University of East Anglia

Synthesis and transport of DMSP and glycine betaine in model diatoms

by

Ana Bermejo Martinez

A thesis presented for the degree of Doctor of Philosophy

University of East Anglia

School of Biological Sciences

September 2019

Declaration

I, Ana Bermejo Martinez, declare that this thesis is a record of my own work and that I composed it. It has not been submitted in any previous application for a degree in any other university. All quotations have been distinguished by quotation marks and all sources of information are specifically acknowledged.

Abstract

Dimethylsulfoniopropionate (DMSP) and glycine betaine (GBT) are very abundant sulfurous and nitrogenous compatible solutes, respectively. Their catabolites play a key role in the environment being climate active compounds, dimethylsulfide (DMS) and trimethylamine (TMA) or methane. DMSP has many suggested roles in the cell including cryoprotection, oxidative stress relief, osmoprotection amongst others in the marine environments. GBT has been proven to be an effective osmolyte in all kingdoms of life. The methylthiohydroxybutyrate (MTHB) methyltransferase enzyme catalysing the key step in DMSP biosynthesis has recently been identified in bacteria (DsyB) and many phytoplankton (DSYB). However, the majority of diatoms with sequenced genomes lack this DSYB, despite using the same pathway for DMSP biosynthesis. Likewise, very little is known about how diatoms and phytoplankton synthesise GBT despite the biosynthetic pathways being well understood. Nor it is known the mechanism diatoms use to transport these molecules. A novel methyltransferase involved in DMSP synthesis was found in the model diatoms *T. pseudonana* and *P. tricornutum* named DSYD (DMSHB synthase in diatoms). DSYD was present in other diatoms, brown and green algae and a gamma-proteobacterium. This is the first gene identified for DMSP synthesis in alga. Furthermore, a functional methyltransferase able to methylate glycine to glycine betaine named GSDMT (glycine sarcosine dimethylglycine methyltransferase) was found in these two diatoms and many bacteria, highlighting the prevalence of this pathway previously thought to be less common than the choline pathway. In addition, a novel enzyme present in bacteria was able of synthesising both DMSP and GBT. These results stress the close relationship between these two abundant osmolytes and show that the contribution of bacterial DMSP production has been underestimated.

Acknowledgments

I'd like to thank Professor Jonathan D. Todd, for everything I've learnt from him, for his support in the toughest times, for being one of the most brilliant and inspiring minds and for letting me be part of the Lab 1.29's family all these years. I would also like to thank Dr Ben Miller for his supervision and guidance, crucial to keep me sane during my CRISPR odyssey. All my appreciation to past and present members of the Lab 1.29, Murrel's lab, Chen's lab and other excellent collaborators. And, of course, thanks to NRP-DTP BBSRC for the funding.

To my dearest friends, family and cat - not even 100,000 words would be enough to tell you how important you are to me, and how much this thesis would not have been possible without you. You'll always have my love and my gratitude – **Graciñas**.

Contents

List of figures

List of tables

List of appendices

Abbreviations

˚C	Degree Celsius
µM	Micromolar
µm³	Cubic Micrometer
1H NMR	Proton Nuclear Magnetic Resonance
ABC	ATP Binding Cassette
BCAT4	Branched-Chain Aminotransferase 4
BCCT	Betaine Choline Carnitine Transport
BLAST	Basic Local Alignment Search Tool
C	Carbon
CCN	Cloud Condensation Nuclei
cDNA	Complementary DNA
CH_4	Methane
CLAW	Robert **C**harlson, James **L**ovelock, Meinrat **A**ndreae and Stephen **W**atson hypothesis
CO_2	Carbon Dioxyde
DMS	Dimethylsulfide
DMSA	Dimethylsulfonioacetate
DMSHB	4-Dimethylsulfonio-2-Hydroxybutyrate
DMSP	Dimethylsulfoniopropionate
DMSP-ald	3-(dimethylsulfonio)propionaldehyde
DMSP-amin	3-(dimethylsulfonio)propylamine
DMT	Dimethylglycine
EDTA	Ethylendiametetraacetic Acid
Fe	Iron
Fm	Maximum Fluorescence
fpkm	Fragments Per Kilobase Of Transcript Per Million Mapped Reads
Fv	Variable Fluorescence
GBT	Glycine Betaine
GC	Gas Chromatography
GC/MS	Gas Chromatography Mass Spectrometry
Gly	Glycine
h	Hour
H_2O_2	Hydrogen Peroxyde
IPTG	Isopropyl B-D-1-Thiogalactopyranoside
JGI	Joint Genome Institute
L	Litre
LB	Luria Broth
LC/MS	Liquid Chromatography Mass Spectrometry

LFQ	Label-Free Quantitation
m/z	Mass to Charge Ratio
MBM	Marine Basal Medium
Met	Methionine
MFS	Major Facilitator Superfamily
mg	Miligram
MGAT	Methionine:Glyoxylate Aminotransferase
min	Minute
ml	Millilitre
mM	Millimolar
MTHB	4-Methylthio-2-Hydroxybutyrate
MTOB	4-Methylthio-2-Oxobutyrate
N	Nitrogen
$NaNO_3$	Sodium Nitrate
NCBI	National Center For Biotechnology Information
NH_4CL	Ammonium Chloride
nmol	Nanomol
NMR	Nuclear Magnetic Resonance
NO_3^-	Nitrate
PCR	Plymerase Chain Reaction
pg	Picogram
PLP	Pyridoxal Phosphate (PLP)
PLPDE_IV	Pyridoxal 5'-Phosphate Dependent Enzyme Class IV (PLPDE_IV)
Pmf	Proton Motive Force
PO_4^{3-}	Phosphate
PSU	Practical Salinity Unit
Pt	*Phaeodactylum tricornutum*
qPCR	Quantitative PCR
Rbs	Ribosome Binding Site
RT-qPCR	Reverse Transcriptase Quantitative PCR
SAH	S-Adenosyl-L-Homocysteine
SAM	S-Adenosyl-L-Methionine
SDS	Sodium Dodecyl Sulfate
Si_4^+	Silicate
Smf	Sodium Motive Force
SMM	S-Methylmethionine
SO_4^{2-}	Sulfate
TCA	Tricarboxylic Acid Cycle
TMH	Transmembrane Helices

TMA	Trimethylamine
TMS	Transmembrane Helical Spanner
To	*Thalassiosira oceanica*
Tp	*Thalassiosira pseudonana*
Tris-HCL	Tris(Hydroxymethyl)Aminomethane Hydrochloride
TY	Triptone Yeast
UV	Ultraviolet
V	Voltage
vol	Volume

Chapter 1

Introduction

1 Introduction

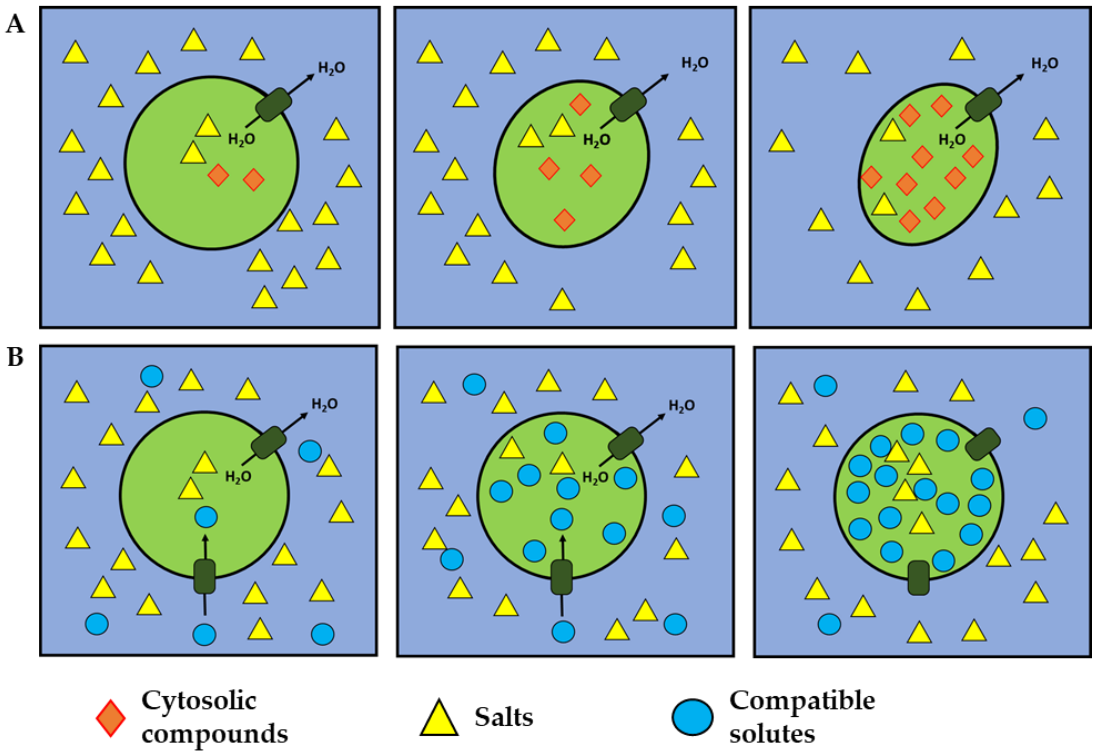
Microorganisms living in marine or saline environments need to cope with the osmotic stress to maintain cell turgor[1,2]. The cytoplasmic membranes of the microorganisms are permeable to water but not to other compounds, leading to water loss and accumulation of cytosolic compounds to balance the osmotic difference[3]. Without mechanisms to cope with the loss of turgor, the normal functioning of the cell and the structure of the proteins would be affected, causing ultimately the plasmolysis of the cell[4,5](Figure 1.1). Two different strategies are used by marine microorganisms to survive in conditions of increased salinity[6,7], an accumulation of inorganic ions and an accumulation of low-molecular-weight organic compatible solutes.

Metabolic function is normally disrupted by high concentration of inorganic ions. For this reason, the prokaryotic strategy involving the accumulation of inorganic ions requires of a specialised adaptation of the proteins and organelles to avoid damaging[6]. This strategy is restricted to a few halophilic microorganisms such as the extremely halophilic *Archaea* of the family *Halobacteriaceae*, the halophilic *Bacteria* of the order Haloanaerobiales, and the extremely halophilic red bacteria *Salinibacter ruber*[7]. These microorganisms accumulate the relatively scarce K^+, and Na^+ as major cations and their enzymes are enriched in acidic amino acids[8].

The second strategy relies on the accumulation of low-molecular-weight organic compounds. These solutes are very soluble molecules, non-charged or neutrally charged (zwitterionic). They are named compatible solutes[1] (also known as "osmolytes") because they can be accumulated in the cytoplasm to high concentrations (in some cases higher than 1 mol/kg water) without interfering with the metabolic processes within the cell[9] (Figure 1.1). Another remarkable characteristic is the ability of osmolytes to stabilize macromolecules. There are many theories to explain how the stabilization occurs, for example, via excluded volume and molecular crowding pressure effects[10], or by excluded solute-water

interactions[11,12]. A wide range of protective functions, along mitigating the osmotic stress, are derived from their stabilizing property[13]. Hence, compatible solutes can also confer thermoprotection, protection against freezing and draught tolerance[14].

Figure 1.1. Effects of high osmotic pressure in the cells. A) Without a coping mechanism, cells loose water and accumulate cytosolic compounds, leading to loss of turgor, disruption of normal cell functioning and, eventually, plasmolysis. B) Cells uptake and/or synthesis compatible solutes, osmotic pressure is balanced preventing water loss and maintaining the normal cell functioning.



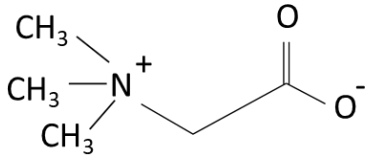
There are three types of compatible solutes, zwitterionic solutes such as glycine betaine (herein GBT), ectoine or dimethylsulphoniopropionate (herein DMSP); non-charged solutes such as carbohydrates, uncharged amino acids (e.g. proline), and peptides; and organic anions such as anionic polyols (e.g. glycerol), and β-glutamate[6,7,9,13,14]. These osmolytes are widely distributed amongst marine microorganisms, which usually can accumulate an array of them. The cocktail of solutes can be formed by either several osmolytes with the same net charge or a combination of anions and zwitterions[13]. In response to a change in osmotic pressure, marine microorganisms can uptake compatible solutes from the medium, as an initial

and quicker response, but they can also synthesise some of them. When the compatible solutes are no longer required, they are degraded or excreted to the medium. The transcript and transport systems are finely regulated by the osmolyte concentration outside and inside the cell[15] to control the pool of compatible solutes in the cytosol[3,16].

The two major osmolytes in marine environments and the subjects of this thesis are GBT and DMSP. These two compatible solutes are major sources of carbon, energy, and nitrogen or sulfur for phytoplankton and marine bacteria. Furthermore, they are the precursors of the two important climate affecting gasses methane and dimethylsulfide (DMS)[14].

1.1 Glycine betaine

Figure 1.2. Glycine betaine molecular structure.



GBT (Figure 1.2) is a zwitterionic quaternary ammonium compound ($C_5H_{11}NO_2$) derived from the methylation of the amino acid glycine. Indeed, the methyl groups in the amino nitrogen are responsible for the effective osmoprotection ability of GBT[5]. It is the most widespread osmolyte in nature being present in all kingdoms of life[6], and widely distributed in marine environments[2,6,17]. Moreover, GBT is the main compatible solute in halophilic eubacteria[18] living in complex microbial ecosystems where they can uptake it from the medium[19]. It is also the main osmolyte in halophilic photosynthetic organisms, including marine algae and microalga[20], cyanobacteria[21], and coastal halophilic higher plants[22]. This nitrogen compound is being synthesised and released to the environment by excretion or cell lysis. Subsequently, GBT can be taken up by marine microorganisms from the environment conferring them osmoprotection[23]. Despite GBT being so abundant, little is known about its role and

regulation in marine environments, where N is thought to be the major limiting nutrient[24,25].

1.1.1 Glycine betaine transport

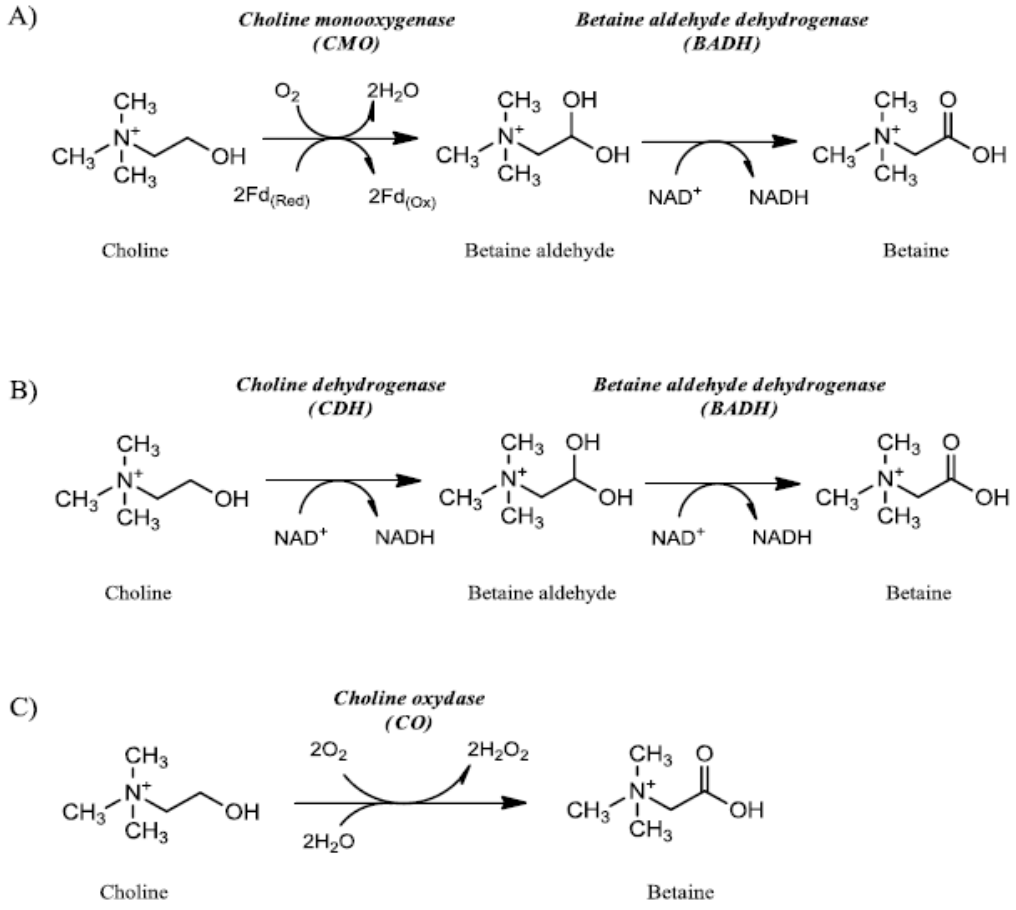
Marine microorganisms have high affinity systems to uptake GBT[23,26] and the precursor choline[27] from the medium. Inside the cell, GBT can be used as carbon, nitrogen and energy source as well as a compatible solute[14,28]. There are multiple transport systems involved in the uptake of GBT. Amongst all the transport systems, the main transporter is the physiological and environmentally relevant ProU[23]. ProU has affinity for GBT and for multiple structurally related substrates such as proline[29] and DMSP. This transport belongs to the family ATP-binding cassette (ABC) uptake system or traffic ATPases[29,30] and it is present in Prokaryotes[29,31,32] and Eukaryotes[33].

1.1.2 Glycine betaine synthesis

In addition to the uptake, organisms can also synthesise GBT *de novo.* There are three different biochemical pathways for GBT synthesis. In the first pathway choline, a compound derived from the methylation of ethanolamine and its derivatives[34], is transformed into GBT. In the second pathway, the amino acid glycine is converted to GBT through a series of methylations[35]. The last pathway comprises the deacetylation of carnitine, however the enzymes involved in this pathway have not been characterised yet[36], so no further description will be developed in this thesis.

Most organisms including most bacteria[16,28,37–39], yeast, plants[17] and animals[40,41] synthesise GBT via the choline pathway[42,43] (Figure 1.3). Choline can be oxidized in two steps to glycine betaine, with betaine aldehyde as the intermediate. In plants, choline is oxidised to betaine aldehyde by a Rieske-type soluble iron-sulfur protein termed choline monooxygenase (CMO); then, betaine aldehyde dehydrogenase (BADH), an NAD+-dependent enzyme, oxidises betaine aldehyde to GBT. Both CMO and BADH are located in the stroma of the chloroplast and are inducible by osmotic stress[43]. In animals and several bacterial species there are two different enzymes

Figure 1.3. Choline-GBT pathway. Three different set of enzymes are involved in this pathway. In plants (A), choline monooxygenase (CMO) oxidises choline to betaine aldehyde, using a federroxin as cofactor, and a NAD+-dependent betaine aldehyde dehydrogenase (BADH) converts betaine aldehyde into GBT. In *Escherichia coli* (B) the first step is carried out by NAD+-dependent choline dehydrogenase, and the second step by a similar BADH to the plants. The choline oxidase of *Athotrobacter globiformis* (C) carries out the whole reaction from choline to GBT.



capable of carrying out the first oxidation step, a membrane-bound choline dehydrogenase (CDH)[37,40], and a soluble choline oxidase (CO)[28,40]. CDH catalyses a four electron oxidation of chosline to betaine aldehyde, using molecular oxygen as electron acceptor[44]. Apparently, CDH does not require soluble cofactors[42,44]. In some bacteria CDH and CO can oxidise both choline to betaine aldehyde and this to GBT[16]. Nevertheless, the second step is, generally, performed by a BADH similar to that characterised for plants[45–47]. GBT biosynthetic pathway from choline requires the

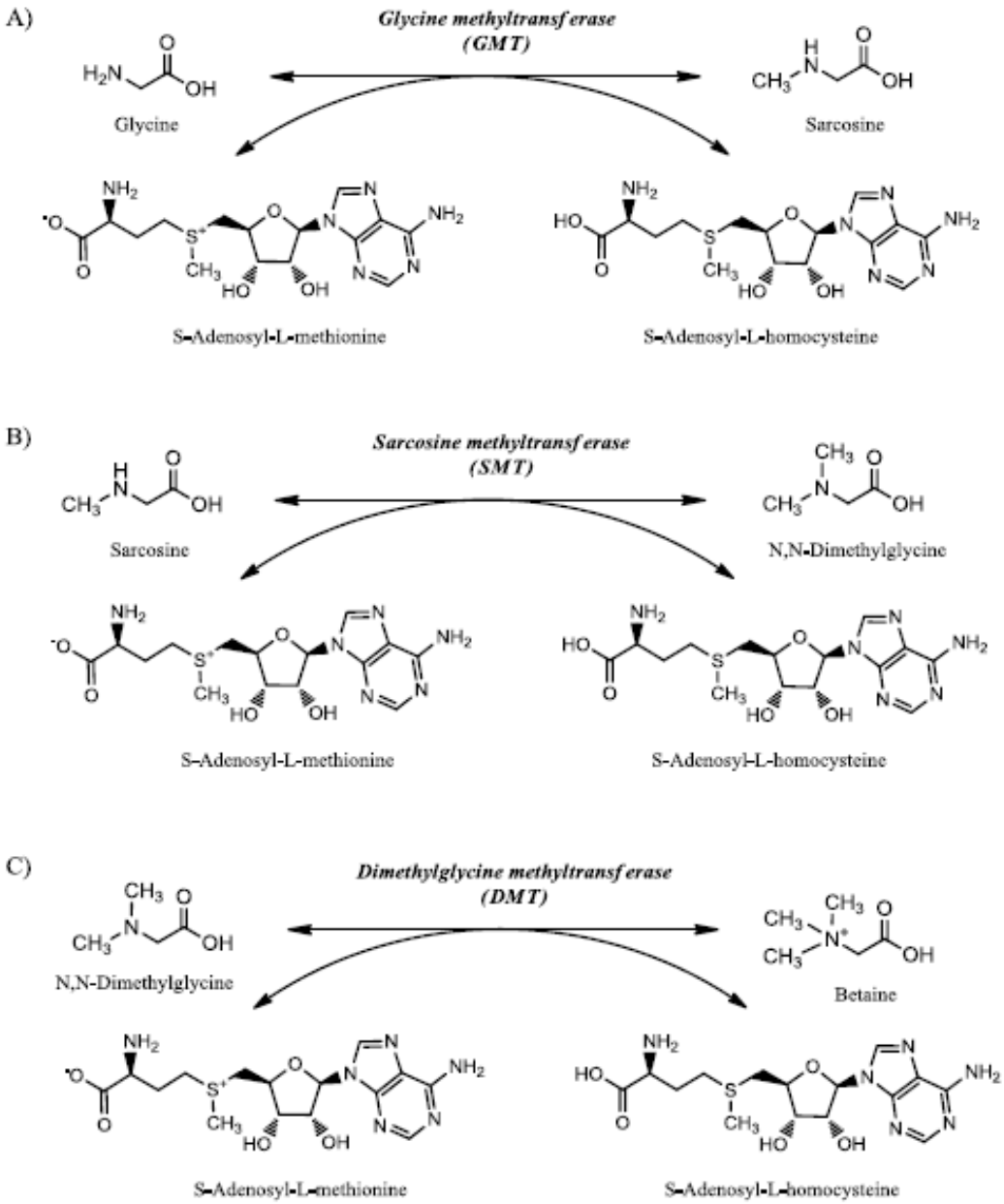
presence of the precursors, choline or the toxic betaine aldehyde intermediate, in the growth medium to occur. These substances are transported into the cells by osmotically regulated proteins[16].

The methylation pathway is found in a number of halotolerant or anoxigenic phototrophic bacteria, including cyanobacteria; in some methanogens and halophilic archaea; in *Actinopolyspora halophila* within the aerobic heterotrophic bacteria[48–51], and in the extremophilic red alga *Galderia sulphuraria*[50,52]. In this alternative pathway, glycine sarcosine dimethylglycine methyltransferase (GSDMT) pathway[35,53], the amine glycine, a simple carbon source, is converted into GBT through a three-step methylation in the N position of glycine[48,51]. In the first step, glycine is methylated to sarcosine (*N*-monomethylglycine); in the second step, sarcosine is methylated to N,N–dimethylglycine; and in the third step, N,N–dimethylglycine is methylated to GBT. The methyl donor is S-adenosylmethionine (SAM) producing S-adenosylhomocysteine[35,48–50] (Figure 1.4). Glycine and SAM are compounds with key roles in the cells. Glycine is synthesised from precursors of glycolysis and SAM, a compound which requires high energy levels to be synthesised, participates in several metabolic reactions, such as the synthesis of methionine, phosphatidylcholine or the modification of amino acids and DNA[51].

To date, two classes of enzymes involved in the GSDMT pathway have been characterised. In the first class, two enzymes are responsible for catalysing the reactions, glycine sarcosine methyltransferase (GSMT) and sarcosine dimethyltransferase (SDMT). GSMT catalyses the methylation from glycine or sarcosine to sarcosine or dimethylglycine, respectively. A similar enzyme with glycine N-methyltransferase activity is also present in mammalian cells, but, in this case, sarcosine is not further methylated to GBT[51,54]. The SDMT methylates sarcosine to dimethylglycine, as well as dimethylglycine to GBT[35,48–50]. Despite the partial substrate overlap between these two enzymes, it was shown that they only have specificity for one of the intermediates[50].

The second enzyme class, present in organisms such as *A. halophila*[54] and *Methanohalophilus portucalensis*[35,55], is formed by a single multi-domain protein (GSDMT). This enzyme has the ability to methylate all the intermediates, therefore, carrying out the entire reaction from glycine to GBT. In comparison, despite sharing

Figure 1.4. GSDMT pathway. GBT biosynthesis from glycine involved three methylation steps. SAM is used as methyl donor. Firstly, an enzyme with glycine methyltransferase activity (GMT) methylates glycine to sarcosine (A). Sarcosine is, in turn, methylated to N,N-Dimethylglycine by the sarcosine methyltransferase (SMT) activity of the enzyme (B). Finally, GBT is produced by the Dimethylglycine methyltransferase activity of the enzyme (C).



the same metabolic activity, GSMT/SDMT seems to have higher catalytic efficiency than the single protein GSDMT[35].

GBT biosynthesis is well characterised in most organisms including those marine dwelling. Nonetheless, very little is known about the GBT synthesis pathway(s) existing in eukaryotic phytoplankton, despite these organisms being key GBT producers[56].

1.1.3 Catabolism

As previously mentioned, microorganisms can degrade GBT. The aerobic catabolic pathway of GBT was first described in *Rhizobium meliloti* and it involves a three-step demethylation reaction, from GBT to glycine, with dimethylglycine (DMG) and sarcosine as intermediates[28]. GBT is demethylated to dimethylglycine by a betaine homocysteine methyltransferase, encoded by *gbcAB* in *Pseudomonas*[28,57,58]. In the second step, the demethylation of dimethylglycine to sarcosine, is carried out by a DMG oxidase similar to eukaryotic dehydrogenase[59] encoded in *Pseudomonas* by the operon *dgcAB*[58]. The operon *soxBDAG* corresponds to a sarcosine oxidase which demethylates sarcosine to glycine[59]. A transcription factor has also been described in *Pseudomonas*, GbdR, which regulates the GBT catabolic genes in response to the concentrations of GBT and dimethylglycine in the cell[58].

In anaerobic conditions the degradation of GBT occurs via an alternative pathway yielding trimethylamine (TMA). The Stickland-type oxidation-reduction reaction forming trimethylamine and acetate from GBT was first shown in *Chlostridium sporogenes*[60]. This alternative pathway occurs in anoxic sediments of hypersaline environments, where GBT is abundant as an osmolyte[61], and is carried out by sulfate reducing bacteria[62]. Then, trimethylamine can be further metabolised to the climate affecting gas methane[63]. More recently, it has been found that a novel family of fermentative bacteria named *Candidatus* 'Betaina sedimentti' is the key actor of the degradation of GBT and it is a family of bacteria prevalent in many hyperosmotic

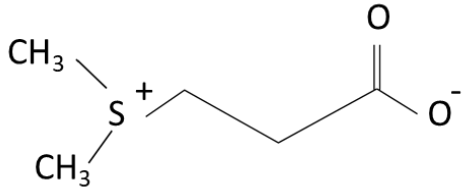
environments including saltmarshes[64]. Furthermore, GBT can also be a substrate for direct methanogenesis by the demethylation to N,N-dimethylglycine by some methanogenic Methanococcoides[65] (Figure 1.6).

Considering that the degradation of GBT and related quaternary amine molecules have been estimated to account for 90% of methane emissions in coastal marine sediments[66,67], it becomes of great importance to understand the dynamics of the production and degradation of GBT in these environments to further understand the production of such a potent greenhouse gas.

1.2 Dimethylsulfoniopropionate (DMSP)

DMSP (Figure 1.5) is a zwitterionic tertiary sulfonium compound ($C_5H_{10}O_2S$) first isolated from the red algae *Polysiphonia fastigiata*[68]. It is considered to be the predominant osmolyte in the euphotic zone of sea waters[69–71]. Similarly to GBT, this compatible solute can be released to the environment where it can be taken up by marine microorganisms[23]. This molecule has been shown to confer osmoprotection to

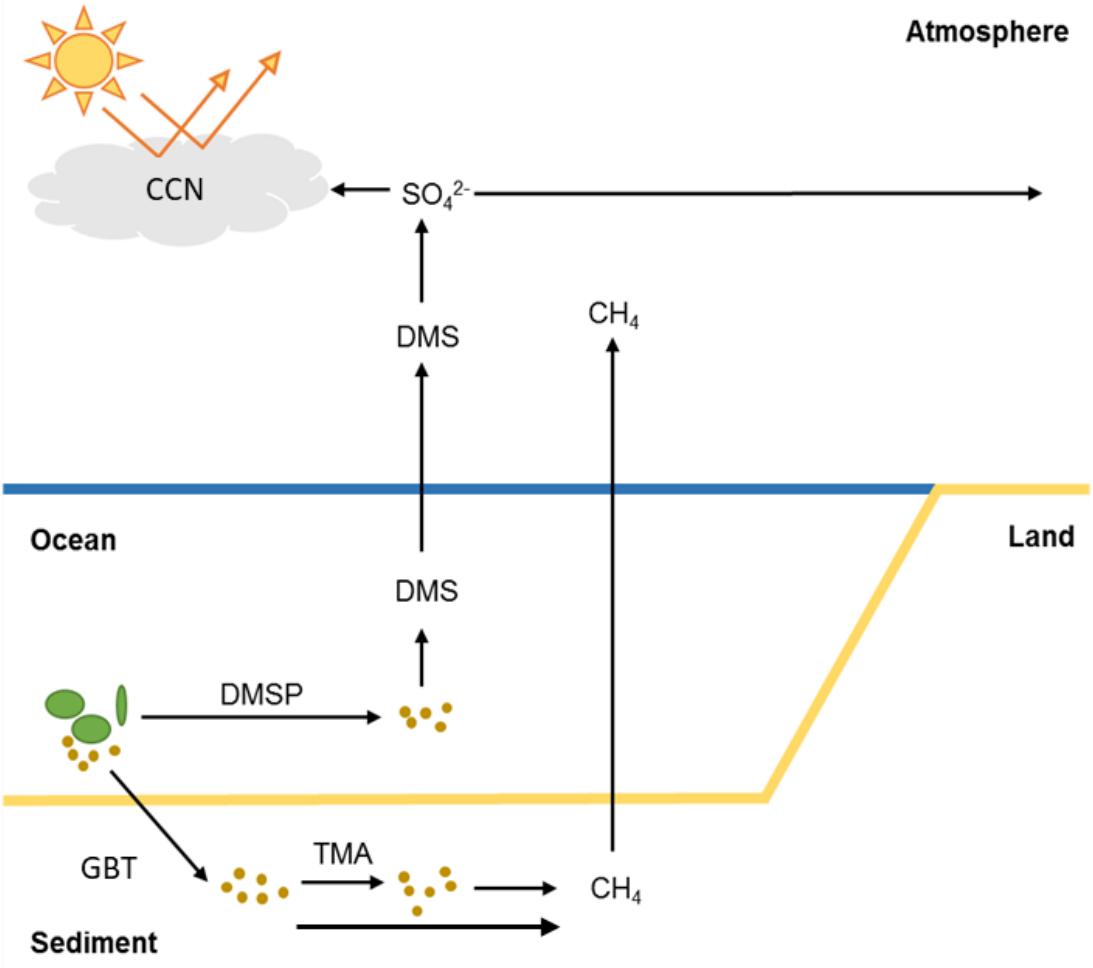
Figure 1.5. Dimethylsulfoniopropionate molecular structure.



bacteria[23,71,72], and is also predicted to have an osmoprotectant role in micro-algae, and nearshore plants[73,74]. Nevertheless, DMSP has a range of predicted functions in the marine environment besides protecting against osmotic stress[75]. Marine alphaproteobacteria can use DMSP as sulfur source, some of them even completely rely on it. Such is the case of *Candidatus 'Pelagibacter ubique'*, member of the widely distributed clade SAR11 which is deficient in assimilatory sulphate reduction genes[76]. DMSP has been shown to be an effective cryoprotectant[77] and antioxidant[78]. Importantly, DMSP is also the precursor of dimethylsulphide (DMS) a volatile compound key in the sulfur cycle. DMS is also hypothesised (CLAW hypothesis) to

have a climate cooling effect for being part of cloud condensation nuclei[79] (Figure 1.6), nonetheless, this hypothesis is under debate[80]. This DMSP cleavage product may also be an effective defence against grazing[81].

Figure 1.6. Effects of DMSP and GBT degradation in the environment. DMSP produced by marine alga and bacteria is released to the environment, some phytoplankton and bacteria degrade DMSP and release the gas dimethylsulphide (DMS) which enters the atmosphere. DMS reacts to form a sulfate aerosol that form cloud condensation nuclei (CCN), which causes a local cooling effect through albedo. Clouds travel inland and then biogenic sulfur returns to land through wet deposition. GBT produced by marine organisms in sediments can be degraded by bacteria to trimethylamine (TMA) and then to methane. Some methanogenic bacteria are capable of producing methane from GBT. Methane travels to the atmosphere where acts as a potent greenhouse gas.



1.2.1 DMSP transport

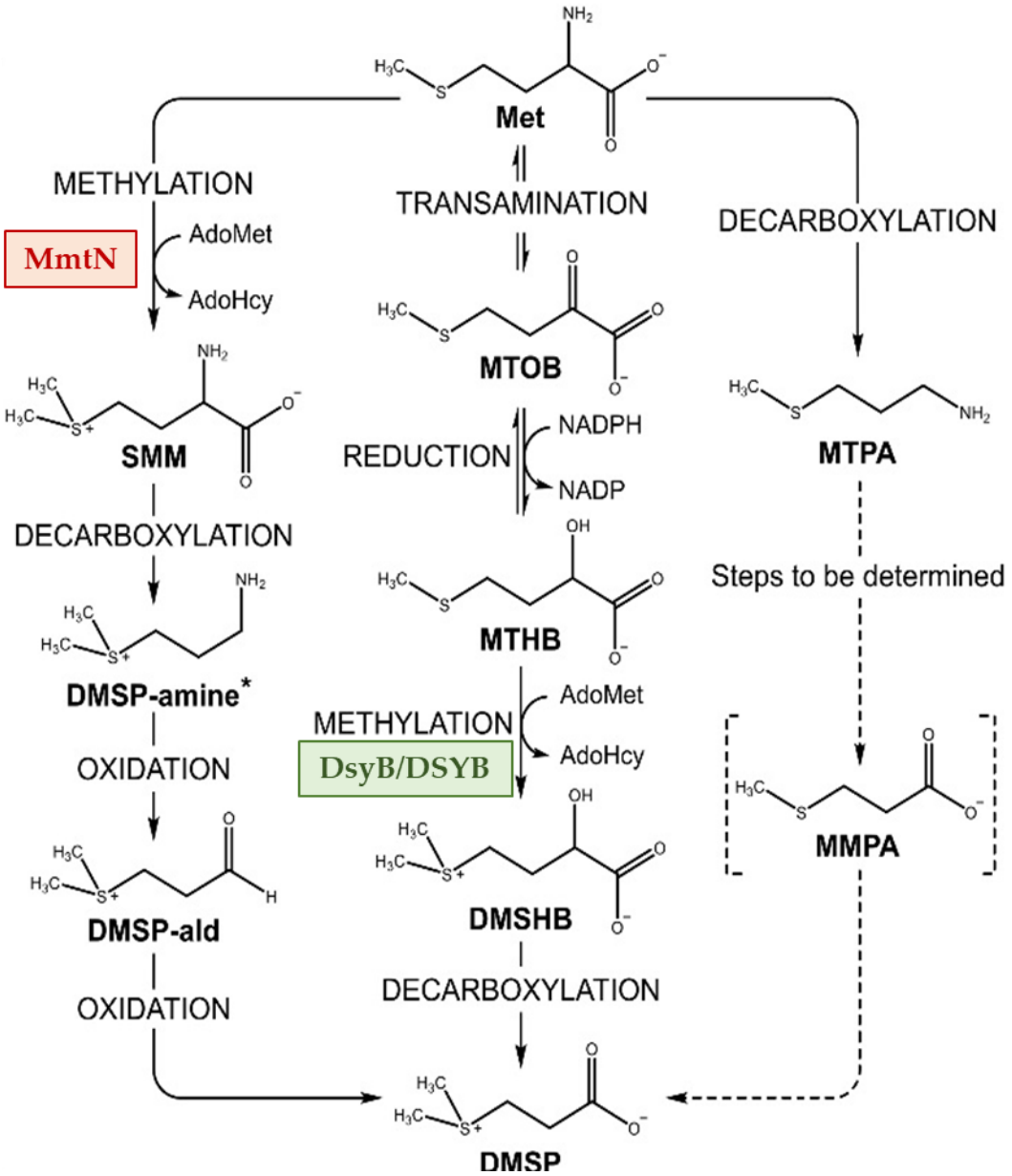
Bacteria[23,72], cyanobacteria and phytoplankton[26] use the same transport systems for the uptake of GBT and DMSP[82] from the environment[73]. The two main transporter systems are the ABC (ATP binding cassette) transport[83] previously mentioned in the GBT section, and the BCCT (betaine choline carnitine transport) proteins[26,71].

1.2.2 DMSP synthesis

Until very recently, it was thought that only eukaryotic organisms including coastal plants[84], green and red algae[68,85], phytoplankton (diatoms[86], coccolithophores[70] and dinoflagellates[87]), and corals[88] had the biosynthetic capacity to produce DMSP. However, Curson *et al.* demonstrated that marine bacteria are capable of synthesising this sulfur molecule *de novo*[89], and Williams *et al.* further showed that DMSP producing bacteria are important and abundant in coastal sediments[90]. Three biosynthetic pathways to DMSP have been described in plants, green algae and bacteria, and dinoflagellates (Figure 1.7).

The ability of the plants *Wollastonia biflora* (Compositae)[84] and *Spartina alterniflora* (Gramineae)[91] to synthesise DMSP *de novo* has been characterised. The initial precursor in both plants is L-methionine. L-methionine (Met) is methylated to S-methyl-L-methionine (SMM)[84,91] by the enzyme S-adenosyl-L-methionine:L-methionine S—methyltrasferase[92]. The central steps in the biosynthetic pathway of these two plants are different. The central step in *W. biflora* consists in a pyridoxal 5′-phosphate (PLP) dependent transamination-decarboxylation sequence obtaining 3-(dimethylsulfonio)propionaldehyde (DMSP-ald)[93]. In *S. alerniflora*, SMM is decarboxylated to 3-(dimethylsulfonio)propylamine (DMSP-amin) and further converted into DMSP-ald by oxidative deamination[91,94]. The last step, common for

Figure 1.7. DMSP biosynthetic pathways. The three pathways, named after their first step, start with L-methionine. The methylation pathway is found in higher plants and bacteria. In bacteria, the methylation of L-Met to SMM is carried out by MmtN. The transamination pathway is the most widely distributed, utilised by macroalgae, phytoplankton, corals and bacteria. In eukaryotes, the committed step of the MTHB methylation, is carried out by DSYB, and in prokaryotes by DsyB. The decarboxylation pathway has only been found in one dinoflagellate (Adapted from Curson *et al.*, 2017).



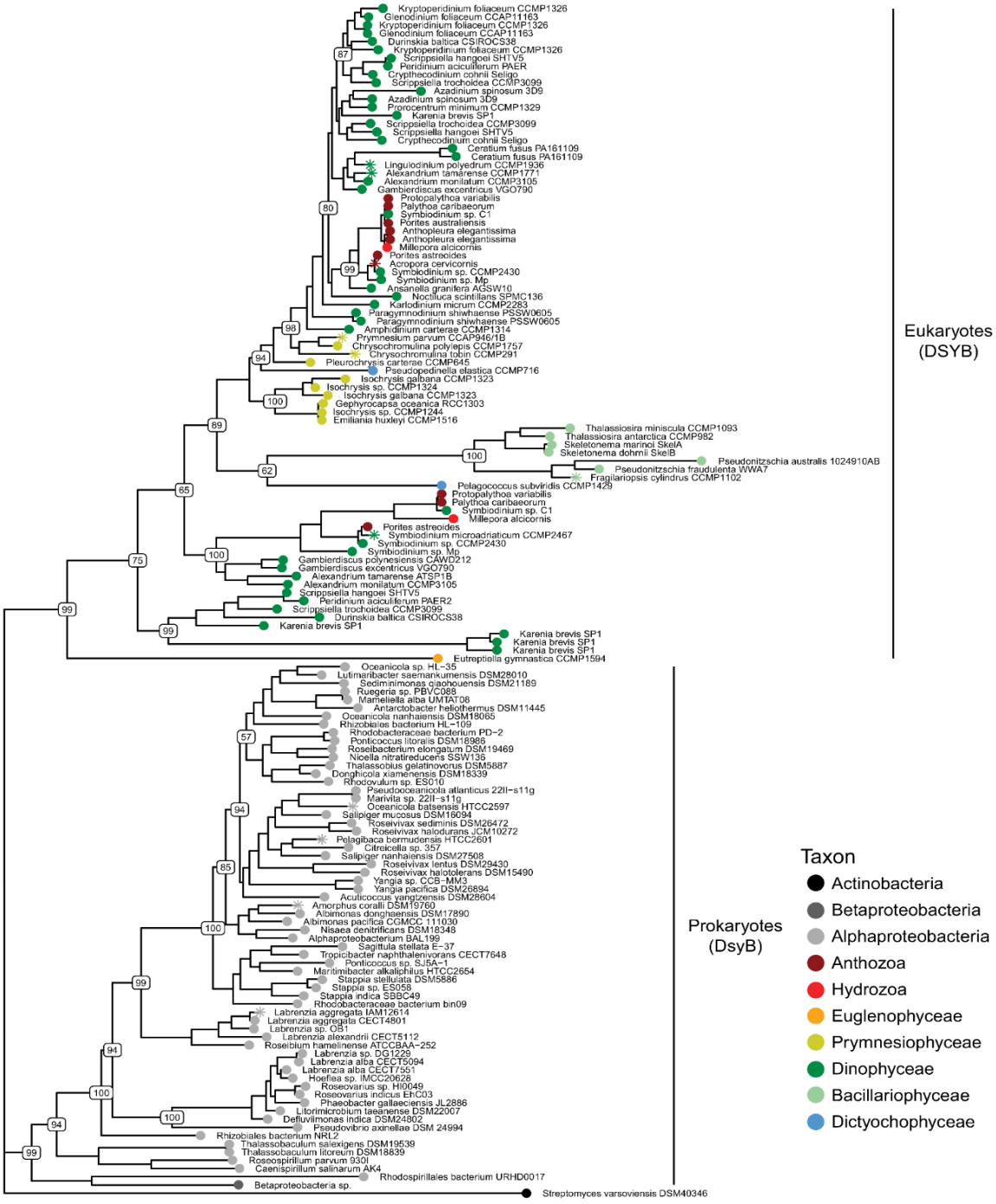
both plants, comprises the oxidation of DMSP-ald to DMSP[91,92]. Moreover, Williams *et al.* discovered that some DMSP producing bacteria, including alphaproteobacteria,

gammaproteobacteria and actinobacteria, are capable of synthesising SMM from L-Met, previously thought to be an ability reserved to plants[90]. Using *Novosphingobium* sp. BW1 as a model organism, the methionine methyltransferase MmtN was identified and proved to be responsible for the conversion of L-Met to SMM in bacteria synthesising DMSP from the methylation pathway[90] (Figure 1.7).

Another DMSP synthesis pathway, only found in one dinoflagellate, also starts with Met, however, different intermediates are involved. Firstly, Met is decarboxilated to 3-(methylthio)propylamine (MTPA) by L-methionine decarboxylase (Figure 1.7). The remaining steps of the pathway are yet to be determined[95].

In macro algae, DMSP is also derived from the amino acid L-methionine[85]. Studies in *Ulva intestinalis* showed that Met is transformed to 4-(methylthio)-2-oxobutanoic acid (MTOB)[96] by 2-oxoglutarate-dependent[97] transamination, then MTOB is reduced to 2-hydroxy-4-(methylthio)butanoic acid (MTHB)[96] using NAD(P)H. MTHB is S-methylated[96], with S-Adenosyl-L-methionine as the methyl group donor (SAM)[97], giving 4-(dimethylsulfonio)-2-hydroxybutanoate (DMSHB), which is subsequently turned into DMSP by oxidative decarboxylation[96]. The methylation from MTHB to DMSHB catalysed by the MTHB methyltransferase enzyme is specific to DMSP producers[97] (Figure 1.7). The same pathway is present in diatoms and coccolithophores[98], corals[88], and bacteria[89], making it the most widespread pathway of the DMSP biosynthetic pathways. Although candidate gene involved in the biosynthesis of DMSP were proposed from studies in the sea-ice diatom *Fragilariopsis cylindrus*[98], none of those genes were proven to be functional. Hence, the first characterised gene for DMSP synthesis proved to encode for a functional MTHB methyltransferase is *dsyB* from the bacterium *Labrenzia aggregata* LZB033[89]. Homologues of DsyB, an acetylserotonin O-methyltransferase, belonging to the family of S-adenosyl methionine-dependent methyltransferases, were found to be present in other alphaproteobacterial species[99] (Figure 1.8). A metagenomic study performed by Williams *et al.* also showed that *dsyB* genes were abundant in marine

Figure 1.8. Maximum likelihood tree of DsyB/DSYB proteins. A total of 145 amino acid sequences were identified by searches in the NCBI, JGI IMG and iMicrobe MMETSP databases. Taxonomic class of the species are colour-coded as indicated in the key. Tested functional proteins are displayed with an asterisk. Numbers mark bootstrap support for nodes (Curson, *et al.* 2018).



sediments[89]. Some of those DsyB homologues were proven to be functional[99]. Finding DsyB facilitated the identification of the functional MTHB-methyltransferase, DSYB,

in marine phytoplankton including *Chrysochromulina tobin* and *Prymnesium parvum*[99]. Homologues of this phytoplankton DMSHB synthase were also found in macroalgae, diatoms such as *Fragilariopsis cylindrus*, prymnesiophytes and prasinophytes[99] which are some of the greatest DMSP producing organisms[100] (Figure 1.8). An evolutionary study of eukaryotic and prokaryotic DSYB/DsyB suggests that DsyB was first originated in alphaproteobacteria and then transferred to eukaryotic organisms by an endosymbiosis event or by horizontal gene transfer (HGT)[99]. Not all phytoplankton produce DMSP, and those which produce the sulfur osmolyte do it at different concentrations. DSYB was shown to be a good indicator to whether an organism would be a major or a minor producer[99]. In addition, degenerate primers targeting DsyB and MmtN have been successfully used to estimate the contribution of bacterial DMSP production in microcosmos experiments[90]. Noteworthy, some DMSP producing phytoplankton lack *DSYB*, such as the model diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* and the genes involved in the synthesis of DMSP by these organisms are still poorly understood[99,101].

1.2.3 DMSP catabolism

The degradation of DMSP can occur via the demethylation pathway or via the cleavage pathway. Along the DMSP demethylation pathway the methyltransferase DmdA demethylates DMSP, using tetrahydrofolate as a methyl donor, yielding 3-methylmercaptopropionic acid[102]. The coenzyme A ligase DmdB is activated by ATP and it converts this free amino acid in coenzyme A ester, followed by its catalysis to methylthioacryloyl-CoA by FAD-dependent dehydrogenase DmdC. The enoyl-CoA hydratase DmdD adds water and the hemithioacetal liberates methanethiol (MeSH) giving the coenzyme A thioester of malonyl semialdehyde. The thioester is hydrolysed and it decarboxylates to acetaldehyde and carbon dioxide[103].

DMSP can also be cleaved by DMSP lyases yielding the climate regulating gas DMS. Prokaryotic DMSP lyases are very diverse, they belong to different polypeptide families, and they vary in size and amino acid sequences[104]. The enzymes DddL[105],

DddP[106], DddQ[107], DddW[71], and DddY[108] lyse DMSP to DMS and acrylate. The enzyme DddD is the only one able to catalyse the hydrolysis of DMSP to DMS and 3-hydroxypropionate[109]. Eukaryotic organisms can also lyase DMSP, the enzyme responsible for this reaction, Alma1, was recently described[110].

The CLAW hypothesis postulates that when DMS is released from the cleavage of DMSP to the environment, the volatile gas escapes to the atmosphere where it is oxidised forming SO_2, rapidly converted into non-sea-salt sulfate aerosols. These aerosol particles would act as cloud condensation nuclei (CNN). This type of CNN would then change the reflectance (albedo) of clouds, creating a local cooling effect, linking DMS fluxes originated from algal blooms with climate regulation[79]. However, this hypothesis relies on the non-sea-salt sulfate aerosols being the main source of CCN. In 2011, a new study showed that there are many other sources of CCN more relevant than the non-sea-salt sulfate aerosols, such as bubble bursting at the ocean surface[80]. Although this study does not disregard the possibility of a direct link between DSM production and CCN formation, it does question the importance of the proposed feedback loop between biogenic source of CNN, atmospheric chemistry, cloud physics and climate postulated by the CLAW hypothesis[80].

1.3 Relation between GBT and DMSP

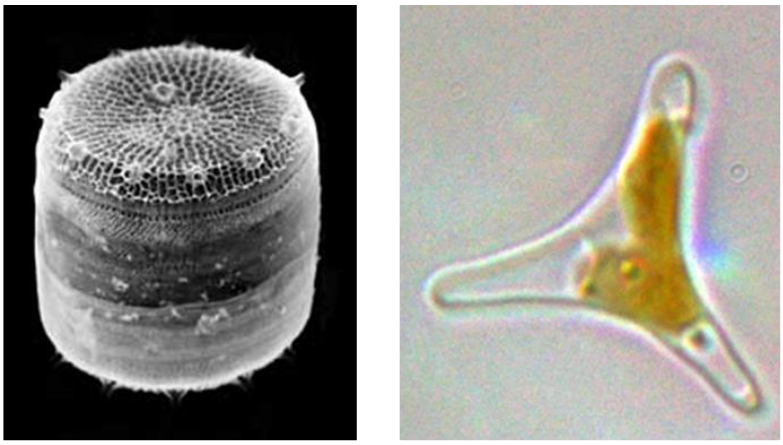
Andreae (1986) hypothesised a reciprocal relationship between GBT and DMSP due to the similar structure of these molecules. Thus, in the oceans, where nitrogen is limiting and there is an abundance of sulphate, synthesis and uptake of DMSP would be favoured over GBT, allowing the N to be used for the synthesis of other organic compounds[111]. By comparison, in high N conditions, GBT production would be higher than DMSP[112]. For instance, a study conducted in the model diatom *Thalassiosira pseudonana* revealed an inverse relationship between GBT and DMSP, GBT being preferably produced in nitrogen-rich conditions[112]. However, numerous studies showed no simple correlation between DMSP and GBT, although GBT synthesis appears to be correlated to N availability[56,112–114]. This relation is also

apparent in the shared transport system, as GBT inhibits DMSP uptake and DMSP inhibits GBT uptake showing an intimate link between the fates of these two compounds[39]. In addition, there is no direct evidence that GBT and DMSP function as osmolytes in the ubiquitous eukaryotic phytoplankton, despite this being most likely.

1.4 Model organisms

Diatoms are eukaryotic phytoplankton belonging to the division Heterokontophyta (Stramenopiles)[115]. Diatoms are aquatic microorganisms present in waters worldwide[116] representing the basis of the trophic pyramid in these ecosystems[115]. Their importance lies in their active role in several biogeochemical cycles such as

Figure 1.9. Microscopy image of *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*. *T. pseudonana's* photo was taken by Nils Kröger, Universität Regensburg, and *P. tricornutum* was taken from the Joint Genome Institute (JGI).



Thalassiosira pseudonana *Phaeodactylum tricornutum*

carbon (C), nitrogen (N), phosphorus (P), silicon (Si) and iron (Fe)[117], being responsible, for instance, for approximately 20% of photosynthesis on the planet[116] and around 40% of carbon fixation in the Ocean[117].

Thalassiosira pseudonana (Figure 1.9) is a diatom that belongs to the bi/multipolar centric paraphyletic group (Mediophyceae) due to its characteristic radially patterned valves. On the other hand, *Phaeodactylum tricornutum* (Figure 1.9) belongs

to the pennate group (Bacillariophyceae)[118], and it has a cell wall poor in silica and different morphotypes[119]. The cytology, physiology and biochemistry information of these two model diatoms is available[115,120], and the whole genome of nuclear, plastid and mitochondrial DNA of *T. pseudonana* was sequenced in 2004[121], and of *P. tricornutum* in 2008[122]. In addition, these model diatoms have a short generation time, are easy to culture and to maintain[115,120]. Furthermore, they are genetically tractable, which allows gene silencing and overexpression of existing or foreign genes[123–125], and mutations using CRISPR/Cas9 technique[126,127].

Most importantly, both diatoms are known DMSP and GBT producers, and the synthesis of these two important osmolytes are regulated by environmental conditions such as nitrogen scarcity[128,129]. Moreover, they lack the known DSYB gene present in other phytoplankton[99], and very little is known about how they synthesise GBT[130]. These features make *T. pseudonana* and *P. tricornutum* excellent models to study the genetics and regulation of the synthesis and transport of DMSP and GBT in DSYB-lacking diatoms.

1.5 Research gaps and aims of the thesis

The uptake and synthesis of compatible solutes are two commonly used coping mechanisms adopted by organisms to overcome stressors such as high osmotic pressure, low temperatures or oxidative stress[10]. The sulphurous molecule DMSP and the nitrogenous compound GBT are amongst the most abundant compatible solutes synthesised and imported by many marine organisms such as phytoplankton[99,113,131] and marine bacteria[71,89,90]. In addition, the degradation product of these DMSP and GBT play a key role in the environment being the climate active compounds dimethylsulphide (DMS) [79,80] and trimethylamine (TMA)[64] or methane[65], respectively. The most widespread pathway for DMSP synthesis is via the transamination pathway, present in alga[96], phytoplankton[99], corals[88] and bacteria[89]. DSYB (phytoplankton) and DsyB (bacteria) are the only known enzymes involved in this pathway and they carry out the methylation of MTHB to DMSHB, thought to be the

committing step[89,99]. However, the majority of diatoms with sequenced genomes lack this DSYB, despite using the same pathway for DMSP biosynthesis[99]. Likewise, GBT synthesis has been well characterised in many organisms[16,44,83,132] but how diatoms synthesise this compatible solute is not fully understood. Similarly, although transport of DMSP and GBT has been reported in phytoplankton[131] and the transport of these osmolytes is well known in bacteria[71], how phytoplankton transport these molecules remains unknown. Therefore, this thesis aims to:

1. Understand how diatoms transport DMSP and GBT.
2. Identify and characterise the gene responsible for the methylation of MTHB to DMSHB in DSYB- lacking diatoms.
3. Elucidate which GBT synthesis pathway diatoms utilize, and which genes are responsible for this process.
4. Study the regulation of the transport and synthesis of the two compatible solutes by selected environmental conditions.
5. Explore the distribution of homologues to the diatom DMSP and GBT synthesis enzymes in other organisms.

The ability of the DSYB-lacking model diatoms *T. pseudonana* and *P. tricornutum* to uptake[131] and to regulate the synthesis[98] of DMSP and GBT compatible solutes will be used to explore the genetics behind their transport and synthesis. First, physiological conditions modulating the production of these two osmolytes will be studied by analysing their metabolite production, transcriptomics and proteomics. Their genome and the regulated genes and proteins by the relevant conditions will be further investigated to identify candidate genes of interest. Molecular and phylogenetic studies will be performed to confirm functionality and distribution of any novel genes involved in the transport and synthesis of DMSP and GBT.

Chapter 2

Materials and methods

2 Materials and methods

2.1 Chemical synthesis

Dimethylsulfoniopropionate (DMSP) was synthesized from DMS (Sigma-Aldrich) and acrylic acid (Sigma Aldrich) as described in Todd *et al.*, 2010[133]. DL-DMSHB was chemically synthesised as described by Curson *et al.*, 2017[89]. Met, MTOB, MTHB, glycine, sarcosine, dimethylglycine, choline chloride and glycine betaine (GBT) are commercially available and were obtained from Sigma-Aldrich.

2.2 Media preparation and growth conditions

2.2.1 Media and general growth of diatoms and bacteria

P. tricornutum CCAP 1055/1 and *T. pseudonana* CCMP 1335 were grown in F/2[134] medium made with enriched seawater artificial water[135] at 35 practical salinity unit (PSU) unless otherwise stated. Both diatom cultures were grown at 22 °C with a light intensity of 120 μE $m^{-2}s^{-1}$ and a light/dark cycle of 16 h light/8 h dark. Where necessary, media for algal growth were modified according to the requirements of the experimental conditions being tested. Where strains were not already known to be axenic, cultures were treated with multiple rounds of antibiotic treatment prior to experiments. Antibiotics used were streptomycin (400 μg ml^{-1}), chloramphenicol (50 μg ml^{-1}), gentamicin (20 μg ml^{-1}) and ampicillin (100 μg ml^{-1}).

Escherichia coli was grown at 37 °C (unless otherwise stated) in Luria-Bertani[136] (LB) complete medium or in M9 salts minimal media[136] supplemented with 50% glycerol, 2 mM $MgSO_2$, 0.1 mM $CaCl_2$, 0.1 mM thiamine and 0.1 mM methionine. When indicated, the salinity of M9 salts minimal was increased by addition of 0.6 M NaCl. *Rhizobium leguminosarum* was grown at 28°C in tryptone yeast[137] (TY) complete medium or Y[137] minimal medium with 10 mM succinate as carbon source and 10 mM NH_4Cl as nitrogen source. *Gynuella sunshinyii* and *Mameliella alba* were grown in YTSS[138], or MBM[139] (Marine Basal Medium) 35 PSU (practical salinity units) unless

Table 2-1. List of strains used in this study.

Strain	Description	Reference
Phaeodactylum tricornutum CCAP 1055/1	Monoclonal culture derived from strain CCMP632 originally isolated in 1956 off Blackpool (U.K.)	De Martino *et al.*, 2007[140]
Thalassiosira pseudonana CCMP 1335	Clone collected in 1958 from Moriches Bay (Long Island, New York)	Hasle & Heimdal (1970)[141]
Gynuella sunshinyii	wild type, *dsy*D$^+$	Chung *et al.* 2015[142]
Mameliella alba	wild type, *gsdmt*/*dsy*D$^+$	Chen *et al.* 2018[143]
E coli BL21	Strain used for overexpression of cloned genes in pET and pUC vectors	Studier and Moffat (1986)[144]
E. coli 803	Strain used for routine transformations	Wood (1966)[145]
E. coli JM101	Strain for expression of lacZ gene in blue-white screen	Yanisch-Perron *et al.*, (1985)[146]
E. coli MKH13	*proP* and *proU* deficient strain	Haardt, M. *et al.* (1995)[147]
E. coli Rosetta	Strain to express eukaryotic proteins with codons seldom used in *E. coli*	Merk Millipore
Rhizobium leguminosarum	Streptomycin-resistant derivative of wild type strain 3841 expression of genes cloned in plasmid pLMB509	Young *et al.* (2006)[148]

otherwise stated, at 30°C. MBM was supplemented with 10 mM mixed carbon source from a 1 M stock of 200 mM succinate, glucose, pyruvate, sucrose and glycerol, and 0.5 or 10 mM NH_4Cl as nitrogen source. Where indicated, the salinity of MBM was adjusted by altering the amount of sea salts (Sigma-Aldrich) added, and nitrogen

levels were altered through the adjustment in volume of NH_4Cl added as the nitrogen source. GBT and DMSP and their pathway intermediates, were only added to M9 salts or Y medium where indicated in experiments that specifically addressed the effect of adding such compounds. Where necessary, antibiotics were added to media at the following concentrations: streptomycin (400 μg ml^{-1}), kanamycin (20 μg ml^{-1}), spectinomycin (200 μg ml^{-1}), gentamicin (20 μg ml^{-1}), ampicillin (100 μg ml^{-1}), rifampicin (400 μg ml^{-1}). Strains used in this study are listed in Table 2-1.

2.2.2 Growth of diatoms under non-standard conditions.

All diatom cultures described here were regularly monitored. All samples were taken in exponential growth phase unless otherwise specified. Standard growth conditions were a temperature of 22 °C, light intensity of 120 μE m^{-2} s^{-1}, salinity of 35 practical salinity units (PSU) and nitrogen concentration of 882 μM. For decreased salinity, the amount of salts added to the artificial seawater were adjusted to give a salinity of 1 or 5 PSU. For reduced nitrogen concentration cultures, the F/2 medium contained 30 μM, all samples were taken at the onset of stationary phase for nitrogen limited treatments. For changes in temperature, *T. pseudonana* CCMP 1335 and *P. tricornutum* CCAP 1055/1 cultures were grown at 22°C and then transferred to 8°C or 30°C and 12°C or 27°C, respectively. Samples were taken after 24 hours, 7 days and 14 days. Low CO_2 and Fe limitation experiments were performed under continuous light. For low CO_2 treatment, initial cultures were grown to a density of 5x105 cells ml^{-1}, cells were filtered and transferred to fresh media. Cultures were kept in DURAN® bottles 95% filled and sealed with PARAFILM®. Samples were taken before transferring the cells to the DURAN® bottle, after 5 days and after 10 days of incubation. For Fe limitation, standard grown cultures were transferred to fresh media containing no Fe. Samples were taken before transferring the cells to no Fe media, after 5 days and after 15 days of incubation. For H_2O_2 experiments, 0.25 mM, 0.75 mM or 2 mM of H_2O_2 were added to the cultures and samples were taken immediately before the addition of H_2O_2, after 0.5 hours and after 3 hours.

2.3 Diatom sampling methods

Diatom growth monitoring and sampling methods are the same as described by Curson *et al.*, 2018[99].To measure growth of diatom cultures, samples were removed and appropriately diluted in artificial seawater. Cell counting was done using a Multisizer 3 Coulter counter (Beckman Coulter). The effect of stress on photosystem II was determined by measuring Fv/Fm values using a Phyto-Pam phytoplankton analyser (Heinz Walz, Germany). Samples for metabolite detection by GC or LC-MS were obtained by filtering 25 ml of culture onto 47 mm GF/F glass microfiber filters (Fisher Scientific, UK) using a Welch WOB-L 2534 vacuum pump. Excess liquid was removed by blotting the filters on paper towel then filters were stored at -80°C in 2 ml centrifuge tubes. Samples for RNA were obtained by filtering 100 ml of culture onto 47 mm 1.2 μ m RTTP polycarbonate filters (Fisher Scientific, UK) , then filters were stored at -80°C in 2 ml centrifuge tubes Samples for western blotting and proteomics were obtained by centrifuging 50 ml of culture at 600 g for 10 min in a 50 ml centrifuge tube, the supernatant removed by pipetting and cells were transferred in the residual liquid to a 2 ml centrifuge tube and centrifuged at 600 g for 5 mins. All residual liquid was then removed, and the pelleted cells were stored at -80°C.

2.4 Reverse transcription quantitative polymerase chain reaction (RT-qPCR)

2.4.1 RNA extraction from diatoms

For each culture, RNA was extracted as follows: 1 ml Trizol reagent (Sigma-Aldrich), prewarmed at 65 °C, was added directly to the frozen phytoplankton filter followed by 600 mg of < 106 μ m glass beads (Sigma-Aldrich). Cells were disrupted using an MP FastPrep-24 instrument set at maximum speed for 2 cycles, the first one for 60 sec and the second for 30 sec. Samples were left to recover for 5 min at 22 °C and then centrifuged at 13,000 *g* and 4 °C for 2 min. The supernatant was transferred to a 2 ml screwcap tube containing 1 ml 95% ethanol. RNA was extracted using a Directzol

RNA MiniPrep kit (Zymo Research, R2050), according to the manufacturer's instructions. Samples were eluted in 50 µl of nuclease-free water. Genomic DNA was removed by treating samples with TURBO DNA-free DNAse (Ambion) according to the manufacturer's protocol. RNA was quantified using a Qubit 3.0 Fluorometer, following the protocol of the Qubit Broad Range RNA Assay Kit (Thermo Fisher Scientific), and was stored at -80 °C until needed. Quality of DNA was assessed by visualising 3 µl of sample using agarose gel electrophoresis.

2.4.2 Reverse transcription reaction

Complementary DNA from RNA samples was produced by reverse transcription of 1 µg DNA-free RNA using the QuantiTect Reverse Transcription Kit (Qiagen) following manufacturer's protocol. No reverse transcriptase and no template controls were performed to confirm that samples were DNA-free and that the reactions were free of contaminants.

2.4.3 Primer design for qPCR

Primers for qPCR (Table 2-2) were designed, using Primer3Plus[149] to amplify a 100-300 bp region, with an optimum melting temperature of 60 °C. Melting temperature difference between primers in a pair was 2 °C and GC content was kept between 40% and 60%. The primer pairs were checked to avoid stable homo- and heterodimers as well as hairpin structures using the IDT (Integrated DNA Technologies) Oligoanalyzer 3.1 tool (https://www.idtdna.com/calc/analyzer). Primer efficiencies were all 90–110% and within recommended limits.

The β -actin gene was used in this study as housekeeping gene to normalise the expression of *T. pseudonana* genes[150]. To normalise the expression of *P. tricornutum* genes, exportin 1-like protein (XPO1) was used as housekeeping gene[151].

Table 2-2. List of primers used for qPCR.

Gene	Primer name	Sequence (5′ to 3′)	Conc (nM)	Step
T. pseudonana				
269095	DsyD FOR	GAAGCCAACTCGTTCGACAT	400	2
	DsyD REV	GGGAGAAAGCCAATCATTCA	200	
262307	BCCT FOR	CAC GAT CTG GGT CAT CAC AG	400	3
	BCCT REV	GAA AGC GTC AGT CCA GAA GG	200	
20797	GSDMT FOR	ACACCTCCGTACCCGATGAGG	400	3
	GSDMT REV	CACGCCAAGCCTCACGAAGAA	400	
β-Actin[150]	Actin FOR	AGCCCAACCTTACTGGATTGGAGA	100	2
	Actin REV	TGTGAACAATCGAAGGTCCCGACT	100	
P. tricornutum				
48704	DsyD FOR	CTGCGGCAACTCACTACAAA	400	2
	DsyD REV	CAATGTCGTCATCCCACTTG	200	
48315	BCCT FOR	CAC GAT CTG GGT CAT CAC AG	400	3
	BCCT REV	GAA AGC GTC AGT CCA GAA GG	200	
20301	GSDMT FOR	AGT CCT CAA CAC GGA AAT GG	400	2
	GSDMT REV	ATT ATT CCA TCG CTC GCA AC	400	
Exportin 1-like protein[151]	Pt24186 FOR	GAG GTC CTG TGC GAG AAC AA	500	2
	Pt24186 REV	GGC AAG AAC TTG TGC ACC AG	500	

2.4.4 Quantitative polymerase chain reaction (qPCR)

qPCR was performed using a StepOnePlus instrument (Applied Biosystems). Quantification was performed using a standard SensiFAST SYBR Hi-ROX Kit (Bioline) following the manufacturer's instructions. Reactions (20 μl) contained 10 ng cDNA, and primer concentrations optimised for each pair (Table 2-2). The annealing/elongation temperature was of 60 °C. Two step or three step cycling programme was selected according to the primer set, the number of steps performed per primer pair is indicated in Table 2-2. The two step cycling consisted on 40 cycles

of 95 °C for 5 sec and 60 °C for 15 sec. Alternatively, the three step cycling consisted on 40 cycles of 95°C for 5 sec, 60 °C for 10 sec and 72 °C for 10 sec. A single gene was quantified per run, with three biological replicates and three technical replicates. Standard curves were included in each run to calculate the reaction efficiency (five points in 1:2 dilutions starting from 50 ng cDNA).

2.4.5 Post-run analysis

Analysis of the melting temperatures were performed. For each condition and gene, the cycle threshold (Ct) values of the technical and biological replicates were averaged. Analysis of the post-run melting curve was also performed. Manually detected outliers were excluded from further analysis. The relative expression ratio was calculated following Pfaffl, M. M., 2011[152] method. Results are expressed as normalized fold change relative to the standard conditions.

2.5 Whole transcriptome sequencing and analysis

T. pseudonana CCMP 1335 cultures for transcriptome sequencing were grown in artificial sea water (NEPC) and subjected to a salinity shift and nutrient limitation treatments. All experiments were performed in quadruplicates. For the salinity shift experiment, starter cultures grown at low salinity (PSU 1) were diluted into to fresh low salt media (PSU 1) and normal salinity (PSU5). After 24 hours, samples were diluted again into fresh low or normal salinity media and samples were taken for NMR analysis and RNA sequencing (T1). After 7 days of incubation, samples were taken for NMR analysis and RNA sequencing (T2). For the nutrient limitation experiment, cultures were grown with limiting concentrations of nitrate (10 μM NO_3^-) or silicate (50 μM Si^{4+}). After 7 days of incubation, corresponding to mid-stationary growth phase, samples were taken for NMR analysis and RNA sequencing. To proof that nitrate or silica were the limiting nutrient in each treatment, standard concentrations of nitrate or silica were added back to the media. In both treatments, growth was measured regularly using a Multisizer 3 Coulter counter (Beckman

Coulter) and photosynthetic health was assessed using phytoPAM ED (WALZ) spectrometer. RNA was extracted from the cultures as previously specified (see 'RNA extraction from diatoms' section) and RNA was sequenced by Illumina HiSeq 2000 system. Analysis of the results involved calculating the fpkm mean for each transcript in each treatment. The ratio was calculated by dividing the treatment result by the results from the control samples. For salinity 35 PSU/1 PSU and for nutrient limitation NO3/SiNO3. It was considered as upregulated values >0, and downregulated <0. Results were ratified by qPCR.

2.6 Whole proteome sequencing and analysis

P. tricornutum CCAP 1055/1 samples grown under low salinity and at the onset of stationary phase were sent for proteomic analysis at the University of Warwick. Label Free Quantitation (LFQ) method was used for protein quantification. Absolute and relative abundance of each protein in each replicate and treatment was used to calculate the 2-fold change using the standard culture samples as control. Student T-test adjusted for q-value was performed to identify significant changes of a protein content between treatments.

2.7 Protein immunogold labelling

2.7.1 Analysis of antibody specificity by western blotting.

A polyclonal rabbit IgG was designed against *P. tricornutum* CCAP 1055/1 candidate BCCT transporter and DSYD using the OptimumAntigen software (GenScript Ltd.). The specificity of these antibodies was tested using protein extracted from *P. tricornutum* CCAP 1055/1 cells grown at standard conditions (used as positive control) and at low salinity (used as negative control). 20 ml of diatom cultures were centrifuged at 600 *g* for 10 min and supernatant was decanted. Pellet was mixed with remaining supernatant and transferred into 2 ml screw cap Eppendorf tubes and the centrifuged again at 600 *g* for 10 min. Supernatant was aspirated and pellet was

resuspended in 50 µl of lysis buffer (50 mM Tris-HCl pH 6.8 buffer and 2% SDS) and 2 µl of proteinase inhibiter. Cells were mixed with the lysis buffer by vortexing briefly. Samples were incubated for 30 min at 22 °C and then centrifuged at 4 °C and 16,000 *g* for 30 min. Supernatant was collected and kept at 4 °C. Protein concentration was measured using Qubit 3.0 Fluorometer, following the protocol of the Qubit Protein Assay Kit (Thermo Fisher Scientific). Proteins were transferred to a PVDF (Amersham Hybond-P, GE Healthcare) membrane following protein resolution on a 15 % (v/v) acrylamide gel by semidry western blot as outlined by Mahmood and Yang[153]. After 1 hour blocking with 5% (w/v) skimmed milk powder in TBS (20 mM Tris, 150 mM NaCl, pH 7.5), the anti-BCCT or DSYD antibody was added at a final concentration of 0.386 µ g ml^{-1}. Specific interactions were left to form overnight at 4 °C, before the membrane was washed 4 × 10 min with TBST (TBS + 0.1 % (v/v) Tween 20). TBST (20 ml) was added with 3 µl anti- rabbit IgG-alkaline phosphatase at 1 mg ml^{-1} (Sigma). Following 1 h incubation, the membrane was washed as before with two 10 min TBS washes. Colorimetric detection with NBT/BCIP (Thermo Fisher) was used to detect the target protein as per the manufacturer's instructions. All SDS–PAGE gels were run with BioRad Precision Plus Dual Colour protein size standards and stained with Coomassie using InstantBlue Protein stain (Expedeon).

2.7.2 Immunogold labelling

Cells from *P. tricornutum* CCAP 1055/1 were cryoimmobilized using a Leica EMPACT High-Pressure Freezer (Leica Microsystems), freeze-substituted in an EM AFS (Leica Microsystems) and embedded in Lowicryl HM20 resin (EMS, Hatfield, USA), as in Perez-Cruz *et al.*[154]. Gold grids containing Lowicryl HM20 ultrathin sections were immunolabeled with a specific primary antibody to *P. tricornutum* CCAP 1055/1 BCCT or DSYD (polyclonal rabbit IgG, GenScript), whose stock concentration was 0.550 mg ml^{-1}, and this was diluted 1:15,000. Secondary antibody was an IgM anti-rabbit coupled to 12 nm diameter colloidal gold particles (Jackson) diluted 1:30. As controls, preimmune rabbit serum was used as primary antibody, or the gold-conjugated secondary antibody was used without the primary antibody. Sections

were observed in a Tecnai Spirit microscope (FEI, Eindhoven, The Netherlands) at 120 kV.

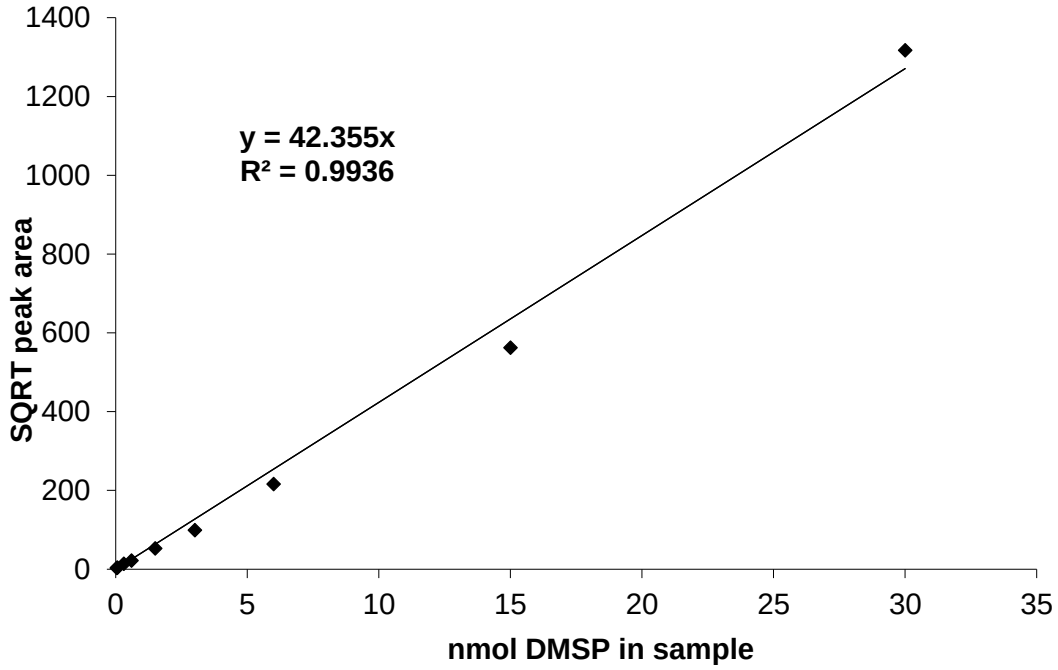
2.8 Metabolite analysis

2.8.1 Detection of DMSHB/DMSP and quantification of DMSP by GC

To detect DMSP or DMSHB by gas chromatography (GC) assays, headspace DMS originated from the alkaline lysis of DMSP or DMSHB was measured. All measurements were performed using a flame photometric detector (Agilent 7890 A GC fitted with a 7693 autosampler) and an HP-INNOWax 30 m × 0.320 mm capillary column (Agilent Technologies J&W Scientific). 300 µl of liquid samples were sealed with PTFE/rubber crimp caps into 2 ml glass serum vials. Methanol extraction method was used for to quantify DMSP from the diatom samples. Methanol extraction involved soaking the filtered samples (see 'Diatom sampling methods' section) in 1 ml 100% methanol. Samples were then stored for 24h overnight at 22 °C to allow the extraction of cellular metabolites. 200 µl of the methanol extract was transferred into a 2 ml glass vial and 100 µl of 10 M NaOH was added. Immediately, vials were crimped and incubated at 22 °C for 24 hours in the dark and monitored by GC. Calibration curves were produced by alkaline lysis of DMSP standards in 100% methanol (Fig. 2.1).

