AN ABSTRACT OF THE DISSERTATION OF

Jing Sun for the degree of Doctor of Philosophy in Molecular and Cellular Biology presented on February 17, 2014

Title: Metabolism of One Carbon (C1) and Methylated Compounds in SAR11 Bacteria

Abstract approved:

Stephen J. Giovannoni

SAR11 *Alphaproteobacteria* are the most abundant aerobic chemoheterotrophs in ocean surface waters. Previous studies have indicated SAR11 cells play an important role in marine carbon cycling and consume up to half of some common dissolved organic compounds, such as amino acids. During sequencing of the first SAR11 genome, genes for the oxidation of one carbon (C1) and methylated compounds were observed, which is very uncommon in proteobacteria, with the exception of methylotrophic species. The goal of this research was to study the metabolism of C1 and methylated compounds in SAR11, and to explore the implications of this for understanding the cycling of dissolved organic matter (DOM) in the oceans. Metabolic reconstruction from genome sequences and physiological experiments showed that SAR11 can utilize a variety of C1 and methylated compounds. The data showed that SAR11 strain HTCC1062 can oxidize methanol,

formaldehyde, methylamine and methyl groups from glycine betaine (GBT), trimethylamine (TMA), trimethylamine N-oxide (TMAO), and dimethylsulfoniopropionate (DMSP). Interestingly, in addition to demethylating DMSP, strain HTCC1062 cleaved DMSP to the climatically active gas dimethyl sulfide (DMS) using a novel DMSP lyase. We propose that SAR11 cleaves DMSP to DMS when intracellular DMSP concentrations rise above a threshold, but below that threshold DMSP is channeled through the demethylation pathway to produce methanethiol that is used as a source of sulfur for growth. Investigation of transcriptional responses to five C1 compounds and methylated compounds revealed only a single case in which a gene was upregulated. The gene annotated as monomeric sarcosine oxidase (SAR11_1304) was upregulated in response to TMAO addition. SAR11_1304 was cloned and overexpressed in *Escherichia. coli,* and in vitro enzymatic assays showed that this protein has TMAO demethylase and dimethylamine monooxygenase activity, indicating it functions in the degradation pathway for TMAO. Overall, the data presented here reveal diverse metabolism in SAR11 for utilization of a variety of C1 and methylated compounds. This research provides insight into the metabolism of SAR11 and enhances understanding of the broad success of this ecologically significant organism. This work also indicates a surprisingly large role for C1 and methylated compounds in marine systems, and shows that broad insights into carbon cycling by marine microbes can be obtained by physiological experimentation based on genome predictions.

Metabolism of One Carbon (C1) and Methylated Compounds in SAR11 Bacteria

by
Jing Sun

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented February 17, 2014
Commencement June 2014

Doctor of Philosophy dissertation of Jing Sun presented on February 17, 2014.

APPROVED:

Major Professor, representing Molecular and Cellular Biology

Director of the Molecular and Cellular Biology Program

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes the release of my dissertation to any reader upon request.

Jing Sun, Author

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincerest appreciation to Dr. Stephen J. Giovannoni for his intelligent guidance and generous support through my study and research. He opens the door for me and helps me grow in this field.

I would also like to thank my supervisory committee members, Dr. Daniel D. Rockey, Dr. Joseph S. Beckman, Dr. Mark E. Dolan, Dr. Alan T. Bakalinsky and Dr. Katja Duesterdieck-Zellmer for their broad knowledge, valuable advice and help in various ways. I would like to extend my appreciation to Dr. Kimberly H. Halsey, who has provided me advices and technical support.

Big thanks to the nice people in the Giovannoni Lab, the Molecular and Cellular Biology Program, the Department of Microbiology, and the CGRB Core Labs.

Last, but not least, I am also grateful to my parents, Yixian and Ling, my husband Jiazhen, and my lovely daughter Hannah for their unconditional love, understanding, support, patience and encouragement all the time, despite the long distance. I would like to thank all my friends for their encouragement and help in my life.

CONTRIBUTION OF AUTHORS

Dr. Stephen J. Giovannoni assisted with research design and Jing Sun primarily performed the research for chapters 2, 3 and 4. Jing Sun and Dr. Stephen J. Giovannoni analyzed most of the data and wrote the manuscripts.

Chapter 2: Jing Sun performed most of the experiments and analyzed the data. Dr. Laura Steindler provided training of the isotope experiments. Dr. J. Cameron Thrash and Dr. Daniel P. Smith helped with the bioinformatics analyses. Dr. Kimberly H. Halsey, Amy E. Carter, Zachary C. Landry and Dr. Stephen J. Giovannoni performed the field experiments in Bermuda.

Chapter 3: Jing Sun performed most of the experiments. Dr. Michael C. Qian and Dr. Yanping Qian assisted with data collection, and contributed to Figure 3.1. Dr. J. Cameron Thrash and Dr. Ben Temperton helped with the bioinformatics analyses. Dr. Andrew W B Johnston’s group found the possible DMSP lyase, they cloned, expressed and characterize this lyase. Staff in PNNL performed the proteomics analyses.

Chapter 4: Jing Sun designed research, performed research, data analysis, and writing. Jeff Morre in the Mass Spectrometry Facility of OSU measured the enzymatic reaction products. Dr. Stephen J. Giovannoni assisted with the writing and contributed reagents.

TABLE OF CONTENTS

Page

1 General Introduction 1

1.1 Abstract 1

1.2 Marine DOC 2

1.3 C1 and Methylated Compounds in the Oceans 3

1.4 Sources of C1 and Methylated Compounds in Surface Seawater 5

1.4.1 Atmospheric Source 5

1.4.2 Phytoplankton Source 6

1.5 The Microbial Removal Processes of C1 and Methylated Compounds 8

1.5.1 Metabolism of C1 and Methylated Compounds in Methylotrophic Bacteria 9

1.5.1.1 Assimilatory Metabolism 9

1.5.1.2 Dissimilatory Metabolism 11

1.5.2 Metabolism of C1 and Methylated Compounds in the Most Abundant Marine Chemoheterotrophs 13

1.6 Goals of This Study 16

2. One Carbon Metabolism in SAR11 Pelagic Marine Bacteria 20

2.1 Abstract 21

2.2 Introduction 22

2.3 Results and Discussion 25

TABLE OF CONTENTS (Continued)

Page

2.4 Conclusions 36

2.5 Materials and Methods 38

2.6 Acknowlegements 45

3 Dimethylsulfoniopropionate (DMSP) Metabolism in SAR11 Pelagic Marine Bacteria .. 56

3.1 Abstract 57

3.2 Introduction 58

3.3 Results and Discussion 61

3.4 Conclusions 72

3.5 Materials and Methods 73

3.6 Acknowlegements 78

4 Transcriptional Regulatory Responses to C1 and Methylated Compounds and A Predicted Enzyme in Trimethylamine N-oxide Metabolism of the Globally Distributed Marine Bacterium *Candidatus* Pelagibacter ubique HTCC1062 85

4.1 Abstract 86

4.2 Introduction 87

4.3 Results 88

4.4 Discussion 91

4.5 Materials and Methods 94

5 General Conclusions 109

TABLE OF CONTENTS (Continued)

Page

Bibliography 112

Appendices 130

Appendix 1 132

Appendix 2 135

LIST OF FIGURES

Figure Page

Figure 1.1 Outline of C1 and methylated compound metabolism in methylotrophs 18

Figure 1.2 Linear pathways for formaldehyde oxidation in methylotrophic bacteria.....19

Figure 2.1 Demethylation and C1 oxidation regions of the strain HTCC1062 genome...49

Figure 2.2 Proposed C1 and methylated compound oxidation pathways in SAR11 Group Ia 51

Figure 2.3 Phylogenetic tree of Fe-ADH proteins.....52

Figure 2.4 Phylogeny of SAR11 AMT proteins 53

Figure 2.5 ^{14}C-labeled compound utilization by HTCC1062 in culture.....54

Figure 2.6 Utilization of ^{14}C-labeled C1 and methylated compounds by bacterioplanktonin the western Sargasso Sea.....55

Figure 3.1 DMSP metabolism in HTCC1062.....81

Figure 3.2 DMSP catabolic pathways and homologs identified in eight SAR11 genomes 82

Figure 3.3 Utilization of C3 compounds in HTCC1062.....83

Figure 3.4 The relative abundance of SAR11 DMSP catabolism genes in GOS data.....84

Figure 4.1 Expression levels of reference gene *recA* under different cultural conditions 102

Figure 4.2 Comparison of differential expression of gene (SAR11_1304) with microarray (red) and RT-qPCR (blue) in addition of different compounds.....103

LIST OF FIGURES (Continued)

Figure Page

Figure 4.3 Proposed metabolic pathway for MSOX (SAR11_1304) in HTCC1062 104

Figure 4.4 Expression and purification of MSOX from *E. coli* 105

Figure 4.5 Enzymatic assay for the monooxgenase activities of MSOX 106

Figure 4.6 Detection of enzymatic activity products of MSOX 107

Figure 4.7 Potential substrates for MSOX (encoded by SAR11_1304) 108

LIST OF TABLES

Table Page

Table 1.1 Potential substrates for metabolism of C1 and methylated compounds in marine ecosystems....17

Table 2.1 Distribution of genes involved in C1 metabolism among three SAR11 Ia genomes....46

Table 2.2 ATP response of starved cells to addition of various alcohols....47

Table 2.3 ATP response of starved cells to addition of C1 and methylated compounds .48

Table 3.1 Mass balance calculated from Figure 3.2 79

Table 3.2 Proteins with differential expression > 1.5 fold between HTCC1062 cultures amended with methionine vs cultures amended with DMSP as determined by quantitative ITRAQ proteomics 80

Table 4.1 The gene upregulated after adding TMAO 100

Table 4.2 Oligonucleotides used in RT-qPCR 101

LIST OF APPENDIX FIGURES

Figure Page

Figure S2.1 The abundance of SAR11 C1 metabolism genes in GOS data, relative to SAR11 *recA* genes 132

Figure S2.2 Distribution of C1 gene homologs throughout the *Alphaproteobacteria* 133

Figure S2.3 Culture experiments to determine the concentrations of C1 and methylatedcompounds for ATP and radioisotope assays 134

Figure S3.1 Comparison of SAR11 DddK and DddQ-like polypeptides with other cupin DMSP lyases 135

Figure S3.2 Stained SDS-PAGE image showing partical purification of histidine-tagged DddK 136

Figure S3.3 Estimating the Km and Vmax of the DddK by Michaelis-Menten Kinetics . 137

Figure S3.4 Volcano plot of Table 3.2 138

Figure S3.5 Culture experiments to determine the concentrations of C3 compounds .. 139

Metabolism of One Carbon (C1) and Methylated Compounds in SAR11 Bacteria

Chapter 1

1. General Introduction

1.1 Abstract

Marine dissolved organic carbon (DOC) is one of Earth's major carbon reservoirs, and is a critical component of the global carbon cycle. There is a class of compounds, known as one carbon (C1) compounds and methylated compounds, among the broad range of marine DOC. Many of these compounds, such as methanol and dimethyl sulfide (DMS), are volatile. These volatile organic compounds (VOCs) are not detected by normal methods for measuring DOC in seawater. However, they represent a potentially important, underexplored component of the carbon cycle. For more than a decade, biogeochemists have been trying to answer questions about what DOC compounds are important in the surface ocean and what organisms are significantly involved in producing them. Currently, there are two major questions, the first is regarding the major source of these compounds: are they from the atmosphere or phytoplankton? The other question is about the fate of these compounds: it has been thought that the major organisms involved in these

processes are methylotrophs. However, genomics, metagenomics and proteomics have fostered investigation of DOC metabolism by bacterioplankton, and the mounting evidences indicate that the most abundant marine chemoheterotrophs make significant contributions to metabolize these compounds in the oligotrophic ocean. Here, I review and discuss current knowledge about these two controversies. We conclude by suggesting opportunities for future research directions.

1.2 Marine DOC

Marine DOC is a broad and complex mixture of organic molecules that have diverse origins in the world's ocean systems. Marine DOC, which contains as much carbon as the CO_2 in atmosphere, is considered to be one of the largest reservoirs of actively cycled organic matter on the planet (Hedges, 1992). Understanding the factors influencing the synthesis and remineralization of DOC in the ocean is of great significance to the global carbon cycle. Marine DOC is also significant in marine food webs, and it affects the light penetration (Pace and Cole, 2002), the air-sea gas exchange, and the availability of trace metals and other nutrients to biota (Reinthaler, 2008).

DOC is most commonly defined as the organic carbon in seawater that has been passed through a filter with a pore size of 0.45 microns, though 0.22 micron filters have also been used to quantify DOC (Gao et al., 2012). DOC can be measured via several different techniques, but the key principle is to first convert inorganic carbon to CO_2 with

acid, and to remove the resulting CO_2 by sparging. The remaining organic carbon is then oxidized by high temperature combustion, UV/persulfate oxidation, ozone, or UV fluorescence, and the CO_2 generated by the oxidation process is then measured by infrared absorption (Kolka et al., 2008).

1.3 C1 and Methylated Compounds in the Oceans

Among the broad range of marine DOC compounds, there is a class of compounds that contain no carbon-carbon bonds, known as one carbon (C1) compounds (Chistoserdova et al., 2005). There are a large number of C1 compounds in the ocean (Table 1.1). These compounds are potential substrates for C1 metabolism and play significant roles in global carbon cycling and atmospheric chemistry (Chistoserdova et al., 2010; Chistoserdova et al., 2005; Neufeld et al., 2007). In addition, organic matter contains many methylated compounds (Table 1.1). Many osmolytes, which are accumulated by cells because their properties lower the chemical activity of water and stabilize protein structure (Schein, 1990), are methylated. Important examples of these are dimthylsulfonylpropionate (DMSP), glycine betaine (GBT), and trimethylamine N-oxide (TMAO) (Barrett and Kwan, 1985; Diaz et al., 1992; Stefels, 2000). When these compounds are catabolized, methyl groups are removed, and oxidized by pathways similar to those that are employed by cells to oxidize C1 compounds.

Most C1 compounds are volatile. For example, methanol is the predominant volatile organic compound (VOC), second in abundance to methane in the atmosphere (Jacob et al., 2005; Neufeld et al., 2007). The oceans represent an extremely large reservoir of this C1 compound due to its high solubility in water (Jacob et al., 2005). Researchers estimate that the oceans contain approximately 230 Tg (Tg=teragrams or 10^{12} g) of methanol, which is 66 times larger than the estimated atmospheric reservoir (Galbally and Kirstine, 2002). According to atmospheric concentrations and solubility, the estimated range of methanol concentrations in ocean surface waters is from 100 nM (Singh et al., 2003) to 300 nM (Galbally and Kirstine, 2002).

Several methylated compounds are volatile as well. Dimethyl sulfide (DMS) is an important VOC and the primary form of atmospheric sulfur (Stefels et al., 2007). Yoch (Yoch, 2002) estimated that the total annual flux of DMS released from the oceans is 28-45 Tg of sulfur $year^{-1}$, which represents about half of the global biogenic sulfur flux to the atmosphere (Andreae, 1990). Photooxidation products of DMS serve as cloud nucleation agents. In addition, DMS can be transformed by bacterioplankton to DMSO sulfate, tetrathionate, thiosulfate, and methanethiol (Reisch et al., 2011b; Stefels et al., 2007).

Developing appropriate analytical techniques for measuring the numerous VOCs in the environment, and especially in seawater, is challenging (Kameyama et al., 2009;

Sinha, 2007; Williams et al., 2004). Currently, a number of direct mass spectrometric or spectroscopic methods allow sensitive and real-time detection of VOCs. Proton transfer reaction mass spectrometry (PTR-MS) is the state-of-the-art technique for the measurement of the majority of C1 and methylated compounds in seawater and air, but it is only effective with these compounds that have a higher proton affinity than water (Christian et al., 2003). For this reason, solid-phase microextraction (SPME), coupled to gas chromatography with mass spectrometry (GC-MS) is often used, for the measurement of C1 compounds in seawater (Bravo-Linares and Mudge, 2009; Kos and Ariya, 2006). With the combination of these two different techniques, increasing numbers of C1 and methylated compounds have been identified in seawater. This new data is stimulating exploration of metabolic pathways for C1 oxidation and further study of the impacts of these biological processes on the global carbon budget.

1.4 Sources of C1 and Methylated Compounds in Surface Seawater.

1.4.1 Atmospheric Source

Due to the development of new methodologies, the importance of C1 and methylated compounds in the global carbon cycle is becoming widely recognized. Hence, a major challenge in the coming decade of research will be to elucidate the sources of these ubiquitous but poorly constrained constituents of DOC in marine ecosystems. A group of

scientists proposed that atmospheric inputs (from gas exchange, particulate/aerosol deposition and rainwater) are the primary sources of C1 compounds in surface seawater (Duce et al., 1983; Willey et al., 2000). Both Williams (Williams et al., 2004) and Carpenter (Carpenter et al., 2004) observed that several C1 and methylated compounds can be removed from air via deposition to the sea surface, and the latter team found that the concentration of DMS and methanol deposited to the ocean surface is dependent on wind speed. Donald (Donald et al., 2007) implied that atmospheric deposition-driven increases in organic matter solubility (such as sulfate and chloride) may have increased the export of DOC to the oceans. Most recently, Yang (Yang et al., 2013) directly measured the flux of methanol at an average of -10 $\mu mol \cdot m^{-2} \cdot d^{-1}$ from the atmosphere to the ocean over the Atlantic Ocean area, accounting for 18-23% of the total removal of atmospheric methanol. However, they also pointed out that the atmosphere may not be the sole source of seawater methanol. The methanol concentration in deeper water (~500 m) remained significantly high, suggesting that its production is not limited to the near surface. Therefore, there appear to be “missing” sources of methanol in the oceans.

1.4.2 Phytoplankton Source

Another group of scientists have expressed an alternative hypothesis that the primary source of marine C1 and methylated compounds is phytoplankton, which are the most abundant organisms on Earth. Their primary production is equal to terrestrial primary

production (Engel et al., 2013; Myklestad, 2000). During photosynthesis, phytoplankton convert atmospheric CO_2 into biologically available forms of carbon, which are then used for cell materials and energy. This carbon, in the form of carbohydrates, polysaccharides, proteins, long-chain lipids or amino acids, is released into the pool of marine DOC through extracellular release (Larsson and Hagström, 1979), release during predation by grazing organisms (Dafner and Wangersky, 2002), *in situ* production via cell lysis due to viral infections (Wilhelm and Suttle, 1999), or senescence (Myklestad, 2000).

Most marine algae contain sulfated polysaccharides that have a large number of methyl groups. These residues are potential precursors to C1 compounds (Heikes et al., 2002). Phytoplankton can also synthesize large amounts of osmolytes, such as GBT and DMSP (Keller et al., 1999). The quantity of DMSP produced by phytoplankton is estimated to be ~10^9 tons per year, globally (Curson et al., 2011), which may account for 10% of the total carbon fixation in parts of the ocean (Archer et al., 2001; Simó, 2001). Phytoplankton are also considered to be the strongest emitters of DMS in the oceans, and laboratory experiments have shown that phytoplankton are capable of producing DMS (Keller et al., 1989).

Even though increasing evidence has shown that many C1 and methylated compounds are excreted by phytoplankton, their concentrations in surface seawater are poorly correlated with phytoplankton activity measured using the concentration of

chlorophyll (Chla) as a proxy (Sinha, 2007). In addition to challenging methodologies, another explanation for the slow progress in this field is that the determination of these compounds depends not only on the rate of production, but also on their rate of removal. The production rates of different compounds show variations among different phytoplankton species. It appears that some algal groups, such as the coccolithophorids, the most abundant species in subtropical zones, excrete DMS at the highest rate per unit biomass (Charlson et al., 1987). The dominant removal processes for these compounds include ventilation to the atmosphere, photochemical breakdown and microbial metabolism in the seawater (Charlson et al., 1987). However, the rate of microbial removal is affected by the density of bacteria that are able to metabolize these compounds, which we will discuss it in detail in the following section. Due to these uncertainties, the atmospheric and phytoplankton inputs to marine DOC cannot be accurately compared, and the most important sources of these compounds in surface seawater remains an unsolved problem.

1.5 The Microbial Removal Processes of C1 and Methylated Compounds

The major mechanism for the removal of marine DOC is consumption by chemoheterotrophic bacteria (Pomeroy, 1974; Sell, 1994). In the ocean, the base of the microbial food web is generated by the transfer of carbon and associated elements from marine DOC to bacteria (Chróst and Rai, 1994). It is estimated that more than half of the

carbon fixed during photosynthesis by marine phytoplankton passes through DOC and the microbial food web (Azam et al., 1983; Fuhrman 1992). These biological processes and cycling of DOC are most rapid in the upper ocean water column where there is sufficient sunlight for photosynthesis (Jumars et al., 1989).

1.5.1 Metabolism of C1 and Methylated Compounds in Methylotrophic Bacteria

It has been thought that the major players in global cycling of C1 and methylated compounds are methylotrophic bacteria (Chistoserdova et al., 2004). This group of organisms possess a large number of specialized enzymes that enable them to utilize diverse reduced C1 compounds (e.g. methanol or methane) and other methylated compounds that do not have C-C bonds (e.g. GBT or TMA) as their sources of carbon and energy (De Marco, 2004). Phylogenetically, the majority of well characterized methylotrophic bacteria are found in alpha-, beta-, and gamma-proteobacteria and in Gram-positive bacteria (Lidstrom, 2006). Many representatives of methylotrophic genera, such as *Methylobacterium extorquens, Methylococcus capsulatus*, *Paracoccus denitrificans*, and *Methylobacillus flagellatus*, are easily and rapidly cultivated in the laboratory. So they are commonly used in genetic, biochemical and physiological studies of C1 metabolism (Chistoserdova et al., 2005).

1.5.1.1 Assimilatory Metabolism

Generally, in methylotrophs, the first step in the metabolism of C1 compounds and the methyl groups of methylated compounds is oxidation to formaldehyde, a central metabolic intermediate in these metabolic processes. Then formaldehyde is further assimilated for biosynthesis (Assimilation) or oxidized to CO_2 for energy formation (Dissimilation) (Fig. 1.1). There are two unique pathways in methylotrophic bacteria for the assimilation of reduced C1 compounds into common central metabolites: the ribulose monophosphate (RuMP) cycle or the serine cycle (Anthony, 1982; Lidstrom, 2006; Schrader et al., 2009). In the RuMP cycle pathway, formaldehyde is fixed with ribulose 5-phosphate (Ru5P) to form hexulose 6-phosphate (Hu6P) by the enzyme hexulose phosphate synthase (HPS). Hu6P is then isomerized to fructose 6-phosphate (F6P) by phosphohexulose isomerase (PHI), and a series of transformations occur that regenerate Ru5P (Fig. 1.1). This pathway was first described by Quayle's group (Johnson and Quayle, 1965; Kemp and Quayle, 1967), and so far it has been found in beta- and gamma-proteobacteria, as well as in Gram-positive bacteria (Lidstrom, 2006).

In contrast, methylotrophs that use the serine cycle for carbon assimilation belong to alpha-proteobacteria (Lidstrom, 2006). The serine cycle initiates with the fixation of methylene tetrahydrofolate (CH_2=THF) and glycine to produce serine. This three carbon compound then undergoes a series of interconversions to phosphoenolpyruvate, which is

carboxylated to form malate. The malate is then cleaved into glyoxylate, which is converted back into glycine, then completing the cycle (Fig. 1.1).

1.5.1.2 Dissimilatory Metabolism

In methylotrophic bacteria, oxidation of formaldehyde to CO_2 is a key step in energy metabolism. Two main types of pathways have been described: one is a cyclic pathway initiated by the condensation of formaldehyde with RuMP, proceeding essentially as the assimilatory RuMP pathway with participation of glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase (Schrader et al., 2009). This pathway was found to play a dissimilatory role in *M. flagellatus* (Chistoserdova et al., 2000). The other type is a distinct linear pathway that involves a formaldehyde dehydrogenase or C1 moieties linked to cofactors, such as the thiol-compounds (glutathione (GSH) or mycothiol (MySH)), or the folates (tetrahydrofolate (THF) and tetrahydromethanopterin (H_4MPT)) (Lidstrom, 2006; Vorholt, 2002; Yurimoto et al., 2005) (Fig. 1.2). The simplest forms of this linear pathway convert formaldehyde to formate by a series of steps catalyzed by NAD-linked formaldehyde dehydrogenases (Anthony, 1982; Lidstrom, 2006) (Fig. 1.2A).

Among the cofactor-dependent pathways, MySH-linked formaldehyde dehydrogenase was found in several Gram-positive bacteria, such as *Amycolatopsis methanolica* and *Rhodococcus erythropolis* (Misset-Smits et al., 1997; Norin et al., 1997).

This zinc-dependent enzyme is a homotrimer of 40-KDa subunits (van Ophem et al., 1992), which catabolizes S-formhyl-MySH formation from S-hydroxyl methyl MySH. No further hydrolytic activity for S-formyl-MySH has been found yet (Fig. 1.2B). Some Gram-negative bacteria, including *Paracoccus* and *Rhodobacter* (Barber and Donohue, 1998; Ras et al., 1995) (Fig. 1.2C), contain another thiol, GSH, instead of MySH (Vorholt, 2002). Formaldehyde oxidation linked to GSH is considered to be the most widespread enzymatic system for formaldehyde conversion and plays roles in formaldehyde dissimilation and detoxification (Gutheil et al., 1997). Three key enzymes for this conversion have been identified in *Paracoccus denitrificans* (Vorholt, 2002). This pathway starts with formation of formaldehyde and GSH producing S-hydroxymethyl-GSH by GSH-dependent formaldehyde-activating enzyme (Gfa) (Goenrich et al., 2002), then oxidation of S-hydroxymethyl-GSH to S-formyl-GSH via the NAD- and GSH-dependent formaldehyde dehydrogenase (Ras et al., 1995). In the subsequent enzymatic reaction of the pathway, S-formyl-GSH hydrolase regenerates GSH and produces formate (Harms et al., 1996).

Two additional linear formaldehyde oxidation pathways are linked to folates. The H_4MPT dependent pathway was discovered initially in the alpha-proteobacterium *M. extorquens* (Chistoserdova et al., 1998) and is widespread in all of the methylotrophic proteobacteria (Vorholt, 2002). In this process, condensation of formaldehyde is

accelerated by a formaldehyde-activating enzyme (Fae) to form methylene-H_4MPT, which is utilized in further oxidation reactions to dissimilate formaldehyde and generate energy (Fig. 1.2D). The other pathway involves a THF and is found in a variety of biosynthetic processes of methylotrophic and nonmethylotrophic organisms (Maden, 2000). This pathway was considered to be the major dissimilatory route for formaldehyde oxidation in serine cycle methylotrophic bacteria (Marison and Attwood, 1982) besides *M. extorquens*, which also possess this pathway, but as a minor one (Chistoserdova et al., 1998). In this pathway, formaldehyde spontaneously reacts with THF to generate methylene-THF, and this compound can be oxidized to the final product CO_2 through the reactions catalyzed by methylene-THF dehydrogenase, methenyl-THF cyclohydrolase, formyl-THF synthetase and formate dehydrogenase (Fig. 1.2D).

1.5.2 Metabolism of C1 and Methylated Compounds in the Most Abundant Marine Chemoheterotrophs.

Although the methylotrophs and their C1 metabolism have been well studied, these microorganisms, which have the virtue that they can be easily cultured in the laboratory, may not reflect the real microbiota that oxidizes C1 compounds in nature. Recent studies indicate that uncultured or poorly cultured microorganisms that are not well characterized may represent the major removers of DOC in the oceans. Proteomic analysis of the microbial community from surface waters during the summer upwelling off the Oregon coast revealed that 1% of protein-coding sequences identified were closely matched to

the abundant *Betaproteobacteria* clade OM43 (Sowell et al., 2011). In that study, a surprisingly high value (2.3%) of all peptide spectra identified in OM43 were best matched to the methanol dehydrogenase (XoxF). The potential C1 metabolism has also been found in the SAR116 clade of *Alphaproteobacteria*, one of the most common microorganisms detected in the oceans by gene cloning technology (Giovannoni and Stingl, 2005; Mullins et al., 1995). The genome analysis predicted that the SAR116 clade can oxidize a variety of C1 and methylated compounds, such as methanol, formaldehyde, formate, formamide and methanesulfonate. The genome encodes a glycine/serine hydroxymethyltransferase, suggesting the conversion of C1 carbon to biomass by an assimilatory serine cycle. It also encodes a DMSP demethylase and a dimethyl sulfoxide (DMSO) reductase (Oh et al., 2010).

SAR11, a chemoheterotrophic alpha-proteobacterial clade, was first discovered from the Sargasso Sea in 1990 by applying 16S ribosomal RNA genes to map the diversity of marine bacteroplankton populations (Giovannoni et al., 1990). Since then, nearly every survey of microbial diversity in the marine environment, regardless of the method employed, attests the global dominance of SAR11 (Carlson et al., 2009; Morris et al., 2002; Treusch et al., 2009). SAR11 accounts for 35% of total prokaryotes in the surface water and 18% in the mesopelagic zones, and it has been estimated that there are 2.4×10^{28} SAR11 cells worldwide and that they could comprise up to 50% of the biovolume of all

bacterioplankton cells (Morris et al., 2002). Electron microscopy images of SAR11 isolates indicate that the organism is a curved rod, about 0.8 to 1 μm long, and 0.2 to 0.3 μm wide (Rappé et al., 2002). *Candidatus* Pelagibacter ubique HTCC1062, a representative SAR11 Ia ecotype, was first cultivated in 2002 by dilution to extinction methods (Connon and Giovannoni, 2002), and sequenced in 2005 (Giovannoni et al., 2005). Genomic analysis of HTCC1062 strain revealed an extremely small genome of only ~1.3 Mbp and 1354 predicted protein-encoding genes (Giovannoni et al., 2005) ranking it among the smallest free-living bacterial genomes known (Grote et al., 2012).

The genome of HTCC1062 is highly streamlined, with few paralogs, no pseudogenes, and small intergenic spacer regions (Giovannoni et al., 2005). Another consequence of genome streamlining is the reduction in metabolic pathways that are common in other organisms, which has led to unusual nutrient requirements for this marine microorganism. Many strains lack a glycolysis operon and have a limited ability to oxidize carbohydrates (Schwalbach et al., 2009). Another example is the lack of a complete assimilatory sulfate reduction pathway, which has rendered HTCC1062 dependent on reduced sulfur sources, such as DMSP and methionine (Tripp et al., 2008). SAR11 clade can also assimilate amino acids, glucose, pyruvate, and osmolytes such as glycine betaine, and taurine (Schwalbach et al., 2009; Tripp et al., 2008). These studies indicate that SAR11 have a significant impact on a diverse set of DOC compounds.

1.6 Goals of This Study

The genomic analysis predicted a suite of genes for demethylation and C1 oxidation in SAR11 clades, however, no physiological evidence has been reported yet. Therefore, this dissertation examines the metabolism of C1 and methylated compounds in SAR11. These unusual SAR11 metabolic pathways are experimentally explored and discussed in detail in Chapter 2. Among the diversity of C1 and methylated compounds oxidized by SAR11, the methylated sulfur compound DMSP is of particular interest due to the fact that it is the main precursor of the climatically important gas DMS. The metabolism of this ubiquitous marine osmolyte by SAR11 is tested and elucidated in Chapter 3. In Chapter 4, microarray-based screening is performed with the aim of identifying genes that respond to different C1 and methylated compounds and the results are validated by quantitative real-time PCR. Based on the results from experiments above, a monomeric sarcosine oxidase of HTCC1062, was proposed to catalyze oxidation of methylated amines. This enzyme is cloned and expressed in *Escherichia. coli*, and its activities are identified and characterized in Chapter 4.

