

**Phosphonate biogeochemical cycling in the marine environment:
from an ocean scale to a molecular scale**

by

Marianne Acker

B.S. Université Paris Diderot (Paris 7), 2013
M.S. Institut Universitaire Européen de la Mer, 2015

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
and the
WOODS HOLE OCEANOGRAPHIC INSTITUTION

May 2021

© Marianne Acker. All rights reserved.

The author hereby grants to MIT and WHOI permission to reproduce and distribute publicity paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Signature of Author.....

Joint Program in Oceanography/Applied Ocean Science and Engineering
Massachusetts Institute of Technology
and Woods Hole Oceanographic Institution
March 5th, 2021

Certified by

Dr. Daniel J. Repeta
Thesis Supervisor
Woods Hole Oceanographic Institution

Accepted by

Dr. Colleen M. Hansel
Chair, Joint Committee for Chemical Oceanography
Woods Hole Oceanographic Institution

Phosphonate biogeochemical cycling in the marine environment: from an ocean scale to a molecular scale

by

Marianne Acker

Submitted to the Massachusetts Institute of Technology and Woods Hole Oceanographic Institution and to the Department of Earth, Atmospheric and Planetary Sciences on March 5th, 2021 in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Chemical Oceanography

ABSTRACT

The existence of a marine phosphorus (P) redox cycle was recently confirmed when phosphonates, organophosphorus compounds with P in the (III) oxidation state, were found in high molecular weight dissolved organic matter. Although some features of the P redox cycle have come to light since the discovery of phosphonates, many aspects of phosphonate production, cycling and fate remain unknown. To address these gaps in our understanding, we studied phosphonate cycling in the Eastern Mediterranean Sea, a chronically P-limited basin, using ^{33}P and enzymatic assays. We showed that phosphonate production was low but consumption was high, suggesting that phosphonate production and consumption may be spatially or temporally decoupled. We also explored phosphonate production in the model marine cyanobacterium *Prochlorococcus SB*. Using ^{31}P NMR, we found *Prochlorococcus SB* allocates ~50% of its cellular P to phosphonates. Allocation of P to phosphonates was conserved under P-limitation, and further investigation revealed phosphonates were associated with proteins. The discovery of phosphonoproteins in *Prochlorococcus SB* opens new perspectives on the biochemical function of phosphonates and their role in P-cycling. Finally, we developed a new P-targeted method to characterize marine organophosphorus compounds using liquid chromatography coupled to electrospray ionization and inductively coupled plasma mass spectrometry.

Thesis supervisor: Dr. Daniel J. Repeta

Title: Senior Scientist, Marine Chemistry and Geochemistry Department, Woods Hole Oceanographic Institution

ACKNOWLEDGMENTS

This work was supported by the Simons Foundation under grant numbers POP49476 and 721227 [D. Repeta], the Gordon and Betty Moore Foundation under the grant number 6000 [D. Repeta] and the National Science Foundation OCE under the grant number 1634080 [D. Repeta].

This PhD thesis is the result of many years of work that would not have been possible without a lot of significant people and I would like to acknowledge and thank them for their support and companionship throughout the years.

Starting with my advisor Daniel Repeta. Dan took a chance on me by giving me the opportunity to join his lab as a visiting student despite my wonky English and my weird French accent. Despite what he knew, he then offered me a PhD position. Dan took the time to read every single thing I wrote and gave me great feedbacks which taught me a lot. He made me practice all my presentations and helped improve them beyond what I imagined. He taught me a way of thinking about science completely different from what I had learned. Thanks to Dan's guidance I really grew as a scientist. Dan also really supported me as a person.

Working with Dan also gave me the chance to meet amazing lab mates who were a daily support. Lydia Babcock Adams with whom I shared many flights across the US to conferences and cruises and who, despite my terrible French accent, always paid attention and answered with kind words or useful advice. Benjamin Granzow and Jinxhuan Li who were always willing to help in the lab or pitch in new ideas. Thank you for letting me crash your office once in a while. Past members of the lab also influenced me greatly. Rene Boiteau, who encouraged me to apply to the Joint Program and of course Randie Bundy. Randie took care of me from Day 1, brought me on countless adventures and gave me a great example to look upon to. More importantly, she showed me that you could be successful while enjoying life and being kind, respectful and compassionate. With Randie, I met an awesome scientist, a beautiful person, a great friend, a mentor for life and Mattias Cape who has been a terrific friend as well.

I was also fortunate enough to be welcomed for a great amount of time in The Chisholm lab, filled with knowledgeable and overall great people. Among them, I would like to thank Allison Coe, the *Prochlorococcus* supergrouper who patiently taught me how to culture, Kevin Dooley, Christina Bliem, Aldo and Tyler who transferred my cultures on several occasions, Paul Berube who was always of great advice, Raphael Laurenceau for all our French and expat chats and of course Shane Hogle for its introductory classes in genomics as well as its mentorship and patience while we were writing our paper. This lab wouldn't exist without the great PI that

Penny is and I feel so lucky that Penny agreed to foster me but also took of her precious time to offer me honest guidance and career advice.

Penny was part of my thesis committee along with Benjamin Van Mooy, David Karl and Ingrid Obernosterer. I would like to thank Ben for hosting me in his lab to perform the radioactive work and his thought provoking questions, Dave for having me at UH to learn how to measure methane (with the training of Samuel Wilson), Ingrid for her advices, encouraging words and our discussions on French academia, and all of them for the time they took to discuss my science. I would like to extend my gratitude to Elizabeth Kujawinski who was part of my PhD milestones by perfectly chairing my general oral exam, my thesis proposal oral and my PhD defense.

The Joint Program allowed me to meet great people starting with my cohort Jennifer Karowleski and Tyani Huang as it was comforting to share all the hurdles on our way. Joleen Heiderich whom I met on Day 1 has been a terrific dance partner and European friend. It also brought me Kalina Grabb who started as my mentee and became a friend. With Kalina, I found a partner for winter surfing and adventures, someone to share my deepest struggles with, someone who supported me and cared for me every step of the way. Kalina became very quickly a true friend that I attend to keep for a long time!

WHOI is a tight community thanks to which I met numerous people who supported me throughout that journey. Everyone in the Watson building and more specifically Mary Szarowski, our Watson mum who has been the best at helping with all administrative issues I encountered, Harriet Alexander with whom I shared surfing adventures and long chats, Gretta who always have had an attentive ear and brought me chocolate on many occasions and Emilie Skoog for her constant cheering and new friendship. I am also extremely thankful to have met Matthew McIlvin as he is a very talented mass spectrometrists (the mass spec whisperer) while being a kind human and have been of a huge help, especially in those past month, when developing the method. Carl Johnson also contributed a lot to my thesis with his unevaluable help with the NMR. I also would like to thank the entire Chemistry department led by Bernhard and Mary, Sheila who printed all my posters with such a good care as well as all the faculty with a special mention to Colleen Hansell and Jeff Seewald our JCCO chairs, the postdocs (Jaci Saunders, Kevin Becker, Alina Ebling, Mike Mazzotta, Nathalie Cohen, Sarah Hu especially) and the students I have met during those 5 years.

Of course my time in Woods Hole would not have been the same without my officemates and roommates. Our international office with Luis Valentin and Noelle Held then Koko Kunde warmed my heart and became my safe and happy work space. Noelle and I also shared a house and just like that we started to spend 24/7 together. We supported and took care of each other throughout our PhDs. Her calm facing the storms, her efficiency, the way she handled multiple complicated tasks at the same time inspired me so much and I am so happy she became one of

my best friend. Brittany Widner joined our household and her strong will, emotional awareness, vulnerability and kindness encouraged me to develop my own. Our texts everyday have been a tremendous support this past year! My newest housemates Kali Horn and Vashan Wright were so tolerant and understanding in those past crazy months that I will always be extremely grateful for them. Kali paid much attention to my wellbeing which meant a lot to me and I for sure will keep her close to my heart in the years to come.