For detection of DMSHB/DMSP from bacterial pellets, samples were resuspended in 200 ml of water, 100 µl of 10 M NaOH was added and vials were crimped immediately. Then, samples were heated at 80 °C for 10 mins (to release DMS from DMSHB). Control sample of DL-DMSHB in water was used as a standard. The detection limit for headspace DMS from DMSP was 0.015 nmol in water and 0.15 nmol in methanol, and from DMSHB was 0.3 nmol in water.

Figure 2.1. Calibration curve of DMS headspace measurements by GC from DMSP dissolved in methanol. Eight point calibration curve used to measure DMSP concentration in methanol extraction samples from DMS originated via alkaline lysis.



2.8.2 Detection of DMSHB, DMSP and GBT in bacteria cultures by LC/MS

Liquid Chromatography Mass Spectrometry (LC/MS) was used to ratify whether the DMS detected by GC in the bacteria samples originated from DMSP or DMSHB as well as to detect the presence of GBT. To obtain samples from bacteria cultures, 2 ml of culture were centrifuged at 18,000g for 3 min. Supernatant was removed and pellets were stored at -80°C. Frozen pellets were resuspended in 350 µl of the extraction solvent (80% LC/MS grade acetonitrile) and then centrifuged at 18,000g for 4 min. 300 µl of the supernatant was transferred into a glass vial. LC-MS was carried out using a Shimadzu Ultra High Performance Liquid Chromatography (UHPLC) system formed by a Nexera X2 LC-30AD Pump, a Nexera X2 SIL-30AC Autosampler, a Prominence CTO20AC Column oven, and a Prominence SPD-M20A Diode array detector; and a Shimadzu LCMS-2020 Single Quadrupole Liquid Chromatograph

Mass Spectrometer. Samples were analysed in hydrophilic interaction chromatography (HILIC) mode using a Phenomenex Luna NH_2 column (100 x 2 mm with a particle size of 3 μm) at pH 3.75. Mass spectrometry spray chamber conditions were capillary voltage 1.25 kV, oven temperature 30 ºC, desolvation temperature 250 ºC and nebulising gas flow 1.50 L min-1. Solvent A is 5% acetonitrile + 95% 5 mM ammonium formate in water. Solvent B is 95% acetonitrile + 5% 100 mM ammonium formate in water. Flow rate was 0.6 ml min-1 and gradient (% solvent A/B) was t = 1 min, 100% B; t = 3.5 min, 70% B; t = 4.1 min, 58% B; t = 4.6 min, 50% B; t = 6.5 min, 100% B; t = 10 min, 100% B. The injection volume was 15 μl. All samples were analysed immediately after being extracted. The targeted mass transition corresponded to $[M+H]^+$ of DMSP (*m*/*z* 135), DMSHB (*m*/*z*165) and GBT (*m*/*z* 118) in positive mode. DMSP, DMSHB and GBT standards were run in the extraction solvent.

2.8.3 Quantification of DMSP, GBT and choline by LC/MS

Metabolite extraction and quantification of DMSP and GBT from diatom samples was conducted following the method described by Beale, R. & Airs, R., 2016[155]. A Dionex 3400RS HPLC, coupled to an AmazolSL quadrupole ion trap MS (Bruker Scientific) via an electrospray ionisation interface was employed. Data analysis was carried out using the Bruker Compass software package, using DataAnalysis for peak identification and characterization of lipid class, and QuantAnalysis for quantification of the relative abundance of DMSP, GBT and choline.

2.8.4 Detection of DMSP and GBT by Nuclear Magnetic Resonance (NMR)

T. pseudonana CCMP 1335 cultures analysed by NMR are the same as the cultures used for whole transcriptome analysis (see 'Whole transcriptome sequencing and analysis' section). Metabolites were extracted and analysed by NMR by Dr. Gwenaelle Le Gall at the School of Pharmacy facilities, University of East Anglia.

1H NMR was used to identify the presence, absence, and concentration of several metabolites in *T. pseudonana*. Samples were prepared for 1H NMR spectroscopy by mixing the pellet with 200 μL of phosphate buffer (0.2MNa2HPO4, 0.038 M NaH2PO4 [pH 7.4]) made up in 100% D2O and containing 0.06% sodium azide, and 1.5 mM DSS (sodium 2,2-dimethyl-2-silapentane- 5-sulfonate) as a chemical shift reference. The sample was mixed, and 500 μL was transferred into a 5-mm NMR tube for spectral acquisition. The 1H NMR spectra were recorded at 600MHz on a Bruker Avance spectrometer (Bruker BioSpin GmbH, Rheinstetten, Germany) running Topspin 2.0 software and fitted with a cryoprobe and a 60-slot autosampler. Each 1H NMR spectrum was acquired with 128 scans, a spectral width of 8,012.8 Hz, an acquisition time of 2.04 s, and a relaxation delay of 2.0 s. The "noesypr1d" presaturation sequence was used to suppress the residual water signal with a low-power selective irradiation at the water frequency during the recycle delay and a mixing time of 100 ms. Spectra were transformed with a 0.3-Hz line broadening, manually phased, baseline corrected, and referenced by setting the DSS methyl signal to 0 ppm.

2.9 In silico assays and sequence analysis.

2.9.1 Identification of BCCT, DSYD and GSDMT enzymes and phylogenetic trees.

BLAST searches to identify homologues of *P. tricornutum* CCAP 1055/1 and *T. pseudonana* CCMP 1335 proteins were performed using BLASTp at the National Centre for Biotechnology Information (NCBI) database. BLASTp tool at NCBI was used to produce phylogenetic trees of the results originated by the searches. Phylogenetic and molecular evolutionary analyses of functional DSYD and GSDMT along with selected non-functional or distantly related methyltransferases were conducted using MEGA version X[156]. Protein sequence alignments were performed using ClustalW (https://www.genome.jp/tools-bin/clustalw).

2.9.2 Sequence optimization and gene synthesis

Genes selected to be synthesised were codon-optimised for *E. coli* expression and modified to avoid the presence of *Xba*I, *Nde*I and *Eco*RI within their sequences using Invitrogen GeneArt®. Genes were synthesise using the facilities provided by the John Innes Centre.

2.9.3 Protein sequence analysis

BCCT transporter, DSYD, GSDMT and DSYD/GSDMT secondary structure and transmembrane helical spanners (TMS) were predicted using Phyre2[157]. Signal peptide prediction of GSDMT, DSYD and BCCT proteins were performed using ASAFind[158], HECTAR[159] and SignalP-5.0[160].

2.10 *In vitro* genetic manipulation

2.10.1 Polymerase Chain Reaction (PCR)

Genes were amplified using polymerase chain reaction (PCR) in a Thermal Cycler. Standard 50 µl PCR mixes contained 25 µl MyFi™ DNA Polymerase 1 µl template, 2 µl of 20 pmol of 50:50 forward and reverse primers in 22 µl of H_2O. For smaller reaction volumes, the components were adjusted proportionally. Oligonucleotide primers used in this study were synthesised by Eurofins Genomics (Table 2-3). To amplify genes from diatoms, complementary DNA was used as template. DNA from bacterial colonies were also amplified. For colony PCR, a sterile toothpick was used to gently touch a colony, the toothpick was introduced into PCR tubes containing 20 µl sterile water. Tubes were microwaved for 10 seconds, and 1 µl of the lysed mixture was used in the PCR mix.

Table 2-3. List of primers used in this study and their use. The restriction enzymes sites are underlined.

Primer name	Sequence (5′ to 3′)	Use
M13 uni (-43)	AGGGTTTTCCCAGTCACGACG TT	Universal forward primer used to amplify inserts in pLMB509
M13 rev (-29)	AGGGTTTTCCCAGTCACGACG TT	Universal reverse primer used to amplify inserts in pLMB509
T7	TAATACGACTCACTATAGGG	Universal forward primer used to amplify inserts in pET21a, pET16b and pBluescript
T7 Terminal	GCTAGTTATTGCTCAGCGG	Universal reverse primer used to amplify inserts in pET21a, pET16b and pBluescript
Tp_20797F	ATATACATATGATGGCACCCA AC ACCTCCACCAGC	Forward primer used for cloning of *T. pseudonana GSDMT* into pET16b
Tp_20797R	GATGGATCCTAGGAACTTGGA C TTTCGTCAACGACGG	Reverse primer used for cloning of *T. pseudonana GSDMT* into pET16b
Tp269095NdeFOR2	CAACACATATGGCTCCCAACA CCACACCC	Cloning of *T. pseudonana* CCMP 1335 DSYD into pET16b and pET21a
Tp_ 269095BamREV	CGTTGATGTGGATCCAGAGTC AGGAGTC	Cloning of *T. pseudonana* CCMP 1335 DSYD into pET16b and pET21a
Pt_48704NdeFOR	GCGCCATATGACTGCGGCAAC TCAC	Cloning of *P. tricornutum* CCAP 1055 DSYD into pET16b and pET21a
Pt_48704BamREV	CGACGGATCCGTATCTGCGTT GG	Cloning of *P. tricornutum* CCAP 1055 DSYD into pET16b and pET21a

2.10.2 Agarose gel electrophoresis

PCR products were visualised using agarose gel electrophoresis. Agarose gel contained 1 or 1.5% (w/v) agarose, for DNA or RNA visualization respectively, using 1x TAE Buffer. Agarose was melted in the buffer and cool down before adding 3 μl Ethidium Bromide (10 mg/ml). 1 KB Plus DNA ladder (Invitrogen) was loaded into the wells alongside the samples to be used as size reference. Gels were typically run

at 90 V for 60 min, and the separation of DNA or RNA fragments was visualised using a UV gel imaging doc.

2.10.3 PCR purification (Roche)

PCR amplified DNA was recovered using the Roche High Pure PCR Product Purification Kit following manufacturer's instructions. The purified PCR product was eluted from the column using 50 μl sterile water, collected in a 1.5 ml microcentrifuge tube.

2.10.4 Gel extraction (Qiagen)

DNA run in an agarose gel electrophoresis was recovered by gel extraction using QIAquick Gel (QIAGEN) following manufacturer's instructions. The DNA was eluted using 50 μl of sterile water collected in a 1.5 ml microcentrifuge tube.

2.10.5 Cloning into pLMB509 and pET vectors

PCR amplified DNA and synthesised genes in plasmids were cloned or subcloned into pLMB509 or pET vectors. Oligonucleotide primers used for molecular cloning were synthesized by Eurofins Genomics (Table 2-3). Routine restriction digestions and ligations for cloning were performed as described in Downie *et al.*[161]. Sequencing of plasmids and PCR products was performed by Eurofins Genomics. Candidate DSYD were subcloned into pLMB509[162], a taurine-inducible plasmid for the expression of genes in *Rhizobium leguminosarum*. Other candidate genes were cloned or subcloned into isopropylthiogalactoside (IPTG)-inducible *E. coli* expression plasmid pBluescript or pET21a or pET16b (Merk Millipore). Restriction enzymes used were *Nde*I and *Bam*HI or *Eco*RI. All plasmid clones are described in Table 2-4.

Table 2-4. List of plasmids used in this study.

Plasmid	Description	Reference
pRK2013	Helper plasmid used in triparental mating	Figurski and Helinski (1979)[163]
pUC18	IPTG inducible vector for expression of cloned genes in *E. coli*	Norrandera[162], J. *et al.* (1983)[164]
pLMB509	Taurine inducible vector for expression of cloned genes in alphaproteobacteria	Tett *et al.* 2012
pET21a	IPTG inducible vector for expression of cloned genes in *E. coli*	Merck Millipore
pET16b	IPTG inducible vector for expression of cloned genes in *E. coli*	Merck Millipore
pBluescript	IPTG inducible vector for expression of cloned genes in *E. coli*	Merck Millipore
pBIO2287	*P. tricornutum* CCAP 1055/1 BCCT chimera synthesised and cloned in pBluescript	This study
pBIO2288	*T. pseudonana* CCMP 1335 11247 methyltransferase synthesised and cloned in pLMB509	This study
pBIO2289	*T. pseudonana* CCMP 1335 11247 methyltransferase synthesised and cloned in pET21a	This study
pBIO2291	*T. pseudonana* CCMP 1335 *DSYD* PCR amplified and cloned in pET21a	This study
pBIO2292	*P. tricornutum CCAP 1055/1 DSYD* PCR amplified and cloned in pLMB509	This study
pBIO2293	*P. tricornutum CCAP 1055/1 DSYD* PCR amplified and cloned in pET21a	This study
pBIO2294	*T. oceanica* CCMP1005 *DSYD* synthesised and cloned in pLMB509	This study
pBIO2295	*T. oceanica* CCMP1005 *DSYD* synthesised and cloned in pET21a	This study
pBIO2296	*T. oceanica* CCMP1005 20742 methyltransferase synthesised and cloned in pLMB509	This study
pBIO2297	*T. oceanica* CCMP1005 20742 methyltransferase synthesised and cloned in pET21a	This study
pBIO2298	*G. sunshinyii dsyD* synthesised and cloned in pLMB509	This study
pBIO2299	*G. sunshinyii dsyD* synthesised and cloned in pET21a	This study
pBIO2300	*O. tauri DSYD* synthesised and cloned in pLMB509	This study
pBIO2301	*O. tauri DSYD* synthesised and cloned in pET21a	This study
pBIO2302	*A. halophila* SAM-methyltransferase synthesised and cloned in pLMB509	This study
pBIO2303	*A. halophila* SAM-methyltransferase synthesised and cloned in pET21a	This study
pBIO2304	*N. exalbescens* SAM-methyltransferase synthesised and cloned in pLMB509	This study
pBIO2305	*N. exalbescens* SAM-methyltransferase synthesised and cloned in pET21a	This study
pBIO2306	*M. bouillonii* SAM-methyltransferase synthesised and cloned in pLMB509	This study
pBIO2307	*M. bouillonii* SAM-methyltransferase synthesised and cloned in pET21a	This study

pBIO2308	*S. vulgare DSYD* synthesised and cloned in pLMB509	This study
pBIO2309	*S. vulgare DSYD* synthesised and cloned in pET21a	This study
pBIO2310	*P. haitanensis* SAM-methyltransferase synthesised and cloned in pLMB509	This study
pBIO2311	*P. haitanensis* SAM-methyltransferase synthesised and cloned in pET21a	This study
pBIO2314	*T. pseudonana* CCMP 1335 *GSDMT* PCR amplified and cloned in p16b	This study
pBIO2315	*P. tricornutum* CCAP 1055/1 *GSDMT* synthesised and cloned in pUC57	This study
pBIO2316	*E. siliculosus* candidate GBT synthase synthesised and cloned in pUC57	This study
pBIO2317	*E. litoralis* DSYD/GSDMT synthase synthesised and cloned in pUC57	This study
pBIO2318	*F. cylindrus* CCMP1102 candidate GBT synthase synthesised and cloned in pUC57	This study
pBIO2319	*T. thiocyanaticus* DSYD/GSDMT synthase synthesised and cloned in pUC5	This study
pBIO2320	*M. alba* DSYD/GSDMT synthase synthesised and cloned in pUC5	This study
pBIO2321	*T. oceanica* GBT synthase synthesised and cloned in pUC57	This study
pBIO2322	*Geitlerinema* sp. GBT synthase synthesised and cloned in pUC57	This study

2.11 *In vivo* genetic manipulation

2.11.1 *E. coli* competent cells

A starting culture of 5 ml LB was inoculated with *E. coli* (803/JM101) and incubated overnight at 37°C. An aliquot of the starter culture was inoculated in 100 ml LB (1:100) and incubated 37°C, 200 rpm until reaching an OD_{600} 0.2 – 0.4. Culture was transferred into sterile Falcon™ 50 ml Conical Centrifuge Tubes and centrifuged at 6,000 rpm for 10 min at 4°C. Supernatant was removed and pellets were kept on ice in the following steps. Pellets were carefully mixed with 10 ml ice cold 0.1 M $CaCl_2$ and left on ice for at least 60 min. Samples were centrifuged at 6,000 rpm for 10 min at 4°C and the supernatant removed. Pellet was resuspended in 2 ml of 0.1M $CaCl_2$. Cells were left on ice for at least 3 h and then stored in the fridge overnight.

2.11.2 Heat shock transformation into *E. coli* 803 or JM101

Up to 100 ng of DNA was added to 100 µl *E. coli* competent cells and incubated on ice 1 h. A tube with cells with no DNA no cells added was used as negative control. Samples were heat shocked at 42 °C for 2 min and transferred to ice for 2 minutes. 500 µl LB was added to the cells and incubated at 37 °C for 60 – 90 minutes. Cells were plated on LB agar containing ampicillin (100 µg ml^{-1}). For molecular cloning IPTG and X-gal was also added to the medium and white/blue screening was performed.

2.11.3 Conjugation by triparental mating

Plasmids were transferred from *E. col*i to *Rhizobium leguminosarum* J391 by conjugation, in a triparental mating using the kanamycin resistant helper plasmid pRK2013[163]. *E. coli* containing the plasmid to be transferred, *E. coli* containing the helper plasmid pRK2013 and the heterologous host *R. leguminosarum* were mixed in rich TY media plates and incubated at 28 °C overnight. The bacteria mix was then plated on TY media plates containing selective antibiotics and incubated at 28 °C. Successful crosses confirmed using colony PCR.

2.11.4 Plasmid extraction from bacteria cultures

2.11.4.1 Phenol chloroform extraction using Miniprep Qiagen kit

1.5 ml of overnight grown culture was centrifuged at 18,000 *g* for 2 min and supernatant was removed by aspiration. Pellet was resuspended in 250 µl Buffer P1 then BufferP2 was added and mixed by inversion. After a 5 min incubation at room temperature, 350 µl Buffer P3 was added and mixed by inversion. Samples were left on ice for 5 minutes, and then centrifuged at 18,000 *g* for 10 min. The supernatant was collected and 400 µl Phenol:Chloroform:Isoamyl Alcohol 25:24:1 (v/v) was added and vortexed to mix. Samples were centrifuged at 14,000 *g* for 2 min. The top aqueous layer was collected in a new microcentrifuge tube containing 700 µl of 100% ethanol.

Tubes were mixed by inversion and centrifuged at 18,000 *g* for 15 min. The supernatant was discarded, and 500 µl of 70% ethanol was added. Samples were centrifuged at 18,000 *g* for 5 min. The pellet was air-dried for 10 min and resuspended in up to 50 µl nuclease-free water.

2.11.4.2 Plasmid extraction by QIAprep Spin Miniprep Kit (QIAGEN)

Plasmids sent for sequences were extracted from ~3 ml of overnight culture using QIAprep Spin Miniprep Kit (QIAGEN) following manufacturer's instructions. The column was eluted with 35 – 50 µl nuclease-free water added to the membrane.

2.11.4.3 Plasmid extraction by Plasmid Midiprep Kit (QIAGEN)

Plasmid Midiprep kit (QIAGEN) was used for high quality, high concentration plasmid extractions. Plasmids were recovered from overnight grown 100 ml of culture and QIAGEN-tip 100 column was used following manufacturer's instructions. Plasmid DNA was eluted in 50 – 100 µl of nuclease-free water.

2.12 *In vivo* assays

2.12.1 BCCT transporter assay +/- DMSP and the different strains

Synthesised BCCT transporter chimera clones were transformed into *E. coli* MKH13, 803 and Rosetta cells. Cultures were grown in rich (LB) media overnight at 37 °C. From the starter cultures, aliquots were inoculated in triplicates in 5 ml fresh rich or and minimal media (M9 salts) (1:100) with or without addition of DMSP and incubated at 37 °C for 2h. Cells containing an empty plasmid were used as negative controls. 1 ml of culture was centrifuged at 18,000 *g* for 2 min and supernatant was removed. Pelleted cells were washed x3 by resuspending in sterile distilled water and centrifuging at 18,000 *g* for 2 min. Washed cells were resuspended in 200 µl of water and then analysed by GC.

2.12.2 Glycine betaine synthesis via glycine pathway assays

Candidate glycine/sarcosine/dimethylglycine methyltransferases clones were transformed into *E. coli* BL21. Starter cultures were grown overnight at 37 °C in rich media (LB). 1 ml of culture was centrifuged at 18,000 *g* for 2 min, the supernatant was removed, and the pellet was resuspended in minimal media (M9 salts), this step was repeated x3 to wash the cells. In the final washing step, cells were resuspended in 1 ml of minimal media. 100 µl of cultures were inoculated in triplicates in 5 ml of minimal media containing 0.2 mM IPTG (Sigma-Aldrich) and either 0.5 mM MTHB, 0.5 mM glycine/sarcosine/dimethylglycine or nothing added. *E. coli* BL21 cells containing empty plasmids were used as negative controls. Cultures were incubated at 30 °C for 48 h before 3 ml samples were collected for metabolite analysis by GC and LC/MS.

2.12.3 DMSHB synthase activity assays.

Candidate DMSHB synthetases were subcloned into pLMB509 and transformed into *E. coli* 803 cells. Clones were then transferred into *R. leguminosarum* by triparental mating conjugation. *R. leguminosarum* containing the *DSYD* homologues or an empty pLMB509 plasmid (negative control) was grown overnight in TY media at 28 °C. 1 ml of culture was centrifuged at 18,000 *g* for 2 min, the supernatant was removed, and the pellet was resuspended in minimal media (Y media), this step was repeated x3 to wash the cells. In the final washing step, cells were resuspended in 1 ml of minimal media. 100 µl of cultures were inoculated in triplicates in 5 ml of Y minimal media containing 5 mM taurine (to induce expression), and either 0.5 mM MTHB, 0.5 mM glycine/sarcosine/dimethylglycine or nothing added. Cultures were incubated at 28 °C for 48 h before 3 ml samples were collected for metabolite analysis by GC and LC/MS.

2.12.4 Phenotyping *E. coli* expressing TpGSDMT

GSDMT from *T. pseudonana* CCMP 1335 was PCR amplified, cloned into pET16a and mobilised into *E. coli* BL21. Starter cultures of *E. coli* BL21 containing pET16b-*TpGSDMT* or just pET16b (as negative control) were grown overnight at 37 °C in rich media (LB). 3 ml of culture was centrifuged at 18,000 *g* for 2 min, the supernatant was removed, and the pellet was resuspended in minimal media (M9 salts), this step was repeated x3 to wash the cells. In the final washing step, cells were resuspended in 900 µl of minimal media. 300 µl of washed cells were inoculated in 100 ml of high salt minimal media (salinity adjusted by addition of 0.6 M of NaCl) containing 0.2 mM IPTG (Sigma-Aldrich), in triplicates Cultures were incubated at 30 °C for 173 h and growth was monitored at regular intervals by measuring the OD_{600}. After the last measurement, samples were collected for metabolite analysis by LC/MS.

2.13 *In vitro* assays

The enzyme activity of DsyD was measured by determining the production of S-adenosyl-L-homocysteine (SAH) using high performance liquid chromatography (HPLC). Reaction system of MmtD contained 5 mM S-adenosyl methionine (Ado-Met), 5 mM 4-methylthio-2-hydroxybutyrate (MTHB) and 0.5 µM recombinant DsyD. Britton-Robinson buffer in the pH range 3.0-9.0 with a final concentration of 20 mM was used to determine the optimum pH for DsyD activity. The optimum temperature for DsyD activity was determined by incubating the reaction mixtures in 20 mM Britton-Robinson buffer (pH 7.0) at 5-50°C. The kinetic parameters of DsyD were determined by non-linear analysis with 0.5 µM DsyD, 5 mM Ado-Met and 0.01-2 mM MTHB at the optimum pH and temperature.

Components of the reaction solutions were separated using a reversed-phase C18 column (4.6×250 mm, 5 µm particle size; Waters, USA) connected to a HPLC system (Dionex, USA). The samples were eluted with a linear gradient of 1-20% (v/v)

acetonitrile in 20 mM ammonium acetate (pH 5.5) over 15 min at a flow rate of 1 ml/min. The detection wavelength was 260 nm.

2.14 Statistics

Statistical methods for RT–qPCR are described in the relevant section above. All measurements for DMSP and GBT production (in diatoms) are based on the mean of three biological replicates per condition tested, and error bars are shown from calculations of standard deviations, all experiments were performed at least twice unless otherwise stated. To identify statistically significant differences between standard and experimental conditions, a two-tailed paired Student's t-test ($P < 0.05$) was applied to the data, using Microsoft Excel.

Chapter 3

Physiological characterization of DMSP and GBT production by model diatoms

3 Physiological characterization of DMSP and GBT production by model diatoms

3.1 Conditions affecting DMSP and GBT production by *T. pseudonana* and *P. tricornutum*.

As explained in the introduction chapter, there are a series of environmental conditions named salinity, nutrient limitation, oxidative stress and temperature, known to affect the production of DMSP, GBT or both in different producer organisms. As it was described, the production of these metabolites is highly species specific and can vary hugely even within the same genus[81,99]. In addition, it is challenging to isolate the effects of individual conditions on growth and production of GBT and DMSP as in most cases there are more than one condition causing and effect at the same time. In the following introductory section, an attempt to dissect the effects of salinity, nutrient limitation and other conditions known to affect the production of GBT and DMSP by the two model diatoms, *Thalassiosira pseudonana* (Tp) and *P. tricornutum* (Pt) will be discussed in more detail.

3.1.1 Salinity

3.1.1.1 Effect of salinity in DMSP and GBT production in *T. pseudonana*

Molecular phylogenetic analysis suggested that *T. pseudonana* has likely evolved from fresh water ancestors[165], and posteriorly, colonised waters worldwide. This centric diatom is found in fresh waters[166], brackish waters, as well as marine environments[167], being able to survive a wide range of salinities ranging from 0 PSU[168] to higher than 50 PSU[128,169,170].

T. pseudonana grown in salinities of 5 or lower PSU show a slow growth rate compared to higher salinities[171]. On the other hand, increased salinities (up to 30 PSU) do not significantly affect the growth rate[171] nor other aspects of *T. pseudonana*'s

physiology such as reproduction or silica frustule formation, opposite to what it is observed in other diatoms[172,173]. However, when subjected to dramatic changes, e.g. when *T. pseudonana* is transferred from 35 PSU to 62 PSU, this diatom struggles to adapt to and survive at such a high salinity[128]. The optimum salinity for *T. pseudonana* varies depending on the origin of the *T. pseudonana* clone, for example *T. pseudonana* isolated from fresh water ballast from the Chengjiang River has an optimum PSU of 20[171], *T. pseudonana* isolated from brackish and fresh water optimum PSU ranges from 8 to 32 whereas *T. pseudonana* from oceanic waters optimum ranges from 14 to 32 PSU[172].

These disparities evidence that *T. pseudonana* is a highly euryhaline organism with great capacity to adapt to different salinities and, as reviewed in the introduction (see Chapter 1) one of the proposed mechanisms used for this adaptability is the synthesis and uptake of DMSP and GBT.

Two previous studies on the effect of salinity on the intracellular DMSP concentration in *T. pseudonana* showed an increase on intracellular DMSP concentration in cultures grown at a marine environment like salinity (35 PSU) compared to cultures grown at lower salinity (10 PSU)[128,174]. In addition, Hockin shows that there is no difference in intracellular DMSP levels in cultures grown at salinities between 31 and 62 PSU[128], although a decrease in the intracellular DMSP concentration was observed as an initial response to osmotic stress (after 24 h) of cultures transferred from 35 PSU to 50 PSU[175].

As previously mentioned in Chapter 1, the Met transamination pathway is the proposed DMSP biosynthetic pathway in diatoms. Met aminotransferase, MTOB reductase and MTHB methyltransferase activity assays performed using cell extracts of *T. pseudonana* grown at low and high salinity also indicated that salt potentially upregulates the three first steps of the transamination pathway[101], which would correlate with the increased concentration of DMSP in the cells.

Similarly, the concentration of GBT and sarcosine, a GBT precursor, in *T. pseudonana* cells are also higher in increased salinity[130,175]. Albeit GBT and DMSP not being the only known osmoprotectants present in *T. pseudonana*, regulation of these two important osmolytes suggest they play a role in the ability of *T. pseudonana* to adapt and survive in waters worldwide.

3.1.1.2 Effect of salinity in DMSP and GBT production in *P. tricornutum*

P. tricornutum is considered non abundant coastal species found in unstable environments, where rapid changes on salinity and temperature may occur, such as estuaries and rock pools[176]. However, it is possible to find different *P. tricornutum* strains in freshwater, brackish and marine waters[177] and it has been reported to grow in salinities ranging from 0 to 45 PSU[172]. Nevertheless, the growth rate of this species is two times lower at 0 PSU compared to other salinities, with 25 PSU being the optimum level observed[172]. Curiously, salinity has a profound effect on the cell morphotype[176] which directly impacts the buoyancy capacity and the vertical movement of the cells in the natural environments[178]. As with *T. pseudonana* (described above), *P. tricornutum* synthesises compatible solutes such as proline, GBT or DMSP. In *P. tricornutum*, proline is the most abundant zwitterion, followed by GBT and then by DMSP[129]. Furthermore, Dickson and Kirst *et al.* reported a linear increase on intracellular concentrations of those osmolytes with increasing salinity in nutrient replete conditions[179]. Interestingly, a candidate choline-glycine betaine transporter (XM 002182735.2) was identified to be regulated by salinity by Osborn H. L. and Hook S. E[180] suggesting that these organisms might import these molecules in addition to synthesising them. It should be noted though that the locations of such transporters and their direction of transport across a membrane will determine their function, e.g. they may be involved in the import to an organelle such as the chloroplast.

3.1.2 Nitrogen limitation

3.1.2.1 Effect of nitrogen limitation in DMSP and GBT production in *T. pseudonana*

Adaptability to N limitation is critical for *T. pseudonana*, as N availability affects key metabolic pathways involved in growth and overall survival. N scarcity causes a reduction in the accumulation of free nitrates, amino acids and proteins. It also affects photosynthesis increasing oxidative stress by causing an excessive accumulation of C components and a reduction in electron transfer[181]. N starvation also causes changes in the C metabolism[181] and the lipid metabolism[175].

A reduction in the levels of N to levels that are limiting to growth triggers a decrease in the synthesis of GBT and its' precursor sarcosine[175] as well as other N compound counterparts such as proline and homarine (free amino acids)[112,175,181]. Although a direct link between N limitation and DMSP accumulation has not been found in many species, in *T. pseudonana* the synthesis of this sulphurous molecule is upregulated when cells reach the onset of N depletion[75,112,174,175,181]. In this case, it is likely DMSP is playing a multifunction role, as an osmoprotectant as well as alleviating the oxidative stress produced by low N [78,181], as a mean for dealing with excess C and S, or as a way of maintaining a low equilibrium of Met pools in the cells and preventing its accumulation derived from the degradation of proteins[75]. Indeed, Met and MTOB concentrations are reduced in N starved conditions[175], possibly due to a shift in the equilibrium of the reversible conversion of Met into MTOB, favouring DMSP synthesis. Moreover, DMSP synthesis also serves to reintroduce nitrogen from the amino group from methionine into the system and contribute to the production of new amino acids[182].

3.1.2.2 Effect of nitrogen limitation in DMSP and GBT production in *P. tricornutum*

Similarly to *T. pseudonana*, nitrogen limitation causes a series of metabolic changes in *P. tricornutum* such as accumulation of carbon compounds[183]. However, the most studied effect is the *de novo* synthesis and increased accumulation of lipids, mostly triacylglycerides due to its biotechnological implications in the biofuel industry[129,184,185]. In addition, there is an increased turnover of phosphatidylcholines[184] and there is a deprioritization of photosynthesis[185] and a reduction of the TCA cycle[129,186].

There are reported evidences of DMSP accumulating when *P. tricornutum* is N starved, and also dimethylsulfoniobutyrate[129,186]. On the other hand, Burrows *et al.* hypothesised that GBT could be used when N is scarce as a pool for choline production or as a N source to allow cell division when exogenous N source is not available from the media[184].

3.1.3 Other conditions: oxidative stress and temperature

3.1.3.1 Effect of temperature in *T. pseudonana* and *P. tricornutum*

T. pseudonana is able to grow at a temperature ranging from 7 ˚C[187] to 32.5 ˚C, and its optimum temperature is relatively high, from 25.99 ˚C to 27.16 ˚C[188]. In contrast, *P. tricornutum* has been found in environments with temperatures as low as 0 ˚C and up to 25 ˚C, although it grows best at temperatures between 11 – 26 ˚C[189]. The growth rate of *P. tricornutum* was shown to peak at 23 ˚C. Temperatures above 32.5 ˚C and 27 ˚C are lethal for *T. pseudonana* and *P. tricornutum* respectively[188].

There is a direct relation between growth rate and cell size[114,190]. Both *P. tricornutum* and *T. pseudonana* present a higher growth rate and smaller cell volumes as temperature increases [114]. However, at temperatures between 23˚C and 25˚C *P. tricornutum* presents a lower division rate and larger cell volume[190].

A comprehensive study on the effect that temperature has on the production of GBT and DMSP in both *T. pseudonana* and *P. tricornutum* was done by Spielmeyer and Pohnert in 2012[114], transferring cultures grown at 15 °C to higher temperature (20.5 °C). Their results show that GBT is dependent of temperature and it has a direct relationship, at higher temperature higher GBT concentrations (~3-fold) in both model diatoms. Conversely, DMSP concentration decreases with higher temperatures (~2-fold) in both model diatoms although the difference is less pronounced in *P. tricornutum*.

3.1.3.2 Effect of oxidative stress in *T. pseudonana* and *P. tricornutum*

There are many factors that can cause oxidative stress in diatom cells, for instance light exposure and intensity[191], or limitation of nutrients such as iron[151], nitrogen[181] or CO_2[78]. The concentration of H_2O_2, a reactive by product of photosynthesis and the major precursor for •OH radicals, increases when *T. pseudonana* is exposed to those stressors. Likewise, DMSP concentrations are also higher (20- to 60-fold) under those conditions in both model diatoms[78]. In this study, exogenous H_2O_2 will be added to *T. pseudonana* and *P. tricornutum* cultures to investigate osmolyte accumulation in these strains.

3.1.4 Aims of the chapter

There is a lot of data available on how GBT and DMSP is regulated by salinity, nitrogen availability, temperature and oxidative stress and one aim of this chapter is to confirm the effects previously published. Furthermore, I aim to investigate the effects of the addition of GBT and DMSP, as well as the DMSP precursor MTHB, and glycine betaine precursors choline and glycine on the regulation of these pathways.

Information obtained from these experiments to gauge the regulation of GBT and DMSP will be used as the basis to identify candidate genes of interest for the synthesis and/or cycling of the environmentally important molecules. This involved selected

samples from these cultures being used in following chapters to analyse the global transcription and translation profiles of these diatoms under conditions that affect DMSP and/or GBT production and uptake.

3.2 Effects of environmental conditions in growth and DMSP concentrations.

3.2.1 Obtaining axenic cultures

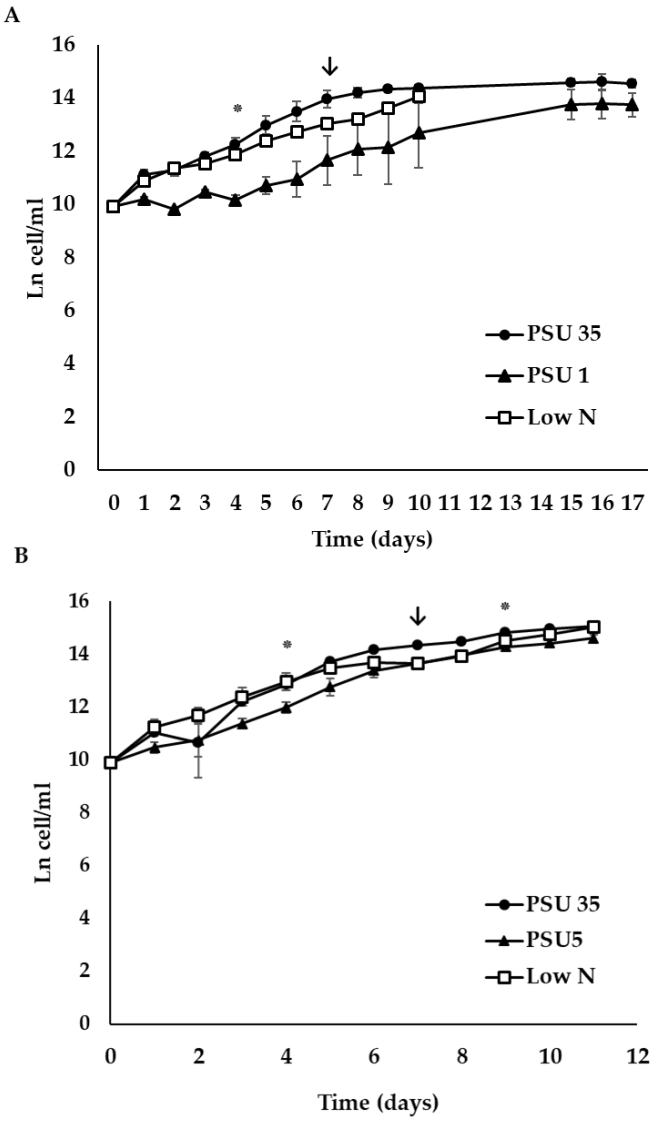
Firstly, *Thalassiosira pseudonana* CCMP 1335 and *Phaeodactylum tricornutum* CCAP 1055/1 cultures were treated to ensure the cultures were axenic. To confirm that the cultures were axenic they were treated with a cocktail of antibiotics containing streptomycin (400 μg ml^{-1}), chloramphenicol (50 μg ml^{-1}), gentamicin (20 μg ml^{-1}) and ampicillin (100 μg ml^{-1}). Cells were grown overnight at 22 °C with a light intensity of 120 μE $m^{-2}s^{-1}$ and a light/dark cycle of 16 h light/8 h dark. An aliquot of these antibiotic treated cultures was transferred to fresh standard media (1:40). Once the cells reached exponential growth, they were transferred again to fresh media.

A small aliquot of the cultures was plated in F/2 medium made with enriched artificial seawater[135] in 1% agarose and incubated at 28 °C for a week. After a week no bacterial growth appeared on the plate. In addition, samples were examined under the microscope and no bacterial cells were observed. The negative results obtained on those two tests suggest that the cultures were axenic hence ready to be used on the following experiments.

3.2.2 Growth under different salinities

P. tricornutum and *T. pseudonana* stock cultures are maintained at standard conditions (see Chapter 2). Starter cultures of each model diatom were inoculated into standard media or low salinity media (1 PSU for *T. pseudonana* and 5 PSU for *P. tricornutum*) allowing the cells to acclimate to the change in salinity. These cultures were

Figure 3.1. Ln of cell number per ml of cultures of (A) *T. pseudonana* CCMP 1335 and (B) *P. tricornutum* CCAP 1055/1. Cultures grown in F/2 medium mixed in artificial seawater in standard salinity and nitrogen (PSU 35), low salinity and standard N concentration (PSU 1 and PSU 5), or at normal salinity and low N concentration (Low N). Standard cultures contain 882 µM of NO_3^- and low N 30 µM of NO_3^-. Results are shown as ln of the cell/ml means ± standard deviation of 3 independent cultures. * indicates sampling points for DMSP measurements, ↓ indicate sampling point and addition of 882 µM of NO_3^- to the media.



monitored daily as described in the method chapter 2. When they reached exponential phase, cell density of 10^6 cells/ml for *T. pseudonana* and $5x10^6$ cell/ml for

P. tricornutum, 2×10^4 cells were transferred to 500 ml of fresh media with either standard or low salt concentration in triplicates.

Culture growth was monitored regularly until they reached stationary phase (Figure 3.1). Cultures of both diatoms at low salinity showed a lower growth rate than the cultures grown at standard salinity (Table 3-1). Reduced growth rates at lowered salinity was also previously described for *T. pseudonana*[192] and *P. tricornutum*[172]. *T. pseudonana* cultures grown at 35 PSU reached a maximum cell density of ~1.8 $x10^6$ cells/ml and 1.86 $x10^8$ μm^3/ml after 10 days of incubation. In contrast, *T. pseudonana* cultures grown at 1 PSU reached a maximum of 1 $x10^6$ cells/ml cell density and 6.49 $x10^7$ μm^3/ml after 15 days. Similarly, cultures of *P. tricornutum* at 35 PSU and 5 PSU reached a cell density of 3 $x10^6$ cells/ml and ~2 $x10^8$ μm^3/ml after 10 days and 12 days respectively.

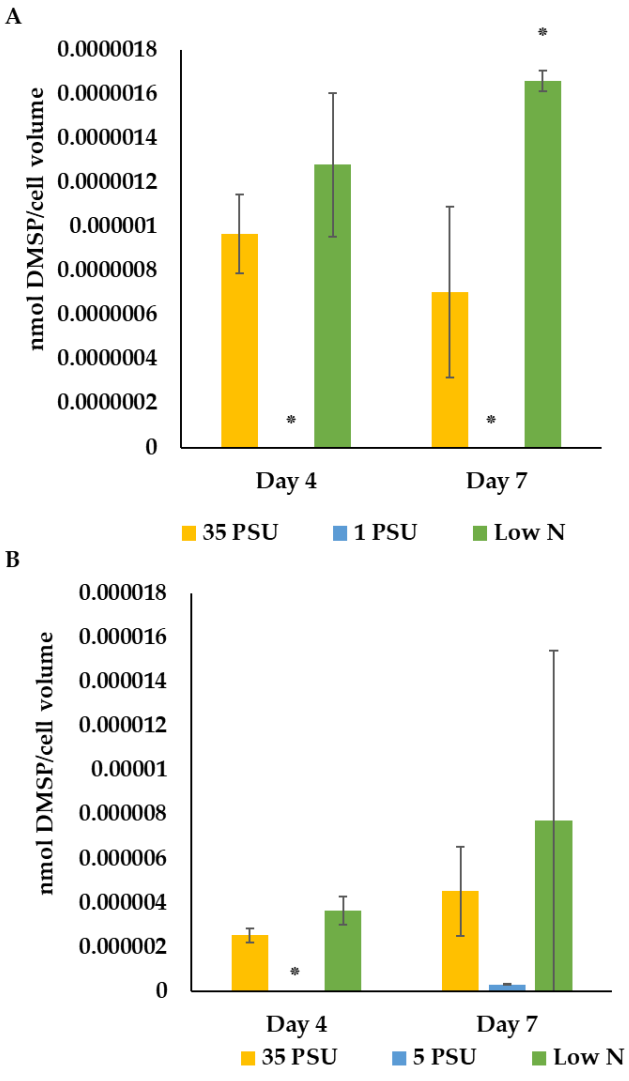
Table 3-1. Growth rates of cultures of *T. pseudonana* CCMP 1335 and *P. tricornutum* CCAP 1055/1. Cultures grown in F/2 medium mixed in artificial seawater in standard salinity and nitrogen (35 PSU, 882 µM of NO_3^-), low salt (*T. pseudonana* 1 PSU, *P. tricornutum* 5 PSU, 882 µM of NO_3^-) and low N concentration (35 PSU and 30 µM of NO_3^-).

	Standard	Low salt	Low N
T. pseudonana	0.41	0.34	0.31
P. tricornutum	0.66	0.51	0.59

Changes in salinity caused a disparity in the growth of both model diatoms. Intracellular DMSP concentration was monitored in parallel to verify that salinity also influenced osmolyte composition. DMSP content was measured by GC (see Chapter 2) in the early exponential phase and in mid-exponential phase, which corresponded with day 4 and day 7 of incubation as indicated with a * and a black arrow in (Figure 3.1). DMSP concentration was significantly lower in *T. pseudonana* cultures grown at decreased levels of salt. However, only DMSP concentration of *P.*

tricornutum cultures grown at low salt and sampled in Day 4 were statistically significantly lower than the control (Figure 3.2). Intracellular concentration of DMSP in *T. pseudonana* cells grown at 1 PSU was undetectable in Day 4 and Day 7 and statistically different to the DMSP concentration from the standard cultures. On the other hand, intracellular concentration of DMSP in *P. tricornutum* cells grown at low

Figure 3.2. Intracellular DMSP concentration of (A) *T. pseudonana* CCMP 1335 and (B) *P. tricornutum* CCAP 1055/1. Cultures grown in F/2 medium mixed in artificial seawater in standard salinity and nitrogen (PSU 35), low salinity and standard N concentration (PSU 1 and PSU 5), or at normal salinity and low N concentration (Low N). Standard cultures contain 882 µM of NO_3^- and low N 30 µM of NO_3^-. Results are shown as means ± standard deviation of 3 independent cultures. T-tests were performed and * indicates significant difference (p ≤0.05).



salinity sampled in day 4 was undetectable. In Day 7, however, DMSP concentration was 16-fold lower but statistically not significantly different to the DMSP concentration from standard cultures ($p > 0.05$) (Figure 3.2).

3.2.3 Growth in nitrogen depleted conditions

The effect limited nitrogen had on growth and metabolite production by *P. tricornutum* and *T. pseudonana* was also studied. From the previously described starter cultures grown in standard conditions, 2×10^4 cells were inoculated, in triplicates, into low nitrogen media (see chapter 2) which contains only 30 μM of NO_3^- instead of 882 μM of NO_3^- present in the standard media as described by Hockin *et al.*, 2012[181].

Regular monitoring of the cell growth showed that low nitrogen cultures reached stationary phase significantly sooner than standard cultures (Figure 3.1), likely caused by the depletion of N in the media. Both diatoms also exhibited a lower growth rate when concentration of N was lower than the standard (Table 3-1). To corroborate that N was limiting the growth, 882 μM of NO_3^- was added to the media after sampling in Day 7(marked with a ↓ in Figure 3.1). Indeed, once nitrogen was added back to the media, depleted cultures started to grow once again and quickly recovered until matching the growth of standard cultures (Figure 3.1).

The F_v/F_m value, which reflects the photosynthetic capacity of the cells, is used as diagnostic of the health of diatom (values ≥0.6) and other phytoplankton cultures[193]. The F_v/F_m of stressed cultures is considerably lower than the one for healthy and non-stressed cultures (≤0.6)[181] Thus, this value could provide further evidence that the low nitrogen cultures were stressed as a result of nitrogen depletion. Unfortunately, due to technical issues, F_v/F_m values were not possible to be obtained during these experiments. However, the stress of the cells could be also assed visually. Standard cultures and low nitrogen cultures before reaching stationary phase and with nitrogen added back to the media appeared a healthy green-brown colour derived from the active photosystems. In contrast, the low nitrogen cultures at the onset of N

limitation lost the colour, likely due to the effect of nitrogen deficiency has on the photosynthetic apparatus[194] (data not shown). Noteworthy, as reflected in the introduction of this chapter, growth patterns of standard cultures and reduced nitrogen cultures agree with previous studies[181,195].

Intracellular DMSP concentration was also measured by GC (see Chapter 2) at different stages of the growth curves, sampling points are as above (Figure 3.1). Figure 3.2 illustrates the intracellular DMSP concentration of both *T. pseudonana* and *P. tricornutum* cultures in those two time points. As expected from previous studies[129,181], both model diatoms grown under the low nitrogen treatment showed comparable intracellular DMSP concentrations to those of the standard cultures in early exponential phase. However, *T. pseudonana* cultures grown with lower concentrations of N had a significantly different concentration of DMSP (22.5-fold higher) than the cultures grown in standard conditions in the samples taken on Day 7. On the other hand, DMSP concentration in *P. tricornutum* cultures grown at low N did not present a statistically significant increase compared to the standard cultures ($p > 0.05$). This could be caused by a slow response to N deficiency by *P. tricornutum*. In order to test this, *P. tricornutum* was grown again in standard conditions and in low nitrogen conditions but, this time, samples were taken 5 days after the low N cultures entered stationary phase. Cultures were sampled after a longer exposure to nitrogen limitation and DMSP concentration in *P. tricornutum* was analysed by LC/MS (see "3.3.4 DMSP production under different conditions").

3.2.4 Other conditions

Other conditions known to affect DMSP production in diatoms and specifically in *T. pseudonana* and *P. tricornutum* cultures are temperature and oxidative stress[78,114]. With help from PhD student Peter Paolo Rivera, *T. pseudonana* and *P. tricornutum*'s responses to heat shock and temperature acclimation along with responses to

oxidative stress induced by low iron, low CO_2 or addition of H_2O_2 were also investigated.

3.2.4.1 Temperature

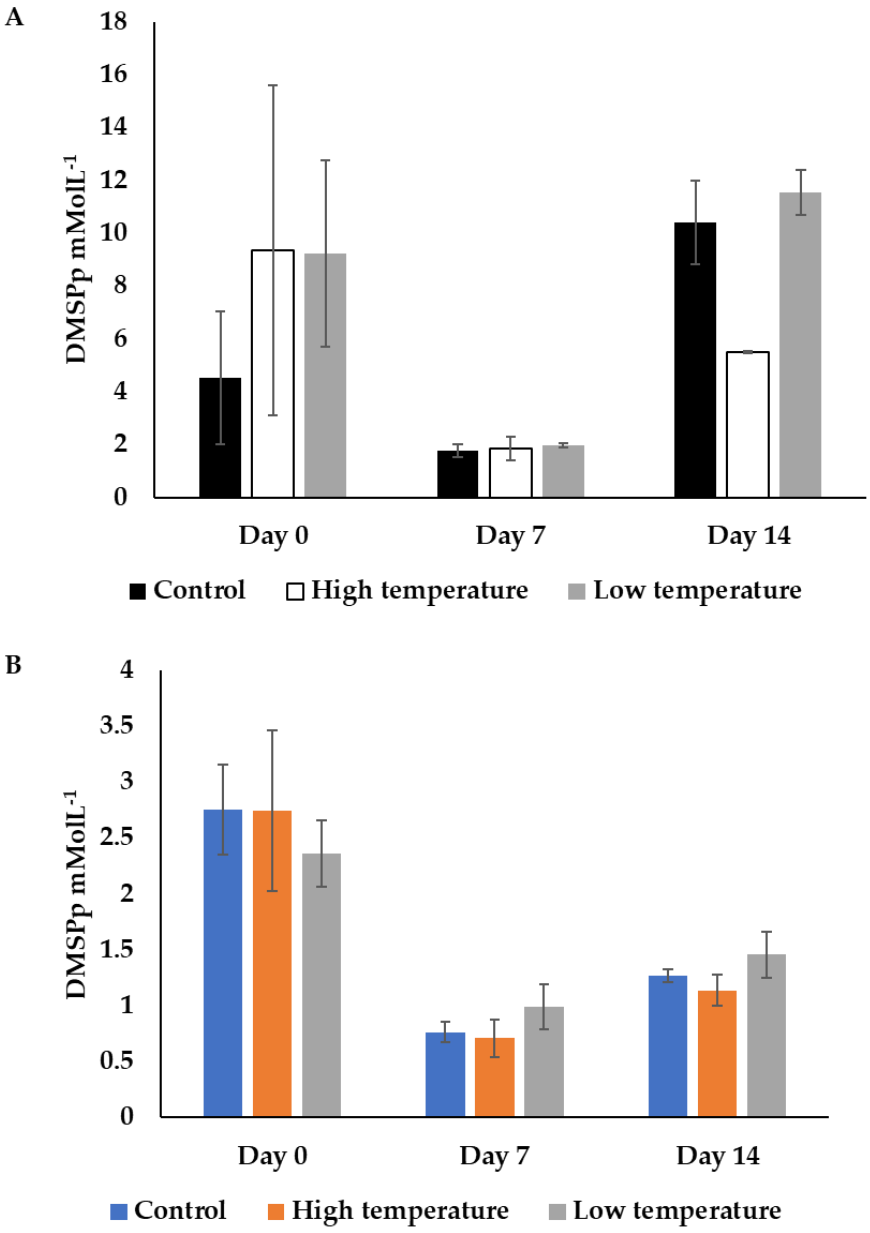
Changes in temperature have been previously linked to DMSP production in other the sea-ice diatom *Fragilariopsis. cylindrus*[98]. *P. tricornutum* and *T. pseudonana* cultures in triplicates were switched from the standard temperature to either the lowest or the highest temperatures the model strains could withstand. Hence, *T. pseudonana* was transferred from 22 °C to 8 °C and to 30 °C and *P. tricornutum* was transferred from 22 °C to 12 °C and 27 °C. Samples for cell growth and DMSP measurements were taken on the first day (Day 0), after 7 days (Day 7) and after 14 days (Day 14) (Figure 3.3).

No significant difference was found in the intracellular DMSP concentrations of any of the temperature treatments in either *T. pseudonana* or *P. tricornutum* cultures (Figure 3.3). As temperature did not seem to regulate DMSP production in the tested strains, it is possible that that this osmolyte is likely playing other roles in the cell rather than protection against temperature changes. Due to the lack of regulation observed here this condition was discarded for future experiments.

3.2.4.2 Oxidative stress

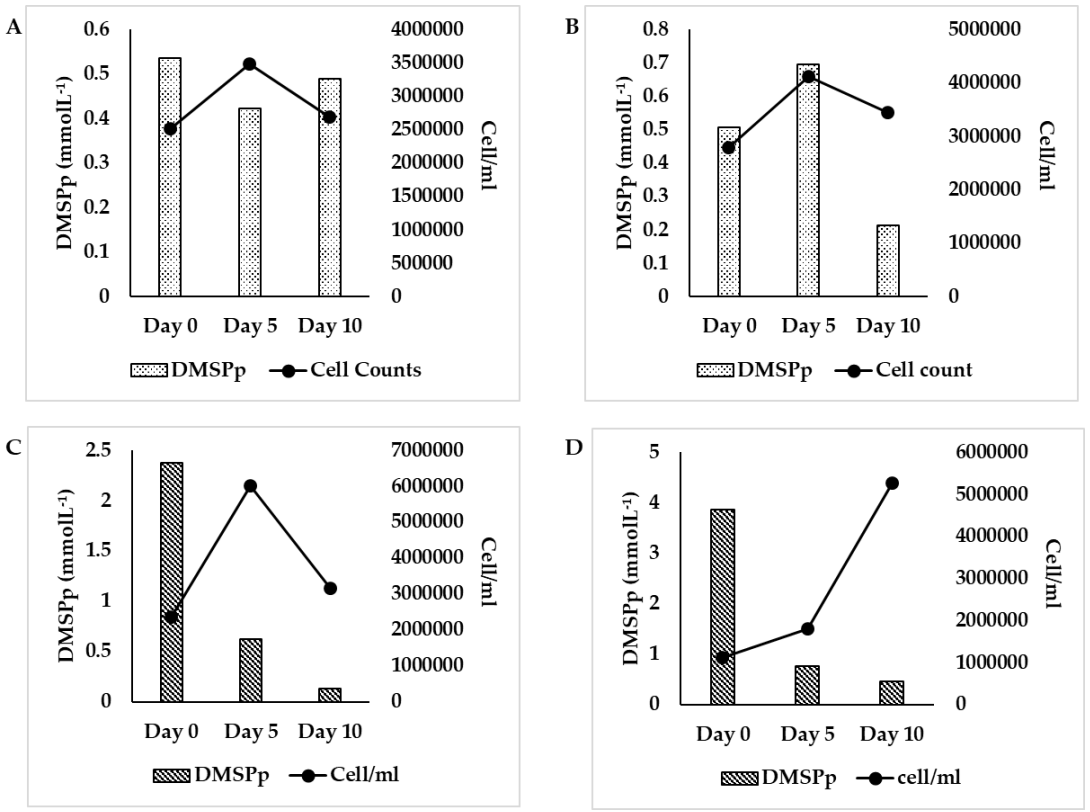
Studying the effects of oxidative stress in diatom cultures is challenging as the causes of this stress are multiple and most are derived from an imbalance of the photosystems[78]. For instance, Sunda *et al.* 2002[78] lowered the CO_2 levels or the Fe levels of the cultures, this reduction caused an increase in the accumulation of reactive oxygen species such as H_2O_2. They measured the APX activity (an enzyme involved in the removal of H_2O_2) to confirm the cells were, indeed, undergoing oxidative stress. With the intention to replicate those results, both conditions were tested in a trial run as specified in the methods chapter (see Chapter 2). In this case, a single replicate was used due to the complexity and large number of samples. None

Figure 3.3. Concentration of DMSP particulate (mMolL^{-1}) per L of culture of (A) *T. pseudonana* CCMP 1335 and (B) *P. tricornutum* CCAP 1055/1. Cultures grown in F/2 medium mixed in artificial seawater in standard salinity and nitrogen, (A) *T. pseudonana* at 22 °C (control), 30 °C (high temperature) and 8 °C (low temperature), (B) *P. tricornutum* at 22 °C (control), 27 °C (high temperature) and 12 °C (low temperature). Results are shown as means ± standard deviation of 3 independent cultures.



of the preliminary results show an increase in DMSP concentration (Figure 3.4). These results are conflicting with those already published[78], but it is important to notice that

Figure 3.4. Concentration of DMSP particulate (mMolL^{-1}) and cell count of (A, B) *T. pseudonana* CCMP 1335 and (C, D) *P. tricornutum* CCAP 1055/1. Cultures grown in F/2 medium mixed in artificial seawater in standard conditions and (A) *T. pseudonana* and low CO_2, (B) *T. pseudonana* and low Fe, (C) *P. tricornutum* and low CO_2, (D) *P. tricornutum* and low Fe.

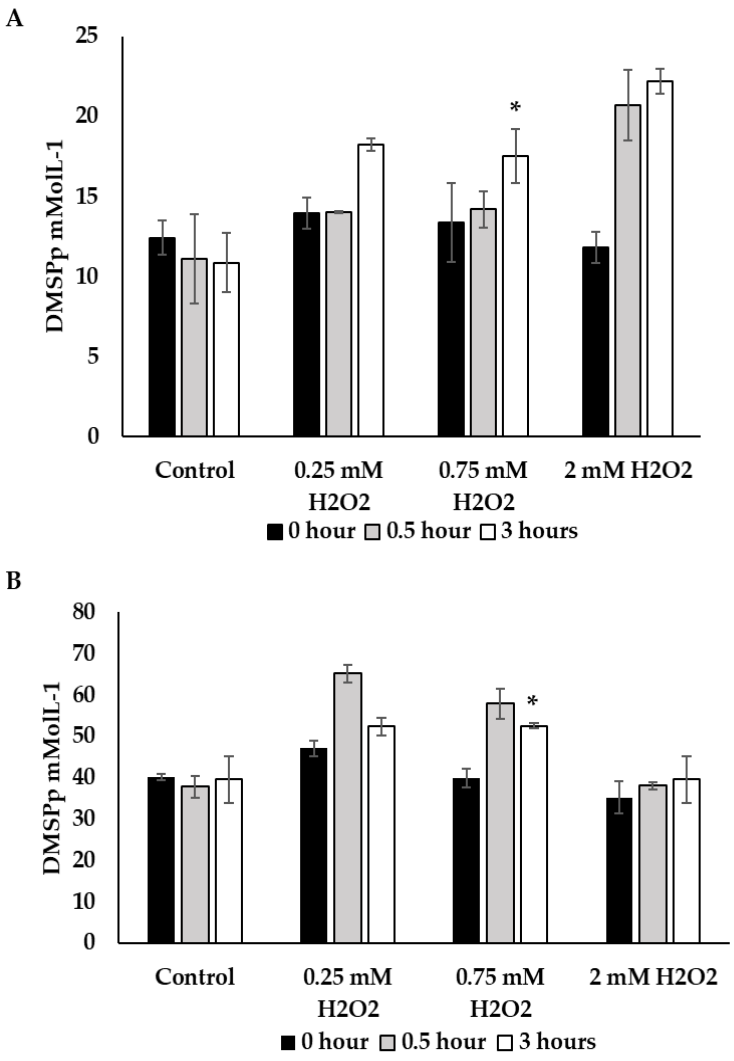


conclusions cannot be drawn from single replicate experiments.

An alternative approach was taken to evaluate the effect of oxidative stress in the model diatoms. In a second experiment, *P. tricornutum* and *T. pseudonana* cultures were grown in triplicates and none, 0.25 mM, 0.75 mM or 2 mM of H_2O_2 were added to the media. Samples were taken immediately after the addition of H_2O_2, after 0.5 h and 3 h to measure the growth of the samples and the intracellular DMSP concentration.

DMSP concentration in *T. pseudonana* and *P. tricornutum* samples taken 3 h after the addition of 0.75 mM of H_2O_2 showed a small yet statistically significant increase of intracellular DMSP, 1.61-fold and 1.24-fold respectively, compared to the control samples taken after 3h but with no addition of H_2O_2 (Figure 3.5). These significant

Figure 3.5. Concentration of DMSP particulate (mMolL^{-1}) (A) *T. pseudonana* CCMP 1335 and (B) *P. tricornutum* CCAP 1055/1. Cultures grown in F/2 medium mixed in artificial seawater in standard conditions and no addition (control) of H_2O_2 or with addition of 0.25 mM, 0.75 mM or 2 mM of H_2O_2. Samples were taken immediately after the addition of H_2O_2 (0 h), after 0.5 h and after 3 h. Results are shown as means ± standard deviation of 3 independent cultures, statistical differences are marked with * ($p \leq 0.05$).



differences suggest that DMSP concentration in these two model diatoms might indeed be regulated by oxidative stress as shown by Sunda *et al.*[78]. The inconsistency of the results obtained in the first experiment with those published by Sunda *et al.*[78] are likely due to using a single replicate. Unfortunately, results of the second experiment are not directly comparable to those performed by Sunda *et al.*[78], as the procedures to induce oxidative stress are different. It would be interesting to get these experiments done again by a different laboratory to ensure the repeatability of the results and the accuracy of the conclusions. Furthermore, as the purpose of the experiments described in this section was to identify the conditions that clearly and consistently regulated DMSP production, hence, oxidative stress was not investigated further.

3.3 DMSP and GBT production by *T. pseudonana* and *P. tricornutum* under different environmental conditions

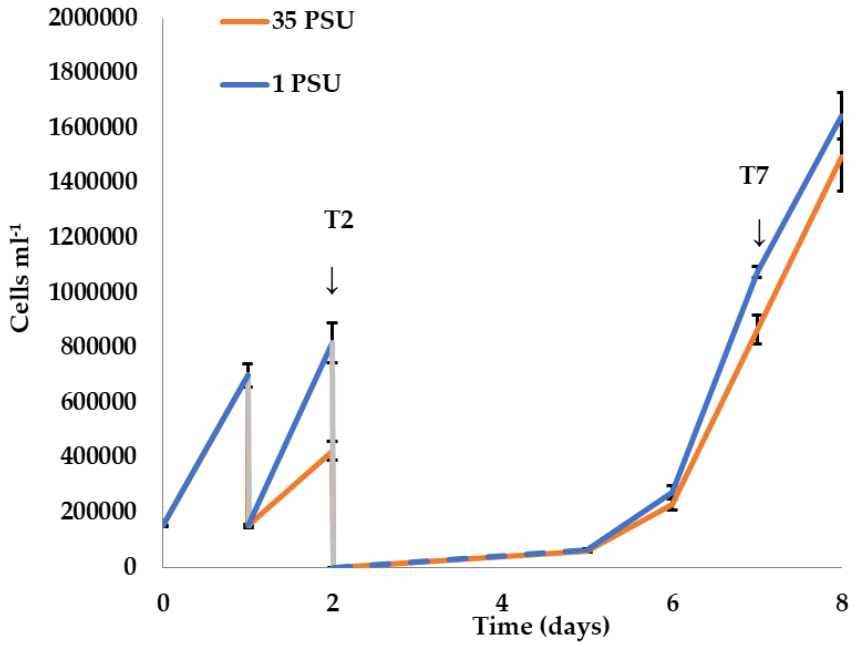
3.3.1 Metabolite analysis of *T. pseudonana* by ^{1}H-NMR

The following experiments were done to analyse the abundance of osmolytes accumulated within *T. pseudonana* in response to changes in salinity and nitrogen levels.

3.3.2 Salinity shift experiment

Working with Dr Curson and Dr Lyon, *T. pseudonana* was subjected to a salinity shift from low salinity (1 PSU) to normal salinity (35 PSU). Starter cultures of *T. pseudonana* were grown in artificial seawater (NEPC) and low salinity from an initial cell density of $2x10^4$ cells/ml. After 24 hours, cells from the starter cultures were transferred to fresh media containing either low or normal concentration of salt with a final cell density of $2x10^4$ cells/ml. After 24 hours from the salinity shift (T2) the cultures were diluted again. Diluted samples were left to grow until they reached exponential

Figure 3.6. Cell count (cell/ml) of *T. pseudonana* cultures subjected to a salinity shift. *T. pseudonana* cultures grown in artificial seawater (NEPC) and low salinity (PSU 1) then transferred to low or normal salinity (35 PSU). After 24 hours from the salinity shift (T2) the cultures were diluted again. Diluted samples were left to grow until they reached exponential growth phase (T7). Results are shown as means ± standard deviation of 4 independent cultures.



growth phase (T7 and sample point). Samples from T7 were taken for whole transcriptome analysis and metabolomics analysis to elucidate long term responses to salinity shifts and to find genes of interest. *T. pseudonana* growth was monitored 24 hours after sampling to confirm that the cells were in exponential growth phase (Figure 3.6).

RNA extracted from the samples was sent for sequencing following the Illumina HiSeq 2000 system. RNA sequencing results will be used in Chapter 4 to identify genes involved in synthesis and transport of GBT and DMSP.

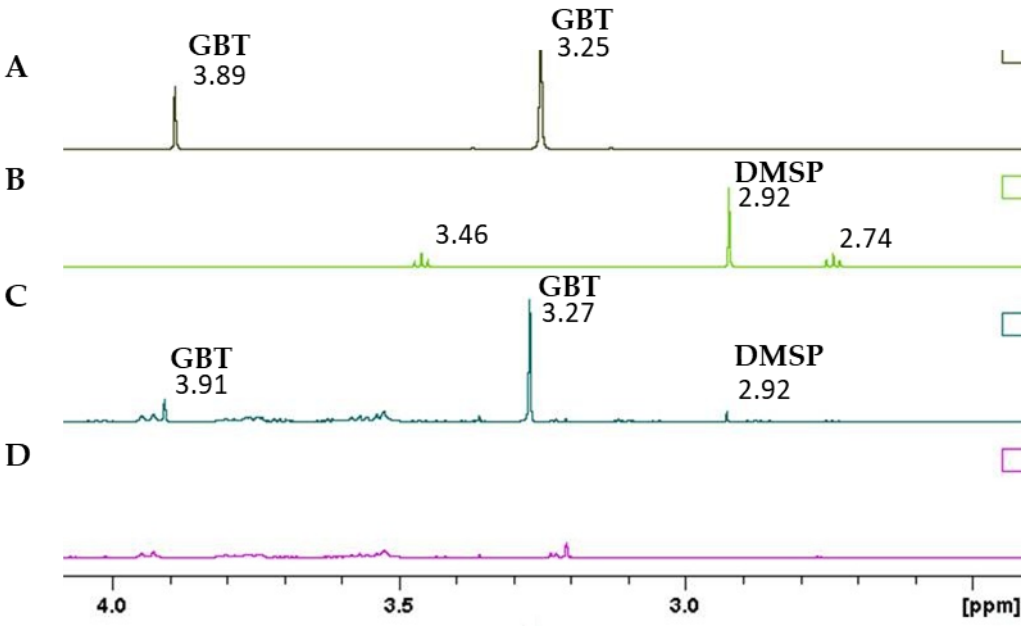
Metabolites extracted from T7 samples were analysed by ^{1}H NMR to identify the major osmolytes. The most abundant osmolytes found in samples grown at normal salinity (35 PSU) were GBT followed by DMSP. Proline and homarine were also

present but in lower concentrations. On the other hand, no osmolytes were detected in samples grown at low salinity (Figure 3.7). Presence of GBT and DMSP in normal salinity has been previously reported[114]. Furthermore, the accumulation of DMSP in cells grown in standard conditions and the absence of this compound in the low salinity cultures supports the results found in the section "3.2.2 Growth under different salinities" (Figure 3.2). Accumulation of GBT and DMSP in increased osmotic stress supports the role of these compatible solutes as osmoprotectants.

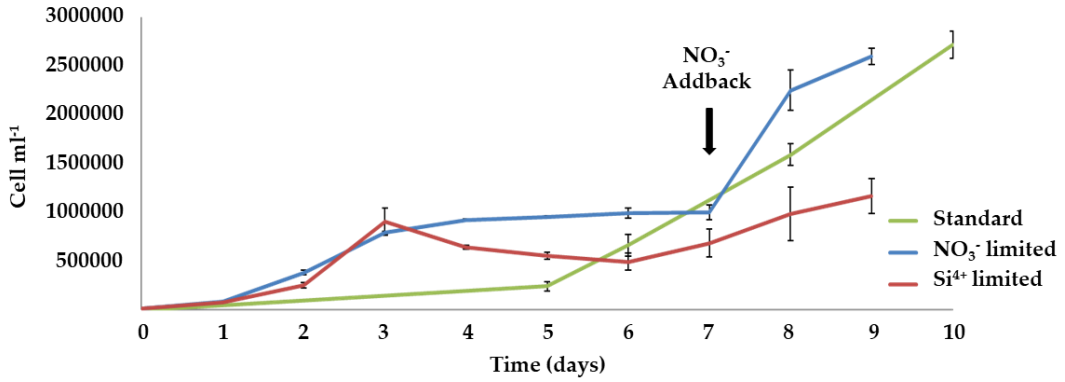
3.3.3 Nutrient limitation experiment

Nutrient limitation experiments were designed and performed by Dr Strauss. To investigate the osmolytes and genes regulated by nutrient depletion, starter cultures grown in standard conditions to exponential growth phase were transferred to 2L of fresh artificial seawater (NEPC) in standard conditions or with limiting concentrations of nitrate (10 μM NO_3^-) or silicate (50 μM Si^{4+}). Initial cell density of Figure 3.8. Nutrient limitation experiment with *T. pseudonana* cultures.

Figure 3.7. Metabolite analysis of *T. pseudonana* T7 samples analysed by ^{1}H NMR. (A) GBT standard, (B) DMSP standard, (C) normal salinity (35 PSU) samples and (D) low salinity (1 PSU) samples.

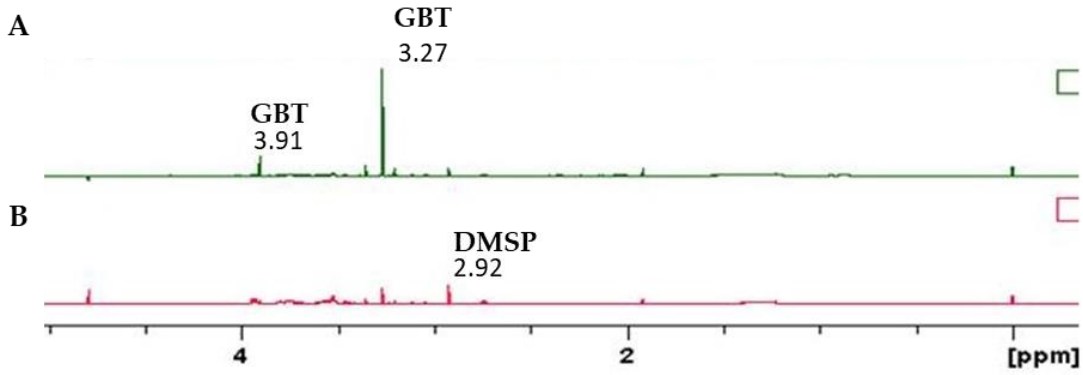


Cultures were grown in artificial seawater (NEPC) with standard and limiting concentrations of nitrate (10 µM NO_3^-) or silicate (50 µM Si^{4+}).



experimental cultures was 10^4 cells/ml. Samples were taken for RNA analysis and osmolyte analysis at T7 point which corresponded with mid-stationary growth phase of nitrate and silicate limited cultures (Figure 3.8). To ensure nitrate was limiting, standard concentrations of nitrate (882 µM NO_3^-) was added back to the media (NO_3^- addback, T7) after sampling. Indeed, cells in nitrogen depleted cultures in stationary growth phase at the moment of sampling restored their growth to comparable levels to the standard cultures after adding back NO_3^-(Figure 3.8).

Figure 3.9. Metabolite analysis of *T. pseudonana* T7 samples analysed by ^{1}H NMR. (A) Silicate limited (50 µM Si^{4+}) and (B) nitrate limited (10 µM NO_3^-) cultures.



Metabolites extracted from the silicate and nitrate limiting cultures at T7 time point were analysed by proton NMR. Silicate limited cultures were used as controls over the standard cultures to allow comparison between treatments of cultures at the same growth phase, in this case stationary. In silicate limited cultures, the predominant osmolyte was GBT, whereas DMSP was undetectable. Conversely, in cultures with depleted levels of nitrate no GBT was detected and DMSP was the predominant compatible solute (Figure 3.8) as expected from previous studies[181]. This shows the interplay of these two osmolytes depending on nitrogen availability in *T. pseudonana*

3.3.4 Metabolite analysis of *P. tricornutum* by LC/MS

In the previous growth curve (Figure 3.1), *P. tricornutum* was sampled two days after they entered stationary growth phase and it was observed that DMSP concentration was not significantly different between the standard conditions and the low nitrogen samples. Here, *P. tricornutum* was grown exactly as before, either in standard conditions, low salt (5 PSU) or limited NO_3^-. Standard and low salt cultures were sampled in mid-exponential phase (day 4) but the nitrate limitation cultures were samples 5 days after the low nitrogen samples entered the stationary growth phase (day 11). Low nitrogen samples were clearly stressed and had a bleached appearance, an addition of two times the F/2 NO_3^- concentration (882 μM) restored the normal growth. Increasing the length of the nitrogen starvation had a significant effect on DMSP accumulation in *P. tricornutum*, as it is shown in (Figure 3.11).

In addition, *P. tricornutum* was also grown in standard conditions and addition of 0.5 mM of choline, GBT, glycine, DMSP or MTHB. Samples were collected when cultures reached mid-exponential growth phase, cell density of 2.5-5x10^6 cells/ml. These compounds are known intermediates of either the Met transamination pathway for the synthesis of DMSP or for GBT synthesis. These experiments are to examine the effect of product/substrate/intermediate addition on the relevant synthesis process,

e.g. is DMSP synthesis induced by increased availability of the methyltransferase substrate MTHB.

Samples were analysed by LC/MS using the method published by Beale *et al.*[155] (see method in Chapter 2). As well as the samples collected for metabolite, *P. tricornutum* samples were collected for transcription analysis by RT-qPCR. Furthermore, standard and low salinity treatments were taken for proteomic analysis (see Chapter 2 for methods). Proteomic and transcriptomic results will be further discussed in subsequent chapters.

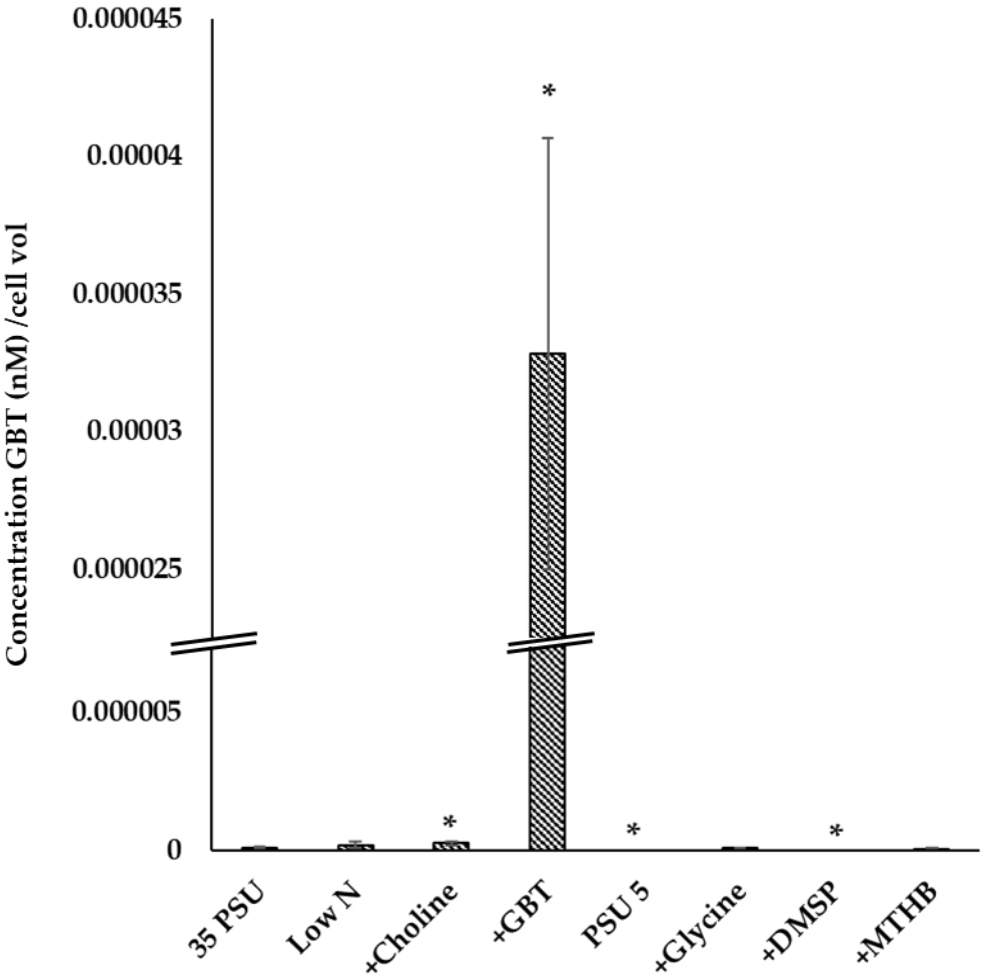
3.3.4.1 GBT production under different conditions

P. tricornutum does not produce a large amount of GBT and the production is not markedly regulated with an average concentration per cell volume of 8.66×10^{-08} nM when grown under standard conditions. GBT concentration increases marginally yet significantly by addition of choline, a precursor of GBT via the choline pathway, whereas the addition of glycine, the precursor via the alternative methylation pathway has no effect on the accumulation of intracellular GBT concentration. External addition of GBT has the greatest effect on the amount of GBT detected in the samples, this could be explained if *P. tricornutum* is actively importing GBT from the environment (Figure 3.10).