In summary, understanding and fully characterizing C1 and methylated compounds metabolism in SAR11 will provide critical knowledge needed to refine current models of the ocean carbon cycle so that we may better understand the broad success of SAR11.

Table 1.1 Potential substrates for C1 metabolism in marine ecosystems.

Substrate		Chemical formula	Potential sources and biogeochemical roles
Methanol		CH_3OH	Terrestrial plants, phytoplankton and atmosphere (Carpenter et al., 2004; Nemecek-Marshall et al., 1995; Sinha et al., 2007). One of the most abundant VOCs in the low atmosphere (Jacob et al., 2005).
Formaldehyde		CH_2O	Photolysis of humic material (Keiber et al., 1990) and phytoplankton (Nuccio et al., 1995)
Formate		HCOOH	Photochemical degradation of DOC
Methylated amines	methylamine	CH_3NH_2	Breakdown products of TMAO， oxidation of dietary choline by bacteria in the gut of marine animals (Seibel and Walsh, 2002).
	dimethylamine	$(CH_3)_2NH$	
	trimethylamine	$N(CH_3)_3$	
Methylated osmolytes	GBT	$(CH_3)_3N^+CH_2CO_2^-$	Osmolyte: salinity tolerance, stabilization of protein structure and function. DMSP supports marine bacterial carbon and sulfur assimilation (Kiene and Linn, 2000). Precursor of DMS.
	TMAO	$(CH_3)_3N^+O^-$	
	DMSP	$(CH_3)_2S^+CH_2CH_2COO^-$	
Methylated sulfur	DMS	CH_3SCH_3	Climate-active gas. Breakdown product of DMSP
	Methanethiol	CH_3SH	Breakdown product of DMSP and precursor of DMS

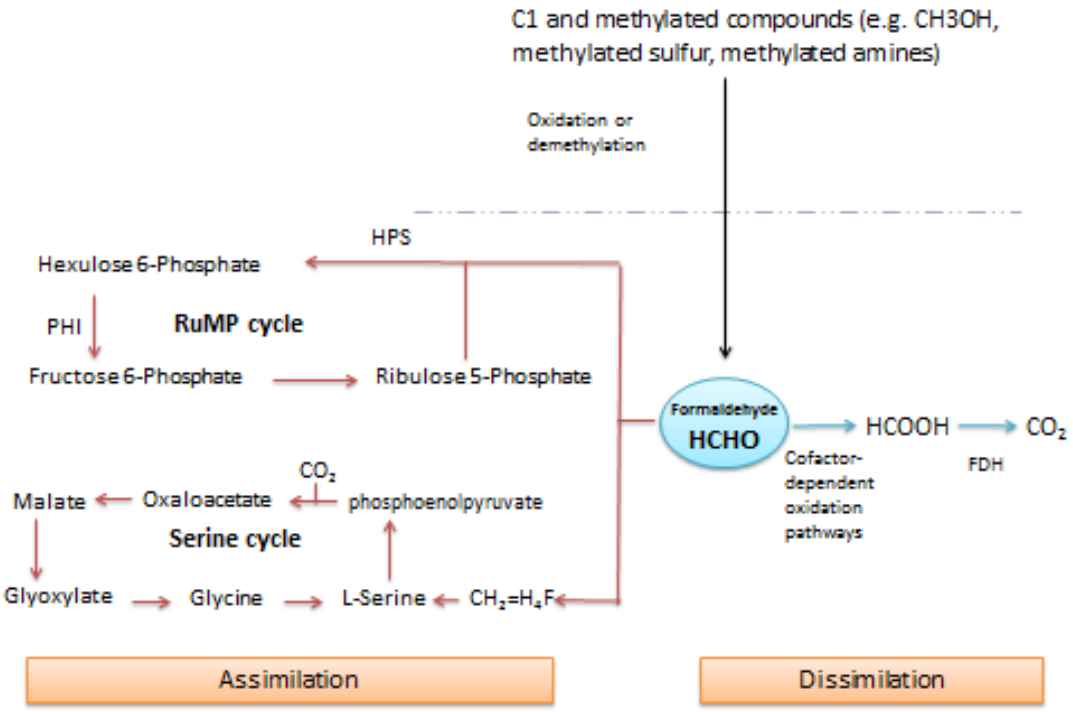


Figure 1.1 Outline of C1 and methylated compound metabolism in methylotrophs. FDH, formate dehydrogenase; RuMP, the ribulose monophosphate pathway; HPS, hexulose phosphate synthase; PHI, phosphohexulose isomerase

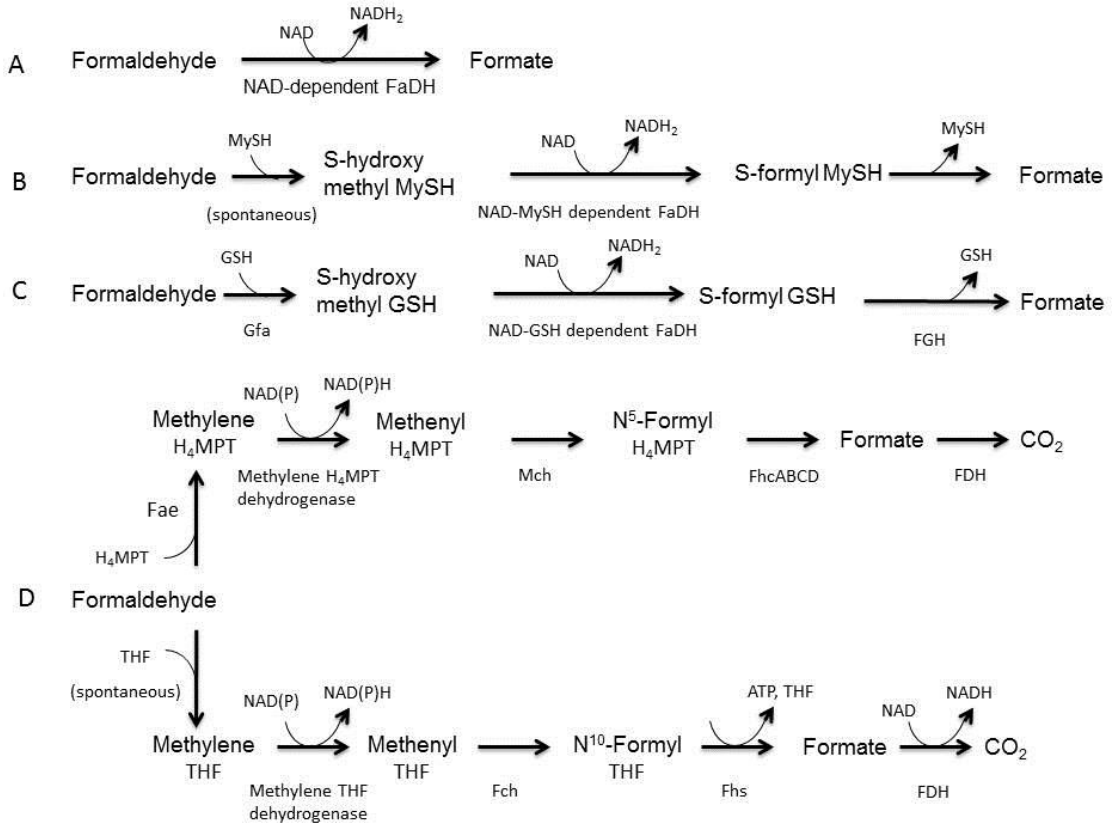


Figure 1.2 Linear pathways for formaldehyde oxidation in methylotrophic bacteria. A, NAD-linked formaldehyde dehydrogenase (FaDH); B, Mycothiol (MySH)-linked FaDH; C,glutathione (GSH)-linked FaDH; D, the two formate-linked pathways, one (upper) involving tetrahydromethanopterin (H4MPT) and the other (lower) involving tetrahydrofolate (THF). Note: Gfa, GSH-denpendent formaldehyde-activating enzyme; FGH, S-formyl GSH hydrolase; Fae, formaldehyde-activating enzyme; Mch, Methenyl H_4MPT cyclohydrolase; FhcABCD, formyltransferase/hydrolase complex; FDH, formate dehydrogenase; Fch, methenyl-THF cyclohydrolase; Fhs, formyl-THF synthetase. (Generate from Lidstrom 2006)

Chapter 2

2. One Carbon Metabolism in SAR11 Pelagic Marine Bacteria

Jing Sun, Laura Steindler, J. Cameron Thrash, Kimberly H. Halsey, Daniel P. Smith, Amy E. Carter, Zachary C. Landry and Stephen J. Giovannoni

Published: PLOS ONE. doi:10.1371/journal.pone.0023973

2.1 Abstract

The SAR11 *Alphaproteobacteria* are the most abundant heterotrophs in the oceans and are believed to play a major role in mineralizing marine dissolved organic carbon. Their genomes are among the smallest known for free-living heterotrophic cells, raising questions about how they successfully utilize complex organic matter with a limited metabolic repertoire. Here we show that conserved genes in SAR11 subgroup Ia (*Candidatus* Pelagibacter ubique) genomes encode pathways for the oxidation of a variety of one-carbon compounds and methyl functional groups from methylated compounds. These pathways were predicted to produce energy by tetrahydrofolate (THF)-mediated oxidation, but not to support the net assimilation of biomass from C1 compounds. Measurements of cellular ATP content and the oxidation of ^{14}C-labeled compounds to $^{14}CO_2$ indicated that methanol, formaldehyde, methylamine, and methyl groups from glycine betaine (GBT), trimethylamine (TMA), trimethylamine N-oxide (TMAO), and dimethylsulfoniopropionate (DMSP) were oxidized by axenic cultures of the SAR11 strain *Ca.* P. ubique HTCC1062. Analyses of metagenomic data showed that genes for C1 metabolism occur at a high frequency in natural SAR11 populations. In short term incubations, natural communities of Sargasso Sea microbial plankton expressed a potential for the oxidation of ^{14}C-labeled formate, formaldehyde, methanol and TMAO that was similar to cultured SAR11 cells and, like cultured SAR11 cells, incorporated a much larger percentage of pyruvate and glucose (27-35%) than of C1 compounds (2-6%) into

biomass. Collectively, these genomic, cellular and environmental data show a surprising capacity for demethylation and C1 oxidation in SAR11 cultures and in natural microbial communities dominated by SAR11, and support the conclusion that C1 oxidation might be a significant conduit by which dissolved organic carbon is recycled to CO_2 in the upper ocean.

2.2 Introduction

C1 metabolism takes place through a network of interrelated biochemical reactions that involves the transfer of one-carbon units from one compound to another. C1 units can be donated in the form of methyl ($-CH_3$), methylene ($-CH_2-$), methenyl (-CH=), formyl (-CHO) and formimino (-CH=NH) groups (Henderson, 1979; McDowell, 2000). A few specialized bacteria oxidize methyl groups and C1 compounds, such as methanol, formaldehyde, formate and methylamine, to derive energy and cellular carbon. The most well known of these organisms are methylotrophs, which assimilate C1 carbon into biomass via the ribulose monophosphate (RuMP) or serine cycle pathways (Chistoserdova et al., 2003; Chistoserdova et al., 2010; Chistoserdova et al., 1998; Giovannoni et al., 2008; Vorholt, 2002). Less well known are organisms that have C1 oxidation pathways for energy production, but lack pathways for the net synthesis of biomass from C1 precursors (Bicknell and Owens, 1980).

Marine dissolved organic carbon (DOC) includes a diverse array of C1 and

methylated compounds that are potential substrates for C1 oxidation. The most common methylated compounds in marine environments are osmolytes such as GBT, TMAO, and DMSP (Barrett and Kwan, 1985; Diaz et al., 1992; Stefels, 2000). Methanol is a major component of oxygenated volatile organic chemicals in the oceans and atmosphere (Heikes et al., 2002; Singh et al., 2003). Air measurements over the Pacific Ocean indicate that sea surface methanol concentrations are about 100 nM and that central ocean regions are net sinks for methanol deposited from the atmosphere (Dixon et al., 2010; Heikes et al., 2002; Singh et al., 2003). Formaldehyde is ubiquitous in seawater. Likely sources of seawater formaldehyde are atmospheric deposition from industrial emissions and the photo-oxidation of atmospheric hydrocarbons (Goode et al., 2000; Zimmerman et al., 1978), and the photo-oxidation of dissolved organic carbon in the ocean surface (Mopper et al., 1991). The metabolism of methylated compounds in mammals also produces formaldehyde (Jones et al., 1978). Formaldehyde is a key reactive intermediate in bacterial metabolism of C1 growth substrates like methane or methanol (Studer et al., 2002; Vorholt et al., 1999; Wilson et al., 2008), and it is also a central intermediate of GBT methyl group oxidation (Meskys et al., 2001). Due to its nonspecific reactivity with proteins and DNA, formaldehyde is toxic to cells, and thus many studies have examined mechanisms by which organisms can remove this potentially lethal compound (Craft et al., 1987; Grafstrom et al., 1983).

Marine bacteria of the SAR11 clade are the most abundant aerobic, free-living,

heterotrophic bacteria in ocean surface waters (Morris et al., 2002; Rappe and Giovannoni, 2003). SAR11 was first discovered in the Sargasso Sea in 1990 (Giovannoni et al., 1990) and is now considered to be one of the most successful organisms on the planet. *Candidatus* Pelagibacter ubique strain HTCC1062, the first cultured SAR11 strain, has one of the most compact genomes known for a free-living organism (1,308,759 base pairs), and has dispensed with many functions that are common in other free-living bacteria, apparently in response to selective pressure for small genome size (Giovannoni et al., 2005). The SAR11 clade is divided into subclades that have distinct temporal and spatial distributions in the environment that are believed to represent ecotypes or species (Carlson et al., 2009). Most strains now being studied in laboratories belong to the Group Ia subclade, which is common in the ocean euphotic zone (Carlson et al., 2009).

In this study, we observed a suite of genes for demethylation and C1 oxidation in three SAR11 Ia genomes (*Ca.* P. ubique strains HTCC1062, HTCC1002, and HTCC7211) for which no role in SAR11 had yet been identified. Genomic and metagenomic data were used to explore the potential for C1 metabolism in the three SAR11 isolates and natural SAR11 populations, and experiments with pure cultures of cells were used to confirm predictions. Together, our data demonstrate that SAR11 cells are capable of oxidizing a range of C1 compounds and methyl groups from methylated compounds to produce energy.

2.3 Results and Discussion

Comparative Genome Analyses and Proposed Metabolic Pathways

Genes in the HTCC1062 genome associated with C1 oxidation are shown in Figure 2.1. The proposed metabolic pathways for these genes are discussed in detail below (Fig. 2.2), although for clarity we include only the HTCC1062 gene identifiers (e.g, “SAR11_”) in the text. To evaluate the conservation and diversity of C1 oxidation pathways in SAR11 we examined the distribution of genes for C1 metabolism among three SAR11 genomes of the Group Ia subclade, and these genes were found in all three SAR11 Ia genomes (Table 2.1). A survey of the global ocean survey (GOS) metagenomic data for SAR11 genes involved in C1 metabolism also supported the conclusion that these genes occur frequently in SAR11 genomes. Averaged across ocean sampling locations worldwide, the ratio of SAR11 C1 genes, relative to SAR11 *recA*, ranged from 0.2 to 2.1 (Fig. S2.1).

Central to the process of C1 and methylated compound oxidation in SAR11 Ia is the tetrahydrofolate (THF)-linked oxidation pathway, which oxidizes C1 units to CO_2, yielding energy in the form of reduced nucleotides and ATP (Fig. 2.2). The gene products of *metF* (SAR11_1264)*, folD* (SAR11_0307) and *fhs* (SAR11_1285) are predicted to catalyze early steps in the methyl-THF linked oxidation pathway (Fig. 2.2A). The genes *fdhF* (SAR11_0679), *fdsB* (SAR11_0680) and *fdhD* (SAR11_0681; Fig. 2.1A) are predicted to encode subunits of formate dehydrogenase (FDH), which catalyzes the final step in the

pathway, the oxidation of formate to CO_2 (Chistoserdova et al., 2007; Chistoserdova et al., 2004) (Fig. 2.2A). Gene *mobA* (SAR11_0684) and *moeA* (SAR11_0685) are predicted to encode proteins for the synthesis of a molybdenum cofactor, which is required for the activity of most bacterial molybdoenzymes, such as FDH (Guse et al., 2003; Hasona et al., 1998; Wang et al., 2002).

Putative genes that encode C1 oxidation activities are found in many bacterial phyla, archaea and eukaryotes. We used the phylogenomic pipeline HAL (Robbertse et al., 2011) to study the distribution of C1 oxidation genes in the class *Alphaproteobacteria* and found evidence for many of these genes throughout the genomes investigated (Fig. S2.2). For example, homologs of *fdhF* and *fdhD* were observed in 77 and 58 of 127 genomes, respectively. However, outside of the methylotrophs, which can grow using C1 compounds as a sole carbon and energy source, C1 oxidation has received relatively little attention. FDH genes have been reported in *Rhizobium japonicum* and *Agrobacterium tumefaciens* (Chen et al., 2010; Manian et al., 1984; Singh and Singh, 1983). Interestingly, several non-methylotrophs, including *A. tumefaciens*, are capable of utilizing methylamine as a nitrogen source, but not as a sole carbon source. As with SAR11 Ia, these bacteria contain the proposed THF-linked C1 oxidation pathway and lack specialized pathways for the assimilation of C1 compounds (Bicknell and Owens, 1980; Chen et al., 2010). These observations suggest that, as we report below for SAR11 Ia strains, *R. japonicum* and *A. tumefaciens*, and probably many other bacteria, may

utilize C1 and methylated compounds for energy production.

All three SAR11 Ia genomes possess an alcohol dehydrogenase (ADH) gene (SAR11_1287; Fig. 2.1B) that was initially annotated as methanol dehydrogenase. We propose that this enzyme catalyzes the oxidation of short chain alcohols, including methanol, to the corresponding aldehydes (Fig. 2.2B). In the case of methanol, the oxidation product is formaldehyde, which can be converted to CO_2 by the methyl-THF linked oxidation pathway described above. This gene encodes a member of the iron-containing alcohol dehydrogenase (Fe-ADH, PF00465) protein family found in many microorganisms. Oxidation of methanol rarely is observed in non-methylotrophic bacteria that lack the classical pyrroloquinoline-quinone (PQQ)-containing methanol dehydrogenase. Methanol dehydrogenase activity has been confirmed in ADH proteins from some *Firmicutes* and Archaea (Antoine et al., 1999; de Vries et al., 1992; Liu et al., 2009). A phylogenetic tree of the Fe-ADH proteins from three SAR11 Ia strains (Fig. 2.3, light green) and closely related homologs from other organisms shows that the proteins with demonstrated methanol dehydrogenase activity (arrows) are more highly diverged than their close relatives, and that the SAR11 Fe-ADH proteins branch nearby (Fig. 2.3). As we report below (Tables 2.2 and 2.3), physiological evidence from ATP assays supports the conclusion that the SAR11 Fe-ADH is not specific for methanol oxidation, but can oxidize other short chain primary alcohols as well. Similar conclusions were reported previously for other bacterial ADH genes in the same family (Arfman et al., 1997;

Bystrykh et al., 1993; Vandecasteele, 2008). Only short-chained, primary alcohols (methanol, ethanol, and 1-propanol) stimulated ATP production in HTCC1062.

A predicted GBT degradation pathway operon was identified in the SAR11 Ia genomes. In this pathway, the three N-methyl groups of GBT are removed in sequence by three enzymes: betaine-homocysteine methyltransferase (BHMT), dimethylglycine dehydrogenase, and sarcosine dehydrogenase (encoded separately by SAR11_1173, SAR11_1253, and SAR11_1221 in HTCC1062). These methyl groups have different predicted fates. The first methyl group is transferred to methionine by BHMT, an enzyme that is common and well studied in mammals, but is rarely found in prokaryotes (Barra et al., 2006; Serra et al., 2002) The second and third methyl groups are predicted to be transferred to THF in reactions that are coupled to partial oxidation and then further oxidized to CO_2 by the THF-linked oxidation pathway described above (Fig. 2.2A, C). A previous study showed that the uptake and metabolism of GBT by bacteria-sized organisms in seawater was correlated with salinity (Kiene, 1998). Consistent with that study, metagenomic analysis showed relatively constant ratios of SAR11 GBT gene frequencies in ocean populations, and absence of these genes in metagenomes from fresh water (Data not shown).

Genes for methylamine dehydrogenase are missing from SAR11 Ia genomes, but they contain genes necessary to encode an alternative, glutamate-mediated pathway for methylamine oxidation (Latypova et al., 2010). Latypova and colleagues showed that

N-methylglutamate synthase (NMGS) and N-methylglutamate dehydrogenase (NMGD), encoded by *mgsABC* and *mgdABCD* in *Methyloversatilis universalis*, participate in methylamine oxidation (Latypova et al., 2010). When the *M. universalis* NMGS protein sequences were blasted against the predicted HTCC1062 proteome, it returned a *glxBCD* operon (SAR11_1313-1315) as the top hit. This operon is annotated as a glutamate synthase and shares gene order with the *M. universalis* NMGS operon. We hypothesize that this operon may encode a holoenzyme that has NMGS activity. NMGS catalyzes transfer of a methyl group from methylamine to glutamate to produce N-methyglutamate as a product.

We postulate that the next step in the pathway, oxidation of N-methyglutamate to glutamate and 5, 10-methylene-THF, with the reduction of FAD to $FADH_2$, is catalyzed by the products of SAR11 genes that are annotated as heterotetrameric forms of sarcosine oxidase (Fig. 2.2D). Two paralogous operons of four sarcosine oxidase genes, *soxBDAG* (SAR11_1064-1067) and $soxB_2D_2A_2G_2$ (SAR11_1281-1284), are located in separate locations of HTCC1062 genomes with the same gene orders and conserved functional domains as those encoding NMGD in *M. universalis* (Fig. 2.1C). Metagenome surveys showed that these two sets of paralogs are highly represented in SAR11 populations globally (Fig. S2.1). We postulate that one of these sets of paralogs encodes the sarcosine oxidase holoenzyme, and the other encodes NMGD; however, which operon encodes which function is unclear. Heterotetrameric sarcosine oxidases (a.k.a sarcosine

dehydrogenase) participate in the GBT degradation pathway by converting sarcosine to glycine. In HTCC1062, SAR11_1304 is predicted to encode the monomeric form of sarcosine oxidase, and is adjacent to two genes annotated as an aminomethyltransferase (AMT) and a GBT transporter (SAR11_1303 and SAR11_1302; Fig. 2.1E). The presence of two paralogous sarcosine oxidase operons and an individual gene encoding a monomeric sarcosine oxidase is noteworthy considering that paralogs are rare in highly compact SAR11 genomes (Giovannoni et al., 2005). We postulate that SAR11 Ia strains may use these genes for the reactions described above and possibly for the demethylation of unknown substrates in oxidation reactions that are likely to produce the intermediate 5, 10-methylene-THF (Chlumsky et al., 1995).

A different pathway for methylamine demethylation that involves the intermediate gamma-glutamylmethylamide was also postulated by Latypova and colleagues (Latypova et al., 2010). We found the homolog of gamma-glutamylmethylamide synthetase (GMAS) from this pathway in HTCC1062, which was annotated as *glnT*, a glutamine synthetase III (SAR11_1316, Fig. 2.1C) and shares the conserved Gln-synt_C domain (Pfam PF00120.17) with GMAS. However, the complete pathway has not been identified and the enzyme was considered to be not essential for oxidation of methylamine in *M. universalis* (Latypova et al., 2010).

The three SAR11 Ia genomes include a family of paralogous AMT genes (Fig. 2.1E). Phylogenetic analysis provided strong support for placing the four AMT genes from

HTCC1062 into three functionally distinct subgroups (Fig. 2.4). SAR11_0666 encodes the glycine cleavage system T-protein (GcvT), part of the glycine cleavage multi-enzyme complex (GCV), which is present in most bacteria and mitochondria (Scott et al., 2008). GCV catalyzes the degradation of glycine to form 5, 10-methylene-THF, CO_2 and NH_3 (Dworkin et al., 2006) (Fig. 2.2F). SAR11_0246 encodes DmdA, a member of the AMT protein family that catalyzes removal of the first methyl group from DMSP to produce methylmercaptopropionate (MMPA) in a pathway for DMSP catabolism (Howard et al., 2006; Reisch et al., 2008) (Fig. 2.2G). However, the enzyme that subsequently demethylates MMPA to mercaptopropionate (MPA) has not been identified.

SAR11_1265 and SAR11_1303 are annotated as probable AMTs with unknown substrates. Three genes (*opuAB, opuAA,* and *opuAC*) are located near SAR11_1303 and encode putative homologs of the GBT transport system permease protein, GBT transport ATP-binding protein and substrate-binding region of ABC-type GBT transport system (Fig. 2.1E). Therefore, we postulate that SAR11_1303 may encode an AMT involved in GBT metabolism. Notably, SAR11_1265 is adjacent to the putative *metF* (SAR11_1264), encoding a homolog of methylene-THF reductase. MetF in *Methylobacterium chloromethanicum* CM4 enables the oxidation of methyl-THF to methylene-THF (20). SAR11_1265 is closely related to SAR11_0246 (*dmdA*) (Fig. 4), and was observed at similar abundances in GOS metagenomic analysis (Fig. S2.1), leading us to speculate that this gene might catalyze removal of the second methyl group in the DMSP

degradation pathway. All of the four AMTs tended towards a 1:1 ratio with single-copy gene *recA* (Fig. S2.1), suggesting that these genes may be part of the SAR11 Ia core genome.

Another gene, SAR11_0621, was initially annotated as an AMT, but later was identified as *ygfZ, which encodes a* folate-binding protein. Studies on YgfZ in *Escherichia coli* and *Arthrobacter globiformis* revealed THF-binding folds in the structure of this protein, but otherwise little similarity to the GcvT family of AMT proteins (Scrutton and Leys, 2005; Teplyakov et al., 2004). Recent studies indicated that YgfZ in *E. coli* has methylase activity (Lin et al., 2010) or may be involved in the regulation of C1 metabolism (Teplyakov et al., 2004). There is currently no direct evidence for a role in C1 metabolism for YgfZ in SAR11.

Some of the three SAR11 Ia genomes had additional genes implicated in C1 metabolism (Table 2.1). Genes encoding glutathione-dependent formaldehyde activating enzyme (GFA), glutathione-dependent formaldehyde dehydrogenase (GD-FALDH) and S-formyl-glutathione hydrolase (FGH) are present in HTCC7211, but not in HTCC1062 or HTCC1002. In *Paracoccus denitrificans*, these proteins catalyze the well-studied glutathione (GSH) dependent pathway that converts formaldehyde to CO_2 (Goenrich et al., 2002; Harms et al., 1996; Vorholt, 2002). We note that some of the predicted C1 oxidation functions of these genes are redundant with genes for the THF-linked oxidation pathway that are also present in the HTCC7211 genome; however, unlike strain

HTCC1062, strain HTCC7211, an isolate from the oligotrophic Sargasso Sea, is predicted to be able to oxidize formaldehyde by using GFA to catalyze the first step in the pathway.

Methane monooxygenases, which catalyze the first reaction of methane oxidation pathways, were not present in any SAR11 genome, indicating that these cells are unlikely to be methane oxidizers. Karl and coworkers reported that methane is produced from methylphosphonate in seawater by the activity of the microbial C-P lyase pathway (Karl et al., 2008). This pathway is present in the genome of HTCC7211, but there are no known metabolic pathways by which methyl groups from methylphosphonate can be diverted into THF-linked C1 oxidation.

Direct Experimental Evidence for C1 Oxidation in SAR11 Strain HTCC1062

Genome analysis suggested that C1 metabolism evolved in SAR11 Ia for energy production, rather than as a means to accumulate biomass. Testing this hypothesis was challenging because SAR11 Ia cells require a variety of unusual organic growth factors and cannot be cultured on “sole” carbon sources, a common paradigm in microbiology. Some required compounds, notably glycine, can be oxidized to produce energy as well as being needed for biomass production (Tripp et al., 2009). Thus, we were not surprised when the addition of various C1 and methylated compounds did not change growth rates or yields of cultures (data not shown). To test for C1 oxidation activities predicted by

genome analysis, we turned to direct measurements of energy production and substrate oxidation.

Measurements of cellular ATP content supported the conclusion that HTCC1062 can produce energy from a wide variety of C1 and methylated compounds (Table 2.3). ATP levels were assayed in HTCC1062 cells that were first grown in the presence of C1 and methylated compounds to be certain that pathways involving C1 oxidation genes were induced. Concentrations of compounds used for ATP assays and ^{14}C experiments (below) were determined in advance by growth assays, as described in the materials and methods. Pyruvate, which was shown previously to be actively metabolized as an energy and carbon source by HTCC1062 cells (Schwalbach et al., 2009), was used as a positive control. ATP content increased in cells incubated with all of the methylated compounds tested with the exception of formate, relative to negative controls. Methanol and TMAO caused the greatest increases in ATP content (3- and 2.4-fold, respectively). We speculate that formate did not enhance ATP levels because it was not transported into cells.

Consistent with predictions from the genome sequence, radioisotope studies with ^{14}C-[methyl]-GBT, ^{14}C-TMA, ^{14}C-methanol and ^{14}C-formaldehyde demonstrated that HTCC1062 oxidized methyl groups from methylated compounds and C1 compounds to CO_2, but incorporation of the compounds into biomass was negligible (Fig. 2.5). Among these compounds, the oxidation rate was fastest with 5 μM ^{14}C-TMA (3.82 nmol × 10^{10}

cells^{-1} × h^{-1}), while 100 nM ^{14}C-formaldehyde was oxidized to $^{14}CO_2$ at the lowest rate (0.01 nmol × 10^{10} cells^{-1} × h^{-1}), likely reflecting the lack of formaldehyde activating enzyme in HTCC1062. The rate of 20 μM methanol oxidized to $^{14}CO_2$ was 0.50 nmol × 10^{10} cells^{-1}× h^{-1}.

The ability of cultured *Ca.* P. ubique to oxidize C1 compounds, and the conservation of genes for C1 metabolism in streamlined SAR11 genomes, suggests that C1 oxidation pathways may contribute significantly to the energy budget of these cells in nature. To examine this hypothesis, we measured the potential rates at which ^{14}C-labeled C1 and methylated compounds were oxidized and incorporated into biomass by cells concentrated from the Sargasso Sea upper euphotic zone during the period of summer stratification. Long-term time series measurements have shown that SAR11 cells range from 30-40% of cells at this ocean site, and that the distinctive microbial community that forms in the upper euphotic zone during the summer is dominated by the SAR11 Ia subclade (Carlson et al., 2009; Morris et al., 2002). The data from bacterioplankton populations collected in the western Sargasso Sea revealed rates of carbon assimilation and oxidation of the same order observed with pure SAR11 cultures in the laboratory (Fig. 2.6 and (Schwalbach et al., 2009)). Similar methanol oxidation rates (~0.8 - 2.5 nmol × 10^{10} cells^{-1} × h^{-1}) were measured in off-shelf northeast Atlantic seawaters (Dixon et al., 2010), which are also dominated by SAR11 cells (Morris et al., 2002). Notably, in the natural bacterioplankton community less than 6 % of the ^{14}C-methanol, 2 % of

^{14}C-formate, and 3 % of ^{14}C-formaldehyde were assimilated into biomass, with the remainder oxidized to CO_2 by the natural community. In contrast, 35 % of glucose and 27 % of pyruvate were assimilated into bacterioplankton biomass in the field experiments (Fig. 2.6), similar to the fraction of organic carbon assimilation into biomass observed previously with growing SAR11 cultures and typical of bacterial organic carbon assimilation in general (Schwalbach et al., 2009). One noteworthy difference between the results obtained with cultured strain HTCC1062 and measurements made in the field is the higher rate of formaldehyde oxidation in the field study. During comparative genome sequence analysis the protein GFA (formaldehyde activating enzyme) was noted in the HTCC7211 genome (Table 2.1), an isolate obtained from the Sargasso Sea. It may be that ecotypes of SAR11 found in the Sargasso Sea summer upper euphotic zone microbial community have a potential to oxidize formaldehyde that is not found in coastal isolates. Overall, these findings suggest that energy production by C1 oxidation, rather than methylotrophy, in which a large fraction (i.e., 62 %; (Ben-Bassat et al., 1980)) of the C1 compound is incorporated into biomass, is the predominant mode of C1 oxidation in surface waters of temperate oligotrophic oceans.