One of the best support I received throughout my PhD came from Boston Team Handball club. This team changed a lot across the years but always was a second family. Interacting with people from all over the world with different nationalities, languages, backgrounds, cultures, religions etc. really brought new perspectives in my life. Thanks to this team, I met lifelong friends: Natalya, Audrey, Hannah, Amir and his family, Bruna and so many others. The other essential part was the French Connection. At WHOI with Emilie, Thomas and David. Merci pour ces soirées apéros/repas et surtout merci pour le rangement. At MIT with Maxime, Timothee, Julien and Zizi. Merci pour ces bons moments à parler politique ou à fermer le Phoenix Landing et autres bars bostoniens. Ma Zizi, merci pour tout ce que tu sais déjà. Tu m'as changé, pour le meilleur.

I wouldn't have been able to achieve this PhD without my support system back home. Coming back to France and visit my childhood (and not so childhood friends) brought me so much joy and kept me going in the hardest times. Merci les monos pour ces étés follichons plein de voile et de virées au bar, merci Alex de m'avoir porté chance, merci Aude ma « sœur » et cette amitié loin des yeux mais près du cœur, merci Mathilde, ma BBF, et ces rendez-vous annuels à la plage. Merci Louise pour ces apéros post-confinement, merci Eva pour ta joie, tes convictions et ta gentillesse, merci Bertille de toujours faire partie de ma vie et de me donner confiance en moi. Merci les Kekettes pour nos retrouvailles récentes qui font remonter tant d'agréables souvenirs. Et surtout : merci mes Sans Culottes. Grandir, évoluer, changer avec vous fut une source de bonheur constant et toutes ces années passées ensemble font en grande partie de moi qui je suis aujourd'hui. Malgré les carottes et autres incompréhensions, j'ai toujours senti votre soutien et votre amitié dans cette dure épreuve. Ma Bibi, merci pour ces coups de téléphone à pas d'heure et tes histoires hilarantes. Ma Kikette, merci pour cette joie de vivre qui irradie partout où tu vas, ce petit grain de folie qui remplit de joie mais surtout de t'être toujours soucie de moi. Ma Vava, merci d'avoir toujours cru en moi et d'avoir été capable de passer au-dessus de tout. Ma Claudia, merci pour ces vacances au ski, en Espagne, à Malte, pour cette légèreté que tu apportes dans tout ce que tu fais et de faire partie de ma vie depuis tant d'années. Ma Gogo, merci pour ces

paquets, ces cadeaux, ces vacances et autres attentions qui me sont allés droit au cœur, merci d'avoir su être là quand j'en avais besoin et merci pour cette grande et formidable amitié.

I would also like to thank my therapist back home, Uta who has really helped me go through those challenging month with more consciousness and appreciation for myself and my work.

Finally, I would like to thank my family. First my grandparents who taught me invaluable life lessons and cared for me on so many occasions. Mamy avec ses gros yeux mais son sourire chaleureux, Papy avec ses cours de maths et ses yeux rieurs, Pépé avec sa franchise et ses farces et enfin Mémé avec sa tête de bois et ses bonnes confitures et petits plats. Tous avec leurs forts caractères, tous droits dans leurs bottes et tous qui m'ont transmis cette authenticité et entèreté. Mes cousins, Quentin, Marie et Sébastien, pour ces vacances en famille et ces après-midi à se courir après. Mes oncles et ma tante. Bien sûr, mon frère Vincent et ma sœur Louise et mes parents Danielle et Renaud. Merci pour ces Skypes et ces moments en famille qui ont été si essentiels pour me ressourcer. Merci de m'avoir poussé à donner le meilleur de moi-même et à devenir, pour vous, ma meilleure version. Merci d'avoir toujours été présents et de m'avoir donné votre amour à votre façon. Je n'aurai jamais les mots suffisants pour décrire la fierté que j'ai à faire partie de cette famille et l'amour que je ressens à votre égard. Cette thèse, elle est pour vous.

Table of contents

ABSTRACT	3
ACKNOWLEDGMENTS	4
CHAPTER 1. INTRODUCTION.....	16
1.1 MOTIVATION AND BRIEF OVERVIEW OF THE PHOSPHORUS CYCLE.....	17
1.2 PHOSPHONATES.....	19
1.2.1 Phosphonate relevance in P biogeochemical cycling	19
1.2.2 Study of phosphonate at a basin scale.....	20
1.2.3 Study of phosphonate at a cellular scale	21
1.2.4 Molecular level characterization of organic phosphorus in marine samples.....	21
1.3 REFERENCES	23
CHAPTER 2: PHOSPHONATE BIOGEOCHEMICAL CYCLING IN THE EASTERN MEDITERRANEAN SEA.....	26
2.1 INTRODUCTION	27
2.2 MATERIAL AND METHODS	28
2.2.1 Cruise track and sampling.....	28
2.2.2 Methane sampling and measurements.....	29
2.2.3 Methane and ethylene supersaturation and air-sea gas exchange.....	30
2.2.4 C-P lyase activity sampling and measurements	30
2.2.5 ³³ P incubations sampling and treatments	31
2.2.6 ³³ P phosphonate separation.....	32
2.2.7 ³³ P _i uptake and turnover rates calculations	33
2.2.8 ³³ P phosphonate fraction and production rate calculations	33
2.3 RESULTS.....	34
2.3.1 Dissolved methane concentrations and stable carbon isotope values	34
2.3.2 C-P lyase activity.....	37
2.3.3 P _i concentrations, uptake and turnover rate	38
2.3.4 Phosphonate production rates.....	38
2.3.5 Phosphonate diversity across stations	39
2.4 DISCUSSION	40
2.4.1 Phosphonate consumption and production in the EMS compare to other oligotrophic areas.....	40
2.4.1.1 Phosphonate consumption.....	40
2.4.1.2 Phosphonate production	41

2.4.2 Comparison between phosphonate consumption proxies.....	43
2.4.3 Comparison between phosphonate production and consumption via C-P lyase.....	44
2.5 CONCLUSION.....	45
2.6 REFERENCES.....	47
CHAPTER 3. PHOSPHONATE PRODUCTION BY MARINE MICROBES: SOURCES AND POTENTIAL FUNCTION.....	51
3.1 PREFACE	52
3.2 ABSTRACT	52
3.3 INTRODUCTION	53
3.4 RESULTS AND DISCUSSION	55
3.4.1 The taxonomic distribution of phosphonate producers and consumers	55
3.4.2 The abundance of phosphonate producing and consuming microbes in the oceans	57
3.4.3 Phosphonate production in <i>Prochlorococcus</i> SB.....	60
3.4.4 Phosphonates and P storage	62
3.4.5 Macromolecular forms of phosphonates.....	64
3.4.6 Functional roles of phosphorylated glycoproteins	65
3.5 CONCLUSION: BIOGEOCHEMICAL IMPLICATIONS - A MICROBIAL PHOSPHONATE LOOP WITHIN THE MARINE PHOSPHORUS CYCLE	66
3.6 MATERIAL AND METHODS	67
3.6.1 Genomic data sources	67
3.6.2 Homology searches	68
3.6.3 Multiple sequence alignments, phylogenetic inference, and topological comparison	69
3.6.4 Gene enrichment analysis.....	70
3.6.5 Identification of PepM sequences within <i>Prochlorococcus</i> genomic islands	70
3.6.6 Estimation of PepM prevalence in <i>Prochlorococcus</i> and SAR11 genomes.....	71
3.6.7 Metagenome read classification	71
3.6.8 Biotic and abiotic data associated with metagenomes.....	72
3.6.9 Random Forest Regression	72
3.6.10 Beta-Binomial Regression and Generalized Additive Models	73
3.6.11 <i>Prochlorococcus</i> cultures under P-replete and P-deficient conditions.....	74
3.6.12 Cell harvest and treatment	74
3.6.13 Nuclear Magnetic Resonance	75
3.6.14 Elemental composition of <i>Prochlorococcus</i> SB and MIT 9301.....	75
3.6.15 Separation of cellular macromolecular components	75
3.6.16 Protein extraction and precipitation	76