Contrary to what it is observed in other species[82,175], N availability did not have an effect on the amount of GBT accumulated within *P. tricornutum*. Conversely, the concentration of this compatible solute is lowered to undetectable levels by growth under low salinity and by the addition of DMSP to the media (Figure 3.10).

The regulation of GBT by salinity shows that GBT is likely part of a larger array of compatible solutes produced by *P. tricornutum* to overcome osmotic stress, but the low concentrations found in this diatom indicate that is not likely playing a major part unless, for example, it is concentrated in specific organelles as was shown to be

Figure 3.10. Analysis of intracellular GBT concentration in *P. tricornutum* CCAP 1055/1 by LC/MS. Cultures grown in F/2 medium mixed in artificial seawater in standard salinity and nitrogen (35 PSU), initial concentration of 30 µM of NO_3^- (low N), low salt (5 PSU) and addition of 0.5 mM of choline, GBT, glycine, DMSP or MTHB. Results are shown as concentration of GBT per cell volume means ± standard deviation of 3 independent cultures, statistical differences are marked with * (p ≤0.05).

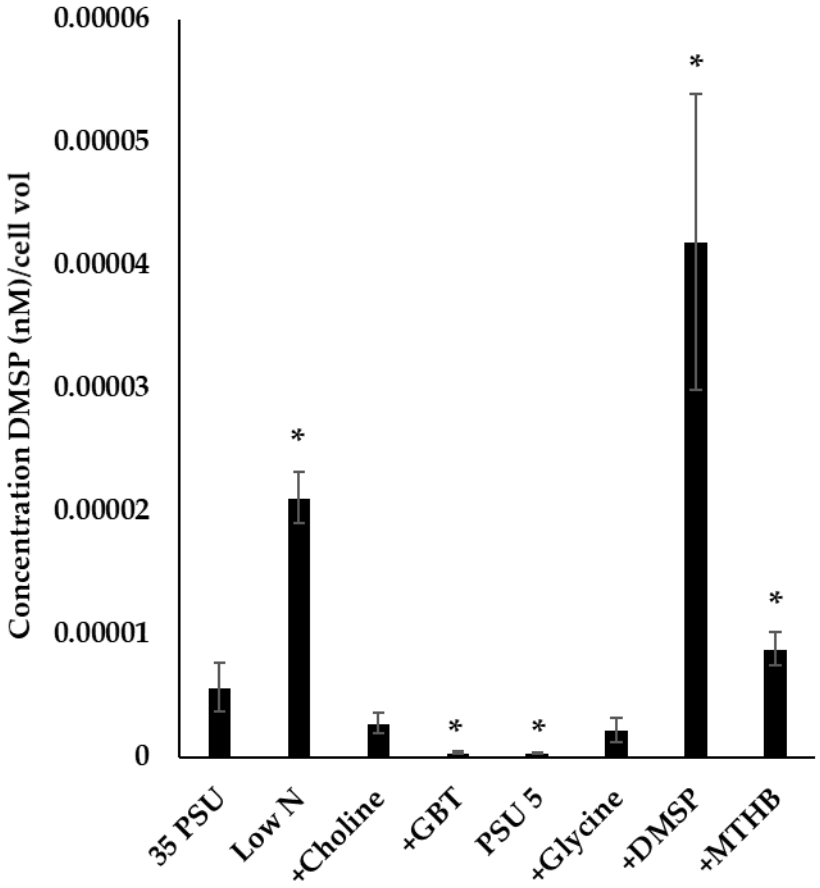


the case for DMSP in *P. parvum*[99]. For instance, a previous study suggested that proline is the main osmolyte followed by DMSP[175]. The suppression of GBT synthesis by the addition of DMSP to the media indicates that these two osmolytes might be playing a similar role in the cell and DMSP can fulfil the role GBT had in the cell they can be substitute one for the other.

3.3.4.2 DMSP production under different conditions

Intracellular DMSP concentration in *P. tricornutum* is far higher than GBT with an average of 5.67293 x10^{-06} nM per cell volume in standard conditions. Addition of the DMSP precursor MTHB significantly increases the amount of DMSP in the cell, in agreement with diatoms producing DMSP through the transamination pathway[96] and with these organisms being able to internalise this compound[99]. Likewise, adding

Figure 3.11. Analysis of intracellular DMSP concentration in *P. tricornutum* CCAP 1055/1 by LC/MS. Cultures grown in F/2 medium mixed in artificial seawater in standard salinity and nitrogen (35 PSU), initial concentration of 30 µM of NO_3^- (low N), low salt (5 PSU) and addition of 0.5 mM of choline, GBT, glycine, DMSP or MTHB. Results are shown as concentration of DMSP per cell volume means ± standard deviation of 3 independent cultures, statistical differences are marked with * (p ≤0.05).



DMSP to the media significantly increases the concentration of accumulated DMSP, showing the ability of *P. tricornutum* to uptake DMSP from the media[131]. Furthermore, sampling the N limited cultures 5 days after entering stationary growth phase had a significant effect on DMSP synthesis increasing it 3.73-fold (Figure 3.11).

There are two conditions that significantly decreases the synthesis of DMSP, these are addition of GBT to the media and growth in lowered salinity. The DMSP accumulated (Figure 3.11) was reduced by 13-fold when GBT was available. In addition, DMSP levels were also downregulated by 15.77-fold when *P. tricornutum* was grown under low salinity compared to normal standard conditions. The reciprocal relationship between GBT and DMSP as well as the regulation of DMSP synthesis by salinity further supports the role of DMSP as an osmoprotectant in *P. tricornutum*.

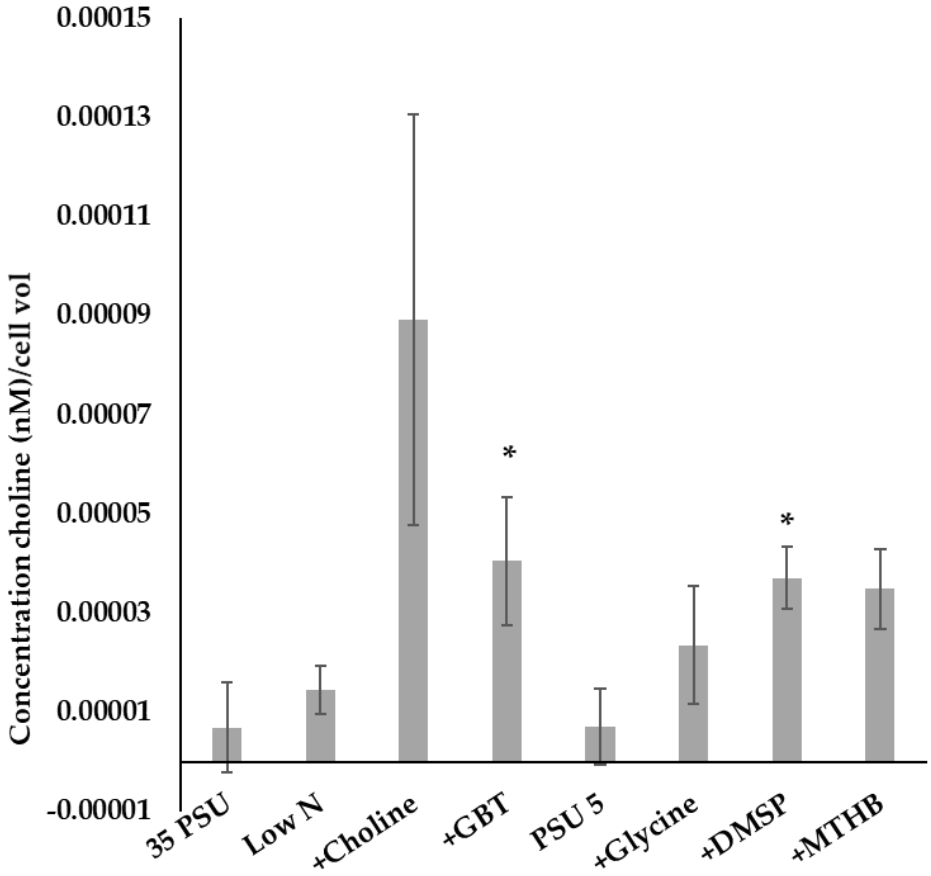
3.3.4.3 Choline production under different conditions

Choline is accumulated in the cell in an average concentration of 6.98 $x10^{-6}$ nM per cell volume in standard conditions, which is 10^2 times higher than the concentration of GBT and in a comparable amount to DMSP. Adding GBT and DMSP to the media had the greatest effect on the intracellular concentration of choline. Choline accumulates in the cell 5.8-fold more when GBT is added to the media and 5.3-fold more when DMSP is added to the media (Figure 3.12). If GBT is added to the media, then the uptake might override the synthesis of GBT so choline would accumulate rather than being used for GBT synthesis. This same phenomenon can explain why choline is accumulated when DMSP is added to the media, as GBT synthesis is downregulated so choline accumulates.

3.4 Discussion and concluding remarks

The two model diatoms *T. pseudonana* CCMP 1335 *and P. tricornutum* CCAP 1055/1 were grown in standard conditions (35 PSU), at low salinity (1 PSU, *T. pseudonana* or 5 PSU, *P. tricornutum*) and with low initial concentration of nitrate. Their growth was

Figure 3.12. Analysis of intracellular choline concentration in *P. tricornutum* CCAP 1055/1 by LC/MS. Cultures grown in F/2 medium mixed in artificial seawater in standard salinity and nitrogen (35 PSU), initial concentration of 30 µM of NO_3^- (low N), low salt (5 PSU) and addition of 0.5 mM of choline, GBT, glycine, DMSP or MTHB. Results are shown as concentration of choline per cell volume means ± standard deviation of 3 independent cultures, statistical differences are marked with * ($p \leq 0.05$).



monitored regularly, and their metabolite content was analysed. Furthermore, changes in metabolite production caused by the addition of DMSP, GBT and their precursors to *P. tricornutum* standard cultures were explored. Finally, the effect of subjecting *T. pseudonana* and *P. tricornutum* cultures to high/low temperature and oxidative stress in the intracellular DMSP concentration was explored.

Previously reported DMSP concentration in *T. pseudonana* and *P. tricornutum* standard cultures was 0.005 pgcell^{-1} and 0.0129 pg cell^{-1}, repectively[196]. Here, we report an average concentration of 0.1286pg cell^{-1} volume in standard cultures of *T. pseudonana* and 0.003752 pg cell^{-1} volume in *P. tricornutum* cultures (Figure 3.2).

3.4.1 Changes in salinity affect the growth and metabolite synthesis in *T. pseudonana* and *P. tricornutum*.

Lowered salinity has been shown to affect the growth rate of these two model diatoms[171,172]. These results were replicated in our growth curves in which both diatoms showed lower growth rate when grown at low salt concentration compared to normal salinity (Table 3-1).

Metabolite production was also monitored, in mid- and late exponential growth phase. *T. pseudonana* did not produce any DMSP in reduced salinity in mid- or late exponential growth phase (Figure 3.2), nor GBT in mid-exponential growth phase (Figure 3.7). The absence of GBT and DMSP in cells grown in low osmotic pressure was previously reported[128,174], and it supports the hypothesis of these two molecules playing a role as osmoprotectants in *T. pseudonana*.

On the other hand, the concentration of DMSP and GBT in *P. tricornutum* cultures grown at low salinity and mid-exponential growth phase was 15.77-fold lower and undetectable, respectively, compared to standard cultures (Figure 3.10, Figure 3.11). In late exponential growth phase, however, the intracellular concentration of DMSP was not statistically different to the control samples (Figure 3.2).

3.4.2 Nitrogen availability conditions the growth rate and metabolite synthesis in *T. pseudonana* and *P. tricornutum*.

T. pseudonana and *P. tricornutum* present lower growth rates in cultures with reduced N in the media (Figure 3.1). The regulation of the intracellular concentration of the metabolites when nitrogen is limiting differ between *T. pseudonana* and *P. tricornutum*.

Previous studies showed that the concentration of DMSP was higher in *T. pseudonana* cultures under nitrogen limitation and the intracellular concentration of GBT was significantly lower or undetectable[175]. This reciprocal pattern between DMSP and GBT accumulation was also observed in our experiments when comparing cultures at the onset of stationary growth phase caused by nitrate depletion and cultures in stationary phase caused by silica scarcity (Figure 3.9).

Nitrogen availability seems to affect *P. tricornutum* in a different manner than *T. pseudonana.* For instance, no statistical difference was found in the intracellular concentration of DMSP in *P. tricornutum* low nitrogen cultures at the onset of stationary phase compared to the mid-exponential phase of the standard cultures (Figure 3.2). The difference becomes statistically significant if low nitrate *P. tricornutum* cultures are exposed to nitrogen limitation for an extended period of time (3.73-fold higher than standard cultures), such as 5 days after entering stationary growth phase (Figure 3.11). In previous studies, an accumulation of DMSP was also observed in cultures maintained under N starvation[129,186]. Besides, the intracellular concentration of GBT and choline does not change significantly when N is depleted. This is not entirely surprising as GBT is not amongst the most relevant osmolytes in *P. tricornutum*. In fact, N availability has an effect in a different nitrogen-based compound, proline, which is the main osmolyte in standard conditions and significantly downregulated when nitrogen is scarce[129].

3.4.3 Temperature changes do not affect DMSP synthesis, but oxidative stress causes an increase in DMSP production in the model diatoms.

3.4.3.1 Temperature does not affect the intracellular DMSP concentration of *T. pseudonana* and *P. tricornutum* grown in batch cultures.

In 2012, Spielmeyer and Pohnert reported that the intracellular concentration of GBT was directly proportional to temperature, and that there was an inverse correlation between temperature and intracellular DMSP[114]. These findings are contrary to the results obtained in the experiments carried out as part of this thesis. Here, cultures of *T. pseudonana* and *P. tricornutum* kept at the standard temperature (22 °C) were either kept at the standard temperature (control) or transferred to low or high temperatures, 8 °C and 30 °C for *T. pseudonana*; 12 °C and 27 °C for *P. tricornutum.* Then samples were taken to measure the cell density and the DMSP particulate concentration on the first day, after seven days and after 14 days. None of the measurements yielded a significant decrease in DMSP content at elevated temperature nor they showed any changes at lowered temperature (Figure 3.3). Nonetheless, it is important to consider the differences in the experimental set up. Whilst our experiments were carried out in batch culture and at different growth phases, Spielmeyer and Pohnert grew the diatoms in incubators keeping the cells in exponential growth phase over seven generation before sampling for metabolite analysis[114].

3.4.3.2 H_2O_2 induced oxidative stress triggers an increase in intracellular DMSP concentration in *T. pseudonana* and *P. tricornutum.*

An attempt to replicate Sunda *et al.* experiments, in which *T. pseudonana* and *P. tricornutum* cultures were grown in low Fe or low CO_2 induced oxidative stress[78], was made. Sunda *et al.*, reported an increase in the concentration of DMSP suggesting a role as an antioxidant in these organisms[78]. Test run of the experiment performed and described in this chapter showed no difference in DMSP concentration between the control samples and those under oxidative stress induced by low Fe or low CO_2 (Figure 3.4). Consequently, a different approach was taken to study the effect of the accumulation of reactive oxygen species in the production of DMSP in the model diatoms.

The new approach involved the addition of different concentrations of H_2O_2 to standard cultures of *T. pseudonana* and *P. tricornutum.* Samples were taken to measure

the cell density and DMSP particulate at 0, 0.5 and 3 hours. In both species, samples taken 3 hours after the addition of 0.75 mM H_2O_2 showed a statistically significant increase in the concentration of DMSP. It is also worth mentioning that cultures exposed to 2 mM H_2O_2 did not survive. It is likely that lower concentrations of H_2O_2 did not cause a significant stress in the cells and that it took as long as 3 hours to the cells to accumulate a significant amount of DMSP (Figure 3.5).

3.4.4 Exogenous GBT, DMSP and their precursors affect metabolite composition in *P. tricornutum*.

P. tricornutum cultures were grown in standard condition and in standard condition with 0.5 mM of DMSP, GBT, choline, glycine and MTHB added to the media. Samples were taken to measure the concentration of DMSP, GBT and choline per cell by LC/MS.

Addition of DMSP increases significantly the concentration of DMSP (Figure 3.11) and choline (Figure 3.12) in the cell, whereas it decreases the concentration of GBT (Figure 3.10). The increase of intracellular concentration of DMSP illustrates the ability of *P. tricornutum* to uptake this metabolite from the media[131]. On the other hand, the accumulation of choline in the cell could be explained if in this diatom, GBT was synthesised via the choline pathway. Thus, the decrease in intracellular GBT when DMSP is exogenously supplied would cause the accumulation of choline. The effect of DMSP on GBT concentration does not occur in the other model diatom *T. pseudonana*[130].

Indeed, the addition of choline increases the concentration of GBT (Figure 3.10) supporting the hypothesis of choline being the precursor of this compatible solute. Moreover, if GBT is added to the media, *P. tricornutum* exhibit the ability to also uptake this metabolite from the media[131] (Figure 3.10). The accumulation of GBT in the cell leads to a rise in the accumulation of choline (Figure 3.12). Being the precursor of GBT, if this molecule is in excess in the cell the synthesis could be downregulated

causing the accumulation of the precursor choline. Conversely, the abundance of GBT in the cell could activate its degradation generating choline[184].

In addition, exogenous GBT also has an effect on intracellular DMSP concentration which decreases significantly compared to the standard cultures. It is likely that GBT and DMSP could be playing a similar role in the cell, although it would be interesting to see how proline, the main osmolyte in the pennate diatom[129], responds to GBT and DMSP addition.

It is notable that supplying choline to the cultures did not translate in an increase in its intracellular concentration (Figure 3.12). This is likely due to its conversion to GBT or to phosphatidylcholines[184]. Contrary to what it would be expected if GBT was synthesised via the methyltransferase pathway, glycine has no effect in GBT concentration. Finally, addition of MTHB causes an increase in DMSP concentration (Figure 3.11). This is predictable as MTHB is a precursor of DMSP from the transamination pathway, which has been showed to be the main pathway in diatoms[96]. The increase in DMSP concentration as a result of the supplied MTHB was also reported by Kageyama *et al.*[101].

Chapter 4

Identification of candidate DMSP and GBT synthesis genes in DSYB lacking DMSP producing diatoms

4 Identification of candidate DMSP and GBT synthesis genes in DSYB lacking DMSP producing diatoms

4.1 Introduction

In this chapter I aim to identify genes for DMSP and GBT synthesis and transport in the model diatoms *T. pseudonana* and *P. tricornutum*.

Diatoms synthesise DMSP via the transamination pathway, the enzymes involved in this pathway, although not identified, have been characterised and they include 2-oxoglutarate-dependent aminotransferase, NADPH-dependent reductase, *S*-adenosylmethionine (SAM)-dependent methyltransferase, and oxidative decarboxylase enzymes[96,97] (Figure 1.7).

There are two possible pathways for GBT synthesis (Figure 1.3, Figure 1.4) and it has been suggested that the methyltransferase pathway is present in the diatom *F. cylindrus*[98] and *T. pseudonana*[128]. In the previous chapter, the metabolite analysis of *P. tricornutum* suggested that the pennate diatom could use primarily the choline pathway (see Chapter 3). The main enzymes in these pathways, previously described in chapter 1, include choline dehydrogenase/oxidase and betaine aldehyde dehydrogenase for the choline pathway[16] and glycine/sarcosine/dimethylglycine methyltransferase as a single enzyme or as a combination of the glycine/sarcosine methyltransferase for the first step and a sarcosine/dimethylglycine[51] or dimethylglycine methyltransferase[49] for the second. In addition, transport of DMSP and GBT have been reported in phytoplankton[131] but no transporter have been characterised yet in these organisms.

4.1.1 Types of enzymes involved in the DMSP synthesis transamination pathway.

4.1.1.1 Transamination

The transamination reaction involves a transfer of the amino group of an amino acid to a keto acid to produce a new amino acid. For example, the amino group of L-methionine (Met) can be transfer to several oxo acids including oxaloacetic acid, pyruvic acid, phenylpyruvic acid, 2-oxobutyric acid, 2-oxovaleric acid, 2-oxoheptanoic acid, oxoglutaric acid and glyoxylic acid forming 4-methylthio-2-oxobutanoate (MTOB) and using pyridoxal phosphate as cofactor[197]. The newly formed amino acid would depend on the acceptor, if the amino group is transferred to pyruvic acid it would form L-alanine[198], but if it is transferred to glyoxylic acid it would result in a glycine molecule[199]. Enzymatic assays done with algae extracts showed that the aminotransferase involved in DMSP synthesis has a strong preference for 2-oxoglutarate as the amino acceptor and the second-best amino donor after L-methionine was L-glutamate[97].

Methionine branched-chain aminotransferase 4 (BCAT4, UniProt ID Q9LE06) from *Arabidopsis* is involved in the methionine elongation pathway of aliphatic glucosinolate biosynthesis, and it has been proposed to carry out the first step, the transamination of Met to MTOB[200]. This protein sequence will be used to search for similar aminotransferases in the two model diatoms.

Another aminotransferase thought to be involved in the transamination of Met to MTOB is methionine:glyoxylate aminotransferase (MGAT), in *Brassica carinata*. MGAT was also thought to be involved in the biosynthesis of allyglucosinolate from methionine, however, in some of the nonglucosinolate producing species a MGAT isoenzyme is present[201]. Unfortunately, no sequences for these two types of MGATs were elucidated to the best of the author's knowledge.

The structure and enzymatic properties of the methionine aminotransferase, gene *ybdL* from *E. coli*, have been described and compared to the aromatic amino acid aminotransferase (ArATPh) from *Pyrococcus horikoshii.* Both of them have the ability of transaminate Met, however, the first has a much higher affinity for Met as substrate than the latter[202]. These two enzymes will also be considered as potentially involved in the initial step of DMSP synthesis.

Candidate aminotransferases have been suggested by Lyon *et al.* in *F. cylindrus* including the unknown aminotransferase (FRACYDRAFT_273803)[98]. Bromke *et al.* have also identified a candidate gene for this first step of the transamination pathway in *T. pseudonana*, THAPSDRAFT_260934 (XP_002285992.1)[175]. However, none of these have been functionally ratified.

4.1.1.2 Reduction

The second step of the transamination pathway consists in the reduction of MTOB to MTHB. The enzyme that carries out this reaction have NADP- NADH- dependent activity but it is more likely that *in vivo* the preference would be for the NADP-dependent reductase activity[97]. In *F. cylindrus* a predicted flavoprotein, NADPH-dependent FMN reductase FRACYDRAFT_173405 (OEU09668.1) was significantly upregulated 1.50-fold by hypersaline conditions. This candidate enzyme will be used to search for homologues in *T. pseudonana* and *P. tricornutum*. As for the transamination step, this candidate enzyme have not been functionally ratified.

4.1.1.3 Methylation

The *S*-adenosyl dependent methylation of MTHB to DMSHB is thought to be the committing and key step in DMSP synthesis as it is the first non-reversible step and the reaction only present in DMSP-producing organisms[97]. No homologues of the *dsyB* or *DSYB* genes are present in the model diatoms *T. pseudonana* and *P. tricornutum*. In addition, Curson *et al.* tested candidate methyltransferases identified

in *F. cylindrus* and *Acropora millepora* and showed they were not functional[99]. Finding the SAM-methyltransferase catalysing the third and essential step in DMSP biosynthesis will be one of the major focus of this thesis.

4.1.1.4 Oxidative decarboxylation

The last step of the pathway is the oxidative decarboxylation of DMSHB to DMSP. Although no function has been ratified yet, there are two proposed candidate decarboxylases identified in *F. cylindrus* , a pyridoxyl-dependent decarboxylase (FRACYDRAFT_238865) and diaminopimelate decarboxylase (FRACYDRAFT_263016)[98]. These candidate genes will be used as probes.

4.1.2 Types of enzymes involved in GBT synthesis.

4.1.2.1 Choline pathway

The catalytic step from choline to betaine aldehyde is carried out in many organisms by an oxygen and NAD^+-dependent choline dehydrogenase[37,203,204]. The choline dehydrogenase present in *E. coli*, named *betA* (WP_115194274.1), was used to find any candidate potentially catalysing this reaction in the model diatoms. *betB* was also used to identify candidate betaine aldehyde dehydrogenases, the enzyme that oxidises betaine aldehyde to GBT[205].

To cover the set of enzymes present in the choline pathway, the sequence of the choline oxidase from *Arthrobacter globiformis, codA*, (AAS99880.1) was also used to probe for an alternative enzyme in the choline pathway[206].

4.1.2.2 Glycine/sarcosine/dimethylglycine methyltransferase pathway

The methylation of glycine to GBT via sarcosine and dimethylglycine can be performed either by one single enzyme or two enzymes. In diatoms, although the pathway has not been ratified, candidate genes for one single SAM-dependent methyltransferase likely involved in GBT biosynthesis have been identified in *F.*

cylindrus[98] and *T. pseudonana*[128]. The one identified in *F. cylindrus* is a putative sarcosine-dimethylglycine methyltransferase FRACYDRAFT_ 212856 (OEU23352.1) upregulated by 2.30-fold in hypersaline conditions. The SAM-dependent methyltransferase identified in *T. pseudonana* is a predicted protein THAPSDRAFT_20797 (XP_002286764.1) 3.60-fold upregulated by high salt[128].

4.1.3 Types of enzymes involved in DMSP and GBT transport.

There are multiple transport systems involved in the uptake of GBT. Amongst all the transport systems, the main transport system is the physiological and environmentally relevant ProU[23]. This transport system composed of three cistrons (proV, proW, and proX) belongs to the family ATP-binding cassette (ABC) uptake system or traffic ATPases[29,30] and it is present in Prokaryotes[29,31,32] and Eukaryotes[33]. The ProU operon from *E. coli* has affinity for GBT and for multiple structurally related substrates such as proline[29] and DMSP[23]. ProP is a secondary H^+ symporter also present in *E. coli* with the ability to transport GBT and DMSP[23]. A different transporter shown to be capable of transporting different betaines such as GBT and DMSP is the BCCT (betaine choline carnitine transport) proteins such as DddT from *Halomonas*[133].

Lavoie *et al.* have identified putative BCCT transporter similar to DddT transporter from *Halomonas* in a number of *Thalassiosira spp.* including the putative BCCT transporter THAPSDRAFT_262307 (XP_002289511.1) in *Thalassiosira pseudonana*[207].

4.1.4 Aims of the chapter

The main aim of this chapter is to identify candidate genes encoding enzymes that could potentially be involved in the synthesis or transport of the two main osmolytes DMSP and GBT. Previously published candidate genes or genes with functions matching those of the ones involved in our pathways of interest in a range of organisms will be used as probes to search for candidate genes in the genomic database of the two model diatoms. In addition, whole transcriptomic and proteomic

sequences obtained from the cultures grown in Chapter 3 will be interrogated to find new candidate genes of interest too.

4.2 Identification of genes of interest in *T. pseudonana*

In addition to a bioinformatic approach, RNA extracted from *T. pseudonana* grown in high/low salinity or N deplete or replete conditions was sequenced. Transcript abundance from the low salinity treatment were compared to those from the high salinity and the N depleted to the N replete treatment to identify transcripts regulated in the same way as DMSP and/ or GBT is produced. This was done by searching the resulting data for genes of interest and to provide an insight into the candidate genes regulation.

4.2.1 Analysis of genes potentially involved in DMSP synthesis.

4.2.1.1 Transamination

Using BLASTp homologues of BCAT4 from *Arabidopsis*, YbdL from *E. coli* and O59096 from *P. horikoshii* were identified in *T. pseudonana*. Three aminotransferases were found in *T. pseudonana* genome with very low similarity to the probed sequences, THAPSDRAFT_21813 (XP_002288882.1) had an e value of $-1e^{-64}$ and 35% identity to BCAT4, THAPSDRAFT_27811 (XP_002289692.1) had an e value of $-1e^{-52}$ and 31% to YbdL, whereas THAPSDRAFT_12030 (XP_002295148.1) had an e value of $-7e^{-77}$ and 38% identity to O59096 from *P. horikoshii.*

THAPSDRAFT_21813 is a predicted protein annotated as a branched-chain amino acid aminotransferase, classified as a Pyridoxal 5'-Phosphate Dependent Enzyme class IV (PLPDE_IV). Its closest homologues with e values of 0 are hypothetical protein THAOC_30858 (EJK50198.1) from *T. oceanica* 68% identity, predicted protein PHATRDRAFT_50793 (XP_002178511.1) from *P. tricornutum* 62% identity, branched-chain amino acid aminotransferases FisN_5Lh102 (GAX13983.1) and *Fistulifera solaris* with 62% and 69% identity respectively.

THAPSDRAFT_27811 is a putative aminotransferase, with an aspartate aminotransferase like region, from a family belonging to pyridoxal phosphate (PLP)-dependent aspartate aminotransferase superfamily (fold I). This superfamily of enzymes catalyses the substitution of an O from a C=O double bound for a N forming a C=N bound, this reaction is called Schiff base. Top hits in BLASTp, e values <-150, are to unnamed protein product VEU42286.1 from *Pseudo-nitzschia multistriata* e value $2e^{-170}$ and 56.35% identity, PLP-dependent transferase FRACYDRAFT_190993 (OEU13090.1) from *F. cylindrus* e value $6e^{-186}$ and 58% identity and hypothetical protein THAOC_33597 (EJK47669.1) e value $2e^{-152}$ and 55% identity.

Finally, THAPSDRAFT_12030 has provisionally been annotated as an aspartate aminotransferase like protein also belonging to a PLP-dependent aspartate aminotransferase superfamily (fold I), and it has been suggested that it may contain a signal peptide. The closest hits to THAPSDRAFT_12030 with e value 0 are to 68% identity unnamed protein product VEU41084.1 from *Pseudo-nitzschia multistriata*, 67% identity aspartate-prehenate aminotransferase FRACYDRAFT_238714 (OEU17070.1) from *F. cylindrus*, 81% identity a partial hypothetical protein THAOC_05268 (EJK73125.1), 66% identity bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase FisN_22Lh121 (GAX11436.1) from *F. solaris* and to 65% partial predicted protein PHATRDRAFT_bd870 (XP_002176258.1) from *P. tricornutum*.

In addition to the candidates found by using known aminotransferases as probes for BLASTp searches in NCBI database, the aminotransferase identified in *F. cylindrus* by Lyon *et al.*[98] was also used to find candidates in *T. pseudonana* and *P. tricornutum.* The closest homologous in *T. pseudonana* is an aspartate aminotransferase, THAPSDRAFT_31394 (XP_002285992.1). The aminotransferase is widely annotated as LL-diaminopimelate aminotransferase, an enzyme involved in lysine biosynthesis in plants[208].

Finally, Bromke *et al.* have also identified a candidate gene for this first step of the transamination pathway THAPSDRAFT_260934 (XP_002285992.1)[175]. This enzyme has another homologue in *T. pseudonana*, a predicted protein 88.64% identical and e value 0, THAPSDRAFT_20816 (XP_002286782.1). A homologue is also present in *T. oceanica* with 64% of identities and e value 0 annotated as hypothetical protein THAOC_13089 (EJK66011.1).

None of these candidates have been further investigated hence whether any of them is involved in the transamination step remains unknown. However, each of the identified candidates were searched in our *T. pseudonana* whole transcriptomic sequencing (see Chapter 3) to observe their regulation by salinity and nitrate availability (Table 4-1).

Table 4-1. List of candidate aminotransferases in *T. pseudonana* and regulation of their transcription by salinity and nitrogen limitation. Regulation of the transcription of the candidate genes from whole transcriptomic sequencing of *T. pseudonana* cultures grown at decreased salinity (Low salt) and in N limitating conditions (Low N).

Description	Prot ID	Low salt	Low N
Branched-chain amino acid aminotransferase	21813	-	-
Putative aminotransferase	27811	1.07	1.75
Aspartate aminotransferase like protein	12030	0.42	1.62
Aspartate aminotransferase	31394	0.44	0.66
Uncharacterised protein[175]	260934	6.18	7.26
Predicted protein	20816	0.79	0.82

Out of all the candidate aminotransferases, the uncharacterised protein THAPSDRAFT_260934 is the most markedly up regulated by both increased salinity and decreased nitrogen. The putative aminotransferase THAPSDRAFT_27811 is very moderately regulated by salt and mildly regulated by nitrogen. These two aminotransferases are the most likely candidates to be involved in DMSP synthesis. However, it should be noted that transcriptional regulation may not be relevant if

regulation is mediated at the protein level or if the only regulated step is the methylation.

4.2.1.2 Reduction

FRACYDRAFT_173405 (OEU09668.1) from *F. cylindrus* was used to search for homologues in *T. pseudonana* genome. A predicted NADPH-dependent FMN reductase THAPSDRAFT_21067 (XP_002287064.1) was identified. The closest hit to this predicted protein from *T. pseudonana*, with an e value -103 and 66.2% identity, is the predicted protein NADPH-dependent FMN reductase PHATRDRAFT_37671 (XP_002181778.1) from *P. tricornutum*.

According to our proteomic database of *P.* tricornutum grown in standard conditions or at low salinity (PSU 5), see Chapter 3, this candidate is downregulated by both increased salinity and nitrogen limitation by 0.94 and 0.14-fold respectively. Therefore, this candidate reductase is unlikely part of the transamination pathway. Furthermore, its inclusion in this study is only based on proteomic regulatory data. There is not functional data linking this gene to DMSP synthesis.

4.2.1.3 S-Methylation

The whole transcriptome datasets from *T. pseudonana* subject to a salinity shift and nitrogen limitation treatments (see Chapter 3) was searched for SAM-dependent methyltransferases regulated by both conditions. Only those regulated as expected from an enzyme involved in DMSP synthesis, upregulated by increased salinity and by N starvation, were selected.

There are two methyltransferases upregulated by both salinity and low nitrogen in *T. pseudonana.* The first is the predicted protein SAM dependent carboxyl methyltransferase THAPSDRAFT_11247 (XP_002296978.1), which is 205.98-fold up in high salt compared to low salt and 2.59-fold higher in low N. The unnamed protein product VEU38650.1 from *Pseudo-nitzschia multistriata* has a 59% identity and e value

of -169 to THAPSDRAFT_11247, and S-adenosyl-L-methionine-dependent methyltransferase FRACYDRAFT_207357 (OEU19039.1) from *F. cylindrus* has 57% identity and e value of -165 to this candidate methyltransferase.

The second is the methyl transferase-like protein THAPSDRAFT_269095 (XP_002291473.1), upregulated 49.23-fold by salt and 88.22-fold by deplete N. *T. oceanica, P. tricornutum, F. solaris and F. cylindrus* have homologues to this protein. In *T. oceanica,* two hypothetical proteins THAOC_31798 (EJK49332) and THAOC_20742 (EJK59074.1) 75% identical and e value of $-3e^{-126}$ and 49.59% and e value of $2e^{-79}$, respectively. The predicted protein (XP_002183266.1) from *P. tricornutum* is 57.66% identical and e value of $3e^{-92}$. *F. solaris* has also two hits to the *T. pseudonana* methyltransferase, a glyoxylate/hydroxypyruvate reductase (GAX25165.1) with an e value of $1e^{-60}$ and 41.15% of identities; and a hypothetical FisN_16Lh025 (GAX26208.1) e value $7e^{-53}$ and 40.34% of identities. The S-adenosyl-L-methionine-dependent methyltransferase (OEU16655.1) is 40% identical to *T*. pseudonana's one with an e value of $6e^{-45}$.

The regulation of these two SAM-methyltransferases, THAPSDRAFT_11247 and THAPSDRAFT_269095, makes them strong candidates to be involved in DMSP synthesis. Functional analysis of both enzymes was performed and showed that indeed, THAPSDRAFT_269095 is a functional MTHB methyltransferase named DSYD (**D**MSHB **sy**nthase in **D**iatoms). Analysis and results of the functional tests are found in Chapter 6.

4.2.1.4 Oxidative decarboxylation

Only one candidate for the oxidative decarboxylation step was found searching the whole transcriptome datasets from the salinity and nutrient experiment in *T. pseudonana* (see Chapter 3). The predicted protein THAPSDRAFT_20613 (XP_002286584.1) is a Type III pyridoxal 5-phosphate (PLP)-dependent enzyme diaminopilate decarboxylase. It is upregulated 1.38-fold by increased salinity and

1.26 times by low nitrogen. The closest hits are found in the diatoms *P. tricornutum, F. solaris, P. multistriata, T. oceanica* and *F. cylindrus* and the microalgae *Nannochloropsis salina* (Table 4-2).

Table 4-2. List of homologues of THAPSDRAFT_20613 (XP_002286584.1) and percentage of identity.

Description	Organism	% Ident	Accession number
Diaminopimelate decarboxylase	*P. tricornutum*	70.93	XP_002181752.1
Diaminopimelate decarboxylase	*F. solaris*	72.59	GAX25863.1
Diaminopimelate decarboxylase	*F. solaris*	73.02	GAX09512.1
Unnamed protein product	*P. multistriata*	70.89	VEU44478.1
Hypothetical protein	*T. oceanica*	72.63	EJK58949.1
Alanine racemase C-terminal domain-like protein	*F. cylindrus*	74.17	OEU12704.1
Hypothetical protein NSK_000524	*N. salina*	60	TFJ88172.1

4.2.2 Analysis of genes potentially involved in GBT synthesis.

A search in the genome database of *T. pseudonana* using the known choline dehydrogenase from *E. coli* BetA[37] shows that this diatom has a candidate choline dehydrogenase THAPSDRAFT_41650 (XP_002292146.1). This candidate is 44.30% identical to the functional choline dehydrogenase from *E. coli* (e value $2e^{-152}$) and 32.14% identical to the functional choline oxidase from *A. globiformis*[206] (e value $2e^{-67}$). A further search in the NCBI database using the diatom candidate as a probe was performed. The closest homologues of the *T. pseudonana* candidate choline dehydrogenase returned belong to a range of α-proteobacteria species (Table 4-3).

Transcription of the candidate choline dehydrogenase in *T. pseudonana* is upregulated by 1.96-fold in high salinity and it is downregulated by 0.75-fold by reduced nitrogen according to our RNA sequencing results. The regulation of this candidate gene is similar to the regulation of GBT in *T. pseudonana*. However, further tests would be

necessary to proof whether this methyltransferase plays a role or not in GBT synthesis in the centric diatom.

Table 4-3. Organisms containing choline dehydrogenase homologues to THAPSDRAFT_41650 and their percentage of identity. In all cases, e value was equal to 0.

Organism	Accession number	% Identity
Roseibacterium elongatum	WP_025312387.1	51.38
Phyllobacteriaceae bacterium SYSU D60012	WP_119392947.1	50.75
Rhodobacteraceae bacterium EL53	WP_121996696.1	52.24
Phaeobacter gallaeciensis	WP_113823002.1	52.24
Nesiotobacter exalbescens	WP_102867507.1	51.49
Nesiotobacter exalbescens	WP_028482931.1	51.30
Loktanella sp. S4079	WP_045995726.1	51.67
Mesorhizobium oceanicum	WP_072603176.1	51.50
Mesorhizobium sp. B2.3	WP_126701110.1	52.24
Ruegeria marisrubri	WP_068347879.1	52.53
Pannonibacter phragmitetus	WP_094462413.1	51.03

On the other hand, S-adenosyl-L-methionine-dependent methyltransferase THAPSDRAFT_20797 (XP_002286764.1) has been already suggested as a candidate GBT synthesis gene via the alternative methylation pathway, regulated by salinity and nitrogen in the same manner as the intracellular GBT concentrations[128]. It has low identity to published methyltransferase, GSMT from *Methanohalophilus portucalensis*[209], it is 30.24% identical ($2e^{-29}$). Likewise, *T. pseudonana* candidate GSDMT is 28.47% identical ($3e^{-25}$) to the SMT domain from *Aphanothece halophytica* and a 28.47% of identity (e value $3e^{-25}$ to the candidate Gsdmt from *Actinopolyspora halophila*.

T. pseudonana from GSDMT (TpGSDMT) has homologues in the diatoms *T. oceanica*, *P. multistriata*, *P. tricornutum* and *F. cylindrus*. The closest homologue is the one found in *T. oceanica* THAOC_12055 (XP_002286764.1) which is 77% identical and e value of 0. The homologues found in the other diatoms range between 30-40% identity with e values between -49 and -112. Other homologues of this candidate TpGSDMT are found in the gammaproteobacterial *Thioalkalivibrio* sp. ALJ24 (Table 4-4).

Table 4-4. Closest homologues of TpGSDMT found using BLASTp to search in NCBI.

Organisms	Accession ID	E value	Per. Ident.
Thalassiosira oceanica	EJK66971.1	0	77.70
Pseudo-nitzschia multistriata	VEU45206.1	5E-112	34.18
Phaeodactylum tricornutum CCAP 1055/1	XP_002180089.1	5E-62	39.86
Fragilariopsis cylindrus CCMP1102	OEU23352.1	3E-49	31.96
Thioalkalivibrio sp. ALJ24	WP_018935996.1	5E-34	34.72
Thioalkalivibrio sp. ALJ24	WP_017940942.1	1E-33	33.68
Thioalkalivibrio sp. ALJ24	WP_018869113.1	1E-33	35.09

The RNA sequence database produced during this study (see Chapter 3) ratifies the already observed regulation, THAPSDRAFT_20797 is within the 15 most upregulated genes in increased osmotic stress, by 166-fold, and downregulated by N limitation by 0.35-fold.

The regulation of the candidate *GSDMT* is more likely to explain the fluctuation of GBT in *T. pseudonana* cells when they are grown in different salt concentrations and nitrogen depletion than a candidate choline dehydrogenase. In Chapter 7, the functionality of this methyltransferase candidate is further explored, results indicate that TpGSDMT is a functional glycine sarcosine dimethylglycine methyltransferase involved in GBT synthesis in this diatom.

4.2.3 Analysis of genes potentially involved in GBT and DMSP transport.

There are no genes in *T. pseudonana* encoding proteins with significant similarity to functional ProU operon from *Salmonella tiphymurium*[210]. However, *T. pseudonana* contains several ABC transporter proteins similar to the ProV which is the proposed to be responsible for energy coupling the ProU transport complex. However, the percentages of identity are low and e values are ≤-25. In addition, *T. pseudonana* does not appear to contain the GBT-binding protein ProX nor the membrane component

ProW and the candidate ABC transporters have a low homology to ProV, 31.88% and 31.87% identity and e values of $2e^{-25}$ and $9e^{-25}$ respectively.

Table 4-5. Closest homologues of candidate ProP (THAPSDRAFT_268228) found using BLASTp to search in NCBI.

Organisms	Accession ID	E value	Perc Identity
T. pseudonana	XP_002292228.1	0.0	0.071.36%
P. multistriata	VEU40486.1	7e-163	56.84%
Fragilariopsis cylindrus CCMP1102	OEU23485.1	1e-154	56.21%
T. oceanica	EJK49112.1	1e-150	62.13%
T. oceanica	EJK70558.1	1e-139	52.49%
F. cylindrus	OEU16229.1	1e-128	48.80%
P. tricornutum CCAP 1055/1	XP_002186229.1	8e-124	45.77%
T. oceanica	EJK67433.1	5e-122	48.19%
T. oceanica	EJK47044.1	1e-113	43.60%

A candidate gene similar to ProP was identified through searches in the whole transcriptome dataset (see Chapter 3). The candidate ProP transporter gene, THAPSDRAFT_268228 (XP_002287760.1), is downregulated by reduced nitrogen 0.63-fold and mildly upregulated 1.13-fold by salt. It belongs to a major facilitator superfamily (MFS). Homologues of this candidate ProP are present in other diatoms including *T. oceanica*, *P. tricornutum*, *F. cylindrus* and *P. multistriata* (Table 4-5).

Table 4-6. Closest homologues of candidate BCCT (THAPSDRAFT_262307) found using BLASTp to search in NCBI.

Organisms	Perc Identity	Accession ID	E value
Fragilariopsis cylindrus CCMP1102	OEU16937.1	0	70.19%
P. tricornutum CCAP 1055/1	XP_002182771.1	5E-122	55.43%
T. oceanica	EJK50030.1	5E-119	62.54%
Hondaea fermentalgiana	GBG27519.1	4E-115	50.00%
Hondaea fermentalgiana	GBG27521.1	4E-109	48.94%
Hondaea fermentalgiana	GBG27520.1	3E-107	49.72%

The candidate BCCT transporter THAPSDRAFT_262307 previously found to be present in *T. pseudonana*[207] appears within the top most regulated transcripts by salinity in the *T. pseudonana* transcriptomic dataset (see Chapter 3). This candidate

transporter is 86.59 times more transcribed at high salt than at low salt. Nitrogen availability does not have a big effect on the transcription of this transporter as the depletion of N only causes it to increase 1.51-fold. Homologues to THAPSDRAFT_262307 are found in *F. cylindrus*, *P. tricornutum* and *T. oceanica* as well as in the protist *Hondaea fermentalgiana* (Table 4-6).

As this candidate BCCT transporter is one of the strongest regulated genes by salinity, it was further studied in Chapter 5.

4.3 Identification of genes of interest in *P. tricornutum*

Samples collected from *P. tricornutum* cultures grown at high/low salt (see Chapter 3) were sent for proteomic analysis to Dr Joseph Christie-Oleza at Warwick University, who has a comprehensive database of *P. tricornutum* proteins available to compare the proteomic results (Appendix 1-3).

To identify candidate genes of interest three different approaches were followed. Firstly, homologues of previously published genes of interest will be used to interrogate *P. tricornutum*'s genome. Secondly, found sequences will be compared to the candidate genes found in *T. pseudonana* and their regulation by salt was established using the proteomic results (see Chapter 3). Lastly, when possible, the number of candidate genes were expanded by searching for relevant enzyme activities in the proteomic results.

4.3.1 Analysis of genes potentially involved in DMSP synthesis.

4.3.1.1 Transamination

P. tricornutum contains a sequence very similar to the candidate aminotransferase found in *F. cylindrus*[98], the predicted protein PHATRDRAFT_22909 (XP_002183711.1) which is 78.47% identical with a e value of 0. This candidate gene is also 81.77% (e value of 0) identical to the protein found in *T. pseudonana* THAPSDRAFT_260934.

When BACT4 from *Arabidopsis* was used as a probe in BLASTp, the two aminotransferases with highest identity were predicted protein PHATRDRAFT_32849 (XP_002178037.1), e value $7e^{-72}$ and 44% identity, and predicted protein PHATRDRAFT_50793 (XP_002178511.1), e value $1e^{-61}$ and 35% identity. The second candidate was already identified as a homologue of the *T. pseudonana* candidate aminotransferase THAPSDRAFT_21813. On the other hand, homologues to YbdL have a low percentage of identity. These are predicted protein PHATRDRAFT_10257 (XP_002186243.1), e value $3e^{-61}$ and 33% identity, and predicted partial protein PHATRDRAFT_10073 (XP_002177979.1), e value $4e^{-47}$ and 29% identity. Lastly, the closest hit to O59096 from *P. horikoshii* was partial predicted protein PHATRDRAFT_bd870 (XP_002176258.1), e value 8-71 and 37% identity, previously identified as homologue to the candidate gene in *T. pseudonana* THAPSDRAFT_12030.

PHATRDRAFT_32849 (XP_002178037.1) is a predicted protein belonging to the BCAT_beta_family, it is predicted to catalyse the transamination of the branched-chain amino acids leucine, isoleucine and valine. Identical proteins with e values <-100 are hypothetical protein THAOC_19234 (EJK60420.1) 62% identity, and hypothetical protein THAOC_23586 (EJK56512.1) 66% identity from *T. oceanica.*

PHATRDRAFT_50793 (XP_002178511.1) is a predicted PLPDE_IV protein, and it appears in the proteomic results as non-regulated. Hits with the e value 0 are previously mentioned FisN_5Lh102 (GAX13983.1) and FisN_5Hh102 (GAX17454.1) from *Fistulifera solaris* with 75% and 76% identity respectively. Unnamed protein product (VEU34356.1), 68% identity, from *P. multistriata*. And previously described aminotransferase candidate in *T. pseudonana*, predicted protein THAPSDRAFT_21813 (XP_002288882.1) 65% identity.

PHATR_10257 (XP_002186243.1) is a predicted aspartate aminotransferase that belongs to the (fold I) superfamily (PLP)-dependent aspartate aminotransferase. The

two proteins with more similarities and e values 0 are 68% identity FisN_34Hh035 (GAX11244.1) and 66% identity FisN_34Lh035 (GAX13433.1) from *F. solaris* both annotated as kynurenine-oxoglutarate transaminase/ cysteine-S-conjugate beta-lyase/ glutamine-phenylpyruvate transaminase.

PHATRDRAFT_10073 (XP_002177979.1) is a partial predicted protein included in the aspartate aminotransferase family which, in turn, belongs to the (fold I) PLP-dependent superfamily of aspartate aminotransferases. There are six proteins with high similarity (e value 0) including FisN_1Lh722 (GAX26452.1) and FisN_1Hh722 (GAX15285.1), both 77% identity and from *F. solaris* described as kynurenine-oxoglutarate transaminase/ cysteine-S-conjugate beta-lyase/ glutamine-phenylpyruvate transaminase. These are followed by a partial PLP-dependent transferase from *F. cylindrus* FRACYDRAFT_157973 (OEU15425.1) and an unnamed protein VEU33419.1 from *P. multistriata*, both with 76% identity. A different aminotransferase from the previously identified ones in *T. pseudonana* is the 71% aminotransferase THAPSDRAFT_460 (XP_002290611.1), which is worth noting that contains two regions, one identified as AspB, Aspartate/methionine/tyrosine aminotransferase, and another region belonging to the (PLP)-dependent aminotransferase superfamily (fold I). Finally, hypothetical protein THAOC_18873 (EJK60722.1) has 70% of identities.

The predicted protein PHATRDRAFT_bd870 (XP_002176258.1) is classified as part of the aspartate aminotransferase family, also belonging to the (PLP)-dependent aspartate aminotransferase superfamily (fold I). The closest homologues with e value 0 are the same as for THAPSDRAFT_12030 (XP_002295148.1), the identity to these proteins are 72% to FisN_22Lh121 (GAX11436.1) from *F. solaris*, 69% to FRACYDRAFT_238714 (OEU17070.1) and 68% to VEU41084.1 from *Pseudo-nitzschia multistriata*. It also includes FisN_22Hh121 (GAX16021.1), with 70% identity, which is a bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase.

Unfortunately, the only aminotransferase present in the proteomic database used in this study is PHATRDRAFT_bd870 (XP_002176258.1) and it is not significantly regulated by salinity.

4.3.1.2 Reduction

The previously identified NADPH-dependent FMN reductase PHATRDRAFT_37671 (XP_002181778.1), closest hit is also the *T. pseudonana* candidate THAPSDRAFT_21067 (XP_002287064.1). And it has 52% of identities ($1e^{-83}$) to the candidate reductase from *F cylindrus* FRACYDRAFT_173405 (OEU09668.1)[98].

Searching in the proteomic database, three candidate reductases were significantly downregulated in low salt compared to high salt. PHATRDRAFT_44546 (XP_002178319.1) was downregulated by -3.10-fold in log2, PHATRDRAFT_44630 (XP_002178364.1) was downregulated -2.16-fold in log 2 and PHATRDRAFT_13107 (XP_002180919.1) by -1.76-fold in log 2 by lowered salinity.

The closest homologue of predicted protein PHATRDRAFT_44546 is found in *Symbiodinium microadriaticum*, putative aldo-keto reductase 1 (OLP84158.1) which is 47% identical (e value $6e^{-74}$). PHATRDRAFT_44546 contains a predicted oxidoreductase region related to aryl-alcohol dehydrogenase.

PHATRDRAFT_44630 is a predicted protein with a short-chain dehydrogenase/reductase region and a NAD(P) binding site. Homologues to this candidate are present in other significant diatoms such as *F. solaris, F. cylindrus* and *T. oceanica*, also in the dinoflagellate *S. microadriaticum*, in the algae *A. anophagefferens* and the corals *A. digitifera*, *S. pistillata* and *P. damicorns* (Table 4-7).

The predicted protein PHATRDRAFT_13107 is highly similar to the peptide-methionine (S)-S-oxide reductases GAX25009.1 and GAX26757.1 from *F. solaris* 81.76% (8e-102) and 82.35 percentage of identity respectively. Other hits with less homology are ≤ 60% identical and e values of <-70 to peptide methionine sulfoxide

reductases in higher organisms such as the lycophyte *Selaginella moellendorffii* or the earth moss *Physcomitrella patens.*

Table 4-7. Closest homologues of candidate reductase (PHATRDRAFT_44630) found using BLASTp to search in NCBI.

Organisms	Accession ID	E value	Perc Ident
Fistulifera solaris	GAX21686.1	2E-137	65.70%
Fistulifera solaris	GAX14906.1	4E-135	63.54%
F. cylindrus CCMP1102	OEU16237.1	1E-111	56.93%
Fistulifera solaris	GAX21685.1	6E-110	56.83%
Fistulifera solaris	GAX14905.1	3E-109	56.46%
T. oceanica	EJK76894.1	4E-105	57.35%
Fistulifera solaris	GAX14904.1	6E-99	51.27%
Fistulifera solaris	GAX21684.1	3E-97	53.33%
T. oceanica	EJK65460.1	9E-83	61.22%
Symbiodinium microadriaticum	OLQ05117.1	2E-75	48.00%
Aureococcus anophagefferens	XP_009032436.1	6E-68	47.04%
Acropora digitifera	XP_015751846.1	4E-64	44.40%
Stylophora pistillata	XP_022787306.1	1E-61	42.60%
Pocillopora damicornis	XP_027042147.1	5E-60	42.24%

4.3.1.3 S-Methylation

In the previous section, candidate DSYD proteins were identified in *T. pseudonana* by looking at methyltransferases regulated by salinity and nitrogen in the sequenced RNA database. A similar approach was used to identify DSYD candidate genes in *P. tricornutum*, in this case, proteomic results from salinity experiments were searched for methyltransferases significantly downregulated in low salinity as it would be expected if involved in DMSP production by *P. tricornutum* (see Chapter 3).

The only methyltransferase meeting the regulation criteria was the single methyltransferase domain protein UniProt ID B7G7F7 (PHATRDRAFT_48704). This protein was found to be statistically significantly downregulated -2.43-fold in log2 (q-value <0.01) in cultures grown in low salinity compared to the control. Closest homologues are the hypothetical protein K0RRP8 (THAOC_31798) from *Thalassiosira oceanica*, 58.28% identity e value $6e^{-128}$ and the hypothetical protein K0S2L3

(THAOC_20742), 47.20% and e value of $3e^{-85}$. Most importantly, it is 57.66% identical with an e value of $1e^{-90}$ to the previously identified methyl transferase-like protein B8C5X1 (THAPSDRAFT_269095) from *T. pseudonana* (see section "4.2.1.3 S-Methylation"). In this thesis, it was found that PHATRDRAFT_48704 was a functional DSYD enzyme. Functional analysis performed are further discussed in Chapter 6.

4.3.1.4 Oxidative decarboxylation

The only candidate oxidative decarboxylase identified in *P. tricornutum* through searches in the proteomic database was the diaminopimelate decarboxylase diaminopimelate decarboxylase PHATRDRAFT_21592.

This decarboxylase is a homologue to the candidate for this step found in *T. pseudonana* (Table 4-2). It also belongs to the Type III pyridoxal 5-phosphate (PLP)-dependent enzyme diaminopilate decarboxylase family. Unlike the *T. pseudonana*'s candidate, the decarboxylase is not regulated by salinity in *P. tricornutum.* Homologues of this protein are found in other diatoms including the aforementioned candidate from *T. pseudonana* (Table 4-8).

Table 4-8. Closest homologues of candidate reductase (PHATRDRAFT_21592) found using BLASTp to search in NCBI.

Organisms	Accession ID	Perc Ident
Fistulifera solaris	GAX09512.1	80.30%
Fistulifera solaris	GAX25863.1	79.65%
Pseudo-nitzschia multistriata	VEU44478.1	71.69%
Thalassiosira pseudonana CCMP1335	XP_002286584.1	70.93%
Fragilariopsis cylindrus CCMP1102	OEU12704.1	76.29%
Thalassiosira oceanica	EJK58949.1	64.33%

4.3.2 Analysis of genes potentially involved in GBT synthesis.

4.3.2.1 Choline pathway

A candidate choline dehydrogenase was found in *P. tricornutum* by doing a Blastp search in NCBI database, UniProt ID B7FSU6 (PHATRDRAFT_1341). This protein has a conserved domain from the superfamily of FAD/NAD (binding) proteins. It also has two conserved domains glucose-methanol-choline oxidoreductase domain in the N and C-terminal. This family include a variety of proteins such as choline dehydrogenases.

It has 35% identity and $2e^{-94}$ to the *betA* gene from *E. coli*. The top hits with e values equal to 0.0 are to *Pseudo-nitzschia miltistrata* (VEU45243.1) with 79% identity and to *Fragilariopsis cylindrus* CCMP1102 (OEU15591.1) with 77% identity. Proteomic results show that this protein is not significantly regulated by salt. As seen in Chapter 3, addition of choline increases the accumulation of GBT in the cell, it would be interesting to explore the functionality and regulation of this candidate gene in future investigations.

4.3.2.2 Methylation pathway

The candidate glycine-sarcosine-dimethylglycine methyltransferase (TpGSDMT) from *T. pseudonana*, was used to search the NCBI database (BLASTp) targeting *Phaeodactylum tricornutum* CCAP 1055/1 yielding a 39.86% identical protein, e value $3e^{-66}$, UniProt ID B7FZ54 (PHATRDRAFT_12280), which is annotated as partial sequence. This sequence was used to BLASTp against *P. tricornutum* genome and producing 7 hits, protein ID 35747, 20301, 45916, 27426, 12280 and 5243. Despite the different identifier numbers, all sequences refer to the same protein or to a fraction of the same protein. Until the sequences are unified, hereafter the candidate GBT synthesis gene will be solely referred to by protein ID PHATRDRAFT_20301.

The top hit of the candidate enzyme is to *Pseudo-nitzschia multistriata* (VEU45206.1), 50.77% identity and e value of 0.0, is followed by *Thalassiosira oceanica* (THAOC_12055) 40% identity and e value of $4e^{-146}$, and by the previously identified TpGSDMT in *Thalassiosira pseudonana* (THAPSDRAFT_20797), which is 40% identical and e value of $1e^{-143}$. The proteomic results are incomplete and lack the information for the gene ID PHATRDRAFT_20301.

This protein has two SAM-dependent methyltransferase domains and in this thesis, it has been functionally characterised as a GSDMT, further details are found in Chapter 7.

4.3.3 Analysis of genes potentially involved in GBT and DMSP transport.

Alike *T. pseudonana*, *P. tricornutum* contains 12 sequences with a percentage of identity rounding 30% and e values of ≤25 to the ProV portion of the ABC like GBT transporter and it lacks the other components that compose the ProU transport system. Furthermore, one homologue to ProP was found, the predicted protein PHATR_33590 (XP_002186229.1) which contains a region belonging to a major facilitator superfamily and is 28% identical to ProP from *E. coli* (e value $2e^{-45}$).

Table 4-9. Closest homologues of candidate reductase (PHATRDRAFT_21592) found using BLASTp to search in NCBI. E value for all the hits was 0.

Organism	Accession number	Perc Ident
Fragilariopsis cylindrus CCMP1102	OEU16937.1	49.27%
Hondaea fermentalgiana	GBG27521.1	43.56%
Hondaea fermentalgiana	GBG27520.1	43.02%
Hondaea fermentalgiana	GBG27519.1	41.82%
Thalassiosira oceanica	EJK50030.1	40.14%

As shown in the previous section, the candidate BCCT transporter found in *T. pseudonana* has a homologue in *P. tricornutum*, predicted protein PHATRDRAFT_48315 (XP_002182771.1). Unfortunately, the proteomic database

done in this study does not contain this protein, so it was not possible to analyse how it is regulated by this means. Homologues of this candidate BCCT transporter (Table 4-9) are the same as those identified for the candidate BCCT from *T. pseudonana* (Table 4-6).

4.3.4 Other proteins of interest

The top 50 proteins upregulated or downregulated by decreased salinity (Appendix 1-3) were identified using the whole proteome sequencing database from *P. tricornutum* cultures grown at either normal salinity (35 PSU) or low salinity (5 PSU).

4.4 Discussion and concluding remarks

In this chapter, candidate genes were identified in the model diatoms *T. pseudonana* (Appendix 1-1) and *P. tricornutum* (Appendix 1-2). These genes are potentially involved in the transamination pathway for DMSP biosynthesis, GBT synthesis via the choline or the methylation pathway and transport proteins that could be responsible for DMSP and GBT movement.

4.4.1 DMSP synthesis

The transamination step involves the removal of the amino group from Met to form MTOB. Aminotransferases capable of carrying out this transamination were found in plants[200] and bacteria[202], and candidate enzymes were proposed to be involved in DMSP synthesis in the diatom *F. cylindrus*[98] and *T. pseudonana*[175]. In this study, several aminotransferases were identified by comparing the known aminotransferases against *T. pseudonana* and *P. tricornutum* genome (Appendix 3). Two of the candidates found in *T. pseudonana* stand out amongst the other aminotransferases as their expression is regulated by salt and nitrogen in the same manner as DMSP intracellular concentrations. In particular, THAPSDRAFT_260934 although annotated as an aspartate aminotransferase, is distinctly upregulated by both

conditions whereas THAPSDRAFT_27811 has a more modest regulation. On the other hand, *P. tricornutum* has also a set of putative aminotransferases that could be taking part in DMSP synthesis. However, it also has a close homologue to the candidate aminotransferase found in *F. cylindrus and T. pseudonana* (THAPSDRAFT_260934) the uncharacterised protein PHATRDRAFT_22909 (XP_002183711.1). None of these candidate aminotransferases have been tested for functionality and further investigation is required.

The next step in the transamination pathway is the reduction of MTOB to MTHB performed by a NADP- NADH- dependent enzyme[97]. Although no candidates were identified in *T. pseudonana*, a candidate reductase found to be downregulated by lowered salinity in *P. tricornutum* contained a short-chain dehydrogenase/reductase region and a NAD(P) binding site PHATRDRAFT_44630 (XP_002178364.1). Although this candidate reductase has not been ratified as a functional protein, its homologues are found in other DMSP producing diatoms and flagellates as well as in corals (Appendix 4). It is strongly recommended to test whether this enzyme is indeed involved in DMSP synthesis.

The methylation of MTHB to DMSHB is the committing and key step of the transamination pathway as it is the first non-reversible enzymatic reaction and it is unique to DMSP producers. MTHB is methylated by a SAM-dependent methyltransferase enzyme[97], known methyltransferases able to catalyse this step in other organism are strongly regulated by the conditions known to affect DMSP production[99]. In *T. pseudonana*, two methyltransferases were found to be regulated by salt and nitrogen, the predicted protein SAM dependent carboxyl methyltransferase THAPSDRAFT_11247 (XP_002296978.1) and the methyl transferase-like protein THAPSDRAFT_269095 (XP_002291473.1). Homologues of both enzymes are found in other DMSP producing diatoms. In *P. tricornutum*, only one methyltransferase was downregulated by decreased salinity, PHATRDRAFT_48704 (XP_002183266.1), this candidate methyltransferase is in turn 57.66% ($1e^{-90}$) identical to the candidate from

T. pseudonana THAPSDRAFT_269095. This candidate protein has a very low homology to the known DsyB from *Labrenzia aggregata* (28.08%, 1e-22). Further work performed as part of this thesis on THAPSDRAFT_269095 and PHATRDRAFT_48704 has ratified them as functional DSYD enzymes. Further details are in Chapter 6.

Finally, a candidate decarboxylase was found in *T. pseudonana* and *P. tricornutum*. In *T. pseudonana*, a Type III pyridoxal 5-phosphate (PLP)-dependent enzyme diaminopilate decarboxylase was upregulated by salt and by N limitation, THAPSDRAFT_20613 (XP_002286584.1). Its homologue in *P. tricornutum* is a diaminopimelate decarboxylase, 71% identical, PHATRDRAFT_21592 XP_002181752.1 (Appendix 5). Again, as this candidate has not been characterised or tested, further work is required to elucidate whether it is responsible for the last step of the transamination pathway or not.

4.4.2 GBT synthesis

Results in chapter 3 suggest that although both *T. pseudonana* and *P. tricornutum* synthesises GBT *de novo*, they have different intracellular concentrations. In both diatoms GBT concentration decreases at lower salinities, however GBT synthesis is regulated differently by N availability. Whereas in *T. pseudonana*, GBT synthesis is strongly downregulated N availability, in *P. tricornutum* it seems that there is a basal production of GBT that is only enhanced when choline is added to the media (Figure 3.10).

T. pseudonana has candidate enzymes for both pathways, THAPSDRAFT_41650 (XP_002292146.1) is a choline dehydrogenase (Appendix 6) mildly regulated by salinity and nitrogen, however, as the change in GBT intracellular concentration is larger than the fold change for this candidate enzyme, it is unlikely that the candidate choline dehydrogenase is the major responsible for GBT synthesis in this diatom for these conditions. Certainly, THAPSDRAFT_20797 (XP_002286764.1) is a candidate methyltransferase likely capable of performing the methylation of Gly to GBT. This

methyltransferase is markedly regulated by both salinity and nitrogen; hence it is more likely the enzyme responsible for GBT synthesis in *T. pseudonana* for these two conditions.

On the other hand, *P. tricornutum* also has a candidate choline dehydrogenase (PHATRDRAFT_1341) (Appendix 6) and a candidate methyltransferase (PHATRDRAFT_20301). The regulation of the candidate methyltransferase is unknown, and the candidate choline dehydrogenase is not regulated by salt and it seems to be expressed at low levels in this condition. The regulation of the candidate choline dehydrogenase agrees with the non-regulation of GBT concentration in the cell and it is likely contributing to the GBT pool. However, as the regulation of the methyltransferase is so far unknown, it is not possible to completely rule out the possibility of both enzymes synthesising GBT in *P. tricornutum* (this is further explored in Chapter 7). The coexistence of both pathways within one organism has already been characterised in *Actinopolyspora halophila*[211].

Functionality tests performed in candidate glycine sarcosine dimethylglycine methyltransferases from both diatoms established that these enzymes are functional. The tests, further characterization and regulation of the diatom GSDMT are described in Chapter 7. Further investigation is required to establish the functionality and the role of the candidate choline dehydrogenases in the model diatoms.

4.4.3 Transport

T. pseudonana and P. tricornutum unlikely use an ABC-like transport system as the one found in E. coli[33] as they lack two out of the three components of the ProU systems[29]. Furthermore, T. pseudonana and P. tricornutum have sequences with very low homology to the ProP transporter (Appendix 8). On the other hand, both model diatoms contain candidate BCCT transporters (Appendix 7).

Lavoie *et al.* already highlighted the presence of THAPSDRAFT_262307 as a likely candidate BCCT transporter, upregulated by increased salinity but not regulated by N availability. In *P. tricornutum*, the candidate BCCT transporter is PHATRDRAFT_48315 (XP_002182771.1). Unsuccessful attempts have been made during this thesis to test the functionality of PHATRDRAFT_48315 and further investigation is needed to ratify its involvement in the transport of these compatible solutes (see Chapter 5).

4.4.4 Conclusion and future work

In this chapter candidate genes were identified but in order to ratify any of the candidates it would be necessary to clone all the candidates and perform the necessary experiments to proof whether they are functional or not. In addition, it is possible that the actual enzymes for the first two steps and last step of the DMSP synthesis pathway have not been reported in this chapter as only the regulated proteins were studied. There is a possibility that the enzymes driving these steps are constitutively expressed and not regulated, with only the methyltransferase step being regulated, which may control DMSP synthesis rates. In addition, the annotation of the genomes of the model diatoms and the characterization of the proteins might contain errors that could lead to overlook some of the sequences.

Chapter 5

Identification and characterization of BCCT transporter

5 Identification and characterization of BCCT transporter

5.1 Introduction

Betaine/Carnitine/Choline Transporter family proteins belong to the APC superfamily[212]. One common feature of proteins in this family is they possess 12 transmembrane helical spanners (TMS) comprised of approximately 20 amino acid residues[213]. However, some non-characterised BCCT proteins have additional domains or membrane-spanning segments in addition to the core 12 TMS[214].

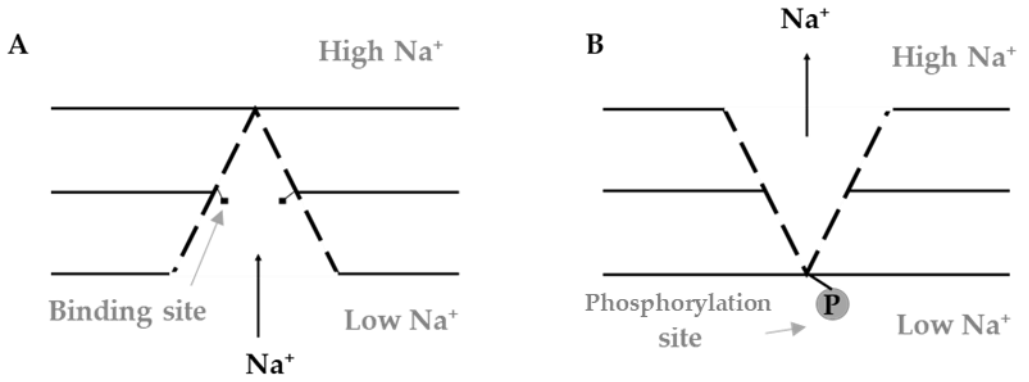
BCCT proteins are present in bacterial, archaeal and in some eukaryotic organisms, although the eukaryotic BCCTs have not been physiologically characterised to date[214]. Several of the known BCCT carrier proteins have been functionally and structurally characterised. These include OpuD from *Bacillus subtilis*[32], DddT from *Halomonas* and *Marinomonas*[71], the choline transporter BetT and CaiT from *E. coli*[34,215] and the well-studied BetP from *Corynebacterium glutamicum*[216,217]. The activity of the BCCT transporter is regulated by osmolarity with the exception of CaiT which lacks the N- and C-terminal extensions responsible for osmosensing[215].

All the proteins in this family share the ability to transport quaternary ammonium derivatives [R-N+(CH3)3] such as choline, betaine and carnitine amongst others[215], but they have also been shown to be able to transport other structurally similar molecules such as dimethylsulfonioacetate (DMSA) or DMSP[71].

The BCCT transport proteins present the inward- outward-facing conformation described by Jardetzky[216], whereby the polypeptides present an opening in one side of the membrane, a sodium molecule binds to a central binding side causing a dephosphorylation. The dephosphorylation, in turn, triggers a conformational change so the pump closes the entrance and opens up on the other side of the

membrane freeing the molecule (Figure 5.1). The recovery of the pump could be triggered by the exit of the substrate or by K+ entering into the cavity[218].

Figure 5.1. Allosteric pump mechanism described by Jardetzky, 1966. A) Sodium conformation B) Phosphorous conformation



In the case of the BCCT transporters, the activation of the pump is driven by the binding of Na^+, sodium motive force (sfm) such as in BetP[217], or of H^+, proton motive force (pmf) such as in BetT to the central binding site. The Na^+ or H^+ as well as the substrate only interacts with the central binding site in the cavity of the protein scaffold. This central binding site is coated with aromatic side chains, in BetP this is formed by tryptophan residues called the aromatic box[216].

The crystal structure and the functional studies performed on BetP from *Corynebacterium glutamicum* show the presence of a glycine motif, GxGxG, in the TM3 conserved in high affinity betaine transporters[216]. Changes in this motif lead to changes in substrate specificity and in the activation responds to increased osmolarity in BetP[217].

5.2 Aims of the chapter

The aim of the chapter is to check whether the identified BCCT transporters in diatoms function in the import and/or export of GBT and DMSP or not. I also investigate the distribution of this type of transporter in other organisms, if it is

specific for DMSP producing organisms, how the expression of this gene is regulated and where it is localised in the cell.

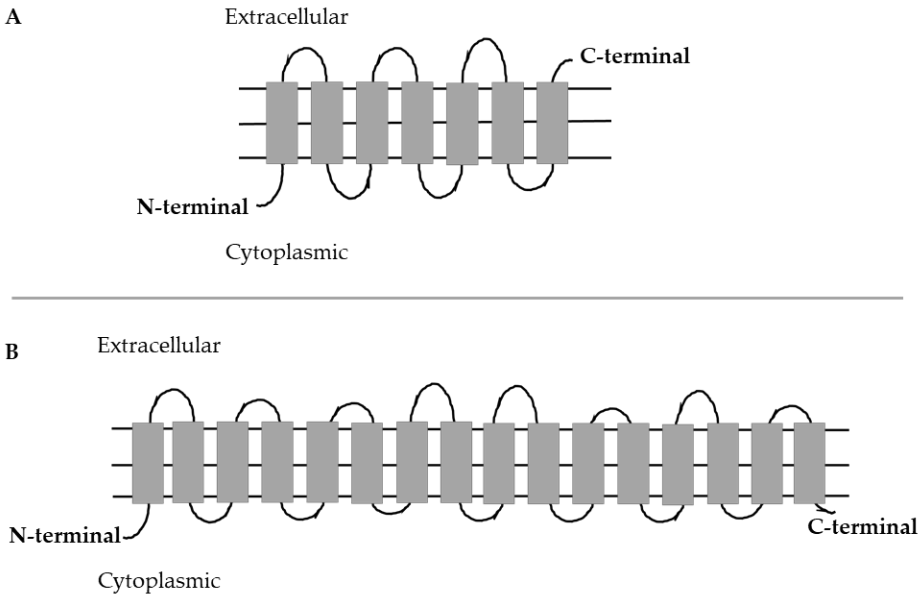
5.3 Characterization BCCT transporter

As seen in previous chapter (Chapter 4), candidate BCCT transporters where found in the model diatoms *F. cylindrus*, *P. tricornutum* and *T. pseudonana*. This is especially interesting since it is known that osmolytes such as glycine betaine and DMSP are being transported from the environment into the cells but the mechanism, although predicted, it has not been proven yet[131].

5.3.1 Analysis of the primary and secondary structure

Predictions using Phyre2[157] indicate the candidate *P. tricornutum* and *F. cylindrus* BCCT transporter has 16 predicted transmembrane helices (TMH), whereas *T. pseudonana* candidate has 8 TMH and FcBCCT candidate are predicted to have 16

Figure 5.2. Predicted TMH of the candidate BCCT transporters from A) *T. pseudonana* and B) *P. tricornutum* or *F. cylindrus*



TMH. (Figure 5.2). Thus *P. tricornutum* and *F. cylindrus* have four extra TMS compared to the canonical 12 TMS and this suggests that the TpBCCT may be incomplete as it is much shorter than other BCCTs.

The sequences of the candidate BCCT from diatoms were aligned to the characterised BCCT proteins OpuD from *Bacillus subtilis*, BetP from *C. glutamicum*, BetT from *E. coli* and DddT from *Halomonas* and the Gly motif was highlighted (Figure 5.3). This motif is conserved between transporters with the same substrate[217]. Previous studies have shown that the substitution of the third Gly for aspartate in BetT from *E. coli* allows this protein to bind to choline[217] rather than to glycine betaine, hence, changing substrate specificity. On the other hand, the substitution of this third Gly for alanine in BetP did not affect the activity at low osmotic pressure but alters the activation profile[217].

Figure 5.3. Gly motif of characterised BCCT proteins OpuD from *Bacillus subtilis*, BetP from *C. glutamicum*, BetT from *E. coli* and DddT from *Halomonas*.

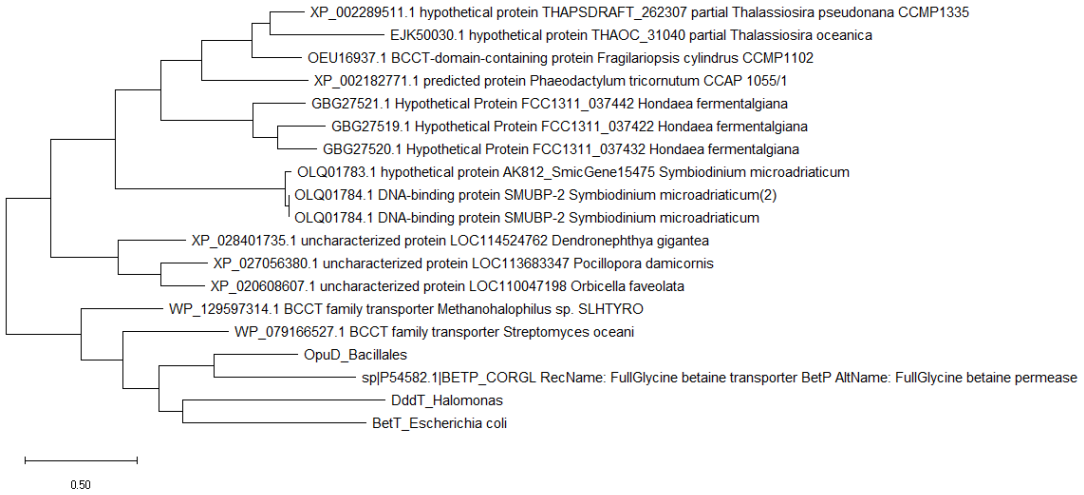
```
OpuD_B. subtilis      WFAMLFSAGMGIGLVFYGAAEPISHY
BetP_C.glutamicum     WISMMFAAGMGIGLMFYGTTEPLTFY
BetT_E.coli           WAAMLFAAGIGIDLMFFSVAEPVTQY
DddT_Halomonas        WVAMIFAGGIGIAIVNWAWVEPIYYF
BCCT_T.pseudonana     YFAMLFSAGVGVGLFFYGVSEPLWHQ
BCCT_F.cylindrus      YFAMIFSAGVGVGLFFYGVSEPLFHR
BCCT_P.tricornutum    YFAMIFAAGVAVGLFVFGVAEPLWHQ
                      : :*:*:.*:.: :. :.  **:
```

Figure 5.3 shows the Gly motif in the selected sequences, and it shows that the third Gly is substituted by an alanine in DddT from *Halomonas*. Moreover, it is worth noting that it has been shown that a replacement of the second Gly by alanine had a dramatic effect on BetP, decreasing the substrate uptake rate and showing no activation at high osmolalities[217]. Figure 5.3 also shows that the second Gly is substituted by an alanine in the candidate BCCT from *P. tricornutum* (Figure 5.3). Although further functional experiments are required, this substitution could be causing the inactivation of the candidate BCCT transporter in this pennate diatom.