2.4 Conclusions

It is a paradox that the SAR11 clade evolved small genomes while becoming the most successful heterotrophs known, because DOC is a complex substrate that would

appear to require complex metabolic pathways for oxidation. The most obvious solution to this conundrum is substrate specialization by SAR11 and other heterotrophic species in microbial communities, dividing the oxidative side of the carbon cycle into niches. In this report we show that a considerable part of the SAR11 Ia genomes is devoted to C1 metabolism that produces energy but not biomass. C1 and methylated compounds have not generally been regarded as a large fraction of the marine carbon cycle, but there have been a number of reports of C1 compounds and C1 oxidation activity in marine systems (Dixon et al., 2010; Kiene, 1998). We speculate that C1 metabolism evolved in SAR11 because it gave these cells the ability to catabolize a variety of compounds – ranging from low molecular weight photolysis products to osmolytes carrying multiple methyl groups, such as DMSP and GBT. DMSP catabolism in SAR11, and its use as a source of reduced sulfur for growth, has been the subjects of previous studies that did not examine the fate of methyl groups (Reisch et al., 2008; Tripp et al., 2008).

The concept of cells producing energy but not biomass from C1 compounds has received little attention. The best analogy is the term “carboxidovory”, which was coined to describe the metabolism of cells that produce energy by oxidizing carbon monoxide, and to distinguish them from “carboxidotrophs” (King and Weber, 2007; Sorokin et al., 2010), which can use carbon atoms from carbon monoxide for the net assimilation of biomass. Based on this precedent, we propose a new term, “methylovores”, to distinguish cells, such as SAR11 and likely many other bacteria, that can utilize C1 compounds as a

source of energy, from methylotrophs, which are able to use C1 compounds as sole sources of energy and carbon.

Phylogenomic studies have placed the mitochondria deep within the *Alphaproteobacteria*, close to *Rickettsiales* and the SAR11 clade (Thrash et al., 2011). Although respiration is a central feature of mitochondria that is assumed to have driven the original endosymbiotic event, mitochondria make complex contributions to cellular metabolism, including vitamin biosynthesis (Atteia et al., 2009), and the metabolism of C1 compounds (Appling, 1991; Pike et al., 2009), which evidence supports as being present in the protomitochondrion (Gabaldon and Huynen, 2007). In light of the data presented here it is not unreasonable to speculate that the C1 metabolism of mitochondria was contributed by free-living methylovorous ancestors of modern SAR11.

The findings we report show that SAR11 Group Ia strains can produce cellular energy from a broad range of C1 and methylated compounds by oxidative metabolism that we here refer to as methylovory to distinguish it from methylotrophy. The data we present also show that the potential for methylovory was highly expressed in a natural oligotrophic ocean surface microbial community dominated by SAR11. Based on these findings, we speculate that C1 oxidation pathways contribute significantly to the marine carbon cycle and in part explain the broad success of SAR11.

2.5 Materials and Methods

Bioinformatics analysis and phylogeny

Gene calls and functional assignment of SAR11 Group Ia genomes were performed by an automated pipeline at Oregon State University's Center for Genome Research and Biocomputing (http://bioinfo.cgrb.oregonstate.edu/microbes/index.html). Conserved domains were determined with Pfam (Finn et al., 2010). KEGG (Kanehisa et al., 2010) and MicrobesOnline (Dehal et al., 2009) were used for predicting physiological and metabolic pathways.

For the phylogeny of the Fe-ADHs, Pfam was used to confirm the orthology of the Fe-ADH domain in the three SAR11 strains with that of the known methanol dehydrogenase genes (Pfam Fe-ADH PF00465). The phylogenetic tree was then constructed using these six sequences and those of over 70 of the 100 top hits identified using PSI-BLAST with the HTCC1062 gene (YP_266695) as the query sequence. Sequences were aligned using MUSCLE (Edgar, 2004a, b), and manually edited to remove columns with >85% gaps. Substitution modeling was completed using ProtTest (Abascal et al., 2005). The alignment was then analyzed with RAxML (Stamatakis, 2006) using the WAG substitution model and nodal support was estimated based on 100 bootstrap replications.

Phylogeny of the AMTs was performed by MicobesOnline website and MacVector software. The top 10 matched protein sequences of each AMT (YP_265671, YP_266089, YP_266673, YP_266710) were downloaded using BLAST tools from MicrobesOnline.

AMTs were aligned using the ClustalW module of MacVector version 10.5.2, and Phylogenetic trees were constructed using the neighbor-joining method. One-hundred bootstrap replicates were used to estimate the robustness of branches in the phylogenetic tree.

Homology assessment of methylovory genes was completed using the HAL pipeline (Robbertse et al., 2006; Robbertse et al., 2011) (http://aftol.org/pages/Halweb3.htm; http://sourceforge.net/projects/bio-hal/) followed by manual examination of predicted orthologous clusters. For this study, the HAL pipeline directed the following analyses: Protein sequences from 127 publically available *Alphaproteobacteria* genomes (for a complete list, see (Thrash et al., 2011)) were imported in FASTA format from IMG (http://img.jgi.doe.gov) and subjected to all vs. all BLASTP with the output E-values provided to the program MCL (van Dongen, 2000), which grouped proteins into orthologous clusters using 13 inflation parameters from 1.1 – 5.0. For homology assessment in Figure S2.2, we examined clusters formed at the relatively conservative 3.0 inflation parameter, and thus these numbers can be considered as lower estimates of total homologs.

Metagenomic analysis

The reciprocal best BLAST (RBB) approach (Wilhelm et al., 2007) was used to identify environmental fragments in the global ocean survey (GOS) database originating

from SAR11 and encoding genes related to C1 and methyl group oxidation. The amino acid sequences for the genes discussed in this paper were searched with BLASTP (e-value threshold = 1×10^{-5}) against peptide sequences called from GOS metagenomic nucleotide reads. The taxonomies of matching peptides were determined by querying the peptide against the NCBI non-redundant protein database (NR) using BLASTP (e-value threshold = 10). Sequences were only retained if this second search against NR found one of the three *Ca.* P. ubique strains to be the best match to the environmental peptide. Peptides were lastly searched with BLASTP (e-value threshold = 1×10^{-5}) against all *Ca.* P. ubique strain HTCC1062 amino acid sequences to ensure that the initial *Ca.* P. ubique query protein, rather than a paralog, was the best match to the environmental peptide sequence. The number of environmental peptides passing the RBB test, or hits, per gene at each GOS sampling site were normalized to the length of the single-copy essential gene *recA* and the number of hits to *recA* at the same sampling site using the formula $P_g = (L_{recA} / L_g) \times (H_g / H_{recA})$, where P_g is the percentage abundance related to *recA* for gene *g*, L_{recA} is the length of *recA* in base-pairs (bp), L_g is the length of gene *g* in bp, H_g is the number of hits for gene *g*, and H_{recA} is the number of hits to *recA* (Biers et al., 2009). GOS sampling sites that had fewer than five hits to *recA* were removed from downstream analysis.

ATP measurements:

HTCC1062 cultures for ATP assays were grown in autoclaved, filtered seawater amended with 1 mM NH_4Cl, 100 µM KH_2PO_4, 1 µM $FeCl_3$, 80 µM pyruvate, 40 µM oxaloacetate, 40 µM taurine, 1 µM GBT, 50 µM glycine, 50 µM methionine, and excess vitamins (Schwalbach et al., 2009). C1 and methylated compounds to be tested were added as follows: 1 µM GBT, 20 µM formate, 10 µM methylamine, 5 µM TMAO, 1 µM DMSP, 20 µM methanol and 100 nM formaldehyde. The concentration of GBT used in the assay was the same as routinely used in *Ca.* P. ubique growth media (1µM). The concentrations of other compounds utilized in assays were chosen by testing the growth of HTCC1062 with differing concentrations of each compound and selecting concentrations that were not inhibitory (Fig. S2.3). Cultures were grown in the dark at 16 °C. When cultures entered early stationary phase (~3×10^8 cells mL^{-1}), 10 mL of each culture were collected and cells were harvested via centrifugation (50 min at 43,700 *g*, 10 °C). Following centrifugation, cells were washed twice in artificial seawater (ASW) (Schwalbach et al., 2009) and finally resuspended in 5 mL ASW. Then cells were distributed into 1.7 mL tubes (500 µL in each tube) and starved overnight (20 hours in the dark, 16 °C). Test treatments and controls were done in triplicate with identical cell suspensions. After 2 hours incubation in the dark, ATP content was measured using a luciferase-based assay (BactTiter Glo, Promega, Madison, WI) as follows: 90 µL of BactTiterGlo reagent were dispensed into white 96 well plates (Tissue culture-treated, BD Biosciences, San Jose, CA). 20 µL of each sample were added per well, and

luminescence was measured after 4 min using a multi-functional plate reader (TECAN, Infinite M200) with a 1 s integration and 10 ms settle time. An ATP standard curve was used to calculate the concentration of ATP in the samples. Student's t-test was used to assess statistical significance (p-value<0.01) between controls and treatments. The ATP measurements for other alcohols (20 μM ethanol, 20 μM 1-propanol, 20 μM 2-propanol, 20 μM 1-butanol, 20 μM 2-pentanol, and 20 μM iso-amyl alcohol) were performed as described above.

Radioisotope assays:

To distinguish between ^{14}C-labeled compounds incorporated into biomass and oxidized to $^{14}CO_2$, a method was devised for volatile ^{14}C compounds (e.g., ^{14}C-labeled methanol). Trichloroacetic acid (TCA)-precipitation was used to assess ^{14}C-incorporation into cellular material, and $^{14}CO_2$ was precipitated by addition of NaOH, Na_2CO_3 and $BaCl_2$, forming $Ba^{14}CO_3$ and BaOH. Filtration removed unincorporated ^{14}C compounds and BaOH, thus minimizing potential quenching effects of the base.

For experiments with radioisotopes, HTCC1062 was grown in seawater medium as described above for ATP measurements. Cultures were amended with unlabeled test compounds to induce activities (5 μM TMA, 20 μM methanol, or 100 nM formaldehyde; 1 μM GBT was in the seawater medium). Cells were harvested in log phase by centrifugation (1 hour at 43,700 *g*, 10 °C) and resuspended in ASW to about 4×10^7 cells

mL^{-1}. Negative controls ("Killed") were incubated in 10% formalin for 1 hour before the addition of the isotope to the sample. 1 μM ^{14}C-GBT, 5 μM ^{14}C-TMA, 20 μM ^{14}C-methanol, or 100 nM ^{14}C-formaldehyde were added to both "Live" and "Killed" culture samples. Inoculated cultures (4 mL) were aliquoted into 40 mL sealed vials. Samples were incubated at room temperature. At each time point, reagents were added to cultures using syringes inserted through stoppers. For ^{14}C-incorporation, 2.2 mL 100% w/v cold TCA was added; for trapping $^{14}CO_2$, 1 mL 1N NaOH, 0.5 mL 0.1 M Na_2CO_3 and 1 mL 1 M $BaCl_2$ were added. All samples were collected by filtration after incubation at 4 °C for 12 hrs. Filters were transferred to vials containing 15 mL Ultima Gold^{TM} XR scintillation fluid (Perkin-Elmer) and kept in dark overnight before counting (Beckman LS-6500 liquid scintillation counter).

For field studies of radiolabeled compound utilization by bacterioplankton in the western Sargasso Sea, seawater was collected from 10 m at Bermuda Hydrostation S using Niskin bottles, and transferred to an acid-washed, autoclaved polycarbonate carboy. For each experiment, microbial plankton were concentrated from 80 L seawater by tangential flow filtration to a final volume of 600 mL, and isotopic labeling was carried out as described above. The concentrations of the tested compounds were 3 μM glucose, 1.7 μM pyruvate, 0.3 μM formate, 1.1 μM formaldehyde, 50 μM methanol and 0.5 μM TMAO. A high concentration of methanol was used because the specific activity of the labeled compound was very low. Some unexpected evaporation of the radio-labeled compounds

occurred during transport causing some variability between concentrations of the other compounds used. Nevertheless, in all cases the concentrations used were likely to be substantially higher than typical seawater concentrations, and the results represent potential rates of compound utilization.

2.6 Acknowledgments

Jing thanks financial support from the China Scholarships Council (CSC). We thank the crews of the RV Elaka and RV Atlantic Explorer, Craig Carlson, Rachel Parsons and Craig Nelson for their assistance with field studies. We also thank Ludmila Chistoserdova and Marina G. Kalyuzhnaya for stimulating discussions about C1 metabolism.

Table 2.1 Distribution of genes involved in C1 metabolism among three SAR11 Ia genomes. The phylogenomics pipeline HAL and manual searches were used to detect orthologs among the genomes. Genes for C1 oxidation were present in all three genomes. HTCC7211 possesses three genes for the glutathione (GSH) dependent C1 oxidation pathway that are not present in the other two SAR11 Ia genomes.

Genes for C1 oxidation and methylovory		HTCC 1062	HTCC 1002	HTCC 7211
THF-linked oxidation	formate dehydrogenase, alpha subunit (*fdhF*)	+	+	+
	NAD-dependent formate dehydrogenase, beta subunit (*fdsB*)	+	+	+
	formate dehydrogenase, chain D (*fdhD*)	+	+	+
	molybdopterin-guanine dinucleotide biosynthesis protein A (*mobA*)	+	+	+
	molybdopterin biosynthesis protein (*moeA*)	+	+	+
	formate-THF ligase (*fhs*)	+	+	+
	methylene-THF reductase (*metF*)	+	+	+
	bifunctional methylene-THF dehydrogenase-methenyl-THF cyclohydrolase (*folD*)	+	+	+
methanol oxidation	iron-containing alcohol dehydrogenase (Fe-ADH)	+	+	+
methylamine oxidation	glutamine synthetase III (*glnT*)	+	+	+
	putative N-methylglutamate synthase (*glxBCD*)	+	+	+
	putative N-methylglutamate dehydrogenase (*soxBDAG)*	+	+	+
GBT oxidation	betaine-homocysteine methyltransferase (*bhmT*)	+	+	-
	sarcosine dehydrogenase (*sardh*)	+	+	+
	dimethylglycine dehydrogenase (*dmgdh*)	+	+	+
AMTs	glycine system cleavage T-protein (*gcvT*)	+	+	+
	dimethylsulfoniopropionate-dependent demethylase (*dmdA*)	+	+	+
	putative aminomethyltransferase	+	+	+
GSH dependent pathway	glutathione-dependent formaldehyde activating enzyme (*gfa*)	-	-	+
	glutathione-dependent formaldehyde dehydrogenase (GD-FALDH)	-	-	+
	S-formyl-glutathione hydrolase (FGH)	-	-	+

Table 2.2 ATP response of starved cells to addition of various alcohols[a].

Test compounds[b]	samples	Cellular ATP content (Mean ± SD; zeptogram cell^{-1})
* ethanol	T	57 ± 8
	N	14 ± 3
	P	135 ± 15
* 1-propanol	T	47 ± 6
	N	19 ± 4
	P	41 ± 3
2-propanol	T	10 ± 1
	N	11 ± 2
	P	26 ± 1
1-butanol	T	25 ± 7
	N	23 ± 9
	P	35 ± 7
2-pentanol	T	13 ± 2
	N	11 ± 2
	P	25 ± 9
iso-amyl alcohol	T	18 ± 5
	N	14 ± 2
	P	40 ± 8

a. For each assay, cells were grown in media containing the test compound, then washed and starved for 20 hrs. Cellular ATP content was measured after cells were exposed for 2 hrs to the test compound (T), to no compound added (N), and to pyruvate (P; positive control to confirm metabolic activity of cells).

b. Asterisk indicates statistical significance (p-value < 0.01) between “no compound added” and “test compound” treatments.

Table 2.3 ATP response of starved cells to addition of C1 and methylated compounds[a].

Test compounds[b]	samples	Cellular ATP content (Mean ± SD; zeptogram cell^{-1})
formate	T	32 ± 3
	N	29 ± 8
	P	221 ± 4
***** methanol	T	48 ± 0
	N	16 ± 3
	P	160 ± 8
***** formaldehyde	T	33 ± 6
	N	14 ± 1
	P	77 ± 5
***** DMSP	T	23 ± 3
	N	16 ± 1
	P	163 ± 7
***** methylamine	T	27 ± 1
	N	18 ± 0
	P	145 ± 10
***** glycine betaine	T	41 ± 1
	N	23 ± 3
	P	132 ± 3
***** TMAO	T	63 ± 5
	N	26 ± 2
	P	148 ± 10

a. For each assay, cells were grown in media containing the test compound, then washed and starved for 20 hrs. Cellular ATP content was measured after cells were exposed for 2 hrs to the test compound (T), to no compound added (N), and to pyruvate (P; positive control to confirm metabolic activity of cells).

b. Asterisk indicates statistical significance (p-value < 0.01) between “no compound added” and “test compound” treatments.

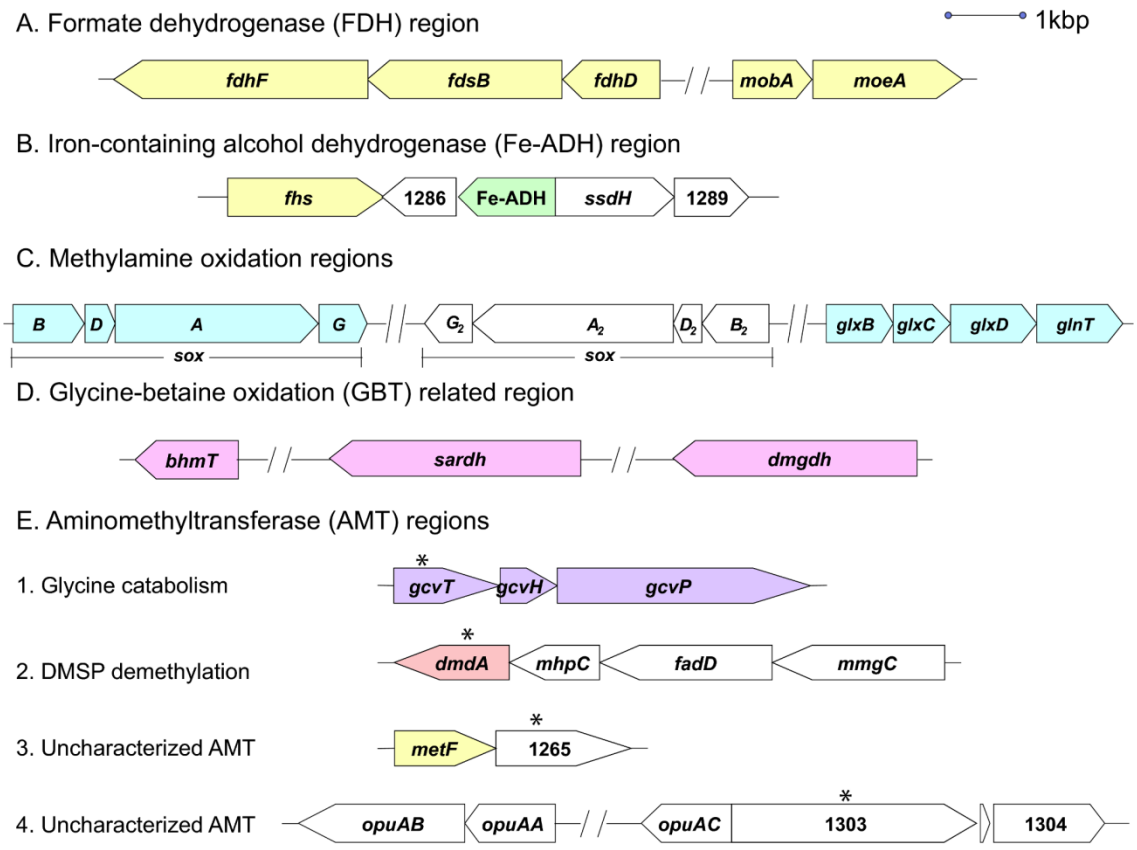


Figure 2.1 Demethylation and C1 oxidation regions of the strain HTCC1062 genome. (A) formate dehydrogenase; (B) methanol metabolism; (C) methylamine oxidation; (D) glycine betaine oxidation; (E) aminomethyltransferases (Asterisk). ***fdhF***, formate dehydrogenase, alpha subunit; ***fdsB***, NAD-dependent formate dehydrogenase, beta subunit; ***fdhD***, formate dehydrogenase, chain D; ***mobA***, molybdopterin-guanine dinucleotide biosynthesis protein A; ***moeA***, molybdopterin biosynthesis protein; ***fhs***, formate-THF ligase; SAR11_1286, putative glutamine amidotransferase; **Fe-ADH,** iron-containing alcohol dehydrogenase; ***ssdH,*** aldehyde dehydrogenase family; SAR11_1289, short chain dehydrogenase; ***soxB***, sarcosine oxidase; ***soxD*** & ***soxD$_2$***, sarcosine oxidase delta chain; ***soxA*** & ***soxA$_2$***, sarcosine oxidase alpha chain; ***soxG*** & ***soxG$_2$***, sarcosine oxidase gamma subunit; ***soxB$_2$***, sarcosine oxidase beta subunit; ***glxBCD***, glutamate synthase; ***glnT***, Glutamine synthetase III (putative gamma-glutamylmethylamide synthetase); ***bhmT***, betaine-homocysteine methyltransferase; ***sardh***, sarcosine dehydrogenase; ***dmgdh***, dimethylglycine dehydrogenase; ***gcvT***, glycine system cleavage T-protein; ***gcvH***, glycine cleavage H-protein; ***gcvP***, glycine cleavage P-protein; ***dmdA***, dimethylsulfoniopropionate-dependent demethylase; ***mhpC***, hydrolase, alpha/beta

hydrolase fold family; ***fadD***, CoA activator for DMSP beta oxidation; ***mmgC***, acyl-CoA dehydrogenase for DMSP beta oxidation; ***metF,*** methylene-THF reductase; ***opuAB***, glycine betaine transport system permease protein; ***opuAA***, glycine betaine transport ATP-binding protein; ***opuAC***, substrate-binding region of ABC-type glycine betaine transport system; SAR11_1265 & SAR11_1303, gcvT-like aminomethyltransferase protein; SAR11_1304, monomeric sarcosine oxidase. Colors correspond to pathways in Figure 2.2.

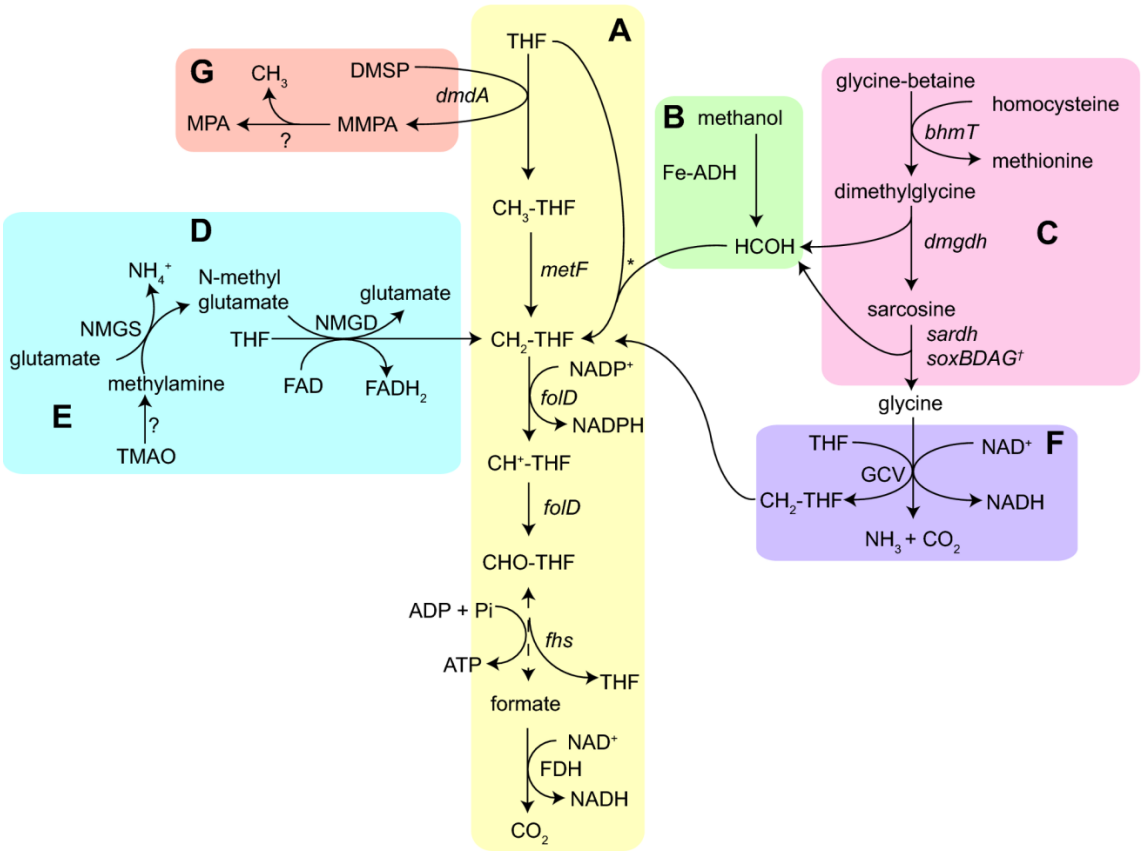


Figure 2.2 Proposed C1 and methylated compound oxidation pathways in SAR11 Group Ia. (A) THF-linked oxidation pathway; (B) methanol oxidation pathway; (C) glycine betaine demethylation and oxidation; (D) methylamine oxidation pathways; (E) TMAO degradation pathway; (F) glycine cleavage pathway; (G) DMSP demethylation.
Note: **?** - unidentified metabolic processes/enzymes; ***** - spontaneous reaction; **†** - two paralogous operons.

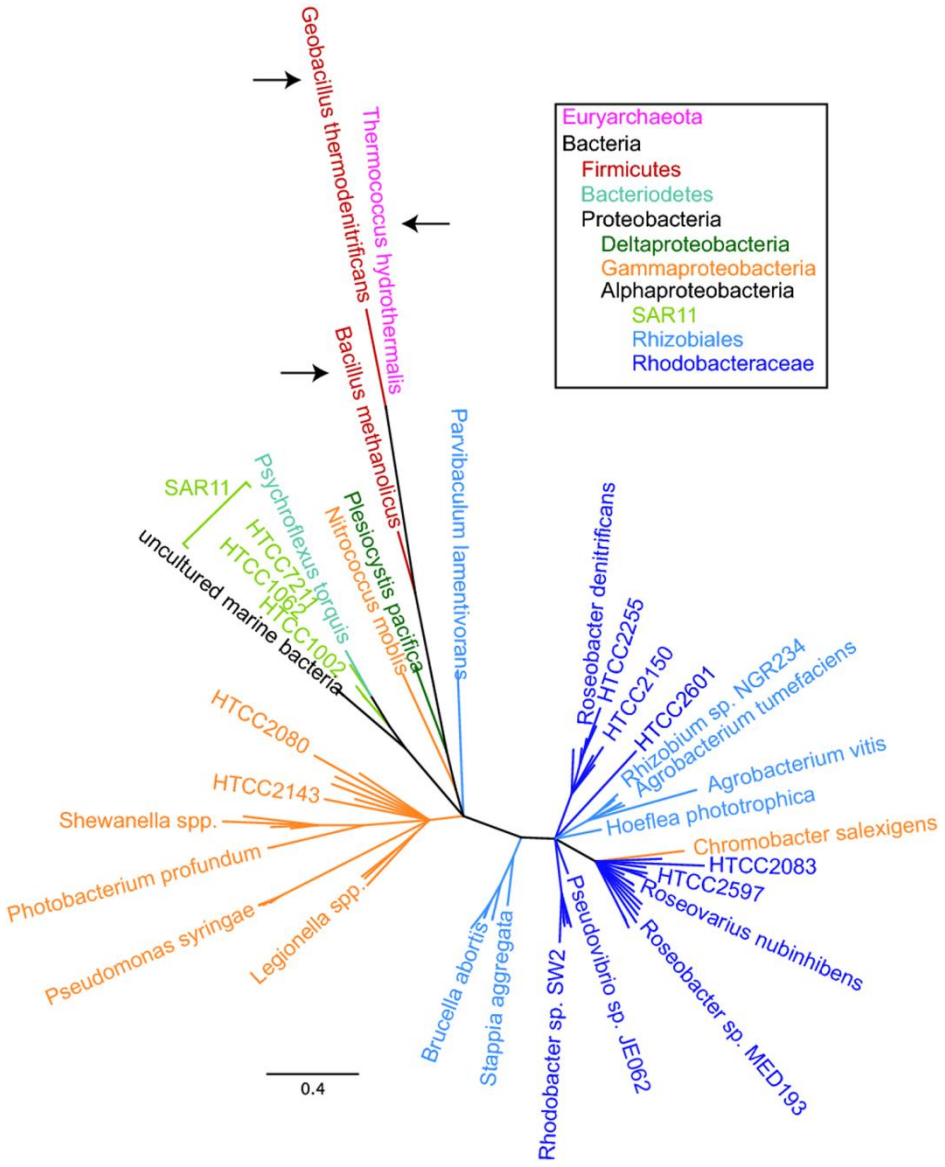


Figure 2.3 Phylogenetic tree of Fe-ADH proteins. Coloration is according to 16S rRNA gene phylogeny, as shown in the boxed legend. Bootstrap values were omitted for clarity; nodes with less than 60% support were collapsed. Arrows indicate Fe-ADH proteins for which methanol dehydrogenase activity has been demonstrated experimentally. Scale bar = 0.4 changes per position.

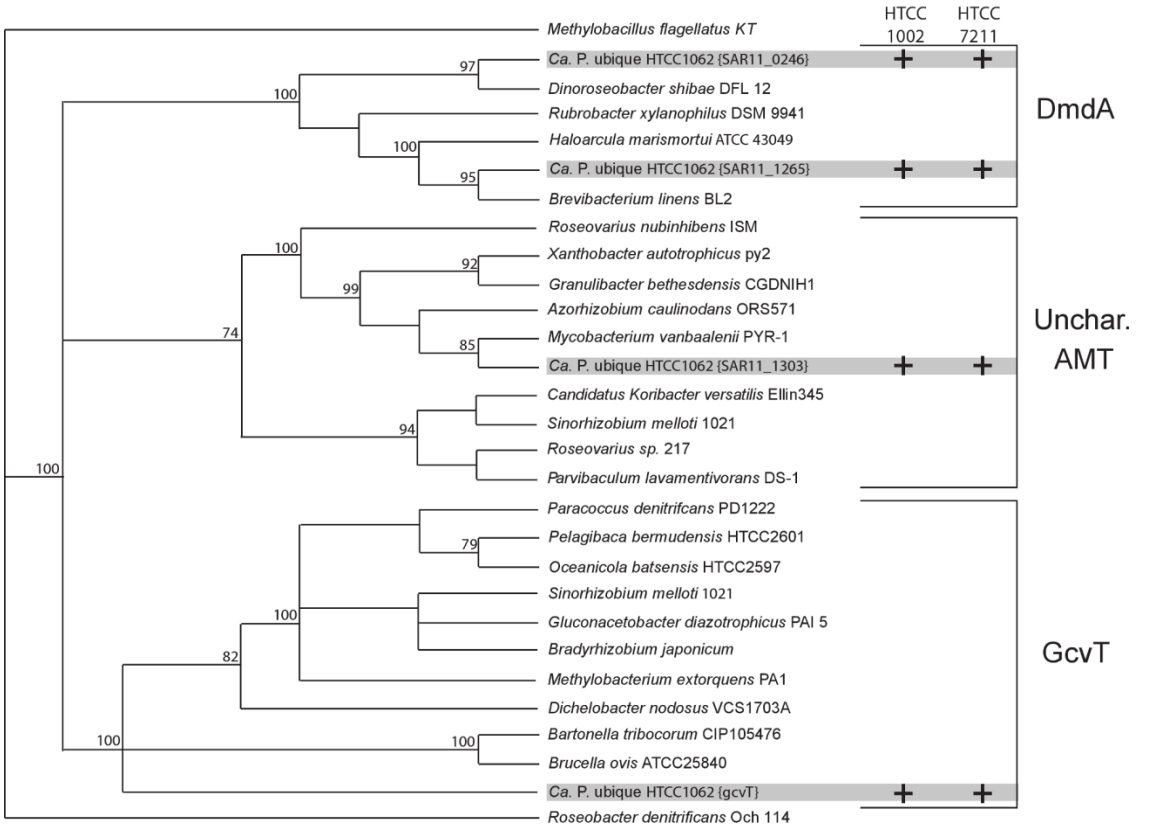


Figure 2.4 Phylogeny of SAR11 AMT proteins. Four paralogous AMTs in HTCC1062 were placed into three functional subgroups: DmdA-like, GcvT, and an AMT of unknown function. All four AMTs were also identified in HTCC1002 and HTCC7211 genomes. This phylogenetic tree was generated using the neighbor-joining method. Bootstrap values are based on 100 iterations.