3.6.17 Dataset/code availability.....	77
3.7 ACKNOWLEDGMENTS	77
3.8 REFERENCES	78
3.9 SUPPLEMENTARY MATERIAL	86
3.9.1 Supplementary Notes.....	86
Supplementary Note S3.1: Taxonomic breadth of marine phosphonate producers and consumers	86
Supplementary Note S3.2: Seasonal dependence of phosphonate producers at BATS	87
Supplementary Note S3.3: Biotic and abiotic covariates shaping the global distribution of phosphonate producers	88
Supplementary Note S3.4: Evidence for phosphonylated surface layer structures by Prochlorococcus SB.....	89
Supplementary Note S3.5: Robust quantification of PepM and marker genes using short metagenomic reads	91
Supplementary Note S3.6: Random forest regression.....	92
3.9.2 Supplementary Tables	94
3.9.3 Supplementary Figures	98
3.9.4 Supplementary references	106
CHAPTER 4. DEVELOPMENT AND OPTIMIZATION OF A METHOD FOR THE IDENTIFICATION OF ORGANOPHOSPHORUS COMPOUNDS USING LIQUID CHROMATOGRAPHY COUPLED WITH INDUCTIVELY COUPLED PLASMA MASS SPECTROMETRY AND ELECTROSPRAY IONIZATION MASS SPECTROMETRY (LC-ICPMS-ESIMS)	110
4.1 INTRODUCTION	111
4.2 PHOSPHONATE SEPARATION BY LIQUID CHROMATOGRAPHY	113
4.2.1 Choice of the column and the standard mixture	113
4.2.2 Optimization of the mobile phase	114
4.3 DETECTION OF PHOSPHORUS USING INDUCTIVELY COUPLED PLASMA MASS SPECTROMETRY (ICPMS)	117
4.3.1 Instrumentation and initial conditions.....	117
4.3.2 Optimization of ICPMS detection sensitivity	118
4.3.2.1 Use of oxygen as a collision gas	118
4.3.2.2 Optimization of dwell time	119
4.3.2.3 Optimization of the oxygen flow.....	120
4.3.2.4 Summary.....	121
4.4 DETECTION OF ORGANOPHOSPHORUS COMPOUNDS BY ELECTROSPRAY IONIZATION MASS SPECTROMETRY (ESIMS)	123

4.4.1 Instrumentation and initial conditions.....	123
4.4.2 ESIMS parameters optimization	123
4.4.3 MS ² of organophosphorus compounds using precursor-product ions	126
4.5 CONCLUSION AND RECOMMENDATIONS FOR FUTURE METHOD IMPROVEMENT	127
4.6 REFERENCES	128
CHAPTER 5. IDENTIFICATION OF ORGANOPHOSPHORUS COMPOUNDS IN MARINE SAMPLES USING LC-ICPMS-ESIMS.....	132
5.1 INTRODUCTION	133
5.2 MATERIAL AND METHODS	134
5.2.1 Solvents, standards and columns.....	134
5.2.2 Sample preparation.....	134
5.2.2.1 Organic phosphorus in Prochlorococcus MIT9313	134
5.2.2.2 Organic phosphorus in Prochlorococcus SB POM and LMWDOM	135
5.2.2.3 North Atlantic POM	135
5.3.3 Liquid chromatography instrumentation conditions.....	136
5.3.4 ICPMS instrumentation and conditions.....	137
5.3.5 ESIMS instrumentation and conditions.....	138
5.3.6 P-compounds characterization and putative annotation	139
5.3 RESULTS AND DISCUSSION	140
5.3.1 <i>Prochlorococcus</i> SB and MIT9313 POM	140
5.3.1.1 ICPMS data and interpretation	140
5.3.1.2 ESIMS data and interpretation	141
5.3.1.3 Conclusion on the <i>Prochlorococcus</i> POM samples	151
5.3.2 <i>Prochlorococcus</i> SB LMWDOM	152
5.3.2.1 ICPMS data and interpretation	152
5.3.4.2 ESIMS data and interpretation	153
5.3.4.3 Conclusion on <i>Prochlorococcus</i> SB LMWDOM	156
5.3.1 North Atlantic POM	156
5.3.1.1 ICPMS data and interpretation	156
5.3.1.2 Identification of P-containing compounds in F4	158
5.3.1.3 Identification of P-containing compounds in F6	161
5.3.1.4 Identification of P-containing compounds in F7	162
5.4 CONCLUSIONS	164
5.5 REFERENCES	166

APPENDIX 1. ETHYLENE CONCENTRATIONS AND SATURATION VALUES IN THE EASTERN MEDITERRANEAN SEA	169
A.1.1 INTRODUCTION.....	170
A.1.2 METHODS	170
A.1.2.1 Ethylene sampling and measurements	170
A.1.2.2 Ethylene saturation and air-sea gas exchange.....	170
A.1.3 RESULTS	171
A.1.4 DISCUSSION.....	173
A.1.5 CONCLUSION	175
A.1.6 REFERENCES.....	175
APPENDIX 2. ABSENCE OF PHOSPHONATE IN <i>Trichodesmium erythraeum</i> IMS101 AND <i>Prorocentrum minimum</i> (CCMP1329)	177
A2.1 INTRODUCTION.....	178
A2.2 MATERIAL AND METHODS	179
A2.2.1 <i>Trichodesmium erythraeum</i> sp. IMS101	179
A2.2.2 <i>Prorocentrum minimum</i>	179
A2.2.3 Nuclear magnetic resonance (NMR)	179
A2.3 RESULTS AND DISCUSSION	180
A2.3.1 Absence of phosphonate production in <i>Trichodesmium erythraeum</i> IMS101	180
A2.3.2 Absence of phosphonate production in <i>Prorocentrum minimum</i>	181
A2.4 REFERENCES.....	183
APPENDIX 3. EXTRACTION EFFICIENCY OF LOW MOLECULAR WEIGHT DISSOLVED ORGANIC PHOSPHORUS (LMWDOP) BY VARIOUS SOLID PHASE EXTRACTION (SPE) COLUMNS.....	185
A3.1 SUMMARY.....	186
A3.2 INTRODUCTION.....	186
A3.3 MATERIAL AND METHODS	187
A3.3.1 SPE columns	187
A3.3.2 Extraction of DOP from seawater.....	187
A3.3.3 UV oxidation	187
A3.3.4 SRP and TDP measurements.....	188
A3.4 RESULTS AND DISCUSSION	188
A3.4.1 North Atlantic seawater initial concentrations	188
A3.4.2 MilliQ blanks	188
A3.4.3 Sargasso seawater	189
A3.5 CONCLUSION	190