5.3.2 Distribution

BLASTp at NCBI was used to perform searches to identify homologues of the *P. tricornutum* candidate BCCT proteins. To create a phylogenetic tree, the closest homologues, above the e value -107 cut off, and a selection of more distantly related proteins within the Pfam family pfam02028 (BCCT, betaine/carnitine/choline family transporter) were selected. The sequences were aligned by ClustalW and a maximum -likelihood phylogenetic was produced (Figure 5.4) using MEGAvX[156] to visualise the relation between the homologues.

Figure 5.4. Maximum Likelyhood phylogenetic tree of closest homologues to the BCCT transporter from *P. tricornutum* and functional transporters OpuD, BetP, DddT and BetT transporters.



The tree shows an early divergence of three branches. The first branch is divided in two groups. The first group englobe the closest homologues to the PtBCCT candidate transporter from the Stramenopiles, eukaryotic picoplankton, and the second sequences from Dinoflagellate. The first group includes the diatoms *P. tricornutum*, *T. pseudonana, T. oceanica* and *F. cylindrus*, and the single-cell protist, *Hondaea fermentalgiana*, from the family of thraustochytriaceaea, class of labyrinthulomycetes usually found in marine and brackish waters[219,220]. In the second group features *Symbiodinium microadriaticum*, photosynthetic unicellular protist belonging the group

of dinoflagellates and it is found in symbiosis with corals[221]. The second branch contains sequences from corals *Dendronephthya gigantea*, *Pocillopora damicornis* and *Orbicella faveolata*. The third and most distantly related BCCT transporters selected from below the cut off value of e^{-107} fall in a clearly distinctive branch of the tree. These include candidate transporters from the actinomycete bacteria *Streptomyces oceani*, the archaeon *Methanohalophilus* sp., and the known functional transporters OpuD, BetP, DddT and BetT (Figure 5.4).

5.3.3 BCCT functionality tests

Candidate BCCT transporter protein from *T. pseudonana*, THAPSDRAFT_262307, sequence is shorter than the other BCCT transporters including the candidate transporters in *F. cylindrus* and *P. tricornutum.* Thus, we did not examine the function of this candidate transporter here.

To test whether the BCCT transporter from *P. tricornutum* was functional or not, attempts were made to express the *P. tricornutum* candidate BCCT transporter in the heterologous host *E. coli*. For that, the sequence of BetT from *E. coli* and PtBCCT transporter were aligned to identify the potential signal peptide from *E. coli*. at the N-terminus.

Figure 5.5. Aligment of N-terminus of BetT and the candidate BCCT transporter from P. tricornutum. In bold, proposed signal peptide of BetT. In grey, sequence replaced by the proposed BetT signal peptide. Conserved amino acids are marked with a *.

```
BetT_Ecoli              ----------------------------------------------------MTDLS
BCCT_Ptricornutum       MGEPSSKTLPRAPEAEVQDDNIEDGNGNLNKETTGTSDSDTDREFDVDKYPVREWEFVIP

BetT_Ecoli              HSREKDKINPVVFYTSAGLILLFSLTTILFRDFSALWIGRTLDWVSKTFGWYYLLAATLY
BCCT_Ptricornutum       GFRDPISINPVVSAIGVIVLWGLAIWCMVDPDGSRETLVGWRGDVTLYFSWLFMGSKAIF
                          *    *****                  * *          *   * *
```

Taking the first conserved amino acids as the reference point, the upstream amino acid sequence from *P. tricornutum* (first 62 amino acid residues) were substituted by

the upstream sequence from *E. coli* (first 7 amino acid residues in bold in Figure 5.5). Hence, the resulting chimeric protein contained the initial 7 amino acids from BetT in the place of the initial 62 amino acids of the *P. tricornutum* protein (Figure 5.6).

Figure 5.6. Amino acid sequence of the chimeric candidate BCCT transporter from Pt. The proposed signal peptide from BetT replacing the non-conserved region of the N-terminus of Pt's BCCT transporter is marked in red.

>Pt BCCT Transporter E. coli chimera
MTDLSHSRDPISINPVVSAIGVIVLWGLAIWCMVDPDGSRETLVGWRGDVTLYFSWLFMGSKAIFFFY
LIYVVFKYGHVKLGRQDEPPEFSTGAYFAMIFAAGVAVGLFVFGVAEPLWHQESHYYANAGYRSQDEI
DMFALNMTVANWGISGWAPYLIVAVAMGLAGHRFNLPMTFRSCFYPILGQYTWGWIGDLIDGFAIVVT
VAGVCTSLGLGAIQIVVGFQYLGWVKDDITQDEVSRVQNATIWVITVIATASVISGLNAGIRILSTIA
FMLGLVLLFLVFVMDDTKYLLNLQVQEVGYYLQHSIFQLNFWTDAFGQIREGGGRAVDGAAAAAWWMD
AWMIFYQAWWVSWSAFVGLFVARISRGRTVSEVIIYSLVAPVAYCIIWFSIWGGVGLRQARQGRELEA
LGGTLFNDTEHFLVPGSTNCYDVPQETLSQDGTVVFENHLLGVTPVCQFDSSQSNTAAFNVLYSFSFP
DSFDTGFGPTLSVLFIISLAIYFATSSDSGSLIVDHLASNGRKNHHWIQRLFWAVTEGAVATALLSAG
GEQALQAVQAASIVCGLPFCFMLCYLLQSIELFCREALIVGDGQDYRIPAQSTFSVPIYGGIFNNMEF
LTSAGSVNPKRIELGMDKATTFHVVEFIKGLFVPFVSLHKVLSDAYPRNSLSNTAVTAVYTVCYYMWI
GIFASLGSKEGLIGWGWLLFFACACILGSVRGGFRARYNVRSNILGDYMASLFFWPQVFTQMRQHCVE
LNLPQDHGDLPSEKEKKLDGSDSDEVAA*

The nucleic acid sequence was codon optimised for expression in *E. coli*, appropriate restriction sites and ribosome biding site were added to the sequence (Appendix 2-1, Appendix 2-2) and then synthesised. The synthesised gene was received in an entry level type commercial plasmid pL0M which is likely based on a pUC19 backbone with *lacZ* promoter[222]. The gene was cloned in the same orientation as the *lacZ* promoter. This construct was mobilised into *E. coli* strains MKH13, 803 and Rosetta.

The transcription of the BCCT gene in this construct was regulated by a *lacZ* promoter, which is known to be a leaky system[223], for this reason and to avoid toxicity effects of expressing a membrane protein in high levels, the samples were not induced. All of the strains containing the chimera and containing a pUC18 empty vector were grown in rich and minimal media with and without the addition of DMSP. Once they were grown, the cells were spun down and the pellets were washed several times with distilled water to eliminate excess of DMSP. If the chimera was functional, it would be expected an accumulation of DMSP in the MKH13 cells, a mutant strain lacking the transport systems *betTIBA, proP and proU*[71], when DMSP was added to the media and no DMSP or only minor traces in any other conditions.

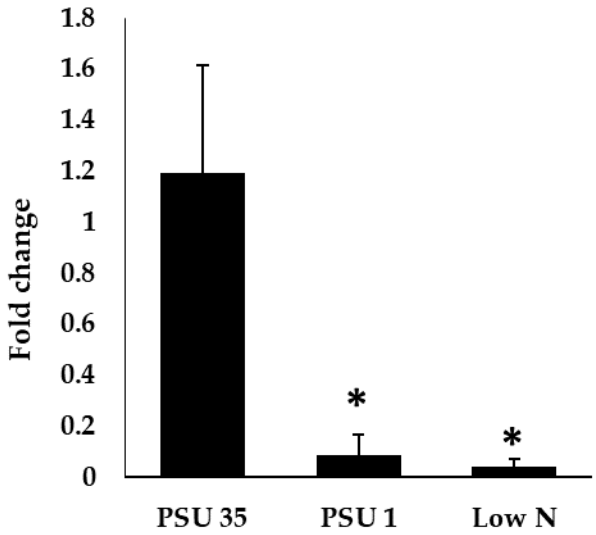
Unfortunately, no transport activity was detected. The same experiments were repeated in the other two different *E. coli* strains 803, which is a strain routinely used for transformations, and Rosetta, a strain used to express eukaryotic proteins with codons seldom used in *E. coli*. Those two strains are capable of importing small quantities of DMSP through their transportation system, thus, an increase in the accumulation of DMSP in the cell expressing the chimeric BCCT transporter would be expected if this were functional compared to cells not expressing the chimera gene. Once again, these tests resulted unsuccessful.

Concomitantly, the codon optimised synthesised chimera was subcloned into pBluescript. Identification of successful subcloned vectors was challenging as the induction of the lacZ operon for blue/white screening was likely toxic for the cells since they would overexpress BCCT transporters severely affecting their membranes. One of the clones contained an insert of the correct size, checked both by amplification of the insert with specific primers and by cutting the insert with restriction enzymes. However, when the clone was sent for sequencing the results showed that the insert had a mutation in one of the amino acids. Unfortunately, no successful clones were obtained, and this work was discontinued.

5.3.4 Regulation of BCCT transport

In Chapter 3, *T. pseudonana* cultures were subjected to a salinity shift and a N depletion treatment. RNA was extracted from those cultures and the whole transcriptome was sequenced. Sequenced results showed that a BCCT transporter was amongst the most upregulated genes being 204.65-fold higher in in increased salinity (35 PSU) compared to low salinity (1 PSU), whereas it was downregulated in N depletion, 0.55-fold difference between samples collected in stationary growth phase of nitrogen starved cultures and silicate starved cultures. Rt-qPCR was used to ratify these data using, using the same RNA as template, qPCR primers targeting *TpBCCT* and β-actin gene[150] to normalise the expression between replicates. As seen

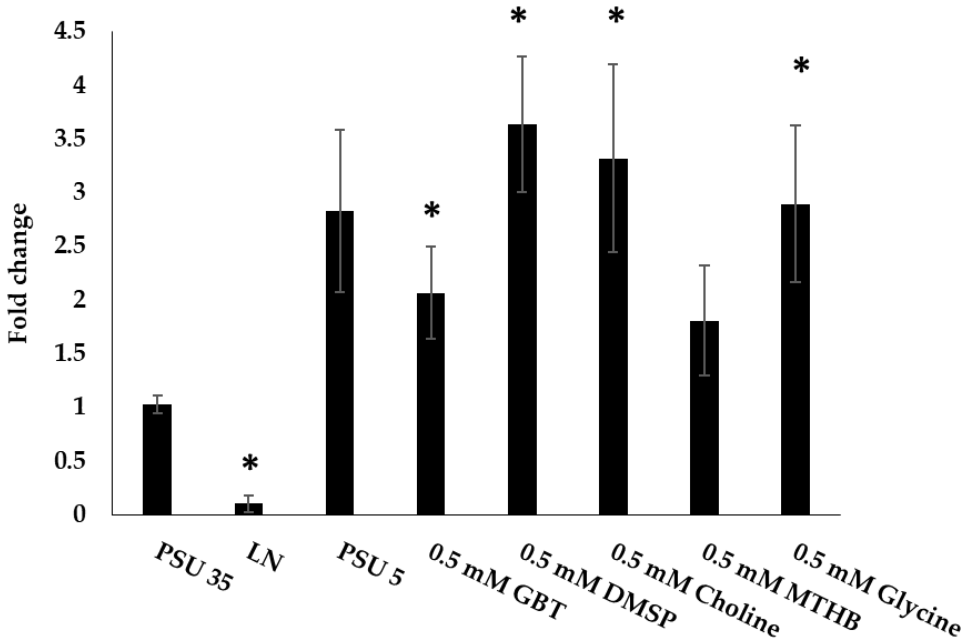
Figure 5.7. Fold change of *TpBCCT* transcription under low salinity and low N. *TpBCCT* transcription was measured by RT-qPCR. Expression is shown as fold change respective to samples grown at standard conditions (PSU35). β-actin gene was used as housekeeping gene to normalise the expression to enable the comparison of the results. Statistically significant results ($p<0.05$) are indicated with a *.



in Figure 5.7, *TpBCCT* is significantly downregulated in lowered salinity and low nitrogen by 0.085 and 0.04-fold, respectively, compared to the standard samples.

In Chapter 3, *P. tricornutum* cultures were also grown in high and low salt concentrations, nitrogen depletion and presence of 0.5 mM DMSP, GBT and their precursors MTHB, glycine and choline. Samples taken from these cultures were used to analyse the transcription of *PtBCCT* using RT-qPCR. qPCR primers were designed to target the candidate transporter and the housekeeping gene exportin 1-like protein[151]. Similarly to results found for the transcription of *TpBCCT*, *PtBCCT* is downregulated by low nitrogen significantly, but it differs as it is not downregulated by lowered salinity compared to normal conditions (Figure 5.8). Perhaps, 5 PSU is not low enough to make a difference in the transcription of this candidate transport, but this needs to be tested by analysing the regulation of this gene under even more decreased salinities[172]. Hypothesis behind the regulation by N will be exposed in the discussion section of this chapter. Moreover, the transcription is significantly higher

Figure 5.8. Fold change of *PtBCCT* transcription under low salinity, low N and addition of GBT, DMSP and their precursors. *PtBCCT* transcription was measured by RT-qPCR. Expression is shown as fold change respective to samples grown at standard conditions (PSU35). Exportin 1-like protein gene was used as housekeeping gene to normalise the expression to enable the comparison of the results. Statistically significant results (p<0.05) are indicated with a *.



in those samples taken from cultures with exogenous source of GBT, DMSP and glycine (Figure 5.8). GB, choline and DMSP are known substrates for BCCT transport[34,71], on the other hand, it has not been shown to be capable of transporting glycine. Nonetheless, it seems that the transcription of this candidate transporter is upregulated by its substrates and the precursor glycine (Figure 5.8).

5.3.5 Cellular localization of PtDSYD.

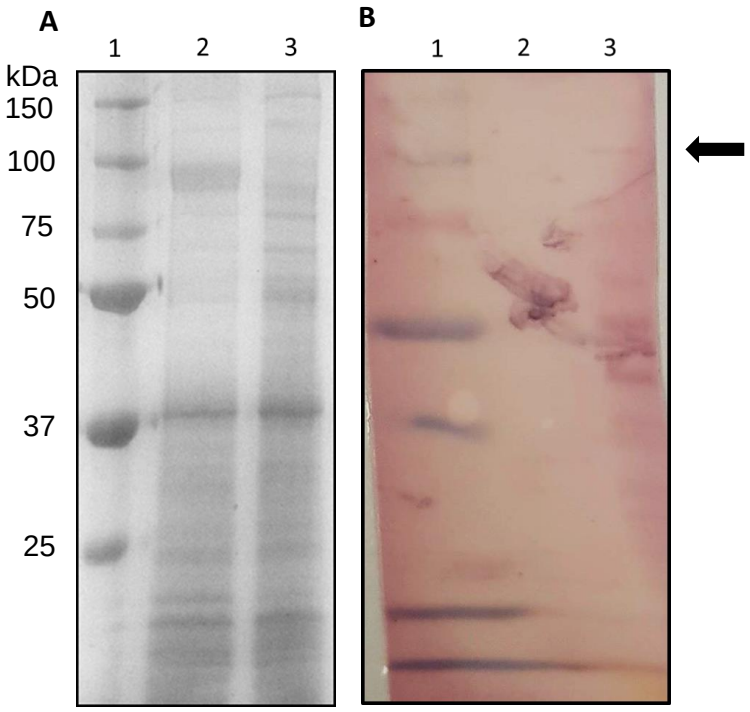
5.3.5.1 *In silico* prediction of targeting motifs.

In silico analysis of the PtBCCT sequence showed no detectable signal peptide, despite being a transporter thus expected to contain a targeting motif to a membrane. The analysis was performed using the available prediction software ASAFind[158], HECTAR[159] and SignalP-5.0[160].

5.3.5.2 Immunogold localization of BCCT in *P. tricornutum* cells.

The candidate BCCT transporter is of a size of estimated 92 kDa. Antibodies against the *P. tricornutum* candidate BCCT protein were designed and tested. With help from Dr. Benjamin Pinchbeck, the specificity of the antibodies was tested by western blot. Western blots were performed in triplicates using samples of *P. tricornutum* grown in either PSU 35, as the positive control, or PSU 5, as the negative control (Figure 5.9).

Figure 5.9. Antibody specificity by western blot. 1) Ladder, 2) Negative control of *P. tricornutum* cell extracts grown at PSU 5, 3) Positive control of *P. tricornutum* cell extracts grown at 35 PSU. A. SDS-PAGE gel of denatured cell extract protein. B. Western blot with antibody 1 against PtBCCT. The arrow marks the band for PtBCCT.

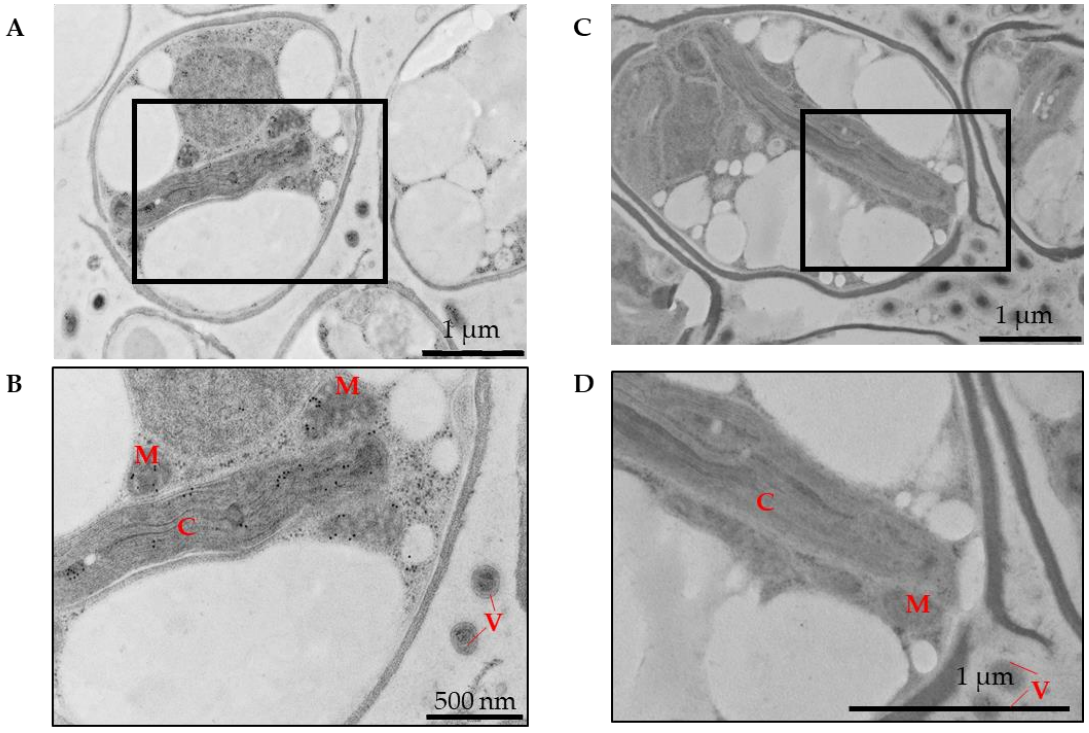


The cell extracts were denatured and run in an SDS-PAGE gel which shows total protein bands from the positive control and negative control cell extracts. The proteins were then transferred to a western blot membrane and the membrane was exposed to the antibody sequence 1. The negative control presented no bands

whereas the positive control showed a band of the approximate correct size of the BCCT transporter, ~100 kDa, marked with an arrow in Figure 5.9. The other bands in the positive control are likely products of degradation of the BCCT protein, however, better images are required without the degradation products and with an additional positive control of a clone of PtBCCT transporter overexpressed in a heterologous host as positive controls.

Immunogold localization of the BCCT transporter was performed in resin *P. tricornutum* cells and electron micrographs were taken (Figure 5.10). The image shows that the transporter is mostly localised on the membranes of the chloroplasts, thylakoid and mitochondria. It is also found in the membranes of unidentified external vesicles.

Figure 5.10. Immunogold localization of BCCT in *P. tricornutum*. Representative electron micrographs of *P. tricornutum* cells showing location of BCCT by immunogold labelling. A, B) Immunostaining of the cell with BCCT antibody and secondary antibody with gold. C, D) Control immunostaining with only secondary antibody. M-mitochondrion, C-chloroplasts, V-external vesicles. Boxes in A and C represent the areas magnified in B and D, respectively.



5.4 Discussion and concluding remarks

Bacterial and Archaea organisms as well as some eukaryotes possess BCCT transporters to import osmolytes[71]. Whereas some of the prokaryotic transporters have been structurally and physiologically characterised[34], eukaryotic transporters have not been functionally studied to date. Candidate BCCT transporters were identified in model diatoms including *F. cylindrus*[98], *T. pseudonana*[128]. In this study, we predict they are also present in the model diatom *P. tricornutum,* the centric diatom *T. oceanica*, and other picoplankton including the Stramenopile *H. fermentalgiana* and the Dinoflagellate *S. adriaticum*. These candidate proteins cluster together and are phylogenetically divergent to other predicted BBCT from bacteria such as *S. oceani* the archaeon *Methanohalophilus* sp. or corals such as *D. gigantea*, *P. damicornis* and *O. faveolata* and from other characterised transporters (Figure 5.4).

Candidate BCCT proteins from *P. tricornutum* and *F. cylindrus* differ from the canonical structure of 12 TMS[214] having up to 4 extra TMS domains. Whereas *T. pseudonana* is significantly shorter with only 8 predicted TMS. It is likely that the publicly available candidate BCCT sequence from *T. pseudonana* is incomplete.

Attempts to prove the functionality of the candidate BCCT transporters from *P. tricornutum* were unsuccessful. There are several reasons that could explain the negative results obtained: i) The C- terminal from *P. tricornutum* sequence play an important role in the functionality of the protein and removing this section in the chimeric protein tempered with the functionality (Figure 5.5); ii) The C-terminal sequence from *E. coli* combined to the chimeric sequence is not enough to target the protein to the bacterial membrane when expressed in the heterologous host *E. coli* (Figure 5.5); iii) the substitution of the third glycine for an alanine residue in the Gly box responsible for the substrate specificity and activation of the protein[34] could imply that the naturally occurring transporter in *P. tricornutum* has a truncated function, another possibility is that a different host is necessary.

An RNA study performed across *T. pseudonana* and *P. tricornutum* showed that the BCCT transport is among the top 20 most transcribed gene in these two model diatoms[224]. An analysis of the transcription of *TpBCCT* performed by whole transcriptome sequencing and RT-qPCR shows that it is highly expressed in normal conditions and downregulated by lowered salinity and N limitation (Figure 5.7). RT-qPCR was also used to examine the transcription of *PtBCCT* which appeared to be regulated by nitrogen availability and presence of the substrates GBT, DMSP, choline and glycine. Salinity, on the other hand, did not have a significative effect on the transcription regulation of *PtBCCT* (Figure 5.8), perhaps because *P. tricornutum* stills requires of osmolytes under the osmotic pressure at 5 PSU. This hypothesis requires further testing.

Despite not having any known predicted target signal peptide, immunogold localisation of PtBCCT situated the transporter in the membranes of the chloroplasts, mitochondria and external vesicles of unknown nature (Figure 5.10). Curson *et al.* 2018 showed that DSYB, the key enzyme in the synthesis of DMSP was also localised in the chloroplasts and mitochondria and that DMSP generally also accumulates in those organelles in *P. parvum*[99]. These suggest that the BCCT transporter is likely playing a role in the distribution of osmolytes within the cell and in between the organelles and the cytoplasm. For instance, in nitrogen limiting conditions DMSP could be maintained in the organelles fulfilling an antioxidant role[78], whereas in increased salinity it could be transported to the cytoplasm to protect the cell against osmotic stress[23].

Unfortunately, some of the results in this chapter are inconclusive, so it is important to highlight that functional tests and the repetition of the western blot are required before reaching firm conclusions. Findings in this chapter are promising and are informative about which transport mechanism is likely to be present in phytoplankton. In addition, as PtBCCT is likely to be placed in the membranes of the organelles (Figure 5.10), the mechanism whereby *P. tricornutum* import osmolytes

from the environment[131] remains unknown as much as the nature of the external vesicles observed under the microscope (Figure 5.10).

Chapter 6

Identification and characterization of DSYD in diatoms

6 Identification and characterization of DSYD in diatoms

6.1 Introduction

Results from this chapter were obtained in collaboration with, Dr Andrew Curson, Dr Chun Yang Li and Dr Elena Mercade in the following sections: Dr Andrew Curson cloning and functionality check by GC. Dr Chung Yan biochemistry and Dr Elena Mercade immunogold labelling experiments.

DMSP producing diatoms lacking *DSY*B- like genes are likely to contain a novel SAM-dependent methyltransferase. In what follows, I will introduce what is known on the assaying of DsyB/DSYB enzymes since it is relevant to potential MTHB methyltransferase isoform enzymes identified here. In Curson *et al.* 2018[99], candidate DMSHB synthases from phytoplankton were cloned and expressed in the *L. aggregata* LZB033 *dsy*B mutant and in *R. leguminosarum* (J391) as heterologous hosts and their MTHB methyltransferase activity was tested. These two heterologous hosts are capable of catalysing all the steps involved in the transamination pathway, apart from the conversion of MTHB into DMSHB[89] The introduction and expression of the DSYB proteins conferred onto the hosts the ability to produce DMSP from methionine. Interestingly, *DSY*B and *dsy*B like proteins are not functional when expressed in the heterologous host *E. coli*, a host incapable of transforming DMSHB to DMSP[89,99].

DSYB enzymatic assays were performed with purified enzyme with addition of *P. parvum* cell extracts and monitoring the production of DMSHB and/or DMSP from MTHB by GC[99]. Thus, it is likely that the known DsyB enzymes require some unknown co-factors to be functional.[99] DMSHB was not formally detected as the product of the DsyB-like proteins in these assays.

In this chapter, candidate DMSHB synthases/MTHB methyltransferases will be tested in the heterologous host *R. leguminosarum* and, if possible, in *E. coli.* Using *E. coli* as a

heterologous host would perhaps shed some light as to whether these diatom MTHB methyltransferases are indeed producing DMSHB. Enzymatic assays will also be performed with purified enzymes to elucidate whether these proteins require co-factor/s. The analytical methods used were gas chromatography (GC) and liquid chromatography mass spectrometry (LC/MS) to detect DMSHB and/or DMSP.

6.2 Aims of the chapter

In this chapter, the candidate DSYD enzymes (D for diatoms) identified in chapter 4 will be examined for MTHB methyltransferase activity using the heterologous hosts *R. leguminosarum* and *E. coli*. Functional DSYDs will be further characterised. In addition, databases will be searched to explore which organisms possess DSYD-like proteins.

Using the samples described in chapter 3, regulation of the *DSYD* genes under different environmental conditions will be ratified by RT-qPCR. Finally, using immunogold labelling imaging, the localization of DSYD in *P. tricornutum* cells will be observed and compared to the cellular localization of DSYB.

6.3 Characterization of DSYD

In chapter 4, two candidate methyltransferases regulated by both salinity and nitrogen in the same manner as DMSP production were identified in *T. pseudonana*: the predicted protein SAM dependent carboxyl methyltransferase THAPSDRAFT_11247 (XP_002296978.1) and the methyl transferase-like protein THAPSDRAFT_269095 (XP_002291473.1). Whereas in *P. tricornutum*, only the candidate methyltransferase PHATRDRAFT_48704 (XP_002183266.1), homologue to THAPSDRAFT_269095 (XP_002291473.1), meet the criteria of being downregulated by lowered salinity.

THAPSDRAFT_11247 (XP_002296978.1) was codon optimised (Appendix 3-1) for expression in *E. coli* and synthesised with a ribosome binding site and a stop codon

at the end of the sequence. Appropriate restriction sites were also added (Appendix 2-1) to allow subcloning into the alpha-proteobacteria specific vector pLMB509 for expression in *R. leguminosarum* and pET21a for expression in *E. coli*. The resulting sequence was sent to be synthesised and subcloned as described in the materials and methods (Chapter 2). The resulting clones were named pBIO2288 (pLMB509-11247) and pBIO2289 (pET21a-11247).

On the other hand, THAPSDRAFT_269095 and PHATRDRAFT_48704 were PCR amplified from cDNA and were cloned into pLMB509 and pET21a, also as described in the methods chapter (Chapter 2). The resulting clones were renamed as pBIO2291(pET21a-269095) for *T. pseudonana* insert and pBIO2292 (pLMB509-48704) and pBIO2293 (pET21a-48704) for *P. tricornutum* insert.

Subsequently, these clones were conjugated or transformed into their corresponding host organisms: pBIO2288, pBIO2290 and pBIO2292 into Streptomycin-resistant *Rhizobium leguminosarum* strain J391, and pBIO2289, pBIO2291, pBIO2293 into *E. coli* strain BL21.

Protein production was induced by addition of taurine or IPTG (see chapter 2) to the media cultures. The strains containing the candidate methyltransferases where grown in the presence and absence of the substrate MTHB (the precursor of DMSHB), in both rich and minimal media. Then, samples of the cultures were taken and analysed by gas chromatography (GC) for the detection of DMSHB and/or DMSP.

Tested THAPSDRAFT_11247 in *E. coli* and *R. leguminosarum* did not present any methyltransferase activity and this enzyme was dismissed as encoding a MTHB methyltransferase. In contrast, *E. coli* BL21 expressing the homologous genes THAPSDRAFT_269095 from *T. pseudonana* and PHATRDRAFT_48704 from *P. tricornutum* acquired the ability to transform MTHB into DMSHB and/or DMSP. The *P. tricornutum* PHATRDRAFT_48704 was also functional when tested in *R. leguminosarum* (Table 6-1).

The functional methyltransferases (THAPSDRAFT_269095 and PHATRDRAFT_48704) were named as DSYD (**D**MSHB **SY**nthase in **D**iatoms), the letter D was assigned to differentiate this new family of proteins from the known DSYB due to the very low homology presented between these two (see chapter 4).

6.3.1 Identification of functional DSYD in other organisms

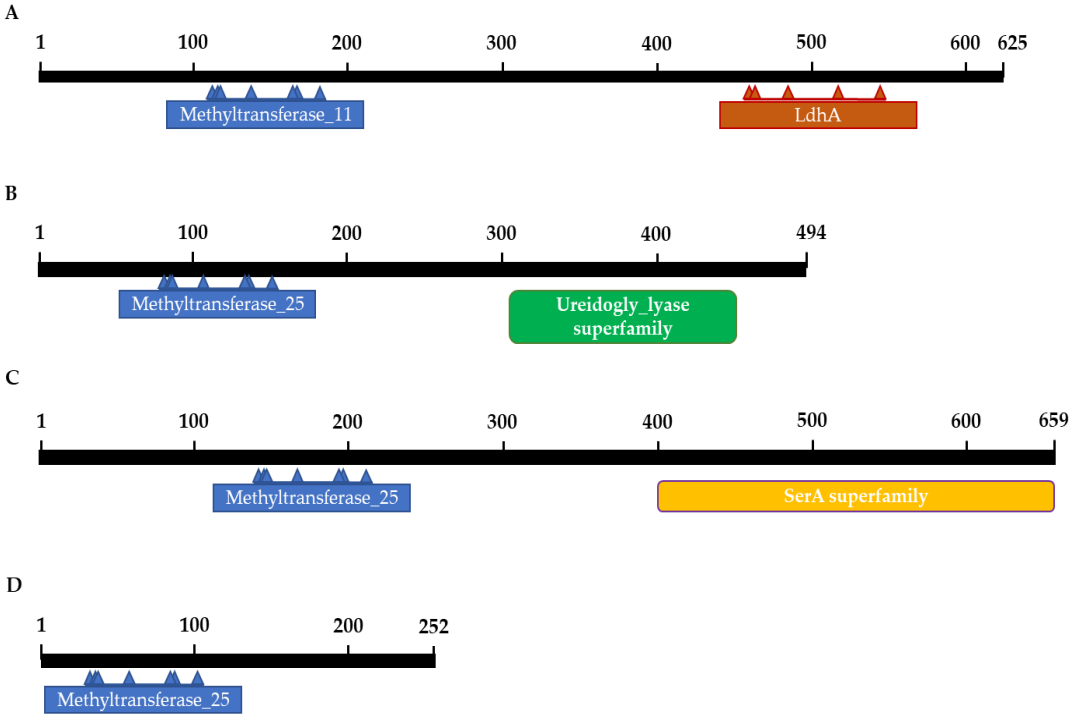
Phylogenetic trees from TpDSYD (Appendix 3-2) and PtDSYD (Appendix 3-3) were produced using the distance tree of results tool at NCBI BLASTp. DSYD protein sequences from *T. pseudonana* and *P. tricornutum* have close homologues in the diatom species *T. oceanica*, *F. solaris* and *F. cylindrus*, but more distantly related methyltransferases to Tp and PtDSYD are annotated in the database as involved in glycine betaine synthesis via the glycine pathway.

DSYD homologues present e values greater than e^{-45} and percentage of identity greater than 40% to the TpDSYD. On the other hand, homology is slightly lower to the PtDSYD, with e values greater than e^{-37} and identity equal or above 34.68%. *F. cylindrus* homologue is a short version of DSYD that lacks N terminus and has no annotated stop codon, therefore is likely misannotated or non-functional.

In addition to the homologues in diatoms, both *P. tricornutum* and *T. pseudonana* had a candidate methyltransferase domain-containing protein from the rhizobacterium *G. sunshinyii* (WP_044616208.1). This bacterial protein presents a decreased homology with a e value of $3e^{-29}$ and 32.19% of identity to TpDSYD and an e value of $3e^{-33}$ and 32.40% identity to PtDSYD. It is worth noting that this bacterial candidate *dsy*D is substantially larger than the TpDSYD (40% coverage) and PtDSYD (53% coverage). *G. sunshinyii* candidate DsyD has an additional domain in the C-terminus, an ureidogly lyase domain (pfam04115) (Figure 6.1). The ureidogly lyase enzymes lyases ureidoglycolate releasing urea[225], however this domain has very low homology to other ureidoglycolate lyases.

Additionally, *T. oceanica* (To) has two homologues to TpDSYD, Prot ID 31798 (ToDSYD1) and ProtID 20742 (ToDSYD2). The candidate ToDSY2 and the DSYD-like methyltransferase from *F. solaris* are also larger than TpDSYD. These two proteins

Figure 6.1. Conserved domains of the candidate DMSHB synthesases from A) *F. solaris* GAX25165.1, B) *G. sunshinyii* WP_044616208.1, C) *T. oceanica* EJK59074.1 and D) *T. pseudonana* XP_002291473.1. The numbers indicate the amino acid coordinates.



contain a NAD binding domain (pfam02826) in the C-terminus from the SerA superfamily (Figure 6.1). This domain is a component of a larger lactate dehydrogenase (LdhA) or related 2-hydroxyacid dehydrogenase (COG1052) family (Figure 6.1). It is similar to the GRHPR protein of Homo sapiens ($6e^{-27}$), an enzyme with hydroxypyruvate reductase, glyoxylate reductase, and D-glycerate dehydrogenase activities. A hypothesis is that these extra domains might be involved in other parts of the DMSP synthesis pathways, such as being involved in the reduction of MTOB. This has not been further investigated but it would be interesting to pursue in the future.

The methyltransferases from *T. oceanica,* To DSYD1 and To DSYD2, and the predicted outlier *G. sunshinyii* (WP_044616208.1) were codon optimised (Appendix 3-1) for expression in *E. coli*, *Nde*I *Eco*RI/*Bam*HI sites were added for subcloning and then were synthesised. The synthesised genes were cloned into the plasmid pUC57. Clones were renamed as pBIO2294 (pUC57-*ToDSYD1*), pBIO2295 (pUC57-*ToDSYD2*), pBIO2296 (pUC57-*gsdsyD*) and transformed into the heterologous host *E. coli* BL21 strain. Heterologous hosts containing DSYD-like proteins were grown in minimal media and addition of 0.5 mM MTHB, then samples were taken and analysed by GC for DMSHB/DMSP detection.

Table 6-1. List of functionality of candidate DSMHB synthases. Candidate genes were cloned in pET21a or pLMB509 and transformed or conjugated into *E. coli* or *R. leguminosarum*. Bacteria containing the clones were grown in minimal media and addition of MTHB. Samples were taken and total content of DMSP/DMSHB was detected by GC. 'Yes' inditicates successful methylation of MTHB , 'No' indicates that no activity was detected in the host. If tests were not performed in that host it is marked as '-'.

Organism	Accession number	*E. coli* BL21	*R. leguminosarum* J391
P. tricornutum	XP_002183266.1	Yes	Yes
T. pseudonana	XP_002291473.1	Yes	-
*T. oceanica*_1	EJK49332.1	Yes	-
*T. oceanica*_2	EJK59074.1	No	-
G. sunshinyii	WP_044616208.1	Yes	-
O. tauri	CEG01383.1	-	Yes
A. halophila	SDW35278.1	No	No
N. exalbescens	WP_028481482.1	No	No
M. bouillonii	WP_075904519.1	No	No
S. vulgare	Sample_MAC_c89409. graph_c1_seq1	-	Yes
P. haitanensis	Pyropia haitanensis Unigene84_All	No	No

DSYD1 from *T. oceanica* (ID 31798) and the DsyD from *G. sunshinyii* conferred *E. coli* BL21 strain the ability to produce DMSHB/DMSP from MTHB. On the other hand, candidate DSYD 2 from *T. oceanica* (ID 20742) was not functional when expressed in *E. coli* (Table 6-1). It would be interesting to test this enzyme in a different host such as *Rhizobium leguminosarum* in case this methyltransferase requires an additional co-

factor not present in *E. coli*[99], as well as to test for functions related with its extra NAD binding domain.

Another NCBI BLASTp search was conducted using *T. oceanica* DSYD1 and a phylogenetic tree was produced using the distance tree of results tool at NCBI BLASTp (Appendix 3-4). Sequences were selected from across the tree including the methyltransferase type 11 from the unicellular green algae *Ostreococcus tauri* (CEG01383.1), the candidate sarcosine/dimethylglycine N-methyltransferase from the proteobacteria belonging to the family Halomonadaceae *Aidingimonas halophila* (SDW35278.1) and the SAM-dependent methyltransferase from the Rhodobacteraceae *Nesiotobacter exalbescens* (WP_028481482.1). The hypothetical protein WP_075904519.1 from the cyanobacteria *Moorea bouillonii* was also selected. None of the candidates showed methyltransferase activity in either of the tested heterologous hosts excepting TpDSYD, PtDSYD, ToDSYD1, GsdsyD and the DSYD from the unicellular green algae *O. tauri* (Table 6-1), which clustered with the DsyD from *G. sunshinyii* (Appendix 3-4). DSYD from *O. tauri* is the first gene involved in DMSP synthesis ever identified and characterised in a green alga.

DSYD homologues are likely present in another alga. To investigate this, DSYD functional proteins were used as tBLASTn against transcriptome assemblies (TSA). Homologues were identified in the brown algae *Sargassum vulgare* (Sample_MAC_c89409.graph_c1_seq1) and from the red algae *Pyropia haitanensis* (Pyropia haitanensis Unigene84_All), with a percentage of identity to DSYD from *O. tauri* of 39.72 and 29.60% respectively. To test whether these proteins have MTHB methyltransferase activity, the candidate *DSYD*s from the brown algae *S. vulgare* and from the red algae *P. haitanensis* were codon optimised (Appendix 3-1), synthesised and tested by heterologous expression.

S. vulgare DSYD was subcloned into pLMB509 (pBIO2302) and proved to be functional when tested in *R. leguminosarum* J391 (Table 6-1). Unfortunately, candidate

DSYD from *P. haitanensis* did not show any activity when subcloned into pLMB509 (pBIO2303) and mobilised into J391 (Table 6-1).

6.3.2 Distribution of homologues to functional DSYD

Neighbour joining distance tree of the results from the searches in the NCBI database using BLASTp and TpDSYD (Appendix 3-2), PtDSYD (Appendix 3-3) and DSYD from the brown algae *Sargassum vulgare* (Appendix 3-5) were produced.

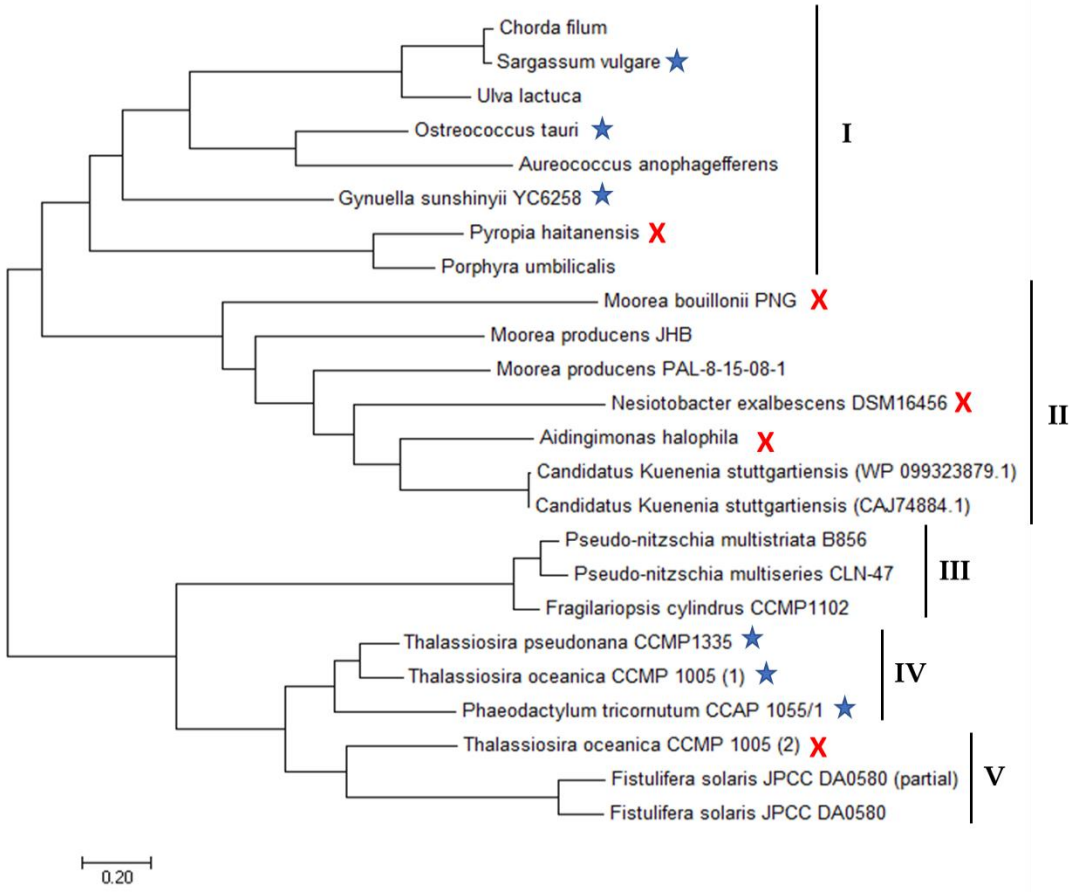
Homologues of TpDSYD are distributed in seven different Clades. Clades I, II, III and IV are comprised of methyltransferases from a variety of bacteria. In Clade I is found a methyltransferase from the alpha-proteobacteria *Erythrobacter litoralis* capable of synthesising both GBT and DSYD and that is further discussed in Chapter 7. The non-functional DSYD-like protein from *Nesiotobacter exalbescens* stands out in Clade II. Clade III and VI is predominately formed by actinobacteria from the genus *Streptomyces* and *Nocardiopsis*, whereas Clade V contains a methyltransferase from the cyanobacteria *Moorea producens* which is closely related to *Moorea bouillonii*, a cyanobacteria with a non-functional DSYD-like protein. Finally, Clade VII uniquely formed by diatom sequences, including the functional DSYD from *T. oceanica*, *P. tricornutum* and *T. pseudonana* itself. According to this tree, functional DsyD from *G. sunshinyii* is completely divergent from the other homologues. Strikingly, none of the functional DSYD sequences from alga are included within the top 100 most similar proteins to TpDSYD (Appendix 3-2). The phylogenetic tree originated form the top 100 most related sequences to PtDSYD is divided in six different clades. Similarly, the DSYD-like methyltransferases from diatoms are also clustered together (Clade VI). Clades I - IV englobe methyltransferases from prokaryotic origin. For example, the methyltransferase from *E. litoralis* appears in Clade II and its functional counterpart from *Mamelliela alba* (also discussed in Chapter 7) appears in Clade III, together with the non-functional DSYD-like protein from *N. exalbescens*. This tree, however, includes alga sequences, for example, from the red algae *Porphyra umbilicallis* which clusters with the functional DsyD from *G. sunshinyii* (Appendix 3-3). Analysis of the

phylogenetic tree produced using the functional DSYD from *Sargassum vulgare* further suggests that there is a division between the diatom sequences and the other DSYD-like methyltransferases. The tree is divided into five different clades, Clade II-IV of them are constituted by bacteria with a predominance of actinobacteria in Clade II, with methyltransferase from the genus *Streptomyces* sp., and Clade III from the genus *Actinopolyspora* and *Saccharomonospora*. Methyltransferases from methanoarchaeon, such as *Methanohalophilus portucalensis* are grouped in the Clade V. Most remarkedly, Clade I comprises the functional DSYD from *S. vulgare*, from *G. sunshinyii* and from the green algae *O. tauri*. None of the closest homologues to the functional DSYD from *S. sargassum*, belong to diatoms (Appendix 3-5). To further understand the relatedness between the functional DSYD, a Maximum Likelihood phylogenetic tree was produced using Mega v.X[156] (Figure 6.2).

Sequences from the functional DSYD, candidate DSYD with no MTHB methyltransferase activity when expressed in the heterologous hosts used in this study as well as non-tested candidate DSYD (Appendix 3-6) were aligned and a phylogenetic tree was produced. The apparent dichotomy between the diatom sequences and the sequences from other organisms, such as macroalga, cyanobacteria and proteobacteria, observed in the analysis of the phylogenetic trees from TpDSYD, PtDSYD and DSYD from *S. vulgare* is also noticeable. In the first branch, there are two groups of enzymes. The group I contains functional DSYD from the unicellular green algae *O. tauri*, the DSYD from the brown algae *S. vulgare* and the DSYD with the extra domain in the C-terminus from *G. sunshinyii*. These functional DSYD also group with candidate methyltransferases from the brown alga *Chorda filum*, the green algae *Ulva lactuca* and the heterokont algae *Aureococcus anophagefferens*. Additionally, this group also includes the non-functional methyltransferase and the non-tested candidate DSYD from the red alga *Pyropia haitanensis* and *Porphyra umbilicallis*, respectively, showing that they are closely related. Group II includes bacterial sequences, such as the non-functional DSYD-like proteins from *Nesiotobacter exalbescens* and *Aidingimonas halophila* and the cyanobacteria *Moorea bouillonii*. Groups III, IV and V

are formed exclusively by diatom sequences. In the Group III, three non-tested methyltransferases from *F. cylindrus* and *P. multistriata* group together. The three tested and functional DSYD from *T. pseudonana*, *P. tricornutum* and *T. oceanica* constitute Group IV. Finally, Group V embody the ToDSY2 and the candidate methyltransferase from *F. solaris*. Both enzymes contain an additional NAD binding domain in the C-terminus (Figure 6.1).

Figure 6.2. Maximum Likelihood phylogenetic tree of the methyltransferases discussed in this study. The phylogenetic tree shows functional DSYD in blue methyltransferases tested for DMSP syntheses that were non-functional in the tested host and sequences from non-tested DSYD homologues. Group I correspond to functional and non-functional DSYD from macroalga and the bacteria *G.* sunshinyii. non-tested methyltransferases. Group II to non-functional bacterial enzymes. Groups III, IV and V englobe diatom sequences, non-tested (III), functional (IV) and diatom sequences with an extra NAD-binding domain (V). Functional proteins are marked with stars, non-functional proteins are marked with a cross.

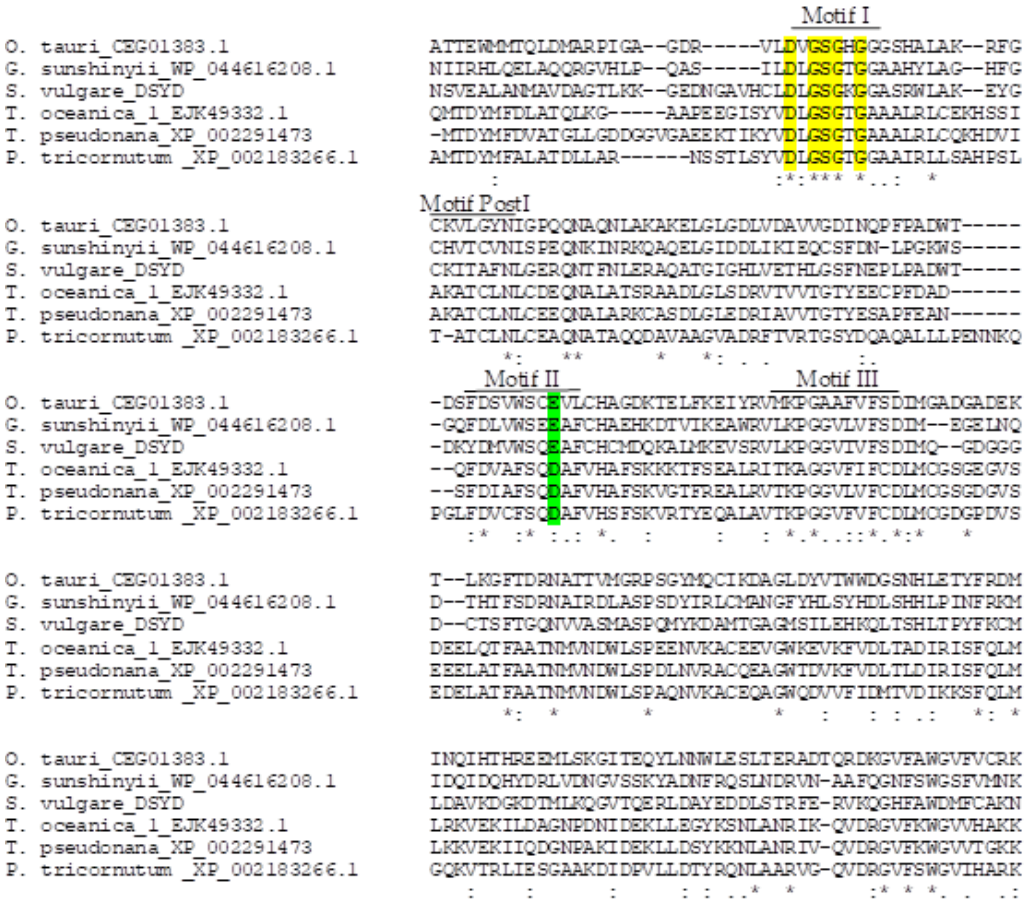


Adding more tested sequences to this tree would allow to clarify whether there has been a clear and significant evolutionary divergence between the DMSHB synthases from diatoms lacking the DSYB enzyme[99] and those other novel DMSHB synthases found in macroalga and bacteria. Furthermore, the similarity between the non-functional methyltransferases to the methyltransferases involved in GBT synthesis suggest that those enzymes might be involved in the synthesis of the nitrogenous osmolyte. Further work needs to be carried out to investigate the substrate specificity of the functional and non-functional DMSHB synthases studied in this chapter.

6.3.3 Analysis of the secondary structure of the functional methyltransferases

Using Phyre2, the secondary structure of TpDSYD was predicted. In this methyltransferase, the first β strand (β1) is followed by the conserved SAM-methyltransferase Motif I DxGSGxG, in which the GxGxG sequence causes an extended turn[226]. Motif Post I is not conserved at the primary structure level among SAM-Methyltransferases but it forms the conserved second β strand (β2). Motif II and Motif III usually are the fourth and fifth β strands (β4, β5), however, the predicted secondary structure of TpDSYD suggests that Motif II and Motif III are the third and fourth β strands (β3, β4). Furthermore, the Motif II interacts with both the substrate and with the methyl group for the methylation, hence it is likely to be involved with the substrate specificity.[226] Tested functional DSYDs have the most common conserved amino acids except for the semiconserved negative amino acid residues Asp (D) and Glu (E) (Figure 6.3). There is a conserved histidine (H) after the motif II that is conserved amongst other methyltransferases, including those involved in GBT synthesis (Chapter 7). This amino acid is preceded by a leucine (L), an amino acid with a hydrophobic chain, in the GBT and GBT/DSYD proteins whereas in DSYD it is preceded by a sulfur containing amino acid cysteine (C) in *G. sunshinyii*, *O. tauri* and *S. vulgare* and by a valine (V), also an amino acid with hydrophobic chain, in diatoms.

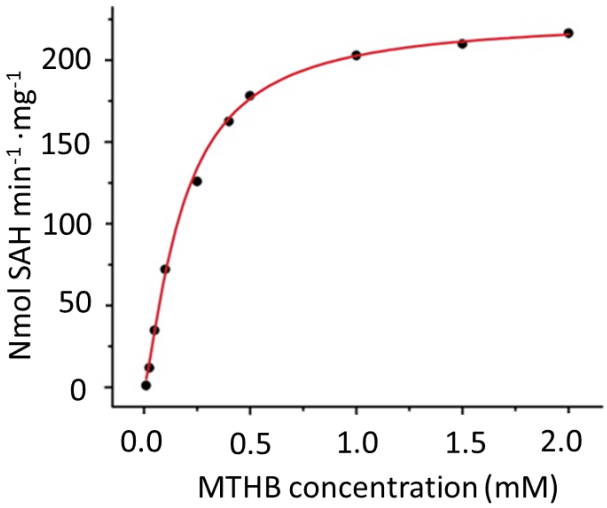
Figure 6.3. Amino acid alignment of functional DSYD enzymes and the (*) conserved amino amino acids (:) semiconserved and (.) amino acids within the same group.



6.3.4 *In vitro* characterization of PtDSYD methyltransferase activity

Cultures of *E. coli* BL21 containing pBIO2293 (pET21a-PtDSYD) were grown in rich LB medium and the expression of PtDSYD was induced by addition of IPTG, PtDSYD was purified. The enzyme activity was measured using a HPLC to monitor the production of S-adenosyl homocysteine (SAH) resulted from the demethylation of SAM. Optimal DSYD activity was determined by testing temperature and pH conditions. The activity from each condition are given as a percentage of the highest activity achieved. To find the optimal temperature, the reaction mixtures were incubated at temperature intervals of 10 °C, from 0 °C to 60 °C, for 30 min. For optimal pH levels, DSYD activity was observed at pH values between pH 5.0 and pH 10.0

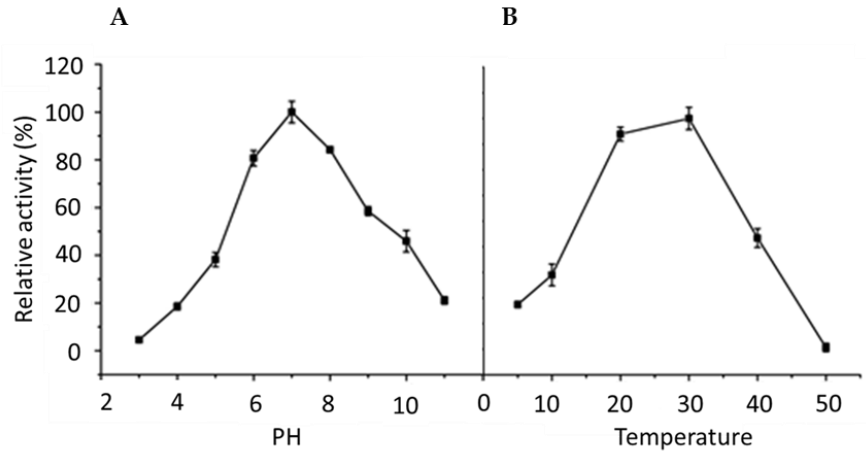
Figure 6.4. Non-linear fit curve for MTHB methylation by PtDSYD. K_M was 186±9 µM Vmax value was 225.46±3.94 nmol·min^{-1}·mg^{-1}.



(Figure 6.5). K_M values were determined by nonlinear analysis based on the initial rates (see Chapter 2 for detailed method).

Enzymatic assays were performed at the optimum pH 6.5 and temperature 30°C (Figure 6.5). The concentration of the SAH produced from the demethylation of SAM per minute of reaction and mg of purified DSYD was plotted against the concentration of the enzyme substrate, MTHB, added to the reaction. The kinetic

Figure 6.5. Effect of pH and temperature on the enzymatic activity of PtDSYD. Activity was defined as the percentage of the highest activity achieved A) at pH 6.8 and B) at 30°C.



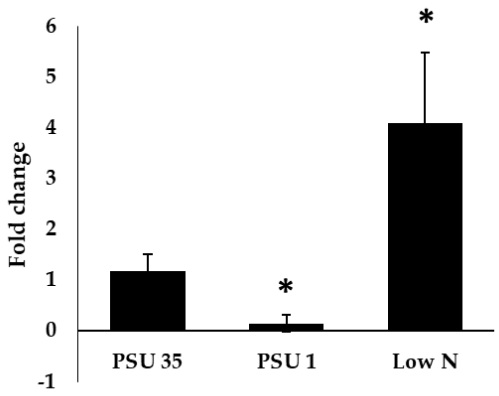
parameter (K_M) value of PtDSYD for MTHB was 186±9 μM (Figure 6.4) which is 2.1-fold higher than the K_M value reported of DSYB for MTHB (88.2 μM)[99].

6.3.5 Transcriptional regulation of *DSYD* genes in the model diatoms.

6.3.5.1 Transcriptional regulation of *DSYD* by salinity and N availability in *T. pseudonana.*

In Chapter 3, the synthesis and transport of DMSP and GBT under increased or decreased salinity and limiting nitrogen cultures of *T. pseudonana* were explored by the analysis of small metabolites by NMR. Samples were also taken for transcriptomic analysis by whole RNA sequencing. Sequencing results showed that 82 genes were

Figure 6.6. Fold change of *TpDSYD* transcription under low salinity and low N measured by RT-qPCR. Expression is shown as fold change respective to samples grown at standard conditions (35 PSU). β-actin gene was used as housekeeping gene to normalise the expression to enable the comparison of the results. Statistically significant results (p<0.05) are indicated with a *.



upregulated, and 92 genes were downregulated by increased salinity. *DSYD* was one of the most upregulated genes in increased salinity (35 PSU) compared to low salinity (1 PSU). When nitrogen was limiting in the culture, 64 genes were upregulated, and 237 genes were downregulated compared to cultures with excess of nitrate in the medium. DSYD was amongst the most upregulated genes, increasing by 49.23-fold when cultures grown in low salt were transferred to normal salinity, and by 88.22-

fold when nitrogen became limiting compared to nitrate replete cultures. RNA sequencing results were ratified by RT-qPCR using primers targeting *DSYD* and using the β-actin[150] gene to normalise the expression between treatments.

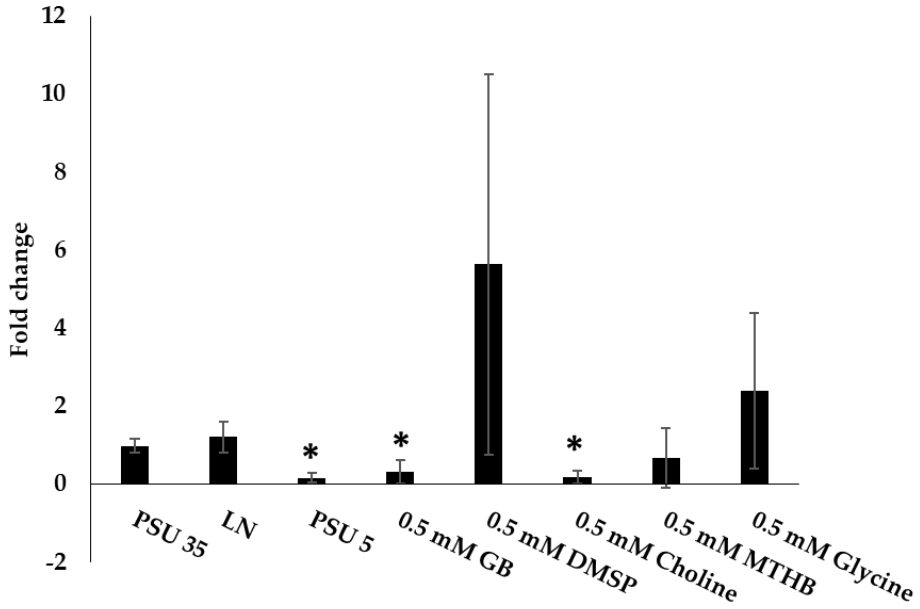
RT-qPCR results show that the transcription of the *TpDSYD* changes significantly at low salinity and low N when compared to the standard condition (35 PSU). At low salinity, *DSYD* transcription is 19-fold lower than in standard cultures. Conversely, transcription of this gene when nitrogen is limiting is almost 9-fold higher than the standard treatment (Figure 6.6). These results ratify the regulation observed through RNA-sequencing. Moreover, *TpDSYD* is regulated in the same manner as the accumulation of DMSP in the diatom.

6.3.5.2 Transcriptional regulation of *DSYD* by salinity and N availability and addition of DMSP, GBT and their precursors in *P. tricornutum.*

P. tricornutum samples were taken from cultures grown in standard conditions (35 PSU), low salt (5 PSU), nitrogen limitation and addition of 0.5 mM GBT, DMSP and their precursors (see chapter 3). RNA was extracted for analysis of the transcription of *DSYD* by RT-qPCR. Primers designed to target *DSYD* were designed and the gene encoding exportin 1-like protein[151] was used as a housekeeping gene to normalise the gene expression.

Results show that transcription is 6-fold and 3-fold lower in the cultures grown at low salinity and addition of GBT compared to standard cultures (Figure 6.7). This regulation is expected, and it agrees with the metabolite analysis results by LC/MS which illustrates a significant decrease in the accumulation of DMSP in the cell in those conditions (Chapter 3). However, there are two conflicting results between transcriptional regulation and metabolite regulation. LC/MS results show a significant increase in DMSP concentration in the cells subjected to a long period of nitrogen starvation (see Chapter 5), however, this increase in metabolite is not reflected in an upregulation of *DSYD* (Figure 6.7). Likewise, *DSYD* transcription

Figure 6.7. Fold change of *PtDSYD* transcription under low salinity, low N and addition of GBT, DMSP and their precursors. *PtDSYD* transcription was measured by RT-qPCR. Expression is shown as fold change respective to samples grown at standard conditions (35 PSU). Exportin 1-like protein gene was used as housekeeping gene to normalise the expression to enable the comparison of the results. Statistically significant results ($p<0.05$) are indicated with a *.



appears to be significantly downregulated by the addition of choline (Figure 6.7), whereas no significant difference was found in the metabolite analysis (Chapter 3). These discrepancies could be explained by the effect of post-transcriptional and post-translational regulation or effects in the metabolite transport that contributes to the accumulation of intracellular DMSP concentrations.

6.3.6 Intracellular localization of DSYD.

DMSP production in higher plants occurs in chloroplasts via the transamination pathway[93]. Curson *et al.*[99] proposed that DMSP was also compartmentalised, being synthesised in the chloroplasts and the mitochondria in DSYB containing phytoplankton. In this section, localization of DSYD in the model diatom *P. tricornutum* will be explored.

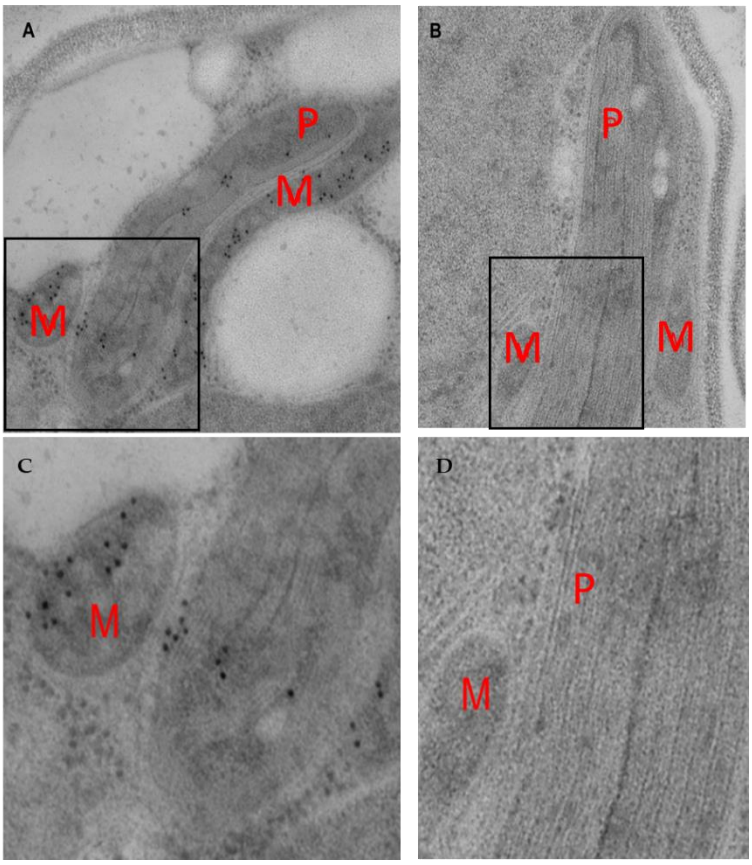
6.3.6.1 *In silico* prediction of signal peptides.

Firstly, the amino acid sequence of PtDSYD was searched for known signal peptides targeting diatom organelles. For this, three programs were used ASAFind[158], HECTAR[159] and SignalP-5.0[160]. None of these methods predicted the presence of a signal peptide in PtDSYD.

6.3.6.2 Immunogold localization of DSYD in *P. tricornutum* cells

Once again, *P. tricornutum* was used as the model diatom to perform immunogold labelling of DSYD over other diatoms as silification only occurs in one valve of the

Figure 6.8. Immunogold localization of DSYD in *P. tricornutum*. Representative electron micrographs of *P. tricornutum* cells showing location of DSYD by immunogold labelling. A, C) Immunostaining of the cell with DSYD antibody and secondary antibody with gold. B, D) Control immunostaining with only secondary antibody. M-mitochondrion, P-chloroplasts. Boxes in A and B represent the areas magnified in C and D, respectively.



cell[227] allowing the penetration of the gold particles and the visualization of the hybridised primary and secondary antibodies.

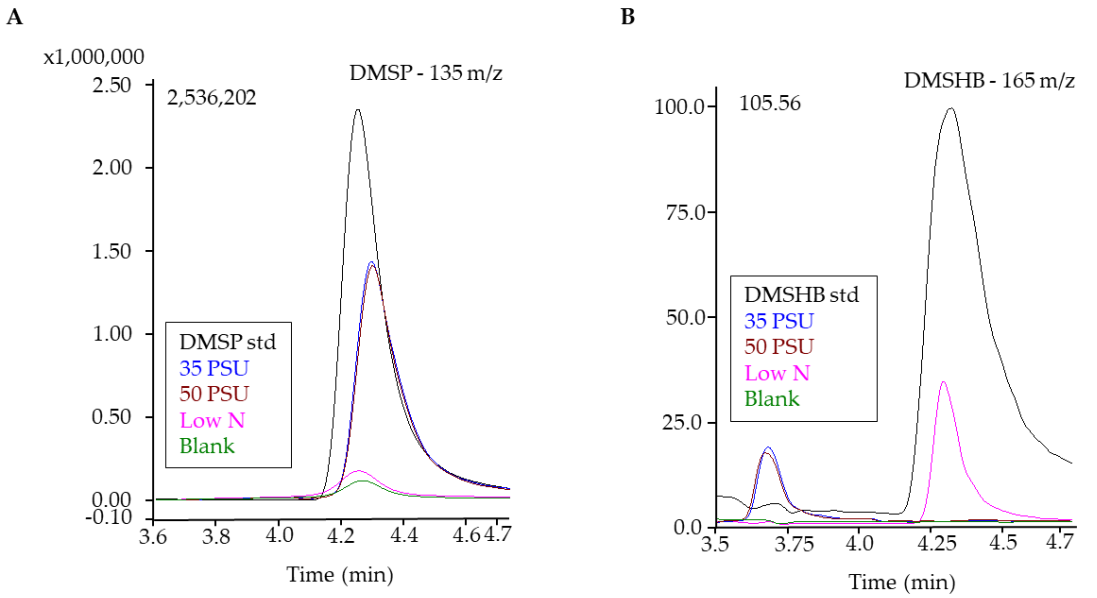
A primary antibody to *P. tricornutum* DSYD was designed (polyclonal rabbit IgG, GenScript) and tested for specificity by western blot. To ensure specificity, cell extracts from *E. coli* containing pBIO2293 overexpressing PtDSYD were used as a positive control and cell extracts from *P. tricornutum* grown in standard conditions were used to verify that the antibody to PtDSYD was specific. Immunogold localization was performed as in Perez Cruz *et al.*[154]. Cells from *P. tricornutum* were cryoimmobilized and embedded in Lowicryl HM20 resin (EMS, Hatfield, USA). Gold grids containing Lowicryl HM20 ultrathin sections were immunolabeled with the specific primary antibody to *P. tricornutum* DSYD a secondary antibody (IgM anti-rabbit) coupled to colloidal gold particles. The gold-conjugated secondary antibody without the primary antibody was used as control. Sections were then visualised using a transmission electron *microscope (Chapter 2)*.

Despite the negative results of the *in silico* analysis, immunogold labelling of DSYD showed that DSYD accumulates in the chloroplasts, mitochondria (Figure 6.8) and in the unidentified vesicles (data not shown). The unidentified vesicles were shown to also have candidate BCCT transporter in the membranes (Chapter 5). The localization of DSYD in *P. tricornutum*, supports the proposition of DMSP being mostly produced in those organelles[99]. It is likely that DMSP can be transported from these organelles to the cytosol playing a multifunctional role in the cell depending on its needs, for example, protecting against oxidative stress or as an osmoprotectant in the cytosol or in the organelles.

6.3.7 *G. sunshinyii*, the bacteria containing DsyD, produces DMSHB and DMSP

G. sunshinyii strain obtained from Dr Ryun Chung Young from National University, Jinju Koreaguinella[142]. Previously we had shown that the *G. sunshinyii* DSYD enzyme had MTHB methyltransferase activity. Here we studied whether the

Figure 6.9. Detection of DMSP and DMSHB in *G. sunshinyii* samples by LC/MS. A) DMSP chromatogram (m/z 135) and B) DMSHB chromatogram (m/z 165) of *G. sunshinyii* grown in minimal media and: brown, high salt (50 PSU); blue, normal salinity (35 PSU); pink, low N (0.5 mM NH_4Cl). In blak, 50 µM of standard DMSP or DMSHB.



gammaproteobacterium could produce DMSP. When grown in Marine Basal Minimal media and standard conditions, *G. sunshinyii* likely produced DMSP as determined by GC analysis. To further confirm this, *G. sunshinyii* was grown in Marine Basal Minimal media at 35 PSU, 50 and with limited nitrogen (0.5 mM NH_4Cl). We could not get this strain to grow at lower salinities. Samples of those cultures were taken and analysed by LC/MS. *G. sunshinyii* produced the same amount of DMSP at normal and high salinity, however, contrary to what it was previously observed in the model diatoms, it appears that in low N DMSP accumulation is almost none existent with similar levels than the background noise as observed in the blank. DMSHB does not appear to accumulate at normal or high salinity, which is probably the result of it being converted into DMSP. Interestingly, DMSHB did accumulate in the cells in low nitrogen conditions and it is not converted into DMSP. This is surprising and this accumulation of DMHB over DMSP and in N limitation is the first evidence of a switch between these two molecules. Quite why DMSHB only

accumulated in low N conditions is not understood. Perhaps the enzyme converting DMSHB to DMSP in *G. sunshinyii* is strongly dependent on N, possibly containing a N-dependent co-factor.

6.4 Discussion and concluding remarks

Unfortunately, by the time of the writing of this thesis, Kageyama *et al.*[101] identified and published the methyltransferase THAPSDRAFT_269095 as the responsible for DMSHB production in *T. pseudonana*. In this publication DSYD was named as TpMMT. That nomenclature was rejected in this study and DSYD was maintained as MMT is the abbreviation used for Methionine methyltransferase, a SAM-dependent methyltransferase that adds a methyl group to methionine forming to S-methyl-L-methionine (SMM). SMM is common in plants and is also known to be the responsible of the first step in the DMSP synthesis via the methylation pathway in plants[91] and also in bacteria[90]. In addition, Kageyama *et al.*[101] identified homologues in *T. oceanica* and *P. tricornutum*. *Thalassiosira oceanica*'s EJK49332 is 76% identical to *T. pseudonana* and *Phaeodactylum tricornutum*'s XM_ 002183230, 57%. However, they do not show that these homologues have MTHB methyltransferase activity. In this thesis, I demonstrate that those two methyltransferases are indeed functional. Hereafter, DMSHB synthase in *dsy*B/*DSY*B lacking diatoms and its homologues will be referred to as DsyD (**D**MSHB **Sy**nthase **D**).

6.4.1 DSYD is responsible for MTHB methylation in diatoms.

Two candidate MTHB methyltransferases were identified in *T. pseudonana* and *P. tricornutum*, THAPSDRAFT_269095 and PHATRDRAFT_48704, respectively (Chapter 4). Both candidates were PCR amplified and cloned into pET21a, pBIO2291(pET21a-269095), for *T. pseudonana* and pET21a and pLMB509, pBIO2292 (pLMB509-48704) and pBIO2293 (pET21a-48704), for *P. tricornutum*. Clones were then conjugated or transformed into the heterologous hosts *E. coli* and *Rhizobium leguminosarum* and tested for MTHB methyltransferase activity. Both candidates are

functional DMSHB synthase activity (Table 6-1) and, hereafter, are referred to as TpDSYD and PtDSYD.

Homologues of TpDSYD and PtDSYD are also found in other sequenced diatoms such as *Thalassiosira oceanica*, *P. multistriata*, *F. solaris* and *F. cylindrus*. Two of those homologues, identified in *T. oceanica* were also tested for DSYD activity. *Thalassiosira oceanica*'s EJK49332 conferred the ability to methylate MTHB to the heterologous host *E. coli* (Table 6-1), hereafter ToDSYD1. On the other hand, the alternative candidate identified in this diatom (ToDSYD2), did not show any DSYD activity.

Functional TpDSYD, ToDSYD1 and PtDSYD present a single methyltransferase domain whereas non-functional DSYD from *T. oceanica* (ToDSYD2), and the non-tested candidate DSYD from *F. solaris* have two different domains, a methyltransferase domain in the N terminus and a NAD binding domain (pfam02826) domain belonging to the superfamily SerA in the C-terminus (Figure 6.1). In *T. oceanica,* the C-terminus domain is a component of a larger lactate dehydrogenase family (LdhA). In both cases, it is predicted to have a dehydrogenase or reductase activity. A hypothesis is that these extra domain might be involved in other parts of the DMSP synthesis pathway, such as being involved in the reduction of MTOB[97]. This has not been further investigated.

It would be interesting to test ToDSYD2 enzyme in a different host such as *Rhizobium leguminosarum* in case this methyltransferase requires an additional co-factor not present in *E. coli*[99]. It would also be interesting to explore the functionality of the extra NAD binding domain and their potential implication in DMSP synthesis.

6.4.1.1 *TpDSYD* and *PtDSYD* are regulated by different environmental conditions

Whole transcriptome sequencing of *T. pseudonana* exposed to a salinity shift from low salt (1 PSU) and high salt (35 PSU) showed that *TpDSYD* transcripts was 49.23-fold higher at increased salinity, and one of the most upregulated genes in this condition.

This upregulation was ratified by RT-qPCR which revealed downregulation by 19-fold of the expression of *TpDSYD* in cultures grown at 1 PSU compared to standard conditions (Figure 6.6). Similarly, whole transcriptome sequencing of *T. pseudonana* samples taken under silica or nitrogen limitation showed that *TpDSYD* transcription increased by 88.22-fold in the nitrogen depleted cultures compared to the silicate depleted cultures. Furthermore, RT-qPCR was used to compare the expression of *TpDSYD* from cultures at the onset of N starvation to cultures grown in standard conditions and exponential growth phase. Once again, *TpDSYD* transcription in reduced nitrogen samples was almost 9-fold higher than the standard treatment (Figure 6.6). The correlation between the transcription of *TpDSYD* and the accumulation of DMSP in the cell denote that *TpDSYD* is playing a role in the synthesis of this osmolyte in the centric diatom.

RNA was extracted from samples taken at mid-exponential phase of *P. tricornutum* cultures grown at standard, low salt and addition of 0.5 mM of DMPS, GBT, glycine, choline and MTHB. RNA samples were also taken from nitrogen starved cultures (Chapter 3). *PtDSYD* transcription regulation was then observed by RT-qPCR. Downregulation of *PtDSYD* by 6-fold and 3-fold occurred in cultures grown at low salinity (5 PSU) and addition of GBT compared to standard cultures (Figure 6.7). This downregulation is understood as DMSP concentration in those two conditions is also significantly lower than in standard cultures (Chapter 3). On the other hand, there are two discrepancies between DMSP concentration and the regulation of *PtDSYD*. Firstly, addition of choline does not affect the intracellular concentration of DMSP whereas it significantly decreases the transcription of this gene. Plus, DMSP concentration is significantly higher in cultures depleted of nitrogen whilst *PtDSYD* transcription is not affected by this condition (Figure 6.7). These discrepancies could be explained by the effect of post-transcriptional and post-translational regulation or, perhaps, by effects in the transport of the metabolite transport which could contribute to the accumulation of intracellular DMSP. However, caution should be taken when interpreting these results as the error bars (representing the standard deviation of the

cultures) are very large. It would be advised to repeat this experiment to be able to reach firm conclusions.