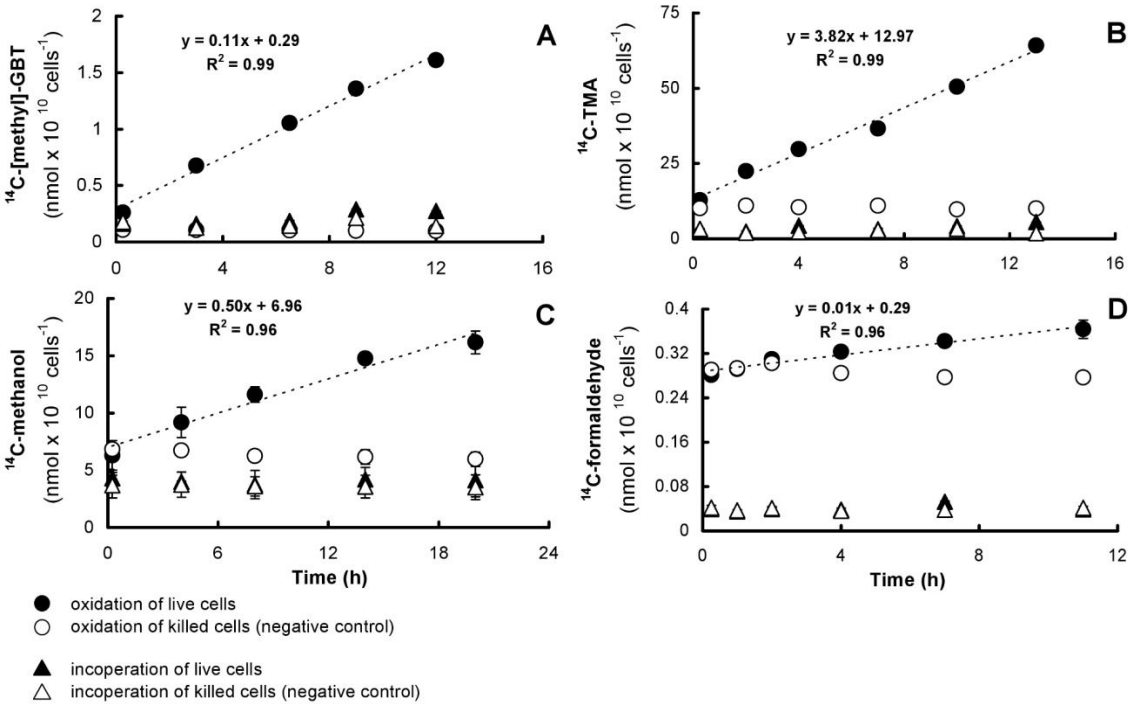


Figure 2.5 ^{14}C-labeled compound utilization by HTCC1062 in culture. HTCC1062 Cells from log phase were collected and resuspended in artificial seawater media (ASW). Radioisotope assays were conducted at room temperature (22 °C) in ASW amended with (A) 1 μM ^{14}C-[methyl]-GBT; (B) 5 μM ^{14}C-TMA; (C) 20 μM ^{14}C-methanol; or (D) 100 nM ^{14}C-formaldehyde. Where not visible, error bars are smaller than the size of the symbols.

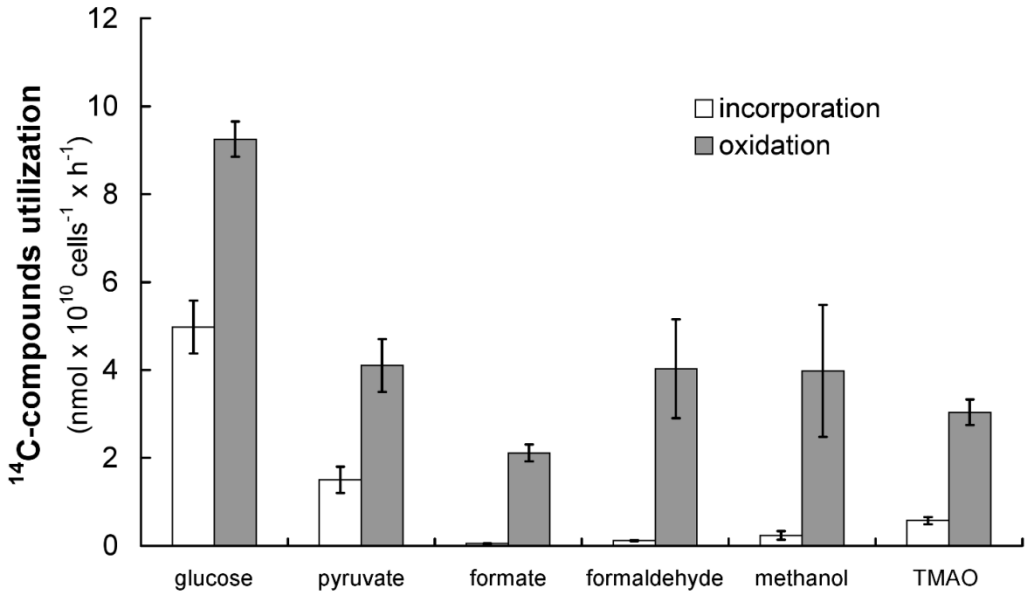


Figure 2.6 Utilization of ^{14}C-labeled C1 and methylated compounds by bacterioplankton in the western Sargasso Sea. The oxidation and incorporation rates were calculated from the initial linear part of each curve. Rate of ^{14}C-compound oxidation to $^{14}CO_2$ (■); rate of ^{14}C-compounds incorporation into biomass (□).

Chapter 3

3. Dimethylsulfoniopropionate (DMSP) Metabolism in SAR11 Pelagic Marine Bacteria

Jing Sun, Jonathan D Todd, J. Cameron Thrash, Yanping Qian, Michael C. Qian, Ben Temperton, Jiazhen Guo, Emily K Fowler, Joshua Aldrich, Carrie D. Nicora, Mary S. Lipton, Richard D. Smith, Andrew W B Johnston and Stephen J. Giovannoni

In preparation for submission to: Environmental Microbiology

3.1 Abstract

Dimethylsulfoniopropionate (DMSP) is ubiquitous in ocean surface waters, where it is an important source of sulfur and carbon for chemoorganotrophic bacteria, and a precursor of the climatically active gas dimethyl sulfide (DMS). SAR11 *Alphaproteobacteria*, the most abundant chemoheterotrophic bacteria in the oceans, have been shown to contribute significantly to mineralizing and assimilating DMSP by demethylation, resulting in formation of methanethiol. Here we demonstrated for the first time that SAR11 strain HTCC1062 can also cleave DMSP to DMS, with as much as 59% of DMSP uptake channeled through the cleavage pathway. HTCC1062 was able to use methanethiol, but not DMS, as a source of reduced sulfur for growth. Bioinformatic analyses predicted three putative DMSP lyases among members of SAR11, including a novel enzyme termed DddK. The corresponding cloned *dddP, dddQ* and *dddK* genes conferred DMSP cleavage ability to *E. coli* and the purified DddK enzyme cleaved DMSP into DMS plus acrylate *in vitro.* Genes for metabolism of DMSP via the cleavage pathway intermediates acrylate or 3-hydroxypropionate were found throughout SAR11. Comparative metagenomic analyses indicate that SAR11 genes for both the cleavage and demethylation pathways are abundant throughout the world's oceans. Due to the vast population sizes of SAR11, our observations have compelling implications for the global sulfur cycle.

3.2 Introduction

Dimethylsulfoniopropionate (DMSP) is an important molecule in marine carbon and sulfur cycling. It is produced in large quantities (~10^9 tons per year, globally (Curson et al., 2011)) by phytoplankton, for which it has a variety of roles. It has been documented as an osmolyte (Kirst, 1990), and can serve additional functions, such as oxygen scavenging (Sunda et al., 2002) and toxicity (Wolfe et al., 1997). Following deterioration of phytoplankton upon cell death, DMSP can be metabolized by a variety of bacterioplankton via two different pathways: cleavage and demethylation. The cleavage pathway produces dimethyl sulfide (DMS), which is a volatile compound and the primary form of biogenic, atmospheric sulfur (Stefels et al., 2007) and either of the three-carbon (C3) compounds acrylate or 3-hydroxypropionate (3-HP), depending on the particular DMSP lyase (Curson et al., 2011; Moran et al., 2012). Thus far, six unique, non-homologous DMSP lyase analogues have been described (Curson et al., 2011; Moran et al., 2012). Five of these, *dddL*, *dddP*, *dddQ*, *dddW*, and *dddY* catalyze the production of DMS + acrylate from DMSP. A sixth, *dddD*, produces DMS + 3-HP (Todd et al., 2007).

It has been estimated, based on bulk seawater turnover rates from around the world, that only about 10% of total DMSP is converted to DMS (Kiene et al., 2000). In addition to incorporation and accumulation as a potential osmolyte (Kiene et al., 2000), DMSP can

be transformed via the demethylation pathway, the genetics of which have recently been elucidated in the model organism *Ruegeria pomeroyi*, a member of the *Alphaproteobacteria* clade known as the Roseobacters (Reisch et al., 2011b). The first step in DMSP demethylation is catalyzed by a DMSP demethylase (*dmdA*), which produces methylmercaptopropionate (MMPA) via a tetrahydrofolate (THF) - mediated reaction. MMPA is further catabolized in three sequential coenzyme-A mediated steps by the gene products of *dmdB*, *dmdC*, and *dmdD*, generating methanethiol (MeSH), acetaldehyde, CO_2, and CoA as the final products. MeSH can be converted to methionine by cystathionine-ɣ-synthetase (Kiene et al., 1999), or oxidized to formaldehyde and sulfide, both of which can be oxidized by other pathways (Ghosh and Dam, 2009; Reisch et al., 2011a). The recent discovery of the demethylation pathway genes *dmdA*, *dmdB*, and *dmdC* led to appreciation of the important role of the demethylation pathway when it was shown that these genes are found in most SAR11 genomes and are highly abundant in ocean metagenomes (Moran et al., 2012).

Alphaproteobacteria of the SAR11 clade are the most abundant organisms in the ocean. With estimated global population sizes of 10^{28} cells (Morris et al., 2002), they play significant roles in marine biogeochemical cycles, including those of carbon and sulfur. The genome of HTCC1062 is one of the smallest for a free-living organism (1.31 Mbp) (Giovannoni et al., 2005). Several unusual instances of metabolic pathway reduction

have been described in microorganisms with streamlined genomes. For example, SAR11 genomes lack complete assimilatory sulfate reduction pathways, which makes them dependent on reduced sulfur sources such as DMSP and methionine (Tripp et al., 2008). Consistent with this finding, others have shown that HTCC1062 contains a functional *dmdA* homolog which, when expressed in *E. coli*, was capable of demethylating DMSP to MMPA (Reisch et al., 2008), and that in the Sargasso Sea, SAR11 cells actively assimilate DMSP (Malmstrom et al., 2004). Recently, Grote and co-workers showed that genome streamlining is a common feature of SAR11 strains, and that these genomes retain a higher percentage of genes in the core genome (genes present in all known sequenced strains) at that level of divergence than any previously examined clade (Grote et al., 2012). All seven sequenced SAR11 strains lack a complete assimilatory sulfate reduction pathway and most share the DMSP demethylation genes *dmdA*, *dmdB*, and *dmdC* (Grote et al., 2012).

In this study we report that axenic cultures of HTCC1062 fed with DMSP produced both DMS and MeSH, and explore the underlying mechanisms with a combination of comparative genomics, quantitative proteomics and *in vitro* biochemical assays. Orthologous substitutions of three different lyase enzymes appeared across SAR11, including a new DMSP lyase, DddK, which we describe here for the first time. These findings indicate that SAR11 are likely major worldwide producers of DMS.

3.3 Results and Discussion

Physiology of DMSP metabolism

The findings of (Tripp et al., 2008) demonstrated that DMSP could relieve sulfur limitation in HTCC1062, analogously to methionine. Because of this work and the demonstration of a functional DmdA protein in HTCC1062 by Reisch (Reisch et al., 2008), we hypothesized that DMSP sulfur might be either be converted to methionine through the demethylation pathway, metabolized to MeSH, or oxidized to sulfate by an unknown pathway. To test these genomic predictions, we examined the decline of DMSP and the accumulation of sulfur products in washed cell suspensions of HTCC1062 cells in artificial seawater medium (Schwalbach et al., 2009) with 1 µM DMSP as the sole sulfur source. The DMSP concentration of 1 µM used in these experiments is within the range of natural variation: DMSP concentrations in phytoplankton blooms have been measured at concentrations between nanomolar to almost 2 µM (van Duyl et al., 1998). Surprisingly, DMS was the most abundant end-product that was detected, with MeSH appearing as a minor product of DMSP metabolism (Fig. 3.1A). The data showed that both DMS and MeSH were produced by HTCC1062 cultures after adding DMSP, and the amount of DMS and MeSH increased linearly within 18 hours of incubation. Decline in DMSP concentration and DMS production by killed cell controls was either low or non-detectable. Over 80% of the DMSP decrease could be accounted for, with 59% going to DMS, 21% to

MeSH, and ~1% DMSP being required for cellular sulfur demands (estimated according to (Tripp et al., 2008))(Table 3.1). The imbalance in the stoichiometry could be due to measurement errors, oxidation, or to DMSP accumulation within cells. Regardless, the data shows that HTCC1062 cultures were capable of converting significant quantities of DMSP to DMS and methanethiol.

Although unexpected, our data are consistent with the estimate that 10% of DMSP is converted to DMS worldwide (Kiene and Linn, 2000). We observed ~60% of DMSP converted to DMS by HTCC1062 cells in culture. With SAR11 cells generally comprising ~25% of total bacterioplankton worldwide (Morris et al., 2002; Schattenhofer et al., 2009), if we assume 60% conversion of DMSP to DMS in these populations, it would be expected that ~15% of DMSP would be converted to DMS. (Malmstrom et al., 2004) have shown that ~50% of SAR11 cells in a given population are actively incorporating sulfur from DMSP, indicating significant metabolic variation within a given population. Examined in this context, our measurements suggest that a significant percentage of the naturally produced DMS in any given seawater sample may be the result of SAR11 activity.

In addition to the novel finding that HTCC1062 has a DMSP cleavage phenotype, MeSH production by cell suspensions demonstrated that the demethylation pathway was simultaneously active (Fig. 3.1A), however, the amount of MeSH is much less than the DMS production. We therefore hypothesized that HTCC1062 could use MeSH as a sole

sulfur source. To test this, cultures were grown in artificial seawater medium with 100 nM methionine, or DMSP, or MeSH, or DMS, or no added sulfur source. Cultures amended with MeSH grew to higher densities than those grown on DMS or without added sulfur, although growth was not as great as in cultures grown with equivalent amounts of methionine (Fig. 3.1B). We speculate that the volatility of MeSH and its susceptibility to oxidation might have reduced the amount available to cells. Thus, we conclude that the demethylation pathway in HTCC1062 channels reduced sulfur from DMSP to metabolism, and operates simultaneously with a lyase pathway that releases DMSP sulfur as DMS, a molecule these cells are unable to use.

Comparative genomics of DMSP metabolism pathways

Cleavage pathway: Although HTCC1062 produced DMS, hidden Markov model (HMM) searches found no homologs of known DMSP lyases in the genome, consistent with previous reports (Moran et al., 2012; Reisch et al., 2011b). However, inspection of the genome of this strain showed that gene SAR11_0394 encoded a polypeptide whose C-terminal region was predicted to include a cupin, a very widely distributed protein fold that resembles a small barrel (Dunwell et al., 2004). Of the DMSP lyases identified to date, three (DddL, DddQ and DddW) also have a C-terminal cupin with only very limited sequences identities to each other, and to the corresponding region of the SAR11_0394

gene product (Fig. S3.1). We therefore set out to test if this gene did encode a functional DMSP lyase by arranging for its *de novo* synthesis, with a ribosomal binding site and codon usage optimized for expression in *E. coli* pET16*.* The resulting plasmid was transformed into *E. coli* strain BL21, which, when assayed for its ability to cleave DMSP, was shown to cleave this substrate into DMS plus acrylate, as demonstrated by gas chromatography and NMR respectively (data not shown).

In light of these observations, the SAR11_0394 gene was termed *dddK.* It was confirmed that the DddK polypeptide was a DMSP lyase by purifying a C-terminal His-tagged version of it, which was isolated to near-purity (Fig. S3.2) and assayed *in vitro* as described in supplementary material. Using NMR it was shown that, as expected, the DddK polypeptide cleaved the DMSP into DMS plus acrylate, in equimolar ratios. Measurements of the kinetics of this reaction showed that there was a V_{max}, (3.5 nmol DMS min^{-1} g protein $^{-1}$) but a very high K_m (96 mM) (Fig. S3.3).

The Km value of DddK for DMSP is much higher than that of the DMSP demethylase DmdA (13.2 ± 2.0 mM) (Reisch et al., 2008). This observation may explain the relationship between the DMSP cleavage and DMSP demethylation pathways in SAR11 cells. In such a model, SAR11 uses DMSP to fulfill sulfur requirements through demethylation to methanethiol and ultimately methionine, and any additional sulfur is shunted through the DMSP cleavage pathway to DMS. DMSP cleavage by DddK may act

as a kinetically-triggered ‘safety valve’, and would not occur until cells had crossed a minimum DMSP concentration threshold that fulfilled their sulfur requirements. Above this threshold (perhaps 96 mM), increasing amounts of DMSP would create a linear increase in DMS production, similarly to the results obtained for *R. pomeroyi* (Gonzalez et al., 1999), in which similarly high intracellular concentrations of DMSP (70 mM) have been observed (Reisch et al., 2008).

We noted that a DddK-like polypeptide (74% identity) also occurs in the deduced proteome of two other genome-sequenced SAR11 strains, the HIMB5_0000473 gene product of strain HIMB5 and Pu1002_04381 of HTCC1002; no close homologues were seen in any other organisms. However, other SAR11 strains do have genes that encode homologues of other families of enzymes with known DMSP lyase activity. Two strains had DddP-type lyases in the M24 family of metallo-peptidases (Todd et al., 2012), encoded by PB7211_1082 and HIMB59_00005110 in SAR11 strains HTCC7211 and HIMB59 respectively; these were 47% identical to each other and ~50% identical to the ratified DddP2 lyase of *Oceanimonas doudoroffii.* One SAR11 strain (HIMB5) also had a gene (HIMB5_00000220) whose product had low–level sequence similarity (28% identity) to the DddQ DMSP lyase of *Ruegeria pomeroyi,* the similarity being greatest around the cupin domains of these two polypeptides (Fig. S3.1).

The PB7211_1082 (*dddP*-like) and HIMB5_00000220 (*dddQ*-like) genes were

synthesized and cloned as described and cloned into pET16 as described above for *dddK*, and the resultant *E. coli* transformants assayed for DMSP lyase activity. The rates of DMS production with the cloned *dddQ*-like gene were similar to those with the *dddK* (above) but, under the conditions used, the cloned PB7211_1082 (*dddP*-like) showed only 20% of that activity.

Genes that encode enzymes for the metabolism of acrylate were present in most SAR11 strains (Fig. 3.2). Acrylate can be metabolized to 3-HP by the action of AcuNK (Todd et al., 2012). However, while AcuK is found in all strains, AcuN is not considered a core gene among SAR11. 3-HP can be oxidized to an intermediate, malonate semialdehyde (mal-SA) and then acetyl-CoA, by an alcohol dehydrogenase (DddA) and mal-SA dehydrogenase (DddC), respectively. The predicted homologs for both *dddA* and *dddC* are found in all strains. *yhdH*, a homolog of *acuI* which has recently been implicated in reductive 3-HP metabolism in *Rhodobacter sphaeroides* and *R. pomeroyi* and proposed as part of a novel pathway that converts acrylate to propionyl-CoA via acrylyl-CoA (Schneider et al., 2012; Todd et al., 2012). This gene has homologs in all strains but HIMB59. The enzyme that converts acrylate to acrylyl-CoA has recently been identified as a propionate-CoA ligase (PrpE) in *R. pomeroyi* (Reisch et al., 2013). PrpE carries out multiple functions, which also involved in a third pathway for acrylate degradation via transformation to propionate and propionyl-CoA by acrylate reductase and PrpE, respectively. Acrylate reductase is missing but PrpE is present in all SAR11

strains.

Demethylation pathway: Consistent with physiological data showing production of methanethiol, HTCC1062 and all other sequenced SAR11 strains, except those in the more distantly related group IIIa (Grote et al., 2012; Vergin et al., 2013), contain homologs of the *dmdABC* genes found in *Ruegeria pomeroyi* (Fig. 3.2). However, none of the sequenced SAR11 strains had homologs of *dmdD*, which is reflected in the low abundance of this gene in ocean metagenomic databases (Moran et al., 2012). As in HTCC1062, *dmdD* is not required for complete demethylation of DMSP to MeSH in *R. lacuscaerulensis* (Reisch et al., 2011). Both strains may have a novel methylthioacryloyl-CoA hydratase, or analogous enzyme that carries out the same function. The overall presence of the DMSP demethylation pathway in most SAR11 strains is consistent with the distribution of other C1 and methyl group oxidation pathways (Sun et al. 2011), which are found in all strains except those of the subclade IIIa (Grote et al., 2012). The *dmdA* reaction apparently provides SAR11 with another means of obtaining energy via the tetrahydrofolate pathway for C1 oxidation. All SAR11 strains contain homologs of cystathionine-ɣ-synthetase, predicted to catalyze the conversion of MeSH to methionine in bulk seawater samples (Kiene et al., 1999). The presence of this homolog is consistent with the observed ability of HTCC1062 to grow with MeSH as the sole sulfur source.

The utilization of the DMSP three carbon moieties

DMSP lyases are known either to produce the three-carbon compounds acrylate (DddL,Q,W,Y)) or 3-HP (DddD) as products of DMSP cleavage (Curson et al., 2011). During metabolic reconstruction, we noted that the HTCC1062 genome has annotated genes for all steps in the degradation of 3-HP to propionyl-CoA, as well as the gene *acuI*, which catalyzes acrylate conversion to propionyl-CoA (Fig. 3.2) and which is conserved in most SAR11 genomes. Therefore we tested the efficacy of 3-HP, acrylate, and propionate as substitutes for pyruvate in pure cultures of HTCC1062. Acrylate and propionate resulted in considerably increased cell growth relative to negative controls, but with 3-HP, the enhancement of growth was extremely limited, but statistically significant (student's t test, n=3, $P<0.05$), However, none of these compounds were as effective as pyruvate (Fig. 3.3).

Expression of DMSP pathway genes in quantitative proteomic data

Comparative analysis of the HTCC1062 proteomes of cultures grown in the presence of DMSP by ITRAQ failed to show major upregulation of proteins with predicted involvement of DMSP metabolism (Fig. 3.2) compared to cultures grown in the presence of methionine. 12 proteins were identified as having a differential expression of > 1.5-fold between DMSP and Methionine treatments (Table 3.2, Fig. S3.4). Only two proteins in the

predicted putative pathways of DMSP cleavage and demethylation were identified in the significantly differentially expressed proteins, but with small changes in protein abundance: DmdC was 25% more abundant in cultures amended with DMSP than those amended with methionine and DddC was 20% more abundant in cultures amended with methionine than those amended with DMSP.

Extensive proteomic analysis under a wide range of conditions in previous work suggests that the proteome of HTCC1062 is largely constitutive, thus the limited response shown by ITRAQ analysis in this study is not unexpected. The relative increase of DddC expression under methionine treatment was surprising as this protein catalyzes a key step in the demethylation of DMSP to methionine via MMPA and MeSH. It may be implied that DddC is a multi-functional protein. As MetF catalyzes the conversion of CH_3-THF to CH_2-THF, upregulation of MetF in the presence of DMSP is consistent with increased concentrations of CH_3-THF resulting from conversion of DMSP to MMPA by DmdA. Similarly, GcvT is required alongside FolD in the conversion of CH_2-THF to CHO-THF (Fig. 3.2). SAR11_1724 is a protein of unknown function containing a YGGT domain conserved among integral membrane proteins of unknown function. Surprisingly, there was no evidence of expression of DddK in the proteomic data, perhaps indicating that this protein has poor extraction efficiency, poor ionization and/or is constitutively expressed at very low levels in HTCC1062. Analysis of previous MS/MS datasets for HTCC1062 (Smith et al., 2010; Smith et al., 2013) also yielded little evidence of DddK

expression. Low, constitutive expression of DddK would likely be consistent with a putative role as a ‘safety valve’ for the export of excess DMSP.

Quantative proteomics provided evidence of upregulation of PepQ (SAR11_0687) under DMSP growth conditions. As structurally similar creatinases have previously been found to have DMSP lyase activity (Kirkwood et al., 2010), SAR11_0687 was synthesized, cloned and overexpressed in *E. coli* as described previously. However, SAR11_0687 showed no evidence of DMSP lyase activity (data not shown), therefore it is unlikely that this protein is responsible for DMSP cleavage in HTCC1062.

Cultures amended with DMSP showed a 70% increase in the production of proteorhodopsin compared to those grown on methionine. Recent work by Feng and colleagues showed that proteorhodopsin abundance in *Psychroflexus torquis* is dependent on salinity level and the growth of this microorganism is light stimulated under osmotic pressure (Feng et al., 2013). We therefore proposed that the presence of DMSP may affect the salinity level of the media, which result in the increase of proteorhodopsin in SAR11 cells.

Metagenomic analysis of DMSP pathway genes

Previous work demonstrated the ubiquity of the demethylation pathway genes in ocean surface waters, which was largely attributed to the presence of these genes in

SAR11 (Reisch et al., 2011b). We compared the abundance of the genes for DMSP cleavage with those for demethylation (*dmdABC*) in the Global Ocean Survey (GOS) dataset (Fig. 3.4). DMSP lyases were much less abundant than *dmdABC* or the single-copy marker *recA*. These data support the interpretation that either the cleavage pathway is less important than the demethylation pathway, or undiscovered DMSP lyase analogs are present in other SAR11 strains. Genes for metabolism of acrylate and 3-HP were in higher abundance than the lyases, and similar in abundance to demethylation genes and *recA*, supporting the interpretation that they might be very common or perhaps universal in SAR11 genomes. It is noteworthy that genes for sulfate assimilation are absent from most SAR11 genomes, potentially providing a strong evolutionary driver for retention of the demethylation pathway (Grote et al., 2012; Tripp et al., 2008).

Data presented here support the conclusion that both the cleavage and demethylation pathways function in HTCC1062, and may be constitutively expressed. Because the cells were grown on 1 µM DMSP prior to harvesting, one interpretation of the immediate production of both DMS and MeSH is that both the demethylase and DMSP lyase genes were induced in the culture. The concentration of DMSP used in these experiments, although ecologically relevant, is much higher than needed to satisfy the sulfur requirements of the cells (Tripp et al., 2008), and thus DMS production could be an energetically favorable route to disposing any DMSP surplus. However, most proteins involved in both DMSP catabolic pathways were detectable in the proteomes of cells that

were grown using methionine as a sulfur sources, and there were no significant changes in the abundance of these proteins, relative to RecA protein, over a wide range of experimental conditions. Evidence for constitutive expression of both pathways in SAR11 was also presented by (Vila-Costa et al., 2010), who tracked transcripts from enrichment cultures of BATS seawater with 25 nM added DMSP. Despite increases in transcription from some microbial groups, neither the SAR11 nor Roseobacter species showed significant changes in gene expression. However, the results of that study may not accurately reflect the potential for regulation of DMSP metabolism in SAR11 cells because the enrichment experiment was performed on a very short time scale (30 min) relative to SAR11 growth rates, which typically are 0.40-0.58 d^{-1} (Morris et al., 2002) (Rappé et al., 2002). We conclude that the data now available favor the model that expression of both the cleavage and demethylation pathways are constitutively expressed in HTCC1062, raising questions about the potential for kinetically determined regulation of sulfur flux through the alternate pathways.

3.4 Conclusions

Here we report the unexpected result that SAR11 strain HTCC1062 was capable of degrading DMSP by both the demethylation and cleavage pathways, with production of the two volatiles, DMS and MeSH. Three DMSP lyase genes, which appear to be analogously substituted among SAR11, *dddP*, *dddQ*, and a novel lyase, *dddK*, can confer

a Ddd+ phenotype to *E, coli* and, in the case of *dddK,* its product was purified and shown to catalyze DMSP cleavage to DMS *in vitro*. The varied distribution of lyase genes across multiple different genomes, as well as the relative abundance of these genes and those of catabolic pathways for C3 metabolism in global ocean datasets, suggest that native SAR11 populations harbor a suite of analogous DMSP lyases and share the DMS production phenotype.

Because of the possible role of DMSP/DMS cycling on cloud nucleation and the vast population sizes of SAR11, the information procured by this research will contribute to refinement of biogeochemical models that predict impacts on climate change on weather and to understanding the biotransformations in the global sulfur cycle. Identifying these metabolic pathways and the genes that comprise them will lead to more accurate interpretations of genomic and metagenomic data pertaining to sulfur metabolism in future studies.

3.5 Materials and Methods

Bioinformatics analysis and proposed DMSP metabolic pathways

We expanded on the knowledge obtained in (Grote et al., 2012) by doing additional homology searches for DMSP metabolism genes using profile hidden markov models (HMMs) (Eddy, 2004). Because they are constructed with a range of probabilistic values for a given site in a protein, profile HMMs are superior to BLAST for finding

distantly related homologs (Eddy, 1998). In this workflow, representative genes for the reactions in Fig. 3.2 were obtained from the original publications (Boden et al., 2011; McDevitt et al., 2002; Todd et al., 2012), searches of NCBI, and E.C. number searches based on figures from (Curson et al., 2011; Reisch et al., 2011a), and references therein. These representative sequences were then searched (using hmmscan of the HMMER3 package (Eddy, 2011) against a database of profile HMMs created for over 436,000 protein families built with Markov clustering (Sharpton et al., 2012). Each SFam contains hundreds to thousands of representative sequences, and thus is an ideal training dataset for construction of profile HMMs. SFam HMMs with lowest expect values to the representative sequences were then searched against our SAR11 genomes. Homologs were classified based on HMMs having comparative expect values to both the representative sequence and a SAR11 gene sequence.