A3.6 REFERENCES.....	190
----------------------	-----

Tables of Figures

Figure 1.1: ^{31}P NMR spectrum of HMWDOM collected by ultrafiltration.....	18
Figure 2.1: Cruise track and stations sampled during PERLE 2 cruise.....	29
Figure 2.2: Vertical profiles of dissolved CH_4 concentrations.....	36
Figure 2.3: Vertical profiles of C-P lyase activity.....	37
Figure 2.4: ^{33}P radiochromatograms of the phosphonate fraction.....	40
Figure 3.1: Genomic potential for phosphonate production and consumption in nearly 13000 single-cell genomes from the tropical ocean	56
Figure 3.2: Phosphonate biosynthesis genes in surface ocean metagenomes.....	59
Figure 3.3: Phosphonate production by <i>Prochlorococcus</i> cultures	61
Figure 3.4: Distribution of phosphonate biosynthesis and utilization pathways in bacterial genomes.....	63
Figure 3.5: Microbial cycling of phosphonates in the upper ocean – an hypothesis.....	67
Figure S3.1: <i>Prochlorococcus</i> and SAR11 phosphonate producers	98
Figure S3.2: Phosphonate consumers and producers in 28 GORG-Tropics samples.....	99
Figure S3.3: Environmental factors driving the relative abundance of PepM in <i>Prochlorococcus</i> , SAR11, and Bacteria/Archaea	100
Figure S3.4: Time series of PepM relative abundance at the Hawaii Ocean Time-series (HOT) and the Bermuda Atlantic Time Series (BATS).....	101
Figure S3.5: Relationship between <i>Prochlorococcus</i> , SAR11, and combined Bacteria/Archaea PepM relative abundance and the top five biotic/abiotic covariates for each taxonomic group	102
Figure S3.6: Variations of <i>Prochlorococcus</i> SB P/C, Phn/C and Ph/C ratios between in exponential and stationary phase.....	103
Figure S3.7: Polysaccharide and phosphonate biosynthesis in <i>Prochlorococcus</i> and SAR11 isolate genomes.....	104
Figure S3.8: Distribution of phosphonate biosynthesis gene clusters within <i>Prochlorococcus</i> genomic islands	105
Figure S3.9: Empirical determination of similarity cutoffs for identifying PepM-derived metagenomic short sequencing reads.....	105
Figure 4.1: Chromatogram of the P standards mixture with various % of AcN.....	115
Figure 4.2: Chromatogram of the P standards mixture with various concentrations of AmFm.	116
Figure 4.3: Influence of the introduction of O_2 the P signal sensitivity.....	119
Figure 4.4: Influence of the dwell time on the $^{31}\text{P}^{16}\text{O}^+$ signal.....	120
Figure 4.5: Collision cell O_2 flow optimization for $^{31}\text{P}^{16}\text{O}^+$ signal	121
Figure 4.6: Changes in LOD with retention time.	122
Figure 4.7: Ion path of the Orbitrap Fusion.....	126

Figure 5.1: <i>Prochlorococcus</i> MIT9313 and SB POM sample ICPMS traces.....	141
Figure 5.2: AMP fractionation pattern.....	144
Figure 5.3: 5-phosphoarabinonic acid fragmentation pattern.....	147
Figure 5.4: D-Glucuronic-1-phosphate fragmentation pattern.....	148
Figure 5.5: Cytidine diphosphate fragmentation pattern.....	150
Figure 5.6: <i>Prochlorococcus</i> SB LMWDOM sample ICPMS trace.....	152
Figure 5.7: Chemical structures of the m/z fragments obtained from the mass m/z 249.0324 fragmentation pattern.....	155
Figure 5.8: Biofest Fractions 4, 6 and 7 ICPMS traces.....	157
Figure 5.9: 4-amino-1-hydroxy-butylphosphonic acid fragmentation pattern.....	158
Figure 5.10: 2,3-dihydroxypropyl-3-(methylamino)propyl hydrogen phosphate fragmentation pattern.....	159
Figure 5.11: F4 ICPMS and ESIMS traces.....	160
Figure 5.12: F6 ICPMS and ESIMS traces.....	162
Figure 5.13: (2-amino-1-hydroxyethyl)-(2,3-dihydroxypropoxy)phosphinic acid fragmentation pattern.....	163
Figure 5.14: F7 ICPMS and ESIMS traces.....	164
Figure A.1.1: Vertical profiles of dissolved C ₂ H ₄ concentrations.....	172
Figure A.2.1: ³¹ P NMR spectrum of soluble phosphorus after 2N KOH hydrolysis of <i>Trichodesmium erythraeum</i> IMS101 cells.....	180
Figure A.2.2: ³¹ P NMR spectrum of <i>Prorocentrum minimum</i> whole cells.....	181

Table of Tables

Table 2.1: Mixed layer depth (MLD), wind speed at 10 m above sea level, methane fluxes, production, excess and supersaturation values based on the MLD.....	35
Table 2.2: Pi concentrations, turnover rates and uptake rates.....	38
Table 2.3: Percentage of phosphate of uptake dedicated to the production of phosphonate (% Phn) and associated phosphonate production rates.....	39
Table 2.4: MLD methane production and calculated phosphonate degradation, C-P lyase activity and averaged uncorrected and Pi corrected phosphonate production rates all in pM d ⁻¹	44
Table S3.1: Model coefficients from beta-binomial regression of PepM abundance.....	94
Table S3.2: Functional enrichment of polysaccharide biosynthesis genes.....	95
Table S3.3: <i>Prochlorococcus</i> SB phosphonate (Phn) and phosphate (Ph) cellular proportions.....	96
Table S3.4: <i>Prochlorococcus</i> SB phosphorus (P), phosphonate (Phn) and phosphate (Ph) to carbon ratios.....	96
Table S3.5: Comparison of genome, PepM, and random tree topologies.....	97
Table 4.1: Influence of the mobile phase AcN content on the retention times of the P-standards.....	115
Table 4.2: Influence of the different mobile phase compositions on the sensitivity.....	116

Table 4.3: Influence of the mobile phase AmFm concentration on the retention times of the P-standards	117
Table 4.4: Summary of the signal/noise ratio obtained for the different tested parameters	122
Table 4.5: Electrospray parameters optimization.....	124
Table 4.6: Influence of Orbitrap resolution, injection times and mass range on the S/N of the different P standards.....	125
Table 5.1: Summary of the liquid chromatography conditions used for each sample analyzed	137
Table 5.2: ESIMS(-) source, Orbitrap and Ion Trap parameters adjusted for each samples ...	138
Table 5.3: <i>Prochlorococcus</i> MIT9313 and SB nucleotide ESIMS(-) data	143
Table 5.4: Putatively annotated compounds present in the <i>Prochlorococcus</i> MIT9313 and SB POM samples based on their mass and their fragmentation pattern after injection onto the ESIMS(-)	145
Table 5.5: Identification of compounds present in the <i>Prochlorococcus</i> SB LMWDOM sample based on their mass and their fragmentation pattern after injection onto the ESIMS(-)	153
Table 5.6: Retention times of the ICPMS peaks for the 3 fractions of interest	157
Table A.1.1: Ethylene fluxes, production, excess and supersaturation values based on the MLD and MLD methane:ethylene ratio.	173
Table A.3.1: Summary of SRP, TDP and calculated DOP concentrations of Sargasso seawater and blanks before and after the 4 different SPE columns.....	189
Table A.3.2: Summary of the amount of DOP retained by the 4 different SPE columns	189

CHAPTER 1. INTRODUCTION

1.1 MOTIVATION AND BRIEF OVERVIEW OF THE PHOSPHORUS CYCLE

Earth's climate is influenced in part by carbon dioxide (CO₂) atmospheric concentrations (Barnola et al., 1987) which are greatly buffered by the ocean and its biological carbon pump (Martin, 1990; Siegenthaler & Sarmiento, 1993). As marine photoautotrophs fix carbon, they drive a CO₂ flux from the atmosphere into the surface ocean. Most of this carbon is then remineralized by heterotrophs and can be respired, transferred to the dissolved organic carbon pool or sink. Upon sinking, organisms transfer carbon to deep water and sediments where it becomes sequestered. Therefore, the quantity of CO₂ removed from the atmosphere depends in part on the strength of the ocean's biological carbon pump. Thus, understanding the processes affecting the biological carbon pump is one key to appreciate how the ocean might drive and respond to climate change. Broadly, the strength and efficiency of the biological carbon pump is driven by physical (light, turbulences, temperature) and chemical (nutrient concentrations and bioavailability) forcings (Coles et al., 2017; Louca et al., 2018; Zakem et al., 2020). Physical and chemical forcings ultimately select for specific metabolic functionalities, which require different nutrients in various proportions. In response, microorganisms' metabolisms affect nutrient distributions and concentrations. Thus, if the nutrient distributions are ultimately shaped by global and abiotic processes happening on long time scales (decades to thousands years), biotic processes are responsible for changes on much shorter times (hours to weeks). Altogether, abiotic and biotic processes lead to complex and highly interconnected nutrient biogeochemical cycles that oceanographers seek to understand from a global to a molecular perspective.