It is likely *TpDSYD* and *PtDSYD* are the genes carrying the methyltransferase step in the biosynthesis of DMSP in the model diatoms. Nonetheless, further analysis to corroborate this would be required, such as creating knock out mutants and observing if DMSP production is disrupted. Moreover, the phenotype of these mutants could help to elucidate the role of this compatible solute in the cell.

6.4.1.2 PtDSYD has a lower affinity to MTHB than *Prymnesium parvum* DSYB.

Cultures of *E. coli* BL21 containing pBIO2293 (pET21a-PtDSYD) were grown in rich LB medium and the expression of PtDSYD was induced by addition of IPTG. PtDSYD was then purified and an enzymatic assay to monitor the production of S-adenosyl homocysteine (SAH) resulted from the demethylation of SAM, was performed. Optimal DSYD activity was determined by testing temperature and pH conditions. The activity from each condition are given as a percentage of the highest activity achieved. PtDSYD was most active at 30 °C and pH6.8.

To calculate the K_M value for MTHB, enzymatic assays were set at the optimum temperature and pH. Then, SAH production per mg of protein and per minute was calculated for different concentration of MTHB, DSYD's substrate. The obtained K_M value was 186±9 μM (Figure 6.4) which is 2.1-fold higher than the K_M value reported of DSYB for MTHB (88.2 μM)[99]. Therefore, PtDSYD exhibits a lower affinity for MTHB than the previously characterised DSYB from *P. parvum*[99]. This is not entirely surprising as *P. parvum* is a species of haptophyte, which is one the greatest DMSP producers, while *P. tricornutum*, as a diatom, is not considered as a significant DMSP producer in the environment[174].

6.4.1.3 PtDSYD is localised in the organelles

In silico analysis of the functional diatom DSYD to identify signal peptides[158–160] did not recognise any sequence or motif targeting this enzyme to the organelles. However, immunogold localisation using specific antibodies against PtDSYD showed that, in *P. tricornutum*, DSYD is localised in the mitochondria, chloroplast (Figure 6.8) and also in the unidentified vesicles described in Chapter 5. DMSP production has been reported to occur in chloroplast in higher plants[228] albeit using a different synthesis pathway, and the characterised DSYB from the haptophyte *P. parvum* was also found to be localised in those organelles[99]. A plausible hypothesis is that DMSP is being produced in the organelles and the concentration between the compartments and the cytosol is regulated by the BCCT transporter, likely giving DMSP and the other osmolytes different functions in the cell depending on the environmental conditions.

6.4.2 DSYD is also present in alga and bacteria.

6.4.2.1 DMSP producing *G. sunshinyii* has a functional dsyD

A homologue to the functional DSYD was found in the alphaproteobacteria *G. sunshinyii*. This homologue contains two conserved domains, the N-terminus domain corresponds to a methyltransferase whereas the C-terminus domain is predicted to be a ureidoglycolate lyase domain (Figure 6.1). This domain has very low homology to other ureidoglycolate lyases[225] and it is possible that the prediction is incorrect as errors in database have been reported in a previous study[229]. We hypothesised that this domain has a similar role in the synthesis of DMSP to the extra domain found in ToDSYD2 and DSYD from *F. solaris. In vivo* assays of this candidate DSYD showed that it was indeed a functional MTHB methyltransferase (Table 6-1).

G. sunshinyii strain was obtained from Dr Ryun Chung Young from National University, Jinju Koreaguinella[142] to test whether it was a DMSP producing bacterium. When *G. sunshinyii* was grown in minimal media and normal (35 PSU) or

increased salinity (50 PSU) it produced similar concentrations of intracellular DMSP (Figure 6.9). It was not possible to study DMSP production at lower salinities as this strain is not able to grow at lower osmotic pressure. Surprisingly, when this strain was grown in minimal media and low N (0.5 mM NH_4Cl), it stopped producing DMSP and accumulated DMSHB instead (Figure 6.9). This is the first evidence of a switch between these two molecules. Quite why DMSHB only accumulated in low N conditions is not understood. Perhaps the enzyme converting DMSHB to DMSP in *G. sunshinyii* is strongly dependent on N, possibly containing a N-dependent co-factor. For instance, the pyridoxal phosphate (PLP)-dependent aminotransferase superfamily, an enzyme hypothesised to carry out the last step of the transamination pathway (Chapter 4) requires N to catalyse the Schiff base reaction.

The discovery of the *dsy*B[89] and now *dsy*D, shows that it is very likely that the bacterial production of DMSP is likely underestimated.

6.4.2.2 DSYD is the first DMSP synthesis enzyme proven to be functional in alga

Alga are important DMSP produces, and they share the same transamination biosynthesis pathway[230] with phytoplankton[99] and some bacteria[89]. Here, we report that the unicellular green algae *Ostreococcus tauri* and the brown algae *Sargassum vulgare* contain functional DSYD homologues (Table 6-1). However, DSYD homologue from *Pyropia haitanensis* did not show MTHB methyltransferase activity in the heterologous hosts tested in this study (Table 6-1). Nevertheless, this is the first report of a functional enzyme involved in DMSP synthesis described in alga. DSYD-like proteins are also found in another alga such as *Chorda filum*, *Ulva lactuca* and the heterokont *Aureococcus anophagefferens* (Figure 6.2). It would be a priority to test whether these candidate DSYD are also functional.

6.4.3 Diatom DSYD is divergent from DSYD found in other organisms.

Amino acid sequences from functional DSYD, non-functional DSYD-like enzymes and non-tested candidate DSYD methyltransferases were aligned. The alignment was used to produce a Maximum Likelihood phylogenetic tree (Figure 6.2). In this tree, it is apparent that the three clades containing diatom sequences (Clade III, IV, V) cluster together and are divergent from the alga DSYD, *G. sunshinyii* DsyD and other prokaryotic methyltransferases. Amongst the diatom sequences, PtDSYD, TpDSYD and ToDSYD are closely related forming their own clade (Clade IV). It is worth noting that DsyD from *G. sunshinyii* is associated to the alga DSYD and DSYD-like proteins (Clade I) rather than to the diatom sequences (Figure 6.2). DSYD-like proteins from prokaryotes including *Moorea bouillonii*, *N. exalbescens* and *A. halophila*, which contain non-functional DSYD, are represented in Clade II. These prokaryotic sequences are, in turn, related to the methyltransferases from *E. litoralis* and *M. alba* involved in both DMSP and GBT synthesis, further discussed in Chapter 7.

An analysis of the methyltransferase motifs of the functional DSYD shows that the primary structure of diatom DSYD and DSYD from other organisms are also divergent. It stands out a variation of the amino acid preceding a conserved histidine (H) found after the motif II. In other methyltransferases, including those involved in GBT and GBT/DMSP synthesis (Chapter 7), H is preceded by a leucine (L), an amino acid with a hydrophobic chain. On the other hand, H is preceded by a different hydrophobic chain containing amino acid, valine (V), in diatoms and by a sulfur-containing amino acid cysteine (C) in the in DSYD from in *G. sunshinyii*, *O. tauri* and *S. vulgare* (Figure 6.3). Attention is brought to this amino acid variation as Motif II has been suggested to be involved in substrate specificity[226], but further biochemical characterization is essential to establish whether this amino acid change is responsible for the affinity of these methyltransferases for MTHB or for GBT precursors.

Chapter 7

Identification and characterization of GSDMT in diatoms

7 Identification and characterization of GSDMT in diatoms

7.1 Introduction

Glycine betaine can be synthesised *de novo* via two different pathways, the choline pathway[16] and the glycine pathway[231] (Figure 1.3, Figure 1.4). Whereas the enzymes involved in the synthesis of GBT via the choline pathway have been widely studied[46], the glycine or methylation pathway remains with a few properly ratified and characterised enzymes[232–236].

The glycine pathway can be carried out by a multiple-domain methyltransferase (Gsdmt), capable of adding a methyl group to glycine, sarcosine and dimethylglycine using SAM as the methyl donor or by two separate enzymes[234]. Substrate specificity of the two separate enzymes carrying out GBT synthesis via the methylation pathway vary depending on the organism. Some enzymes are able to add a methyl group to glycine and to sarcosine (Gsmt)[232] and other enzymes can only methylate dimethylglycine (Dmt)[54] or both sarcosine and dimethylglycine (Sdmt)[232]. Even if the catalytic capacity of converting sarcosine into dimethylglycine overlaps in the enzymes present in one organism, one of the enzymes will have a stronger specificity for sarcosine than the other[236].

7.1.1 Gsdmt, an elusive multiple domain methyltransferase

Glycine sarcosine dimethylglycine N-methyltransferase (Gsdmt) is a fusion protein with two methyltransferase domains able to catalyse the complete conversion of glycine to glycine betaine. This type of protein has been reported to be present in two organisms, the methanoarchaeon *Methanohalophilus portucalensis*[237] and the aerobic extreme halophile eubacteria *Actinopolyspora halophila*[234]. However, no gene encoding for Gsdmt or amino acid sequence in *M. portucalensis* have been identified and reports of this multi-domain protein are based only in kinetics using purified enzyme

separated from crude protein extracts[237]. Then again, although the Gsdmt gene and protein sequence from *A. halophila* are available and it has been suggested to be the main actor in GBT synthesis in this organism[211], no functional analysis have been performed in this enzyme to date.

7.1.2 Gsmt and Sdmt/Dmt, two enzymes to perform the glycine pathway.

Many organisms have a two-enzyme system to carry out the three methylation steps of glycine to GBT. Glycine sarcosine *N*-methyltransferase (Gsmt) and sarcosine dimethylglycine *N*-methyltransferase (Sdmt) from the extreme halophile *Ectothiorhodospira halochloris* were successfully expressed in the heterologous host *E. coli*[51] and later they were characterised[231]. It was found that Gsmt could not methylate sarcosine and that this compound had an inhibitory effect on Sdmt causing inhibition of the methylation of newly formed dimethylglycine to GBT[232]. A similar two-enzyme system constituted by a Gsmt/Sdmt was also found in *M. portucalensis*[231], this system is additional to the Gsdmt enzyme previously mentioned. This system was also cloned and expressed in *E. coli* conferring osmoprotection to the heterologous host thanks to the acquired ability to synthesise GBT from glycine. Furthermore, it was shown that SAH and GBT had an inhibitory effect on Gsmt[231]. Crystallisation of Gsmt from *M. portucalensis*, the only methyltransferase involved in GBT synthesis structurally analysed, allowed the identification of key amino acids and tertiary conformation for SAH binding as well as a substrate binding pocket and a betaine binding pocket[209]. The cyanobacteria *Aphanothece halophytica* and *Synechococcus* sp. WH8102, and the Gram-negative bacterium *Myxococcus xanthus* have a different two-enzyme system, GSMT/DMT, in which the second enzyme has affinity only for dimethylglycine[236,238,239]. *A. halophytica* Gsmt was inhibited by acetate and dimethylglycine and Dmt was inhibited by n-butyric acid.

7.2 Aims of the chapter

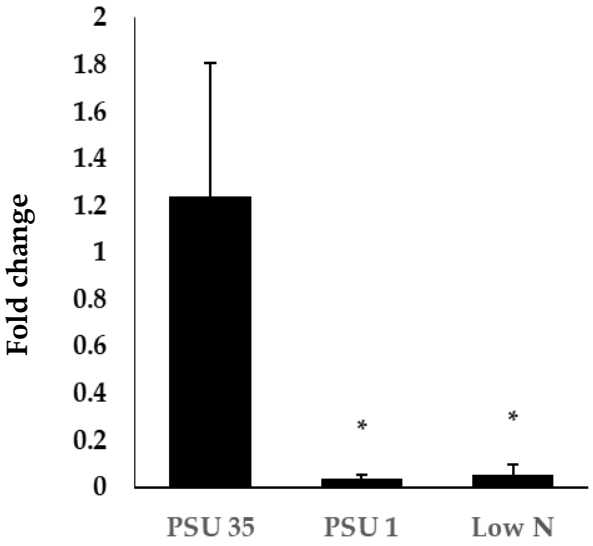
Homologues of the described Gsdmt have been found in diatoms and it has been suggested that glycine pathway might be utilised to produce GBT in these organisms[98,181]. This chapter aims to study the regulation of *GSDMT* from *P. tricornutum* and *T. pseudonana* and relate the regulation to the concentration of the nitrogenous osmolyte in the cell to better establish the importance of this enzyme for GBT synthesis in these organisms. It also seeks to characterise the candidate GSDMT from *T. pseudonana*, the first enzyme for GBT synthesis in diatoms and the first multi-domain protein ever described. GSDMT from *T. pseudonana* with confirmed activity were used as a probe to find homologues in other organisms and tests were performed to confirm whether those homologues are functional or not and whether they are relevant in nature. Primary and secondary structure of functional methyltransferases will be studied and their relatedness to each other and to functional DsyD/DSYD from chapter 6 will be observed. Given the relatedness of GSDMT and DSYD, it was hoped work in this chapter might enable us to better distinguish between these two important enzymes and processes.

7.3 Transcriptional regulation of *GSDMT* genes in the model diatoms.

7.3.1 Transcriptional regulation of *GSDMT* by salinity and N availability in *T. pseudonana.*

As previously discussed in chapter 4, Tp*GSDMT* was shown to be upregulated by 166-fold when *T. pseudonana* was grown at high salinity, conversely, it was shown to be downregulated by 0.35-fold when *T. pseudonana* was grown under nutrient deplete conditions. To ratify these results, RNA from the T7 of the salinity shift experiment and nutrient limitation experiment (see chapter 3) was used to perform RT-qPCR. Expression of *TpGSDMT* was studied using qPCR primers targeting *TpGSDMT* and β-actin[150] gene as housekeeping gene. The expression of the gene at 1 PSU and low N

Figure 7.1. Fold change of *TpGSDMT* transcription under low salinity and low N. *TpGSDMT* transcription was measured by RT-qPCR. Expression is shown as fold change respective to samples grown at standard conditions (PSU35). β-actin gene was used as housekeeping gene to normalise the expression to enable the comparison of the results. Statistically significant results ($p<0.05$) are indicated with a *.



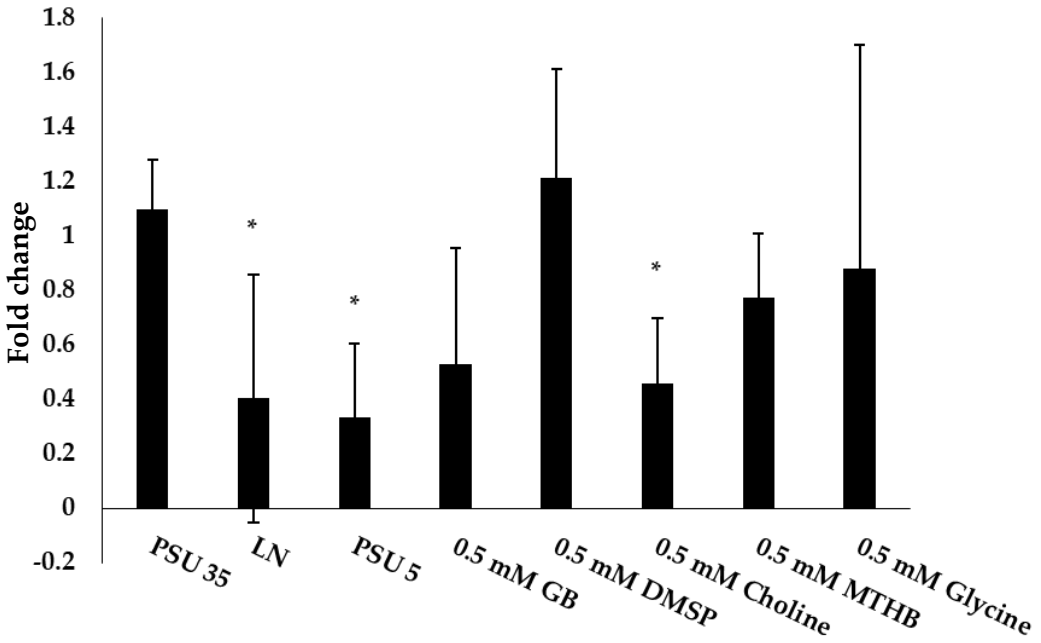
was compared to the standard 35 PSU or normal salinity samples as indicated in chapter 2. RT-qPCR results showed that *TpGSDMT* was significantly downregulated by both low salinity and low N (Figure 7.1), supporting the results found by RNA-seq. Moreover, the regulation of this GBT synthesis gene correlates with the accumulation of GBT in the cell under those tested conditions, namely low salinity and low N, in which GBT was below the detection limit of the NMR (See chapter 3).

7.3.2 Transcriptional regulation of *GSDMT* by salinity, N availability, DMSP, GBT and their precursors in *P. tricornutum.*

RNA was extracted from *P. tricornutum* samples grown in standard conditions (35 PSU), low salt (5 PSU), stationary growth phase in low nitrogen cultures and addition of 0.5 mM GBT, DMSP and their precursors (see chapter 3). The RNA was used to study the expression and regulation of *GSDMT* from *P. tricornutum* in the aforementioned conditions. Primers designed to target *GSDMT* was designed and

the gene encoding exportin 1-like[151] protein was used as a housekeeping gene to normalise the gene expression (Figure 7.2).

Figure 7.2. Fold change of *PtGSDMT* transcription under low salinity, low N and addition of GBT, DMSP and their precursors. *PtGSDMT* transcription was measured by RT-qPCR. Expression is shown as fold change respective to samples grown at standard conditions (PSU35). Exportin 1-like protein gene was used as housekeeping gene to normalise the expression to enable the comparison of the results. Statistically significant results (p<0.05) are indicated with a *.



Transcription of *GSDMT* was significantly reduced in low salt and low nitrogen compared to standard condition. This is consistent with what we observed for Tp*GSDMT*. In chapter 3, it was shown that the concentration of GBT was lower at low salt, corresponding to the transcription, however, GBT concentration did not change by lowered N (Figure 7.2). Downregulation of *GSDMT* could lead to a no significant change in accumulation of GBT due to posttranslational modifications or if the *GSDMT* gene is not the main enzyme responsible for GBT in Pt. The possibility of an alternative pathway for GBT synthesis from choline is also supported by the downregulation of *GSDMT* when choline is added to the media, as it was also shown that addition of choline did increase the intracellular concentration of GBT. Furthermore, addition of DMSP decreases significantly the accumulation of GBT but

does not affect *GSDMT* transcription (Figure 7.2), suggesting that DMSP could potentially be regulating an alternative GBT synthesis pathway. Certainly, *P. tricornutum* has a candidate choline dehydrogenase (PHATRDRAFT_1341), although this protein has not been tested.

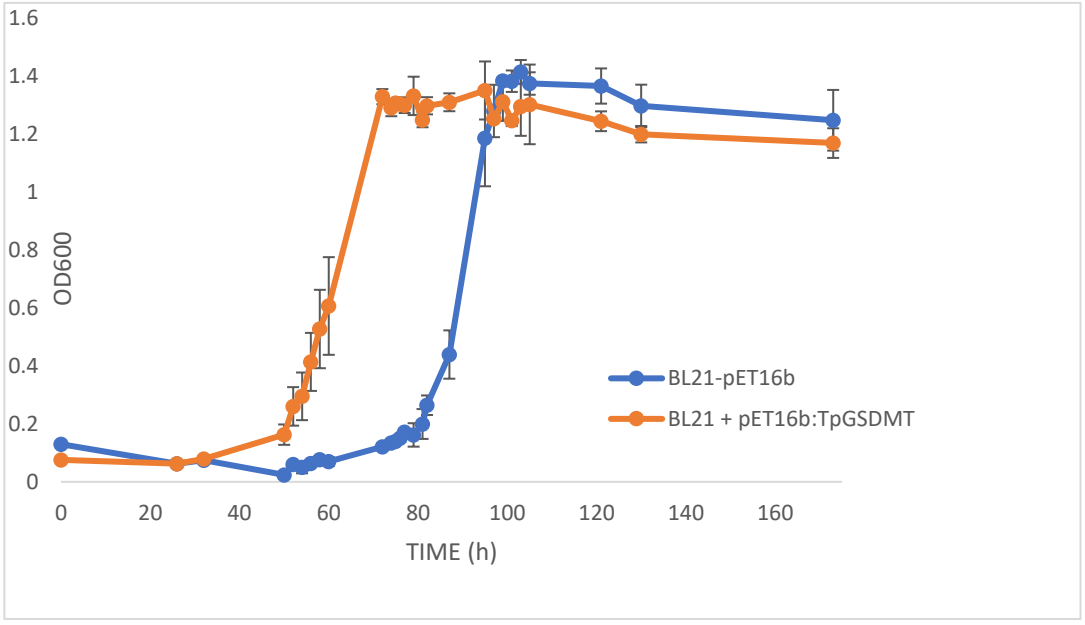
7.4 Characterization of TpGSDMT

T. pseudonana is likely to synthesise GBT via the glycine pathway using a double domain methyltransferase. THAPSDRAFT_20797 (XP_002286764.1) has been identified as the candidate S-adenosyl-L-methionine methyltransferase in chapter 4.

Primers were designed to amplify the candidate *GSDMT* from complementary DNA (cDNA) flanked by the restriction sites *Nde*I and *Bam*HI (Table 2-3). The insert was then cloned into pET16a expression vector for *E. coli* and transformed into *E. coli* BL21.

E. coli BL21 containing pET16a:*TpGSDMT* or *E. coli* BL21 containing the empty plasmid (pET16b) as control, were grown in triplicates in high salt (0.6 M NaCl)

Figure 7.3. Salt tolerance of E. coli and E. coli containing TpGSDMT. Samples were grown in high salt minimal media (0.6 M NaCl) and supplemented with the precursors of GBT via the glycine pathway (0.5 mM glycine). Growth was monitored at regular intervals over a period of 173h.



minimal media with the addition of the precursor of GBT, glycine, and the addition of methionine, the precursor of the methyl donor AdoMet. *E. coli* can only synthesise GBT *de novo* via the choline pathway and only if choline is exogenously added to the media. If functional, Tp*GSDMT* would confer *E. coli* the ability to produce GBT from Gly, hence, conferring osmoprotection to the cells.

As shown in Figure 7.3, *E. coli* BL21 containing *TpGSDMT* had a shorter lag phase than the control. This suggests that *GSDMT* is encoding for a functional enzyme able to transform glycine to glycine betaine and, consequently, conferring osmoprotection to BL21. Control sample undergoes a longer lag phase as it does not have glycine betaine available. *E. coli* suffers changes in its metabolism in order to produce another set of osmoprotectants that allow it to survive high concentrations of salt[240]. This potentially explains why the growth of the control culture eventually equals that to the cells producing GBT.

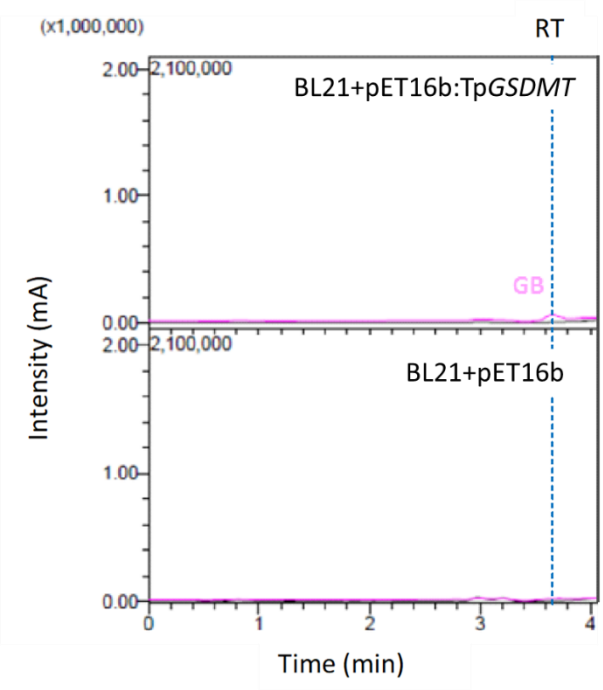
After the last measurement, samples were sacrificed and metabolites were analysed by LC/MS. *E. coli* BL21+pET16b:*TpGSDMT* produced a peak corresponding to GBT whereas *E. coli* BL21+pET16b did not produced a peak for this molecule (Figure 7.4). This indicates that it was indeed the GBT synthesised by the enzyme GSDMT what conferred osmoprotection to *E. coli* BL21 (Figure 7.3).

7.5 Diversity and functionality of methyltransferases involved in glycine betaine synthesis via the glycine pathway.

7.5.1 Distribution of homologues of TpGSDMT

A BLASTp search in the NCBI database was conducted using the functional GSDMT from *T. pseudonana* as probe. Top hit results show this protein is not unique to the central diatom *T. pseudonana*. Homologous to this enzyme are also present in other model diatoms such in *T. oceanica*, *Pseudo-nitzschia multistriata*, *Phaeodactylum tricornutum* CCAP 1055/1 and *Fragilariopsis cylindrus* CCMP1102 (Table 7-1).

Figure 7.4. LC/MS chromatogram showing glycine betaine (GB) in *E. coli* BL21 containing pET16b:TpGSDMT. Samples correspond to the cultures grown in high salt and GBT precursors. The retention time (RT) for GBT is marked by a line. E. coli with the empty plasmid was used as negative control.



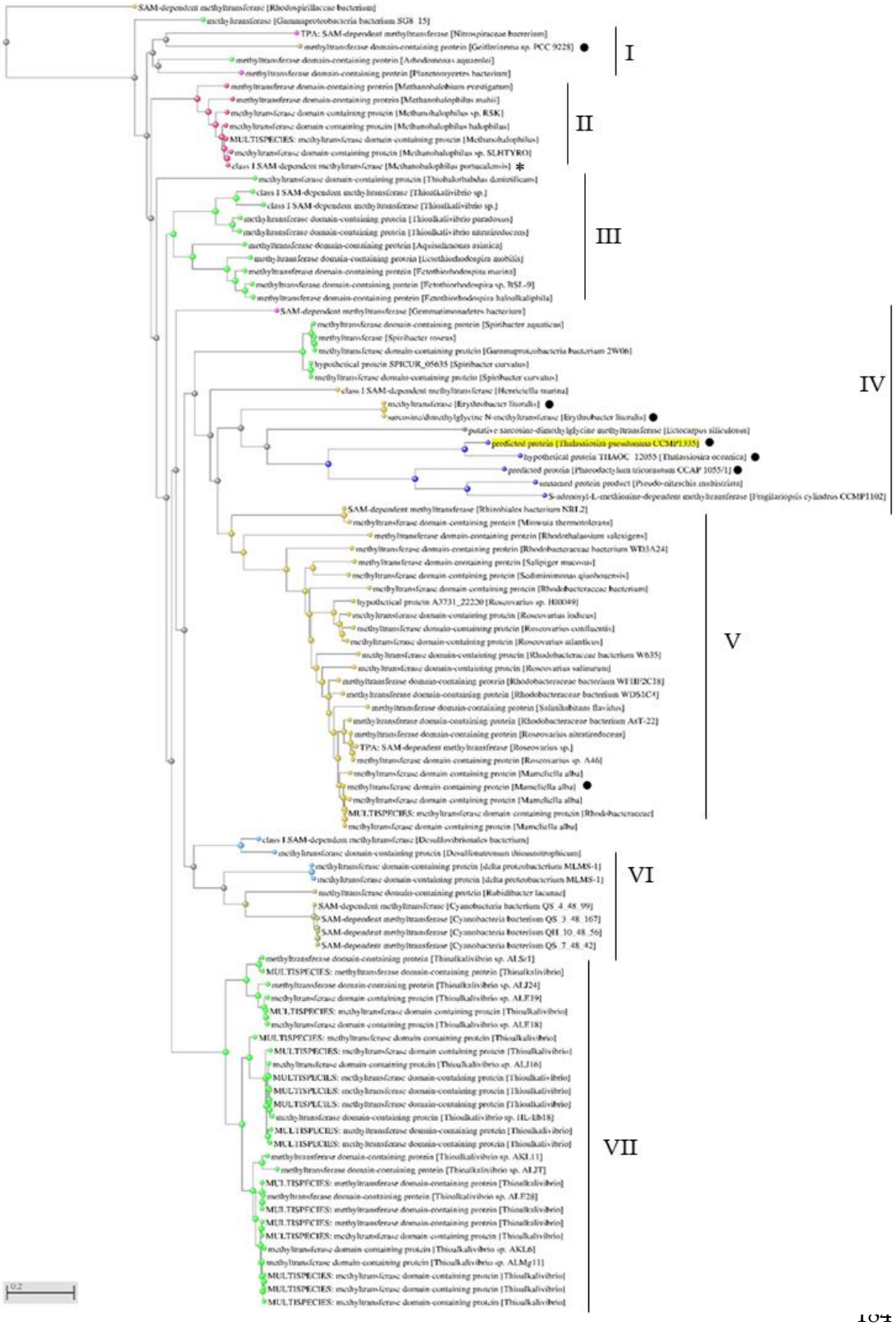
Candidate diatom GSDMT cluster together and other methyltransferases predicted to be involved in GBT synthesis appear in organisms including archaea, bacteria, cyanobacteria and brown alga (Figure 7.5).

Table 7-1. List of the closest homologues to TpGSDMT found by BLASTp search in NCBI database.

Organism	Prot ID	E value	% Id
Thalassiosira oceanica	EJK66971.1	$e^{0.0}$	77.70
Pseudo-nitzschia multistriata	VEU45206.1	$6e^{-112}$	34.18
Phaeodactylum tricornutum CCAP 1055/1	XP_002180089.1	$6e^{-62}$	39.86
Fragilariopsis cylindrus CCMP1102	OEU23352.1	$3e^{-49}$	31.96
Thioalkalivibrio sp. ALJ24	WP_018935996.1	$6e^{-34}$	34.72

The Neighbour Joining Distance Tree produced with the closest 100 sequences to TpGSDMT returned by NCBI BLASTp show the methyltransferase domain-containing proteins or SAM dependent methyltransferases cluster in seven distinctive groups and three sequences that do not pair with others. The first branch divides into a SAM-dependent methyltransferase from the unclassified alpha-proteobacteria *Rhodospirillaceae* and the branch which contain most of the other hits. The secondary branch divides into the unclassified gammaproteobacteria bacterium SG8_15 and a tertiary branch. The tertiary branch diverges between Cluster I, compose of gamma-proteobacteria, bacteria and a cyanobacteria (*Geitlerinema* sp. PCC 9228) and a subtree. The subtree divides into the Cluster II and a further subtree. Cluster II is formed by methyltransferase domain-containing protein from euryarchaeotes including the Class I SAM-dependent methyltransferase from *Methanohalophilus portucalensis.* The methyltransferase from *M. portucalensis* was the first identified and characterised enzyme involved in glycine betaine synthesis via the glycine pathway[35]. The next subtree contains a methyltransferase domain-containing protein from the gamma-proteobacteria *Thiohalorhabdus denitrificans* and a subtree containing the Clusters III to VII. Methyltransferase domain-containing proteins from gamma-proteobacteria species such as *Thioalkalivibrio* sp. or *Ectothiorhodospira* sp. group in Cluster III. This cluster diverges from a further subtree branch which, in turn, divides into Cluster VII and another subgroup containing the remaining clusters. Cluster VII is formed by divergent methyltransferase domain-containing proteins from the gamma-proteobacteria *Thioalkalivibrio* sp. The following subtree branch shows an unpaired SAM-dependent methyltransferase from the unclassified bacteria *Gemmatimonodetes bacterium* and another subtree branch. This last subtree divides into two, the first part of the branch contains Clusters III, IV and V. These three clusters are differentiated from Cluster VI. Cluster III is formed by enzymes from gamma-proteobacteria from species such as *Spiribacter* sp., and Cluster V contains SAM-dependent methyltransferase from alpha-proteobacteria including *Roseovarius* or *Rhodobacteraceae*.

Figure 7.5. Neighbour Joining Distance Tree of Tp GSDMT homologues found by BLASTp in NCBI. Tested methyltransferases are marked with a •, published gsdmt is marked with a *, highlighted in yellow TpGSDMT.



Interestingly, Cluster IV contains candidate methyltransferases from a mixture of organisms, including the alpha-proteobacteria *Henriciella marina* and *Erythrobacter litoralis*, the brown algae *Ectocarpus siliculosus* and the diatoms *T. pseudonana*, *T. oceanica*, *P. tricornutum*, *P. multistriata* and *F, cylindrus* (Figure 7.5). Noteworthy, *T. pseudonana*, *T. oceanica* and *P. multistriata.* SAM-dependent methyltransferases contain two methyltransferase domains (methyltransferase_25, Pfam 13649, AdoMet_Methyltransferase superfamily). The rest of the hits produced by BLASTp where half the size and aligned only to the second methyltransferase domain.

A few representatives were selected from this tree to test whether they were functional glycine betaine synthases. These are the putative dimethylglycine methyltransferase from the brown algae *Ectocarpus siliculosus* (CBN80020.1), from the diatoms *T. oceanica* (hypothetical protein, THAOC_12055) and *F. cylindrus* (S-Adenosyl-L-dependent methyltransferase, OEU2335.1), the methyltransferase domain-containing protein from the cyanobacteria *Geitlerinema* sp. PCC_9228 (methyltransferase domain-containing protein, WP_071515314.1), the methyltransferase (AOL23288) from the alphaproteobacteria *Erythrobacter litoralis*, and the alphaproteobacteria *Mameliella alba* (methyltransferase domain-containing protein, WP_043147237.1) (Figure 7.5).

The candidate methyltransferase domain-containing protein (WP096366830.1) from *Thiohalobacter thiocyanaticus* was also identified by BLASTp search using the candidate GSDMT from *T. oceanica*. Furthermore, as candidate methyltransferase from *P. tricornutum* (XP_002180089.1) appears incomplete in the NCBI database, the sequence available at the Joint Genome Institute database (JGI) was codon optimised and used for testing instead.

7.5.2 Characterization of candidate methyltransferases involved in GBT synthesis.

Selected candidates were codon optimised and synthesised (Appendix 4-1). A ribosome binding site was added at the beginning of the sequence, restriction sites

were also added flanking the sequence and a stop codon. Clones of the synthesised sequences and the empty vector pUC18 to be used as control were transformed into *E. coli* BL21.

E. coli BL21 containing the candidate methyltransferases were first grown in rich media. Washed aliquots of these cultures were inoculated in minimal media (see chapter 2 for detailed method) and tested for GBT and DMSP synthesis when the appropriate substrates were provided. To test whether these candidates were capable of GBT synthesis, 0.5 mM of glycine, sarcosine and dimethylglycine were added to the media. To test whether any of these enzymes were capable DMSP synthesis, 0.5 mM MTHB was added to the media. Control samples did not have any additional substrates added to the minimal media.

Table 7-2. GBT, DMSP production by TpGSDMT homologue methyltransferases. GC and LCMS results of *E. coli* BL21 containing methyltransferase candidates from *E. litoralis*, *T. oceanica*, *F. cylindrus*, *Geitlerinema* sp., *M. alba*, *T. thiocyanaticus* and *P. tricornutum*. Detection of DMSP was done by GC and detection of GBT by LC/MS. Positive results are indicated with a +, negative results with a -.

Organism	Prot ID	DMSP	GBT
E. siliculosus	CBN80020.1	-	-
T. oceanica	THAOC_12055	-	+
F. cylindrus	OEU2335.1	-	-
Geitlerinema sp.	WP_071515314.1	-	+
E. litoralis	AOL23288	+	+
M. alba	WP_043147237.1	+	+
T. thiocyanaticus	WP096366830.1	+	+
P. tricornutum	XP_002180089.1	-	+

Aliquots of these cultures were taken for analysis by both GC and LC/MS. Surprisingly, the GC analysis showed that three of these candidates were capable of producing DMSP/DMSHB when expressed in *E. coli* BL21 and MTHB was added to

the media. These proteins belonged to the alpha-proteobacteria *Erythrobacter litoralis*, *Mameliella alba* and to the gamma-proteobacteria *Thiohalobacter thiocyanaticus*.

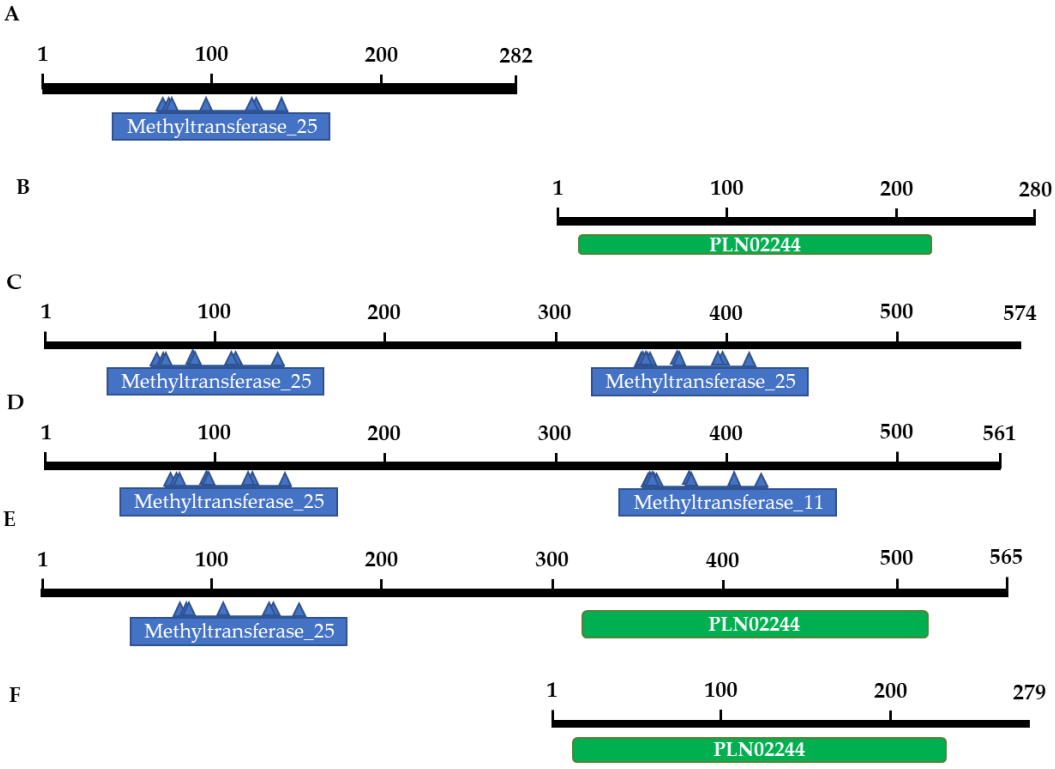
LC/MS analysis showed that the tested diatom *GSDMT* from *T. oceanica* and *P. tricornutum* conferred the ability of synthesising GBT from glycine, sarcosine and/or dimethylglycine to *E. coli*. In addition, it was also confirmed that the methyltransferases from the alphaproteobacteria *M. alba* and *E. litoralis* and the gammaproteobacteria *T. thiocyanaticus* were capable of producing GBT from the glycine pathway precursors and also DMSP from MTHB. This is surprising as *E. coli* does not have the ability of transforming DMSHB into DMSP, therefore, it is likely that these enzymes are capable of carrying out both the methylation and the decarboxylation reaction transforming MTHB into DMSP (Table *7-2*). This process has never been described before and it requires further investigation. It is also noteworthy that tested functional GBT synthesis genes produced a peak of GBT even when the precursors where not added to the media due to the presence of the precursor in *E. coli* cell. However, peaks where greater with the external addition of the precursors. Exceptionally, when MTHB was added to the media, the ability of the protein from *M. alba* to produce GBT was completely supressed.

7.5.3 *In silico* analysis of the methyltransferases involved in GBT synthesis and the novel multifunctional enzymes.

The sequences of the functional GSDMT enzymes from the diatoms *T. pseudonana*, *T. oceanica* and *P. tricornutum*, the cyanobacteria *Geitlerinema* sp. PC 9228, the published Gsdmt from *Actinopolyspora halophila* and the three multifunctional enzymes from *M. alba*, *T. thiocyanaticus* and *E. litoralis* (Appendix 4-3) were aligned using ClustalW Multiple alignment and motifs were searched.

Geitlerinema sp PC 9228, *E litoralis* and *T thiocyanaticus* align with DMT from *Synechococcus* sp. Strain WH8102. On the other hand, diatom sequences, *A. halophila* and *M. alba* have two conserved methyltransferase domains (Figure 7.6). The N-terminus domain aligns with the glycine/sarcosine N-methyltransferase or Gsmt

Figure 7.6. Conserved domains and alignment of functional GBT synthesases. (A) Gsmt and (B) Dmt from *Synechococcus* sp. Strain WH8102, (C) GSDMT from *T. pseudonana*, (D) DsyD/Gsdmt from *M. alba*, (E) Gsdmt from *A. halophila* and (E) GBT synthase from *Geitlerinema* sp. PC9228. Numbers indicate the amino acid ...



from *M. portucalensis* (F6KV61) and from *Synechococcus* sp. strain WH8102 (Q7U4Z8), also with the Gsmt (Q9KJ22) from *Halorhodospira halochloris*. The conserved domain is a Methyltransferase_25 (Figure 7.6). On the other hand, the C-terminus domain aligns to Dimethylglycine N-methyltransferase or Dmt (Q7U4Z9) from *Synechococcus* sp. (strain WH8102) and Sdmt (Q9KJ21) from *H. halochloris* (Figure 7.6). However, the predicted conserved domain varies between species, in *T. pseudonana* is a Methyltransferase_25, whereas in *M. alba* is a Methyltransferase_11. In *Synechococcus* sp. Strain WH8102, *Geitlerinema* sp. PC 9228 and *A. halophila* the predicted domain belongs to tocopherol o-methyltransferase family (PLN02244) (Figure 7.6). Despite the very low homology of these domains to the published Gsmt and Sdmt, it is likely

that the double-domain proteins tested in this study have a similar catalytic capacity as TpGSDMT and can carry out the three methylation steps from glycine to GBT.

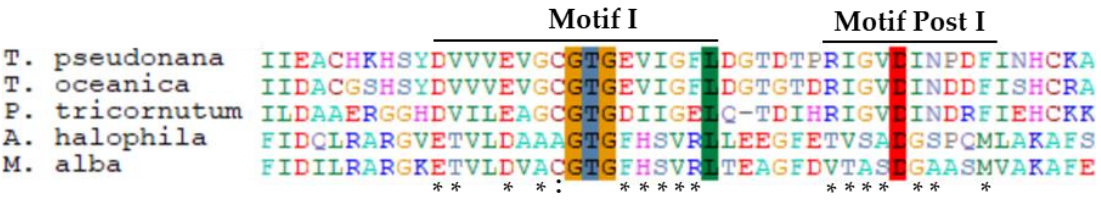
7.5.3.1 Analysis of the secondary structure of the functional methyltransferases.

The secondary structure of representative GSDMT from *T. pseudonana*, DsyD/Gsdmt from *M. alba* and DsyD/Gsdmt from *E. litoralis* were analysed using Phyre2 and motifs were manually identified.

7.5.3.2 Motif recognition of the N-terminus methyltransferase domain in the double-domain functional methyltransferases

In 2016, Lee *et al.*[209] crystallised and resolved the structure of the glycine-sarcosine methyltransferase from *M. portucalensis*. The tertiary structure allowed to identify key amino acid residues involved in GBT, SAH and substrate binding pockets in the tertiary structure. A comparison of those residues with the aligned methyltransferases of this study showed key differences between the bacterial and diatom primary structure. The D35, I38, D39, W40, N100 and H104 corresponded to the betaine binding site in *M. portucalensis*[209], all but H104 were conserved in the bacteria *A. halophila* in which there is a substitution of the positively charged amino acid with N, a polar uncharged residue and this residue appears as a gap in *M. alba*. Diatoms lack any of the key betaine binding site amino acids apart from N104 which is also present in the N-terminus of the *GSDMT* from *T. oceanica*. The amino acid residue R43 from *M. portucalensis* and L132 have been suggested to be involved in the formation of the SAH binding pocket[209]. R43 is conserved in diatoms except in *M. alba* which has a substitution of this positively charged amino acid with a G. Furthermore, L132 that is conserved in bacteria but in diatoms a V or a C is present instead, which also contain hydrophobic side chains. Finally, the amino acid residues forming the substrate binding pocket (N134, H138, R167, Y206 and M218)[209] are conserved in *A. halophila* and *M. alba* but only N134 is conserved also in all diatom sequences. The positive amino acid H138 appears as M, a hydrophobic side chain containing amino

Figure 7.7. ClustalW multiple alignment of N-terminus Motif I and Motif Post I from functional double-domain methyltransferases. Alignment where highlighted amino acids are conserved amongst all of the sequences and * indicates a difference in the conserved amino acids between bacterial and diatom sequences, : indicates a semi-conserved region where only one of the sequences has a different amino acid.



acid, in *T. oceanica* and *T. pseudonana*, and as N, a polar uncharged amino acid, in *P. tricornutum*. Positive residue R167 is changed for a hydrophobic W and hydrophobic M218 for the negatively charged D or an E in diatoms. Y206 is substituted for another hydrophobic side chain amino acid (A) in *P. tricornutum* and a polar uncharged amino acid in the other two diatoms (T).

In other characterised methyltransferases, Motif I has the signature GxGxG sequence[226], this sequence is also present in the eukaryotic GSDMT. In *A. halophila* and *M. alba*, the first G has been substituted by an A (Figure 7.7). However this change does not change the secondary structure in which the sequence GTG follow the first β strand (β1) strand and causes the formation of an alpha helix[226] (Appendix 4-2).

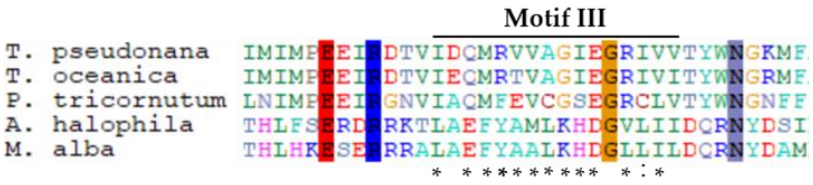
Figure 7.8. ClustalW multiple alignment of N-terminus Motif II from functional double-domain methyltransferases. Alignment where highlighted amino acids are conserved amongst all of the sequences and * indicates a difference in the conserved amino acids between bacterial and diatom sequences, : indicates a semi-conserved region where only one of the sequences has a different amino acid.

As a general rule, Motif Post I is not conserved at the primary structure but it is conserved at the secondary structure level[226]. Indeed, Motif Post I correspond to the second β strand (β2) (Appendix 4-2).

Motif II is predicted to be involved in substrate specificity[226]. The last amino acid of the predicted Motif II is a conserved W (Figure 7.8).

Predicted Motif III, has a conserved G (Figure 7.9). Despite Motif II and Motif III being reported as usually corresponding to the fourth and fifth β strands (β4, β5) in other methyltransferases[226], in *T. pseudonana* and *M. alba* it correspond to β 3 and β 5 (Appendix 4-2).

Figure 7.9 ClustalW multiple alignment of N-terminus Motif III from functional double-domain methyltransferases. Alignment where highlighted amino acids are conserved amongst all of the sequences and * indicates a difference in the conserved amino acids between bacterial and diatom sequences, : indicates a semi-conserved region where only one of the sequences has a different amino acid.



In all of the motifs there is a clear difference between the eukaryotic and the prokaryotic sequences, and GSDMT from *T. pseudonana*, *T. oceanica* and *P. tricornutum* are very similar to each other. Likewise, the N-terminus methyltransferase domain of *M. alba* and *A. halophila* are more similar to each other than to the diatom sequences.

7.5.3.3 Motif recognition of the C-terminus methyltransferase domain in the double-domain functional methyltransferases and single domain enzymes.

As previously mentioned, *Geitlerinema* sp. PC 9228, *T. thiocyanaticus* and *E. litoralis* and the C terminus of the double domain proteins aligned with Sdmt (Q7U4Z9) from *Synechococcus* sp. (strain WH8102) and Sdmt (Q9KJ21) from *H. halochloris*. These

methyltransferases have a conserved region at the beginning of the sequence. This motif has not been previously described in the literature and hereafter will be referred as Motif Pre I (Figure 7.10).

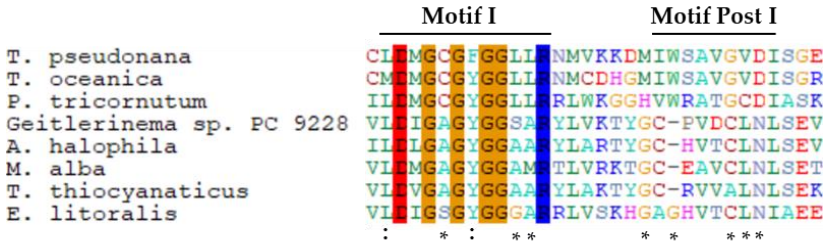
Figure 7.10. ClustalW multiple alignment of Motif Pre I from functional GSDMT, Gsdmt and multifunctional methyltransferases. Alignment where highlighted amino acids are conserved amongst all of the sequences and * indicates a difference in the conserved amino acids between bacterial and diatom sequences, : indicates a semi-conserved region where only one of the sequences has a different amino acid.

Motif Pre I consists of two tyrosine residues (YY) followed by 3 non-conserved amino acid, a semi conserved D residue with the exception of *Geitlerinema* sp. PC 9228 which D is replaced by a S. Then after another three non-conserved amino acid residues it follows the conserved region FYxxxWGxxxxHxG. However, diatom sequences seem to differ from the bacterial sequences in their conserved amino acids in the interspace between the WG and the H. Diatom has a conserved T before the H (WGxxxTH) whereas the bacterial conserved region has a conserved G and a D (WGGxDxHxG) (Figure 7.10). The YY represent the end of the first predicted alpha helix, whereas the FY are conserved amino acid residues in the middle of the second alpha helix. The sequence IHVGI correspond to the first β strand (β1) of the single domain protein from *E. litoralis*, and the first β strand of the C-terminus methyltransferase domain (β1C) in the double domain proteins with the sequence LHIGR in *M. alba* and THIGR in *T. pseudonana* (Appendix 4-2).

Motif I was identified as DxGxGxGGxxR, it has the same structure as the previously characterised SAM-methyltransferases (DxGxGxG which causes an extended turn after the second β strand (β2 and β2C)[226]. In addition, it also has two semi-conserved

amino acids, the first is before the D, which is an L apart from in *T oceanica* which is a M, both amino acids are hydrophobic amino acids. The second semi-conserved is after the second G, which is followed by an Y in all by in *T. pseudonana* which contains a F instead, these are also hydrophobic amino acids. Motif I from diatom methyltransferases diverge from the bacterial one and it has more conserved amino acids, these being xDMGCGxGGLLR. On the other hand, bacterial Motif I has the following sequence LDxGxGYGGxxR (Figure 7.11). Motif Post I is not conserved amongst diatom and bacterial methyltransferases, however, some conservation exists if compared separately between diatom and bacteria enzymes. Diatom Motif Post I

Figure 7.11. ClustalW multiple alignment of Motif I and Motif Post I from functional GSDMT, Gsdmt and multifunctional methyltransferases. Alignment where highlighted amino acids are conserved amongst all of the sequences and * indicates a difference in the conserved amino acids between bacterial and diatom sequences, : indicates a semi-conserved region where only one of the sequences has a different amino acid.

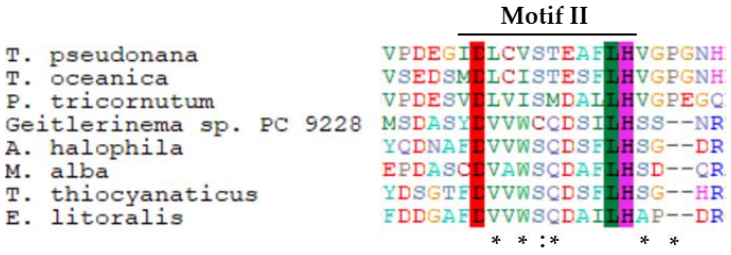


has the sequence xxWxxxGxD whereas the bacteria methyltransferases have GxxxxxLN (Figure 7.11). In both cases, this Motif Post I constitutes the third β strand (β3 or β3C)[226].

SAM-methyltransferases Motif II has been proposed to be involved in substrate specificity[226]. The functional methyltransferases analysed in this study have three conserved amino acids in the motif II, a D in position 2, and an LH 11 and 12 respectively. Furthermore, there is an E or D residue in the position 8, which are amino acids that have not been described as part of the Motif II in any other methyltransferase except than in the DSYD enzymes (see chapter 6). Once again, the number of conserved amino acids in the diatom Motif II are different to the amino

acid sequence of the bacterial methyltransferase Motif II. In the diatom GSDMT, the conserved primary structure follows the pattern x**D**Lxxxxxxx**LH**VG and the two extra amino acids, not present in the other methyltransferases, P and G/E. On the

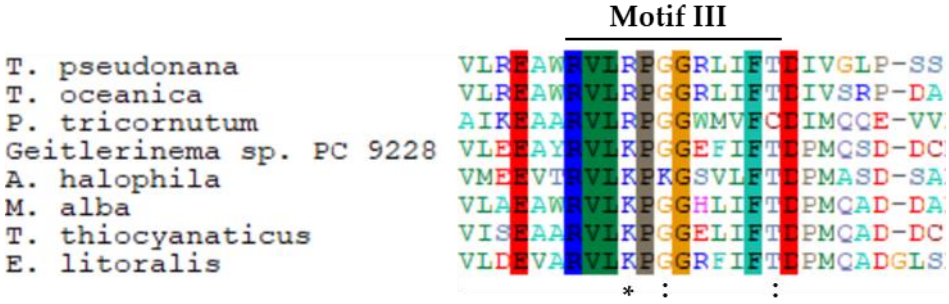
Figure 7.12. ClustalW multiple alignment of Motif II from functional GSDMT, Gsdmt and multifunctional methyltransferases. Alignment where highlighted amino acids are conserved amongst all of the sequences and * indicates a difference in the conserved amino acids between bacterial and diatom sequences, : indicates a semi-conserved region where only one of the sequences has a different amino acid.



other hand, bacterial Gsdmt or DsyD/Gsdmt follow the pattern x**D**VxWQxxx**LH** as shown in Figure 7.12, *Geitlerinema* sp. PC 9228 methyltransferase has a C in the sixth position of the motif II compared the S present in any of the other methyltransferases.

The predicted Motif II and III (Figure 7.12, Figure 7.13) of *E. litoralis* and the C-terminus methyltransferase domain of *T. pseudonana* correspond to an alpha helix region and the fifth β strand (β5 or β5C), whereas in *M. alba* they are predicted to be

Figure 7.13. ClustalW multiple alignment of Motif III from functional GSDMT, Gsdmt and multifunctional methyltransferases. Alignment where highlighted amino acids are conserved amongst all of the sequences and * indicates a difference in the conserved amino acids between bacterial and diatom sequences, : indicates a semi-conserved region where only one of the sequences has a different amino acid.



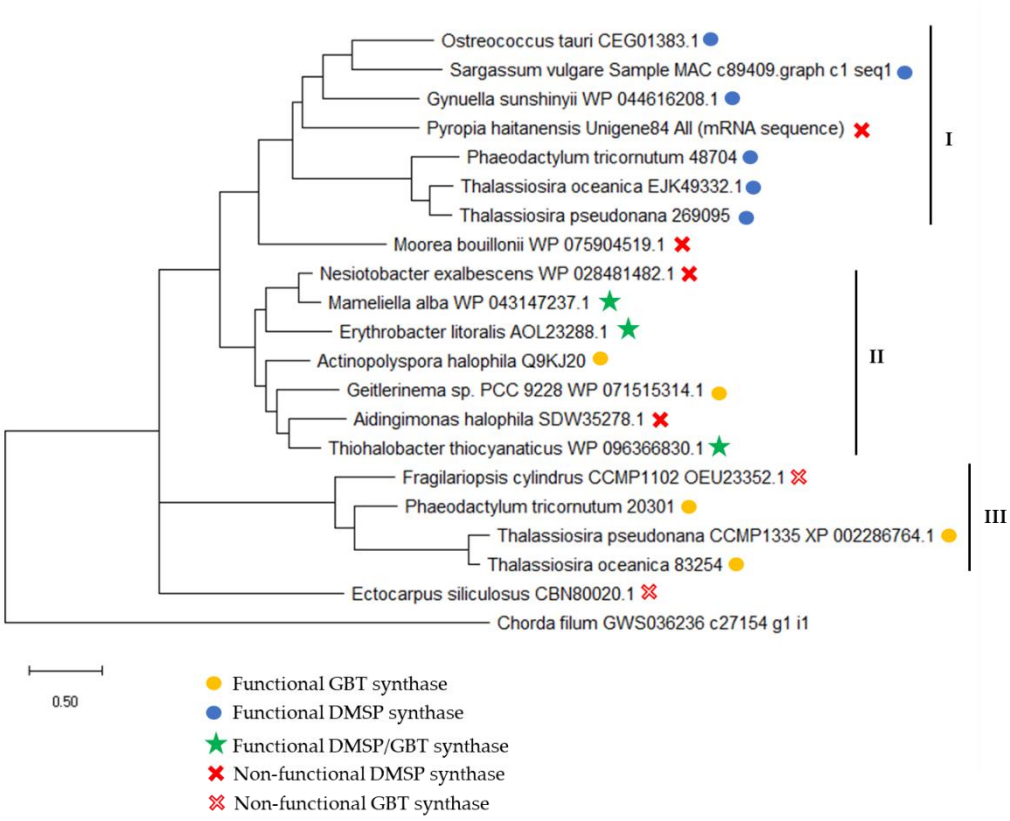
the fifth β strand(β5C) and the sixth β strand (β6C) respectively. The P and the M residues that bacterial sequences have after D, the last conserved amino acid in Motif III, have been previously suggested to be involved in substrate specificity[54], however, this possibility was lately ruled out after functional studies of the *Galderia sulphuraria* methyltransferase[50].

7.5.4 Phylogenetic analysis of the functional methyltransferases in this study.

All the methyltransferases analysed in these study, including those involved in DMSP synthesis and in GBT synthesis or both, and some non-tested methyltransferases (Appendix 4-3) were aligned by ClustalW and a Maximum Likelihood phylogenetic tree was produced to observe their relatedness using Mega v.X[241].

The phylogenetic tree shows that tested functional DMSP synthesis described in chapter 6 cluster together, named Clade I. These DsyD and DSYD proteins are present in Eukaryotes, including diatoms and brown alga, and in at least one confirmed prokaryote, the gamma-proteobacteria *G. sunshinyii*. These proteins, however, have not been tested for GBT synthesis so it is not possible to completely disregard the possibility of these enzymes also being able to synthesise GBT. Clade II englobes a group of bacterial methyltransferases with mixed functions. As seen in this chapter, three of them are capable of synthesising both DMSP and GBT whereas others are only involved in GBT synthesis. It is worth noting that functional Gsdmt from *Actinopolyspora halophila* has never been tested for DMSP synthesis, likewise, those marked as non-functional DSYD could potentially be involved in GBT synthesis instead, but further tests are required. Finally, Clade III contain GSDMT from diatoms, these methyltransferases have several conserved amino acid residues that are distinctive from the methyltransferases in Clade I and II and are uniquely involved in GBT synthesis (Figure 7.14).

Figure 7.14. Maximum Likelihood phylogenetic tree of the methyltransferases discussed in this study. The phylogenetic tree shows functional DSYD/DsyD in blue, functional multifunctional proteins in green and functional GSDMT/Gsdmt/Sdmt in yellow. Marked with a red cross are those methyltransferases tested for DMSP syntheses that were non-functional and a red and white cross represent those methyltransferases that show no function for GBT synthesis. Clade I correspond to true DMSP synthesases, Clade II to bacterial methyltransferases with mixed functions and Clade III are functional diatom GSDMT.



7.6 GBT and DMSP production by the *alpha*-proteobacteria *M. alba* encoding a novel multifunctional enzyme.

The alphaproteobacteria *M. alba* contains the novel multifunctional enzyme capable of catalysing the synthesis of GBT and DMSP. The ability of the strain to produce those metabolites was tested by growing it in marine basal minimal media with either

replete or limited nitrogen (see chapter 2 for details). Triplicates of the cultures were then sacrificed and analysed by LC/MS.

The metabolite analysis by LC/MS showed that when N is abundant in the media, *M. alba* accumulates GBT and choline, whereas when N availability is restricted it accumulates both GBT and DMSP and choline reservoirs are depleted (Table 7-3).

Table 7-3. Detection of GBT, DMSP and choline by LC/MS in M. alba cultures grown in minimal media and wither N replete or N limiting conditions.

	N replete	N limiting
GBT	+	+
DMSP	-	+
Choline	+	-

M. alba has four quasi-identical copies of a hypothetical choline dehydrogenase (WP_074623315.1, WP_088670827.1, WP_088716211.1, WP_043137804.1), an enzyme likely involved in GBT synthesis via the choline pathway.

7.7 Discussion and concluding remarks

7.7.1 GSDMT from model diatoms is regulated by environmental conditions

The analysis of the transcription of the candidate GSDMT genes from the model diatoms *T. pseudonana* (THAPSDRAFT_20797) and *P. tricornutum* (PHATRDRAFT_20301) showed that these genes are regulated both by N availability and salinity.

Whole transcriptome sequencing of *T. pseudonana* cultures showed an upregulation of the expression of *GSDMT* by 166-fold in normal salinity (35 PSU) when compared to cultures grown at low salt concentration (1 PSU). Conversely, *GSDMT* was downregulated by 0.35-fold in those cultures where N was limiting (samples taken at the onset of stationary phase with initial concentration of 10 μM of NO^{3-}) compared to samples taken at stationary growth phase from cultures grown with deplete N

(initial N concentration of 882 μM of NO^{3-}) and limiting silicate (initial Si concentration of 50 μM Si^{4+}). The same samples were used for RT-qPCR analysis to ratify the RNAseq results. qPCR results showed that indeed, GSDMT is significantly downregulated both by low salinity and N limitation when compared to standard condition (Figure 7.1). As seen in Chapter 3, the regulation of *GSDMT* follows the same pattern as the accumulation of this osmolyte in the centric diatom *T. pseudonana*. It is likely that this is the main pathway for GBT synthesis in these conditions in the model diatom. *T. pseudonana* has a candidate choline dehydrogenase which the whole transcriptome sequence shows as not being expressed at significant levels in neither standard, low salt or N limiting. In this study, it was not possible to clone from cDNA or to measure the transcription of the candidate choline dehydrogenase due to its very low level of transcription.

Likewise, *GSDMT* transcription from the pennate diatom *P. tricornutum* was also downregulated by both lowered salinity and N limitation (Figure 7.2). In this case, samples were taken from cultures grown at exponential growth phase in standard conditions, low salinity (5 PSU) and stationary phase with limiting N (initial N concentration of 30 μM of NO_3^-). Furthermore, exogenous addition of 0.5 mM choline also downregulated the expression of *GSDMT* whereas the addition of GBT, MTHB or glycine did not appear to have any significant effect on its transcription (Figure 7.2). Contrary to the role of *GSDMT* in *T. pseudonana*, *GSDMT* from *P. tricornutum* might not be the main gene responsible for GBT synthesis in the pennate diatom. *GSDMT* is downregulated by decreased salinity as it is the concentration of GBT. Likewise, *GSDMT* is also less transcribed in nitrogen starvation, however, GBT concentration does not significantly change in this condition. Moreover, addition of choline to the media increases the concentration of intracellular GBT. These last two conditions suggest that there might be an alternative synthesis pathway acting in *P. tricornutum*, likely a gene involved in GBT via the choline pathway. According to this hypothesis, the alternative pathway could be supplying the additional GBT observed in the metabolite analysis in low N conditions and in addition of choline. Moreover,

the activation of this pathway could be causing the downregulation of *GSDMT*. Indeed, *P. tricornutum* has a candidate choline dehydrogenase (PHATRDRAFT_1341), this protein does not seem to be regulated by salt as shown in the whole proteome sequencing. The presence of the two pathways for GBT synthesis in one organism and its alternation has been previously described in the bacterium *Actinopolyspora halophila*[211], however further investigation is required to clarify whether this situation also occurs in the model diatoms.

7.7.2 GSDMT from *T. pseudonana* catalyses all the methylation steps in the glycine pathway

TpGSDMT, THAPSDRAFT_20797 (XP_002286764.1), is a functional methyltransferase. When cloned into pET16b and transformed into *E. coli*, it confers to the heterologous host the ability to triple methylate glycine to make GBT using S-adenosyl methionine as methyl donor (Figure 7.4). Accumulation of GBT in *E. coli* protected this strain against high salinity (0.6 M NaCl in minimal media), whereas the control sample, unable to produce GBT, experienced a longer lag phase. Control samples eventually reach the same OD_{600} than *E. coli* containing the *GSDMT* from *T. pseudonana* (Figure 7.3) thanks to the ability of *E. coli* to make metabolic adaptations to osmotic stress. Some of these metabolic adaptations include the production of compatible solutes such as threhalose, glutamate, the membrane-stabilizing isoprenoid ubiquinone-8, hypotaurine, arginine, malate or N-acetylornithine[240]. This is the first enzyme involved in GBT synthesis in diatoms and the first ever multidomain methyltransferase to be cloned and characterised. Kageyama *et al.*[130] also identified this enzyme and showed that it was functional supporting the findings of this study.

7.7.3 Homologues of GDSMT from *T. pseudonana* are found in several organisms from different kingdoms.

A search in the NCBI database using BLASTp was performed to find homologues of *GSDMT* from *T. pseudonana*. Returned sequences showed that methyltransferases

from the diatoms *T. oceanica*, *Phaeodactylum tricornutum* CCAP 1055/1, *P. multistriata* and *Fragilariopsis cylindrus* CCMP1102 are the closest homologues (Table 7-1). However, similar candidate methyltransferases likely to be involved in GBT via the glycine pathway are found in a variety of organisms including alpha- and gamma-proteobacteria, cyanobacteria, euryarchaeotes such as *Methanohalophilus portucalensis* which Gsmt has been previously characterised[209], and the brown algae *E. siliculosus* (Figure 7.5). Those methyltransferases are grouped in seven distinctive clades in which the diatom methyltransferases are found in the Clade IV, shared with the brown algae *E. siliculosus* and other gamma-proteobacteria and alphaproteobacteria (Figure 7.5), although it is worth noting that the functionality of *E. siliculosus* has not been proven (Table 7-2).

7.7.4 Diatom methyltransferases synthesise GBT and are different from prokaryotic methyltransferases involved in GBT or GBT/DMSP synthesis.

The selected methyltransferases from the brown algae *Ectocarpus siliculosus* (CBN80020.1), the diatoms *T. oceanica* (THAOC_12055), *P. tricornutum* (XP_002180089.1) and *F. cylindrus* (S-Adenosyl-L-dependent methyltransferase, OEU2335.1), from the cyanobacteria *Geitlerinema* sp. PCC_9228 (WP_071515314.1), the alphaproteobacteria *Erythrobacter litoralis* (AOL23288) and *Mameliella alba* (WP_043147237.1) and the gammaproteobacteria *Thiohalobacter thiocyanaticus* (WP096366830.1) were tested for GBT and DMSP synthesis (Table *7-2*). The methyltransferases from the diatoms *P. tricornutum* and *T. oceanica*, and the cyanobacteria *Geitlerinema* sp. were positive for GBT synthesis, whereas methyltransferases from *E. siliculosus* and *F. cylindrus* tested negative for both GBT and DMSP synthesis. Surprisingly, the alphaproteobacteria *M. alba* and *E. litoralis* and the gammaproteobacteria *T. thiocyanaticus* tested positive for both GBT and DMSP synthesis (Table *7-2*).

An alignment of the two-domain proteins with Gsmt and Sdmt from *Synechococcus* sp. strain WH8102 (Q7U4Z8 and Q7U4Z9 respectively), and Gsmt and Sdmt from *Halorhodospira halochloris* (Q9KJ22 and Q9KJ21, respectively) suggest that, as TpGSDMT, these enzymes are capable of carrying out the entire methylation pathway from glycine to GBT. It is likely that the first methyltransferase domain methylates glycine and/or sarcosine to sarcosine and dimethylglycine respectively and the second domain sarcosine and/or dimethylglycine to dimethylglycine and GBT respectively.

Motifs characteristic of SAM-methyltransferases were identified by aligning all functional methyltransferases from this chapter and the previously characterised Gsdmt from *Actinopolyspora halophila*[211]. It stood out that despite having some common conserved amino acid residues, in each motif there was a clear difference between the diatom sequences and the prokaryotic sequences. Whereas the N terminus domain of the double-domain proteins followed most of the common SAM-methyltransferases motif patterns(Figure 7.7, Figure 7.8 and Figure 7.9) C-terminus domain and singled-domain proteins presented an additional Motif not previously described in literature and named in this study as Motif Pre I (Figure 7.10). The sequence xHxGx in Motif Pre I correspond to the first β strand (β1) of this methyltransferase domain (Appendix 4-2). Most importantly, differences in this domain's Motif II sequence could explain why GSDMT from diatoms are only capable of methylating glycine, sarcosine, dimethylglycine into GBT, whereas *M. alba*, *T. thiocyanaticus* and *E. litoralis* can catalyse the methylation of MTHB as well.

As a matter of fact, the Motif II of these unspecific proteins are more similar to the functional DSYD (see chapter 6) than to the diatom GSDMT.

Those common amino acids found in the Motif II of functional DSYD and in the bacterial Gsdmt are a V in the fourth position, a W in the fifth, a Q in the seventh or a A/C/S in the thirteenth position (Figure 7.12). It is noteworthy that Gsdmt from *A. halophila* has never been tested for the ability to transform MTHB into DMSP, and due to the sequence similarity, the possibility of this enzyme being indeed multifunctional

should not be ruled out. Furthermore, *Geitlerinema* sp. PC 9228 methyltransferase has been shown to only synthesise GBT, although very similar to the other bacterial methyltransferases in this study, it has a C in the sixth position of the motif II compared the S present in any of the other methyltransferases in this study (including DSYD) (Figure 7.12). Finally, the substitution of the conserved L in the 11th position of the GSDMT and DsyD/Gsdmt proteins (Figure 7.12) for a C or a V in the functional DMSHB synthesases (see chapter 6) constitute the main difference in the primary structure of the DMSP synthesis methyltransferases. This could potentially be the key to the specificity of DSYD for MTHB together with changes in the secondary structure of the proteins. The predicted secondary structures are different amongst species. Motif II and III (Appendix 20) of *E. litoralis* and the C-terminus methyltransferase domain of *T. pseudonana* are predicted to be within an alpha helix region and the fifth β strand (β5 or β5C) respectively, whereas in *M. alba* they are predicted to be the fifth β strand(β5C) and the sixth β strand (β6C) respectively. These secondary structure differs from the characterised SAM-methyltransferase domain found in yeast[226].

7.7.5 Functional DMSHB synthases and diatom GSDMT are evolutionary divergent to the prokaryotic methyltransferases.

Relatedness of the aligned functional DSYD, GSDMT, Sdmt and DsyD/Gsdmt sequences were visualised in a Maximum Likelihood phylogenetic tree. This tree shows that three distinctive groups named Clades. In Clade I all the functional DsyD and DSYD proteins characterised in chapter 6 are grouped together, and Clade III represents all of the diatom GSDMT uniquely involved in GBT synthesis. However, Clade II contains prokaryotic methyltransferases involved in either GBT or both DMSP and GBT synthesis (Figure 7.14).

It is important to notice that the Motif analysis in this study is merely descriptive and it contains a small representation of the enzymes. However, this study proves the complexity of distinguishing between SAM-methyltransferases involved in GBT

synthesis, those involved in DMSP synthesis or those involved the synthesis of both osmolytes at this stage. For instance, DSYD proteins present in Eukaryotes, including diatoms and brown alga, and DsyD and found in at least one confirmed prokaryote, the gamma-proteobacteria *G. sunshinyii* have not been tested for GBT synthesis, therefore it is not possible to completely disregard the possibility of these enzymes also being able to synthesise GBT. Likewise, the functional Gsdmt from *Actinopolyspora halophila* has never been tested for DMSP synthesis nor they have been tested for GBT synthesis those enzymes marked as non-functional DSYD.

It is important that in the future, efforts should be made to functionally characterise more of these methyltransferases and that biochemical analysis are performed on enzymes with point mutations in the Motif II to be able to understand the mechanics of these enzymes. Only then, more in depth bioinformatic analysis in large environmental datasets can be performed.

7.7.6 DsyD/Gsdmt containing alpha-proteobacteria *M. alba* synthesise both GBT and DMSP.

M. alba is an alpha-proteobacteria containing a multifunctional enzyme showed to be able to catabolise the methylation of glycine, sarcosine and or dimethylglycine to GBT (glycine pathway) as well as MTHB to DMSP (unknown mechanism). It also has in its genome candidate choline dehydrogenase homologue to the enzymes involved in GBT synthesis via the choline pathway in *E. coli*[46].

M. alba was grown in minimal media with excess of N or with low N present in the media and samples were taken for metabolite analysis. *M. alba* accumulates GBT and choline when N is abundant in the media, and GBT and DMSP when N is limiting (Table 7-3). Results of this study open the hypothesis of a complex interplay between the two possible GBT synthesis pathways and the production of DMSP. It is possible that in nitrogen replete condition, the synthesis of GBT is carried out by the DsyD/Gsdmt-like protein, hence DMSP is not being synthesised and choline accumulates in the cell. However, when N is limiting, DMSP production would be

favoured over GBT production by the methyltransferase enzyme and GBT would be synthesised via the choline pathway instead. As already mentioned, a relationship between two GBT pathways has been previously characterised in *A. halophila*[211] but it is the first time it has been reported a possible relationship between GBT synthesis and DMSP synthesis as the one found in *M. alba*. These findings also imply that there might be other upstream or downstream genes playing a more important role in the regulation of DMSP synthesis as it depends on substrate availability.

An alternative hypothesis would be that GBT could be only synthesised from choline in *M. alba* and when grown in high N, the abundance of choline produced is so high that not all is converted into GBT, therefore accumulating in the cell. These hypotheses need further investigation.

The work carried in this thesis highlights the imperial need for studying the kinetics of the functional GBT synthases, the newly found enzymes capable of synthesising both GBT and DMSP as well as further characterization of the DSYD genes.

Chapter 8

Discussion and concluding remarks

8 Discussion and concluding remarks

8.1 Research gaps and aims

Diatoms (Bacillariophyceae) are eukaryotic phytoplankton found in waters worldwide, from freshwater to marine environments[167]. Their ability to survive in such a wide range of salinities relay on their capacity to uptake[131] and synthesise[98] compatible solutes. *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* are two model diatoms with sequenced genomes[121,122] characterised for transporting[26] and synthesising a range of osmolytes, including the sulfur-containing zwitterion dimethylsulfoniopriopionate (DMSP) and the nitrogen-based zwitterion glycine betaine (GBT) amongst others[128,129]. DMSP is an abundant molecule in marine environments[104] whereas GBT is a molecule found in organisms across all kingdoms of life[6]. Furthermore, the catabolites of these two molecules are the climate active gas dimethylsulphide (DMS) originated by the lysis of DMSP[105], and trimethylamine[60] or methane[64] generated by the bacterial degradation of GBT.

Alga and phytoplankton synthesise DMSP via the transamination pathway[96]. DSYB (phytoplankton) and DsyB (bacteria) are the only known enzymes involved in this pathway and they carry out the methylation of MTHB to DMSHB, thought to be the committing step[89,99]. However, many DMSP producing diatoms, including the two model diatoms *T. pseudonana* and *P. tricornutum*, lack DSYB. Hence, the genetics behind DSYB-lacking organisms known to produced DMSP via the transamination pathway is yet to be fully comprehended. Similarly, GBT synthesis has been well characterised in many organisms[16,44,83,132], and the transport of DMSP and GBT is well known in bacteria[71] but how diatoms synthesise or transport these compatible solutes is not understood to date.

Consequently, this thesis aims to:

1. Understand how diatoms transport DMSP and GBT.

2. Identify and characterise the gene responsible for the methylation of MTHB to DMSHB in DSYB- lacking diatoms.
3. Elucidate which GBT synthesis pathway diatoms utilize, and which genes are responsible for this process.
4. Gather information about the regulation of the transport and synthesis of the two compatible solutes by environmental conditions.
5. Explore the distribution of homologues to the diatom DMSP and GBT synthesis and transport enzymes in other organisms.

The information provided by this thesis' findings will contribute to further understand which the main actors are and how and why DMSP and GBT are synthesised in the environment. This information will be useful to update current models used to predict the abundance of these osmolytes in the environment. More accurate models result in better predictions on the production of the climate active gases[242] resulted from the degradation of DMSP and GBT.

8.2 Major findings described in this thesis

8.2.1 DMSP and GBT production is regulated differently in *T. pseudonana* and *P. tricornutum.*

The effects of salinity, N availability, temperature changes and oxidative stress on DMSP and GBT synthesis by the two model diatoms *T. pseudonana* CCMP 1335 *and P. tricornutum* CCAP 1055/1 were studied. Moreover, *P. tricornutum* cultures were also exposed to exogenous DMSP, GBT, choline, glycine and MTHB and the accumulation of DMSP, choline and GBT was observed.

Salinity had a significant effect on *T. pseudonana* intracellular DMSP and GBT concentration. *T. pseudonana* did not produce any DMSP in reduced salinity (1 PSU) in mid- or late exponential growth phase (Figure 3.2), nor GBT in mid-exponential growth phase (Figure 3.7). It has been hypothesised that these two compatible solutes are acting as osmolytes alleviating the negative effects of high osmotic pressure. The

presence of GBT and DMSP in cultures grown at normal salinity (35 PSU) and their absence of GBT and DMSP in low salinity supports this hypothesis and previous findings[128,174]. A similar relationship between osmotic pressure and osmolyte accumulation was found in *P. tricornutum* cultures in mid-exponential phase (Figure 3.10, Figure 3.11). However, in late exponential growth phase, the intracellular concentration of DMSP in low salt was not statistically different to the control samples (Figure 3.2).