Measurements of products of DMSP metabolism and decline of DMSP:

HTCC1062 for this study were grown in autoclaved, filtered artificial seawater (ASW) amended with 1 mM NH_4Cl, 100 µM KH_2PO_4, 1 µM $FeCl_3$, 100 µM pyruvate, 50 µM glycine, 1 µM DMSP, and excess vitamins (ASW media-A). Cultures were harvested by centrifugation, washed once and resuspended in ASW. The cells (final concentration is ~1.5×10^6 cells mL^{-1}) are distributed into 20 mL sealed vials (10 mL /vial). Inject 1 µM

DMSP (Final concentration) in vials and incubate in the dark at 16 °C. Biological activities were stopped by 0.1 M sodium azide (100 µL/vial) at 0 min, 20 min, 1 hr, 3 hr, 9 hr and 18 hr incubation. Duplicate samples were refrigerated before analysis. Killed cell cultures (1 vial for each time point) were used as negative controls.

DMS, MeSH and other individual volatile sulfur compounds (hydrogen sulfide, thioacetate) were analyzed using SPME-GC-PFPD method (Fang and Qian, 2005; Vazquez-Landaverde et al., 2006). Briefly, an internal standard solution (EMS in methanol) was added into the vial before sample analysis. An 85 µm Stableflex Carboxen™ / Polydimethylsiloxane (CAR/PDMS) fiber (Supelco Inc.) was used to extract sulfur compounds, and CombiPAL system (CTC Analytics) was used to control the extraction process. Samples were pre-incubated for 5 min at 30ºC with 500 rpm agitation, and then fiber was inserted in the vial headspace for 15 min at 30ºC with 250 rpm agitation. The fiber was desorbed in injector for 6 min at 300ºC. Chromatographic analysis was performed using a Varian CP-3800 GC with a pulsed flame photometric detector. Separation was performed using a DB-FFAP column (30 m x 0.32 mm x 1 µm). DMSP will be quantified by measuring released DMS after hydrolysis in NaOH (0.1 M final concentration, 12 hrs) at room temperature.

Test of DMS & MeSH utilization in HTCC1062

HTCC1062 were cultured in 40 mL clear sealed vials (Thermo scientific, part no. 340-40C/CT) with autoclaved, filtered ASW amended with 1 mM NH_4Cl, 100 μM KH_2PO_4, 1 μM $FeCl_3$, 100 μM pyruvate, 50 μM glycine, 250 μM methanol, excess vitamins and 100 nM four different sulfur sources (DMSP, methionine, DMS and MeSH). Each vial had 10 mL aliquot. Vials were incubated on a shaker in 16℃ room. Cell densities were monitored by Guava (Tripp, 2008).

Test of C3 compounds utilization in HTCC1062

HTCC1062 cells were grown in autoclaved, filtered ASW amended with 100 μM NH_4Cl, 10 μM KH_2PO_4, 100 nM $FeCl_3$, 50 μM glycine, 50 μM methionine, and excess vitamins (ASW media-B). Each compound (3-HP, acrylate or propionate) was tested at a concentration of 10 μM in ASW media-B. The positive control was amended with 10 μM pyruvate. The negative control contained no pyruvate. The concentrations of compounds utilized in assays were chosen by testing the growth of HTCC1062 with differing concentrations of each compound (Fig S3.5) and selecting concentrations that were not inhibitory (data not shown).

Preparation of HTCC1062 ITRAQ samples

HTCC1062 was grown in ASW amended with 100 μM NH_4Cl, 10 μM KH_2PO_4, 0.1

μM $FeCl_3$, 1 mM pyruvate, 500 μM glycine and excess vitamins. Triplicate Samples were amended with 1 μM DMSP, 1 μM methionine and samples with both 1 μM DMSP and 1 μM methionine are treated as positive controls. Cells were all harvested by centrifugation at the same time point in the exponential phase. Prior to harvesting, cultures were treated with chloramphenicol (0.01 g/L) and protease inhibitor cocktail Set II (0.1 mL/L). Cell pellets were immediately stored in - 80℃ prior to ITRAQ (isobaric tag for relative and absolute quantitation) analysis at the Pacific Northwest National Laboratory (PNNL).

Metagenomic analysis

To identify the relative abundance of SAR11 genes involved in DMSP metabolism in surface water metagenomes, predicted proteins encoded by homologs of *acuIKN, dddACKPQ, dmdABC* and *prpE* were identified in all 14 genomes from the SAR11 clade (HTCC1002, HTCC1013, HTCC1062, HTCC7211, HIMB5, HIMB59, HIMB058, HIMB083, HIMB114, HIMB140, HTCC8051, HTCC9022, HTCC9565, IMCC9063) currently in the complete Integrated Microbial Genomes (IMG, http://img.jgi.doe.gov/) database (v. 400). Homologs were determined using a previously described comparative genomics analysis pipeline (Grote et al., 2012). Genes within each cluster were used as queries in a TBLASTN (v. 2.2.22+) search against the Global Ocean Survey (GOS) nucleotide database available from CAMERA (http://camera.calit2.net/) with query filtering disabled and default *e*-value cutoff (-seg no –max_target_seqs 10000000). Nucleotide sequences

returned from this search were used in a reciprocal best-BLAST (RBB) (Wilhelm et al., 2007) filtering step against the amino acid sequences in the complete IMG database, returning the best hit to each nucleotide query (BLASTX, -max_target_seqs 1 –seg no). If the best hit for a nucleotide sequence in the RBB analysis was a protein sequence from the original gene cluster, the nucleotide query was recorded as a successful hit; otherwise it was rejected. To normalize gene abundance, counts to the single-copy marker gene *recA* were also determined in the same manner.

3.6 Acknowledgements

Jing thanks financial support from the China Scholarships Council (CSC). The authors would like to thank Dr. John W. H. Dacey for kindly providing DMSP. We are grateful to Dr. John W. H. Dacey and Dr. Samuel Bennett for the advice of designing the methods of DMSP measurements. We also thank Mary Lipton, Sam Payne, Dick Smith and the researchers of Pacific Northwest National Laboratory (PNNL) for generating and providing the datasets discussed herein.

Table 3.1. Mass balance calculated from Figure 3.2.

	Concentration (nM)	%
Δ DMSP	− 122.12	100%
Δ DMS	+ 72.47	59.3%
Δ MeSH	+ 26.10	21.4%
Δ Cellular sulfur (est.)[a]	+ 0.86	0.7%
Missing sulfur		18.6%

[a]Estimated cellular sulfur demands according to previous studies (Tripp et al., 2008).

Table 3.2 Proteins with differential expression > 1.5-fold between HTCC1062 cultures amended with methionine vs. cultures amended with DMSP as determined by quantitative ITRAQ proteomics. Fold change was calculated using the LIBRA module of the Trans-Proteomic Pipeline and by linear mixed-effects model encompassing a fixed treatment effect and random effect for each peptide associated with the protein. Bold-text indicates proteins enriched in DMSP amended cultures.

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

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

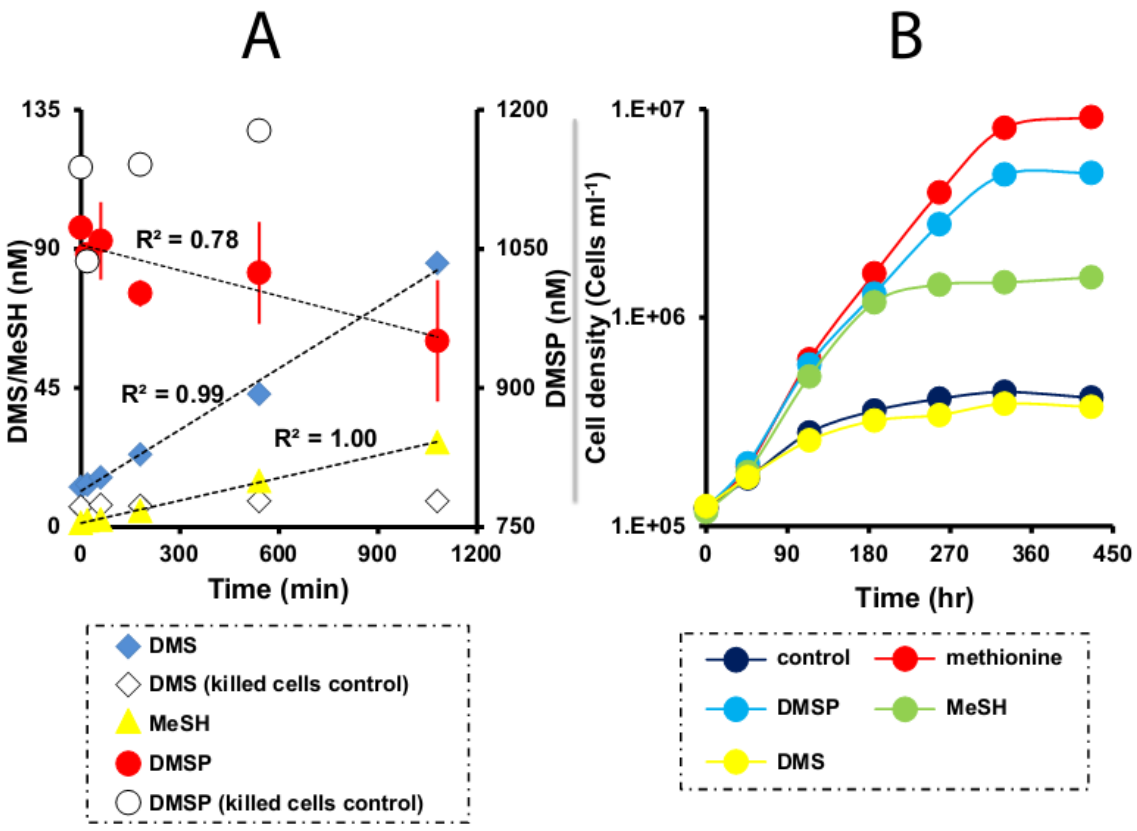


Figure 3.1 DMSP metabolism in HTCC1062. A) DMS and MeSH production from DMSP (Left Y axis) and DMSP decline (Right Y axis) in HTCC1062 culture. Results are the average of duplicate samples and error bars show the range of the duplicates. When the error bars are invisible, they are smaller than the size of the symbols. Killed cells controls were performed in single vials. The signals of MeSH from killed cell controls were not detectable. **B)** Utilization of MeSH and DMS in HTCC1062. Cells were grown in artificial seawater media (ASW). Cultures were incubated in ASW amended with methionine (positive control), DMSP, MeSH, or DMS. A culture without any sulfur source was treated as negative control. Each point represents a single vial, and the experiments were repeated three times, with similar results.

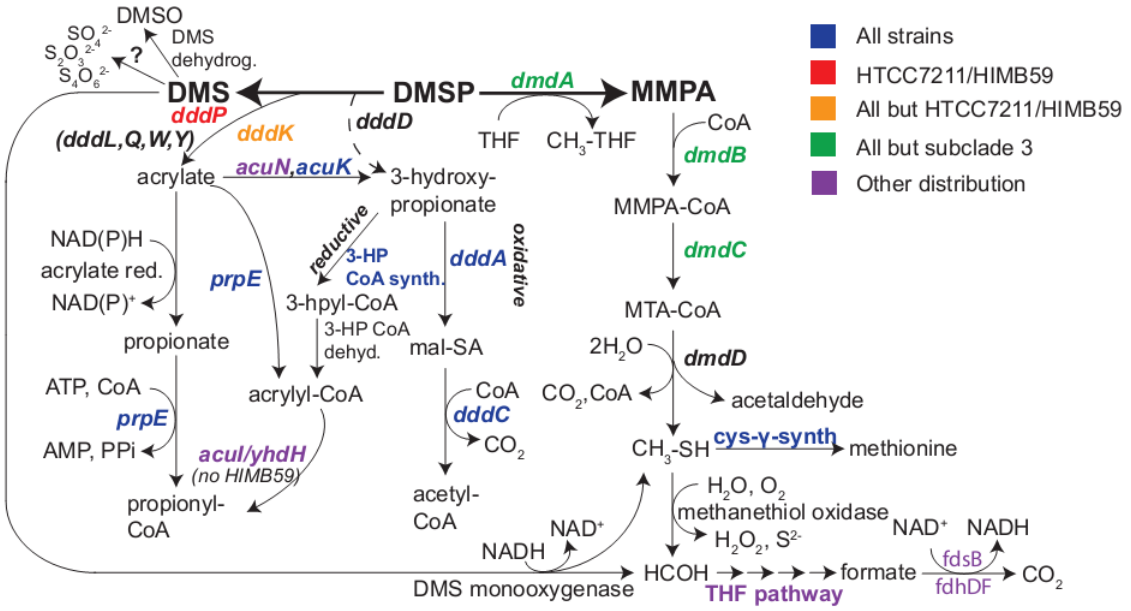


Figure 3.2 DMSP catabolic pathways and homologs identified in eight SAR11 genomes. Dashed line indicates a proposed pathway. Question marks indicate unknown enzymes. Genes in black did not have identified homologs.

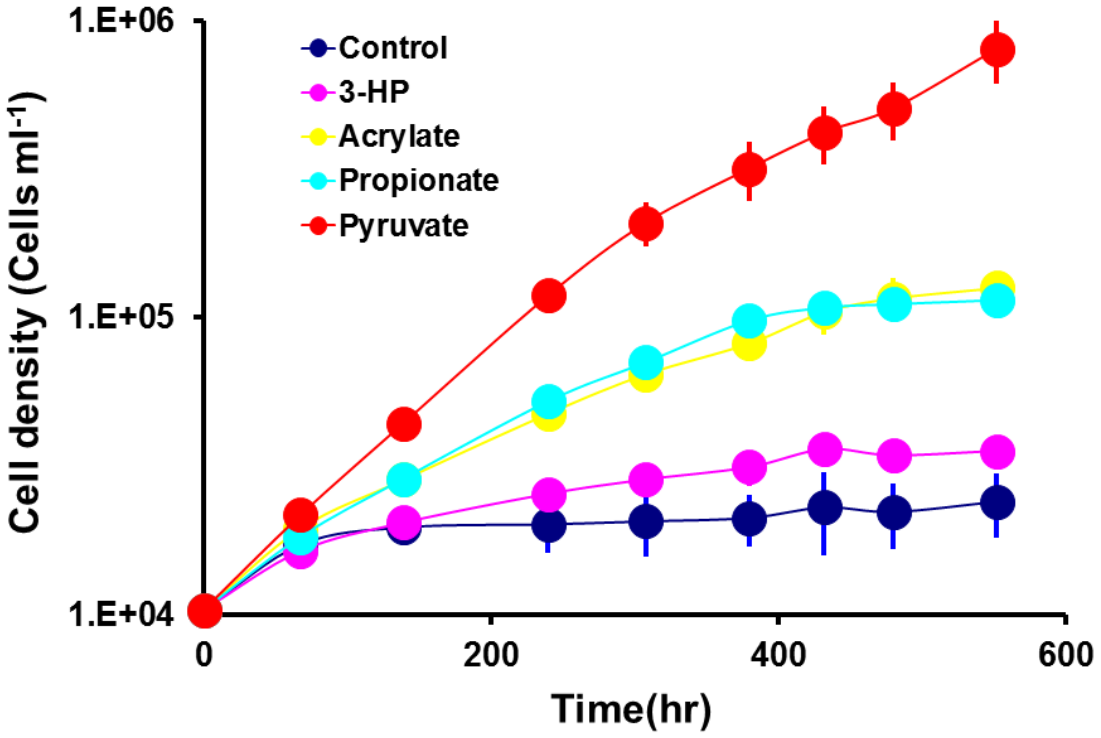


Figure 3.3 Utilization of C3 compounds in HTCC1062. Cells were grown in artificial seawater media (ASW). Cultures were incubated in ASW amended with 10 M pyruvate (positive control), 3-HP, acrylate and propionate. The culture without pyruvate was treated as negative control. Points are the average density of triplicate cultures. Where not visible, error bars are smaller than the size of the symbols.

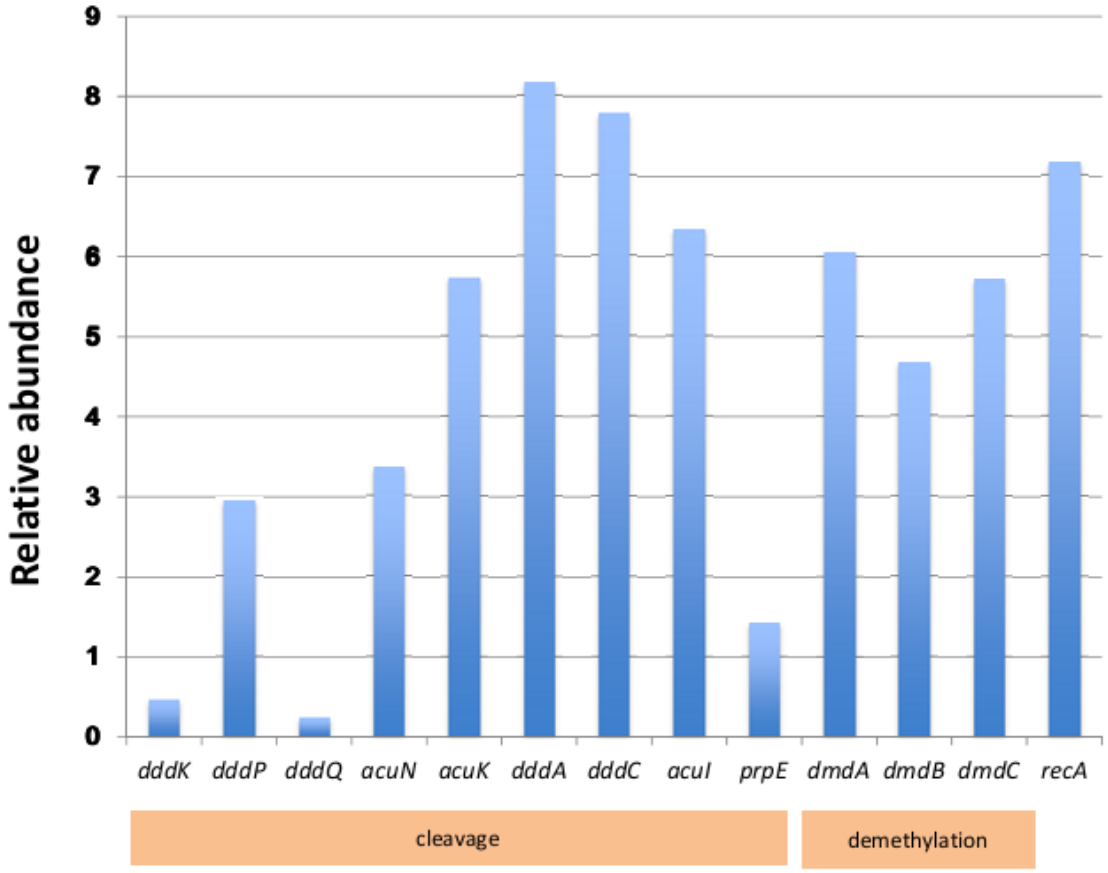


Figure 3.4 The relative abundance of SAR11 DMSP catabolism genes in GOS data. Genes were identified as SAR11 clades by a reciprocal best BLAST (RBB) approach. The frequency of the single copy *recA* gene was used to evaluate the abundance of the DMSP metabolism genes.

Chapter 4

4. Transcriptional Regulatory Responses to C1 and Methylated Compounds and A Predicted Enzyme in Trimethylamine N-oxide Metabolism of the Globally Distributed Marine Bacterium *Candidatus* Pelagibacter ubique HTCC1062

Jing Sun and Stephen J. Giovannoni

4.1 Abstract

The SAR11 clade is one of the most successful groups of microorganisms on Earth. Recent studies have shown that these organisms oxidize a wide range of one carbon (C1) and methylated compounds that are ubiquitous in the oceans. However, the metabolic pathways used to oxidize and assimilate these compounds are complex and have been only partly described. To understand the metabolism of these compounds in SAR11 and to identify candidate genes involved in these pathways, we used a high density oligonucleotide-based microarray to study changes in gene expression in response to five different compounds (trimethylamine N-oxide (TMAO), methylamine, dimethylsulfoniopropionate (DMSP), methanol, and glycine betaine (GBT)) in SAR11 strain HTCC1062. Only one differentially expressed gene SAR11_1304, annotated as a monomeric sarcosine oxidase (MSOX), was identified in the TMAO treatments, while no changes of gene expression were observed in treatments with the other four compounds. The upregulation of gene SAR11_1304 in response to TMAO was confirmed by SYBR green I based quantitative reverse transcription PCR. SAR11_1304 was cloned and overexpressed in *Escherichia coli*. The purified protein showed TMAO demethylase and dimethylamine (DMA) monooxygenase activities, suggesting that this enzyme catalyzes multiple functions in the SAR11 TMAO degradation pathway.

4.2 Introduction

SAR11 *Alphaproteobacteria* (the order *Pelagibacterales*) are the most abundant heterotrophs in the oceans (Morris et al., 2002). Due to their small genomes and the vast population sizes, SAR11 are regarded as important model organisms for studying the oxidation of dissolved organic matter in the ocean environments. In previous work, we found that SAR11 can oxidize a variety of one carbon (C1) and methylated compounds, including trimethylamine N-oxide (TMAO), dimethylsulfoniopropionate (DMSP), glycine betaine (GBT) (Sun et al., 2011). However, the specific metabolic pathways by which these compounds are metabolized in SAR11 are still unknown.

Microarrays are a powerful tool for studying whole-cell responses at the transcriptional level to different treatments and provide information about the potential functions of genes and the proteins they encode. Several platforms are currently available, the most common of which is high density oligonucleotide-based Affymetrix GeneChip®arrays. Usually a single Affymetrix chip contains thousands of spots, each representing a single gene and collectively the entire gene repertoire of an organism. This technique allows comparisons of the expression levels of thousands of genes simultaneously, providing a systems perspective on regulation of gene expression. In this chapter, we aimed to identify genes that function in mineralizing several C1 and methylated compounds. To achieve this, differential expression of genes in strain HTCC1062 exposed to a variety of compounds was screened by Affymetrix microarray

technology, and gene expression was confirmed with an alternate experimental approach.

4.3 Results

Microarray Results

Initial screening of gene expression responses to C1 and methylated compounds was undertaken to provide insight into metabolic pathways used for catabolism of a set of C1 and methylated compounds that are thought to be significant in marine environments. We tested changes in expression in response to addition of five different C1 and methylated compounds (TMAO, methylamine, DMSP, methanol, and GBT), all of which are known to be used by SAR11 strain HTCC1062 (Sun et al., 2011; Tripp et al., 2008). Microarray data showed that, across the entire set of six experimental variables, only one gene (SAR11_1304) was significantly upregulated. When treated with TMAO, the fold change was 6.63 ± 0.38 (Table 4.1). This TMAO-induced gene is annotated as a putative monomeric sarcosine oxidase. No changes of gene expression were observed in the other treatments (data not shown). The other research by running this Affymetrix data through the limma package is performed, and the results supported that SAR11_1304 is the top-hit in TMAO data, however, this gene is 15-fold more abundant in the presence of TMAO, suggesting the limma test is more sensitive than the standard analysis approach (Temperton, unpublished).

Quantitative reverse transcription polymerase chain reaction (RT-qPCR) validation

Commonly employed microarray techniques can produce false positives, calling into question on the reliability and accuracy of the expression data of individual genes. By incorporating a validation step, through the use of qPCR, this issue can be eliminated. QPCR is also cost- and time-effective, because usually only a small subset of all genes represented on microarray are differentially expressed. Therefore, qPCR is often used as the final step in any microarray protocol to ensure the accuracy and repeatability of the data prior to further analysis.

In this study, relative quantification was performed to measure and validate the fold change in expression of gene SAR11_1304. For accuracy, it is essential to normalize *real-time PCR* data to a fixed *reference gene.* RecA is a common housekeeping gene that has been used as reference gene in studies on various bacteria (Takle et al., 2007). The *recA* gene was (SAR11_0641) selected for RT-qPCR validation based on its constant expression across all samples in microarray data (coefficient of variation =0.7 %). RT-qPCR showed that this gene was constitutively expressed across all growth conditions (Fig. 4.1). Therefore, we concluded that it would serve as a suitable reference gene for accurate normalization of RT-qPCR data.

The RT-qPCR results confirmed observations made with microarrays. SAR11_1304 was upregulated when TMAO was present, and showed similar expression changes 8.55 ± 1.30 (Fig. 4.2). The data also revealed that the fold changes of this gene were slightly

higher in RT-qPCR than in microarrays, which might be due to the differences in the detection methods and probe design between the two techniques.

Cloning, Expression and Characterization of MSOX

The microarray data revealed that SAR11_1304 was upregulated by TMAO, but not by methylamine, which suggested its role in TMAO metabolism. Therefore we postulated that the MSOX encoded by SAR11_1304 functions in TMAO demethylation or DMA oxidation or both metabolic pathways (Fig. 4.3).

To identify the function of SAR11_1304, we cloned this gene from HTCC1062 into vector pET28b, overexpressed it in *E. coli*, and purified the resulting polypeptide by affinity chromatography. Further analysis of these fractions by SDS-PAGE and western blotting revealed a substantially pure protein of ~50 kDa (Fig. 4.4).

DMA monooxygenases, such as the predicted SAR11_1304 monomeric sarcosine oxidase, are NADPH-dependent enzymes that catalyze the following reaction: $(CH_3)_2NH + NADPH + H^+ + O_2 \rightarrow CH_3NH_2 + H_2O + CH_2O + NADP^+$ (Chen et al., 2010).This gene contains an NADPH binding domain (GXSXXA). Therefore, we tested the NADPH oxidizing activity of MSOX in the presence and absence of DMA. The results were consistent with the prediction that the reaction catalyzed by SAR11_1304 is NADPH dependent, the decrease of NADPH in absorbance at 340 nm was observed when adding DMA, but not TMAO (Fig. 4.5). This observation support the conclusion that gene

SAR11_1304 encodes a protein that has a DMA monooxygenase activity.

When the reaction products were measured by HILIC-ESI-TOF-MS, the data showed that in addition to the substrate TMAO, DMA was present in the reaction mixture (Fig. 4.6), suggesting a TMAO demethylase activity ($(CH_3)_3NO \rightarrow (CH_3)_2NH + CH_2O$). MMA could not be detected in this experiment due to the limitations of the analytical technique (see Materials and Methods).

4.4 Discussion

In this study, our goal was to understand the metabolic pathways and the molecular mechanisms behind the oxidation of C1 and methylated compounds in SAR11. We used microarray technology to screen the potentially functional gene expression at the transcriptional level. Only one gene (SAR11_1304) was found to be upregulated in the presence of TMAO, and no significant differential expression was observed in the other treatments.

The low levels of gene regulation in this organism might be due to its streamlined metabolisms and adaption to oligotrophic metabolism of dissolved organic compounds found in ambient background dissolved organic matter (Serra et al., 2002). Most SAR11 genes are constitutively expressed (e.g. proteorhodopsin, (Steindler et al., 2011)) in order to support their growth and reproduction in complex marine environments. Similar observations were reported in previous studies that compared proteomic and microarray results. Smith et al. showed that mRNA abundance is generally uncorrelated with

changes in protein abundance in SAR11, suggesting that post-transcriptional mechanisms might be acting (Smith et al., 2010). Therefore, quantitative proteomics analyses could lead to the observation of additional regulation of genes involved in the metabolism of C1 and methylated compounds.

Sarcosine oxidase (SOX, EC 1.5.3.1) is a flavoprotein that catalyzes the oxidative methylation of sarcosine (*N*-methylglycine) by molecular oxygen and water to yield glycine, formaldehyde and hydrogen peroxide (Trickey et al., 1999). The native enzyme occurs in heterotetrameric and monomeric forms. In previous research, we noticed the presence of a family of paralogous *sox* genes in SAR11 genomes. The two *sox* operons (*soxBDAG and* $soxB_2D_2A_2G_2$) encode heterotetrameric forms of sarcosine oxidase, some of which might be linked to the pathway for the degradation of GBT through the intermediate sarcosine (Chapter 2; (Sun et al., 2011)). SAR11_1304 is annotated as a monomeric sarcosine oxidase, but its function was not clear. In this study, SAR11_1304 was observed to be upregulated in the presence of TMAO, and the enzymatic assays and direct product measurements demonstrated that SAR11_1304 is involved in metabolism of TMAO, which can be first demethylated to DMA, and then oxidized to methylamine.

After our study was initiated, Lidbury (Lidbury et al., 2014) discovered and identified a different TMAO demethylase (Tdm) in SAR11 strain HIMB59, which is a closely related homolog of gene SAR11_1303 in HTCC1062, and thus most likely has a similar function. SAR11_1303, annotated as an aminomethyltransferase, is adjacent to SAR11_1304 on

the same stand. The function of SAR11_1303 is still unknown, but it shares ~61% sequence similarity at the amino acid level to the Tdm from SAR11 strain HIMB59 and 48% to Msil_3603 from *Methylocella silvestris*, both of which have been identified as TMAO demethylases that convert TMAO to DMA, releasing formaldehyde (Chen et al., 2010). However, in our microarray data, SAR11_1303 expression wasn't significantly increased in the present of TMAO ($P > 0.01$), instead, SAR11_1304 was upregulated by TMAO.

As yet, the biochemistry of TMAO metabolism is incompletely understood. Two scenarios can potentially explain these observations. First is the possibility that both enzymes have TMAO demethylase activity. But, in vivo, Tdm (encoded by SAR11_1303) is the more active enzyme and catalyzes the first step of TMAO demethylation, with the primary function of SAR11_1304 being DMA monooxygenase activity. The enzymatic kinetics (Km and Vmax) of these two proteins for the TMAO demethylation reaction haven't been studied yet, but such experiments in the future could shed light and explain which enzyme is likely to function in the first step of TMAO oxidation in nature.

The second hypothesis is that SAR11_1304 may have other substrates. Other researchers found the high expression of the SAR11_1304 protein and mRNA transcripts when they studied HTCC1062 responses to nitrogen limitation (Smith et al., 2013). Structural analysis with the program I-TASSER confirmed that SAR11_1304 is a monomeric sarcosine oxidase with structural homologies to both amine demethylases and deaminases. Amine demethylases convert secondary amines ($R\text{-}NH\text{-}CH_3$) and

tertiary amines (R-N(CH_3)$_2$) to primary amines (R-NH_2) and deaminases liberate ammonia from compounds with primary amine groups (Smith et al., 2013). Therefore, we speculated that SAR11_1304 encodes a promiscuous enzyme that may be involved in the metabolism of some of the broad range of methylated amine compounds (Fig. 4.7) that are predicted to be transported and metabolized by HTCC1062.

4.5 Materials and Methods

Growth Media and Harvesting

HTCC1062 cells were grown in autoclaved, filtered artificial seawater (ASW), containing 1 mM NH_4Cl, 100 µM KH_2PO_4, 1 µM $FeCl_3$, 80 µM pyruvate, 40 µM oxaloacetate, 40 µM taurine, 50 µM glycine, 50 µM methionine and excess vitamins. Five different C1 and methylated compounds (i.e. 5 µM TMAO or 10 µM methylamine or 1 µM DMSP or 20 µM methanol or 1 µM glycine betaine) were tested. Culture growth was tracked daily by staining cells with SYBR Green and counting on a Guava EasyCyte flow cytometer (Tripp, 2008). Cells used in microarray experiments were collected during log phase. 40 ml of culture from each biological replicate were harvested via centrifugation (1 hour at 20 000 rpm, 10°C), re-suspended in 700 µL of RNAprotect Bacteria Reagent (Qiagen #76506) and allowed to sit for 15 min. Cells were centrifuged again (30 min at 40 000 g, 10°C). After decanting the supernatant, samples were stored frozen at −80°C prior to extraction.