P is essential for every living organism as it is part of genetic material (DNA and RNA), lipids, proteins, and metabolites involved in energy transfer (de Duve, 1991). Inorganic phosphate (Pi) is the most bioavailable form of P, and Pi concentrations are low and can reach undetectable levels (< 1 nM) in oligotrophic gyres which account for 40 % of the surface ocean (Polovina et al., 2008). In those regions, dissolved organic phosphorus (DOP) can constitute 80-95% of the total P pool and sustain over half of the primary production by acting as an alternative source of P for marine organisms with the appropriate hydrolytic enzymes (Karl & Björkman, 2015). However, very little is known about organic P. With the molybdenum blue method (Murphy & Riley, 1962), DOP can be quantified indirectly by difference between soluble reactive phosphorus (SRP) and total dissolved phosphorus (TDP). Although those results are extremely useful to visualize distributions on a global scale, the limits of detection prevent us from capturing finer scale dynamics in the surface ocean (Martiny et al., 2019). Moreover, it does not inform us of the various organophosphorus compounds present. Thanks to ³¹P nuclear magnetic resonance (NMR) we know that DOP, which can be further divided in high and low molecular weight DOP (HMWDOP and LMWDOP respectively), is mainly constituted of phosphate di- and mono-esters, poly-phosphates and phosphonates. HMWDOP is made of approximately 75% phosphate, 20% phosphonate and 5% of polyphosphate esters (Figure 1.1) (Kolowitz et al., 2001) but only few key compounds have been identified at a molecular scale (Repeta et al., 2016) and the rest remains unknown due to the lack of peak resolution. LMWDOP is even more

of an unknown as most techniques employed to isolate LMWDOM (mostly through SPE extraction) are not efficient for DOP compounds (Johnson, Kido Soule, & Kujawinski, 2017). However, thanks to the combination of reverse osmosis and electro dialysis a mixture of LMWDOM and HMWDOM was isolated and the ^{31}P NMR spectrum showed that this DOM fraction contains the same P functional groups even though phosphonate relative abundance is this time around 5-10% (Young & Ingall, 2010). The lack of suitable methods to identify DOP compounds on a molecular scale and investigate their biogeochemical cycling is a big obstacle in our understanding of DOP bioavailability and the overall P biogeochemical cycling.

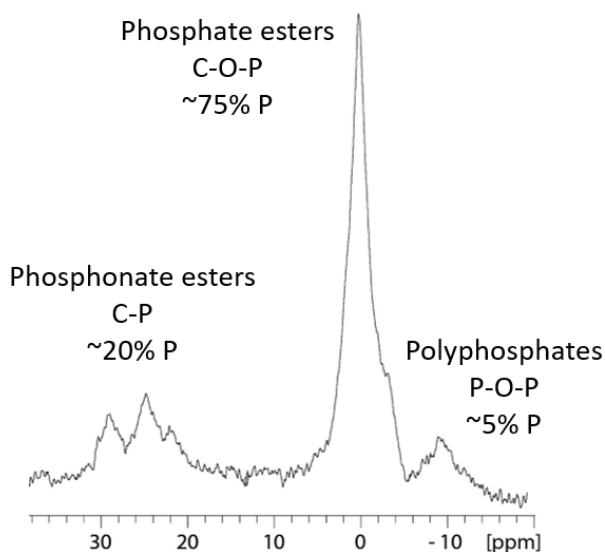


Figure 1.1: ^{31}P NMR spectrum of HMWDOM collected by ultrafiltration showing the distribution of different type of P functional groups. From Repeta *et al.* 2016.

In this thesis, I present my efforts to gain a better understanding of P biogeochemical cycling with a focus on phosphonate in oligotrophic regions. I aimed to take a holistic approach by studying organophosphorus on multiple scales in order to shed light on some key questions around P and phosphonates. First, I studied phosphonate production and consumption rates in the Eastern Mediterranean Sea using different field experiments in order to better understand the connection between phosphonates sources and sinks as well as the effect of P concentrations on those processes. This work is the focus of Chapter 2. In order to identify more phosphonates sources and investigate the cellular function of phosphonates, I studied *Prochlorococcus* as a model microbe and combined culture work with bioinformatics. This is the topic of Chapter 3. Finally, I developed a new liquid chromatography coupled to mass spectrometry method to enable organophosphorus characterization. I developed that method with the prospects of using it to better characterize the breadth of the organophosphorus compounds in marine samples and to

be able to start understanding phosphonates biogeochemical cycle, as well as their impact on the overall P cycle. This method along with the optimization work is presented in Chapters 4 and 5.

1.2 PHOSPHONATES

1.2.1 Phosphonate relevance in P biogeochemical cycling

In the ocean, an important fraction of P in HMWDOP (20-25%) is in the form of phosphonates (Kolowitz et al., 2001). Phosphonates are organophosphorus compounds with a C-P bond instead of the more common C-O-P bond found in phosphates. In phosphonates, the P atom is therefore in the +III oxidation state, which drastically changes the chemistry, conformation and biochemistry of those compounds. Indeed, phosphonates, due to their tetrahedral conformation, can mimic phosphates or carboxylic acids in the active sites of some enzymes and disrupt important metabolic reactions. It is for their potency that nature and humans have used phosphonates as antibiotics, herbicides, flame retardants, chemical warfare agents etc. since their discovery in the 1960s.

In oligotrophic gyres, where Pi is scarce, phosphonates constitute an important alternative source of P. Indeed, the C-P lyase pathway – one of the pathways known to hydrolyze the C-P bond of phosphonates - is part of the Pho regulon and is Pi inducible (Chen et al., 1990). Moreover, it has been shown that the relative abundance of C-P lyase genes was negatively correlated with Pi concentrations (Sosa et al., 2019) which emphasize the importance of phosphonate as a P source in the surface oligotrophic ocean. Phosphonate hydrolysis by the C-P lyase pathway has been of great interest as the hydrolysis of methylphosphonate, one of the most abundant phosphonates in HMWDOM, via this pathway leads to the production of methane (Repetta et al., 2016). This process causes an accumulation of methane in the oxic surface ocean - previously described as the “oceanic methane paradox” (Kiene, 1991; Reeburgh, 2007) - which leads to a flux of methane from the ocean to the atmosphere. Methylphosphonate consumption is responsible for ~ 30 to 50 % of methane emissions from the ocean to the atmosphere (Reeburgh, 2007; Repetta et al., 2016; Weber et al., 2019). In the context of increased anthropogenic forcing, it is thought that in the future, surface waters will become more stratified leading to an expansion of oligotrophic gyres (Polovina et al., 2008). Increased anthropogenic activities could also increase N input, either from anthropogenically fixed N₂ (Battye et al., 2017) or enhanced microbial N₂ fixation due to an increased supply of anthropogenic iron from combustion (Matsui et al., 2018). Altogether, this might cause a decrease in Pi concentrations and a consequent increase in DOP consumption to meet microbial P requirements. Thus, this could cause an increase in methylphosphonate consumption leading to a stronger methane flux to the atmosphere. Methane is a potent greenhouse gas with ~ 20-25 times the radiative forcing strength per molecule of CO₂ (Etminan et al., 2016; Myhre et al., 2013) and is involved in many atmospheric chemical reactions resulting in a positive radiative forcing (Lelieveld et al., 1998).

