Construction of the
Q146E-R	GGAGCCGGCATGCTCGCTGTTGGAATAGA	mutant Gln146Glu
H291A-F	CTGGCAGCTCCAGGCCACCGCCTTCACG	Construction of the
H291A-R	CGTGAAGGCGGTGGCCTGGAGCTGCCAG	mutant His291Ala
N.d dsyBF	GGGTCTAAGGCGTTATTTG	
N.d dsyBR	CAGTCGGCGTAGGAGC	
N.d gyrBF	CTATCACGAAAACACGCTC	For RT-qPCR
N.d gyrBF	GCTTCATACGCCTTGGA	
N.d recAF	AGAATGGCGGCACTTG	
N.d recAF	CTAGGCAGCGAGACTTTG	

1041

62