Messenger RNA Preparation

Each cell pellet was resuspended in 100 µL TE (1 mM Tris; 1 mM EDTA; pH 8) containing 40 µg lysozyme and incubated at room temperature for 5 minutes. Total RNA was extracted using RNeasy MiniElute Cleanup kits (Qiagen #74204) according to the manufacturer's instructions, with the exception of an additional wash step with 700 µL Buffer RW1 (Qiagen #1053394) immediately prior to the prescribed washes with Buffer RPE. Eluted RNA was amplified and biotin-labeled using the MessageAmp-II Bacteria RNA Amplification Kit (Ambion #AM1790) and biotin-11-UTP (Ambion) according to the manufacturer's instructions. Resulting amplified and labeled RNA (aRNA) was screened for length and quality using a Bioanalyzer 2100 (Agilent) and quantified utilizing a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific).

RNA Labeling and Microarray Processing

5 µg of biotinylated aRNA from triplicate samples was fractionated and then hybridized (45°C) overnight to custom '*Candidatus* Pelagibacter ubique' Affymetrix GeneChip arrays that contained probes for HTCC1002, HTCC1062 and HTCC7211 using Affymetrix GeneChip Fluidics Station 450, and Affymetrix GeneChip Hybridization Oven 640. Arrays were then washed following manufacturer's instructions and the resulting images were analyzed using an Affymetrix GeneChip Scanner 3000. Microarray images were processed and raw data were extracted with Affymetrix GeneChip Command

Console (AGCC) Software.

Statistical analysis was conducted using the Excel. P-values were calculated by Student's t-test. The P values were corrected with a multiple test to assess the false discovery rate by means of the publicly available software QVALUE (http://genomine.org/qvalue/) (Storey et al., 2003). Differences between treatments were deemed significant when genes exhibited either a 2 fold change or greater between treatments (Q values ≤ 0.05).

Quantitative Reverse Transcription-PCR

To confirm the expression levels detected in the microarrays, using the same batch of RNA, we performed RT-qPCR using *recA* as a reference gene. Primers were designed using Primer 3 (http://bioinfo.ut.ee/primer3-0.4.0/) and the information is shown in Table 2. Tm of all primers ranged from 57°C to 61°C, and amplicon lengths ranged from 100 to 250 bp. Two genes (*msox* and *recA* genes) successfully amplified product with target size (checked on 1% agarose gels). Efficiencies of the two primer pairs were tested using serial dilutions of cDNA. We also did melting curve analysis on the primers, all of those primers amplified without nonspecific products, and for every gene, melting curves of different samples clustered together, with very little variation. cDNA was synthesized using QuantiTect Reverse Transcription kit (Qiagen).

The amplification was performed in 20 µl final volumes containing 2 µl cDNA

template, 0.2 mM of each respective primer, and 10 µl of SybrGreen Master Mix (Applied Biosystems). All the amplifications were carried out in optical-grade 96-well plates on an ABI Prism® 7500 FAST sequence detection system (Applied Biosystems) with an initial step at 95°C for 5 min followed by 40 cycles of 95°C for 10 s and 60°C for 30 s. The Ct value was determined automatically by the instrument. All samples were analyzed in triplicate. As negative controls, reaction mix without cDNA and reaction mix with RNA reverse transcribed without the addition of transcriptase were used. The Amplification efficiency (E) were calculated using the formula $E = (10^{-1/slope}) - 1$. Confidence intervals were calculated by Student's t test with a significance level of 5%.

Overexpression of msox in Escherichia coli

The *msox* gene from *Ca.* P. ubique HTCC1062 was commercially synthesized with *E. coli* codon usage (GenScript Corporation). The synthesized gene was inserted into the expression vector pET28b using the BamHI/XhoI sites. The resulting plasmids were then transformed into the expression host *E. coli* BL21 (DE3) (Merck Biosciences). To overexpress *msox*, 2 L of *E. coli* cells were grown at 37 °C to an OD_{600} of 0.6. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.5 mM. Cultures were incubated in 16°C and shake overnight.

Cells were harvested by centrifugation at 10,000 × *g* for 10 min. Pellets were resuspended in 20 mL lysis buffer and placed on ice. Cells were then lysed by sonication

(sonication occurs 4 times for 30 s on/ 30 s off on ice at number “4” on the sonicator). Lysed cells were centrifuged at 15,000 × *g* for 1 h to remove cell debris from the supernatant. Overexpressed MSOX was purified using a cOmplete His-Tag purification resin (Roche) as described in the manufacturer’s protocol.

The protein concentration was quantified using a Bio-Rad protein assay kit according to the instructions (Bio-Rad). One-dimensional protein analyses were carried out using a precast NuPAGE Bis-Tris gel (10%; Invitrogen) followed by Coomassie blue staining. The proteins were also blotted onto polyvinylidene fluoride (PVDF) membranes using the Novex XCell System and reagents following the manufacturers protocol (Invitrogen or Life Technologies). Membrane was probed with anti-His pAb primary antibody and anti-rabbit secondary antibody (Santa Cruz) following standard protocols. Membranes were developed by incubating in chemiluminescent reagent (Thermo Scientific) and visualized on an Alpha Innotech imaging system

Enzyme assays

DMA monooxygenase activities were measured by following the decrease in absorbance at 340 nm of substrate-dependent oxidation of NADPH (Sigma-Aldrich). The following concentrations of substrates were used: TMAO or DMA (10 mM) and NADPH (0.25 mM). The assays were performed at room temperature on a UV-visible spectrophotometer. A 100 uL mixture contained 3 µg purified enzyme, 100 mM PIPES

buffer (pH 7.6), and 0.25 mM NADPH. The reaction was initiated by adding substrate, and the decrease in absorbance at 340 nm was recorded continuously for 30 min. All substrates were dissolved in water.

Detection of enzymatic activity products

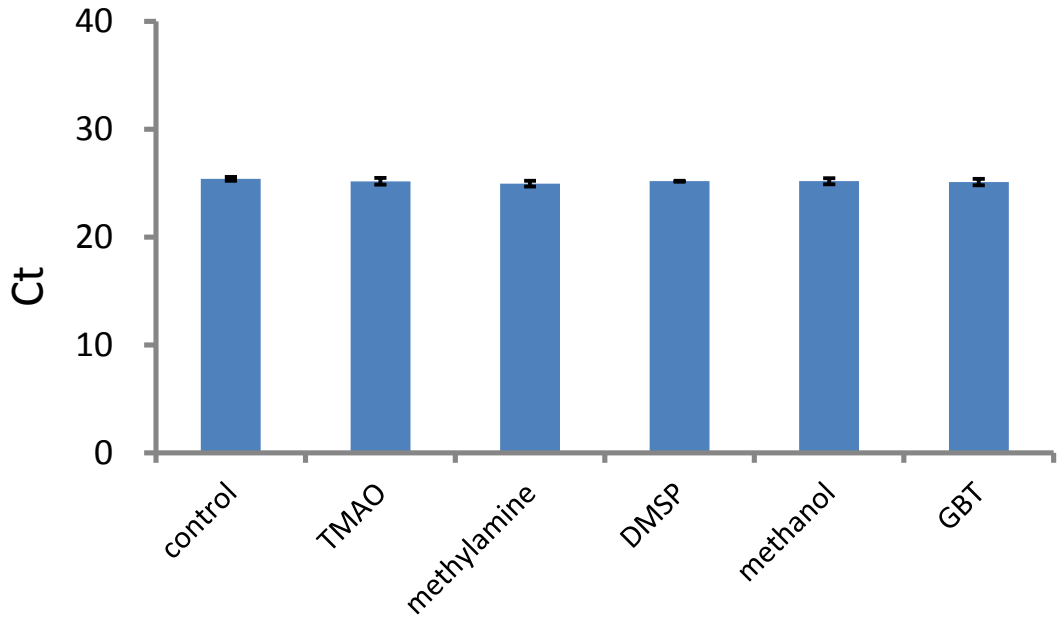
The enzymatic reaction mixture (100 µL) included 10 mM TMAO, 0.25 mM NADPH, 100 mM PIPES buffer and 3 µg protein, and incubate at room temperature for 1 hour. After the reaction were done, an Amicon Ultra 0.5 mL centrifugal filter (3KDa) was used to remove protein, and the rest solution was made 1:10 dilution with water and analyzed on a electrospray ionization/ time of flight mass spectrometer (ESI-TOF-MS; AB SCIEX TripleTOF® 5600 System). The mixture was chromatographed on a hydrophilic interaction chromatography (HILIC) column (100 × 2.1 mm, 5 µm, 200 Å; Merck, Darmstadt, Germany) with a constant flow rate of 0.2 mL/min. Volume of injection was 1 µL. The mobile phase A consisted of 0.1% formic acid in water, while mobile phase B was acetonitrile. A 10-min gradient from 90% to 50% B was used.

Table 4.1. The gene upregulated after adding TMAO. Differences were deemed significant when genes exhibited either a 2 fold change or greater between treatments and controls, and fold change value indicated that differential expression is supported by a Q-value of 0.05 or less.

ID	Annotation	Fold Change	P-value	Q-value
SAR11_1304	monomeric sarcosine oxidase	6.63	< 0.001	0.0027

Table 4.2 Oligonucleotides used in RT-qPCR

Primer Name	Sequence (5’ – 3’)
msox-forward	AAGTTGTTGTTGGTGCTGGTC
msox-reverse	CAGTATTTCCACATTTCAGTTTCA
recA-forward	AACCTGATACTGGCGAACAA
recA-reverse	GCAACAGAGTCAATTACGATAACA



Figue 4.1 Expression levels of reference gene *recA* under different cultural conditions. Gene expression levels (represented by absolute Ct values) of *recA*, each bar represents the mean of three samples from independent RNA isolations from three cultures. Error bars indicate standard deviations.

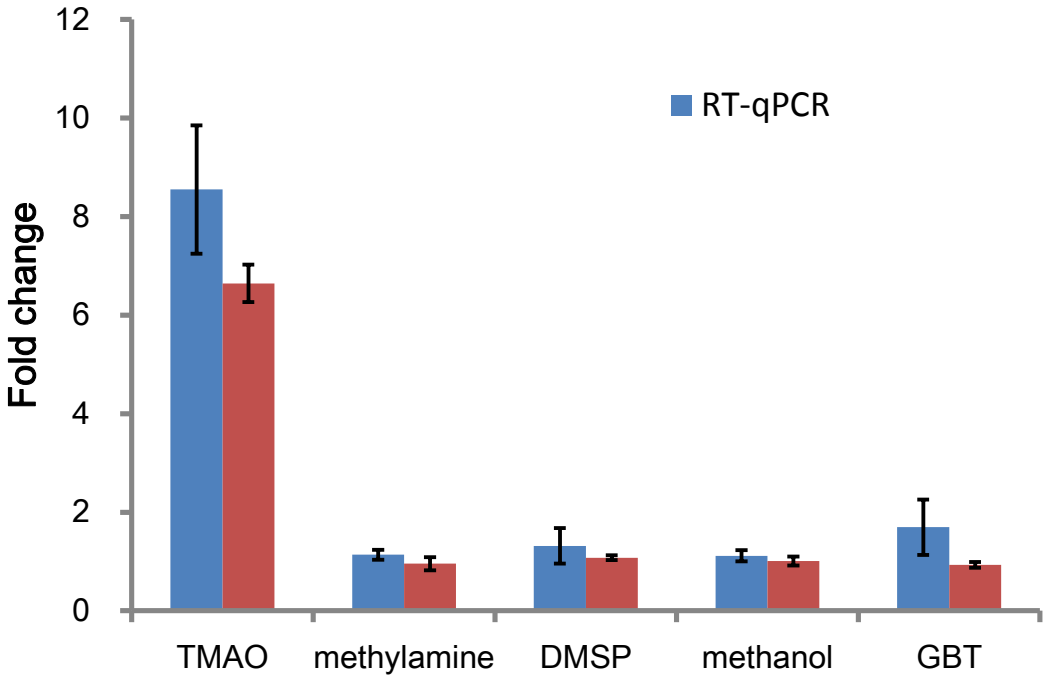


Figure 2. Comparison of differential expression of gene (SAR11_1304) with microarray (Red) and RT-qPCR (blue) in addition of different compounds. RT-qPCR reactions were performed in triplicate and both qPCR and microarray data used the same RNA as the source.

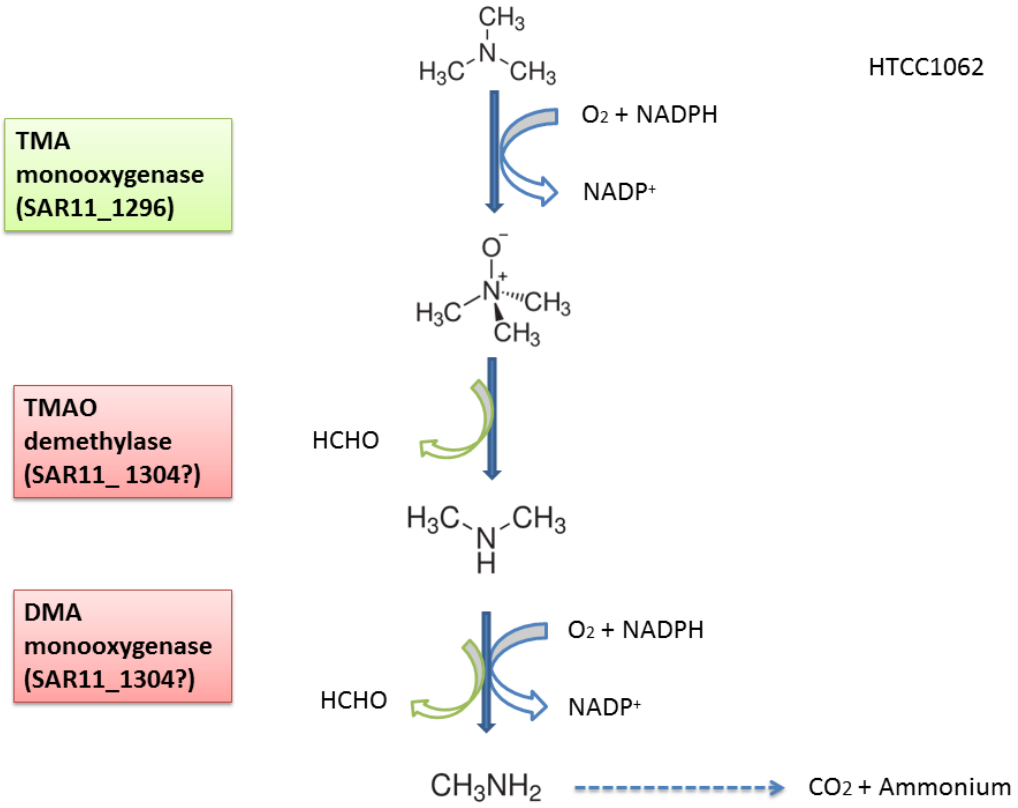


Figure 4.3 Proposed metabolic pathway for MSOX (SAR11_1304) in HTCC1062.

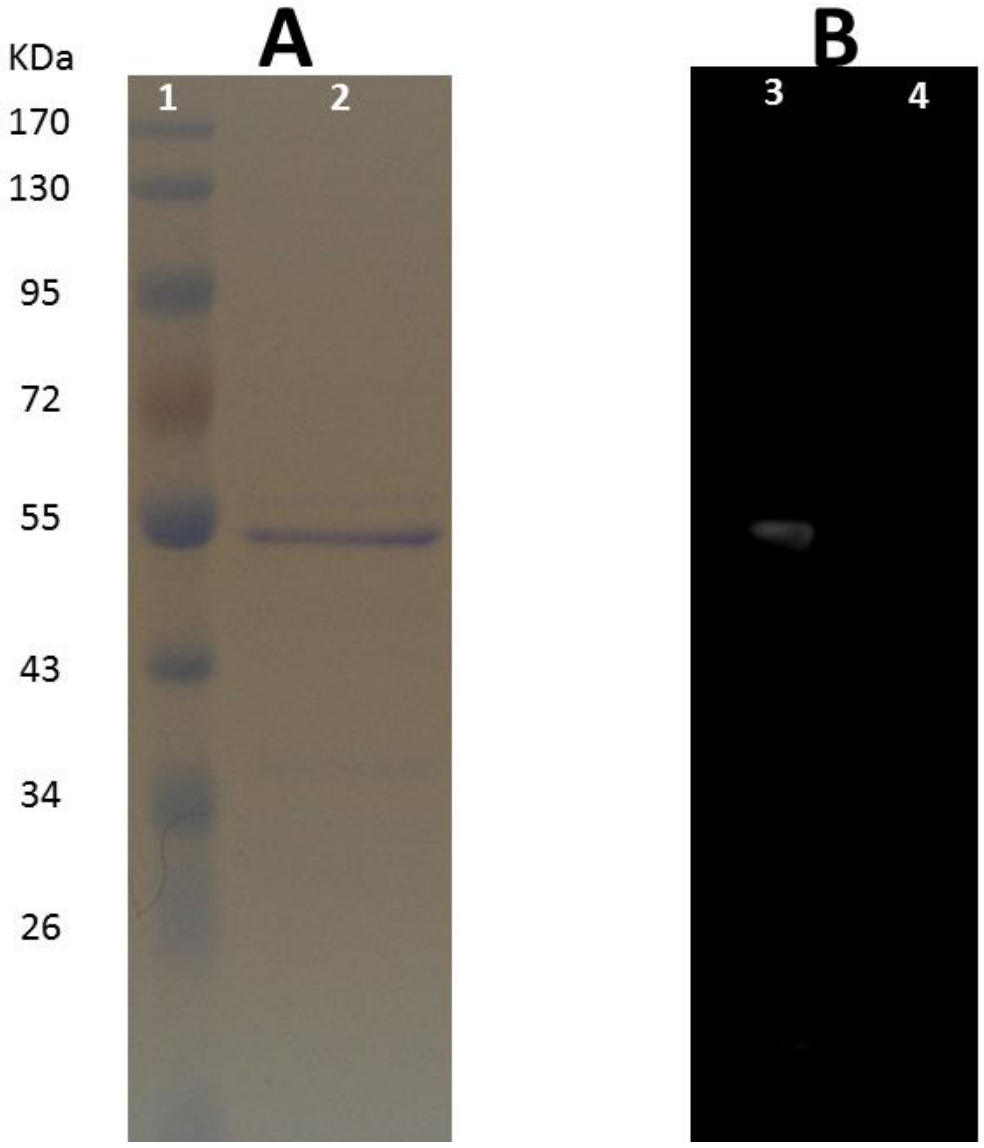


Figure 4.4 Expression and purification of MSOX from *E. coli.* A. SDS-PAGE of purified MSOX protein. Lane 1: broad range protein marker; Lane 2: purified MSOX from recombinant *E. coli*; B. Western blot analysis of cell extracts using His-Tag Antibody. Lane 3: His-tagged protein (MSOX); Lane 4: control from non-recombinant *E. coli*.

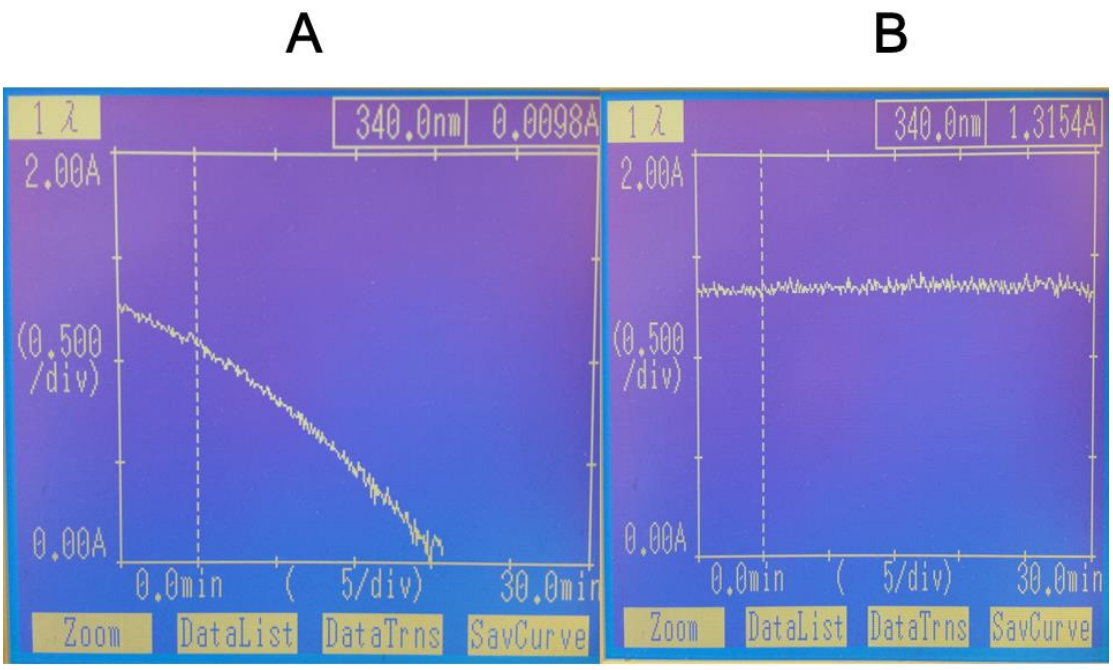


Figure 4.5 Enzymatic assay for the monooxgenase activities of MSOX. This activity is measured by following the decrease in absorbance at 340 nm of substrate-dependent oxidation of NADPH after addition of DMA (A) or TMAO (B). The concentration of the each substrate is 10 mM. The reaction is performed in room temperature.

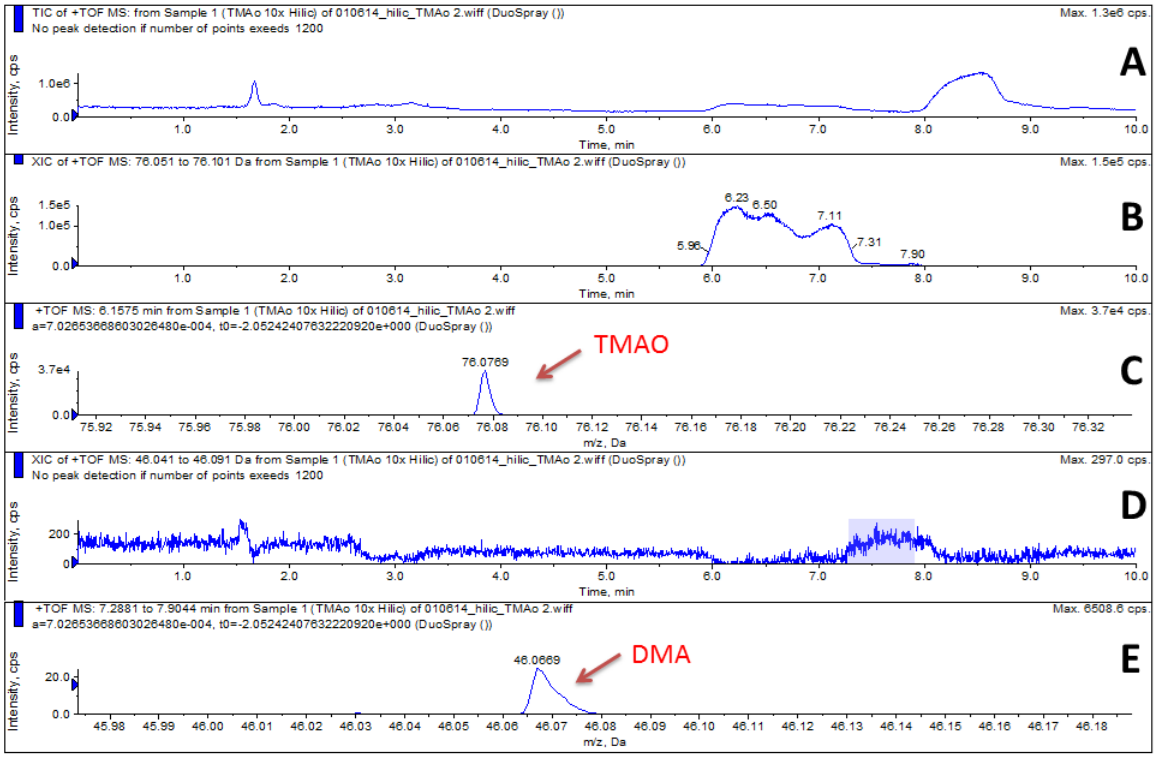


Figure 4.6. Detection of enzymatic activity products of MSOX. The enzymatic reaction mixture included 10 mM TMAO, 0.25 mM NADPH, 100 mM PIPES buffer and 3 µg protein, and incubate at room temperature for 1 hour. After incubation, the protein was removed by centrifugal filer, and the mixture was made 1:10 dilution with water and analyzed by HILIC-ESI/TOF/MS. A and B are the chromatograms of TMAO substrate, C is the mass spectrum of TMAO. D is the chromatogram of enzymatic reaction product DMA and E is the mass spectrum of DMA.

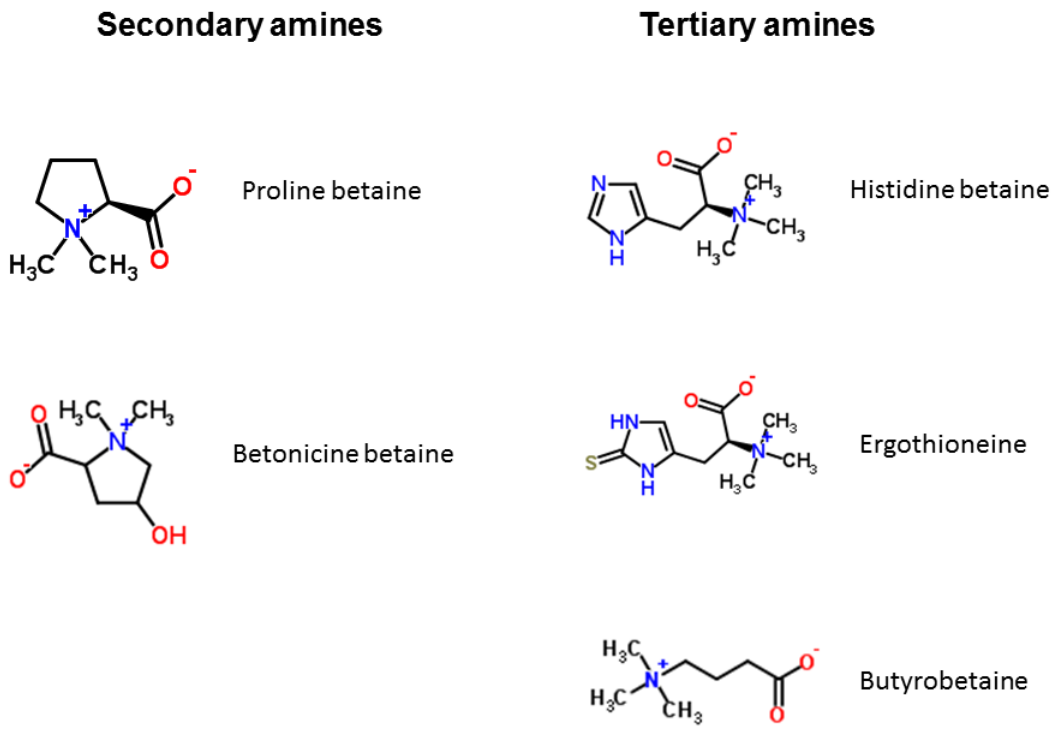


Figure 4.7 Potential substrates for MSOX (encoded by SAR11_1304).

Chapter 5

5. General Conclusions

The study of genomes from new organisms can provide insight into unknown metabolism and geochemical processes. This dissertation displays several powerful examples of how experimental approaches can be used to test hypotheses generated from genomic and metagenomic data, revealing novel microbial metabolic pathways that are important in the environment.

In that spirit, when we first sequenced the SAR11 genome, a suite of genes for demethylation and C1 oxidation was discovered. This observation was surprising because C1 metabolism is rarely reported in non-methylotrophic bacteria, and at the time this study was initiated there were few indications that these compounds were important components of DOM. In chapter 2, we used pure cultures to show for the first time that SAR11 strain HTCC1062 can use a variety of C1 and methylated compounds. Unlike well-known methylotrophic bacteria, HTCC1062 can use these compounds for energy, but not biomass production. After chapter 2 was published, Grote (Grote et al., 2012) further explored and discovered that genes for C1 and methyl group oxidation are found

throughout the SAR11 clade. A new term, methylovory, was coined to describe this phenomenon. Given their highly abundant populations and wide distributions in the oceans, these findings provide the new insight that methylovory is an important mechanism of DOM oxidation.

In previous work, Tripp (Tripp et al., 2008) reported that SAR11 could use the sulfur containing osmolyte DMSP as a source of reduced sulfur for growth, and a pathway for DMSP metabolism by demethylation was described, but no known DMSP lyases were observed in the SAR11 HTCC1062 genome. In chapter 3, our physiological experiments showed that HTCC1062 uses an unidentified DMSP lyase to cleave DMSP to produce the climatically active gas DMS, and that as much as 60% of DMSP uptake is channeled through this cleavage pathway. Simultaneously, via a previously identified demethylation pathway, HTCC1062 produces methanethiol, which we show is used as a source of reduced sulfur for biomass production.

In chapter 4, transcriptomics screening by Affymetrix microarray technology was performed to expand insights into novel enzymes and metabolic pathways for catabolism of five different C1 compounds and methylated compounds that are metabolized by SAR11. One inducible gene (SAR11_1304), annotated as a monomeric sarcosine oxidase (MSOX), was found to upregulated in response to addition of TMAO. Further investigation by cloning, over-expression and characterizion the enzymatic activities indicated that this enzyme catalyzes demethylation and oxidization of TMAO to

methylamine.

Overall, this investigation shows that SAR11 cells can use a diverse range of C1 and methylated compounds. The impact of this discovery is not yet known and will not be fully understood until more information becomes available about the concentrations and turnover rates of C1 and methylated compounds in the oceans. The long range, broader impacts of the novel findings presented in this dissertation are likely to come from future investigations that explore the ramifications of finding these unusual metabolic pathways in the very small genomes of the world most abundant bacterial species. Since these cells have very compact genomes and are very successful, genomic indications of unexpected metabolism provide important clues that can be used to refine our knowledge of carbon cycling by marine bacteria.

Bibliography

Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. Bioinformatics *21*, 2104-2105.

Andreae, M.O. (1990). Ocean-atmosphere interactions in the global biogeochemical sulfur cycle. Marine Chemistry *30*, 1-29.

Anthony, C. (1982). The Biochemistry of Methylotrophs, London: Academic Press.

Antoine, E., Rolland, J.L., Raffin, J.P., and Dietrich, J. (1999). Cloning and over-expression in *Escherichia coli* of the gene encoding NADPH group III alcohol dehydrogenase from *Thermococcus hydrothermalis*. Characterization and comparison of the native and the recombinant enzymes. European Journal of Biochemistry *264*, 880-889.

Appling, D.R. (1991). Compartmentation of folate-mediated one-carbon metabolism in eukaryotes. FASEB Journal *5*, 2645-2651.

Archer, S.D., Widdicombe, C.E., Tarran, G.A., Rees, A.P., and Burkill, P.H. (2001). Production and turnover of particulate dimethylsulphoniopropionate during a coccolithophore bloom in the northern North Sea. Aquatic Microbial Ecology *24*, 225-241.

Arfman, N., Hektor, H.J., Bystrykh, L.V., Govorukhina, N.I., Dijkhuizen, L., and Frank, J. (1997). Properties of an NAD(H)-containing methanol dehydrogenase and its activator protein from *Bacillus methanolicus*. European Journal of Biochemistry *244*, 426-433.

Atteia, A., Adrait, A., Brugiere, S., Tardif, M., van Lis, R., Deusch, O., Dagan, T., Kuhn, L., Gontero, B., Martin, W*., et al.* (2009). A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the alpha-proteobacterial mitochondrial ancestor. Molecular Biology and Evolution *26*, 1533-1548.