Therefore, an increase in methane fluxes from methylphosphonate consumption could cause positive feedback on global warming. It is therefore valuable to understand phosphonate cycling and for that, the sources of phosphonates and factors influencing phosphonate cycling need to be identified. To understand phosphonate cycling, phosphonates sources have to be investigated. All phosphonate biosynthetic pathways, with one exception, start with the conversion of phosphoenol pyruvate to phosphonopyruvate by the enzyme phosphoenolpyruvate mutase (PepM). Very few organisms have been experimentally confirmed as phosphonate producers and we cannot account for the large inventory of phosphonates in DOM (Dyhrman et al., 2009; Metcalf et al., 2012).

1.2.2 Study of phosphonate at a basin scale

Very little phosphonate has been detected in marine particulate organic phosphorus (Benitez-Nelson et al., 2004; Kolowitz et al., 2001; Paytan et al., 2003) which may be explained by the low rates of phosphonate production measured in the North Atlantic (Ebling et al., 2021; Van Mooy et al., 2015) and the apparent lack of phosphonate producers. However, it does not explain why phosphonate relative abundance is important in HMWDOM even in low Pi environment suggesting that phosphonate accumulates despite their role as an alternative P source. To make sense of this discrepancy, more information on which environmental parameters control phosphonate biogeochemical cycling is needed. To start understanding those parameters, we studied phosphonate cycling on a basin scale and report our findings in Chapter 2. This work was done in the ultra-oligotrophic Eastern Mediterranean Sea (EMS) where primary production is chronically P-limited (Krom et al., 2010) making this location a perfect end member to study the influence of P-limitation on phosphonate cycling. There, we measured rates of Pi uptake and turnover as well as phosphonate production rates using ^{33}Pi incubations. Simultaneously, we measured methane and ethylene concentrations to obtain air-sea fluxes in order to calculate phosphonate consumption rates. We also used a fluorescent C-P lyase assay to measure the activity of the C-P lyase pathway. As expected from the metagenomic results obtained by Sosa et al. (2019), we found high methane saturation and sea air flux indicative of high methylphosphonate via the C-P lyase pathway. Meanwhile, the fluorescent assay showed similar patterns to what has been observed in the North Pacific Subtropical Gyre (NPSG) (Granzow et al., 2021) with higher activity found at mid depth in the euphotic zone but unlike methane and ethylene air-sea fluxes, activities were lower in the EMS than in the NPSG. Phosphonate production rates were lower than in the North Atlantic Subtropical Gyre (Ebling et al., 2021) and positively correlated with Pi concentrations indicating that phosphonate production potential is lower in P-limited regions. Those production rates were shown to only sustain 25% of the phosphonate consumption via C-P lyase calculated with methane saturation implying that the phosphonate cycle is not at steady state and suggests that it is decoupled in time and/or space in the EMS. These data populate a very sparse database and provide the first indications of which

environmental parameters affect the phosphonate biogeochemical cycling and its influence on the overall P cycle.

1.2.3 Study of phosphonate at a cellular scale

In order to identify a model marine microbe to study the synthesis of phosphonates under different nutrient conditions and explore their biochemical function, we screened cultured microbes looking for one with phosphonate biosynthetic pathway. A previous genomic survey showed that this trait was widely distributed among various organisms (Yu et al., 2013). Among those, 18 % of the metagenomic reads were attributed to *Prochlorococcus* which is the most abundant photosynthetic organism inhabiting the surface of oligotrophic gyres (Yu et al., 2013). Therefore, we thought that *Prochlorococcus* could be a major phosphonate producer in the oligotrophic ocean which could explain in part the abundance of phosphonates in marine DOM. In Chapter 3, we show that we found one fully sequenced and culturable strain of *Prochlorococcus*, *Prochlorococcus* SB, with the full biosynthetic pathway. Using ^{31}P NMR, we confirmed that this strain produces phosphonates and allocates at least 40% of cellular P to phosphonate production. We looked at the influence of P supply on phosphonate production to test the hypothesis that phosphonates were used for P-storage, expecting their relative abundance to decrease if the P supply decreased. However, we found that their relative abundance increased when the cells became P-limited. We also found that those phosphonates are associated with proteins pointing to their potential functions in *Prochlorococcus*. The discovery of those phosphonoproteins was rather surprising as we expected to find the phosphonate associated with the sugars as they are in HMWDOM. Using genomic information, we found that phosphonate biosynthetic potential was a low abundance but widespread trait among marine microbes. In Chapter 3, we discuss the implications of these findings for phosphonate biogeochemical cycling.

1.2.4 Molecular level characterization of organic phosphorus in marine samples

P distributions are well described thanks to the molybdenum blue method combined with the different oxidation methods and P functional groups have been identified in the marine environment using ^{31}P NMR. However, those methods do not allow us to fully characterize the P pools and to study P-compounds bioavailability and cycling. Thus, our understanding of organophosphorus cycling is hindered by the lack of molecular methods. To address this issue, we developed a new molecular-level method using liquid chromatography coupled to mass spectrometry to characterize organophosphorus. To do so, we took an element targeted approach inspired from the approach used for siderophores discovery (Boiteau & Repeta, 2015) and used liquid chromatography (HPLC) coupled with inductively coupled plasma mass spectrometry (ICPMS) and electrospray ionization mass spectrometry (ESIMS). Because of its extremely high ionization, ICPMS fully atomizes organic molecules to their constituent elements allowing for an

element targeted, here P, approach to organic matter characterization. In contrast, ESIMS is a soft ionization technique that applies a charge to organic molecules with little fragmentation, enabling the detection of molecular ions. By integrating these two MS datasets, masses putatively corresponding to the P compounds can be extracted and their elemental formula determined. However, since P does not have a stable isotope, algorithms typically used to align ICPMS and ESIMS data cannot be used (Baumeister et al., 2018; Boiteau & Repeta, 2015; Durham et al., 2019), and we had to find alternative ways to integrate the data. Here we describe an approach using the detection of a product ion common to phosphoesters and phosphonoesters to find precursor ions of P-containing compounds in ESIMS. For those precursor ions, elemental formulae containing at least one P can be generated and the molecular fragmentation pattern used to identify organophosphorus compounds in marine samples. The method required optimization work specific to the detection of P in order to improve the sensitivity on each mass spectrometer using P-containing standards and this work is presented in Chapter 4. This method was then tested on marine samples –both from cultures and natural seawater- and the results are presented in Chapter 5. Using this method, we were able to identify organophosphorus compounds in two *Prochlorococcus* strains, *Prochlorococcus* SB and MIT9313, and in LMWDOM from the spent culture medium of *Prochlorococcus* SB. We also analyzed particulate matter isolated from the North Atlantic Ocean showing the potential for this method to be applied to environmental samples. We detected phosphite, the reduced form of Pi, in both the *Prochlorococcus* SB LMWDOM sample and the North Atlantic POM samples which is, to our knowledge, the first proof that phosphite is present in the marine environment and most likely serves as an alternative source of P.