N limitation has been shown to cause an increase in the intracellular concentration of DMSP in both *T. pseudonana*[175] and *P. tricornutum* [129,186], and the reduction of GBT production to almost undetectable levels in *T. pseudonana*[175]. DMSP and GBT production by *T. pseudonana* cultures at the onset of stationary growth phase caused by nitrate depletion was compared to this of cultures in stationary phase caused by silica scarcity. Our experiments confirmed the presence of DMSP and absence of GBT in low N and vice versa in N replete cultures (Figure 3.9). On the other hand, a statistically difference in the concentration of intracellular DMSP in *P. tricornutum* is only observed when the cultures are exposed to nitrogen starvation for an extended period of time such as 5 days after entering stationary growth phase (Figure 3.11). Besides, the intracellular concentration of GBT and choline does not change significantly when N is depleted. However, these two compounds are not the major osmolyte in *P. tricornutum*. Proline, on the other hand, is the major osmolyte in this diatom and, as it is a N-based molecule, is indeed downregulated by N limitation[129].

In this thesis, it was found that temperature does not have a significant effect on DMSP synthesis in neither *T. pseudonana* nor in *P. tricornutum* (Figure 3.3). As these results contradict a previous study carried out by Spielmeyer and Pohnert[114], it is necessary to repeat the experiments to elucidate whether temperature is indeed a factor regulating DMSP and GBT in these model diatoms. Furthermore, exposing *P. tricornutum* and *T. pseudonana* to 0.75 mM H_2O_2 for 2h resulted in a significant increase in the accumulation of intracellular DMSP (Figure 3.5), supporting the role of these molecules as antioxidants[78].

Addition of DMSP increases significantly concentration of DMSP (Figure 3.11) and choline (Figure 3.12) in the cell, whereas it decreases the concentration of GBT (Figure 3.10). The increase of intracellular concentration of DMSP illustrates the ability of *P. tricornutum* to uptake this metabolite from the media[131]. On the other hand, the accumulation of choline in the cell could be explained if in this diatom, GBT was synthesised via the choline pathway, at least when choline is available. Thus, the decrease in intracellular GBT when DMSP is exogenously supplied would cause the accumulation of choline. The effect of DMSP on GBT concentration does not occur in the other model diatom *T. pseudonana*[130].

Furthermore, the addition of choline increases the concentration of GBT (Figure 3.10) supporting the hypothesis of choline being the precursor of this compatible solute. Moreover, if GBT is added to the media, *P. tricornutum* exhibit the ability to also uptake this metabolite from the media[131] (Figure 3.10). The accumulation of GBT in the cell leads to a rise in the accumulation of choline (Figure 3.12). Being the precursor of GBT, if this molecule is in excess in the cell the synthesis could be downregulated causing the accumulation of the precursor choline. Conversely, the abundance of GBT in the cell could activate its degradation generating choline[184].

Metabolite analysis of standard *P. tricornutum* cultures with added DMSP, GBT, and precursors suggest that the pennate diatom is capable of taking up DMSP, GBT, choline and MTHB from the media[131]. In addition, increase in the concentration of GBT when choline is added to the media suggest that *P. tricornutum* is capable of synthesising GBT via the choline pathway. As expected, supplying MTHB, an intermediate of the transamination pathway[96], causes an increase in DMSP concentration (Figure 3.11), this effect has also been reported by Kageyama *et al.*[101].

8.2.2 Whole transcriptome and proteome sequencing combined with BLASTp searches allowed the identification of candidate genes involved in GBT and DMSP synthesis

Candidate genes for synthesis and transport of GBT and DMSP were identified in the model diatoms. Candidates were found through literature review and by conducting searches in their genomes using enzymes known to carry the reactions involved in those processes as probes. In addition, whole transcriptome and proteome sequencing allowed to investigate the regulation of the candidate genes by salinity and nitrogen availability.

Using this approach, a candidate BCCT transport, the DMSHB synthase in diatoms (DSYD) and a gene involved in GBT synthesis (GSDMT) were identified and further characterised in this thesis. Other candidate genes have been proposed to be involved in GBT and DMSP synthesis and transport (Appendix 1-1, Appendix 1-2). It is important to bear in mind that the proposed candidate genes other than DSYD and GSDMT need to be functionally tested. In addition, it is possible that the actual enzymes for the first two steps and last step of the DMSP synthesis pathway have not been reported in this chapter as only the regulated proteins were studied. There is a possibility that the enzymes driving these steps are constitutively expressed and not regulated. It has been hypothesised that the methyltransferase is the key step controlling DMSP synthesis rates[89,99]. In addition, the annotation of the genomes of the model diatoms and the domain predictions might contain misannotations that could lead to overlook some of the sequences.

8.2.3 DSYD, a novel DMSHB synthase present in DSYB- lacking diatoms, alga and bacteria

In this study, a methyltransferase capable of adding a methyl group to MTHB, converting it into DMSHB, using SAM as the methyl donor was identified in the diatom *T. pseudonana* (TpDSYD) (Figure 8.1). TpDSYD has also been reported as a functional DMSHB synthase by Kageyama *et al.*[101]. Then, functional homologues were also identified in other two diatoms, *P. tricornutum* (PtDSYD) and *T. oceanica* (ToDSYD1). Strikingly, homologues to the functional diatoms DSYD were also found in the DMSP producing gamma-proteobacteria *G. sunshinyii*, in the unicellular green

algae *Ostreococcus tauri* and the brown algae *Sargassum vulgare*. (Table 6-1). A phylogenetic tree produced with the functional, non-functional and non-tested DSYD-like proteins show that although they are homologues, diatom DSYD are clearly divergent from alga and bacteria DSYD (Figure 6.2).

TpDSYD is downregulated by low salinity and upregulated by low nitrogen. This regulation was found both by RNAseq and RT-qPCR (Figure 6.6). On the other hand, *PtDSYD* is downregulated by lowered salt, addition of choline and GBT (Figure 6.7). These results, altogether with the accumulation of DMSP in standard conditions and the decrease in DMSP concentration when GBT is present in the cell (Figure 3.9, Figure 3.11), support the hypothesis of DMSP acting as an osmoprotectant in these model diatoms[128]. Surprisingly, there was no difference in the transcription of *PtDSYD* in nitrogen starved samples (Figure 6.7), despite observing an accumulation of DMSP in the cell (Figure 3.11). However, error bars in this experiment are large and further work would be required to confirm the expression of this gene in *P. tricornutum*. Regulation of the DMSP biosynthetic pathway was also observed in *G. sunshinyii*. In this bacterium, N limitation regulate the last step of the transamination pathway. Hence, when N is limiting, this bacterium accumulates DMSHB rather than DMSP (Figure 6.9). Although not quite understood, it is likely due to the N requirement of the enzyme predicted to carry out the last step, a pyridoxal phosphate (PLP)-dependent aminotransferase[96].

PtDSYD exhibit the maximum activity at 30 °C and pH 6.8 (Figure 6.5). These parameters were used to calculate the K_M value for MTHB. PtDSYD possesses a K_M value 2.1-fold higher (Figure 6.4) than the previously characterised *P. parvum* DSYB[99]. *P. parvum* is a species of haptophyte, which is one the greatest DMSP producers[99], whereas the pennate diatom does not produce as much DMSP. Thus, PtDSYD's lower affinity to MTHB is not completely unexpected.

P. parvum DSYB was found to be concentrated in mitochondria and chloroplasts[99]. Likewise, PtDSYD was also found to be concentrated in the mitochondria, chloroplast

and some unidentified vesicles (Figure 6.8). This supports the hypothesis of DMSP being produced in those cellular compartments[99]. A candidate BCCT transporter studied in this thesis was also found to be accumulated in the membranes of mitochondria, chloroplasts and the unidentified vesicles. I hypothesise that this transporter could be distributing DMSP and GBT across the cells. For instance, in *T. pseudonana* the candidate BCCT transporter is downregulated in lower salinity and low N whereas it is highly transcribed at standard salinity (Figure 5.7). Similarly, BCCT transporter is downregulated by nitrogen starvation in *P. tricornutum* and upregulated by the addition of GBT, DMSP and the GBT precursors glycine and choline (Figure 5.8). The transporter is equally transcribed in standard cultures, cultures grown at 5 PSU and in additional MTHB (Figure 5.8). In this hypothesis, when N is limiting DMSP would be maintained in the organelles to alleviate the oxidative stress produced by the lack of N[78,129], whereas, in increased salinity, transportation of the compatible solutes to the cytosol would be favoured. Nevertheless, it is worth noting that functionality of the BCCT transporter is still to be proven.

The discovery of DMSP producing bacteria in 2017 allowed the characterization, for the first time, of a DMSP synthesis gene named *dsyB*[89]. Next, homologues of dsyB, termed DSYB, were found to be widespread in phytoplankton[99]. In this work, DSYD is proposed as the main gene involved in DMSP synthesis in DSYB-lacking diatoms and in alga, considered high DMSP producers[97]. Furthermore, DSYD, like DSYB, is also present at least one bacterium.

8.2.4 Diatoms and many bacteria use the methyltransferase pathway to synthesise GBT *de novo*.

The first gene involved in GBT synthesis in phytoplankton was characterised in this study, a functional glycine sarcosine dimethylglycine methyltransferase in the centric diatom *T. pseudonana*. This discovery has been backed by the findings of Kageyama *et al.*[130]. TpGSDMT is not only the first GBT synthase found in phytoplankton but it is

also the first enzyme capable of carrying the three methylation steps to be characterised[234,237] (Figure 8.1). Later, homologue GSDMT from the diatoms *T. oceanica* (ToGSDMT) and *P. tricornutum* (PtGSDMT) were also proven to be functional, as well as the single domain methyltransferase from the cyanobacteria *Geitlerinema* sp.

The regulation of the GBT synthases in the model diatoms *T. pseudonana* and *P. tricornutum* was analysed. In both diatoms, reduced salinity causes a downregulation of the expression of *GSDMT* (Figure 7.1, Figure 7.2). Salinity is also a factor influencing the intracellular concentrations of GBT in these diatoms, supporting its role as an osmoprotectant. At higher salinity, GBT concentration in the cell increases, whereas at low salinity, GBT it decreases, even to undetectable levels in the case of *T. pseudonana* (Figure 3.7). This regulation occurs in *P. tricornutum* too (Figure 3.10), despite GBT not being the main osmolyte in this diatom[129]. Nitrogen limitation causes a downregulation in the transcription of *TpGSDMT* and *PtGSDMT* (Figure 7.1, Figure 7.2). N availability also has a dramatic effect on *T. pseudonana*'s intracellular GBT, lowering it to undetectable levels (Figure 3.9). In contrast, GBT concentration does not change in N starved *P. tricornutum* cultures (Figure 3.10). This, together with regulation by the addition of choline, which cause the downregulation of *PtGSDMT* (Figure 7.2) and the increase in GBT concentration (Figure 3.10), leads to think that *PtGSDMT* might not be the main gene involved in GBT synthesis in this diatom. Indeed, *P. tricornutum* has a candidate choline dehydrogenase (PHATRDRAFT_1341), although it is not regulated by salt according to our whole proteome sequencing database. It is also possible that both pathways are active and alternating, a phenomenon previously described in the bacterium *Actinopolyspora halophila*[211], however further investigation is required to clarify whether this situation also occurs in the model diatom.

Many GSDMT-like proteins or TpGSDMT homologues with only one methyltransferase domain (Figure 7.6) are found in many other organisms, mostly prokaryotes. Homologues include a methyltransferase in the brown algae *E.*

siliculosus or the functionally characterised Gsmt from the methanoarchaeon *Methanohalophilus portucalensis*[209], although the candidate GSDMT from *E. siliculosus* did not show any methyltransferase activity for glycine, sarcosine, dimethylglycine or MTHB (Table 7-2). It was always understood that the choline pathway was the main GBT biosynthesis pathway and the methylation pathway was regarded as an exception. Results from this thesis imply that the synthesis of GBT from glycine might be more prevalent than previously thought. On the other hand, how other phytoplankton synthesise the nitrogenous osmolyte is still to be elucidated.

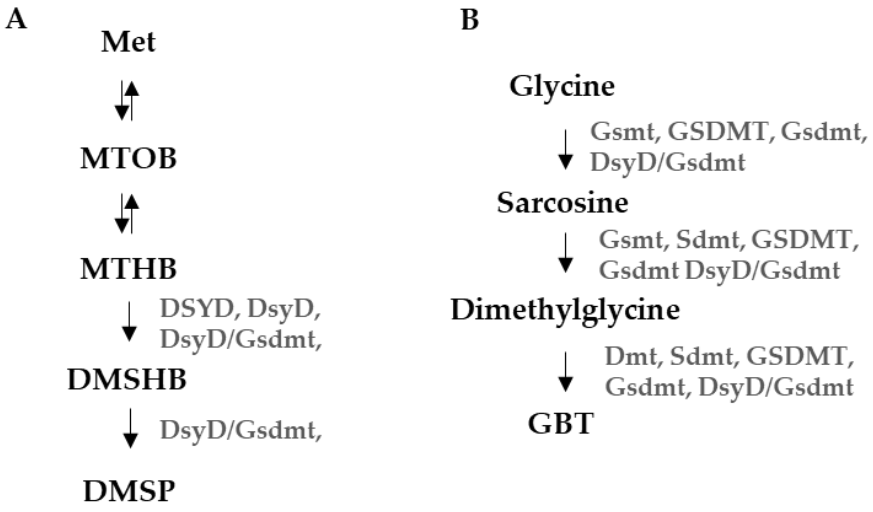
8.2.5 A novel enzyme found in bacteria is capable of both DMSP and GBT synthesis

In the searching for GSDMT homologues some methyltransferases were selected to test whether they were functional or not (Table *7-2*). Selected GSDMT-like methyltransferases were tested for their ability to add a methyl group to GBT precursors glycine, sarcosine and dimethylglycine, as well as to the DMSP precursor MTHB (Figure 8.1). Three of the methyltransferases belonging to the alphaproteobacteria *M. alba* and *E. litoralis* and the gammaproteobacteria *T. thiocyanaticus* were capable of producing both GBT and DMSP (Table 7-1). These three multi-functional enzymes were tested in the heterologous host *E. coli* which is not able to catalyse the last step of the transamination pathway, the decarboxylation of DMSHB to DMSP. The mechanism how these enzymes turn MTHB to DMSP is not yet understood and requires further investigation. These novel enzymes are yet another supporting evidence of the interplay of GBT and DMSP, two analogous molecules.

A Maximum Likelihood phylogenetic tree was produced with sequences from functional GSDMT, Gsdmt, GSDMT/DSYD, DSYD, non-functional methyltransferases and GSDMT-like proteins. In this tree, DSYD proteins are grouped together. Another group englobes a series of prokaryotic methyltransferases with mixed functions including the tested DSYD/GSMT enzymes, the functional GBT

from *Geitlerinema* sp PCC 9228 and *Actinopolyspora halophila* and the non-functional DSYD methyltransferases from *Nesiotobacter exalbescens* and *Aidingimonas halophila*. On the other hand, diatom GSDMT cluster together. It is worth noting that the group including the prokaryotic mixed functional methyltransferases are phylogenetically closer to functional DSYD than to diatom GSDMT (Figure 7.14). Analysis of the primary structure of the functional GSDMT, DSYD/GSDMT and DSYD show that diatom GSDMT are different from DSYD/GSDMT and DSYD from alga and bacteria, which, at the same time are different from the diatom DSYD (Figure 7.12). It is important to bear in mind, that DSYD proteins were not tested for GBT synthesis, conversely Gsdmt from *A. halophila* has not been tested for DMSP synthesis so to what extent the double function enzymes are common is not known yet. Most importantly, focus should be put on charactering the structure and kinetics of these novel proteins.

Figure 8.1. DMSP and GBT biosynthetic pathways and steps carried out by the novel enzymes found in this thesis. A) DMSP transamination pathway, the enzymes DSYD, DsyD and DsyD/Gsdmt catalyse the methylation of MTHB to DMSHB using SAM as methyl donor. DsyD/Gsdmt can catabolize the conversion from DMSHB to DMSP via an unknown mechanism. B) GBT methylation pathway, GSDMT and Gsdmt are capable of catabolising the three methylation steps from glycine to GBT. Gsmt, Sdmt and Dmthave only affinity for one of the GBT precursors, whereas DsyD/Gsdmt can carry the entire reaction or only part of it.



The alpha-proteobacteria containing a multifunctional enzyme *M. alba* is a DMSP and GBT producing bacteria. Apart from the methyltransferase capable of synthesising DMSP and GBT it also has in its genome a candidate choline dehydrogenase homologue to the enzymes involved in GBT synthesis via the choline pathway in *E. coli*[46]. *M. alba* accumulates GBT and choline when N is abundant in the media, and GBT and DMSP when N is limiting (Table 7-3). It would be interesting to know the mechanism used to prefer one osmolyte over the other and whether *M. alba* synthesises DMSP and GBT using the multifunctional enzyme or whether, as in the case of *A. halophila*[211], it has an alternative GBT synthesis pathway from choline.

8.3 Limitations of this study

8.3.1 Limitations of the diatom work

Growth experiments described in Chapter 3 were repeated in several occasions and it was found that there was a great variability in the growth of *T. pseudonana*. In addition, the regulation of the genes of interest in this study have been investigated using RNA sequencing and RT-qPCR. Only the effect of salinity in *P. tricornutum* has been contemplated looking at both transcription and protein levels. It is recommended to support RNA work with other methodologies such as western blotting to have a more accurate vision of the regulation of the methyltransferases involved in synthesis and transport of GBT and DMSP. Moreover, metabolite detection has been performed using two different techniques, NMR and LC/MS, therefore results are not comparable. Lastly, none of the genes from the model diatoms have been knocked out in the original organism, therefore, the involvement of DSYD and GSDMT in the synthesis of the compatible solutes has not been completely demonstrated.

8.3.2 Bioinformatics limitations

NCBI database is very comprehensive, however, it does not contain the genome of many phytoplankton. Therefore, searches in specialised databases such as in the

Marine Microbial Eukaryote Transcriptome Sequencing Project[243] via the sequencing repositories iMicrobe (https://imicrobe.us/#/projects/104) and ENA (European Nucleotide Archive)[244] are needed. As seen in this study, it is very difficult to date to be able to discern between DSYD, DSYD/GSDMT and GSDMT enzymes, and domain predictions could not be accurate, hence, care should be taken before making predictions about abundance of those genes in the environment. Finally, signal peptides or signalling motifs targeting specific organelles in diatoms are still understudied and software is not comprehensive enough.

8.4 Recommendations for future research

8.4.1 Confirming DSYD and GSDMT and BCCT as the main agents for synthesis and transport of DMSP in diatoms, alga and bacteria

Next steps towards the investigation of these novel enzymes include the creation of knock out mutants in diatoms and also in bacteria. During this thesis, attempts were made to mutate the genes using CRISPR/Cas9[126] technique in *T. pseudonana* (data not shown). Unfortunately, due to external circumstances it was not possible to obtain any mutant. It would be advised pursuing to mutate either *T. pseudonana*[126] and/or *P. tricornutum*[127] as it would not only be a good indication of the role of the novel genes in the cell but it could also through some light into the role DMSP is playing in the cell. Mutants can also be performed in the bacteria found to contain the novel enzymes.

Many methyltransferases homologues to DSYD or GSDMT were found in this thesis, and further functional characterization is required. Perhaps, it should be considered to widen the heterologous host spectrum to test those enzymes deemed as non-functional in this study. It is also necessary to carry out protein work to support the regulation of DSYD, GSDMT and BCCT in the model diatoms proposed in this thesis. It would also be interesting to better understand the regulation of Gsdmt and DsyD/Gsdmt in bacteria. Studying the regulation of Gsdmt is specially interesting in

relation to the alternative GBT synthesis pathway from choline, to understand to which extent both pathways coexist, alternate or dominate in *P. tricornutum* and *M. alba*.

8.4.2 Cellular localisation of DMSP and GBT in diatoms

Immunogold labelling experiments place DSYD and BCCT in the mitochondria, chloroplasts and unidentified vesicles. First, identification of the vesicles would be desirable. Secondly, it would also be interesting where is GSDMT accumulated in the cell. To further investigate whether DMSP and GBT is accumulated in different cellular compartments depending on the environmental conditions two methodologies can be proposed. First, organelle extraction combined with western blotting and metabolite analysis would further confirm the findings of the immunogold labelling experiments. Secondly, using NanoSIMS targeting $DMSP^{245}$ and GBT.

8.4.3 Biochemical characterization of DSYD, GSDMT and DSYD/GSDMT

Perhaps one of the most important issues to elucidate is the biochemical characterization of DSYD, DsyD, DsyD/Gsdmt and GSDMT. These would involve the analysis of substrate affinity and specificity for each enzyme and the elucidation of their main differences through site directed mutagenesis and structural analysis. This would allow to predict whether candidate homologues are functional or not. Furthermore, studying the extra domains found in DsyD from *G. sunshinyii* and ToDSYD2 could help the search of the upstream and downstream enzymes of the transamination pathway.

8.4.4 In depth evolution and distribution in nature of the three novel enzymes

Another recommendation for future research include searching for more DSYD, DsyD, DsyD/Gsdmt and Gsdmt homologues in other more comprehensive databases

(such as iMicrobe, ENA[244] or TARA Ocean). Also, carrying out an evolutionary study to understand how these enzymes have spread out from organisms in different kingdoms of life would be interesting. Nevertheless, it is important to bear in mind that the genetics of DMSP production by bacteria and other organisms apart from phytoplankton is still vastly unknown. In addition, the discovery of enzymes capable of synthesising both DMSP and GBT and the lack of reliable biochemical information that would allow to tell them apart makes bioinformatic searches using those enzymes as probes are unreliable. For instance, Nelson *et al.* used homologues of TpDSYD found in NCBI database and selected those diatom sequences with an e-value of $10e^{-10}$ or lower, which included *F. cylindrus*, a methyltransferase shown not to be functional in this thesis. This study calls for caution and for further molecular work that can back up the data mining from metagenome/ transcriptome/ proteomes and metabolomes.

8.4.5 Use of the novel enzymes to improve crops

I was awarded a grant from the Community Resource for Wheat Transformation grant by the National Institute of Agricultural Botany in 2016. As part of this award, the NIAB introduced *TpGSDMT* into wheat. In collaboration with Dr. Ben Miller, the plantlets were genotyped, and the metabolites were analysed. However, GBT detection from wheat was a limitation and an optimised method is required. In the future, studying where does GBT accumulate in the plant and how much compared to wild type as well as whether GBT protects the plants against stressors will likely lead to the development of more nutritional and resistant crops[53].

8.5 Concluding remarks

DMSP and GBT are abundant osmolytes in marine environments. Many bacteria, alga and phytoplankton can both synthesise and uptake these compatible solutes. In addition, DMSP is broken down releasing the gas DMS and GBT is degraded to TMA or methane. DMS and methane are climate active gases, therefore, knowing how,

when and why the precursors are produced could help to understand the influence marine producing organisms have on the climate.

In this thesis, the regulation of the synthesis and transport of DMSP and GBT by model diatoms *T. pseudonana* and *P. tricornutum* was investigated. *T. pseudonana* exhibited a reciprocal relationship between DMSP and GBT depending on nitrogen availability, and a direct relationship between salinity and their synthesis. On the other hand, the relationship was no so clear in *P. tricornutum*. In the pennate diatom, DMSP and GBT were downregulated by low salinity and DMSP was upregulated by nitrogen starvation, however, GBT was not affected by N limitation, despite being a nitrogenous compound. Furthermore, no differences were found in intracellular concentrations of DMSP in cultures subjected to temperature changes, and further experiments are required to establish the regulation of DMSP by oxidative stress.

Searches in their published genomes as well as in whole transcriptome and proteome sequencing of cultures exposed to salinity shifts and nitrogen limitation allowed the identification of several candidate genes likely involved in DMSP and GBT synthesis. Most importantly, candidate methyltransferases involved in GBT and DMSP synthesis in these DSYB-lacking diatoms and candidate transporter were discovered.

A novel methyltransferase involved in DMSP synthesis was found in the model diatoms *T. pseudonana* and *P. tricornutum* named DSYD (DMSHB synthase in diatoms). DSYD was present in other diatoms, brown and green algae and a gamma-proteobacteria. This is the first gene functionally characterised for DMSP synthesis in alga. The regulation of DMSP synthesis and the transcription of *DSYD* suggest that the sulfurous osmolyte could be playing different roles in the cell depending on the environmental conditions. DSYD was localised in the chloroplasts and mitochondria, and a candidate BCCT transporter was also found to be placed in the membranes of those organelles. Hence, when oxidative stress is high, such in low nitrogen, DMSP would be mostly produced and kept in the organelles acting as an antioxidant,

whereas in high salinity, the transporter would distribute DMSP to the cytoplasm acting as an osmoprotectant.

Furthermore, a functional methyltransferase able to methylate glycine to glycine betaine named GSDMT (glycine sarcosine dimethylglycine methyltransferase) was found in these two and other diatoms, such as *Thalassiosira oceanica*, and in bacteria and cyanobacteria. Nonetheless, the effect of choline in *P. tricornutum* indicates that this diatom could be synthesising GBT via the choline pathway or using both pathways. Although the coexistence of the glycine and the choline pathway has been reported in the bacterium *Actinopolyspora halophila*, not much is known about the occurrence of this phenomenon. Furthermore, a novel enzyme present in DMSP producing bacteria was able of synthesising both DMSP and GBT. The presence of two potential GBT synthesis pathways, and a shared DMSP/GBT synthesising enzyme stress the close relationship between these two abundant osmolytes. Most importantly, it highlights how little is known about the interplay between the DMSP and GBT biosynthetic pathways. Also, until further structural and biochemical characterization of the novel enzymes found in this thesis are performed, it is not possible to predict from the primary sequence whether they would be functional DMSP, GBT or DMSP/GBT synthases, corroborating how much the contribution of bacteria towards DMSP synthesis has likely been underestimated

References

1. Brown, A. D. Microbial water stress. *Bacteriol. Rev.* **40**, 803–846 (1976).

2. Csonka, L. N. Physiological and genetic responses of bacteria to osmotic stress. *Microbiol. Rev.* **53**, 121–147 (1989).

3. Poolman, B. & Glaasker, E. Regulation of compatible solute accumulation in bacteria. *Mol. Microbiol.* **29**, 397–407 (1998).

4. Somero, G. N. Protons, osmolytes, and fitness of internal milieu for protein function. *Am. J. Physiol.* **251**, R197–R213 (1986).

5. Pollard, A. & Wyn Jones, R. G. Enzyme activities in concentrated solutions of glycinebetaine and other solutes. *Planta* **144**, 291–298 (1979).

6. Yancey, P. H., Clark, M. E., Hand, S. C., Bowlus, R. D. & Somero, G. N. Living with water stress: evolution of osmolyte systems. *Science* **217**, 1214–1222 (1982).

7. Empadinhas, N. & Da Costa, M. S. Osmoadaptation mechanisms in prokaryotes: Distribution of compatible solutes. *Int. Microbiol.* **11**, 151–161 (2008).

8. Eisenberg, H. & Wachtel, E. J. Structural studies of organelles of bacteria adapted to extreme salt. *Ann. Rev. Bioohys. Chem.* **16**, 69–92 (1987).

9. Galinski, E. A. Compatible solutes of halophilic eubacteria: molecular principles, water-solute interaction, stress protection. *Experientia* **49**, 487–496 (1993).

10. Winzor, C. L., Winzor, D. J., Paleg, L. G., Jones, G. P. & Naidu, B. P. Rationalization of the effects of compatible solutes on protein stability in terms of thermodynamic nonideality. *Arch. Biochem. Biophys.* **296**, 102–107 (1992).

11. Arakawa, T. & Timasheff, S. N. The stabilization of proteins by osmolytes. *Biophys. J.* **47**, 411–414 (1985).

12. Galinski, E. a., Stein, M., Amendt, B. & Kinder, M. The kosmotropic (structure-forming) effect of compensatory solutes. *Comp. Biochem. Physiol.* **117A**, 357–365 (1997).

13. Roberts, M. F. Microorganisms. *Saline Systems* **30**, 1–30 (2005).

14. Welsh, D. T. Ecological significance of compatible solute accumulation by micro-organisms: from single cells to global climate. *FEMS Microbiol. Rev.* **24**, 263–290 (2000).

15. Verheul, A., Glaasker, E., Poolman, B. & Abee, T. Betaine and L-carnitine transport by *Listeria monocytogenes* Scott A in response to osmotic signals. *J. Bacteriol.* **179**, 6979–6985 (1997).

16. Boch, J., Kempf, B., Schmid, R. & Bremer, E. Synthesis of the osmoprotectant glycine betaine in *Bacillus subtilis*: Characterization of the *gbs*AB genes. *J. Bacteriol.* **178**, 5121–5129 (1996).

17. Anthoni, U., Christophersen, C., Hougaard, L. & Nielsen, P. H. Quaternary ammonium compounds in the biosphere-An example of a versatile adaptive strategy. *Comp. Biochem. Physiol.* **99B**, 1–18 (1991).

18. Imhoff, J. F. & Rodriguez-Valera, F. Betaine is the main compatible solute of halophilic eubacteria. *J. Bacteriol.* **160**, 478–479 (1984).

19. Oren, A., Elevi Bardavid, R., Kandel, N., Aizenshtat, Z. & Jehlička, J. Glycine betaine is the main organic osmotic solute in a stratified microbial community in a hypersaline evaporitic gypsum crust. *Extremophiles* **17**, 445–451 (2013).

20. Blunden, G., Smith, B. E., Irons, M. W., Yang, M. H., Roch, O. G. & Patel, A. V. Betaines and tertiary sulphonium compounds from 62 species of marine algae. *Biochem. Syst. Ecol.* **20**, 373–388 (1992).

21. Reed, R. H., Chudek, J. A., Foster, R. & Stewart, W. D. P. Osmotic adjustment

in cyanobacteria from hypersaline environments. *Arch. Microbiol.* **138**, 333–337 (1984).

22. Rhodes, D. & Hanson, A. D. Compounds in higher plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **44**, 357–384 (1993).

23. Cosquer, A., Pichereau, V., Pocard, J. A., Minet, J., Cormier, M. & Bernard, T. Nanomolar levels of dimethylsulfoniopropionate, dimethylsulfonioacetate, and glycine betaine are sufficient to confer osmoprotection to *Escherichia coli*. *Appl. Environ. Microbiol.* **65**, 3304–3311 (1999).

24. Ryther, J. H. & Dunstan, W. M. Nitrogen, phosphorus, and eutrophication in the coastal marine environment. *Science (80-.).* **171**, 1008–1014 (1971).

25. King, G. M. Distribution and metabolism of quaternary amines in marine sediments. in *Nitrogen cycling in coastal marine environments* (eds. Blackburn, T. H. & Sorensen, J.) 143–173 (John Wiley & Sons Ltd, 1988).

26. Vila-Costa, M., Simó, R., Harada, H., Gasol, J. M., Slezak, D. & Kiene, R. P. Dimethylsulfoniopropionate uptake by marine phytoplankton. *Science (80-.).* **314**, 652–654 (2006).

27. Kiene, R. P. Uptake of Choline and its conversion to glycine betaine by bacteria in estuarine waters. *Appl. Environ. Microbiol* **64**, 1045–1051 (1998).

28. Smith, L. T., Pocard, J. A., Bernard, T. & Le Rudulier, D. Osmotic control of glycine betaine biosynthesis and degradation in *Rhizobium meliloti*. *J. Bacteriol.* **170**, 3142–3149 (1988).

29. May, G., Faatz, E., Villarejo, M. & Bremer, E. Binding protein dependent transport of glycine betaine and its osmotic regulation in *Escherichia coli* K12. *Mol. Gen. Genet.* **205**, 225–233 (1986).

30. Chen, C., Malek, A. A., Wargo, M. J., Hogan, D. A. & Beattie, G. A. The ATP-

binding cassette transporter Cbc (choline/betaine/carnitine) recruits multiple substrate-binding proteins with strong specificity for distinct quaternary ammonium compounds. *Mol. Microbiol.* **75**, 29–45 (2010).

31. Cairney, J., Booth, I. R. & Higgins, C. F. *Salmonella typhimurium proP* gene encodes a transport system for the osmoprotectant betaine. *J. Bacteriol.* **164**, 1218–1223 (1985).

32. Kappes, R. M., Kempf, B. & Bremer, E. Three transport systems for the osmoprotectant glycine betaine operate in *Bacillus subtilis*: characterization of OpuD . *J. Bacteriol.* **178**, 5071–5079 (1996).

33. Higgins, C. F. ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.* **8**, 67–113 (1992).

34. Tøndervik, A. & Strøm, A. R. Membrane topology and mutational analysis of the osmotically activated BetT choline transporter of *Escherichia coli*. *Microbiology* **153**, 803–813 (2007).

35. Lai, S. J. & Lai, M. C. Characterization and regulation of the osmolyte betaine synthesizing enzymes GSMT and SDMT from halophilic methanogen *Methanohalophilus portucalensis*. *PLoS One* **6**, e25090 (2011).

36. Kleber, H. P. Bacterial carnitine metabolism. *FEMS Microbiol. Lett.* **147**, 1–9 (1997).

37. Styrvold, O. B., Falkenberg, P., Landfald, B., Eshoo, M. W., Bjørnsen, T. & Strøm, A. R. Selection, mapping, and characterization of osmoregulatory mutants of *Escherichia coli* blocked in the choline-glycine betaine pathway. *J. Bacteriol.* **165**, 856–863 (1986).

38. Landfald, B. & Strom, A. R. Choline-glycine betaine pathway confers a high level of osmolyte tolerance in *Escherichia coli*. *J. Bacteriol.* **165**, 849–855 (1986).

39. Kiene, R. P. Uptake of choline and its conversion to glycine betaine by bacteria in estuarine waters. *Appl. Environ. Microbiol.* **64**, 1045–1051 (1998).

40. Rendinat, G. & Hospital, F. Studies on choline dehydrogenasee: I. EXTRACTION IN SOLUBLE FORM, ASSAY, AND SOME PROPERTIES OF THE ENZYME. *J. Biol. Chem.* **234**, 1605–1610 (1959).

41. Wilken, D. R., McMacken, M. L. & Rodriquez, A. Choline and betaine aldehyde oxidation by rat liver mitochondria. *Biochim. Biophys. Acta* **216**, 305–317 (1970).

42. Lamark, T., Kaasen, I., Eshoo, M. W., Falkenberg, P., McDougall, J., Strom, A. R. DNA sequence and analysis of the bet genes encoding the osmoregulatory choline—glycine betaine pathway of *Escherichia coli*. *Arch. Biochem. Biophys.* **271**, 56–63 (1989).

43. Sakamoto, A. & Murata, N. Genetic engineering of glycinebetaine synthesis in plants: current status and implications for enhancement of stress tolerance. *J. Exp. Bot.* **51**, 81–88 (2000).

44. Gadda, G. & McAllister-Wilkins, E. E. Cloning, expression, and purification of choline dehydrogenase from the moderate halophile *Halomonas elongata*. *Appl. Environ. Microbiol.* **69**, 2126–2132 (2003).

45. Weretilnyk, E. A. & Hanson, A. D. Betaine aldehyde dehydrogenase from spinach leaves: Purification, in vitro translation of the mRNA, and regulation by salinity. *Arch. Biochem. Biophys.* **271**, 56–63 (1989).

46. Velasco-García, R., Villalobos, M. A., Ramírez-Romero, M. A., Mújica-Jiménez, C., Iturriaga, G. & Muñoz-Clares, R. A. Betaine aldehyde dehydrogenase from *Pseudomonas aeruginosa*: Cloning, over-expression in *Escherichia coli*, and regulation by choline and salt. *Arch. Microbiol.* **185**, 14–22 (2006).

47. Falkenberg, P. & Strom, A. R. Purification and characterization of

osmoregulatory betaine aldehyde dehydrogenase of *Escherichia coli*. *Biochim. Biophys. Acta* **1034**, 253–259 (1990).

48. Nyyssölä, A., Reinikainen, T. & Leisola, M. Characterization of glycine sarcosine N-methyltransferase and sarcosine dimethylglycine N-methyltransferase. *Appl. Environ. Microbiol.* **67**, 2044–2050 (2001).

49. Lu, W. D., Chi, Z. M. & Su, C. D. Identification of glycine betaine as compatible solute in *Synechococcus* sp. WH8102 and characterization of its N-methyltransferase genes involved in betaine synthesis. *Arch. Microbiol.* **186**, 495–506 (2006).

50. McCoy, J. G., Bailey, L. J., Ng, Y. H., Bingman, C. A., Wrobel, R., Weber, A. P. M., Fox, B. G. & Phillips, G. N. Discovery of sarcosine dimethylglycine methyltransferase from *Galdieria sulphuraria*. *Proteins Struct. Funct. Bioinforma.* **74**, 368–377 (2009).

51. Nyyssölä, A., Kerovuo, J., Kaukinen, P., Von Weymarn, N. & Reinikainen, T. Extreme halophiles synthesize betaine from glycine by methylation. *J. Biol. Chem.* **275**, 22196–22201 (2000).

52. Schönknecht, G., Chen, W.-H., Ternes, C. M., Barbier, G. G., Shrestha, R. P., Stanke, M., Bräutigam, A., Baker, B. J., Banfield, J. F., Garavito, R. M., Carr, K., Wilkerson, C., Rensing, S. a, Gagneul, D., Dickenson, N. E., Oesterhelt, C., Lercher, M. J. & Weber, A. P. M. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* **339**, 1207–10 (2013).

53. Waditee, R., Bhuiyan, M. N. H., Rai, V., Aoki, K., Tanaka, Y., Hibino, T., Suzuki, S., Takano, J., Jagendorf, A. T., Takabe, T. & Takabe, T. Genes for direct methylation of glycine provide high levels of glycinebetaine and abiotic-stress tolerance in *Synechococcus* and *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 1318–1323 (2005).

54. Waditee, R., Tanaka, Y., Aoki, K., Hibino, T., Jikuya, H., Takano, J., Takabe, T. & Takabe, T. Isolation and functional characterization of N-methyltransferases that catalyze betaine synthesis from glycine in a halotolerant photosynthetic organism *Aphanothece halophytica*. *J. Biol. Chem.* **278**, 4932–4942 (2003).

55. Chen, S. Y., Lai, M. C., Lai, S. J. & Lee, Y. C. Characterization of osmolyte betaine synthesizing sarcosine dimethylglycine N-methyltransferase from *Methanohalophilus portucalensis*. *Arch. Microbiol.* **191**, 735–743 (2009).

56. Keller, M. D., Kiene, R. P., Matrai, P. A. & Bellows, W. K. Production of glycine betaine and dimethylsulfoniopropionate in marine phytoplankton. I. Batch cultures. *Mar. Biol.* **135**, 237–248 (1999).

57. White, R. F., Kaplan, L. & Birnbaum, J. Betaine-homocysteine transmethylase in *Pseudomonas denitrificans*, a vitamin B12 overproducer. *J. Bacteriol.* **113**, 218–223 (1973).

58. Wargo, M. J., Szwergold, B. S. & Hogan, D. A. Identification of two gene clusters and a transcriptional regulator required for *Pseudomonas aeruginosa* glycine betaine catabolism. *J. Bacteriol.* **190**, 2690–2699 (2008).

59. Meskys, R., Harris, R. J., Casaite, V., Basran, J. & Scrutton, N. S. Organization of the genes involved in dimethylglycine and sarcosine degradation in *Arthrobacter* spp.: Implications for glycine betaine catabolism. *Eur. J. Biochem.* **268**, 3390–3398 (2001).

60. Naumann, E., Hippe, H. & Gottschalk, G. Betaine: New oxidant in the Stickland reaction and methanogenesis from betaine and l-alanine by a *Clostridium sporogenes-Methanosarcina barkeri* coculture. *Appl. Environ. Microbiol.* **45**, 474–483 (1983).

61. Mouné, S., Manac'h, N., Hirschler, A., Caumette, P., Willison, J. C. & Matheron, R. *Haloanaerobacter salinarius* sp. nov., a novel halophilic fermentative

bacterium that reduces glycine-betaine to trimethylamine with hydrogen or serine as electron donors; emendation of the genus *Haloanaerobacter*. *Int. J. Syst. Bacteriol.* **49 Pt 1**, 103–112 (1999).

62. Orphan, V., Hinrichs, K., Ussler, W., Paull, C., Taylor, L., Sylva, S. & Al, E. Comparative analysis of methane-oxidizing archaea and sulfate-reducing bacteria in anoxic marine sediments. *Appl. Environ. Microbiol.* **67**, 1922–34 (2001).

63. Neill, a R., Grime, D. W. & Dawson, R. M. Conversion of choline methyl groups through trimethylamine into methane in the rumen. *Biochem. J.* **170**, 529–535 (1978).

64. Jones, H. J., Kröber, E., Stephenson, J., Mausz, M. A., Jameson, E., Millard, A., Purdy, K. J. & Chen, Y. A new family of uncultivated bacteria involved in methanogenesis from the ubiquitous osmolyte glycine betaine in coastal saltmarsh sediments. *Microbiome* **7**, 1–11 (2019).

65. Watkins, A., Roussel, E., Parkes, R. & Sass, H. Glycine betaine as a direct substrate for methanogens (*Methanococcoides* spp.). *Appl. Environ. Microbiol.* **80**, 289–93 (2014).

66. Oremland, R., Marsh, L. & Polcin, S. Methane production and simultaneous sulphate reduction in anoxic, salt marsh sediments. *Nature* **296**, 143–5 (1982).

67. King, G. Metabolism of trimethylamine, choline and glycine betaine by sulfate-reducing and methanogenic bacteria in marine sediments. *Appl. Environ. Microbiol.* **48**, 719–25 (1984).

68. Challenger, F. & Simpson, M. I. Studies on biological methylation. Part XII. A precursor of the dimethyl sulphide evolved by *Polysiphonia fastigiata*. dimethyl-2-carboxyethylsulphonium hydroxide and its salts. *J. Chem. Soc.* **3**, 1591–1597 (1948).

69. Bürgermeister, S., Zimmermann, R. L., Georgii, H.-W., Bingemer, H. G., Kirst, G. O., Janssen, M. & Ernst, W. On the biogenic origin of dimethylsulfide: Relation between chlorophyll, ATP, organismic DMSP, phytoplankton species, and DMS distribution in Atlantic surface water and atmosphere. *J. Geophys. Res.* **95**, 20607 (1990).

70. Turner, S. M., Malin, G., Liss, P. S., Harbour, D. S. & Holligan, P. M. The seasonal variation of dimethyl sulfide and dimethylsulfoniopropionate concentrations in nearshore waters. *Limnol. Oceanogr.* **33**, 364–375 (1988).

71. Sun, L., Curson, A. R. J., Todd, J. D. & Johnston, A. W. B. Diversity of DMSP transport in marine bacteria, revealed by genetic analyses. *Biogeochemistry* **110**, 121–130 (2012).

72. Pichereau, V., Pocard, J. A., Hamelin, J., Blanco, C. & Bernard, T. Differential effects of dimethylsulfoniopropionate, dimethylsulfonioacetate, and other S-methylated compounds on the growth of *Sinorhizobium meliloti* at low and high osmolarities. *Appl. Environ. Microbiol.* **64**, 1420–1429 (1998).

73. Ghoul, M., Minet, J., Bernard, T., Dupray, E. & Cormier, M. Marine macroalgae as a source for osmoprotection for *Escherichia coli*. *Microb. Ecol.* **30**, 171–181 (1995).

74. Yang, G., Li, C. & Sun, J. Influence of salinity and nitrogen content on production of dimethylsulfoniopropionate (DMSP) and dimethylsulfide (DMS) by *Skeletonema costatum*. *Chinese J. Oceanol. Limnol.* **29**, 378–386 (2011).

75. Stefels, J. Physiological aspects of the production and conversion of DMSP in marine algae and higher plants. *J. Sea Res.* **43**, 183–197 (2000).

76. Tripp, H. J., Kitner, J. B., Schwalbach, M. S., Dacey, J. W. H., Wilhelm, L. J. & Giovannoni, S. J. SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* **452**, 741–4 (2008).

77. Nishiguchi, M. K. & Somero, G. N. Temperature- and concentration-dependence of compatibility of the organic osmolyte β-dimethyl-sulfoniopropionate. *Cryobiology* **29**, 118–124 (1992).

78. Sunda, W., Kieber, D. J., Kiene, R. P. & Huntsman, S. An antioxidant function for DMSP and DMS in marine algae. *Nature* **418**, 317–20 (2002).

79. Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature* **326**, 655–661 (1987).

80. Quinn, P. K. & Bates, T. S. The case against climate regulation via oceanic phytoplankton sulphur emissions. *Nature* **480**, 51–56 (2011).

81. Wolfe, G. V, Steinke, M. & Kirst, G. O. Grazing-activated chemical defence in a unicellular marine alga. *Nature* **387**, 894–897 (1997).

82. Keller, M. D., Kiene, R. P., Matrai, P. a & Bellows, W. K. Production of glycine betaine and dimethylsulsopropionate in marine phytoplankton. II. N-limited chemostat cultures. *Mar. Biol.* **135**, 237–248 (1999).

83. Kempf, B. & Bremer, E. Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Arch. Microbiol.* **170**, 319–330 (1998).

84. Hanson, A. D., Rivoal, J., Paquet, L. & Cage, D. A. Biosynthesis of 3-dimethylsulfoniopropionate in *Wollastonia biflora* (L.) DC. Evidence that S-methylmethionine is an intermediate. *Plant Physiol.* **105**, 103–110 (1994).

85. Greene, C. Biosynthesis of dimethyl-beta-propiothetin. *J Biol Chem* **237**, 2251–2254 (1962).

86. Barnard, W. R., Andreae, M. O. & Iverson, R. L. Dimethylsulflde and Phaeocystis puchetii in the southeastern Bering Sea. *Cont. Shelf Res.* **3**, 103–113

(1984).

87. Yoch, D. C. Dimethylsulfoniopropionate : its sources , role in the marine food web , and biological degradation to dimethylsulfide. *Appl Env. Microbiol.* **68**, 5804–5815 (2002).

88. Raina, J.-B., Tapiolas, D. M., Forêt, S., Lutz, A., Abrego, D., Ceh, J., Seneca, F. O., Clode, P. L., Bourne, D. G., Willis, B. L. & Motti, C. a. DMSP biosynthesis by an animal and its role in coral thermal stress response. *Nature* **502**, 677–80 (2013).

89. Curson, A. R. J., Liu, J., Bermejo Martínez, A., Green, R. T., Chan, Y., Carrión, O., Williams, B. T., Zhang, S., Yang, G., Page, P. C. B., Zhang, X. & Todd, J. D. Dimethylsulfoniopropionate biosynthesis in marine bacteria and identification of the key gene in this process. *Nat. Microbiol.* **17009**, (2017).

90. Williams, B. T. *et al.* Bacteria are important dimethylsulfoniopropionate producers in coastal sediments. *Nat. Microbiol.* **4**, 1815–1825 (2019).

91. Kocsis, M. G., Nolte, K. D., Rhodes, D., Shen, T. L., Gage, D. a & Hanson, a D. Dimethylsulfoniopropionate biosynthesis in Spartina alterniflora1. Evidence that S-methylmethionine and dimethylsulfoniopropylamine are intermediates. *Plant Physiol.* **117**, 273–281 (1998).

92. James, F., Pacqut, L., Sparace, S. A., Gage, D. A. & Hanson, A. D. Evidence implicating dimethylsulfoniopropionaaldehyde as an intermediate in dimethylsulfoniopropionate biosynthesis. *Plant Physiol.* **108**, 1439–1448 (1995).

93. Rhodes, D., Gage, D. a., Cooper, A. & Hanson, a. D. S-Methylmethionine conversion to dimethylsulfoniopropionate: evidence for an unusual transamination reaction. *Plant Physiol.* **115**, 1541–1548 (1997).

94. Kocsis, M. G. & Hanson, A. D. Biochemical evidence for two novel enzymes in

the biosynthesis of 3-dimethylsulfoniopropionate in *Spartina alterniflora*. *Plant Physiol.* **123**, 1153–61 (2000).

95. Uchida, A., Ooguri, T., Ishida, T., Kitaguchi, H. & Ishida, Y. Biosynthesis of dimethylsulfoniopropionate in *Crypthecodinium cohnii* (Dinophyceae). in *Kiene, R.P., Visscher, P.T., Keller, M.D., Kirst, G.O. (Eds.). Biological and Environmental Chemistry of DMSP and Related Sulfonium Compounds, Plenum Press, New York,* 97–107 (1996).

96. Gage, D. A., Rhodes, D., Nolte, K. D., Hicks, W. A., Leustek, T., Cooper, A. J. & Hanson, A. D. A new route for synthesis of dimethylsulphoniopropionate in marine algae. *Nature* **387**, 891–894 (1997).

97. Summers, P. S., Nolte, K. D., Cooper, A. J. L., Borgeas, H., Leustek, T., Rhodes, D. & Hanson, A. D. Identification and stereospecificity of the first three enzymes of 3-dimethylsulfoniopropionate biosynthesis in a chlorophyte alga. *Plant Physiol.* **116**, 369–378 (1998).

98. Lyon, B. R., Lee, P. a., Bennett, J. M., DiTullio, G. R. & Janech, M. G. Proteomic analysis of a sea-ice diatom: salinity acclimation provides new insight into the dimethylsulfoniopropionate production pathway. *Plant Physiol.* **157**, 1926–1941 (2011).

99. Curson, A. R. J., Williams, B. T., Pinchbeck, B. J., Sims, L. P., Martínez, A. B., Rivera, P. P. L., Kumaresan, D., Mercadé, E., Spurgin, L. G., Carrión, O., Moxon, S., Cattolico, R. A., Kuzhiumparambil, U., Guagliardo, P., Clode, P. L., Raina, J. B. & Todd, J. D. DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton. *Nat. Microbiol.* **3**, 430–439 (2018).

100. Keller, M. D., Bellows, W. K. & Guillard, R. R. L. Dimethylsulfide production in marine phytoplankton. *Biog. Sulfur Environ.* 167–182 (1989).

101. Kageyama, H., Tanaka, Y., Shibata, A., Waditee-Sirisattha, R. & Takabe, T. Dimethylsulfoniopropionate biosynthesis in a diatom *Thalassiosira pseudonana*: Identification of a gene encoding MTHB-methyltransferase. *Arch. Biochem. Biophys.* **645**, 100–106 (2018).

102. Howard, E. C., Henriksen, J. R., Buchan, A., Reisch, C. R., Bürgmann, H., Welsh, R., Ye, W., González, J. M., Mace, K., Joye, S. B., Kiene, R. P., Whitman, W. B. & Moran, M. A. Bacterial taxa that limit sulfur flux from the ocean. *Science (80-.).* **314**, 649–652 (2006).

103. Howard, E. C., Sun, S., Biers, E. J. & Moran, M. A. Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environ. Microbiol.* **10**, 2397–2410 (2008).

104. Curson, A. R. J., Todd, J. D., Sullivan, M. J. & Johnston, A. W. B. Catabolism of dimethylsulphoniopropionate: microorganisms, enzymes and genes. *Nat. Rev. Microbiol.* **9**, 849–859 (2011).

105. Curson, A. R. J., Rogers, R., Todd, J. D., Brearley, C. A. & Johnston, A. W. B. Molecular genetic analysis of a dimethylsulfoniopropionate lyase that liberates the climate-changing gas dimethylsulfide in several marine alpha-proteobacteria and *Rhodobacter sphaeroides*. *Environ. Microbiol.* **10**, 757–767 (2008).

106. Kirkwood, M., Le Brun, N. E., Todd, J. D. & Johnston, A. W. B. The dddP gene of Roseovarius nubinhibens encodes a novel lyase that cleaves dimethylsulfoniopropionate into acrylate plus dimethyl sulfide. *Microbiology* **156**, 1900–1906 (2010).

107. Todd, J. D., Curson, A. R. J., Kirkwood, M., Sullivan, M. J., Green, R. T. & Johnston, A. W. B. DddQ, a novel, cupin-containing, dimethylsulfoniopropionate lyase in marine roseobacters and in uncultured

marine bacteria. *Environ. Microbiol.* **13**, 427–438 (2011).

108. Curson, A. R. J., Sullivan, M. J., Todd, J. D. & Johnston, A. W. B. DddY, a periplasmic dimethylsulfoniopropionate lyase found in taxonomically diverse species of Proteobacteria. *ISME J.* **5**, 1191–1200 (2011).

109. Todd, J. D., Rogers, R., Li, Y. G., Wexler, M., Bond, P. L., Sun, L., Curson, A. R. J., Malin, G., Steinke, M. & Johnston, A. W. B. Structural and regulatory genes required to make the gas dimethylsulfide in bacteria. *Science (80-.).* **315**, 666–669 (2007).

110. Alcolombri, U., Ben-Dor, S., Feldmesser, E., Levin, Y., Tawfik, D. S. & A., V. MARINE SULFUR CYCLE. Identification of the algal dimethyl sulfide-releasing enzyme: A missing link in the marine sulfur cycle. *Science (80-.).* **348**, 1466–1469 (2015).

111. Andreae, M. O. The ocean as a source of atmospheric sulfur compounds. The role of air–sea exchange in geochemical cycling. *P. Buat-Merard role air-sea Exch. geochemical cycling.(Reidel, New York)* 331–362 (1986).

112. Keller, M. D., Matrai, P. A., Kiene, R. P. & Bellows, W. K. Responses of coastal phytoplankton populations to nitrogen additions: dynamics of cell-associated dimethylsulfoniopropionate (DMSP), glycine betaine (GBT), and homarine. *Can. J. Fish. Aquat. Sci.* **61**, 685–699 (2004).

113. Keller, M. D., Kiene, R. P., Matrai, P. A. & Bellows, W. K. Production of glycine betaine and dimethylsulfoniopropionate in marine phytoplankton. II. N-limited chemostat cultures. *Mar. Biol.* **135**, 249–257 (1999).

114. Spielmeyer, A. & Pohnert, G. Influence of temperature and elevated carbon dioxide on the production of dimethylsulfoniopropionate and glycine betaine by marine phytoplankton. *Mar. Environ. Res.* **73**, 62–69 (2012).

115. Chepurnov, V. A. & Al, E. In search of new tractable diatoms for experimental biology. *Bioessays* **30**, 692–702 (2008).

116. Armbrust, E. V. The life of diatoms in the world's oceans. *Nature* **459**, 185–192 (2009).

117. Sarthou, G. Growth physiology and fate of diatoms in the ocean: a review. *J. Sea Res.* **53**, 25–42 (2005).

118. Mock, T. & Medlin, L. K. Chapter Seven-Genomics and Genetics of Diatoms. *Adv. Bot. Res.* **64**, 245–284 (2012).

119. De Martino, A., Bartual, A., Willis, A., Meichenin, A., Villazn, B., Maheswari, U. & Bowler, C. Physiological and molecular evidence that environmental changes elicit morphological interconversion in the model diatom Phaeodactylum tricornutum. *Protist* **162**, 462–481 (2011).

120. Khiyami, M. A., Al-ghamdi, A. K., Sabir, M. J., Romanovicz, K., Hajrah, N. H., Omri, A. El, Jansen, R. K. & Ashworth, P. Phylogenetic analysis and a review of the history of the accidental phytoplankter , *Phaeodactylum tricornutum* Bohlin (Bacillariophyta). *PLoS One* **13**, e0196744 (2018).

121. Armbrust, E. V. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science (80-.).* **306**, 79–86 (2004).

122. Bowler, C. *et al.* The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239–244 (2008).

123. De Riso, V., Raniello, R., Maumus, F., Rogato, A., Bowler, C. & Falciatore, A. Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res.* **37**, e96 (2009).

124. Siaut, M., Heijde, M., Mangogna, M., Montsant, A., Coesel, S., Allen, A., Manfredonia, A., Falciatore, A. & Bowler, C. Molecular toolbox for studying

diatom biology in *Phaeodactylum tricornutum*. *Gene* **406**, 23–35 (2007).

125. Poulsen, N. & Kröger, N. Silica morphogenesis by alternative processing of silaffins in the diatom *Thalassiosira pseudonana*. *J. Biol. Chem.* **279**, 42993–42999 (2004).

126. Hopes, A., Nekrasov, V., Kamoun, S. & Mock, T. Editing of the urease gene by CRISPR-Cas in the diatom *Thalassiosira pseudonana*. *Plant Methods* **12**, 49 (2016).

127. Stukenberg, D., Zauner, S., Dell'Aquila, G. & Maier, U. G. Optimizing CRISPR/cas9 for the diatom *Phaeodactylum tricornutum*. *Front. Plant Sci.* **9**, 740 (2018).

128. Hockin, N. L. A proteomic approach to metabolism in the diatom *Thalassiosiraa* pseudonana.(2011).

129. Remmers, I. M., D'Adamo, S., Martens, D. E., de Vos, R. C. H., Mumm, R., America, A. H. P., Cordewener, J. H. G., Bakker, L. V., Peters, S. A., Wijffels, R. H. & Lamers, P. P. Orchestration of transcriptome, proteome and metabolome in the diatom *Phaeodactylum tricornutum* during nitrogen limitation. *Algal Res.* **35**, 33–49 (2018).

130. Kageyama, H., Tanaka, Y. & Takabe, T. Biosynthetic pathways of glycinebetaine in Thalassiosira pseudonana; functional characterization of enzyme catalyzing three-step methylation of glycine. *Plant Physiol. Biochem.* **127**, 248–255 (2018).

131. Vila-Costa, M., Simó, R., Harada, H., Gasol, J. M., Slezak, D. & Kiene, R. P. Dimethylsulfoniopropionate uptake by marine phytoplankton. *Science* **314**, 652–654 (2006).

132. Klähn, S. & Hagemann, M. Compatible solute biosynthesis in cyanobacteria. *Environmental Microbiology* vol. 13 551–562 (2011).

133. Todd, J. D., Curson, A. R. J., Nikolaidou-Katsaraidou, N., Brearley, C. A., Watmough, N. J., Chan, Y., Page, P. C. B., Sun, L. & Johnston, A. W. B. Molecular dissection of bacterial acrylate catabolism - unexpected links with dimethylsulfoniopropionate catabolism and dimethyl sulfide production. *Environ. Microbiol.* **12**, 327–343 (2010).

134. Guillard, R. R. L. *Culture of marine invertebrate animals*. (Plenum Press, 1975).

135. Berges, J. A., Franklin, D. J. & Harrison, P. J. Evolution of an artificial seawater medium: improvements in enriched seawater, artificial water over the last two decades. *J. Phycol.* **37**, 1138–1145 (2001).

136. Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, 2001).

137. Beringer, J. E. R factor transfer in *Rhizobium leguminosarum*. *J. Gen. Microbiol.* **84**, 188–198 (1974).

138. Gonzalez, J. M., Whitman, W. B., Hodson, R. E. & Moran, M. A. Identifying numerically abundant culturable bacteria from complex communities: an example from a lignin enrichment culture. *Appl. Environ. Microbiol.* **62**, 4433–4440 (1996).

139. Baumann, P. & Baumann, L. *The Prokaryotes: A Handbook on Habitats, Isolation and Identification of Bacteria*. (Springer-Verlag, 1981).

140. Martino, A. De, Meichenin, A., Shi, J., Pan, K. & Bowler, C. Genetic and phenotypic characterization of *Phaeodactylum tricornutum* (Bacillariophyceae) accessions. *J. Phycol.* **43**, 992–1009 (2007).

141. Hasle, G. R. & Heimdal, B. R. Some species of the centric diatom genus *Thalassiosira* studied in the light and electron microscopes. *Diatomaceae II (J. Gerloff J.B. Cholnoky, eds.). Beihefte zur Nov. Hedwigia* **31: 559-58**, (1970).

142. Chung, E. J., Park, J. A., Jeon, C. O. & Chung, Y. R. *Gynuella sunshinyii* gen. Nov., Sp. Nov., an antifungal rhizobacterium isolated from a halophyte, carex scabrifolia steud. *Int. J. Syst. Evol. Microbiol.* **65**, 1038–1043 (2015).

143. Chen, Z., Zhang, J., Lei, X., Lai, Q., Yang, L., Zhang, H., Li, Y., Zheng, W., Tian, Y., Yu, Z., Xu, H. & Zheng, T. *Mameliella phaeodactyli* sp. nov., a member of the family rhodobacteraceae isolated from the marine algae *Phaeodactylum tricornutum*. *Int. J. Syst. Evol. Microbiol.* **65**, 1617–1621 (2015).

144. Studier, F. & Moffatt, B. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–30 (1986).

145. Wood, W. Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affecting the restriction and modification of DNA. *J. Mol. Biol.* **16**, 118–33 (1966).

146. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–19 (1985).

147. Haardt, M., Kempf, B., Faatz, E. & Bremer, E. The osmoprotectant proline betaine is a major substrate for the binding-protein-dependent transport system ProU of *Escherichia coli* K-12. *Mol. Gen. Genet.* **246**, 783–6 (1995).

148. Young, J. P. W. *et al.* The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome biolo* **7**, R34 (2006).

149. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. & Rozen, S. G. Primer3 – new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).

150. Brown, K. L., Twing, K. I. & Robertson, D. L. Unraveling the regulation of nitrogen assimilation in the marine diatom *Thalassiosira pseudonana*

(Bacillariophyceae): diurnal variations in transcript levels for five genes involved in nitrogen assimilation. *J. Phycol.* **45**, 413–426 (2009).

151. Allen, A. E., Laroche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P. J., Finazzi, G., Fernie, A. R. & Bowler, C. Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 10438–10443 (2008).

152. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).

153. Mahmood, T. & Yang, P. C. Western blot: technique, theory, and trouble shooting. *N. Am. J. Med. Sci.* **4**, 429–434 (2012).

154. Pérez-Cruz, C., Carrión, O., Delgado, L., Martinez, G. & López-Iglesias, C. Mercade, E. New type of outer membrane vesicle produced by the Gram-negative bacterium *Shewanella vesiculosa* M7T: implications for DNA content. *Appl. Environ. Microbiol.* **79**, 1874–81 (2013).

155. Beale, R. & Airs, R. Quantification of glycine betaine, choline and trimethylamine N-oxide in seawater particulates: Minimisation of seawater associated ion suppression. *Anal. Chim. Acta* **938**, 114–122 (2016).

156. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

157. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

158. Gruber, A., Rocap, G., Kroth, P. G., Armbrust, E. V. & Mock, T. Plastid proteome prediction for diatoms and other algae with secondary plastids of

the red lineage. *Plant J.* **81**, 519–528 (2015).

159. Gschloessl, B., Guermeur, Y. & Cock, J. M. HECTAR: A method to predict subcellular targeting in heterokonts. *BMC Bioinformatics* **9**, 1–13 (2008).

160. Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S. & Nelsen, H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).

161. Downie, J. A., Hombrecher, G., Ma, Q. S., Knight, C. D., Wells, B. & Johnson, A. W. B. Cloned nodulation genes of *Rhizobium leguminosarum* determine host range specificity. *Mol. Gen. Genet.* **190**, 359–365 (1983).

162. Tett, A. J., Rudder, S. J., Karunakaran, R. & Poole, P. S. Regulatable vectors for environmental gene expression in alphaproteobacteria. *Appl. Environ. Microbiol.* **78**, 7137–7140 (2012).

163. Figurski, D. H. & Helinski, D. . Replication of an origin-containing derivative of plasmid RK2 dependant on a plasmid function provided in trans. *Proc. U.S. Nat. Acad. Sci* **76**, 1648–1652 (1979).

164. Norrandera, J., Kempe, T. & Messinga, J. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* **26**, 101–106 (1983).

165. Alverson, A. J., Beszteri, B., Julius, M. L. & Theriot, E. C. The model marine diatom *Thalassiosira pseudonana* likely descended from a freshwater ancestor in the genus Cyclotella. *BMC Evol. Biol.* **11**, 125 (2011).

166. Belcher, J. H. & Swale, E. M. F. Species of *Thalassiosira* (Diatoms, Bacillariophyceae) in the plankton of English rivers. *Br. Phycol. J.* **12**, 291–296 (1977).

167. Guillard, R. R. L. & Ryther, J. H. Studies of marine planktonic diatoms: I.

Cyclotella nana Hustedt, and *Detonula confervacea* (Cleve) Gran. *Can. J. Microbiol.* **8**, 229–239 (1962).

168. Maloney, A. E., Shinneman, A. L. C., Hemeon, K. & Sachs, J. P. Exploring lipid 2H/1H fractionation mechanisms in response to salinity with continuous cultures of the diatom *Thalassiosira pseudonana*. *Org. Geochem.* **101**, 154–165 (2016).

169. Belcher, J. H. & Swale, E. M. F. Species of *Thalassiosira* (Diatoms, Bacillariophyceae) in the plankton of English rivers. *Br. Phycol. J.* **12**, 291–296 (1977).

170. Guillard, R. R. L. & Myklestad, S. Osmotic and ionic requirements of the marine centric diatom *Cyclotella nana*. *Helgolander Wiss. Meeresunters* **20**, 104–110 (1970).

171. Baek, S. H., Jung, S. W. & Shin, K. Effects of temperature and salinity on growth of *Thalassiosira pseudonana* (Bacillariophyceae) isolated from ballast water. *J. Freshw. Ecol.* **26**, 547–552 (2011).

172. Brand, L. E. The salinity tolerance of Forty-six marine phytoplankton isolates. *Estuar. Coast. Shelf Sci.* **18**, 543–556 (1984).

173. La Vars, S. M., Johnston, M. R., Hayles, J., Gascooke, J. R., Brown, M. H., Leterme, S. C. & Ellis, A. V. 29Si{1H} CP-MAS NMR comparison and ATR-FTIR spectroscopic analysis of the diatoms *Chaetoceros muelleri* and *Thalassiosira pseudonana* grown at different salinities. *Anal. Bioanal. Chem.* **405**, 3359–3365 (2013).

174. Kettles, N. L., Kopriva, S. & Malin, G. Insights into the regulation of DMSP synthesis in the diatom *Thalassiosira pseudonana* through APR activity, proteomics and gene expression analyses on cells acclimating to changes in salinity, light and nitrogen. *PLoS One* **9**, e94795 (2014).

175. Bromke, M. A., Giavalisco, P., Willmitzer, L. & Hesse, H. Metabolic analysis of adaptation to short-term changes in culture conditions of the marine diatom *Thalassiosira pseudonana*. *PLoS One* **8**, 1–11 (2013).

176. De Martino, A., Bartual, A., Willis, A., Meichenin, A., Villazán, B., Maheswari, U. & Bowler, C. Physiological and molecular evidence that environmental changes elicit morphological interconversion in the model diatom *Phaeodactylum tricornutum*. *Protist* **162**, 462–481 (2011).

177. Martino, A. De, Meichenin, A., Shi, J., Pan, K. & Bowler, C. Genetic and phenotypic characterization of *Phaeodactylum tricornutum* (Bacillariophyceae) accessions. *J. Phycol.* **43**, 992–1009 (2007).

178. Li, W., Gao, K. & Beardall, J. Interactive effects of ocean acidification and nitrogen-limitation on the diatom *Phaeodactylum tricornutum*. *PLoS One* **7**, e51590 (2012).

179. DICKSON, D. M. J. & KIRST, G. O. Osmotic adjustment in marine eukaryotic algae: the role of inorganic ions, quaternary ammonium, tertiary sulphonium and carbohydrate solutes. I. Diatoms and a Rhodophyte. *New Phytol.* **106**, 645–655 (1987).

180. Osborn, H. L. & Hook, S. E. Using transcriptomic profiles in the diatom *Phaeodactylum tricornutum* to identify and prioritize stressors. *Aquat. Toxicol.* **138–139**, 12–25 (2013).

181. Hockin, N. L., Mock, T., Mulholland, F., Kopriva, S. & Malin, G. The response of diatom central carbon metabolism to nitrogen starvation is different from that of green algae and higher plants. *Plant Physiol.* **158**, 299–312 (2012).

182. Bullock, H. A., Luo, H. & Whitman, W. B. Evolution of dimethylsulfoniopropionate metabolism in marine phytoplankton and bacteria. *Front. Microbiol.* **8**, 637 (2017).

183. Abdullahi, A. S., Underwood, G. J. C. & Gretz, M. R. Extracellular matrix assembly in diatoms (Bacillariophyceae). V. Environmental effects on polysaccharide synthesis in the model diatom *Phaeodactylum tricornutum*. *J. Phycol.* **42**, 363–378 (2006).

184. Burrows, E. H., Bennette, N. B., Carrieri, D., Dixon, J. L., Brinker, A., Frada, M., Baldassano, S. N., Falkowski, P. G. & Dismukes, G. C. Dynamics of lipid biosynthesis and redistribution in the marine diatom *Phaeodactylum tricornutum* under nitrate deprivation. *BioEnergy Res.* **5**, 876–885 (2012).

185. Yang, Z.-K., Ma, Y.-H., Zheng, J.-W., Yang, W.-D., Liu, J.-S. & Li, H.-Y. Proteomics to reveal metabolic network shifts towards lipid accumulation following nitrogen deprivation in the diatom *Phaeodactylum tricornutum*. *J. Appl. Phycol.* **26**, 73–82 (2014).

186. Popko, J., Herrfurth, C., Feussner, K., Ischebeck, T., Iven, T., Haslam, R., Hamilton, M., Sayanova, O., Napier, J., Khozin-Goldberg, I. & Feussner, I. Metabolome analysis reveals betaine lipids as major source for triglyceride formation, and the accumulation of sedoheptulose during nitrogen-starvation of *Phaeodactylum tricornutum*. *PLoS One* **11**, e0164673 (2016).

187. Berges, J., Varela, D. & Harrison, P. Effects of temperature on growth rate, cell composition and nitrogen metabolism in the marine diatom *Thalassiosira pseudonana* (Bacillariophyceae). *Mar. Ecol. Prog. Ser.* **225**, 139–146 (2002).

188. Mulholland, M. R., Thomas, M. K., Kudela, R. M., Passow, U., Hayashi, K., Armstrong, E. A., Rynearson, T. A., Fu, F., Hutchins, D. A., Litchman, E., Hu, Z., Boyd, P. W., Yu, E., Strzepek, R. F. & Whittaker, K. A. Marine phytoplankton temperature versus growth responses from polar to tropical waters – outcome of a scientific community-wide study. *PLoS One* **8**, e63091 (2013).

189. Sigaud, T. C. S. & Aidar, E. Salinity and temperature effects on the growth and chlorophyll-α content of some planktonic aigae. *Bol. do Inst. Ocean.* **41**, 95–103 (1993).

190. Fawley, M. W. Effects of light intensity and temperature interactions on growth characteristics of *Phaeodactylum tricornutum* (Bacillariophyceae). *J. Phycol.* **20**, 67–72 (1984).

191. Domingues, N., Matos, A. R., Marques da Silva, J. & Cartaxana, P. Response of the diatom *Phaeodactylum tricornutum* to photooxidative stress resulting from high light exposure. *PLoS One* **7**, e38162 (2012).

192. Styron, C. E., Hagan, T. M., Campbell, D. R., Harvin, J., Whittenburg, N. K., Baughman, G. A., Bransford, M. E., Saunders, W. H., Williams, D. C., Woodle, C., Dixon, N. K. & McNeill, C. R. Effects of temperature and salinity on growth and uptake of 65Zn and 137Cs for six marine algae. *J. Mar. Biol. Assoc. United Kingdom* **56**, 13 (1976).

193. Maxwell, K. & Johnson, G. Chlorophyll fluorescence – a practical guide. *J Exp Bot* **51**, 659–668 (2000).

194. Kuczynska, P., Jemiola-Rzeminska, M. & Strzalka, K. Photosynthetic pigments in diatoms. *Mar. Drugs* **13**, 5847–5881 (2015).

195. Longworth, J., Wu, D., Huete-Ortega, M., Wright, P. C. & Vaidyanathan, S. Proteome response of *Phaeodactylum tricornutum*, during lipid accumulation induced by nitrogen depletion. *Algal Res.* **18**, 213–224 (2016).

196. Hatton, A. D. & Wilson, S. T. Particulate dimethylsulphoxide and dimethylsulphoniopropionate in phytoplankton cultures and Scottish coastal waters. *Aquat. Sci.* **69**, 330–340 (2007).

197. Mapson, B. L. W., March, J. F. & Wardale, D. A. Biosynthesis of Ethylene.

Biochem. J. **115**, 653–661 (1969).

198. Bhagavan, N. V. & Ha, C.-E. Protein and amino acid metabolism. in *Essentials of Medical Biochemistry* (2015).

199. Campbell, L. L. Transamination of amino acids with glyoxylic acid in bacterial extracts. *J. Bacteriol.* **71**, 81–83 (1955).

200. Schuster, J., Knill, T., Reichelt, M., Gershenzon, J. & Binder, S. Branched-chain aminotransferase4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in *Arabidopsis*. *Plant Cell* **18**, 2664–2679 (2006).

201. Chapple, C. C. S., Glover, J. R. & Ellis, B. E. Purification and characterization of methionine:glyoxylate aminotransferase from *Brassica carinata* and *Brassica napus*. *Plant Physiol.* **94**, 1887–1896 (2008).

202. Dolzan, M., Johansson, K., Tegoni, M., Schneider, G. & Cambillau, C. Crystal structure and reactivity of YbdL from *Escherichia coli* identify a methionine aminotransferase function. *FEBS Lett.* **571**, 141–146 (2004).

203. Wang, S., Yao, Q., Tao, J., Qiao, Y. & Zhang, Z. Co-ordinate expression of glycine betaine synthesis genes linked by the FMDV 2A region in a single open reading frame in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* **77**, 891–899 (2007).

204. Rendina, G. & Singer, T. P. Studies on choline dehydrogenase: I. Extraction in soluble form, assay, and some properties of the enzyme. *J. Biol. Chem.* **234**, 1605–1610 (1959).

205. Andresen, P. A., Kaaasen, I., Styrvold, O. B., Boulnois, G. & Strom, A. S. Molecular cloning , physical mapping and expression of the bet genes governing the osmoregulatory choline-glycine betaine pathway of *Escherichia coli*. *J. Gen. Microbiol.* **134**, 1737–1746 (1988).

206. Gadda, G. Kinetic mechanism of choline oxidase from *Arthrobacter globiformis*. *Biochim. Biophys. Acta - Proteins Proteomics* **1646**, 112–118 (2003).

207. Lavoie, M., Waller, J. C., Kiene, R. P. & Levasseur, M. Polar marine diatoms likely take up a small fraction of dissolved dimethylsulfoniopropionate relative to bacteria in oligotrophic environments. *Aquat Microb Ecol* **81**, 213–218 (2018).

208. Hudson, O., Singh, B. K., Leustek, T. & Gilvarg, C. An LL-Diaminopimelate aminotransferase defines a novel variant of the lysine biosynthesis pathway in plants. *Plant Physiol.* **140**, 292–301 (2006).

209. Lee, Y. R., Lin, T. S., Lai, S. J., Liu, M. Sen, Lai, M. C. & Chan, N. L. Structural analysis of glycine sarcosine N-methyltransferase from *Methanohalophilus portucalensis* reveals mechanistic insights into the regulation of methyltransferase activity. *Sci. Rep.* **6**, e38071 (2016).

210. Cairney, J., Booth, I. R. & Higgins, C. F. Osmoregulation of gene expression in *Salmonella typhimurium*: *proU* encodes an osmotically induced betaine transport system. *J. Bacteriol.* **164**, 1224–1232 (1985).

211. Nyyssola, A. & Leisola, M. *Actinopolyspora halophila* has two separate pathways for betaine synthesis. **176**, 294–300 (2001).

212. Vastermark, A., Wollwage, S., Houle, M. E., Rio, R. & Saier, M. H. Expansion of the APC superfamily of secondary carriers. *Proteins Struct. Funct. Bioinforma.* **82**, 2797–2811 (2014).

213. Kappes, R. M., Kempf, B. & Bremer, E. Three transport systems for the osmoprotectant glycine betaine operate in *Bacillus subtilis:* characterization of OpuD. *Journal of Bacteriology* vol. 178 (1996).

214. Ziegler, C. M., Bremer, E. & Krämer, R. The BCCT-family of carriers: from

physiology to crystal structure. *Mol Microbiol.* **78**, 13–34 (2010).

215. Jung, H., Buchholz, M., Clausen, J., Nietschke, M., Revermann, A., Schmid, R. & Jung, K. CaiT of *Escherichia coli*, a new transporter catalyzing L-carnitine/γ-butyrobetaine exchange. *J. Biol. Chem.* **277**, 39251–39258 (2002).

216. Ressl, S., Terwisscha Van Scheltinga, A. C., Vonrhein, C., Ott, V. & Ziegler, C. Molecular basis of transport and regulation in the Na+/betaine symporter BetP. *Nature* **458**, 47–52 (2009).

217. Perez, C., Koshy, C., Ressl, S., Nicklisch, S., Krä Mer, R. & Ziegler, C. Substrate specificity and ion coupling in the Na+/betaine symporter BetP. *EMBO J.* **30**, 1221-1229 (2011).

218. Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* 211, 969-970 (1966).

219. Raghukumar, S. Ecology of the marine protists, the Labyrinthulomycetes (Thraustochytrids and Labyrinthulids). *Eur. J. Protistol.* **38**, 127–145 (2002).

220. Nutahara, E., Abe, E., Uno, S., Ishibashi, Y., Watanabe, T., Hayashi, M., Okino, N. & Ito Id, M. The glycerol-3-phosphate acyltransferase PLAT2 functions in the generation of DHA-rich glycerolipids in *Aurantiochytrium limacinum* F26-b. *PLoS One* **14**, e0211164 (2019).

221. Aranda, M., Li, Y., Liew, Y. J., Baumgarten, S., Simakov, O., Wilson, M. C., Piel, J., Ashoor, H., Bougouffa, S., Bajic, V. B., Ryu, T., Ravasi, T., Bayer, T., Micklem, G., Kim, H., Bhak, J., Lajeunesse, T. C. & Voolstra, C. R. Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Sci. Rep.* **6**, e39734 (2016).

222. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS One*

6, e16765 (2011).

223. Nielsen, B. L., Willis, V. C. & Lin, C. Western blot analysis to illustrate relative control levels of the lac and ara promoters in *Escherichia coli*. **35**, 133–137 (2007).

224. Maheswari, U. *et al.* Digital expression profiling of novel diatom transcripts provides insight into their biological functions. *Genome Biol.* **11**, R85 (2010).

225. Werner, A. K., Romeis, T. & Witte, C. Ureide catabolism in *Arabidopsis thaliana* and *Escherichia coli*. *Nat. Chem. Biol.* **6**, 19–21 (2010).

226. Petrossian, T. C. & Clarke, S. G. Multiple motif scanning to identify methyltransferases from the yeast proteome. *Mol. Cell. Proteomics* **8**, 1516–1526 (2009).

227. Borowitzka, M. A. & Volcani, B. E. The polymorphic diatom *Phaeodactylum tricornutum*: ultrastructure of its morphotypes. *J. Phycol.* **14**, 10–21 (1978).

228. Trossat, C., Rathinasabapathi, B., Weretilnyk, E. A., Shen, T.-L., Huang, Z.-H., Gage, D. A. & Hanson, A. D. Salinity promotes accumulation of 3-dimethylsulfoniopropionate and its precursor *S* -methylmethionine in chloroplasts. *Plant Physiol.* **116**, 165–171 (1998).

229. Percudani, R., Carnevali, D. & Puggioni, V. Ureidoglycolate hydrolase , amidohydrolase , lyase : how errors in biological databases are incorporated in scientific papers and vice versa. **2013**, 1–9 (2013).

230. Gage, D. A., Rhodes, D., Nolte, K. D., Hicks, W. A., Leustek, T., Cooper, A. J. L. & Hanson, A. D. A new route for synthesis of dimethylsulphoniopropionate in marine algae. *Nature* **387**, 891–894 (1997).

231. Lai, S. J. & Lai, M. C. Characterization and regulation of the osmolyte betaine synthesizing enzymes GSMT and SDMT from halophilic methanogen *Methanohalophilus portucalensis*. *PLoS One* **6**, e25090 (2011).

232. Nyyssölä, A., Reinikainen, T. & Leisola, M. Characterization of glycine sarcosine N-methyltransferase and sarcosine dimethylglycine N-methyltransferase. *Appl. Environ. Microbiol.* **67**, 2044–2050 (2001).

233. Von Weymarn, N., Nyyssölä, a., Reinikainen, T., Leisola, M. & Ojamo, H. Improved osmotolerance of recombinant *Escherichia coli* by de novo glycine betaine biosynthesis. *Appl. Microbiol. Biotechnol.* **55**, 214–218 (2001).

234. Nyyssölä, A., Kerovuo, J., Kaukinen, P., Von Weymarn, N. & Reinikainen, T. Extreme halophiles synthesize betaine from glycine by methylation. *J. Biol. Chem.* **275**, 22196–22201 (2000).

235. McCoy, J. G., Bailey, L. J., Ng, Y. H., Bingman, C. A., Wrobel, R., Weber, A. P. M., Fox, B. G. & Phillips, G. N. Discovery of sarcosine dimethylglycine methyltransferase from *Galdieria sulphuraria*. *Proteins Struct. Funct. Bioinforma.* **74**, 368–377 (2009).

236. Lu, W. D., Chi, Z. M. & Su, C. D. Identification of glycine betaine as compatible solute in *Synechococcus* sp. WH8102 and characterization of its N-methyltransferase genes involved in betaine synthesis. *Arch. Microbiol.* **186**, 495–506 (2006).

237. Lai, M. C., Wang, C. C., Chuang, M. J., Wu, Y. C. & Lee, Y. C. Effects of substrate and potassium on the betaine-synthesizing enzyme glycine sarcosine dimethylglycine N-methyltransferase from a halophilic methanoarchaeon Methanohalophilus portucalensis. *Res. Microbiol.* **157**, 948–955 (2006).

238. Kimura, Y., Kawasaki, S., Yoshimoto, H. & Takegawa, K. Glycine betaine biosynthesized from glycine provides an osmolyte for cell growth and spore germination during osmotic stress in *Myxococcus xanthus*. *J. Bacteriol.* **192**, 1467–1470 (2010).

239. Waditee, R., Tanaka, Y., Aoki, K., Hibino, T., Jikuya, H., Takano, J., Takabe, T.

& Takabe, T. Isolation and functional characterization of N-methyltransferases that catalyze betaine synthesis from glycine in a halotolerant photosynthetic organism *Aphanothece halophytica*. *J. Biol. Chem.* **278**, 4932–4942 (2003).

240. Sévin, D. C., Stählin, J. N., Pollak, G. R., Kuehne, A. & Sauer, U. Global metabolic responses to salt stress in fifteen species. *PLoS One* **11**, e0148888 (2016).

241. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

242. Simó, R. & Dachs, J. Global ocean emission of dimethylsulfide predicted from biogeophysical data. *Global Biogeochem. Cycles* **16**, 26-1-26–10 (2002).

243. Keeling, P. J. et al. The marine microbial eukaryote transcriptome sequencing project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **12**, e1001889 (2014).

244. Toribio, A. L. et al. European nucleotide archive in 2016. *Nucleic Acids Res.* **45**, 32–36 (2017).

245. Raina, J.-B., Clode, P. L., Cheong, S., Bougoure, J., Kilburn, M. R., Reeder, A., Forêt, S., Stat, M., Beltran, V., Thomas-Hall, P., Tapiolas, D., Motti, C. M., Gong, B., Pernice, M., Marjo, C. E., Seymour, J. R., Willis, B. L. & Bourne, D. G. Subcellular tracking reveals the location of dimethylsulfoniopropionate in microalgae and visualises its uptake by marine bacteria. *Elife* **6**, e23008 (2017).

Appendices

Appendices

Identification of genes of interest in the model diatoms *T. pseudonana* and *P. tricornutum.*

Genes of interest involved in DMSP and GBT synthesis and transport.

Candidate genes for the DMSP and GBT synthesis as well as candidate transporters for those compatible solutes were identified in *T. pseudonana.*

Appendix 1. List of candidate genes for the synthesis and transport of DMSP and GBT in *T. pseudonana.*

Gene ID	Type of enzyme	Accession number
DMSP synthesis		
THAPSDRAFT_260934	Aminotransferase	XP_002285992.1
THAPSDRAFT_27811	Aminotransferase	XP_002289692.1
THAPSDRAFT_11247	SAM dependent carboxyl methyltransferase	XP_002296978.1
THAPSDRAFT_269095[(a)]	SAM dependent carboxyl methyltransferase	XP_002291473.1
THAPSDRAFT_20613	Decarboxylase	XP_002286584.1
GBT synthesis		
THAPSDRAFT_41650	Choline dehydrogenase	XP_002292146.1
THAPSDRAFT_20797[(b)]	SAM dependent carboxyl methyltransferase	XP_002286764.1
DMSP and GBT transport		
THAPSDRAFT_262307	BCCT	XP_002289511.1
THAPSDRAFT_268228	PROP	XP_002287760.1

[(a)] Functional DSYD (**D**MSHB **sy**nthase in **D**iatoms), involved in DMSP synthesis via the transamination pathway. It adds a methyl group from SAM to MTHB.

[(b)] Functional GSDMT (**g**lycine **s**arcosine **d**imethylglycine **m**ethyl**t**ransferase), involved in GBT synthesis via the methylation pathway. It adds methyl groups to glycine, sarcosine and dimethylglycine using SAM as methyl donor.

Appendix 2. List of candidate genes for the synthesis and transport of DMSP and GBT in *P. tricornutum*.

Gene ID	Type of enzyme	Accession number
DMSP synthesis		
PHATRDRAFT_22909	Aminotransferase	XP_002183711.1
PHATRDRAFT_44630	NADP- NADH- dependent enzyme	XP_002178364.1
PHATRDRAFT_48704[(a)]	SAM dependent carboxyl methyltransferase	XP_002183266.1
PHATRDRAFT_21592	Decarboxylase	XP_002181752.1
GBT synthesis		
PHATRDRAFT_1341	Choline dehydrogenase	XP_002178042.1
PHATRDRAFT_20301[(b)]	SAM dependent carboxyl methyltransferase	XP_002180089.1
DMSP and GBT transport		
PHATRDRAFT_48315	BCCT	XP_002182771.1

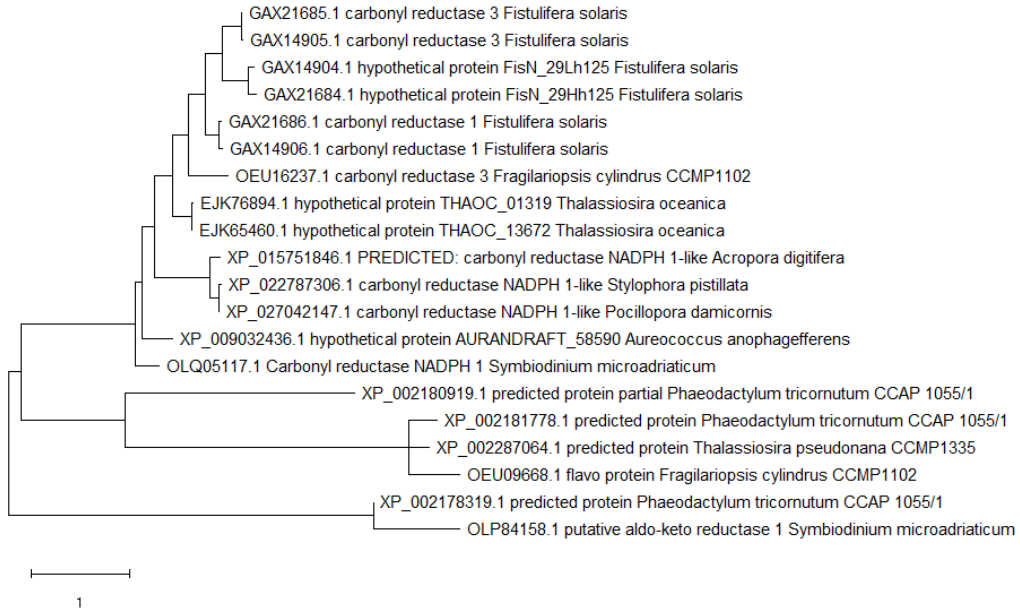
[(a)] Functional DSYD (**D**MSHB **sy**nthase in **D**iatoms), involved in DMSP synthesis via the transamination pathway. It adds a methyl group from SAM to MTHB.

[(b)] Functional GSDMT (**g**lycine **s**arcosine **d**imethylglycine **m**ethyl**t**ransferase), involved in GBT synthesis via the methylation pathway. It adds methyl groups to glycine, sarcosine and dimethylglycine using SAM as methyl donor.

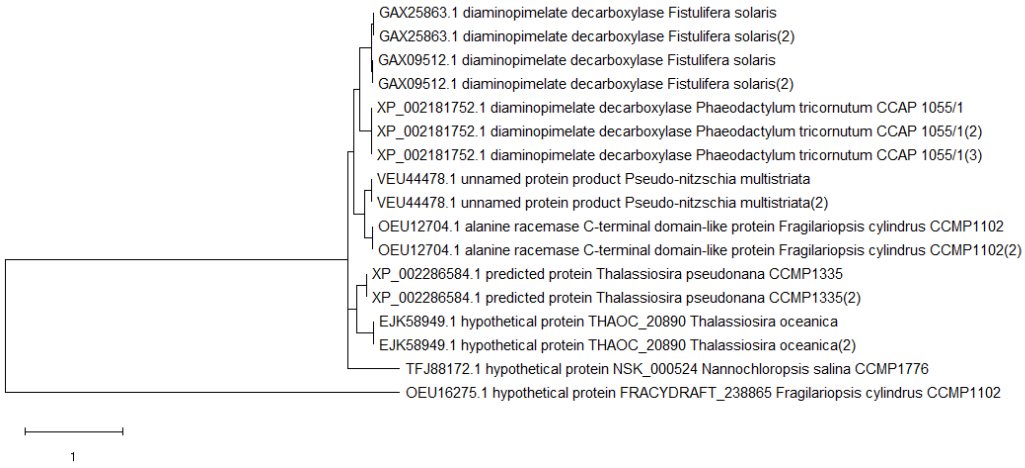
Relatedness between genes of interest involved in DMSP and GBT synthesis and transport and their closest homologues.

Appendix 3. Maximum likelihood tree of candidate aminotransferases in *T. pseudonana* and *P. tricornutum* and their closest homologues. Probe sequences are marked with an

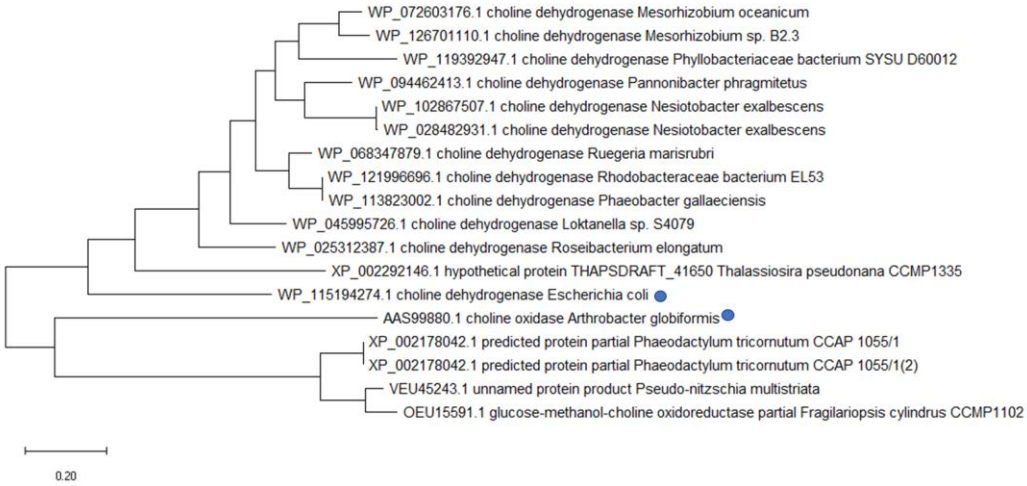
Appendix 4. Maximum likelihood tree of candidate reductase in *T. pseudonana* and *P. tricornutum* and their closest homologues.



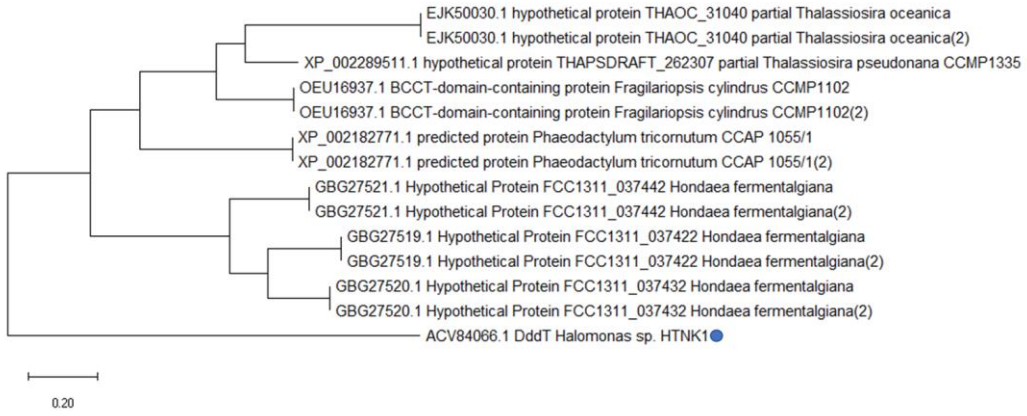
Appendix 5. Maximum likelihood tree of candidate decarboxylase in *T. pseudonana* and *P. tricornutum* and their closest homologues.



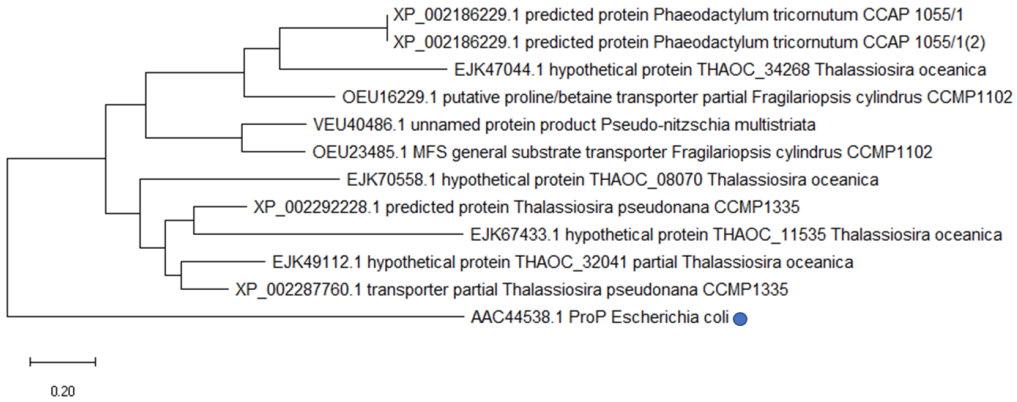
Appendix 6. Maximum likelihood tree of candidate choline dehydrogenase in *T. pseudonana* and *P. tricornutum* and their closest homologues. Probe sequences are marked with an



Appendix 7. Maximum likelihood tree of candidate BCCT-like proteins transporter in *T. pseudonana* and *P. tricornutum* and their closest homologues. Probe sequence are marked with an



Appendix 8. Maximum likelihood tree of candidate ProP-like proteins in *T. pseudonana* and *P. tricornutum* and their closest homologues. Probe sequence are marked with an



Identification of the most regulated proteins in *P. tricornutum* through whole proteome sequencing

Axenic cultures of *P. tricornutum* CCAP 1055/1 were grown in F/2 medium mixed with artificial seawater in standard salinity (35 PSU) or low salinity (5 PSU). Samples were taken in mid-exponential growth phase (Day 4), proteins were extracted and the whole proteome was sequenced and analysed.

Appendix 9. List of top 50 proteins downregulated in low salinity (5 PSU) compared to normal salinity (35 PSU) in *P. tricornutum* extracted from whole proteome sequencing performed in triplicates. Regulation is expressed as fold change in log2. All the results in this table are statistically significant (q ≤ 0.05), q-value was calculated Student's T-test.

Protein name	Pfam	Accession number	FOLD in log2
Predicted protein	CaMKII_AD; SnoaL_2	XP_002183267.1	-6.10
Flavodoxin	Flavodoxin_1	XP_002184899.1	-5.22
Predicted protein		XP_002186410.1	-4.50
Predicted protein		XP_002181775.1	-3.41
Predicted protein		XP_002178308.1	-3.37
Predicted protein		XP_002178510.1	-3.33
Predicted protein	HSP20	XP_002180814.1	-3.21
Predicted protein	HAD_2	XP_002178073.1	-3.14
Predicted protein	Aldo_ket_red	XP_002178319.1	-3.10
Predicted protein		XP_002178980.1	-3.09

Predicted protein		XP_002177414.1	-2.93
Frustulin 3 (Fragment)		XP_002179415.1	-2.80
Predicted protein		XP_002179461.1	-2.70
Predicted protein	Acylphosphatase	XP_002179037.1	-2.68
Predicted protein		XP_002177257.1	-2.65
Protein fucoxanthin chlorophyll a/c protein	Chloroa_b-bind	XP_002184869.1	-2.65
Predicted protein	His_Phos_1	XP_002180492.1	-2.59
Iron starvation induced protein		XP_002183742.1	-2.59
Predicted protein	PDZ	XP_002180353.1	-2.59
Predicted protein	SnoaL_2	XP_002179645.1	-2.52
Peptidylprolyl isomerase (Fragment)	FKBP_C	XP_002180251.1	-2.45
Predicted protein [(a)]	Methyltransf_11	XP_002183266.1	-2.43
Predicted protein	Pirin_C	XP_002176807.1	-2.43
Predicted protein		XP_002183924.1	-2.36
Predicted protein		XP_002183627.1	-2.36
Proteasome endopeptidase complex	Proteasome_A_N	XP_002184677.1	-2.34
Predicted protein		XP_002177138.1	-2.30
Predicted protein		XP_002177882.1	-2.27
Predicted protein	EF-hand_7	XP_002183345.1	-2.27
Predicted protein	PsbP	XP_002182797.1	-2.25
Predicted protein	FAD_binding_6	XP_002179706.1	-2.24
Predicted protein	LSM	XP_002182258.1	-2.23
Predicted protein	PDZ	XP_002182905.1	-2.19
High light induced protein 2		XP_002184741.1	-2.18
Predicted protein	adh_short	XP_002178364.1	-2.16
Predicted protein	adh_short	XP_002179448.1	-2.16
Predicted protein	PITH	XP_002176920.1	-2.16
Predicted protein		XP_002183916.1	-2.14
GrpE protein homolog (Fragment)	GrpE	XP_002182951.1	-2.14
Heat shock protein Hsp20	HSP20	XP_002179810.1	-2.10
Predicted protein		XP_002184228.1	-2.09
Predicted protein	Proteasome	XP_002185913.1	-2.08
Predicted protein	PA28_beta	XP_002182205.1	-2.07
Predicted protein		XP_002179139.1	-2.06
Photosystem II reaction center protein L	PsbL	YP_874371.1	-2.05
Predicted protein		XP_002181087.1	-2.03
Predicted protein	UQ_con	XP_002180700.1	-2.02

Glutathione peroxidase (Fragment)	GSHPx	XP_002180739.1	-2.00
Pyrroline-5-carboxylate reductase	F420_oxidored, P5CR_dimer	XP_002177719.1	-1.99
Predicted protein	DUF1995	XP_002178583.1	-1.99

[(a)] Functional DSYD in *P. tricornutum* downregulated -2.43-fold in log2 labelled as predicted protein and Pfam Methyltransf_11 (XP_002183266.1).

Appendix 10. List of top 50 proteins upregulated in low salinity (5 PSU) compared to normal salinity (35 PSU) in *P. tricornutum* extracted from whole proteome sequencing performed in triplicates. Regulation is expressed as fold change in log2. All the results in this table are statistically significant (q ≤ 0.05), q-value was calculated Student's T-test.

Protein name	Pfam	Accession number	FOLD in log2
Predicted protein	PEPcase	XP_002182829.1	4.79
Predicted protein	HCO3_cotransp	XP_002176305.1	4.73
Predicted protein	CPSase_L_chain; CPSase_sm_chain; GATase; MGS	XP_002183539.1	4.17
Alanine--tRNA ligase	DHHA1; tRNA_SAD; tRNA-synt_2c	XP_002182439.1	4.06
Eukaryotic translation initiation factor 3 subunit A	PCI	XP_002180731.1	4.04
Predicted protein	PEPcase	XP_002181027.1	3.89
Predicted protein (Fragment)		XP_002183955.1	3.87
Biotin carboxylase	ACC_central;Biotin_carb_C;Biotin_lipoyl;Carboxyl_trans;CPSase_L_chain; CPSase_L_D2	XP_002185458.1	3.85
ATP-citrate synthase	ATP-grasp_2;Citrate_synt;CoA_binding;Ligase_CoA	XP_002180032.1	3.70
Predicted protein		XP_002177010.1	3.67
Predicted protein	ClpB_D2-small	XP_002186189.1	3.60
Nitrate reductase	Cyt-b5; FAD_binding_6; Mo-co_dimer;NAD_binding_1;Oxidored_molyb	XP_002183599.1	3.53
Predicted protein	HGTP_anticodon; ProRS-C_1;tRNA-synt_2b	XP_002184639.1	3.34
Transmembrane 9 superfamily member	EMP70	XP_002186115.1	3.30
Predicted protein		XP_002184422.1	3.28
Predicted protein	DUF389	XP_002183180.1	3.23
Predicted protein	Anticodon_1; tRNA-synt_1	XP_002178570.1	3.16
Predicted protein	GATase_2;Glu_syn_central;Glu_synthase;GXGXG	XP_002176769.1	3.14
Predicted protein	AAA	XP_002178361.1	3.13
Predicted protein	AAA	XP_002186016.1	3.08

Predicted protein	Amino_oxidase	XP_002180966.1	3.07
Predicted protein	EFG_C;GTP_EFTU;GTP_EFTU_D2	XP_002184487.1	3.07
Predicted protein	Dynamin_M;GED	XP_002181636.1	3.02
Predicted protein	Adaptin_N;B2-adapt-app_C	XP_002176429.1	3.01
Eukaryotic translation initiation factor 3 subunit B	eIF2A;RRM_1	XP_002179558.1	3.01
Predicted protein	DUF3419	XP_002176772.1	3.00
Predicted protein	AAA;CDC48_N	XP_002181088.1	2.96
DNA-directed RNA polymerase subunit beta	RNA_pol_Rpb2_1;RNA_pol_Rpb2_2;RNA_pol_Rpb2_3;RNA_pol_Rpb2_4;RNA_pol_Rpb2_5;RNA_pol_Rpb2_6;RNA_pol_Rpb2_7	XP_002178926.1	2.96
Predicted protein	APS_kinase;ATP-sulfurylase;PUA_2;Pyrophosphatase	XP_002179808.1	2.94
Predicted protein (Fragment)	DnaJ;DnaJ-X	XP_002183103.1	2.92
Predicted protein (Fragment)		XP_002177362.1	2.89
Predicted protein	Anticodon_1;tRNA-synt_1	XP_002177376.1	2.87
2-oxoglutarate dehydrogenase E1 component	E1_dh;Transket_pyr	XP_002182131.1	2.87
Predicted protein	Patched	XP_002180233.1	2.86
Predicted protein		XP_002176702.1	2.80
Predicted protein	HGTP_anticodon;tRNA-synt_2b;WHEP-TRS	XP_002179281.1	2.80
Ferredoxin-dependent glutamate synthase, fusion of large and small subunits	GATase_2;Glu_syn_central;Glu_synthase;GXGXG	XP_002184279.1	2.79
Alanine--tRNA ligase	DHHA1;tRNA_SAD;tRNA-synt_2c	XP_002183027.1	2.73
Long chain acyl-CoA synthetase	AMP-binding	XP_002186275.1	2.73
Predicted protein	HATPase_c;HSP90	XP_002185608.1	2.68
Predicted protein		XP_002184141.1	2.68
Predicted protein (Fragment)		XP_002182602.1	2.68
Translocator of the inner chloroplast envelope membrane 110k		XP_002185478.1	2.59

Protoporphyrin IX magnesium chelatase, subunit H	CobN-Mg_chel;DUF3479	XP_002185833.1	2.57
Pyrophosphate-dependent phosphofructose kinase	PFK	XP_002182086.1	2.55
Predicted protein	AAA;AAA_2;ClpB_D2-small	XP_002178573.1	2.53
Predicted protein (Fragment)	cNMP_binding	XP_002176836.1	2.51
Predicted protein	FoP_duplication	XP_002182700.1	2.48
Predicted protein	DEAD	XP_002181753.1	2.47
Early-response-to-dehydration protein	DUF221;DUF4463	XP_002183484.1	2.46

Identification and characterization of BCCT transporter Codon optimised BCCT chimera sequence

A BCCT transporter chimera was made and codon optimised for expression in *E. coli*. Added nucleotides, restriction sites and ribosome binding site to allow cloning and expression are shown in Appendix 5-1.

Appendix 11. Sequence added to the codon optimised protein sequence to be synthesised. In green the restriction sites, in red the ribosome binding site sequence, in blue interspaces sequences. Underlined and in italics the start codon and in underlined orange stop codon. Stop codons used are TAG or TAA.

*Xba*I *rbs* *Nde*I *Eco*RI

TCTAGAAATAATTTTGTTTAACTTTAAG**AAGGAG**ATATA**CAT*ATG***...**TAG**GAATTC

Appendix 12. Complete sequence of synthesised BCCT chimera. In green the restriction sites, in red the ribosome binding site sequence, in blue interspaces sequences. Underlined and in italics the start codon and in underlined orange stop codon.

TCTAGAAATAATTTTGTTTAACTTTAAG**AAGGAG**ATATA**CAT*ATG***ACCGATCTGAGCCATAGCCGTGATCCGATT
AGCATTAATCCGGTTGTTAGCGCAATTGGTGTTATTGTTCTGTGGGGTTTAGCAATTTGGTGTATGGTTGATCCG
GATGGTAGCCGTGAAACCCTGGTTGGTTGGCGTGGTGATGTTACCCTGTATTTTAGCTGGCTGTTTATGGGTAGC
AAAGCGATCTTTTTCTTCTACCTGATTTACGTGGTGTTCAAATATGGCCATGTTAAACTGGGTCGTCAGGATGAA
CCGCCTGAATTTAGTACCGGTGCATATTTTGCAATGATTTTTGCAGCCGGTGTTGCAGTTGGTCTGTTTGTTTTT
GGTGTTGCCGAACCGCTGTGGCATCAAGAAAGCCATTATTATGCAAATGCAGGTTATCGTAGCCAGGATGAAATT
GATATGTTTGCCCTGAATATGACCGTTGCCAATTGGGGTATTAGCGGTTGGGCACCGTATCTGATTGTTGCCGTT
GCAATGGGTTTAGCCGGTCATCGTTTTAATCTGCCGATGACCTTTCGTAGCTGTTTTTATCCGATTCTGGGTCAG
TATACCTGGGGTTGGATTGGTGATCTGATTGATGGTTTTGCAATTGTTGTTACCGTTGCCGGTGTTTGTACCAGC
TTAGGTCTGGGTGCAATTCAGATTGTTGTTGGCTTTCAGTATTTAGGCTGGGTGAAAGATGATATTACACAGGAT
GAAGTTAGCCGTGTTCAGAATGCAACCATTTGGGTTATTACCGTTATTGCAACCGCAAGCGTGATTAGCGGTCTG
AATGCAGGTATTCGTATTCTGAGCACCATTGCCTTTATGCTGGGTTTAGTTCTGCTGTTTCTGGTGTTTGTTATG
GACGATACCAAATATCTGCTGAACCTGCAAGTTCAAGAAGTGGGCTATTATCTGCAGCATAGCATTTTCCAGCTG
AACTTTTGGACCGATGCATTTGGTCAGATTCGTGAAGGTGGTGGTCGTGCAGTTGATGGTGCAGCAGCAGCCGCA
TGGTGGATGGATGCATGGATGATCTTTTATCAGGCATGGTGGGTTAGCTGGTCAGCATTTGTGGGCCTGTTTGTT
GCACGTATTAGCCGTGGTCGTACCGTTAGCGAAGTTATTATCTATAGCCTGGTTGCACCGGTTGCCTATTGTATT
ATTTGGTTTAGTATTTGGGGTGGTGTTGGTCTGCGTCAGGCACGTCAGGGTCGTGAACTGGAAGCATTAGGTGGC
ACCCTGTTTAATGATACCGAACATTTTCTGGTTCCGGGTAGCACCAATTGTTATGATGTTCCGCAAGAAACCCTG
AGTCAGGATGGCACCGTTGTTTTTGAAAATCATCTGCTGGGTGTTACACCGGTTTGTCAGTTTGATAGCAGCCAG
AGCAATACCGCAGCATTTAATGTTCTGTATAGCTTTAGCTTTCCGGATAGCTTTGATACCGGTTTTGGTCCGACA
CTGAGCGTTCTGTTTATTATCAGCCTGGCAATCTATTTTGCGACCAGCAGCGATAGCGGTAGCCTGATTGTGGAT
CATCTGGCAAGCAATGGTCGTAAAAATCATCATTGGATTCAGCGTCTGTTTTGGGCAGTTACCGAAGGTGCAGTT
GCCACCGCACTGCTGAGTGCCGGTGGTGAACAGGCACTGCAGGCAGTTCAGGCAGCAAGCATTGTTTGTGGTCTG
CCGTTTTGTTTTATGCTGTGTTATCTGCTGCAGTCCATTGAACTGTTTTGTCGTGAAGCACTGATCGTTGGTGAT
GGTCAGGATTATCGTATTCCGGCACAGAGCACCTTTAGCGTTCCGATTTATGGTGGCATTTTCAACAACATGGAA
TTTCTGACCAGCGCAGGTAGCGTTAATCCGAAACGTATTGAATTAGGTATGGATAAAGCGACCACCTTTCATGTG
GTGGAATTTATCAAAGGTCTGTTCGTTCCGTTTGTGAGCCTGCATAAAGTTCTGAGTGATGCATATCCGCGTAAT
AGCCTGAGTAATACAGCCGTTACCGCAGTTTATACCGTTTGTTATTACATGTGGATTGGCATTTTTGCAAGCCTG
GGTAGTAAAGAAGGTCTGATTGGTTGGGGTTGGCTGCTGTTTTTTGCATGTGCATGTATTCTGGGTAGCGTTCGT
GGTGGTTTTCGTGCACGTTATAATGTTCGTAGCAATATCCTGGGTGATTATATGGCAAGCCTGTTTTTTTGGCCT
CAGGTTTTTACCCAGATGCGTCAGCATTGTGTTGAACTGAATCTGCCGCAGGATCATGGCGATCTGCCGAGCGAA
AAAGAAAAAAAACTGGATGGTTCCGATTCCGATGAAGTTGCAGCA**TAGGAATTC**

Appendix 13. Optimised sequences used for gene synthesis. Overexpressed proteins were tested for MTHB-methyltransferase activity. tested proteins

>XP_002296978.1 predicted protein [Thalassiosira pseudonana CCMP1335]

sTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGAAGTTTAGCATTCTGGTTCTGCTGAGCACCCTGA
GCCTGACCGCAGGTTTTGCAGTTGCAACCGCAGATAGCAGCAGCAGTATTAGCAGCTCAAGCACCACCAGTACCGCAGCC
GGTGGTGCCGGTGTTCATGTTCCGATTGGTAAAGATGGTGAAGGTGAATATACCGCAAGCACCAAAGGTTGCTTTGATGT
TATTGCCACCGCAACACCGCTGATTCTGAGCCAGATTAGCACCCAGCCGCTGCGTCCGTTTGGTATTGGTAGTCCGGCAT
ATAACATTGCAGATTATGGCACAGCAGATGCAGGCACCAGCCTGGGTCTGATGAGCAAAATGATTACCGCAGTTCGTGAT
CGTACCAGCAGCGATAAAGAAGTTGTTATCCATTACGAAGATCAGCTGACCAATGAATGGCAGAGCGTTTTTAATCATGC
CCTGGGTATTAAAGCAGTTACCGATGCCTATGGTAAACCGGTTCCGAATCCGTATGATCTGGAAAATGTTTTTGTTGAAG
CATGCGGTGTGGGCTTTCATAATCAGTGTTATCCGAGCAATAGCGTGGATTTTGGTGTTAGCTTTACCGCCATGCATTGG
CTGAGCCGTTTTCCGAGCAGCCTGGTGGGTAAAGATACCATGCATGCAGCACGTAGCGAAGTTCCGCCTACACCGGAAAA
AGAACAGGCAGCAAGCGATTGGAAAAGCATTCTGAAAGCACGTGCAAAAGAACTGGTTCCTGGTGGTCGTTTTGTTTGTG
TGAATTTCTGCAAAAACACCGACGGTTATTTTCTGGGTCAGACCGATGTTGGTGAAAGCATGTGGGATAGCTTTCAGAGC
GCATGGGATCAACTGAAAAGTGATGGTCTGATTGATGAAGAAGAACGTCTGGGCGTTAGCTTTCCGAACTATTATCGCAC
CAGCGAAGAATTTCTGGATGCAGTTCACGATGAAGAAATTTCAAGCAGCCTGAAAGTTGTGAGCATTGAAGAACGTGTTG
TTCGTTGTCCGTATCGCGAACTGTATACCAGCGGTAAAAGCAATAAAAGTCCGCGTGAATATGCCGAATGGTTTGTTCCG
ACCACCAAAACCTGGTCACATAGCACCTTTAAAAGCGCACTGAAATGCGATAAAAGCGAGGATGAAAAAGAAGCCATTAT
GACCCAGTTTTGGGAGAATTATATGAGCCTGGCAGAGAAAGAACCTGAAAAACATGGTATGGATTATGTGCATGCCTATA
TCGTGTTTGAGAAGATCCTGTAGGAATTC

>EJK49332.1 hypothetical protein THAOC_31798 [Thalassiosira oceanica]
TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGCGTATTGGTCCGTTTCTGACCAGCAGCATTGCATT
TCAAGAACGTTATAACCTGACCATCAATAATATGGCACCGCCTAATGCAGAAGCAGTTGCAGTTCAGGTTGATCAAGAAA
TTTATGGTCGTAGCGGTGATCTGGATAAACACATGGAAGGTGTTAAAGCACAGTATGACACCAAAGAAAAGCTGGAGTTT
TATGCACAGGTTATGGGTGATGGTACAGCCAATATTCATTTTGGCAAATGGGATAACGTGAACATCGATGAAGAAGGTGC
ATACGGTAAAGCAAGCGAGCAGATGACCGATTATATGTTTGATCTGGCAACCCAGCTGAAAGGTGCAGCACCGGAAGAGG
GTATTAGCTATGTTGATTTAGGTAGCGGTACAGGTGCAGCCGCACTGCGTCTGTGTGAAAAACATAGCAGTATTGCAAAA
GCCACCTGTCTGAATCTGTGTGATGAACAGAATGCACTGGCAACCAGCCGTGCAGCAGATCTGGGTCTGAGCGATCGTGT
TACCGTTGTTACCGGCACCTATGAAGAATGTCCGTTTGATGCAGATCAGTTTGATGTTGCATTTAGCCAGGATGCCTTTG
TTCATGCCTTTAGCAAAAAAAAGACCTTTAGCGAAGCACTGCGTATTACCAAAGCCGGTGGTGTTTTTATCTTTTGTGAT
CTGATGTGTGGTAGTGGTGAAGGTGTTTCAGATGAAGAACTGCAGACCTTTGCAGCAACCAATATGGTTAATGATTGGCT
GAGTCCGGAAGAAAATGTTAAAGCCTGTGAAGAAGTTGGCTGGAAAGAAGTGAAATTTGTTGATCTGACCGCAGATATTC
GCATCAGTTTTCAGCTGATGCTGCGTAAAGTTGAGAAAATTCTGGATGCAGGTAACCCGGATAACATTGATGAAAAACTG
CTGGAAGGCTACAAAAGCAATCTGGCCAATCGTATTAAACAGGTGGATCGTGGTGTGTTTAAATGGGGTGTTGTGCATGC
CAAAAAAAGCGATTAGGAATTC

>EJK59074.1 hypothetical protein THAOC_20742 [Thalassiosira oceanica]
TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGCACCGAATCTGATGAATAATGGTGAAGAAACCCT
GTGCATTAGCGATAGCTTTCATTTTAGCAGCTTTTATGCAAGCAGCAGCAATCTGGCAGCACTGCAGGATGATGATAGCG
TTCAGCCGAATGGTCCGCGTAGCCGTGCAGTTAGCGAAGCAGATAGCATTGCACCGACCATGGATGATGGTCCGGTTCAT
AAACCGCTGACCGTTGCAAATCCGGATCGTGTTAAACTGAAATTCTATAGCAATGTGCTGGGCGATAACAGCACCTTTAT
TCATTATGGTAAATGGGATGGCATCGATCTGGATCAGCCTGGTGCCTATGGTATGGCAAGCGAAGCAATGACCGATTATA
TGTATCGTCTGAGCCTGGGTCTGCTGACCCATCGTGCAGAAAGCCGTGATTTTGCCTATGTTGATTTAGGTAGCGGTACA
GGTGCAAGCGCAATTCATCTGGCAGAAAAACATAGCCTGACCATTAGCAAAGCCACCTGTATTAATCTGTGTCACGATCA
GAACATTATTGCAGTTGAACGTGCAGCAGAACGTAATCTGAGCGATCGTATTGAAATTGTGGAAAGCAGCTTTGATGAAA
CCCCGTGTGAAGCAAATCATTATGATCTGGCATTTAGCCAGGATGCGTTTATTCATGCCGTTAGCAAAGAGAAAGCATAC
AAAGAAGCATACCGCATTACCAAACCTGGTGGTGCATTTGTTTTTTGTGATCTGGTTTGTGGTGATAACCCGGATCTGAC
CGTTCAAGAACTGGCACAGTTTGCAGAAAAAAACCGCATTAATGATTGGCTGAATCCGAGCCAGACCATTAAAACCTGTA
AACTGAGCGGTTGGTCCGATGTTAAATTTGTTGATCTGAGCACCGATCTGCGTATTAGCTTTCAGCTGATGCTGCGTAAA
GTTAGCTTTGTTCTGGAACATGGTGATACCGGTAGCAATAGCAGCCGTGTTCTGCTGATGAATTATCGTGATAGCATTTG
TCGTCGCATTACCCAGATTGAACGTGGTGTGTTTAAATGGGGTGTTTTTCATTGTCGTAAACCGGTTGTTATGGATCTGG
TTGTTAAACCGCCTGTTCCGTTTGAAAAAACCAATCATCTGATTCTGGATATGGCCGATGATAATGATGAAGGTTGTGTG
ATGGAAACCAATATTGTGGTTGTGGACATCCTGAAAAAAATGAACAAAGAAACCATCGAAAAACTGCCCAAAACCGTTGA
ACTGCTGATTACCATGAGCGCAGGTCTGGATCATATTGATCTGGATGCATGTAGCCAGAATGGTGAACGTACCCAGGTTC
TGAGCAGCCTGTGTTGTACCCCGAGCCAGTATGTTATTAGCCATGTTGGTATTGTCGTGAAAAATAGCGGTCGTGATGCA

ATTACCAGCCATGTGGTTCAGTATTGTCTGAGCTTTATTATCGTTGGTCTGCGTGATGCACTGAATCAGCTGAGCGTTCC
GTTTCCGAGCAGCGGTTGGAATCTGAATTGGAATTGTGAAGGTAAACCTCTGACCAGCAGCACCATTGCAATTATTGGTC
TGGGTCTGATTAGTAAAGCCCTGATTGAAGAAATTCGCAAAATTGCACCGGAAGCGCGTATTATCTATAATACCCGTACA
CGCGATTTTGACTTCGAGAGCAAATTCAACCTGGAATATATCAGCTGTCTGAATGAGCTGGCACGTCAGTGTGATGTTCT
GGTTCCGATGTGTGCCCTGAATAAACAGACCGAAGATCTGGTGAGCAAAGAAGTTATTAGTAATCTGCAGCCGCATGCCG
GTATTATCAATATGGCACGTGGTAAAGTTGTTGAAACCGATGCCATGACCGAAGCGCTGCAGAGCAAAGCAATCAAATAT
GCAATCCTGGATACCACCTATCCGGAACCGCTGCCGAAAGAACATCCGCTGTGGTCACTGGATAATTGTTTTATCTTTCC
GCATTATGCCAGCGAGTGT**TAG**GAATTC

>XM_003080613.1 Ostreococcus tauri sarcosine-dimethylglycine methyltransferase (ISS) (Ot08g00900) mRNA, complete cds

TCTAGAAATAATTTTGTTTAACTTTAAG**AAGGAG**ATATA**CAT*ATG***GGTTGCGCGCCGCGTGGTCGTCGTGACGCGTCTAC
CGACCGTCGTGCGCGTCGTCGTGGTGCGACCTGGTGCGACGGTCGTGCGCGTGACGGTGACGACGGTGCGACCACCACCA
CCACCACCGGTATCGACGCGGGTTTCGACCGTGCGCAGGCGGACTTCGTTGGTAAACAGTTCGACAAAGTTGAAGTTGGT
AAAAAAGTTCTGGAACAGTACGACGACGCGCAGCAGCGTGCGTTCTACACCGTTGTTATGGGTGGTGGTGGTGACGACAT
CCACTTCGGTATCTACCGTGCGCCGGGTGACGGTGTTCGTGAATCTTCTGCGGCGACCACCGAATGGATGATGACCCAGC
TGGACATGGCGCGTCCGATCGGTGCGGGTGACCGTGTTCTGGACGTTGGTTCTGGTCACGGTGGTGGTTCTCACGCGCTG
GCGAAACGTTTCGGTTGCAAAGTTCTGGGTTACAACATCGGTCCGCAGCAGAACGCGCAGAACCTGGCGAAAGCGAAAGA
ACTGGGTCTGGGTGACCTGGTTGACGCGGTTGTTGGTGACATCAACCAGCCGTTCCCGGCGGACTGGACCGACTCTTTCG
ACTCTGTTTGGTCTTGCGAAGTTCTGTGCCACGCGGGTGACAAAACCGAACTGTTCAAAGAAATCTACCGTGTTATGAAA
CCGGGTGCGGCGTTCGTTTTCTCTGACATCATGGGTGCGGACGGTGCGGACGAAAAAACCCTGAAAGGTTTCACCGACCG
TAACGCGACCACCGTTATGGGTCGTCCGTCTGGTTACATGCAGTGCATCAAAGACGCGGGTCTGGACTACGTTACCTGGT
GGGACGGTTCTAACCACCTGGAAACCTACTTCCGTGACATGATCAACCAGATCCACACCCACCGTGAAGAAATGCTGTCT
AAAGGTATCACCGAACAGTACCTGAACAACTGGCTGGAATCTCTGACCGAACGTGCGGACACCCAGCGTGACAAAGGTGT
TTTCGCGTGGGGTGTTTTCGTTTGCCGTAAACCGCTGGCGTGA**TAG**GAATTC

>Codon optimized. gb|CP007142.1|:1327347-1328831 Gynuella sunshinyii YC6258, complete genome

TCTAGAAATAATTTTGTTTAACTTTAAG**AAGGAG**ATATA**CAT*ATG***AAACAGGTTTCTTACGAAATCTCTTCTCAGGTTCT
GGAACAGTACGACTCTCCGCAGGGTCGTGCGTTCTACCGTCAGGTTATGGGTGACTCTGGTTTCAACATCCACTACGGTA
TCTACCCGTCTGAAAACGAAACCATGAAAACCGCGTCTGAAAACATCATCCGTCACCTGCAGGAACTGGCGCAGCAGCGT
GGTGTTCACCTGCCGCAGGCGTCTATCCTGGACCTGGGTTCTGGTACCGGTGGTGCGGCGCACTACCTGGCGGGTCACTT
CGGTTGCCACGTTACCTGCGTTAACATCTCTCCGGAACAGAACAAAATCAACCGTAAACAGGCGCAGGAACTGGGTATCG
ACGACCTGATCAAAATCGAACAGTGCTCTTTCGACAACCTGCCGGGTAAATGGTCTGGTCAGTTCGACCTGGTTTGGTCT
GAAGAAGCGTTCTGCCACGCGGAACACAAAGACACCGTTATCAAAGAAGCGTGGCGTGTTCTGAAACCGGGTGGTGTTCT
GGTTTTCTCTGACATCATGGAAGGTGAACTGAACCAGGACACCCACACCTTCTCTGACCGTAACGCGATCCGTGACCTGG
CGTCTCCGTCTGACTACATCCGTCTGTGCATGGCGAACGGTTTCTACCACCTGTCTTACCACGACCTGTCTCACCACCTG
CCGATCAACTTCCGTAAAATGATCGACCAGATCGACCAGCACTACGACCGTCTGGTTGACAACGGTGTTTCTTCTAAATA
CGCGGACAACTTCCGTCAGTCTCTGAACGACCGTGTTAACGCGGCGTTCCAGGGTAACTTCTCTTGGGGTTCTTTCGTTA
TGAACAAATCTACCCGTCTGGAACACCCGCACCTGCGTTCTGTTATCGAAGGTCGTAACCTGTGCCGTATCACCGCGGAA
CCGCTGACCCGTGAAAACCTGGCGGAACTGGGTACCCTGTACGCGTACGACCAGCCGCTGGAACAGCACCCGCCGGTTCC
GCAGCACTCTTGGCCGGTTAAATACCCGCAGCGTATGGAAACCGGTCGTGGTCTGGCGCCGCTGGGTACCGACGACATGA
CCATGACCTGGCAGCAGGAAATCCTGCACGCGCGTAACGAAGTTCTGGAACACTGCTACCAGAACATCGCGTACCGTGAC
GAACAGCACGGTTACTTCATCGAATGGTTCAACCAGCACGTTGAATCTGGTCAGAAATTCTACTGCCCGGGTGTTCCGCT
GCTGTACGTTCTGGCGGCGCCGGTTGACCACCCGAAAGCGCAGGACTTCAAAGCGTTCATCGCGGACGGTTCTCACGGTG
TTATCATCAACCCGGGTGTTTGGCACACCAACCCGATCCCGCTGATCGACACCGAAGTTACCCTGACCACCACCCAGTCT
ATCGTTGACGCGTCTTGCGACTGCTCTCTGTCTGCGGAACACAACCAGTGGCTGAACATCACCGTTTCTACCGGTACCGA
CTCTTGA**TAG**GAATTC

>SDW35278.1 sarcosine/dimethylglycine N-methyltransferase [Aidingimonas halophila]

TCTAGAAATAATTTTGTTTAACTTTAAG**AAGGAG**ATATA**CAT*ATG***AACAAGGTGATGGAAAAAGCATATACCGCAGCAGA
TGAAACCGCACGTGAATATTATAATTCAGATGATGCCGACAACTTCTACTTTCATGTTTGGGGTGGTGAAGATATTCATG
TTGGTCTGTATCGTGATGATAAAGAACCGGTTTATGATGCAAGCCGTCGTACCGTTGCACAGATGGCAGCACGTCTGCCG
AAACTGGATGCAACCACACAGGTTCTGGATATTGGTGCAGGTTATGCAGGTAGCGCACGTTATCTGGCCCATGAATATGG
TTGTAGCGTTACAGCCCTGAACATTAGCGAAAAAGAAAATGAACGTGGTCGCGAAAAAAATCGTCAGCAGGGTGTTGATG
ATCGCGTGGAAATTATTGATGGTAGCTTTGAAGATATCCCGTTCGATAACGAAAGCTTTGATGTTGTTTGGAGCCAGGAT
GCAATTCTGCATAGCGGTGATCGTGAACGTGTTCTGGAAGAAGCAGTTCGCGTTCTGCGTCCTGGTGGTCATCTGATTTT
TACCGATCCGATGCAGGATGATGAATGTCCGCGTGAAGCACTGGAACCGATTCTGGCACGTATTCATCTGGAAACCCTGG
GTAGTCCGGGTTTTTATCGTGATACCCTGCGTCGTCTGGGTTTAGAAGAACAGGATTTTGATGATCACAGCGATATGATC
GCAATGCATTATGGTCGCATTCATCAGGTTCTGAGCGAACATGAAGATGATGTTGGTCAGCATGTTAGCCGTGGTTATAT
TGATCGTATGAAACATGGTCTGCAGCATTGGGTTGAAGGTGGTCAGCGTGGTAATCTGCGTTGGGGTATTTTTCATTTTC
GTAAACCGGCACAG**TAG**GAATTC

>WP_028481482.1 SAM-dependent methyltransferase [Nesiotobacter exalbescens]

TCTAGAAATAATTTTGTTTAACTTTAAG**AAGGAG**ATATA**CAT*ATG***AGCAAACAGAATCTGAGCAATGCAGAACTGACCAT
GCGTGATCAGGTTTATGGTGATAATCCGCTGGAAGATCGTGAAACCGATCATTACAAAAAAGAGTACATCAGCACCTTCG
TGGATAAATGGGATGAACTGATTGATTGGGATGGTCGTGCAGAAAGCGAAGGTCAGTTTTTTATCGATATTCTGCGTGCA

CGTGGTAAAGAACGTATTCTGGATGTTGCATGTGGCACCGGTTTTCATAGCGTTCGTCTGATGGAAAATGGTTTTGATGT
TACCGCAGCAGATGGTAGCGCAGCAATGGTTGCAAAAGCATTTAACAATGCACAGAGCCGTGGTCTGATTCTGAAAACCG
TTCAGGCAGATTGGCGTTGGCTGAATCGTGATGTTCATGGTAAATATGATGCCATTATTTGCCTGGGCAATAGCTTTACC
CATCTGTATGAAGAAAGCGATCGTCGTCGTGCACTGGCAGAATTTTATGCAGCACTGAAACATGATGGTGTGCTGATTCT
GGATCAGCGTAATTATGATGCAATGCTGGATCGTGGTTTTACCACCAAACACAAATACTATTACTGCGGTGAACAGGTGA
CCGCAGAACCGGATCATGTTGATGAAGGTCTGCTGCGTATGCGTTATAGCTTTCCGGATGGTAGTGAATATACCCTGAAT
ATGTGTCCGATCCGCAAAAACTATCTGCGTCGTCTGCTGAGCGAAGCAGGTTTTGAACGTGTTCGTACCTATGGTGATTT
CCAAGAAACCTATGCAGAAGATGATCCGGATTTTTTCATTCACGTTGCAGAAAAAAGCGCACTGCATCTGGTTCGTTGGG
GTGGCACCGTTGCCGAAGGTCAAGAAGATATTCGCGATTATACCGAGGATTATTATGATTCAGATGATGCAGCCGTGTTC
TATAGCACCATTTGGGGTGGTGAAGATCTGCATGTTGGTCTGTATGATACCACACAGGATATTCGTAGCGCAAGCGATCT
GACCATTGATCGTATGATTGATACCCTGCCTCCGCTGACCAAAGATAGCCATGTTCTGGATATGGGTGCCGGTTTTGGTG
GTGCAATGCGTCGCCTGGTGAAAAAAACCGGTTGTCAGGCAACCTGTCTGAATATTAGCGAAACCCAGAATGAGTACAAC
CTGCAGAAAATTCGTCAGGCACGTCTGAATGATCGCATCAAAGTTAAACATGGCGTGTTTGAAGATGTGCCGTTTGAGGA
TGCAAGCTTTGATGTTGTTTGGAGCCAGGATGCATTTCTGCATTCAGATCAGCGCAATAAAGTTCTGGCCGAAGCACTGC
GTGTTCTGAAACCTGGTGGTAGCCTGATTTTTACCGATCCGATGCAGGCAGATGAAGTGAATGTTGAAAATCTGCAGCCG
GTTTATGATCGTCTGCGTCTGAATAGCATGGGTAGCTTTCGTTTTTATCGTGAAGCAGCAGAAACCCTGGGTTTTGAAAC
CGTTGGCCTGGATGAAATGACCCATAATATGCGTGCACATTATGCCCGTGTGAAACAAGAAATCGAGAAAAACTATGAAC
TGCTGCGTGAAAAAGGTGCAAGTGCCGAATATCTGGATAAAATGCTGGTTGGTATGGATCATTGGGTTAATGCATGTGAT
AGCGGTAATCTGGCATGGGGTATTCTGCATTTTCGTAAACGTGCA**TAG**GAATTC

>Chorda filum GWS036236_c27154_g1_i1 (transcribed RNA sequence) – brown macroalgae
TCTAGAAATAATTTTGTTTAACTTTAAG**AAGGAG**ATATA**CAT*ATG***ACCGCAGCAACCGAATTTACCAGCGCAGAGCAGAT
ACCACCAGTAAAGCAGCAGTTGAAAGCAAAGTGAACGAACAGTATGATACCGATGCAAGCATGGCCTTTTATGAATATGT
TATGGGTGGTGGCGGTGATGATATTCATTATGGTCTGTTTCTGACCGAAAAAGATGGTCTGAAAGAAAGCAGCCAGAATA
GCGTTGAAGCACTGGCAAATATGGCAGTTGATGCAGGCACCCTGAAAAAAGGTGAAAATAATGCAGCAGTTCACTGCCTG
GATTTAGGTAGCGGTAAAGGTGGTGCAAGCCGTTGGCTGGCAAAAGAATATGGTTGTAAAATCACCGCCTTTAATCTGGG
TGAACGTCAGAATGCATTTAACCTGGAACGTGCACAGGCAACCGGTATTGGTCATCTGGTTGAAACCCATCTGGGTAGCT
TTAATGAACCGCTGCCTGCAGGTTGGACCGATAAATATGATATGGTTTGGAGCCAAGAAGCCTTTTGTCATTGTATGGAT
CAGAAAGCACTGATGGCCGAAGTTAGCCGTGTTCTGAAACCTGGTGGTGTTATTGTTTTTAGCGATATCATGCAAGGTGA
TGGTGGTGGTGATTGTACCAGCTTTACCGGTCAGAATGTTGTTGCCAGCATGGCAAGTCCGGCAATGTATAAAGAAGCAA
TGATTGGCGCAGGTATGAGCATTGTTGAACATAAAGAACTGACCAGCCATCTGACCCCGTATTTCAAATGTATGCTGGAT
GCAGTGAAAGATGGTAAAGCAACGATGCTGGAACAGGGTGTTACCCAAGAACGTCTGGATGCCTATGAAGATGATCTGAG
CACCCGTTTTGAACGTGTTAAACAGGGTCATTTTGCCTGGGATATGTTTTGCGCCAAAAACT**TAA**GAATTC

Sargassum vulgare Sample_MAC_c89409.graph_c1_seq1 (transcribed RNA sequence) – brown macroalgae
TCTAGAAATAATTTTGTTTAACTTTAAG**AAGGAG**ATATA**CAT*ATG***ACCGCAGCAGCAGAATTTACCAGCGCAGCAGCCGA
TACCAGCAATAAAACCGCAGTTGAAAGCAAAGTGAACGAGCAGTATGATACCGATGCAAGCATGGCCTTTTATGAATATG
TTATGGGTGGTGGCGGTGATGATATTCATTATGGTCTGTTTCTGACCGAACAGGATGGTCTGAAAGAAAGCAGCCAGAAT
AGCGTTGAAGCACTGGCAAATATGGCAGTTGATGCAGGCACCCTGAAAAAAGGTGAAGATAATGGTGCAGTTCATTGCCT
GGATTTAGGTAGCGGTAAAGGTGGTGCAAGCCGTTGGCTGGCAAAAGAATATGGTTGTAAAATCACCGCCTTTAATCTGG
GTGAACGTCAGAATACCTTTAACCTGGAACGTGCACAGGCAACCGGTATTGGTCATCTGGTTGAAACCCATCTGGGTAGC
TTTAATGAACCGCTGCCTGCAGATTGGACCGATAAATATGATATGGTTTGGAGCCAAGAAGCCTTTTGTCATTGTATGGA
TCAGAAAGCCCTGATGAAAGAAGTTAGTCGCGTTCTGAAACCTGGTGGTGTTATTGTTTTTAGCGATATCATGCAAGGTG
ATGGTGGTGGTGATTGTACCAGCTTTACCGGTCAGAATGTTGTTGCCAGCATGGCAAGTCCGCAGATGTATAAAGATGCA
ATGACCGGTGCAGGTATGAGCATTCTGGAACATAAACAGCTGACCAGCCATCTGACCCCGTATTTCAAATGTATGCTGGA
TGCAGTGAAAGATGGCAAAGATACCATGCTGAAACAGGGTGTTACCCAAGAACGTCTGGATGCCTATGAAGATGATCTGA
GCACCCGTTTTGAACGTGTTAAACAGGGTCATTTTGCCTGGGATATGTTTTGCGCCAAAAAC**TAA**GAATTC

Pyropia haitanensis Unigene84_All (mRNA sequence) – red macroalgae
TCTAGAAATAATTTTGTTTAACTTTAAG**AAGGAG**ATATA**CAT*ATG***ACCGGCACCAATGGTGCAGCAAGCCCGACCGATAC
CGTTCTGGCACAGTATAATACCGATGAAAGCGCACAGTTTTATAGCGTGATTATGGGTGATGGCACCGCACATGTTCATT
ATGGTGTTTATACCAGTCCGGATGATACCGTGCGTCAGGCAGTTGCAGCAACCAATACCACACTGCGTGTTCTGGCAGAA
AGTGCCGGTGCAGTTTTTCGTCCGGGTGTTCGTGTGCTGGATTTAGGTGCAGGTAATGGTGGTACAGCACATGCACTGGC
AGCAGCCACCGGTTGTAGCGTTACCTGTCTGAATCTGTGTGAAGTTCAGAATGCAGCAAATGAAGCAGCAGCAGCCGCAG
AAGGTCTGAGCAGCCTGATTACCGTTGTTACCGGTAGCTTTGAAGATCTGCCTGCAAATTGGAGCGGTAGTTTTGATGTT
GTTTGGAGCCAGGATGCAATTGTTCATAGCAGCGATAAAGGTCGTGTTTGGGCAGAAGCAGCACGCGTTCTGGTTCCAGG
TGGTTATGTTGTTCTGAGCGATATTATGGCAGCACCGAGCGCAGCAGATGCAGCCCTGCGTGCATTCAAAAAACGTCTGC
ATGTTGATGAACTGCTGACCCTGGATGGTTATGAAGCAGGTCTGCGTGAAGCCGGTGTTAGCACCCTGCGTACCCGTGAT
ATGAGCGGTCAGCTGCTGCCGAATTATCGTCGTATGCTGGAACGTATTACCACCGAACGTGCCCGTCTGACCCAGTGTAG
TGATGCATATCTGGAACAGTATGCCGGTCTGCTGCGTGATAACATTAAAGTTCTGTGTGATGGTGAAGCACAGGCATGGT
GTGCACTGGTTGCACGTAAAAATGGTTTTGCCGCAACCAGCCGTACCGCAAGCGCAGTTGCACCGG**TAA**GAATTC

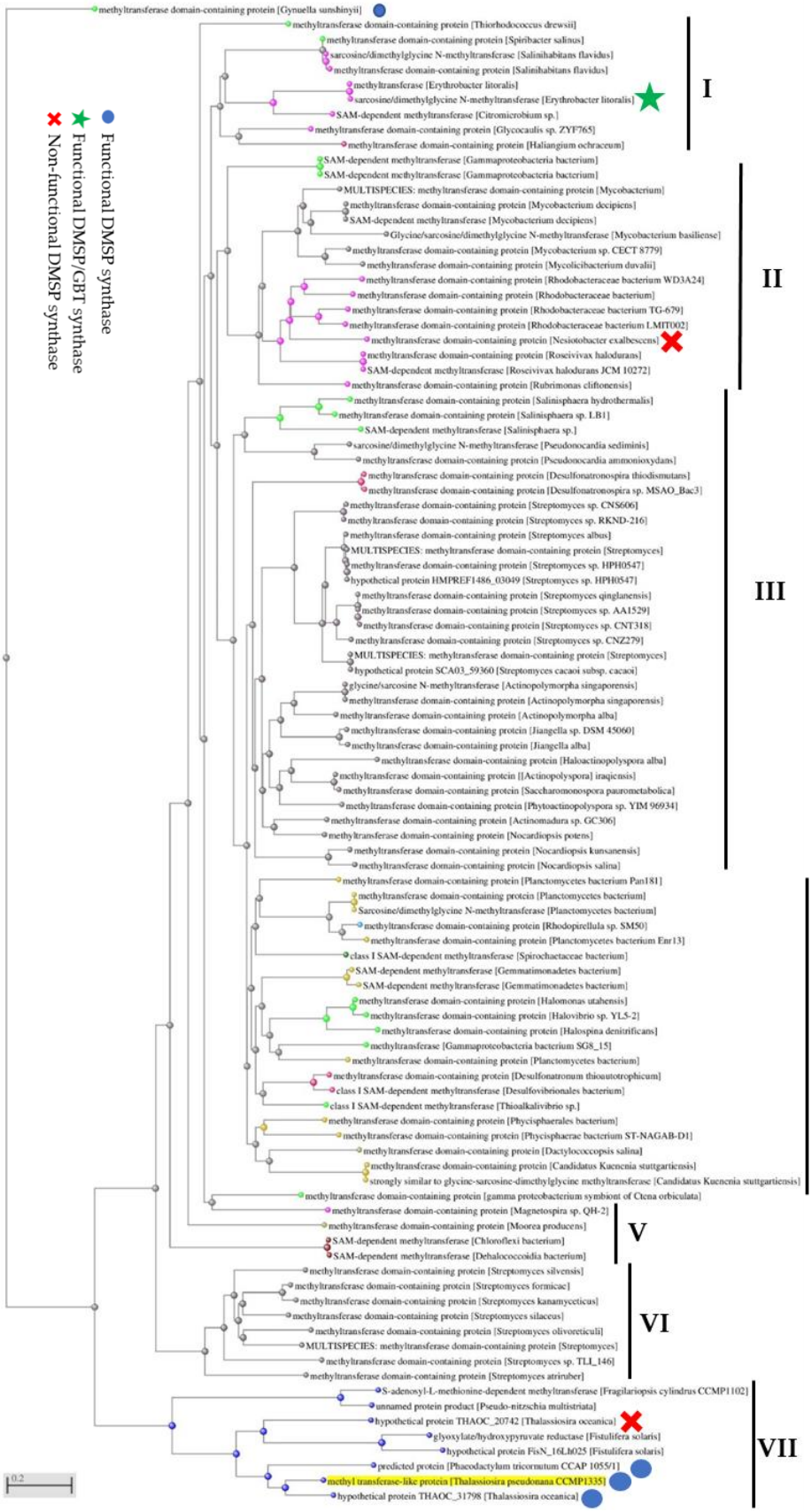
>WP_075904519.1 hypothetical protein [Moorea bouillonii]
TCTAGAAATAATTTTGTTTAACTTTAAG**AAGGAG**ATATA**CAT*ATG***GCAAATCTGAGCGTGATCGATACCTTTAACCAGAG
CTATTTCGATAGCAAAGATATGGACCTGTTTTATCGTCGTGTTAGCGGTGAACATACCCATTGTGGTATTTTTGAACATC
CGAACGAGGATGTGTATATCGCAAAAAAACGTACCACCGAATATATGACCAGCCTGCTGACCCTGGATCGTGATAGCCAT
CTGCTGGATTTAGGTAGCGGTTATGGTGGTGCAGCACGTTATATTGCAAAAGAATTTGGTTGTCAGGTGAGCTGCATTAA

TCTGAGTGAACAGCAGAATGCCATTAACATTGATCGCAATCAAGAAGAAGGTCTGAGCGATCTGGTTCATGTTCATCAGG
GTAGCTTTGAAAAACTGCCGTTTAGCGATAGTAAATTCAATGCAGCATGGGCACAAGATAGCCTGTTCTATAGCGATACC
CAGCTGCAGGCATTTCGTGAAGCACATCGTGTTCTGGTGAAAGGTGGTGAATTTATTGCATGCACCTATTTTTTCGGTGG
CAATTATCCGAGTCCGGAAGAAGTTAATAAAGTCGTTGATTGGTATACCGGTGGTGGCATTCATAAAGTGTATTTTCTGC
ACATCGATGACTATCGTAAAGTGGCAGATGAAATTGGTATGGCAGAAGTGCAGGTTATTGATCTGACCCATCATATTAGC
GTGAACTATTGGCAGATCCTGAAAAAGATGGAAGAAATTCAGGCCGATGAACAGCTGTGGTCAGATGAATTTTTCGAGAA
AAAAAAGCAGCGTCTGCTGGACTGTGCAGAAGTTGGTAAAAGCGGTCTGCTGCAGTGGGGTATTCTGCATTACCGTAAAG
AAAACTAGGAATTC

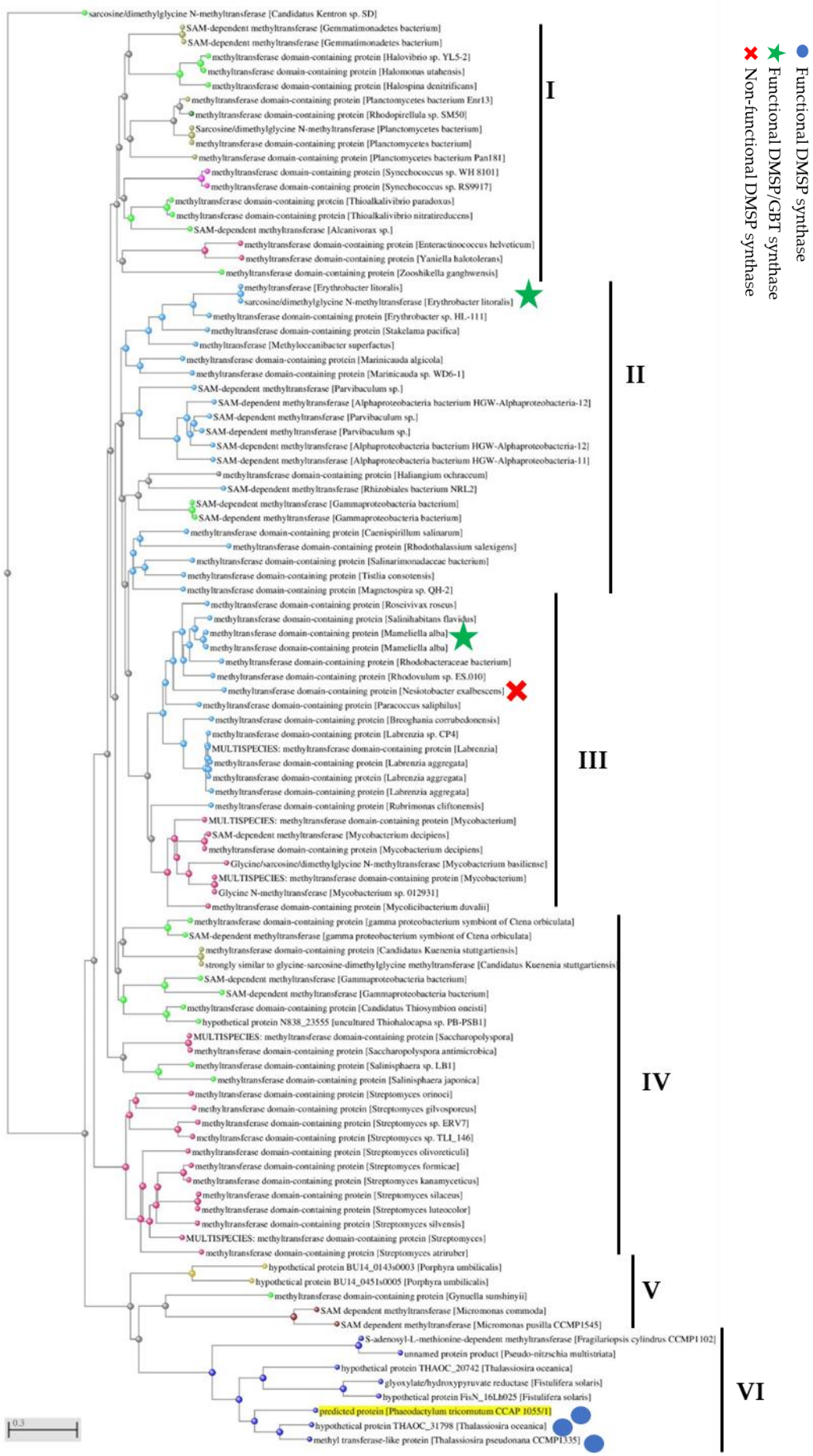
>OSX77567.1 hypothetical protein BU14_0143s0003 [Porphyra umbilicalis]
TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGACCGGCACCAATAGCGGTAATGCACCGGCACAGGC
ACCGGAAGCCGGTGATACCACCAGTACCGTTCTGGCACAGTATGATACACCGGAAAGCGCACAGTTTTATAGCGTTGTTA
TGGGTGATGGTGGTTCCGATGTTCATTATGGTGTTTATACCAAAGCAAGCGATAGCACCCGTGATGCAGCAGAAAATACC
ATTGCAGTTCTGCGTAGCCTGGGTGAAAGTGCCGGTGTTACCTTTGGTCCGGGTGTTAAAGTTCTGGATTTAGGTGCAGG
TAATGGTGGTGCAGCACATAAACTGGCAGCCGAAACCGGTTGTAGCGTTACCTGTCTGAATCTGTGTGCAGAACAGAATG
CAGCAAATCGTAGCGAAGTTGTTGCACGTGATCTGAGCGATCTGGTTACCGTTGTTACCGGTAGCTTTGAAAAACTGCCT
GCAGATTGGACCGCAAGCTTTAACATTGTTTGGAGCCAGGATGCAATTGTTCATAGCAATGCCAAAGATAAAGTTTTTGC
CGAAGCAGCACGTGTTCTGGTTCCAGGTGGTCATGTTGTTATGAGCGATATTATGGCAGGTCCGGCAGCAGATGCAAGCG
TTCTGAGCGCATTTAAAGAACGTCTGCATGTTGATGAACTGCTGACCCCGAGCGCCTATGAAGTTGGTCTGGCAGCAGCG
GGTGTTACCACACTGCGTACCCGTGATCTGTCAGGTCAGCTGGTTACCAATTATCGTCGTATGGTTGGTCGTATTAGCAG
CGAACGTGCCCGTCTGGATCGTTGTAGTGATGCATATCTGGATACCTATGCAGGTCTGCTGGGTGCAAATATTGAAGCAC
TGTGTAATGGTGAAGCACAGGCATGGTTTGCACTGGTTGGTCAGAAAGTTGGTGGTCGTACCACCGCAGCAGCAGCCGCA
GATGCAGATGCCACCGTTGCACCGAGTCCGCCTCTGGTTCGTCGTTGGGTTGCAAGCAAACTGCATGGTGTTCGTGTTAC
CGATAAAAGCGTTCGTTATCATGGTAGCGTTGGTGTTAGCCGTAATCTGATGGCAGCGGCAGGTATTAAAGAATTTGAAG
CAGTTGATGTGGTGAACCTGACCAATGGTGCACGTTGGACCACCTATGCACTGGGTATTGATGCCGATGATGCATTTACC
CTGAATGGTGGCGGTGCACGTCGTGGTGAAACCGGTGATGAATGTGTTCTGATGACCTATGTTCAGAGCGCAGCCTATGA
ACCGGCACGTGTTGCATATTGTGGTGCAGATAATAGCATCATCGATCAGTTTACCTATAGCCATAAAGAAGAAGCAGGCG
AAAACGGTAGCGAAGGTGTGGCAATTGGTAATGGTAGCAGCAATGCAGCCACCGGTCTAGGAATTC

>XP_009035163.1 hypothetical protein AURANDRAFT_59958 [Aureococcus anophagefferens]
TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGAGCGCAGCAGTTCAGAAAACCACAGGTGATTTTAA
GAACAAACAGTTCAGCGAAGAAGAAGTGGCAAGCAAAGTTGTTGAACAGTATGATGAAGTTCATGCCCGTACCTTCTATA
AATACGTTATGGGTGGTGGTGGCTATGATATTCATTATGGTATGTTTCGTACCGCCAGTGATGGTGTTTTTGAAAGCAGC
AAAAATACCAATGCAGCACTGCTGCGTCTGCTGGATCAGACCCGTCCGGTTACCAAAGATAGCGTTGTTCTGGATTTAGG
TAGCGGTCATGGTGGTCTGAGCCATGAAATTGCAACCACCTTTGGTTGTAAAGTGGTGAGCTTTAATATCAGTCCGGAAC
AGAATAACATGAACCTGGAAGAGGCAGCACGTCTGGGTGTTAAAGAACTGATTAGCGTTGTGGAAGGCAATTTTAACGAT
GCAGCAACCTTTCCGCCTAAAAAACTGCCGAAAATCACCCATATTGTTAGCTGCGAAGTTTTTTGTCATGCAGCAAGCAA
ACCGGCACTGCTGAGCGATATCTTTAAAATGCTGGAACCTGGTGGTGCACTGGTTTTTACCGATATTATGGGTGCAGATG
GTGCAAATGAAAAAGCCCTGAAAGATTTTACAGATCGTAATGCCACCACCGAAATGGCACGTCCGAGCGGTTATCTGCAG
CAGATGAAAGATGCAGGTTTTGCACATGTGGGTTTTTTTGATGGTAGTGGTCATCTGCTTCCGTATTTTGCAGCAATGCT
GGATGTTTGTCTGAAACAGGGTGATGATATGGTGAAAGATGGTGTTCCGCGTCCGTATCTGGATAAATGGATTGCAAGCC
TGACCGATCGTGTTAAAATTCAGGGTGACGAAGCAGTTTTTGCATGGGGTCTGTTTAGCGCACGTAAACCGGGTCCGCTG
TATTAGGAATTC

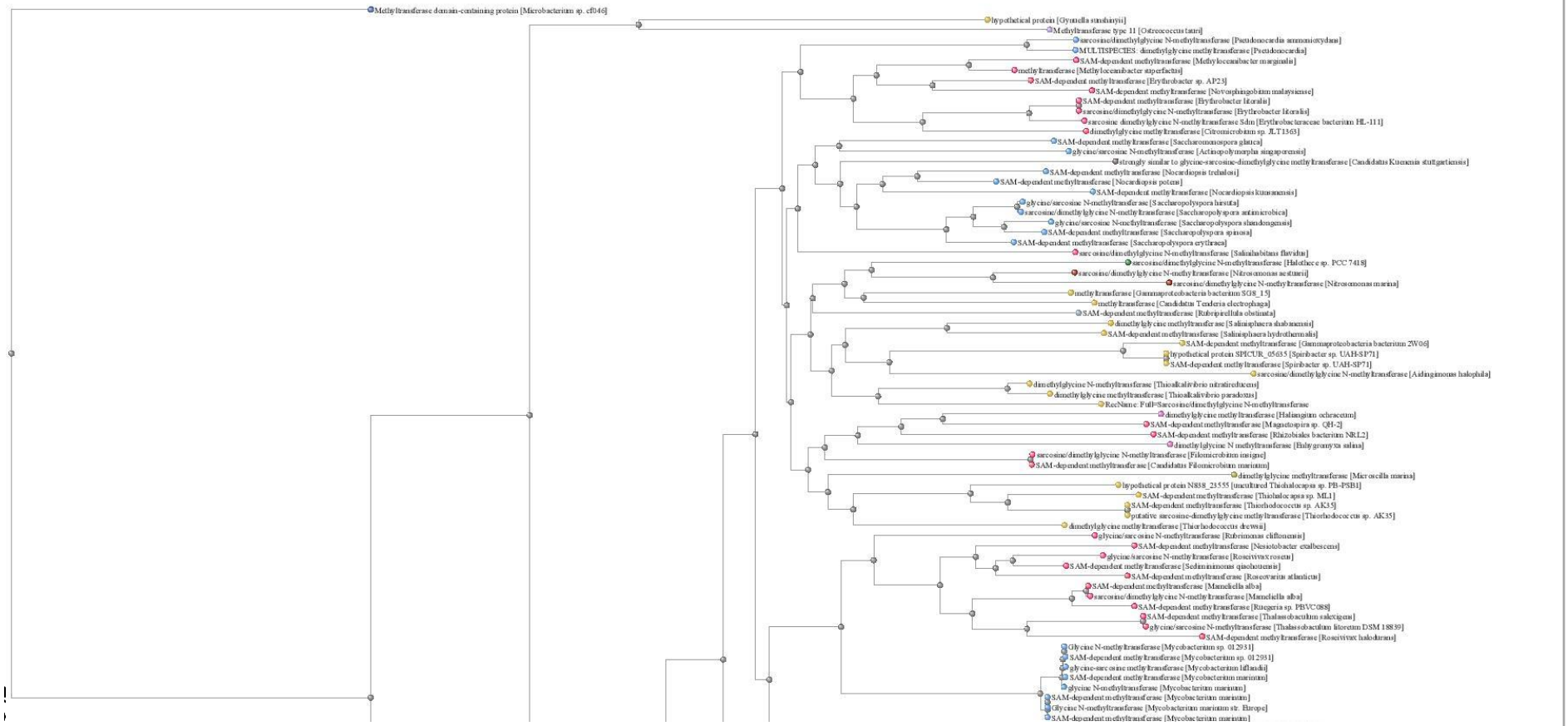
Appendix 14. Neighbour Joining Distance Tree of TpDSYD homologues found by BLAST via NCBI

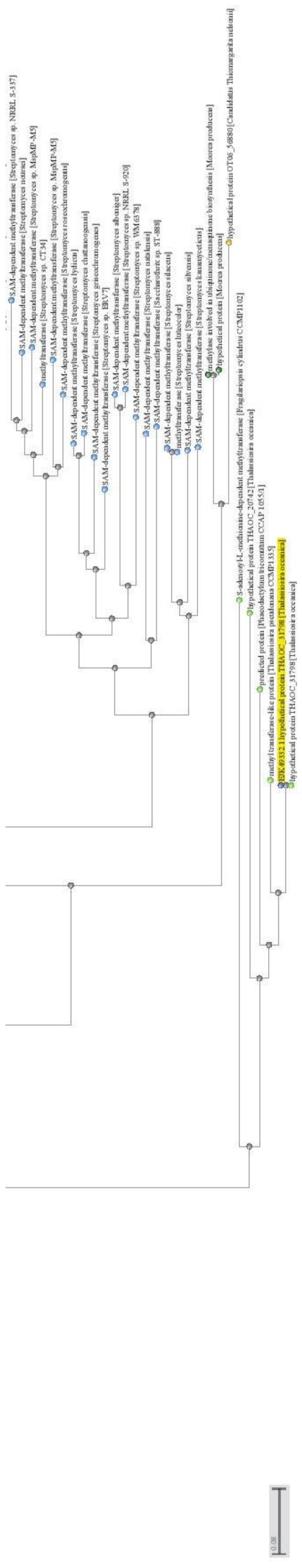


Appendix 15. Neighbour Joining Distance Tree of PtDSYD homologues found by BLASTp in NCBI

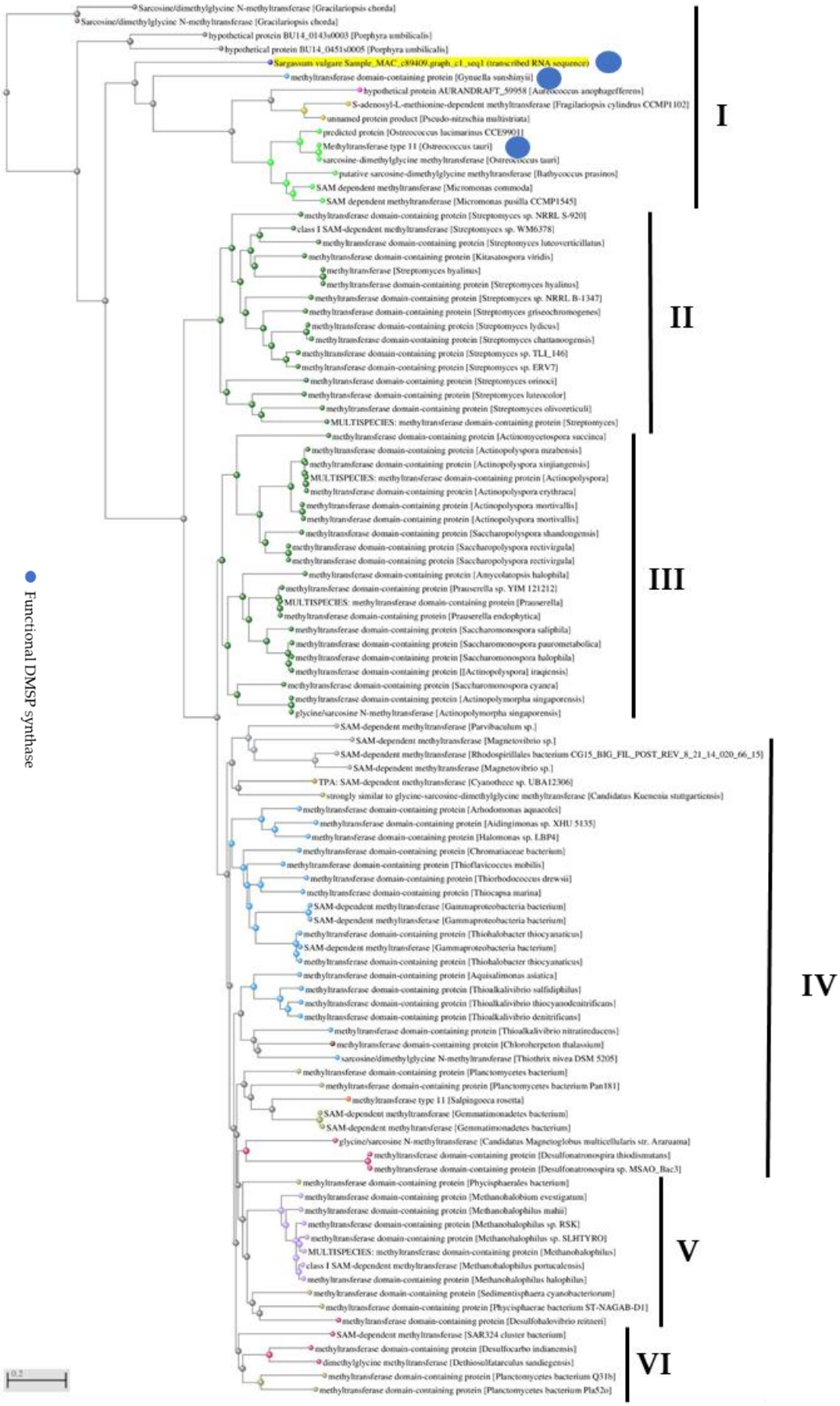


Appendix 16. Neighbour Joining Distance Tree of ToDSYD homologues found by BLASTp in NCBI




SAM-dependent methyltransferase [Streptomyces sp. NRRL S-337]
SAM-dependent methyltransferase [Streptomyces noursei]
SAM-dependent methyltransferase [Streptomyces sp. MspMP-M5]
methyltransferase [Streptomyces sp. CT34]
SAM-dependent methyltransferase [Streptomyces sp. MspMP-M5]
SAM-dependent methyltransferase [Streptomyces roseochromogenus]
SAM-dependent methyltransferase [Streptomyces lydicus]
SAM-dependent methyltransferase [Streptomyces chattanoogensis]
SAM-dependent methyltransferase [Streptomyces griseochromogenes]
SAM-dependent methyltransferase [Streptomyces sp. ERV7]
SAM-dependent methyltransferase [Streptomyces albaniger]
SAM-dependent methyltransferase [Streptomyces sp. NRRL S-920]
SAM-dependent methyltransferase [Streptomyces sp. WM6378]
SAM-dependent methyltransferase [Streptomyces natalensis]
SAM-dependent methyltransferase [Saccharothrix sp. ST-888]
SAM-dependent methyltransferase [Streptomyces silaceus]
methyltransferase [Streptomyces luteocolor]
SAM-dependent methyltransferase [Streptomyces silvensis]
SAM-dependent methyltransferase [Streptomyces kanamyceticus]
methylase involved in ubiquinone/menaquinone biosynthesis [Moorea producens]
hypothetical protein [Moorea producens]
hypothetical protein OT06_56880 [Candidatus Thiomargarita nelsonii]
S-adenosyl-L-methionine-dependent methyltransferase [Fragilariopsis cylindrus CCMP1102]
hypothetical protein THAOC_20742 [Thalassiosira oceanica]
predicted protein [Phaeodactylum tricornutum CCAP 1055/1]
methyltransferase-like protein [Thalassiosira pseudonana CCMP1335]
EJK49352.1 hypothetical protein THAOC_31798 [Thalassiosira oceanica]
hypothetical protein THAOC_31798 [Thalassiosira oceanica]
0.08

Appendix 17. Neighbour Joining Distance Tree of homologues of DSYD from _*Sargassum vulgare* found by BLASTp in NCBI



Appendix 18. Sequences of the functional, non-functional and non-tested DSYD-like proteins aligned and used to visualise their relativeness in a Maximum Likelihood phylogenetic tree.