Azam, F., Fenchel, T., Field, J.G., Gray, J.S., Meyer-Reil, L.A., and Thingstad, F. (1983). The ecological role of water-column microbes in the sea. Marine Ecology-progress Series *10*, 257-263.

Barber, R.D., and Donohue, T.J. (1998). Function of a glutathione-dependent formaldehyde dehydrogenase in *Rhodobacter sphaeroides* formaldehyde oxidation and assimilation. Biochemistry *37*, 530-537.

Barra, L., Fontenelle, C., Ermel, G., Trautwetter, A., Walker, G.C., and Blanco, C. (2006). Interrelations between glycine betaine catabolism and methionine biosynthesis in *Sinorhizobium meliloti* strain 102F34. Journal of Bacteriology *188*, 7195-7204.

Barrett, E.L., and Kwan, H.S. (1985). Bacterial reduction of trimethylamine oxide. Annual Review of Microbiology *39*, 131-149.

Ben-Bassat, A., Goldberg, I., and Mateles, R.I. (1980). Distribution of methanol carbon between assimilation and oxidation pathways in methanol-grown *Pseudomonas* C. Journal of General Microbiology *116*, 213-223.

Bicknell, B., and Owens, J.D. (1980). Utilization of methylamines as nitrogen sources by non-methylotrophs. Journal of General Microbiology *117*, 89-96.

Biers, E.J., Sun, S., and Howard, E.C. (2009). Prokaryotic genomes and diversity in surface ocean waters: interrogating the global ocean sampling metagenome. Applied and Environmental Microbiology *75*, 2221-2229.

Boden, R., Borodina, E., Wood, A.P., Kelly, D.P., Murrell, J.C., and Schafer, H. (2011). Purification and characterization of dimethylsulfide monooxygenase from *Hyphomicrobium sulfonivorans*. Journal of Bacteriology *193*, 1250-1258.

Bravo-Linares, C.M., and Mudge, S.M. (2009). Temporal trends and identification of the sources of volatile organic compounds in coastal seawater. Journal of Environmental Monitoring *11*, 628-641.

Bystrykh, L.V., Arfman, N., and Dijkhuizen, L. (1993). The methanol-oxidizing enzyme systems in Gram-positive methylotrophic bacteria. In microbial growth on C1 compounds (Murrell, J C and Kelly, D P, Eds) (Andover, UK: Intercept Ltd), pp. 245-251.

Carlson, C.A., Morris, R., Parsons, R., Treusch, A.H., Giovannoni, S.J., and Vergin, K. (2009). Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. The ISME Journal *3*, 283-295.

Carpenter, L.J., Lewis, A.C., Hopkins, J.R., Read, K.A., Longley, I.D., and Gallagher, M.W. (2004). Uptake of methanol to the North Atlantic Ocean surface. Global Biogeochemical Cycles *18*, GB4027.

Charlson, R.J., Lovelock, J.E., Andreae, M.O., and Warren, S.G. (1987). Ocean phytoplankton, atmospheric sulfur, cloud albedo and climate. Nature, *326*, 655-661.

Chen, Y., McAleer, K.L., and Murrell, J.C. (2010). Monomethylamine as a nitrogen source

for a nonmethylotrophic bacterium, *Agrobacterium tumefaciens*. Applied and Environmental Microbiology *76*, 4102-4104.

Chistoserdova, L., Chen, S.W., Lapidus, A., and Lidstrom, M.E. (2003). Methylotrophy in *Methylobacterium extorquens* AM1 from a genomic point of view. Journal of Bacteriology *185*, 2980-2987.

Chistoserdova, L., Crowther, G.J., Vorholt, J.A., Skovran, E., Portais, J.C., and Lidstrom, M.E. (2007). Identification of a fourth formate dehydrogenase in *Methylobacterium extorquens* AM1 and confirmation of the essential role of formate oxidation in methylotrophy. Journal of Bacteriology *189*, 9076-9081.

Chistoserdova, L., Gomelsky, L., Vorholt, J.A., Gomelsky, M., Tsygankov, Y.D., and Lidstrom, M.E. (2000). Analysis of two formaldehyde oxidation pathways in *Methylobacillus flagellatus* KT, a ribulose monophosphate cycle methylotroph. Microbiology *146*, 233-238.

Chistoserdova, L., Kalyuzhnaya, M.G., and Lidstrom, M.E. (2010). The expanding world of methylotrophic metabolism. Annual Review of Microbiology *63*, 477-499.

Chistoserdova, L., Laukel, M., Portais, J.C., Vorholt, J.A., and Lidstrom, M.E. (2004). Multiple formate dehydrogenase enzymes in the facultative methylotroph *Methylobacterium extorquens* AM1 are dispensable for growth on methanol. Journal of Bacteriology *186*, 22-28.

Chistoserdova, L., and Lidstrom, M. (2013). Aerobic Methylotrophic Prokaryotes. In The Prokaryotes, E. Rosenberg, E. DeLong, S. Lory, E. Stackebrandt, and F. Thompson, eds. (Springer Berlin Heidelberg), pp. 267-285.

Chistoserdova, L., Vorholt, J.A., and Lidstrom, M.E. (2005). A genomic view of methane oxidation by aerobic bacteria and anaerobic archaea. Genome Biology *6*, 208.

Chistoserdova, L., Vorholt, J.A., Thauer, R.K., and Lidstrom, M.E. (1998). C1 transfer enzymes and coenzymes linking methylotrophic bacteria and methanogenic Archaea. Science *281*, 99-102.

Chlumsky, L.J., Zhang, L., and Jorns, M.S. (1995). Sequence analysis of sarcosine oxidase and nearby genes reveals homologies with key enzymes of folate one-carbon metabolism. The Journal of Biological Chemistry *270*, 18252-18259.

Chróst, R., and Rai, H. (1994). Bacterial Secondary Production. In Microbial Ecology of Lake Plußsee, J. Overbeck, and R. Chróst, eds. (Springer New York), pp. 92-117.

Christian, T.J., Kleiss, B., Yokelson, R.J., Holzinger, R., Crutzen, P.J., Hao, W.M., Saharjo, B.H., and Ward, D.E. (2003). Comprehensive laboratory measurements of biomass-burning emissions: 1. Emissions from Indonesian, African, and other fuels. Journal of Geophysical Research: Atmospheres *108*, 4719.

Connon, S.A., and Giovannoni, S.J. (2002). High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. Applied and Environmental Microbiology *68*, 3878-3885.

Craft, T.R., Bermudez, E., and Skopek, T.R. (1987). Formaldehyde mutagenesis and formation of DNA-protein crosslinks in human lymphoblasts in vitro. Mutation Research *176*, 147-155.

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

Dafner, E.V., and Wangersky, P.J. (2002). A brief overview of modern directions in marine DOC studies Part II-Recent progress in marine DOC studies. Journal of Environmental Monitoring *4*, 55-69.

De Marco, P. (2004). Methylotrophy versus heterotrophy: a misconception. Microbiology *150*, 1606-1607.

de Vries, G.E., Arfman, N., Terpstra, P., and Dijkhuizen, L. (1992). Cloning, expression, and sequence analysis of the *Bacillus methanolicus* C1 methanol dehydrogenase gene. Journal of Bacteriology *174*, 5346-5353.

Dehal, P.S., Joachimiak, M.P., Price, M.N., Bates, J.T., Baumohl, J.K., Chivian, D., Friedland, G.D., Huang, K.H., Keller, K., Novichkov, P.S.*, et al.* (2009). MicrobesOnline: an integrated portal for comparative and functional genomics. Nucleic Acids Research *38*, D396-D400.

Diaz, M.R., Visscher, P.T., and Taylor, B.F. (1992). Metabolism of dimethylsulfoniopropionate and glycine betaine by a marine bacterium. FEMS Microbiology Letters *96*, 61-65.

Dixon, J.L., Beale, R., and Nightingale, P.D. (2010). Microbial methanol uptake in northeast Atlantic waters. The ISME Journal *5*, 704-716

Donald, T.M., John, L.S., Christopher, D.E., Heleen, A.d.W., Martin, F., Tore, H., Anders, W., Brit Lisa, S., Dean, S.J., Jussi, V.*, et al.* (2007). Dissolved organic carbon trends

resulting from changes in atmospheric deposition chemistry. Nature *450*, 537-540.

Duce, R.A., Arimoto, R., Ray, B.J., Unni, C.K., and Harder, P.J. (1983). Atmospheric trace elements at Enewetak Atoll: 1. Concentrations, sources, and temporal variability. Journal of Geophysical Research: Oceans *88*, 5321-5342.

Dunwell, J.M., Purvis, A., and Khuri, S. (2004). Cupins: the most functionally diverse protein superfamily? Phytochemistry *65*, 7-17.

Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E. (2006). The Prokaryotes: Symbiotic associations, biotechnology, applied microbiology 3rd edition edn (Springer).

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

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

Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLoS Computational Biology *7*, e1002195.

Edgar, R.C. (2004a). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

Edgar, R.C. (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research *32*, 1792-1797.

Engel, A., Borchard, C., Piontek, J., Schulz, K.G., Riebesell, U., and Bellerby, R. (2013). CO_2 increases ^{14}C primary production in an Arctic plankton community. Biogeosciences *10*, 1291-1308.

Fang, Y., and Qian, M.C. (2005). Sensitive quantification of sulfur compounds in wine by headspace solid-phase microextraction technique. Journal of Chromatography *1080*, 177-185.

Feng, S., Powell, S.M., Wilson, R., and Bowman, J.P. (2013). Light-stimulated growth of proteorhodopsin-bearing sea-ice psychrophile *Psychroflexus torquis* is salinity dependent. The ISME Journal *7*, 2206-2213.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K.*, et al.* (2010). The Pfam protein families database. Nucleic Acids Research *38*, D211-D222.

Fuhrman, J. (1992). Bacterioplankton roles in cycling of organic matter: the microbial food web. In: Primary Productivity and Biogeochemical Cycles in the Sea (Falkowski, PG, Woodhead, AD, Eds) (New York: Plenum Press), pp. 361-383.

Gabaldon, T., and Huynen, M.A. (2007). From endosymbiont to host-controlled organelle: The hijacking of mitochondrial protein synthesis and metabolism. PLoS Computational Biology *3*, e219.

Galbally, I.E., and Kirstine, W. (2002). The production of methanol by flowering plants and the global cycle of methanol. Journal of Atmospheric Chemistry *43*, 195-229.

Gao, Q., Leck, C., Rauschenberg, C., and Matrai, P.A. (2012). On the chemical dynamics of extracellular polysaccharides in the high Arctic surface microlayer. Ocean Science Discussions *9*, 215-259.

Ghosh, W., and Dam, B. (2009). Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and ecologically diverse bacteria and archaea. FEMS Microbiology Reviews *33*, 999-1043.

Giovannoni, S.J., Britschgi, T.B., Moyer, C.L., and Field, K.G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. Nature *345*, 60-63.

Giovannoni, S.J., Hayakawa, D.H., Tripp, H.J., Stingl, U., Givan, S.A., Cho, J.C., Oh, H.M., Kitner, J.B., Vergin, K.L., and Rappe, M.S. (2008). The small genome of an abundant coastal ocean methylotroph. Environmental Microbiology *10*, 1771-1782.

Giovannoni, S.J., and Stingl, U. (2005). Molecular diversity and ecology of microbial plankton. Nature *437*, 343-348.

Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D., Bibbs, L., Eads, J., Richardson, T.H., Noordewier, M.*, et al.* (2005). Genome streamlining in a cosmopolitan oceanic bacterium. Science *309*, 1242-1245.

Goenrich, M., Bartoschek, S., Hagemeier, C.H., Griesinger, C., and Vorholt, J.A. (2002). A glutathione-dependent formaldehyde-activating enzyme (Gfa) from Paracoccus denitrificans detected and purified via two-dimensional proton exchange NMR spectroscopy. The Journal of biological chemistry *277*, 3069-3072.

Gonzalez, J.M., Kiene, R.P., and Moran, M.A. (1999). Transformation of sulfur compounds by an abundant lineage of marine bacteria in the α-Subclass of the Class Proteobacteria. Applied and Environmental Microbiology *65*, 3810-3819.

Goode, J.G., Yokelson, R.J., Ward, D.E., Susott, R.A., Babbitt, R.E., Davies, M.A., and Hao, W.M. (2000). Measurements of excess O_3, CO_2, CO, CH_4, C_2H_4, C_2H_2, HCN, NO, NH_3, HCOOH, CH_3COOH, HCHO, and CH_3OH in 1997 Alaskan biomass burning plumes by airborne Fourier transform infrared spectroscopy (AFTIR). Journal of Geophysical Research *105*, 22147-22166.

Grafstrom, R.C., Fornace, A.J., Jr., Autrup, H., Lechner, J.F., and Harris, C.C. (1983). Formaldehyde damage to DNA and inhibition of DNA repair in human bronchial cells. Science *220*, 216-218.

Grote, J., Thrash, J.C., Hugget, M.J., Landry, Z.C., Carini, P., Giovannoni, S.J., and Rappé, M.S. (2012). Streamlining and core genome conservation among highly divergent members of the SAR11 clade. MBio *3*, doi: 10.1128/mBio.00252-12

Guse, A., Stevenson, C.E., Kuper, J., Buchanan, G., Schwarz, G., Giordano, G., Magalon, A., Mendel, R.R., Lawson, D.M., and Palmer, T. (2003). Biochemical and structural analysis of the molybdenum cofactor biosynthesis protein MobA. The Journal of Biological Chemistry *278*, 25302-25307.

Gutheil, W.G., Kasimoglu, E., and Nicholson, P.C. (1997). Induction of glutathione-dependent formaldehyde dehydrogenase activity in *Escherichia coli* and Hemophilus influenza. Biochemical and biophysical research communications *238*, 693-696.

Harms, N., Ras, J., Reijnders, W.N., van Spanning, R.J., and Stouthamer, A.H. (1996). S-formylglutathione hydrolase of *Paracoccus denitrificans* is homologous to human esterase D: a universal pathway for formaldehyde detoxification? Journal of Bacteriology *178*, 6296-6299.

Hasona, A., Ray, R.M., and Shanmugam, K.T. (1998). Physiological and genetic analyses leading to identification of a biochemical role for the *moeA* (molybdate metabolism) gene product in *Escherichia coli*. Journal of Bacteriology *180*, 1466-1472.

Hedges, J.I. (1992). Global biogeochemical cycles: progess and problems. Marine Chemistry *39*, 67-93.

Heikes, B.G., Chang, W., Pilson, M.E.Q., Swift, E., Singh, H.B., Guenther, A., Jacob, D.J., Field, B.D., Fall, R., Riemer, D.*, et al.* (2002). Atmospheric methanol budget and ocean implication. Global Biogeochemical Cycles *16*, 1133.

Henderson, J.F. (1979). Teaching one-carbon metabolism. Biochemical Education *7*, 51-52.

Howard, E.C., Henriksen, J.R., Buchan, A., Reisch, C.R., Burgmann, H., Welsh, R., Ye, W., Gonzalez, J.M., Mace, K., Joye, S.B.*, et al.* (2006). Bacterial taxa that limit sulfur flux from the ocean. Science *314*, 649-652.

Jacob, D.J., Field, B.D., Li, Q., Blake, D.R., de Gouw, J., Warneke, C., Hansel, A., Wisthaler, A., Singh, H.B., and Guenther, A. (2005). Global budget of methanol: Constraints from atmospheric observations. Journal of Geophysical Research: Atmospheres *110*, D08303.

Johnson, P.A., and Quayle, J.R. (1965). Microbial growth on C1 Compounds. Synthesis of cell constituents by methane- and methanol-grown *Pseudomonas Methanica*. The Biochemical Journal *95*, 859-867.

Jones, D.P., Thor, H., Andersson, B., and Orrenius, S. (1978). Detoxification reactions in isolated hepatocytes. Role of glutathione peroxidase, catalase, and formaldehyde dehydrogenase in reactions relating to N-demethylation by the cytochrome P-450 system. The Journal of Biological Chemistry *253*, 6031-6037.

Jumars, P.A., Penry, D.L., Baross, J.A., Perry, M.J., and Frost, B.W. (1989). Closing the microbial loop: dissolved carbon pathway to heterotrophic bacteria from incomplete ingestion, digestion and absorption in animals. Deep Sea Research Part A Oceanographic Research Papers *36*, 483-495.

Kameyama, S., Tanimoto, H., Inomata, S., Tsunogai, U., Ooki, A., Yokouchi, Y., Takeda, S., Obata, H., and Uematsu, M. (2009). Equilibrator inlet-proton transfer reaction-mass spectrometry (EI-PTR-MS) for sensitive, high-resolution measurement of dimethyl sulfide dissolved in seawater. Analytical Chemistry *81*, 9021-9026.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Research *38*, D355-D360.

Karl, D.M., Beversdorf, L., Bjorkman, K.M., Church, M.J., Martinez, A., and Delong, E.F. (2008). Aerobic production of methane in the sea. Nature Geoscience *1*, 473-478.

Keiber, R.J., Zhou, X., and Mopper, K. (1990). Formation of carbonyl compounds from UV-induced photodegradation of humic substances in natural waters: Fate of riverine carbon in the sea. Limnology and Oceanography *35*, 1503-1515.

Keller Maureen, D., Bellows Wendy, K., and Guillard Robert, R.L. (1989). Dimethyl sulfide production in marine phytoplankton. In Biogenic Sulfur in the Environment (American Chemical Society), pp. 167-182.

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

Kemp, M.B., and Quayle, J.R. (1967). Microbial growth on C1 compounds. Uptake of [14C] formaldehyde and [14C] formate by methane-grown *Pseudomonas methanica* and determination of the hexose labelling pattern after brief incubation with [14C] methanol. The Biochemical Journal *102*, 94-102.

Kiene, R.P. (1998). Uptake of choline and its conversion to glycine betaine by bacteria in Estuarine Waters. Applied and Environmental Microbiology *64*, 1045-1051.

Kiene, R.P., and Linn, L.J. (2000). The fate of dissolved dimethylsulfoniopropionate (DMSP) in seawater: Tracer studies using ^{35}S-DMSP. Geochimica et Cosmochimica Acta *64*, 2797-2810.

Kiene, R.P., Linn, L.J., and Bruton, J.A. (2000). New and important roles for DMSP in marine microbial communities. Journal of Sea Research *43*, 209-224.

Kiene, R.P., Linn, L.J., González, J., Moran, M.A., and Bruton, J.A. (1999). Dimethylsulfoniopropionate and methanethiol are important precursors of methionine and protein-sulfur in marine bacterioplankton. Applied and Environmental Microbiology *65*, 4549-4558.

King, G.M., and Weber, C.F. (2007). Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. Nature Review Microbiology *5*, 107-118.

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

Kirst, G.O. (1990). Salinity tolerance of eukaryotic marine algae. Annual Review of Plant Physiology and Plant Molecular Biology *40*, 21.

Kolka, R., Weishampel, P., and Fröberg, M. (2008). Measurement and Importance of Dissolved Organic Carbon. In Field Measurements for Forest Carbon Monitoring, C. Hoover, ed. (Springer Netherlands), pp. 171-176.

Kos, G., and Ariya, P. (2006). Determination of a wide range of volatile and semivolatile organic compounds in snow by use of solid-phase micro-extraction (SPME). Analytical and Bioanalytical Chemistry *385*, 57-66.

Larsson, U., and Hagström, A. (1979). Phytoplankton exudate release as an energy source for the growth of pelagic bacteria. Marine Biology *52*, 199-206.

Latypova, E., Yang, S., Wang, Y.S., Wang, T., Chavkin, T.A., Hackett, M., Schafer, H., and Kalyuzhnaya, M.G. (2010). Genetics of the glutamate-mediated methylamine utilization pathway in the facultative methylotrophic beta-proteobacterium *Methyloversatilis universalis* FAM5. Molecular Microbiology *75*, 426-439.

Lidbury, I., Murrell, Colin J. and Chen, Yin (2014). Trimethylamine N-oxide metabolism by abundant marine heterotrophic bacteria. Proceedings of the National Academy of Sciences of the United States of America *111*, 2710-2715. doi: 10.1073/pnas.1317834111.

Lidstrom, M. (2006). Aerobic methylotrophic prokaryotes. In, pp. 618-634.

Lin, C.N., Syu, W.J., Sun, W.S., Chen, J.W., Chen, T.H., Don, M.J., and Wang, S.H. (2010). A role of *ygfZ* in the *Escherichia coli* response to plumbagin challenge. Journal of Biomedical Science *17*, 84.

Liu, X., Dong, Y., Zhang, J., Zhang, A., Wang, L., and Feng, L. (2009). Two novel metal-independent long-chain alkyl alcohol dehydrogenases from *Geobacillus thermodenitrificans* NG80-2. Microbiology *155*, 2078-2085.

Maden, B.E. (2000). Tetrahydrofolate and tetrahydromethanopterin compared: functionally distinct carriers in C1 metabolism. The Biochemical journal *350 Pt 3*, 609-629.

Malmstrom, R.R., Kiene, R.P., Cottrell, M.T., and Kirchman, D.L. (2004). Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the North Atlantic ocean. Applied and Environmental Microbiology *70*, 4129-4135.

Manian, S.S., Gumbleton, R., Buckley, A.M., and O'Gara, F. (1984). Nitrogen fixation and carbon dioxide assimilation in *Rhizobium japonicum*. Applied and Environmental Microbiology *48*, 276-279.

Marison, I.W., and Attwood, M.M. (1982). A possible alternative mechanism for the oxidation of formaldehyde to formate. Journal of General Microbiology *128*, 1441-1446.

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

McDowell, L.R. (2000). Vitamins in Animal and Human Nutrition (Iowa State University Press).

Meskys, R., Harris, R.J., Casaite, V., Basran, J., and Scrutton, N.S. (2001). Organization of the genes involved in dimethylglycine and sarcosine degradation in *Arthrobacter spp*.: implications for glycine betaine catabolism. European Journal of Biochemistry *268*, 3390-3398.

Misset-Smits, M., van Ophem, P.W., Sakuda, S., and Duine, J.A. (1997). Mycothiol, 1-O-(2′ -[N-acetyl-L-cysteinyl]amido-2′ -deoxy-α -D-glucopyranosyl)-D-myo-inositol, is the factor of NAD/factor-dependent formaldehyde dehydrogenase. FEBS Letters *409*, 221-222.

Mopper, K., Zhou, X., Kieber, R.J., Kieber, D.J., Sikorski, R.J., and Jones, R.D. (1991). Photochemical degradation of dissolved organic carbon and its impact on the oceanic carbon cycle. Nature *353*, 60-62.

Moran, M.A., Reisch, C.R., Kiene, R.P., and Whitman, W.B. (2012). Genomic insights into bacterial DMSP transformations. Annual review of marine science *4*, 523-542.

Morris, R.M., Rappe, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., and Giovannoni, S.J. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. Nature *420*, 806-810.

Mullins, T.D., Britschgi, T.B., Krest, R.L., and Giovannoni, S.J. (1995). Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. Limnology & Oceanography *40*, 148-158.

Myklestad, S. (2000). Dissolved Organic Carbon from Phytoplankton. In Marine Chemistry, P.J. Wangersky, ed. (Springer Berlin Heidelberg), pp. 111-148.

Nemecek-Marshall, M., MacDonald, R.C., Franzen, J.J., Wojciechowski, C.L., and Fall, R. (1995). Methanol emission from leaves (enzymatic detection of gas-phase methanol and relation of methanol fluxes to stomatal conductance and leaf development). Plant Physiology *108*, 1359-1368.

Neufeld, J.D., Schafer, H., Cox, M.J., Boden, R., McDonald, I.R., and Murrell, J.C. (2007). Stable-isotope probing implicates *Methylophaga spp*. and novel Gammaproteobacteria in marine methanol and methylamine metabolism. The ISME Journal *1*, 480-491.

Norin, A., Van Ophem, P.W., Piersma, S.R., Persson, B., Duine, J.A., and Jörnvall, H. (1997). Mycothiol-dependent formaldehyde dehydrogenase, a prokaryotic medium-chain

dehydrogenase/reductase, phylogenetically links different eukaroytic alcohol dehydrogenases primary structure, Conformational Modelling and Functional Correlations. European Journal of Biochemistry *248*, 282-289.

Nuccio, J.S., P. J. ; Kieber, R. J. (1995). Biological production of formaldehyde in the marine environment. Limnology and Oceanography *40*, 521-527.

Oh, H.-M., Kwon, K.K., Kang, I., Kang, S.G., Lee, J.-H., Kim, S.-J., and Cho, J.-C. (2010). Complete genome sequence of "*Candidatus* Puniceispirillum marinum" IMCC1322, a representative of the SAR116 clade in the *Alphaproteobacteria*. Journal of Bacteriology *192*, 3240-3241.

Pace, M.L., and Cole, J.J. (2002). Synchronous variation of dissolved organic carbon and color in lakes. Limnology and Oceanography *47*.

Pike, S.T., Rajendra, R., Artzt, K., and Appling, D.R. (2009). Mitochondrial C1-tetrahydrofolate synthase (MTHFD1L) supports the flow of mitochondrial one-carbon units into the methyl cycle in embryos. The Journal of Biological Chemistry *285*, 4612-4620.

Pomeroy, L.R. (1974). The ocean's food web, a changing paradigm. Bioscience *24*.

Rappé, M., Connon, S., Vergin, K., and Giovannoni, S. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature *418*, 630-633.

Rappe, M.S., and Giovannoni, S.J. (2003). The uncultured microbial majority. Annual Review of Microbiology *57*, 369-394.

Ras, J., Van Ophem, P.W., Reijnders, W.N., Van Spanning, R.J., Duine, J.A., Stouthamer, A.H., and Harms, N. (1995). Isolation, sequencing, and mutagenesis of the gene encoding NAD- and glutathione-dependent formaldehyde dehydrogenase (GD-FALDH) from *Paracoccus denitrificans*, in which GD-FALDH is essential for methylotrophic growth. Journal of Bacteriology *177*, 247-251.

Reinthaler, T., Eva Sintes, Gerhard J. Herndl (2008). Dissolved organic matter and bacterial production and respiration in the sea-surface microlayer of the open Atlantic and the western Mediterranean Sea. Limnology and Oceanography *51*, 122-136.

Reisch, C.R., Crabb, W.M., Gifford, S.M., Teng, Q., Stoudemayer, M.J., Moran, M.A., and Whitman, W.B. (2013). Metabolism of dimethylsulphoniopropionate by *Ruegeria pomeroyi* DSS-3. Molecular Microbiology *89*, 774-791.

Reisch, C.R., Moran, M.A., and Whitman, W.B. (2008). Dimethylsulfoniopropionate-dependent demethylase (DmdA) from *Pelagibacter ubique* and *Silicibacter pomeroyi*. Journal of Bacteriology *190*, 8018-8024.

Reisch, C.R., Moran, M.A., and Whitman, W.B. (2011a). Bacterial catabolism of dimethylsulfoniopropionate (DMSP). Frontiers in Microbiology *2*, 172.

Reisch, C.R., Stoudemayer, M.J., Varaljay, V.A., Amster, I.J., Moran, M.A., and Whitman, W.B. (2011b). Novel pathway for assimilation of dimethylsulphoniopropionate widespread in marine bacteria. Nature *473*, 208-211.

Robbertse, B., Reeves, J.B., Schoch, C.L., and Spatafora, J.W. (2006). A phylogenomic analysis of the Ascomycota. Fungal Genetics and Biology *43*, 715-725.

Robbertse, B., Yoder, R.J., Boyd, A., Reeves, J., and Spatafora, J.W. (2011). Hal: an automated pipeline for phylogenetic analyses of genomic data. PLoS Currents *3*.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D., Eisen, J.A., Hoffman, J.M., Remington, K.*, et al.* (2007). The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biology *5*, e77.

Schattenhofer, M., Fuchs, B.M., Amann, R., Zubkov, M.V., Tarran, G.A., and Pernthaler, J. (2009). Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. Environmental microbiology *11*, 2078-2093.

Schein, C.H. (1990). Solubility as a function of protein structure and solvent components. Nature Biotechnology *8*, 308-317.

Schneider, K., Asao, M., Carter, M.S., and Alber, B.E. (2012). *Rhodobacter sphaeroides* uses a reductive route via propionyl coenzyme A to assimilate 3-hydroxypropionate. Journal of Bacteriology *194*, 225-232.

Schrader, J., Schilling, M., Holtmann, D., Sell, D., Filho, M.V., Marx, A., and Vorholt, J.A. (2009). Methanol-based industrial biotechnology: current status and future perspectives of methylotrophic bacteria. Trends in Biotechnology *27*, 107-115.

Schwalbach, M.S., Tripp, H.J., Steindler, L., Smith, D.P., and Giovannoni, S.J. (2009). The presence of the glycolysis operon in SAR11 genomes is positively correlated with ocean productivity. Environmental Microbiology *12*, 490-500.

Scott, D.A., Hickerson, S.M., Vickers, T.J., and Beverley, S.M. (2008). The role of the

mitochondrial glycine cleavage complex in the metabolism and virulence of the protozoan parasite *Leishmania major*. The Journal of Biological Chemistry *283*, 155-165.

Scrutton, N.S., and Leys, D. (2005). Crystal structure of DMGO provides a prototype for a new tetrahydrofolate-binding fold. Biochemical Society Transactions *33*, 776-779.

Seibel, B.A., and Walsh, P.J. (2002). Trimethylamine oxide accumulation in marine animals: relationship to acylglycerol storagej. Journal of Experimental Biology *205*, 297-306.

Sell, A. (1994). Phytoplankton-Excreted Organic Carbon. In Microbial Ecology of Lake Plußsee, J. Overbeck, and R. Chróst, eds. (Springer New York), pp. 81-91.

Serra, A.L., Mariscotti, J.F., Barra, J.L., Lucchesi, G.I., Domenech, C.E., and Lisa, A.T. (2002). Glycine betaine transmethylase mutant of *Pseudomonas aeruginosa*. Journal of Bacteriology *184*, 4301-4303.

Sharpton, T.J., Jospin, G., Langille, M.G.I., Wu, D., Pollard, K.S., and Eisen, J.A. (2012). Sifting through genomes with iterative sequence clustering produces a large, phylogenetically diverse protein-family resource. BMC bioinformatics. doi: 10.1186/1471-2105-13-264

Simó, R. (2001). Production of atmospheric sulfur by oceanic plankton: biogeochemical, ecological and evolutionary links. Trends in Ecology & Evolution *16*, 287-294.

Singh, H.B., Tabazadeh, A., Evans, M.J., Field, B.D., Jacob, D.J., Sachse, G., Crawford, J.H., Shetter, R., and Brune, W.H. (2003). Oxygenated volatile organic chemicals in the oceans: Inferences and implications based on atmospheric observations and air-sea exchange models. Geophysical Research Letters *30*, 1862.

Singh, R.K., and Singh, R.M. (1983). Diverse effects of formate on the assimilatory metabolism of nitrate and nitrite in *Rhizobium*. Journal of Biosciences *6*, 181-184.

Sinha, V.W., J; Riebesell, U; Meyerhoefer, M; Larsen, A; Paulino, A (2007). Air-sea fluxes of methanol, acetone, acetaldehyde, isoprene and DMS from a Norwegian fjord following a phytoplankton bloom in a mesocosm experiment. Atmospheric Chemistry and Physics *7*, 739-755.

Smith, D.P., Kitner, J.B., Norbeck, A.D., Clauss, T.R., Lipton, M.S., SSchwalbach, M., Steindler, L., Nicora, C.D., Smith, R.D., and Giovannoni, S.J. (2010). Transcriptional and translational regulatory responses to iron limitation in the globally distributed marine bacterium *Candidatus* Pelagibacter ubique. PloS One *5*, 1-10.