1.3 REFERENCES

- Barnola, J. M., Raynaud, D., Korotkevich, Y. S., & Lorius, C. (1987). Vostok ice core provides 160,000-year record of atmospheric CO₂. *Nature*, 329(6138), 408–414.
- Battye, W., Aneja, V. P., & Schlesinger, W. H. (2017). Is nitrogen the next carbon? *Earth's Future*, 5(9), 894–904.
- Baumeister, T. U. H., Ueberschaar, N., Schmidt-Heck, W., Mohr, J. F., Deicke, M., Wichard, T., ... Pohnert, G. (2018). DeltaMS: a tool to track isotopologues in GC- and LC-MS data. *Metabolomics: Official Journal of the Metabolomic Society*, 14(4), 41.
- Benitez-Nelson, C. R., O'Neill, L., Kolowith, L. C., Pellechia, P., & Thunell, R. (2004). Phosphonates and particulate organic phosphorus cycling in an anoxic marine basin. *Limnology and Oceanography*, 49(5), 1593–1604.
- Boiteau, R. M., & Repeta, D. J. (2015). An extended siderophore suite from *Synechococcus* sp. PCC 7002 revealed by LC-ICPMS-ESIMS. *Metallomics: Integrated Biometal Science*, 7(5), 877–884.
- Chen, C. M., Ye, Q. Z., Zhu, Z. M., Wanner, B. L., & Walsh, C. T. (1990). Molecular biology of carbon-phosphorus bond cleavage. *The Journal of Biological Chemistry*, 265(8), 4461–4471.
- Coles, V. J., Stukel, M. R., Brooks, M. T., Burd, A., Crump, B. C., Moran, M. A., ... Hood, R. R. (2017). Ocean biogeochemistry modeled with emergent trait-based genomics. *Science*, 358(6367), 1149–1154.
- de Duve, C. (1991). *Blueprint for a Cell — The Nature and Origin of Life* (p. 275). Burlington, NC, USA: Neil Patterson Publishers.
- Durham, B. P., Boysen, A. K., Carlson, L. T., Groussman, R. D., Heal, K. R., Cain, K. R., ... Armbrust, E. V. (2019). Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. *Nature Microbiology*, 4(10), 1706–1715.
- Dyhrman, S. T., Benitez-Nelson, C. R., Orchard, E. D., Haley, S. T., & Pellechia, P. J. (2009). A microbial source of phosphonates in oligotrophic marine systems. *Nature Geoscience*, 2(10), 696–699.
- Ebling, A. E., Acker, M., Repeta, J. D., and Van Mooy B. A. S. (2021). Phosphorus Redox Cycling in the Western North Atlantic. In advanced prep *Limnology and Oceanography*.
- Etminan, M., Myhre, G., Highwood, E. J., & Shine, K. P. (2016). Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing: Greenhouse Gas Radiative Forcing. *Geophysical Research Letters*, 43(24), 12,614–12,623.
- Granzow, B. N., Sosa, O. A., Gonnelli, M., Santinelli, C., Karl, D. M., & Repeta, D. J. (2021). A sensitive fluorescent assay for measuring carbon-phosphorus lyase activity in aquatic systems. *Limnology and Oceanography, Methods / ASLO*, (lom3.10418). doi:10.1002/lom3.10418

- Johnson, W. M., Kido Soule, M. C., & Kujawinski, E. B. (2017). Extraction efficiency and quantification of dissolved metabolites in targeted marine metabolomics: Matrix effects in marine metabolomics. *Limnology and Oceanography, Methods / ASLO*, 15(4), 417–428.
- Karl, D. M., & Björkman, K. M. (2015). Chapter 5 - Dynamics of Dissolved Organic Phosphorus. In D. A. Hansell & C. A. Carlson (Eds.), *Biogeochemistry of Marine Dissolved Organic Matter (Second Edition)* (pp. 233–334). Boston: Academic Press.
- Kiene, R. P. (1991). Microbial Production and Consumption of Greenhouse Gases: Methane, Nitrogen Oxides, and Halomethanes, Edited by JE Rogers and WB Whitman, Library of Congress Cataloging-in-Publication Data, 298pp.
- Kolowitz, L. C., Ingall, E. D., & Benner, R. (2001). Composition and cycling of marine organic phosphorus. *Limnology and Oceanography*, 46(2), 309–320.
- Krom, M. D., Emeis, K. C., & Van Cappellen, P. (2010). Why is the Eastern Mediterranean phosphorus limited? *Progress in Oceanography*, 55(3–4), 236–244.
- Lelieveld, J., Crutzen, P. J., & Dentener, F. J. (1998). Changing concentration, lifetime and climate forcing of atmospheric methane. *Tellus. Series B, Chemical and Physical Meteorology*, 50(2), 128–150.
- Louca, S., Polz, M. F., Mazel, F., Albright, M. B. N., Huber, J. A., O’Connor, M. I., ... Parfrey, L. W. (2018). Function and functional redundancy in microbial systems. *Nature Ecology & Evolution*, 2(6), 936–943.
- Martin, J. H. (1990). Glacial-interglacial CO₂ change: The iron hypothesis. *Paleoceanography*, 5(1), 1–13.
- Martiny, A. C., Lomas, M. W., Fu, W., Boyd, P. W., Chen, Y.-L. L., Cutter, G. A., ... Moore, J. K. (2019). Biogeochemical controls of surface ocean phosphate. *Science Advances*, 5(8), eaax0341.
- Matsui, H., Mahowald, N. M., Moteki, N., Hamilton, D. S., Ohata, S., Yoshida, A., ... Flanner, M. G. (2018). Anthropogenic combustion iron as a complex climate forcer. *Nature Communications*, 9(1), 1593.
- Metcalf, W. W., Griffin, B. M., Cicchillo, R. M., Gao, J., Janga, S. C., Cooke, H. A., ... van der Donk, W. A. (2012). Synthesis of methylphosphonic acid by marine microbes: a source for methane in the aerobic ocean. *Science*, 337(6098), 1104–1107.
- Murphy, J., & Riley, J. P. (1962). A modified single solution method for the determination of phosphate in natural waters. *Analytica Chimica Acta*, 27, 31–36.
- Myhre, G., Samset, B. H., Schulz, M., Balkanski, Y., Bauer, S., Bernsten, T. K., ... Others. (2013). Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations. *Atmospheric Chemistry and Physics*, 13, 1853–1877.

- Paytan, A., Cade-Menun, B. J., McLaughlin, K., & Faul, K. L. (2003). Selective phosphorus regeneration of sinking marine particles: evidence from ^{31}P -NMR. *Marine Chemistry*, *82*(1), 55–70.
- Polovina, J. J., Howell, E. A., & Abecassis, M. (2008). Ocean's least productive waters are expanding. *Geophysical Research Letters*, *35*(3), 270.
- Reeburgh, W. S. (2007). Oceanic methane biogeochemistry. *Chemical Reviews*, *107*(2), 486–513.
- Repeta, D. J., Ferrón, S., Sosa, O. A., Johnson, C. G., Repeta, L. D., Acker, M., ... Karl, D. M. (2016). Marine methane paradox explained by bacterial degradation of dissolved organic matter. *Nature Geoscience*, *9*(12), 884–887.
- Sarmiento, J. L., & Gruber, N. (2006). *Ocean Biogeochemical Dynamics*. Princeton University Press.
- Siegenthaler, U., & Sarmiento, J. L. (1993). Atmospheric carbon dioxide and the ocean. *Nature*, *365*(6442), 119–125.
- Sosa, O. A., Repeta, D. J., DeLong, E. F., Ashkezari, M. D., & Karl, D. M. (2019). Phosphate-limited ocean regions select for bacterial populations enriched in the carbon-phosphorus lyase pathway for phosphonate degradation. *Environmental Microbiology*, *21*(7), 2402–2414.
- Van Mooy, B. A. S., Krupke, A., Dyhrman, S. T., Fredricks, H. F., Frischkorn, K. R., Ossolinski, J. E., ... Sylva, S. P. (2015). Phosphorus cycling. Major role of planktonic phosphate reduction in the marine phosphorus redox cycle. *Science*, *348*(6236), 783–785.
- Weber, T., Wiseman, N. A., & Kock, A. (2019). Global ocean methane emissions dominated by shallow coastal waters. *Nature Communications*, *10*(1), 4584.
- Young, C. L., & Ingall, E. D. (2010). Marine Dissolved Organic Phosphorus Composition: Insights from Samples Recovered Using Combined Electrodialysis/Reverse Osmosis. *Aquatic Geochemistry*, *16*(4), 563–574.
- Yu, X., Doroghazi, J. R., Janga, S. C., Zhang, J. K., Circello, B., Griffin, B. M., ... Metcalf, W. W. (2013). Diversity and abundance of phosphonate biosynthetic genes in nature. *Proceedings of the National Academy of Sciences*, *110*(51), 20759–20764.
- Zakem, E. J., Polz, M. F., & Follows, M. J. (2020). Redox-informed models of global biogeochemical cycles. *Nature Communications*, *11*(1), 1–10.