>Thalassiosira pseudonana CCMP1335
MAPNTTPFDPIATIATEVDPSIYGKSGDISKHVAGVKAQYDTAEKLEFYAQVMGDGTANIHFGKWDNVDLEEEGA
YGKASEQMTDYMFDVATGLLGDDGGVGAEEKTIKYVDLGSGTGAAALRLCQKHDVIAKATCLNLCEEQNALARKC
ASDLGLEDRIAVVTGTYESAPFEANSFDIAFSQDAFVHAFSKVGTFREALRVTKPGGVLVFCDLMCGSGDGVSEE
ELATFAATNMVNDWLSPDLNVRACQEAGWTDVKFVDLTLDIRISFQLMLKKVEKIIQDGNPAKIDEKLLDSYKKN
LANRIVQVDRGVFKWGVVTGKKPLYKFCTRKDSW

>Phaeodactylum tricornutum CCAP 1055/1
MTAATHYKPEVYSKSGEISQHERGVQAQYDTAEKRAFYAQVMGDGTSNIHFGKWDDDIDVTQEGAYGKASDAMTD
YMFALATDLLARNSSTLSYVDLGSGTGGAAIRLLSAHPSLTATCLNLCEAQNATAQQDAVAAGVADRFTVRTGSY
DQAQALLLPENNKQPGLFDVCFSQDAFVHSFSKVRTYEQALAVTKPGGVFVFCDLMCGDGPDVSEDELATFAATN
MVNDWLSPAQNVKACEQAGWQDVVFIDMTVDIKKSFQLMGQKVTRLIESGAAKDIDPVLLDTYRQNLAARVGQVD
RGVFSWGVIHARKAE

>Thalassiosira oceanica CCMP 1005 (1)
MRIGPFLTSSIAFQERYNLTINNMAPPNAEAVAVQVDQEIYGRSGDLDKHMEGVKAQYDTKEKLEFYAQVMGDGT
ANIHFGKWDNVNIDEEGAYGKASEQMTDYMFDLATQLKGAAPEEGISYVDLGSGTGAAALRLCEKHSSIAKATCL
NLCDEQNALATSRAADLGLSDRVTVVTGTYEECPFDADQFDVAFSQDAFVHAFSKKKTFSEALRITKAGGVFIFC
DLMCGSGEGVSDEELQTFAATNMVNDWLSPEENVKACEEVGWKEVKFVDLTADIRISFQLMLRKVEKILDAGNPD
NIDEKLLEGYKSNLANRIKQVDRGVFKWGVVHAKKSD

>Thalassiosira oceanica CCMP 1005 (2)
MAPNLMNNGEETLCISDSFHFSSFYASSSNLAALQDDDSVQPNGPRSRAVSEADSIAPTMDDGPVHKPLTVANPD
RVKLKFYSNVLGDNSTFIHYGKWDGIDLDQPGAYGMASEAMTDYMYRLSLGLLTHRAESRDFAYVDLGSGTGASA
IHLAEKHSLTISKATCINLCHDQNIIAVERAAERNLSDRIEIVESSFDETPCEANHYDLAFSQDAFIHAVSKEKA
YKEAYRITKPGGAFVFCDLVCGDNPDLTVQELAQFAEKNRINDWLNPSQTIKTCKLSGWSDVKFVDLSTDLRISF
QLMLRKVSFVLEHGDTGSNSSRVLLMNYRDSICRRITQIERGVFKWGVFHCRKPVVMDLVVKPPVPFEKTNHLIL
DMADDNDEGCVMETNIVVVDILKKMNKETIEKLPKTVELLITMSAGLDHIDLDACSQNGERTQVLSSLCCTPSQY
VISHVGIVVKNSGRDAITSHVVQYCLSFIIVGLRDALNQLSVPFPSSGWNLNWNCEGKPLTSSTIAIIGLGLISK
ALIEEIRKIAPEARIIYNTRTRDFDFESKFNLEYISCLNELARQCDVLVPMCALNKQTEDLVSKEVISNLQPHAG
IINMARGKVVETDAMTEALQSKAIKYAILDTTYPEPLPKEHPLWSLDNCFIFPHYASEC

>Fistulifera solaris JPCC DA0580 (partial)
MTRSIVFASDHFHFSHFYWKSSQNEEEDTTATSQKSLSRSPPQDLKLKLYSRILHDDENNIHFGKWDHVSLQEKD
AYSQAAHQMTDYMYYDLAMTLIPAHRRTAPNYQCIDLGSGTGASAIRIASSTTNQKGKVVCLNVCPQQNRIAQHN
VTKSNLHNQIQIVEGDFEQVPFPDHAFDLAFSQDAFIHAQSKIQAFREAFRLVRQGGAFVFCDVMAGPTTTDDQV
EQFAQDNAIPDLLRPEQAMQAMRQAGWKDVHCLNVTKDMRISFQLMLQKVLASLETHKHDCDLLEQYCQQIQRRI
EQIDRGLIQWCVFYARKPVALQLLCQPPLPLVNTSEMIVVKKDDSTQPSVVAVDILTPMPRSKIEALPSSVSLLI
TMSAGLDHIDLKACHERKIMVQQA

>Fistulifera solaris JPCC DA0580
MTRSMVVVSDHFHFSHFTWKASPQDSEEETTSTPLKLSSPPPQDIKLKLYSRILHDDENNIHFGKWDRVSLEEKD
AYSQAAHRMTDYMYNDLAIQLIPEHRRNASDFQYIDLGSGTGASALRIASQCTFPGKVVCLNVCPQQNRIAQRAV
AQANLQQTIQIMEGDFEQMPFAENAFDVAFSQDAFIHARRKLQAFREAFRVVRPAGAFVFCDVMAGPMTTDEQVE
QLSRDNAISNLMRPHETVQAMQEAGWKDVHSINVTNDMRISFQLMLQKVHASLETHQQDRDLLELYRQQIQTRID
QIDRGIIQWCVFHGRKPVVLQMLCRPPFPLVNTSELIVVKEQDNDPTSQDPSIVVVDILTPMPRAKVEALPSSVS
LLITMSAGLDHIDMAACHERNITVQQAGRRAITSHVAQYGLALIVLGLRDALSQQQVPFPSQGWNLTWNIPGKPL
EESKIALIGMGTIAKELVRQIRCLAPTAEILYHARSRHETAEAEYHLQYYDDLKEMASHCDILVPLCPLTDQTRN
LINADVLYHLPTHAGLLNLSRGAVVDTEALTHALEAKQFRYAILDTTAPEPLPKDHPLWSLPNCFIFPHFATNTM
AVRRALVDDIQPLVEEYFGLTCSDV

>Fragilariopsis cylindrus CCMP1102
MIRSDSGTELLNAPDLSVPFQISDFEVHGIGGGFVVVNHYGDNSLAAARVNIGVIGEDWLDIDSALLYLHRTADL
TNTKCTTTGLPLSGAGILLYFIDTSLDTEDNRKKLLKSLVYELQTFTKTKVLCLPANLTTNDLIHMYKTNELIEI
AKAADCEEKKNGSDGESDVEVSTFVPLSKTKKKSSFAICTNSFDFPVPKALITEDAIRDNVAAKSAAHSSFGSIV
TLTKAKQWTNARIVLLGVTPLSCELYQLLTRETDTVFLSDPDSSKYDVQHEQLPKTASKIPPRAFIPWDKARDLK
NNWDIIVFCSPVCPLLDASTIPTIRSKAIISVSDDFLPSDEEEREQILLALEKANVFEIADGMSDLGEIAKVYSL
SGNSTLSFNDAFELGSQVMQKKLHLHGIVTSDDVTAKRKFHDLMLHALEEDENARLFGLGTVMHQSSDRMTDWIW
VKASAMCPAFRALTSQNATVKPKGVNYLDMGAGAGAAARWICKQNKKIHVTCIDVCPKQSGENRSLSDEEGLGSQ

IDVVQGSYERLNSDYSNYFDGCMSQDAFIHAFVKHQAFSEALRVTKGGGWLLISDLMRGDGKDGDEEMEIFVKEH
NITDWATPNDCCQMARDAGWAEVRFIDCTAEINVSLHGLLKQIKTMMESGKFDGRNLQLLKTHRARLSSRIGQAD
RGIFKWGIISGRKPYDVVFMSEPPVSPEPRGMINYSVNSLDGDLKFGTDVLVVNIGEKLGYEKIMQLPSTTRLIV
TLSAGLDHICVKAANERGIRIRRAARDAIVKSVADYLLSNVIFGLRNGFQNVGVPFPGKNWNLTWNSEGVDLDCS
KIGFIGMGAIAIETAKRIRSLSKTCELVYHIPEDIRCRFEEGTHRMYHVGIADLLSTCDVVIPMVPLTETTSGLI
NYASFSMMKAIFINMARGKVVETSGMLRALDEGLVRHAILDTTDPEPLPVDHALWTMKNCTITPHFATNTTYVRR
ELVEDIPNQVEDTLEERGILRLEEQRMRLELSEAYRITREFGMDELVWNHISVLLSDGTFLITPGSAMFDDIGPE
DLVKSSGNITADIIHEAVYKTRSDIKAIVHLHTPATVAVSCLEMGFVPLAQEAAPFVGRVSRHPWHGVSNDREEQ
ALLGAAVKDPKFNTLLMENHGFCTFGKTLGEAWVLAYYFDKACQTQLSCLQTGQKINYPSEKVMANAAEQSVLPE
FLPGACEWEALRRMLTRKSRLHR

>Pseudo-nitzschia multistriata B856
MIRSESGTDLLLHSSLSVPFHIIDFEVNGIGGGFVVVNHYGDYGLSAARIDIGVLGDDWRDIDSALLYLHSTASL
TSTKCTTTGLNLGGAGVLLYFCSRNQFTDSNRKTLLAALVEELKGFTKSNILCLPRNLSNEDLTHMYKKNEIVTM
RHDNEGIDNAFVPLAKTKRKGSFAIATESFDYPIPKALITQEENLVAKSAAYSSFGSILTLMKAKEWTSARIVLL
GVTHLALELYQLLKRETDTVFLSNPHPSKEDLQSKIPSDAFISWDEAHKSKEEWDIMVFCSQSCPSLDESVISSI
RTKAVISVSDDLLPMDEKRREEVLLTLVKRDIFEAADGISDLGEIAKVFSLAGGTTFSFNDAFGMGCRVMEKKLH
LHGIVHSEDVTAKRKFHDLMLHALEKDENARLFGLGTVMHQSSDRMTAWLWSRARSMCPSYRAMSSKSATVKPEK
VSYLDMGSGNGAAARWICKQGKKIHITCIDVCAKQSYENRNISDEEGLGSQIDVVQGSYERLNSDYSNFFDGCMS
QDAFIHAFTKRSAFLEAFRVTKGGGWLLISDLMRGDGKDGNEEMEAFVKDHNITSWATPNECVQLATEAGWSEVR
FIDCTTEINVSLHGLLKKIKSMIKSGEYEGRNLQLLKTHRVRLSNRIGQADRGIFKWGIISGRKPYDVVFMSEPP
VSPESRSMINYSVNSLDGDLKFGTDVLVVNIGEKLDYEKIMELPSTTRLIVTLSAGLDHICVKAAEERGIRIRRA
AREAIVKSVADYLLSNVIFGLRNGFQNVGVPFPGASWDLTWNSDGVDLDCSKIGFIGMGAIAIETAKRIRSLSKS
CELVYHVPSDIRCRFEEGTHHMYHVGIADLLSTCDVIIPMVPLTTSTTGLINYSSFSMMKRTAIFINMARGKVVE
TAGMLRALDEGLVRHAILDTTDPEPLPPDHDLWNLKNCTITPHFATNTTYVRKELVEDIPNQVEDTLEERGILRL
EEQRMRHELSDAYRITREFGLDELVWNHISVLLSDGSFLITPGSRMFDDVGPEDLVKSSGNITADIIHEAVYKRR
SDVKAIVHLHTPASVAVSCLEMGFVPLAQEAAPFIGRVSNHPWHGVSNDREEQALLGEAVKNPKTNTLLMENHGF
CTFGKTLGEAWVLAYYFDKACQTQLSCLQTGQKIRYPSEDVLAHAAEQSVLPEFLPGACEWDALRKMLTRKQGRF
RR

>Pseudo-nitzschia multiseries CLN-47
MIRTESGIDLLQHSSLSVPFHIINFDVNDVGGGFVVVNHYGDFGLSAARIDIGVLGEDWTDINSALLYLHRTAGL
TSTKCNSTGLNLGGAGILLYFRSRDVCTDDNRKKLLTALVSELQSFTKTNILCLPTNLTMDDLVHMYRKNEILTV
KQADIEEKKAETDGAGGGGSGGSGSGSENNFTTLNKLKRKGSFAVAAECFEYPIPKALIGEENIAAKSAAHSSFG
SIVTLTKAKQWTNARIVLLGVTHLSLELYRLLTRETDTVFLSHPNQTQESIQDNMPTGTTLTIPPGAFLSWEEAR
STEEQWDIVVFCSPDCPTIDETSISSIRTKAVISVSNDFLPTDQKQREEVLLMLEKGDIFEVLDGISDLGEIAKV
FSFAGDKAFSFNDAFELGSQVMRKKLHLDGIVRSDDVTAKRKFHDLMLHALEKDENARLFGLGTVMHQSSDRMTA
WLWSRARSTCPAYRALSSKNATVKPDGVRYLDMGAGNGAAARWICKQSKKIHITCIDVCAKQSCENRNLSDELGL
GSQIDVVQGTYERLNSDYSNWFDGCMSQDAFIHAYTKRNAFMEAFRVTKGGGWILISDLMRGDGKDGDEEMEVFV
KEHNITNWATPNECVQFATEAGWSEVRFIDCTAEIVVSLHGLLKKIKTMMDSGDYEGRNLKLLQTHRSRLTNRIG
LADRGIFKWGIISGRKPYDVVFMSEPPVPPEPRSMINYSINSLDGDLKFGTDVLVVNIGEKLDYNKIMELPSTTR
LIVTLSAGLDHICVKAANERGIRIRRAAREAIVKSVADYLLSNVIFGLRNGFQNVGVPFPGASWDLTWNSDGVDL
DSSKVGFIGMGAIAIETAKRIRSLSKSCELVYHVPDDIRCRFEEGTHRMYHVGIADLLSTCDVIIPMVPLTETTT
GLINYSSFSMMKKSAIFINMARGKVVETAGMLRALDENLVRHAILDTTDPEPLPPDHELWRLKNCTITPHFATNT
TYVRKELVEDIPNQVEDTLEERGILRLEEQRMRVELSEAYRITREFGMDELVWNHISVLLSDGSFLITPGNRMFD
DIGPEDLVKSSGNITADIIHEAVYNTRSDVKAIVHLHTPATVAISCLEMGFVPIAQEAAPFIGRVSRHPWHGVSN
DREEQALLGAAVKDEKVNTLLMENHGFCTLGKTLGEAWVLAYYFDKACQTQLSCLQTGQKIKYPSEKVLAHAAEQ
SVLPEFLPGACEWQALRNMLTRKNRLRR

>Gynuella sunshinyii YC6258
MKQVSYEISSQVLEQYDSPQGRAFYRQVMGDSGFNIHYGIYPSENETMKTASENIIRHLQELAQQRGVHLPQASI
LDLGSGTGGAAHYLAGHFGCHVTCVNISPEQNKINRKQAQELGIDDLIKIEQCSFDNLPGKWSGQFDLVWSEEAF
CHAEHKDTVIKEAWRVLKPGGVLVFSDIMEGELNQDTHTFSDRNAIRDLASPSDYIRLCMANGFYHLSYHDLSHH
LPINFRKMIDQIDQHYDRLVDNGVSSKYADNFRQSLNDRVNAAFQGNFSWGSFVMNKSTRLEHPHLRSVIEGRNL
CRITAEPLTRENLAELGTLYAYDQPLEQHPPVPQHSWPVKYPQRMETGRGLAPLGTDDMTMTWQQEILHARNEVL
EHCYQNIAYRDEQHGYFIEWFNQHVESGQKFYCPGVPLLYVLAAPVDHPKAQDFKAFIADGSHGVIINPGVWHTN
PIPLIDTEVTLTTTQSIVDASCDCSLSAEHNQWLNITVSTGTDS

>Ulva lactuca
MSSAAENGAPAEGKSTVEKGVNEQYDTDASMAFYEYVMGGGGDDIHYGIFNADSDGLKESSQNSVEALAAMAEAC
GGLKKVTDGEPIKVLDLGSGKGGAARWLAKTYGCHVTCFNLGEKQNEFNKNKAVADGIGELIDTQLGSFNEPLPQ
KWTDQFDMVWSQEAFCHAMDHQALIAEVQRCLKPGGTAVFSDIMQSDNGGDCTSFTGQNVTTELASPQTYKDAIK
TTGMKLCEYKDLTHHLTRYFQCMLEAVKESKLPLQMRGVSLARLEAYEEDLVVRFDAVKQRSFAWGMFSCQNLFS
SDALLNGATYGDQCQDQMEENK

>Ulva lactuca (long)
MSSAAENGAPAEGKSTVEKGVNEQYDTDASMAFYEYVMGGGGDDIHYGIFNADSDGLKESSQNSVEALAAMAEAC
GGLKKVTDGEPIKVLDLGSGKGGAARWLAKTYGCHVTCFNLGEKQNEFNKNKAVADGIGELIDTQLGSFNEPLPQ
KWTDQFDMVWSQEAFCHAMDHQALIAEVQRCLKPGGTAVFSDIMQSDNGGDCTSFTGQNVTTELASPQTYKDAIK
TTGMKLCEYKDLTHHLTRYFQCMLEAVKESKLPLQMRGVSLARLEAYEEDLVVRFDAVKQRSFAWGMFSCQNLFS
SDALLNGATYGDQCQDQMEECINAQYDSETSMEFYKHAMGGGSDYCHYGVFNSPDDDLATATHNSVVLLSELARR
CSSLGLGLKNAVRVLDLGSGKGAAARHLAERFGCHVTCFNLGRNQNAHNIAEASAAGLEHLVNTVLGNFNSGLPA
EWTASFDVVWSQEALCHAKDQTQVLKEVNRVLKPGGVIIFTDIMRTMVSTANLAKMYSSSIDAGDVADAPTLALT
TRQPSSDSLASSSADRSPQRVGDSSTSSPSVPPSATPNTPSATSASRAINIGKAAVSSSPTPGGGFTRNLSVNGL
QTASISSSYGQSMLEGGSAMHDFHFNMGPKAQEAAEAAEAAEMLHVSSPGKYVQTLQAVGLELIDFRDLSEHLSV
FYDAMVDELVSNEAELTAAGISQSRLDAARVDLQHRVAAANHGALAWGVFTARKRSNASALIVGAGPVGACVAMR
MSQSGVANVTLKALSTAAPDVLTMARLAGVTVVEEWDDLVKTHWGSVIVAVKTHWLPAIALDMECTQITFENCCL
AYNGYTQTPEYFSSRRCCRAVVPQSFAFDSTAMLAWDIKHGNTQNPWNLPAGKASQSFVGQLRQSGIRAQCDAHF
HQAAFFKFCVNNSANLLSVIKTRCCRDLVENAECVEIMTKILNETFDVLELDPFYAALVPMHRKGYIDNVIFAGK
GSVASYSGHFPSSHQHYKAKQLVDTRALNGFVVRKARELNTPAPVNEYITRQVIAISNATAVSGIPREEWDIRVK
LSAAKRLIEGMGLGTSVWGHASARMPETSDRMLLAAFGESFGEATASSLRLTPTHPDALGKEGRDTTNITAVVLH
GAVYSQRPDAGCVLHTHSPYCTALAASSLEFDAEIIQDAMQFAGRVAYHAFGGVADENEEMEHVAKVAKDADVLF
LRNHGVIVVGPDVETAFSRLYYLERCCQTQLLVHGAGSKVHPAGADGVKHTVDWYATDGVESAHAEWAAHLRELL
HLEVSAGARAKPFWYE

>Chorda filum
MTAATEFTSAAADTTSKAAVESKVNEQYDTDASMAFYEYVMGGGGDDIHYGLFLTEKDGLKESSQNSVEALANMA
VDAGTLKKGENNAAVHCLDLGSGKGGASRWLAKEYGCKITAFNLGERQNAFNLERAQATGIGHLVETHLGSFNEP
LPAGWTDKYDMVWSQEAFCHCMDQKALMAEVSRVLKPGGVIVFSDIMQGDGGGDCTSFTGQNVVASMASPAMYKE
AMIGAGMSIVEHKELTSHLTPYFKCMLDAVKDGKATMLEQGVTQERLDAYEDDLSTRFERVKQGHFAWDMFCAKN

>Sargassum vulgare
MTAAAEFTSAAADTSNKTAVESKVNEQYDTDASMAFYEYVMGGGGDDIHYGLFLTEQDGLKESSQNSVEALANMA
VDAGTLKKGEDNGAVHCLDLGSGKGGASRWLAKEYGCKITAFNLGERQNTFNLERAQATGIGHLVETHLGSFNEP
LPADWTDKYDMVWSQEAFCHCMDQKALMKEVSRVLKPGGVIVFSDIMQGDGGGDCTSFTGQNVVASMASPQMYKD
AMTGAGMSILEHKQLTSHLTPYFKCMLDAVKDGKDTMLKQGVTQERLDAYEDDLSTRFERVKQGHFAWDMFCAKN

>Pyropia haitanensis
MTGTNGAASPTDTVLAQYNTDESAQFYSVIMGDGTAHVHYGVYTSPDDTVRQAVAATNTTLRVLAESAGAVFRPG
VRVLDLGAGNGGTAHALAAATGCSVTCLNLCEVQNAANEAAAAAEGLSSLITVVTGSFEDLPANWSGSFDVVWSQ
DAIVHSSDKGRVWAEAARVLVPGGYVVLSDIMAAPSAADAALRAFKKRLHVDELLTLDGYEAGLREAGVSTLRTR
DMSGQLLPNYRRMLERITTERARLTQCSDAYLEQYAGLLRDNIKVLCDGEAQAWCALVARKNGFAATSRTASAVA
P

>Ostreococcus tauri
MGCAPRGRRDASTDRRARRRGATWCDGRARDGDDGATTTTTTGIDAGFDRAQADFVGKQFDKVEVGKKVLEQYDD
AQQRAFYTVVMGGGGDDIHFGIYRAPGDGVRESSAATTEWMMTQLDMARPIGAGDRVLDVGSGHGGGSHALAKRF
GCKVLGYNIGPQQNAQNLAKAKELGLGDLVDAVVGDINQPFPADWTDSFDSVWSCEVLCHAGDKTELFKEIYRVM
KPGAAFVFSDIMGADGADEKTLKGFTDRNATTVMGRPSGYMQCIKDAGLDYVTWWDGSNHLETYFRDMINQIHTH
REEMLSKGITEQYLNNWLESLTERADTQRDKGVFAWGVFVCRKPLA

>Aidingimonas halophila
MNKVMEKAYTAADETAREYYNSDDADNFYFHVWGGEDIHVGLYRDDKEPVYDASRRTVAQMAARLPKLDATTQVL
DIGAGYAGSARYLAHEYGCSVTALNISEKENERGREKNRQQGVDDRVEIIDGSFEDIPFDNESFDVVWSQDAILH
SGDRERVLEEAVRVLRPGGHLIFTDPMQDDECPREALEPILARIHLETLGSPGFYRDTLRRLGLEEQDFDDHSDM
IAMHYGRIHQVLSEHEDDVGQHVSRGYIDRMKHGLQHWVEGGQRGNLRWGIFHFRKPAQ

>Nesiotobacter exalbescens DSM16456
MSKQNLSNAELTMRDQVYGDNPLEDRETDHYKKEYISTFVDKWDELIDWDGRAESEGQFFIDILRARGKERILDV
ACGTGFHSVRLMENGFDVTAADGSAAMVAKAFNNAQSRGLILKTVQADWRWLNRDVHGKYDAIICLGNSFTHLYE
ESDRRRALAEFYAALKHDGVLILDQRNYDAMLDRGFTTKHKYYYCGEQVTAEPDHVDEGLLRMRYSFPDGSEYTL
NMCPIRKNYLRRLLSEAGFERVRTYGDFQETYAEDDPDFFIHVAEKSALHLVRWGGTVAEGQEDIRDYTEDYYDS
DDAAVFYSTIWGGEDLHVGLYDTTQDIRSASDLTIDRMIDTLPPLTKDSHVLDMGAGFGGAMRRLVKKTGCQATC
LNISETQNEYNLQKIRQARLNDRIKVKHGVFEDVPFEDASFDVVWSQDAFLHSDQRNKVLAEALRVLKPGGSLIF
TDPMQADEVNVENLQPVYDRLRLNSMGSFRFYREAAETLGFETVGLDEMTHNMRAHYARVKQEIEKNYELLREKG
ASAEYLDKMLVGMDHWVNACDSGNLAWGILHFRKRA

>Porphyra umbilicalis
MTGTNSGNAPAQAPEAGDTTSTVLAQYDTPESAQFYSVVMGDGGSDVHYGVYTKASDSTRDAAENTIAVLRSLGE
SAGVTFGPGVKVLDLGAGNGGAAHKLAAETGCSVTCLNLCAEQNAANRSEVVARDLSDLVTVVTGSFEKLPADWT

ASFNIVWSQDAIVHSNAKDKVFAEAARVLVPGGHVVMSDIMAGPAADASVLSAFKERLHVDELLTPSAYEVGLAA
AGVTTLRTRDLSGQLVTNYRRMVGRISSERARLDRCSDAYLDTYAGLLGANIEALCNGEAQAWFALVGQKVGGRT
TAAAAADADATVAPSPPLVRRWVASKLHGVRVTDKSVRYHGSVGVSRNLMAAAGIKEFEAVDVVNLTNGARWTTY
ALGIDADDAFTLNGGGARRGETGDECVLMTYVQSAAYEPARVAYCGADNSIIDQFTYSHKEEAGENGSEGVAIGN
GSSNAATG

>Aureococcus anophagefferens
MSAAVQKTTGDFKNKQFSEEEVASKVVEQYDEVHARTFYKYVMGGGGYDIHYGMFRTASDGVFESSKNTNAALLR
LLDQTRPVTKDSVVLDLGSGHGGLSHEIATTFGCKVVSFNISPEQNNMNLEEAARLGVKELISVVEGNFNDAATF
PPKKLPKITHIVSCEVFCHAASKPALLSDIFKMLEPGGALVFTDIMGADGANEKALKDFTDRNATTEMARPSGYL
QQMKDAGFAHVGFFDGSGHLLPYFAAMLDVCLKQGDDMVKDGVPRPYLDKWIASLTDRVKIQGDEAVFAWGLFSA
RKPGPLY

>Moorea bouillonii PNG
MANLSVIDTFNQSYFDSKDMDLFYRRVSGEHTHCGIFEHPNEDVYIAKKRTTEYMTSLLTLDRDSHLLDLGSGYG
GAARYIAKEFGCQVSCINLSEQQNAINIDRNQEEGLSDLVHVHQGSFEKLPFSDSKFNAAWAQDSLFYSDTQLQA
FREAHRVLVKGGEFIACTYFFGGNYPSPEEVNKVVDWYTGGGIHKVYFLHIDDYRKVADEIGMAEVQVIDLTHHI
SVNYWQILKKMEEIQADEQLWSDEFFEKKKQRLLDCAEVGKSGLLQWGILHYRKEN

>Moorea producens PAL-8-15-08-1
MNTTYSEAVKTAQTYYDGAETDQLYATFWGGEHITYGIYKSSDEPIHDASKRTVETIAQTLENLAPDSRVIDLGA
GYGGAARYLAKTYGCSVCCLNLSERQNQRNRQLNQEQNLAHLVEVTQGSFEDIPYPDNSFNIVWSQDAILHSSDR
TQVFEEIKRVLQPGGELIFTDPMQKETCPPGLLQPAFDRLGIKDMGSYRFYSQTAQELGFEELHFIDLSENVPIH
YRRFGKEVRERYQEVVTMTSTELADKTLKSIEPWIEYYEKGYMQWGILHFRLR

>Moorea producens JHB
MTTTTANSEQATERARVHYNSENARNLYEIAIGSDTLNLGMYEDDPDRPIAEGMAKTTEWMGAQVQNLNPDFRVL
DLGSGYGPAARHLAGKYGCHVTCLNISEEQNQENERRNREQGLDNLIDIVYGNFKDIPFDDASFDLVWSQDALFH
SDGQDRVLEEAYRVLKPGGQLMFMDILQADDCPDGVLKDSLQRVNIHHGRLGSFHSYTSKAESLGFETINVIDKS
YQLLVHYTKLRDSVISHYDELSQKCTPEFLESSKNGLCQWVESAEKGMFTWGLFHYRRPNA

>Candidatus Kuenenia stuttgartiensis (WP_099323879.1)
MTYSRPETTTRDYYNDQSADRFYYSVWGGEDIHIGIYRYPSESIFAASRRTVETMCAFLPRLNASTRVLDIGAGY
GGSARYLAMQFQCEVTCLNLSDVQNQKNRQQTEEAGLDSLIEVVDGSFENIPFSPNSFEVVWSQDAILHSDDRKK
VLEEVHRVLKPGGVFIFTDPMQSDECPKEALAPVLARIHLESLGSPGVYRKIAAQLGWIELKFEDLSSQLPIHYR
RVLEEITGRETEIVRQCTPEYIERMKTGLGHWIKNGEKGYLKWGIFLFQKPDVFHDSFVGSRPVQ

>Candidatus Kuenenia stuttgartiensis (CAJ74884.1)
MTYSRPETTTRDYYNDQSADRFYYSVWGGEDIHIGIYRYPSESIFAASRRTVETMCAFLPRLNASTRVLDIGAGY
GGSARYLAMQFQCEVTCLNLSDVQNQKNRQQTEEAGLDSLIEVVDGSFENIPFSPNSFEVVWSQDAILHSDDRKK
VLEEVHRVLKPGGVFIFTDPMQSDECPKEALAPVLARIHLESLGSPGVYRQIAAQLGWIELKFEDLSSQLPIHYR
RVLEEITGRETEIVRQCTPEYIERMKTGLGHWIKNGEKGYLKWGIFLFQKPDVFQDSFVGSRPVQ

Appendix 19. Codon optimised sequences of candidate GBT synthases synthesised for functionality tests.

>CBN80020.1 Ectocarpus siliculosus
ATGCCTCCGACCGTTGTTGCAACCAGCGATAAAATGGATGAAATTAGCAGCGTGGCCAAAGATTATTATGATAGC
GATAACGCCTTCAACTTTTATCGTCAGGTTTGGGGTGGTGAACATATTCATGTTGGTCTGTATACCCGTCTGGAA
GGTGAAGATGCAGAACTGAAAGGTGTTGATCGTATTACCAAAGCAAGCAGCCTGTGTACCGAAGAACTGCTGAGC
CGTTGTTTTCCGGCAGGCGCAACCGCAACCAGTGATGGCGAAGATGCACCGCCTGCAAGCGAATGTACCATGATG
GATATGGGTAGCGGTTATGGTGGCACCGCACGTGCAGCAGCAAAAACCCTGGGTTGTAAAGTTGTTTGTATTAAT
GTTGCCAGCCGTGAGAACGAAATTAATGCAGCAATTACAAAAGATGCCGGTCTGGAAGATAAAGTTATTGTTCCG
GGTGAAAAGAGCTTTTTTGAAACCGGTATGCCGGATGCAAGCTGTGATGTTGTTACCAGCCAGGATGCACTGCTG
CATGCCGGTAGCGAACGTCATCGTGCACTGGGTGAAGCAGCACGTGTTCTGAAACCTGGTGGTCGTATGGTTTTT
ACCGATATTATGCAGAGCAAAGAAGCCGATCCGAAAGATCTGCAAGAAGTTTATCAGCGTATTCACCTGGATGAT
ATGGCAACACCGGAAAGCTATGCAAAATGGGGTAAAGCACATGGTCTGGAATTTGTCGAATTTGTGGATATGACC
GATAATCTGGCACTGCATTATGGTGCAGTTCGTGAAGTTCTGGTTAGCCGTCGTGGTAATCTGGATGGTGTTGAA
GATGGCTTTATTGATAATATGGCACGCGGTCTGGATGCATGGACCAGCGCAGCAGGTCGTGATCTGATTCGTTGG
GGTTGTCTGGTGTTTACCAAACCGGAAAAAGCCAGCAATACCCAGAATGCAGCCAATGCAATTGCCGCAACCGCG
AAAGGCTATTATGATTCAGATAATGCGTTTAACTTCTACCGCCAGGTGTGGGGAGGCGAACACATCCACGTGGGC
CTGTATAACAAACTGGAAGGCGAGGATGCCAAATTAGAAGGTGTGGATCGCATCACAAAAGCCTCAAGCCTGGGC
ACAGAGGAACTGCTGTCACGTTGCTTTCCAGCCGGTGCGACCAGCAGCAGCTCAAGCGGTGAAGGTGCACTGCCA
GCCAGCGGTTGTACAATGGTTGATATGGGTTCAGGTTATGGCGGTACAGCCCGTGTTGCAGCCAAAACATTAGGT
TGCAAAGTGGTGTGCATTAACGTTGCAAGTCGTGAAAATGGTGTTAATGCCGCACTGAATAAAGCCGAAGGTTTA
GAAGATATGGTGATTATTCCTGGCGAGAAATCCTTTTTTGAGACAGGCATGCCAGACGCATCATGTGATGTGGTG
ACCTCACAGGATGCCCTGTTACATGCAGGTTCAGAACGCCACCGTGCCTTAGGCGAAGCTGCCCGTGTGTTAAAA
CCAGGCGGTCGCATGGTGTTCACAGACATCATGCAGGCAGAACATGCAAAACCGGAAGATTTACAAGAGGTGTAT
CAGCGCATCCATTTAGATGACATGGCGACTCCGCAGGCCTACGCCAAATGGGGAGAAGCCCATGGCTTAGAGTTC
GTTGAGTTTGTTGACATGACCGCCAATATCGAAACCCATTATGGTAGCGTGCGTGATGTTTTAGTTTCACGCCGT
GGCAACCTGGACGGCGTGGAAGATGATTTCATCGACAACATGGCTCGTGGCCTGGACGCCTGGACCTCAGCAGCT
GGACGCGATCTGATCCGCTGGGGCTATTTAGTTTTTACAAAACCGCCTTGGGTTCAAGCAAGCAAATGTGCAGAA
GCAATGGATACGAAACCGTATGTTCAAGGCTGGAATATGAATGGTCGTCGTGTTGTTTTTACACCGGGTCTGCTG
GCAATTAGCTATGGCACCTGTAGCCTGTGTGTTGTTTGTGGTAGCTTAGAATTTGCACTGGTTGTTGCACGTTTT
CGTAGCGCACGTGTGCATTGTTTTTGTTTTAGCGGTCTGCCTCGTTGT

>83254 Thalassiosira oceanica
ATGCCTCCGAATACCACCGTTGAAAGCGAAAATGATCATGATGAAACCCAGCTGTGGAAAGATCTGAGCGTTGAT
AGCTGGAAAAATAGCGTGCGTAATAAAGATAGCGTGTTCTATGATCTGTATCAGGATACCCGTCGTCGTATTATT
GATGCATGTGGTAGCCATAGTTATGATGTTGTTGTTGAAGTTGGTTGTGGCACCGGTGAAGTTATTGGTTTTCTG
GATGGTACAGGTACAGATCGTATTGGCGTTGATATTAACGATGATTTTATCAGCCATTGCCGTGCCAATTATGTT
GCAGATGGTCTGGAATTTCATGTTGCCGATGCAATGAAACTGGATGAATGGTGGGCAAAAATGGGCTACGACAAA
AAATACAAAGCACCGCTGATGGTTTGTCCGAATAACACCATTATGATTATGCCGGAAGAAATTCGCGATACCGTT
ATTGAGCAGATGCGTACCGTTGCAGGTATTGAAGGTCGTATTGTTATTACCTATTGGAACGGTCGCATGTTTGCA
CATGGTGTGATGGGCTATTATAAGAAAAATAGTGATCTGTGCGGCACCTTTGATCTGACCGAAGAACATGTTGAT
TGGCAGAATAAACGTATTGAAACCCGCACCAGCTATAAAAGCGAATGGCCGACCAGCGAAGATGTTTGTCGTTGG
ATGGCAAGCCTGCATATTGATGTTGATATTGTTGAACCCGAGATGCAAGAAACCCTGGAAATTGATCATATTGCC
GAAGTTGGTATGGGTGTTTATCTGTGGCTGAAAGGTATTGCACCGCAGGATAGCAGCATTGGTAGCGCACGTGAT
TATTATGATAACAAAGATAGCCAGACCTTCTATCGTACCGTTTGGGGTGATCACAATACCCATATTGGTCGTCAT
GATCTGCTGGATAGCGATCCGGAAATGAAAGGTGCAGATCTGTGTGCACGTCTGCGTAAAGCACAGCAGATTCAA
GAGGACCTGTTTATGGAAAATGTGAATAGCCGTTATGGTGGTGCCAAAGTTCGTTGTATGGATATGGGTTGTGGT
TATGGCGGTCTGCTGCGTAATATGTGTGATCATGGTATGATTTGGAGCGCAGTTGGTGTGGATATTAGCGGTCGT
ATGATCGATGCAGCAGCACGTCATAGCAAAAGCTATGATAATATTCGTTTTCTGCGCGAGAGCTATATGCGTACC
AGCGTTAGCGAAGATAGCATGGATCTGTGTATTAGCACCGAAAGTTTTCTGCATGTTGGTCCGGGTAATCATGAA
GCAGTTCTGCGTGAAGCATGGCGTGTTCTGCGTCCTGGTGGTCGTCTGATTTTTACCGATATTGTTAGCCGTCCG
GATGCACCGCCTGAAGCAAGCGTTCTGTATGAACGTATTGGTCTGCAAAGTTTTCAGACCGTGGAAGGTTATTTT
GAAGTTGCCCGTAAATTAGGTTTTGGCGAACTGAACTTTGAAGATCATAGCGATAATGTGAGCGTTCATTATGGC
ACCGTTCTGGAAGCACTGGAAGAAATGTGGTCAAATGGCGAGATCGATATTAAAGAAGAAAGCAAAGATCGCATG
GTTGATGGTCTGAGCAAATGGCGTGACCTGGCACCGAGCTGTCTGCAATGGGGTATTGTTAGTATGCGTAAACTG
GATCAGACCGAACATAGCGTTATTAGTCGTCCGAGCAGCGTTGCAAGCCTGAATCGTCTGAATCTGAAATAA

>WP_071515314.1 Geitlerinema sp. PCC 9228

ATGGCAGATCCGACCACCTATAGCGAAGCAGTTAGCACCGCACGTGAATATTACAATAGCAGCAGCGCAGAAAAC
TTCTATAGCACCATTTGGGGTGGTGAAGATATTCATGTTGGTATCTATCAGAGCGAGAACGATACCATTTTTACC
GCAAGCCAGCGTACCGTTGAAAAAATTGCAGAACAGCTGAAACTGGCACCGAGCACCAAAGTTCTGGATATTGGT
GCAGGTTATGGTGGTAGCGCACGTTATCTGGTTAAAACCTATGGTTGTCCGGTTGATTGTCTGAATCTGAGCGAA
GTTCAGAATGAACGTAATCGCAAACTGAATCAAGAACAGGGTGTTGCCGATAAAATCGATGTGTATGATGGTGAT
TTTGAAGCACTGCCGATGAGTGATGCAAGCTATGATGTTGTTTGGTGTCAGGATAGCATTCTGCATAGCAGCAAT
CGTACAAAAGTGCTGGAAGAAGCATATCGCGTTCTGAAACCTGGTGGTGAATTTATCTTTACCGATCCGATGCAG
AGTGATGATTGTAATCCGGAAGAAATTCAGCCGGTTCTGGATCGTATTCATCTGCCGAGCATGGGTAGCGTTGGT
TTTTATCGTGAAGTTGCAGCACGTCTGGGTTTTGATGAAGTTCAGTTTATTGATATGAGCGAGCAGATCCCGAAT
CATTATGGTAGCGTGCTGAAAGCCGTGAACGAAAATTATGATGCAACCGTTAAACAGTGCGGTGAAGAATATGTT
GAAAATATGCGCAAAGGCCTGAATCACTGGGTGAATTTTGGTAAACAGGGTAAACTGAAATGGGGCATCCTGCAT
TTTCGTAAACGT

>WP_043147237.1 Mameliella alba
ATGAGCGATACCATGGCAGAAAATGCAGAACTGACCATGACCGATCAGGTTTATAGTGATGATCCGCTGGCAGAT
CGTGAAACCGATCATTATCGTAAAGAATATGTGCGCACCTTCGTGGATAAATGGGATGAACTGATTGATTGGGCA
GGTCGTGCAGAAAGCGAAGGTCAGTTTTTTATCGATATTCTGCGTGCACGTGGTAAAGAAACCGTTCTGGATGTT
GCATGTGGCACCGGTTTTCATAGCGTTCGTCTGACCGAAGCAGGTTTTGATGTTACCGCAAGTGATGGTGCAGCA
AGCATGGTTGCAAAAGCATTTGAAAATGGTCGTAGCCGTGGTCTGATTCTGAAAACCGCACAGGCAGATTGGCGT
TGGCTGAATCGTGATATTAATGGTAAATATGATGCCATTGTGTGCCTGGGTAATAGTTTTACCCATCTGCATAAA
GAAAGTGAACGTCGTCGTGCACTGGCAGAATTTTATGCAGCACTGAAACATGATGGTCTGCTGATTCTGGATCAG
CGTAATTATGATGCAATGCTGGATCGTGGTTATACCACCAAACACAAATACTATTATTGCGGTGATCAGGTTACC
GCAGAACCGGATCATGTTGATGAAGGTCTGTGTCGTATGCGTTATAGCTTTCCGGATGGTAGCGAATATACCCTG
AATATGTGTCCGATTCGCAAAAACTATATGCGTCGTCTGCTGAGTGAAGCTGGCTTTGAACGTGTTCGTACCTAT
GGTGATTTTCAAGAAACCTATGCAGAAGATGATCCGGATTTTTTCATTCACGTTGCAGAAAAAAGCGCCATGCAT
CTGGTTCGTTGGGGCACCGGCACCGATGATGGTAAACAGGATATTCGCGATATTACCGAGAACTATTATGATTCC
GATGATGCCGATACCTTCTATAGCACCATTTGGGGTGGTGAAGATCTGCATATTGGTCGTTATGCAGAAACACCG
GATATTCGTACCGCATCAGATCGTACCATTGAAAGCATGATTGATCGTCTGCCTCCGCTGGATGCAAGCGCACGT
GTTCTGGATATGGGTGCAGGTTATGGTGGTGCAATGCGTACCCTGGTTCGTAAAACCGGTTGTGAAGCAGTTTGT
CTGAATCTGAGCGAAACCCAGAATGAATATAACCTGGGTAAAATTCGTGCAGCAAAACTGAGCGATAAAATCAGC
GTGCGTCATGGTGTTTTTGAAGATGTGCCGGAACCGGATGCCAGCTGTGATGTTGCCTGGTCACAGGATGCATTT
CTGCATTCAGATCAACGTGCAAAAGTTCTGGCCGAAGCATGGCGTGTTCTGAAACCTGGTGGTCATCTGATTTTT
ACCGATCCGATGCAGGCAGATGATGCAGATCCGGCAGTTTTACAGCCGGTTTATGATCGTTTACAGCTGAATGAT
CTGGGTAGTCCGCGTTTTTATCGTGAAGCAGCAGAAGCACTGGGTTTTGAAACCGTTGAACAAGAAGAAGCAGTT
AGCGATCTGCGTACCCATTATTTTCGTGTTCGTGAAGAACTGCTGGCCAACTATGAAAAACTGCGTGAAGCCGGT
GCAAGTGCCGAATATCTGGATAAAATGGCAGTTGGTCTGGAAAATTGGGTTAAAGCAGCGGATGATGGTCACCTG
GCATGGGGTATTCAGCATTTTCGTAAACCGGCA

>WP_096366830.1 Thiohalobacter thiocyanaticus
ATGAGCACCACACCGAGCTATAGCGAAGTTGTTGAAACCGCACGTAGCTATTATAATTCAGATGATGCCGACAAC
TTCTACTTTCATGTTTGGGGTGGTGAAGATATTCATATTGGTCTGTATCAGGATGACCAAGAACCGATTGCAGAT
GCAAGCCGTCGTACCGTTGAACGTATTGCAAGCAAACTGGATAATCTGGGTCCTGATAGCTATGTTCTGGATGTT
GGTGCAGGTTATGGTGGTGCAGCACGTTATCTGGCAAAAACCTATGGTTGTCGTGTTGTTGCACTGAATCTGAGC
GAAAAAGAAAATGAACGTGATCGTCAGATGAATCGTGAACAAGGTCTGGATCATCTGATTGAAGTTGTGGATGGC
AATTTTGAAAACCTGCCGTATGATAGCGGCACCTTTGATGTTGTTTGGAGCCAGGATAGCTTTCTGCATAGCGGT
CATCGTGATAAAGTTATTAGCGAAGCAGCACGTGTTCTGAAACCTGGTGGTGAACTGATTTTTACCGATCCGATG
CAGGCAGATGATTGTCCGGAAGATGTTCTGCAACCGGTTTATGATCGTATTCATCTGTCAAGCCTGGGTAGCATT
GGTTTTTATCGCGAACAGGCAGCACGTAATGGTCTGAAAGAAATTGAAATTGAAGATCTGACCGCACGTCTGCCG
GTTCATTATGGTCGTGTTCGTGAAGAACTGACCCGTAATCGTGATGATCTGGTTAATCGTGTTAGCGAAGAATAT
ATTGATCGCATGATTACCGGTTTAGGCCATTGGGTTGATGCAGGTAGCAAAGGTTATCTGAGCTGGGGTATTCTG
CATTTTCGTAAAACCGCAGAT

>AOL23288.1 Erythrobacter litoralis
ATGACCAGTGATGCAATGACCGGTGAAGCCGGTGTTGCAGTTGCACGTGATTATTATGATAGCTCAGATGCCATC
GAGTTCTATAGCACCATTTGGGGTGGTGAAGATATTCATGTGGGCATTTATGATGATACCCGTGATATTCGTGAA
GCAAGCGCAACCACCGTTGATACCATGGCACGTCTGCTGAGCAGCGTTACCGGTGCAAGCGAACGTGTTGCAAGT
GATCGTGGTGCCAGCGAAACCGGTAGCAATCTGAGCCAGGCACATGTTCTGGATATTGGTAGCGGTTATGGTGGT
GGTGCACGTCGTCTGGTTAGCAAACATGGTGCAGGTCATGTTACCTGTCTGAATATTGCAGAAGAAGAAAACGCA
CGTAATCGTAAACTGACAGGTGAAGCAGGTCTGCTGGATCGTATTGATGTTGTTGATGGTAGCTTTGATGCACTG
CCGTTTGATGATGGCGCATTTGATGTGGTTTGGAGCCAGGATGCAATTCTGCACGCACCGGATCGTGGCGCAGTT
CTGGATGAAGTTGCCCGTGTTCTGAAACCTGGTGGTCGTTTTATCTTTACCGATCCGATGCAGGCAGATGGTCTG
AGCGATCCGAGCGCACTGCAACCGATTTATGATCGTATTCATCTGCAAAATCTGGCCAGCTTTGGTTTTTATCGT
GAAGGTCTGAAAGCACGCGGTTTTCAAGAAGTTGAAGTTCAGGATCGTAGCCGTCAGCTGCGTAATCATTATGCA
CGTGTTGCCGAAGAACTGGATAGCCGTCGTGGTGAACTGAGCGCAAGTGATGAATTTGTTGATCGTATGATTGCC
GGTTTAGGTCATTGGGTTCGTGGTGCGGATGAAGGTCGTCTGACCTGGGGTATTATGCTGTTTCGTCGTATT

>OEU23352.1 F. cylindrus
ATGAAAAGCAACGATCTGAGCAAAACCGGTAGCAGCCATATTCGTTGTGATAGCCTGGCAATTTTTGTGTGGTTT
GATAAAAGCAATACCAGCAACGCCAAAGGCTATTATGATAGTGATGATGCCCAGAACTTCTATAACAAAATTTGG
GCAGGCGAAGATGAACTGTTTCATGTTGGTCGTTATGATCTGCTGACCAGCGAAGAAAAAAATGTGTTTCCGATT
GAACAGCAGATTCGTCGTGCCGAAGAACTGCATGAACTGGAACTGATGAACAAAATTCGTCATTATTGTCTGCCG
AGCAGCGAAACCAATTTTGGTCTGCGTGTTGTTGATTTAGGTTGTGGTTATGGTAGCCTGCTGCGTCGTCTGTAT
AAAGAAGGTATGATTTGGAAAGGCACCGGTGTTGATATTAGCCTGAAAATGTGTAAAGAAGCCCGTAAACGTAAT
CAGATTGTTGCAGCAAGTAACACCATTGAAATTCTGGAACAGAGCTATCTGCAAATTAGCGTTGGTAATGAAAGC
GTGGATGTTGTGATTAGCATGGATAGTCTGCTGCATGTGGGTCCTGAACGTCAGCGTGCAGTTATTAAAGAAGTT
TATCGTATGCTGCGTCCTGGTGGTTGGATGATTTTTAGCGATATTATGCAGAACGATAACGCCGATAGCAATGAA
ATGCAGCCGATTTATGATCGTCTGAACCTGAGCAATATGGGCACCATTAGCAACTATAAAAGCGCACTGGAAGAA
AATGGCTTTGCCAATTTTACCACCGATCTGCATAGCGATAACATCAGCAAACATTATGGTCACGTTCTGGACATC
ACCCAGAAAAAAGGTCATGAAATTGGTCTGAGCGACACCTATATCAAAAAAGCAGAAGTTGGTCTGAAAGTCTGG
AAAGAAAATAGTCCGGGTAATATTGTGTGGGGTATTATTGCAGCACAGAAAACCAACAAAATT

>P. tricornutum_20301
ATGGTTCCGCAGATTAACAAAGTGAGCGAAGAGAAAAAATACGAAGAGTTCGATGAACAGAAACTGTGGGCAGAA
ATGGACCCGCTGGCATGGCATTATAGCATTGCACGTAATAGCGTTTTCGAGGATACCTATGAACTGACCCGTACA
CAGATTCTGGATGCAGCAGAACGTGGTGGTCATGATGTTATTCTGGAAGCAGGTTGTGGCACCGGTGATATTATT
GGTGAACTGCAAACCGATATTCATCGTATTGGCGTGGATATTAACGACCGCTTTATTGAGCATTGCAAAAAACAT
CATCAGCACGAAAACATGGAATTTCACGAACTGGATGTGACCAATTTAGGTCAGTGGTGGAAACAGTTTGAGGGC
AAATTCAAAAAACCGCTGGTTATTTGCGTGAACAACACCCTGAATATCATGCCGGAAGAAATTCGTGGTAATGTT
ATTGCCCAGATGTTTGAAGTTTGTGGTAGCGAAGGTCGTTGTCTGGTTACCTATTGGAATGGTAACTTTTTTAGC
CACGCCATCATGAACTTCTATAGCCAGAATGTTGAACTGTGTGGTCCGTTTGATTTTAGCCATGTGGATTGGAAT
AATCGTACCCTGCACGCACCGAGCGGTTATAGCACCCATTGGATGCTGCCTGAAGAAGTTCAGCGTCTGCTGCGT
AGCTATGATGTTAATATTGGTCTGATTGGTGCCGAAATTCAGCATGGTAAAGATCATATCAATACCGCAGGTCTG
AGCATTTTTGCATGGTTTAGCGCAGATTGTAGCTGTGGTGGTAAAAGCTATTATGATAGCGAAGATGCCCAGACC
TTTTATAGTCAGATTTGGGGTGATGCAGAAACCCATATTGGTCGTTATGATCTGCTGACCGATGCCGAAAAAAGC
AGCCTGACCAAAATCCAGCAGATTAGCCGTGCCGAAGAAATCCATGAAGAAAACTTTGTGAAACTGATCGCCAGC
AAATTTCGTAGCACCAGTGATGATATTCCGAAAGTTCGTATTCTGGATATGGGTTGTGGTTATGGTGGCCTGCTG
CGTCGTCTGTGGAAAGGTGGCCATGTTTGGCGTGCCACCGGTTGTGATATTGCAAGCAAAATGTGTGGTAAAGCC
CGTGTTCTGAATACCGAAATGGGTGCAGATCAGGATATTGCCATCCTGGAAGAAAGCTATCTGGGTGTTAGCGTT
CCGGATGAAAGCGTTGATCTGGTGATTAGCATGGATGCACTGCTGCATGTTGGTCCGGAAGGTCAGAAAACCGCA
ATTAAAGAAGCAGCACGCGTTCTGCGTCCTGGTGGTTGGATGGTTTTTTGCGATATTATGCAGCAAGAAGTCGTT
GATCCGGTTGAAATGCAGCCGATTTATGATCGTATTCATCTGACCAAACTGGGCACCGTTAGCAATTATCAAGAA
TGTCTGAGCGAAAGCGGCTTCACCAAATTTGAGTTTGAACCGCATAGCGAAAATGTGGCAAGCCATTATGGCACC
GTTCGTGAAGTTCTGATTGAAAAAAAAGGCGATATTGCAGTCAGCGAAGCCTTTCTGAATAAAATGGAAGCCGGT
CTGGCAGTTTGGAAAGAACTGGCACCGCAGAATATTGTTTGGGGTTTTATGACCGCACAGAAAACGGAAAAAGTG
AACATCTAA

Appendix 20. Predicted secondary structure of A) double-domain methyltransferase from *M. alba*, B) double-domain methyltransferase from *T. pseudonana*, C) single domain methyltransferase from *E. litoralis*. Figure adapted from Phyre2[157].

A

B

C

Appendix 21. Amino acid sequences of functional, non-functional and candidate GBT synthases used to produce a Maximum Likelihood phylogenetic tree to visualised the relatedness.

>CBN80020.1 putative sarcosine-dimethylglycine methyltransferase Ectocarpus siliculosus
MPPTVVATSDKMDEISSVAKDYYDSDNAFNFYRQVWGGEHIHVGLYTRLEGEDAELKGVDRITKASSLCTEELLS
RCFPAGATATSDGEDAPPASECTMMDMGSGYGGTARAAAKTLGCKVVCINVASRENEINAAITKDAGLEDKVIVP
GEKSFFETGMPDASCDVVTSQDALLHAGSERHRALGEAARVLKPGGRMVFTDIMQSKEADPKDLQEVYQRIHLDD
MATPESYAKWGKAHGLEFVEFVDMTDNLALHYGAVREVLVSRRGNLDGVEDGFIDNMARGLDAWTSAAGRDLIRW
GCLVFTKPEKASNTQNAANAIAATAKGYYDSDNAFNFYRQVWGGEHIHVGLYNKLEGEDAKLEGVDRITKASSLG
TEELLSRCFPAGATSSSSSGEGALPASGCTMVDMGSGYGGTARVAAKTLGCKVVCINVASRENGVNAALNKAEGL
EDMVIIPGEKSFFETGMPDASCDVVTSQDALLHAGSERHRALGEAARVLKPGGRMVFTDIMQAEHAKPEDLQEVY
QRIHLDDMATPQAYAKWGEAHGLEFVEFVDMTANIETHYGSVRDVLVSRRGNLDGVEDDFIDNMARGLDAWTSAA
GRDLIRWGYLVFTKPPWVQASKCAEAMDTKPYVQGWNMNGRRVVFTPGLLAISYGTCSLCVVCGSLEFALVVARF
RSARVHCFCFSGLPRC

>jgi|Thaoce1|83254|rna11146
MPPNTTVESENDHDETQLWKDLSVDSWKNSVRNKDSVFYDLYQDTRRRIIDACGSHSYDVVVEVGCGTGEVIGFL
DGTGTDRIGVDINDDFISHCRANYVADGLEFHVADAMKLDEWWAKMGYDKKYKAPLMVCPNNTIMIMPEEIRDTV
IEQMRTVAGIEGRIVITYWNGRMFAHGVMGYYKKNSDLCGTFDLTEEHVDWQNKRIETRTSYKSEWPTSEDVCRW
MASLHIDVDIVEPEMQETLEIDHIAEVGMGVYLWLKGIAPQDSSIGSARDYYDNKDSQTFYRTVWGDHNTHIGRH
DLLDSDPEMKGADLCARLRKAQQIQEDLFMENVNSRYGGAKVRCMDMGCGYGGLLRNMCDHGMIWSAVGVDISGR
MIDAAARHSKSYDNIRFLRESYMRTSVSEDSMDLCISTESFLHVGPGNHEAVLREAWRVLRPGGRLIFTDIVSRP
DAPPEASVLYERIGLQSFQTVEGYFEVARKLGFGELNFEDHSDNVSVHYGTVLEALEEMWSNGEIDIKEESKDRM
VDGLSKWRDLAPSCLQWGIVSMRKLDQTEHSVISRPSSVASLNRLNLK*

>WP_071515314.1 methyltransferase domain-containing protein Geitlerinema sp. PCC 9228
MADPTTYSEAVSTAREYYNSSSAENFYSTIWGGEDIHVGIYQSENDTIFTASQRTVEKIAEQLKLAPSTKVLDIG
AGYGGSARYLVKTYGCPVDCLNLSEVQNERNRKLNQEQGVADKIDVYDGDFEALPMSDASYDVVWCQDSILHSSN
RTKVLEEAYRVLKPGGEFIFTDPMQSDDCNPEEIQPVLDRIHLPSMGSVGFYREVAARLGFDEVQFIDMSEQIPN
HYGSVLKAVNENYDATVKQCGEEYVENMRKGLNHWVNFGKQGKLKWGILHFRKR

>WP_043147237.1 methyltransferase domain-containing protein Mameliella alba
MSDTMAENAELTMTDQVYSDDPLADRETDHYRKEYVRTFVDKWDELIDWAGRAESEGQFFIDILRARGKETVLDV
ACGTGFHSVRLTEAGFDVTASDGAASMVAKAFENGRSRGLILKTAQADWRWLNRDINGKYDAIVCLGNSFTHLHK
ESERRRALAEFYAALKHDGLLILDQRNYDAMLDRGYTTKHKYYYCGDQVTAEPDHVDEGLCRMRYSFPDGSEYTL
NMCPIRKNYMRRLLSEAGFERVRTYGDFQETYAEDDPDFFIHVAEKSAMHLVRWGTGTDDGKQDIRDITENYYDS
DDADTFYSTIWGGEDLHIGRYAETPDIRTASDRTIESMIDRLPPLDASARVLDMGAGYGGAMRTLVRKTGCEAVC
LNLSETQNEYNLGKIRAAKLSDKISVRHGVFEDVPEPDASCDVAWSQDAFLHSDQRAKVLAEAWRVLKPGGHLIF
TDPMQADDADPAVLQPVYDRLQLNDLGSPRFYREAAEALGFETVEQEEAVSDLRTHYFRVREELLANYEKLREAG
ASAEYLDKMAVGLENWVKAADDGHLAWGIQHFRKPA

>WP_096366830.1 methyltransferase domain-containing protein Thiohalobacter thiocyanaticus
MSTTPSYSEVVETARSYYNSDDADNFYFHVWGGEDIHIGLYQDDQEPIADASRRTVERIASKLDNLGPDSYVLDV
GAGYGGAARYLAKTYGCRVVALNLSEKENERDRQMNREQGLDHLIEVVDGNFENLPYDSGTFDVVWSQDSFLHSG
HRDKVISEAARVLKPGGELIFTDPMQADDCPEDVLQPVYDRIHLSSLGSIGFYREQAARNGLKEIEIEDLTARLP
VHYGRVREELTRNRDDLVNRVSEEYIDRMITGLGHWVDAGSKGYLSWGILHFRKTAD

>AOL23288.1 sarcosine/dimethylglycine N-methyltransferase Erythrobacter litoralis
MTSDAMTGEAGVAVARDYYDSSDAIEFYSTIWGGEDIHVGIYDDTRDIREASATTVDTMARLLSSVTGASERVAS
DRGASETGSNLSQAHVLDIGSGYGGGARRLVSKHGAGHVTCLNIAEEENARNRKLTGEAGLLDRIDVVDGSFDAL
PFDDGAFDVVWSQDAILHAPDRGAVLDEVARVLKPGGRFIFTDPMQADGLSDPSALQPIYDRIHLQNLASFGFYR
EGLKARGFQEVEVQDRSRQLRNHYARVAEELDSRRGELSASDEFVDRMIAGLGHWVRGADEGRLTWGIMLFRRI

>OEU23352.1 S-adenosyl-L-methionine-dependent methyltransferase Fragilariopsis cylindrus CCMP1102
MKSNDLSKTGSSHIRCDSLAIFVWFDKSNTSNAKGYYDSDDAQNFYNKIWAGEDELFHVGRYDLLTSEEKNVFPI
EQQIRRAEELHELELMNKIRHYCLPSSETNFGLRVVDLGCGYGSLLRRLYKEGMIWKGTGVDISLKMCKEARKRN
QIVAASNTIEILEQSYLQISVGNESVDVVISMDSLLHVGPERQRAVIKEVYRMLRPGGWMIFSDIMQNDNADSNE

MQPIYDRLNLSNMGTISNYKSALEENGFANFTTDLHSDNISKHYGHVLDITQKKGHEIGLSDTYIKKAEVGLKVW
KENSPGNIVWGIIAAQKTNKI

>jgi|Phatr2|20301|estExt gwp gw1.C chr 80194
MVPQINKVSEEKKYEEFDEQKLWAEMDPLAWHYSIARNSVFEDTYELTRTQILDAAERGGHDVILEAGCGTGDII
GELQTDIHRIGVDINDRFIEHCKKHHQHENMEFHELDVTNLGQWWKQFEGKFKKPLVICVNNTLNIMPEEIRGNV
IAQMFEVCGSEGRCLVTYWNGNFFSHAIMNFYSQNVELCGPFDFSHVDWNNRTLHAPSGYSTHWMLPEEVQRLLR
SYDVNIGLIGAEIQHGKDHINTAGLSIFAWFSADCSCGGKSYYDSEDAQTFYSQIWGDAETHIGRYDLLTDAEKS
SLTKIQQISRAEEIHEENFVKLIASKFRSTSDDIPKVRILDMGCGYGGLLRRLWKGGHVWRATGCDIASKMCGKA
RVLNTEMGADQDIAILEESYLGVSVPDESVDLVISMDALLHVGPEGQKTAIKEAARVLRPGGWMVFCDIMQQEVV
DPVEMQPIYDRIHLTKLGTVSNYQECLSESGFTKFEFEPHSENVASHYGTVREVLIEKKGDIAVSEAFLNKMEAG
LAVWKELAPQNIVWGFMTAQKTEKVNI*

>XP_002286764.1 predicted protein [Thalassiosira pseudonana CCMP1335]
MAPNTSTSTASFQTQESNGSLAHDETQLWKDLSVESWKNSVMNKDSVFYDLYNDTRRRIIEACHKHSYDVVVEVG
CGTGEVIGFLDGTDTPRIGVDINPDFINHCKATYKGDNLEFHVADAMKLDEWWASMGYDKKYKAPLMVCPNNTIM
IMPEEIRDTVIDQMRVVAGIEGRIVVTYWNGKMFAHGVMGFYRKNSDLCGTFDLTEEHVDWDSKKIETRTSYKSE
WPTADDVARWMAALLIDVVIAEPQIVETPEIDHVAEVGMGVYLWLKGISPENSSTGSARDYYDDKDSQTFYRTVW
GENNTHIGRHDLLEQDPEYTDADLCTKLRKAQQLQEEVFMEYVKTFYGESKVRCLDMGCGFGGLLRNMVKKDMIW
SAVGVDISGEMIDAATRLSKGIESLTFYRESYMNTSVPDEGIDLCVSTEAFLHVGPGNHEAVLREAWRVLRPGGR
LIFTDIVGLPSSPPEAKVLYQRIGLQSFQTVDGYFEVAKKLGFGELTFEEHSANVSTHYGTVLEALEELWEKKEI
DIKQESKDRMVDGLTKWRDLAPSCLQWGVISMRKIEQTEHSVVDESPSS

>sp|Q9KJ20|GSDMT_ACTHA Glycine/sarcosine/dimethylglycine N-methyltransferase
OS=Actinopolyspora halophila OX=1850 PE=1 SV=1
MTKSVDDLARGDQAGDEQDPVHREQQTFGDNPLEVRDTDHYMHEYVGGFVDKWDDLIDWKKRYESEGSFFIDQLR
ARGVETVLDAAAGTGFHSVRLLEEGFETVSADGSPQMLAKAFSNGLAYNGHILRVVNADWRWLNRDVHGEYDAII
CLGNSFTHLFSERDRRKTLAEFYAMLKHDGVLIIDQRNYDSILDTGFSSKHTYYYAGEDVSAEPDHIDDGLARFK
YTFPDKSEFFLNMYPLRKDYMRRLMREVGFQRIDTYGDFQETYGEDEPDFYIHVAEKSYRTEDEFVDMYSNAVHT
ARDYYNSEDADNFYYHVWGGNDIHVGLYQTPQEDIATASERTVQRMAGKVDISPETRILDLGAGYGGAARYLART
YGCHVTCLNLSEVENQRNREITRAEGLEHLIEVTDGSFEDLPYQDNAFDVVWSQDSFLHSGDRSRVMEEVTRVLK
PKGSVLFTDPMASDSAKKNELGPILDRLHLDSLGSPGFYRKELTRLGLQNIEFEDLSEYLPVHYGRVLEVLESRE
NELAGFIGEEYRAHMKTGLRNWVQAGNGGSLAWGIIHARA

Research articles published during the duration of the doctoral training.

During the duration of this study, I contributed towards the following publications:

Appendix 22. List of research articles published as part of the doctoral training.

- **Bacteria are important dimethylsulfoniopropionate producers in coastal sediments.** Beth T. Williams, Kasha Cowles, **Ana Bermejo Martínez**, Andrew R. J. Curson, Yanfen Zheng, Jingli Liu, Simone Newton-Payne, Andrew J. Hind, Chun-Yang Li, Peter Paolo L. Rivera, Ornella Carrión, Ji Liu1, Lewis G. Spurgin, Charles A. Brearley, Brett Wagner Mackenzie, Benjamin J. Pinchbeck, Ming Peng, Jennifer Pratscher, Xiao-Hua Zhang, Yu-Zhong Zhang, J. Colin Murrell and Jonathan D. Todd. *Nature Microbiology*, p1-11 (2019)
- **Author Correction: DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton.** Andrew R. J. Curson, Beth T. Williams, Benjamin J. Pinchbeck, Leanne P. Sims, **Ana Bermejo Martínez**, Peter Paolo L. Rivera, Deepak Kumaresan, Elena Mercadé, Lewis G. Spurgin, Ornella Carrión, Simon Moxon, Rose Ann Cattolico, Unnikrishnan Kuzhiumparambil, Paul Guagliardo, Peta L. Clode, Jean-Baptiste Raina & Jonathan D. Todd. *Nature Microbiology*, vol 4, p540–542 (2019)
- **DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton.** Andrew R. J. Curson, Beth T. Williams, Benjamin J. Pinchbeck, Leanne P. Sims, **Ana Bermejo Martínez**, Peter Paolo L. Rivera, Deepak Kumaresan, Elena Mercadé, Lewis G. Spurgin, Ornella Carrión, Simon Moxon, Rose Ann Cattolico, Unnikrishnan Kuzhiumparambil, Paul Guagliardo, Peta L. Clode, Jean-Baptiste Raina & Jonathan D. Todd. Nature Microbiology, vol 3, p430–439 (2018)

- **Sensing iron availability: Via the fragile [4Fe-4S] cluster of the bacterial transcriptional repressor RirA.** Ma Teresa Pellicer Martinez, **Ana Bermejo Martinez**, Jason C. Crack, John D. Holmes, Dimitri A. Svistunenko, Andrew W. B. Johnston, Myles R. Cheesman, Jonathan D. Todd and Nick E. Le Brun. *Chemical Science*, 8, 8451-8463 (2017).
- **Dimethylsulfoniopropionate biosynthesis in marine bacteria and identification of the key gene in this process.** Andrew R. J. Curson, Ji Liu, **Ana Bermejo Martínez**, Robert T. Green, Yohan Chan, Ornella Carrión, Beth T. Williams, Sheng-Hui Zhang, Gui-Peng Yang, Philip C. Bulman Page, Xiao-Hua Zhang & Jonathan D. Todd. *Nature Microbiology*, vol 2, 17009 (2017)