Smith, D.P., Thrash, J.C., Nicora, C.D., Lipton, M.S., Burnum, K.E., Carini, P., Smith, R.D., and Giovannoni, S.J. (2013). Proteomic and transcriptomic analysis of *Candidatus* Pelagibacter ubique describes the first PII-independent response to nitrogen limitation in a free-living alphaproteobacterium. MBio *4*, doi: 10.1128/mBio.00133-12

Sorokin, D.Y., Tourova, T.P., Kovaleva, O.L., Kuenen, J.G., and Muyzer, G. (2010). Aerobic carboxydotrophy under extremely haloalkaline conditions in *Alkalispirillum/Alkalilimnicola* strains isolated from soda lakes. Microbiology *156*, 819-827.

Sowell, S.M., Abraham, P.E., Shah, M., Verberkmoes, N.C., Smith, D.P., Barofsky, D.F., and Giovannoni, S.J. (2011). Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. The ISME Journal *5*, 856-865

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688-2690.

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

Stefels, J., Steinke, M., Turner, S., Malin, G., and Belviso, S. (2007). Environmental constraints on the production and removal of the climatically active gas dimethylsulphide (DMS) and implications for ecosystem modelling. Biogeochemistry *83*, 245-275.

Steindler, L., Schwalbach, M.S., Smith, D.P., Chan, F., and Giovannoni, S.J. (2011). Energy starved *Candidatus* Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. PLoS One *6*, doi: 10.1371/journal.pone.0019725.

Studer, A., McAnulla, C., Buchele, R., Leisinger, T., and Vuilleumier, S. (2002). Chloromethane-induced genes define a third C1 utilization pathway in *Methylobacterium chloromethanicum* CM4. Journal of Bacteriology *184*, 3476-3484.

Sun, J., Steindler, L., Thrash, J.C., Halsey, K.H., Smith, D.P., Carter, A., Landry, Z., and Giovannoni, S.J. (2011). One Carbon Metabolism in SAR11 Pelagic Marine Bacteria. PLoS One *6*, doi: 10.1371/journal.pone.0023973.

Sunda, W., Kieber, D.J., Kiene, R.P., and Huntsman, S. (2002). An antioxidant function for DMSP and DMS in marine algae. Nature *418*, 317-320.

Takle, G., Toth, I., and Brurberg, M. (2007). Evaluation of reference genes for real-time RT-PCR expression studies in the plant pathogen *Pectobacterium atrosepticum*. BMC

Plant Biology *7*, 1-9.

Teplyakov, A., Obmolova, G., Sarikaya, E., Pullalarevu, S., Krajewski, W., Galkin, A., Howard, A.J., Herzberg, O., and Gilliland, G.L. (2004). Crystal structure of the YgfZ protein from *Escherichia coli* suggests a folate-dependent regulatory role in one-carbon metabolism. Journal of Bacteriology *186*, 7134-7140.

Thrash, J.C., Boyd, A., Huggett, M.J., Grote, J., Carini, P., Yoder, R.J., Robbertse, B., Spatafora, J.W., Rappe, M.S., and Giovannoni, S.J. (2011). Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. Scientific Report *1*, doi:10.1038/srep00013

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

Todd, J.D., Rogers, R., Li, Y.G., Wexler, M., Bond, P.L., Sun, L., Curson, A.R.J., Malin, G., Steinke, M., and Johnston, A.W.B. (2007). Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. Science *315*, 666-669.

Treusch, A.H., Vergin, K.L., Finlay, L.A., Donatz, M.G., Burton, R.M., Carlson, C.A., and Giovannoni, S.J. (2009). Seasonality and vertical structure of microbial communities in an ocean gyre. The ISME Journal *3*, 1148-1163.

Trickey, P., Wagner, M.A., Jorns, M.S., and Mathews, F.S. (1999). Monomeric sarcosine oxidase: structure of a covalently flavinylated amine oxidizing enzyme. Structure *7*, 331-345.

Tripp, H.J. (2008). Counting marine microbes with Guava Easy-Cyte 96 well plate reading flow cytometer. Protocol Exchange, doi:10.1038/nprot.2008.29

Tripp, H.J., Kitner, J.B., Schwalbach, M.S., Dacey, J.W., Wilhelm, L.J., and Giovannoni, S.J. (2008). SAR11 marine bacteria require exogenous reduced sulphur for growth. Nature *452*, 741-744.

Tripp, H.J., Schwalbach, M.S., Meyer, M.M., Kitner, J.B., Breaker, R.R., and Giovannoni, S.J. (2009). Unique glycine-activated riboswitch linked to glycine-serine auxotrophy in SAR11. Environmental Microbiology *11*, 230-238.

van Dongen, S. (2000). Graph clustering by flow simulation. Utrecht, Netherlands: Doctoral dissertation, University of Utrecht. Retrieved July 25, 2011, from

website:http://igitur-archive.library.uu.nl/dissertations/1895620/inhoud.htm.

van Duyl, F.C., Gieskes, W.W.C., Kop, A.J., and Lewis, W.E. (1998). Biological control of short-term variations in the concentration of DMSP and DMS during a *Phaeocystis* spring bloom. Journal of Sea Research *40*, 221-231.

van Ophem, P.W., Van Beeumen, J., and Duine, J.A. (1992). NAD-linked, factor-dependent formaldehyde dehydrogenase or trimeric, zinc-containing, long-chain alcohol dehydrogenase from Amycolatopsis methanolica. European Journal of Biochemistry *206*, 511-518.

Vandecasteele, J.-P. (2008). Petroleum Microbiology: Concepts Environmental Implications Industrial Applications (Paris Editions Technip).

Vazquez-Landaverde, P.A., Torres, J.A., and Qian, M.C. (2006). Quantification of trace volatile sulfur compounds in milk by solid-phase microextraction and gas chromatography-pulsed flame photometric detection. Journal of Dairy Science *89*, 2919-2927.

Vergin, K.L., Beszteri, B., Monier, A., Cameron Thrash, J., Temperton, B., Treusch, A.H., Kilpert, F., Worden, A.Z., and Giovannoni, S.J. (2013). High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. The ISME Journal *7*, 1322-1332.

Vila-Costa, M., Rinta-Kanto, J.M., Sun, S., Sharma, S., Poretsky, R., and Moran, M.A. (2010). Transcriptomic analysis of a marine bacterial community enriched with dimethylsulfoniopropionate. The ISME Journal *4*, 1410-1420.

Vorholt, J.A. (2002). Cofactor-dependent pathways of formaldehyde oxidation in methylotrophic bacteria. Archives of Microbiology *178*, 239-249.

Vorholt, J.A., Chistoserdova, L., Stolyar, S.M., Thauer, R.K., and Lidstrom, M.E. (1999). Distribution of tetrahydromethanopterin-dependent enzymes in methylotrophic bacteria and phylogeny of methenyl tetrahydromethanopterin cyclohydrolases. Journal of Bacteriology *181*, 5750-5757.

Wang, W., Zhang, W., Lu, J., Yang, Y., Chiao, J., Zhao, G., and Jiang, W. (2002). MoeA, an enzyme in the molybdopterin synthesis pathway, is required for rifamycin SV production in *Amycolatopsis mediterranei* U32. Applied Microbiology Biotechnology *60*, 139-146.

Wilhelm, L.J., Tripp, H.J., Givan, S.A., Smith, D.P., and Giovannoni, S.J. (2007). Natural

variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. Biology Direct *2*, 27.

Wilhelm, S.W., and Suttle, C.A. (1999). Viruses and nutrient cycles in the sea. Bioscience *49*, 781-788

Willey, J.D., Kieber, R.J., Eyman, M.S., and Avery, G.B. (2000). Rainwater dissolved organic carbon: Concentrations and global flux. Global Biogeochemical Cycles *14*, 139-148.

Williams, J., Holzinger, R., Gros, V., Xu, X., Atlas, E., and Wallace, D.W.R. (2004). Measurements of organic species in air and seawater from the tropical Atlantic. Geophysical Research Letters *31*, doi: 10.1029/2004GL020012.

Wilson, S.M., Gleisten, M.P., and Donohue, T.J. (2008). Identification of proteins involved in formaldehyde metabolism by *Rhodobacter sphaeroides*. Microbiology *154*, 296-305.

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

Yang, M., Nightingale, P.D., Beale, R., Liss, P.S., Blomquist, B., and Fairall, C. (2013). Atmospheric deposition of methanol over the Atlantic Ocean. Proceedings of the National Academy of Sciences, doi: 10.1073/pnas.1317840110

Yoch, D.C. (2002). Dimethylsulfoniopropionate: its sources, role in the marine food web, and biological degradation to dimethylsulfide. Applied and Environmental Microbiology *68*, 5804-5815.

Yurimoto, H., Kato, N., and Sakai, Y. (2005). Assimilation, dissimilation, and detoxification of formaldehyde, a central metabolic intermediate of methylotrophic metabolism. The Chemical record *5*, 367-375.

Zimmerman, P.R., Chatfield, R.B., Fishman, J., Crutzen, P.J., and Hanst, P.L. (1978). Estimates on the production of CO and H_2 from the oxidation of hydrocarbon emissions from vegetation. Geophysical Research Letters *5*, 679-682.

APPENDICES

Appendix 1 is published as a supplemental information documents to: doi:10.1371/journal.pone.0023973. Appendix 2 will accompany the submission of Chapter 3.

Appendix 1

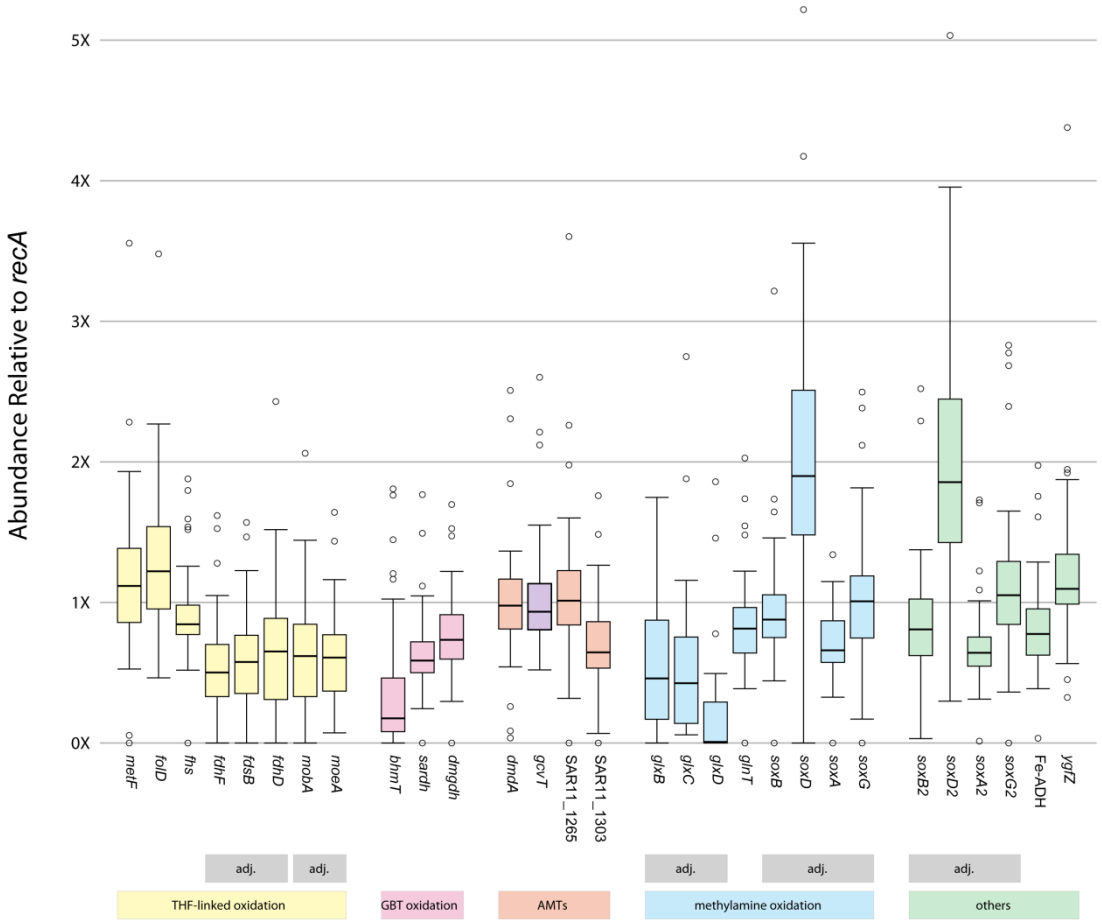


Figure S2.1 The abundance of SAR11 C1 metabolism genes in GOS data, relative to SAR11 *recA* genes. Genes were identified as SAR11 by a reciprocal best BLAST (RBB) approach. SAR11 C1 genes with frequencies less than SAR11 *recA* (< 1x) may indicate that only subpopulations of SAR11 cells possess that gene; genes greater than 1x suggest that multiple copies of that gene are present per cell. Boxes encompass points between the 25th and 75th percentiles, with the median represented as a thick horizontal line. Whiskers span the minimal distance needed to include all points within 1.5 x the interquartile range beyond the interquartile boundary, with points outside of this range rendered individually as circles. For each gene, n = 40. Abbreviations: adj., genomically adjacent genes.

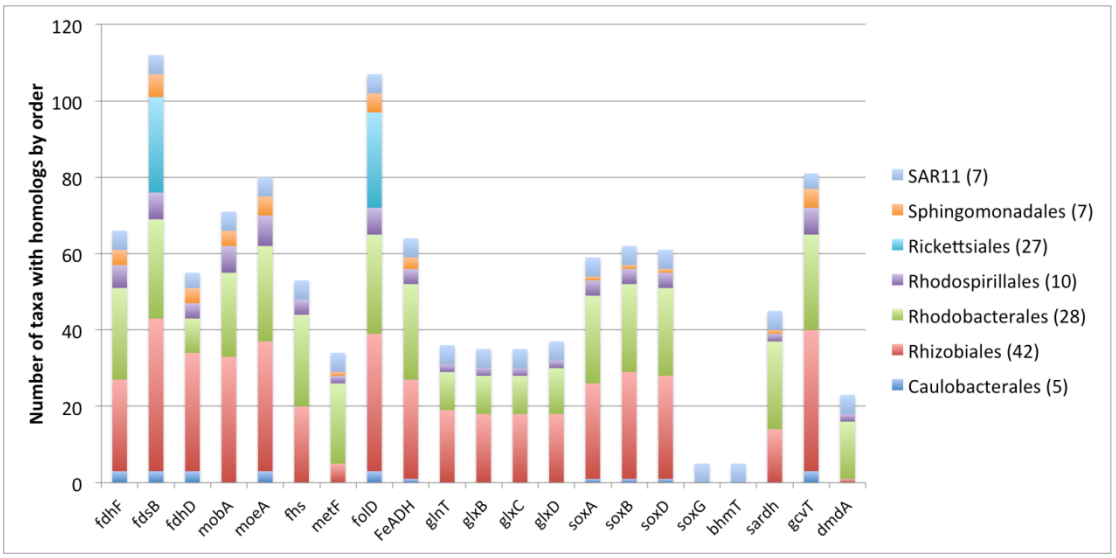


Figure S2.2 Distribution of C1 gene homologs throughout the *Alphaproteobacteria*. The number of genomes containing homologs of C1 oxidation genes reported by gene and divided by Order. The total number of genomes examined for each order is in parentheses.

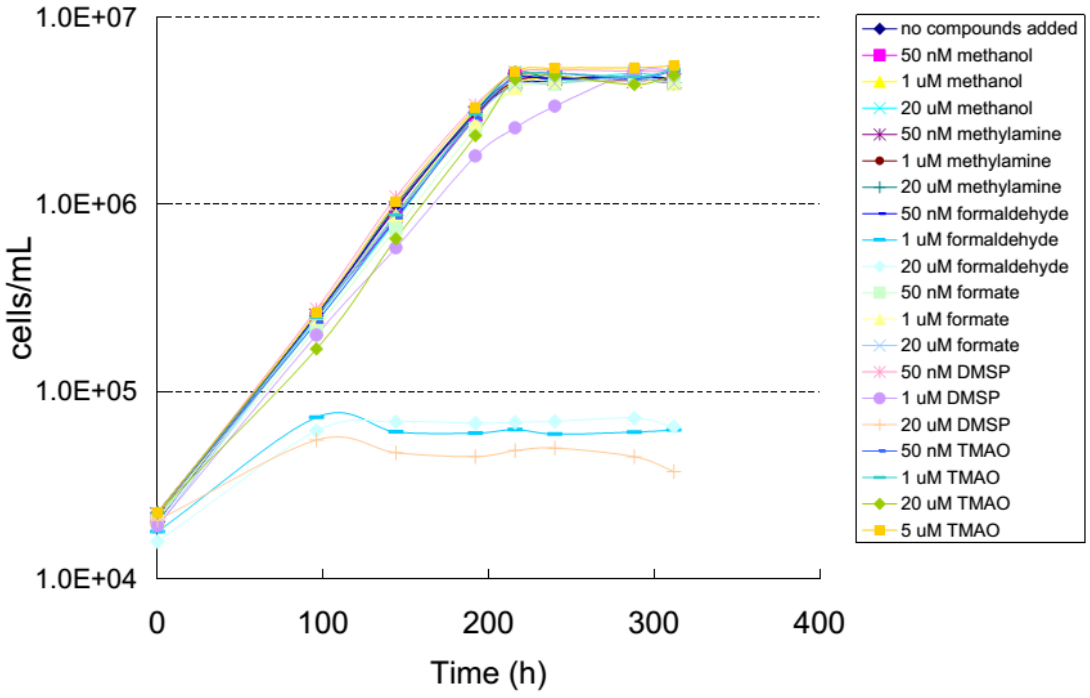


Figure S2.3 Culture experiments to determine the concentrations of C1 and methylated compounds for ATP and radioisotope assays. HTCC1062 cells were cultured in seawater medium amended with 10 µM NH_4Cl, 1 µM KH_2PO_4, 10 nM $FeCl_3$, vitamins, and C1 and methylated compounds at different concentrations.

Appendix 2

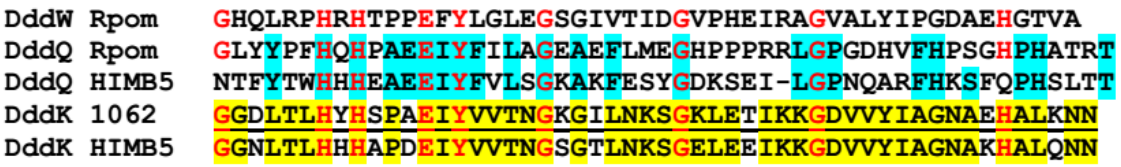


Figure S3.1. Comparison of SAR11 DddK and DddQ-like polypeptides with other cupin DMSP lyases. The amino acid sequences of the cupin domains of the known DddW and DddQ polypeptides in *Ruegeria pomeroyi* are lined up in comparison with the DddK polypeptides of SAR11 strains 1062 and HIMB5 (gene products SAR11_0394 and HIMB5_00004730 respectively) and the DddQ-like product of the HIMB5_00000220 gene of SAR11 strain HIMB5. Residues conserved in all these polypeptides are shown as red letters. Yellow shading denotes identical residues in the two DddK polypeptides and turquoise shading shows those residues in common in the DddQs of SAR11 HIMB5 and of *R. pomeroyi.*

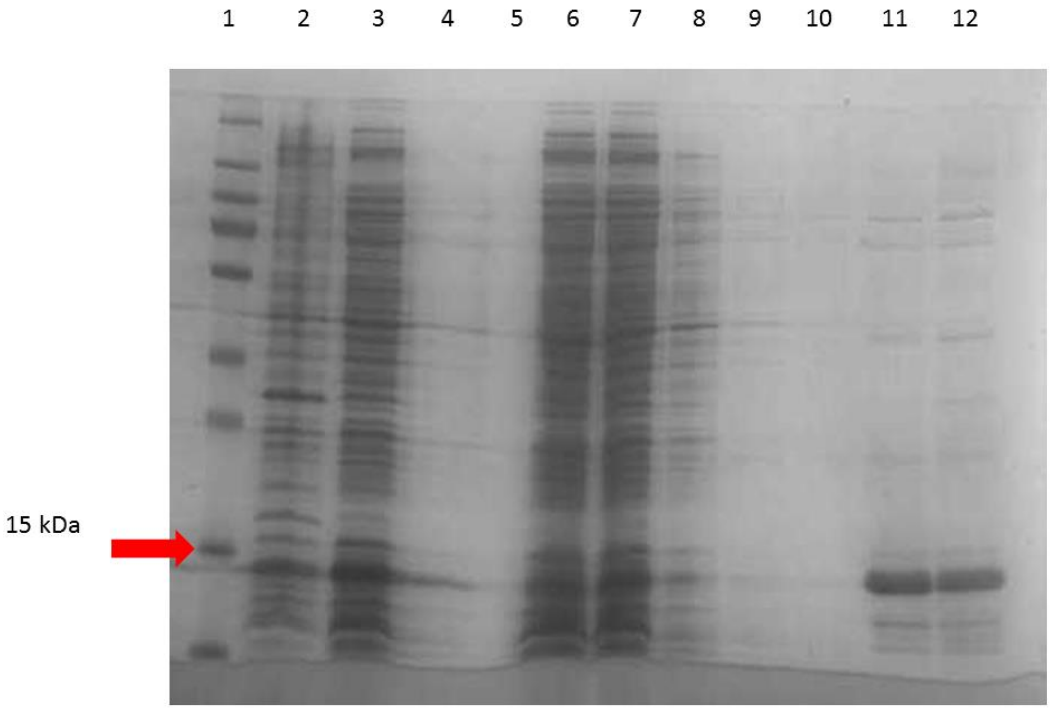


Figure S3.2: Stained SDS-PAGE image showing partial purification of histidine-tagged DddK. Native DddK has a predicted molecular mass of 14.42 kDa. Lane 1 = Precision Plus Protein Dual Colour Standards (Biorad); Lane 2 = Insoluble fraction; Lane 3 = Soluble fraction; Lane 6 = First spin through Ni-NTA column; Lane 7 = Second spin through Ni-NTA column; Lane 8 = First wash; Lane 9 = Second wash; Lane 10 = Third wash; Lane 11 = First elution; Lane 12 = Second elution.

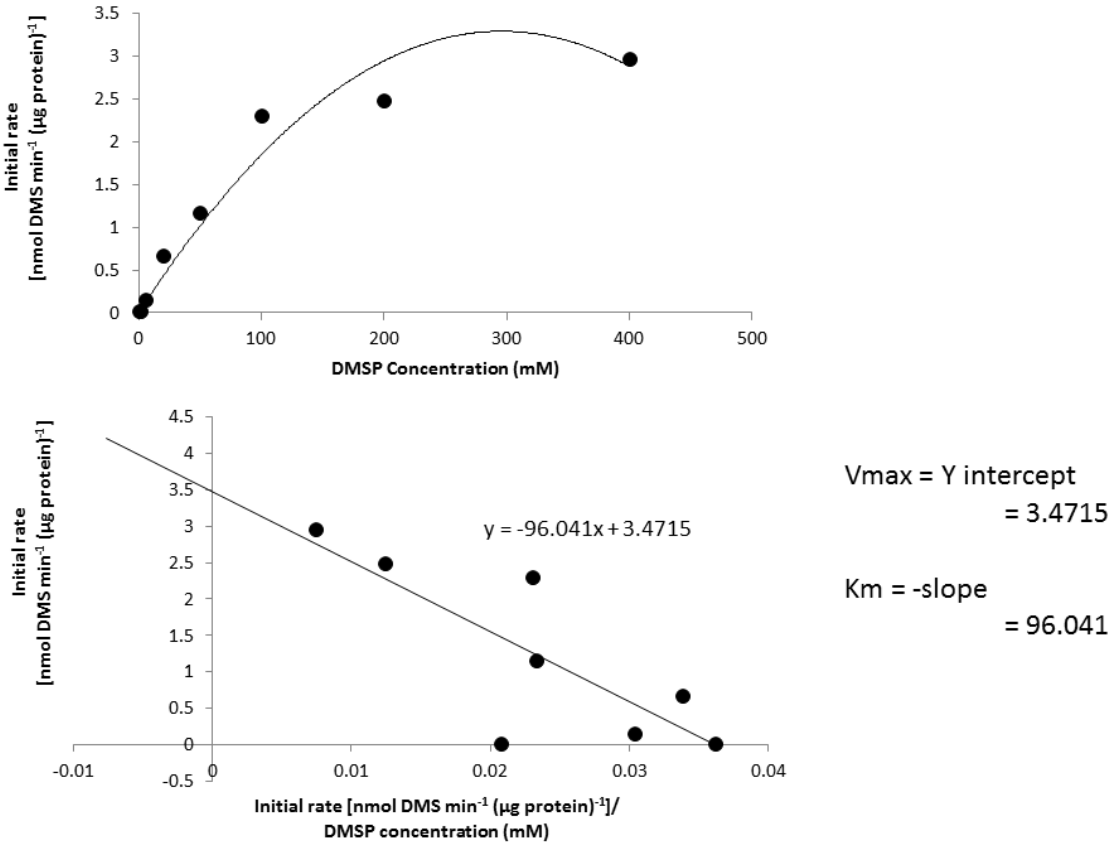


Figure S3.3. Estimating the Km and Vmax of the DddK by Michaelis-Menten Kinetics.

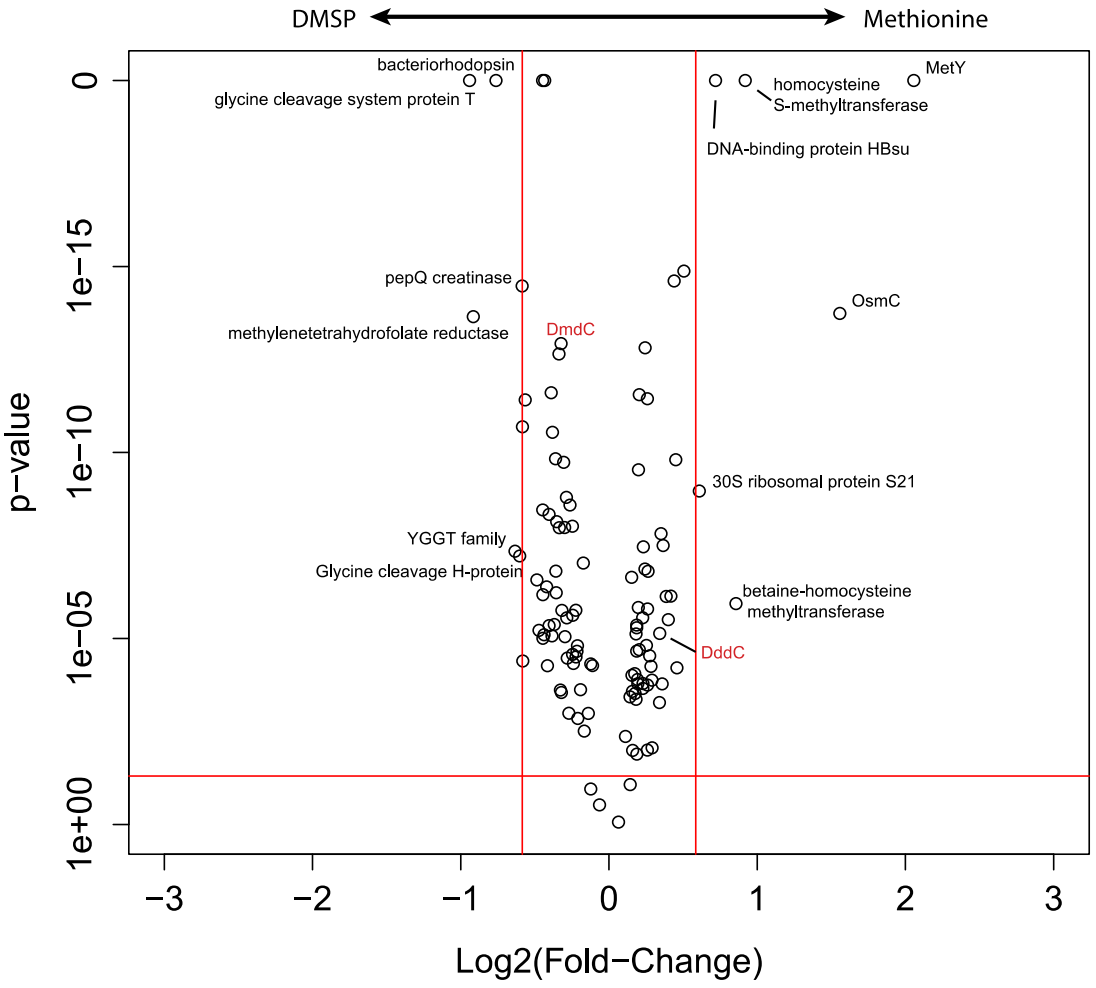


Figure S3.4 Volcano plot of Table 3.2. Horizontal red line indicates a p-value cut-off of 0.05; Vertical red lines indicate boundaries of 1.5-fold difference in expression

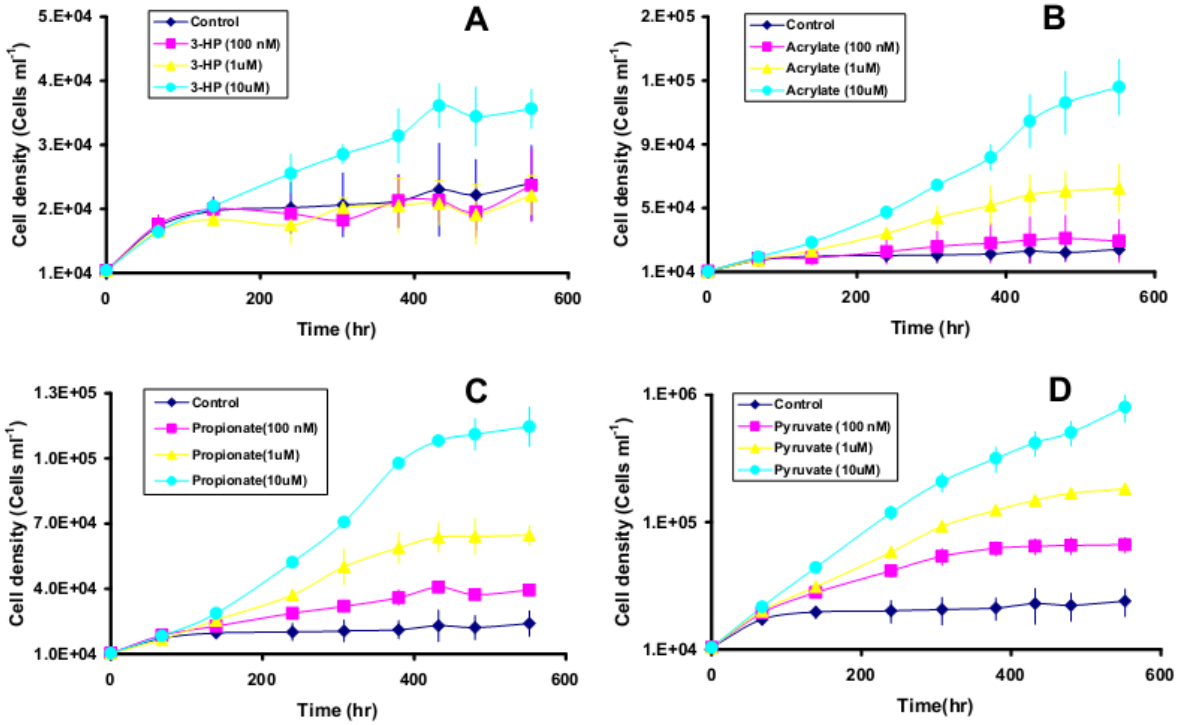


Figure S3.5 Culture experiments to determine the concentrations of C3 compounds. HTCC1062 cells were grown in ASW media-B amended different concentrations of each compound (3-HP (A), acrylate (B) or propionate (C)). The positive control was treated with pyruvate (D). Points are the average density of triplicate cultures. Where not visible, error bars are smaller than the size of the symbols. Note: the vertical (value) axis of (D) is in a logarithmic scale.