CHAPTER 2: PHOSPHONATE BIOGEOCHEMICAL CYCLING IN THE EASTERN MEDITERRANEAN SEA

2.1 INTRODUCTION

Phosphorus (P) plays an important role in cellular growth, structure and energy transfer making it an essential nutrient for all forms of life (de Duve, 1991). Inorganic phosphate (Pi), found as HPO_4^{2-} in seawater, is the most bioavailable form of P but its concentration can be extremely low and limiting for primary production in large areas of the ocean. In oligotrophic surface waters, dissolved organic phosphorus (DOP) is the most abundant form of P (Karl & Björkman, 2015). DOP is mainly composed of phosphate (C-O-P bond), and phosphonate (C-P bond but are esterified through a C-O-P esters (Kolowitz et al., 2001) as well as traces of phosphite esters (Repeta et al., 2016). To meet their P demand, marine microorganisms have evolved specific hydrolytic enzymes that release Pi from these distinct DOP pools (Dyhrman et al., 2007). The Eastern Mediterranean Sea (EMS) is an ultra-oligotrophic basin considered to be a primarily P-limited environment (Krom et al., 2010; Krom et al., 1991). Indeed, in the EMS, Pi concentrations are often below detection limits ($< 1\text{nM}$). DOP concentrations are relatively low in the EMS compared to other oligotrophic regions, but still constitutes $>95\%$ of the P pool (Karl & Björkman, 2015; Krom et al., 2005; Lomas et al., 2010). N:P ratios are high throughout the water column, and there is a very low N_2 fixation (between 0.1 and $0.4 \text{ nmol.L}^{-1}.\text{d}^{-1}$ in Spring-early Summer) (Bonnet et al., 2011; Ibello et al., 2010; Rahav et al., 2013). P-limitation responses have been shown experimentally by the enhancement of bacterial production upon the addition of Pi (Van Wambeke et al., 2002) and the observation of relatively high alkaline phosphatase activity (Duhamel et al., 2011; Mather et al., 2008; Van Wambeke et al., 2002). The EMS also has a high relative abundance of genomic markers for high affinity Pi transporters, phospholipid substitution (Pependorf et al., 2011), DOP transporters and hydrolytic enzymes involved in Pi acquisition such as *phoD*, *phoX*, C-P lyase but also *phnX* (Sosa et al., 2019) compared to other oligotrophic regions.

C-P lyase is a multiprotein complex that degrades a broad suite of alkyl-phosphonates in contrast with other phosphonate-specific hydrolytic pathways (White & Metcalf, 2007). C-P lyase is part of the Pho regulon (Chen et al., 1990), is Pi-controlled and is present in diverse marine microorganisms as a way to access Pi. Recently, it has been shown that the relative abundance of C-P lyase genes is negatively correlated with Pi concentrations as it increases from the North Pacific Subtropical gyre (NPSG) where Pi concentrations are on the order of hundreds nM to the North Atlantic Subtropical Gyre (NASG) where Pi concentrations are on the order of tens and the Mediterranean Sea (Sosa et al., 2019) where Pi concentrations are on the order of few nM. Hydrolysis of methylphosphonate (MPn) and 2-hydroxyethylphosphonate (2-HEP) from HMWDOP via C-P lyase, has been shown to produce methane (CH_4) and ethylene (C_2H_4) respectively under aerobic conditions helping to explain the supersaturation of those gases in surface waters previously described as the “marine methane paradox” (del Valle & Karl, 2014; Karl et al., 2008; Repeta et al., 2016). Therefore, methane and ethylene concentrations and their sea-air fluxes can be used to constrain rates of phosphonate consumption. Recently, a fluorescent phosphonate - 3-(5-(dimethylamino) naphthalene-1-sulfonamido)propylphosphonic acid (n-

DDPh) - has been chemically synthesized and can be used as a substrate to measure C-P lyase activity by quantifying the hydrolysis product 3-(5-(dimethylamino)naphthalene-1-sulfonamido)propane (n-DP) (Granzow et al., 2021).

If phosphonates are being consumed, they have to be produced to compensate for the loss by hydrolysis. So far, the only known first step common to all phosphonate biosynthetic pathways is the transformation of phosphoenolpyruvate to phosphonopyruvate catalyzed by the phosphoenolpyruvate mutase (PepM) enzyme (Horsman & Zechel, 2017; Seidel et al., 1988). PepM is present in low abundance in genomes from very diverse lineages of numerically abundant microbes like *Prochlorococcus* and SAR11. It has been shown experimentally that *Prochlorococcus* could be a major source of phosphonate in oligotrophic regions (Chapter 3). Phosphonates production rates have been measured in the NASG using incubations with ^{33}Pi . Results show a positive relationship between phosphonate production rates and Pi concentration implying that regions with high concentrations of Pi (for example along continental margins) might be a source of phosphonates whereas regions with low Pi concentrations will support very low if any phosphonate production (Ebling et al., 2021; Van Mooy et al., 2015).

By combining measurements of methane and ethylene saturation and air-sea fluxes with the fluorimetric assay to directly quantify C-P lyase activity along with ^{33}Pi incubations to assess phosphonate production rates in the EMS, we aimed to address two main questions. First, how do phosphonate production and consumption rates in the EMS compare to other oligotrophic regions? Because of the chronic P-limitation of the EMS, we expect higher phosphonate consumption rates as well as lower phosphonate production rates than in other oligotrophic regions. Second, is phosphonate cycling unbalanced in the EMS? Due to the expected low production and high consumption, we expect a decoupling of phosphonate production and consumption in surface waters.

2.2 MATERIAL AND METHODS

2.2.1 Cruise track and sampling

This study was conducted during the PERLE 2 cruise on board of the R/V Pourquoi Pas? from February 26th and March 16th 2019 in the Eastern Mediterranean Sea (Figure 2.1). The cruise left from Heraklion (Greece) and went around Crete starting northwest of the island in the Cretean Sea, then moving southwest to South East in the Lybian Sea, then back northeast to finish in Heraklion. A total of 98 stations were surveyed but our study revolved around 5 stations where diverse biogeochemical parameters were measured: Dagon 1 in the well mixed Cretean Sea, Dagon 2 and 3 in the more stratified Lybian Sea, Dagon 4 and Tarhun 8 in the most productive part of the study area on the edge of the cold Rhodos gyre. All samples were collected from 12 L Niskin bottles installed on a rosette equipped with a CTD (CTD SBE 911+) as well as an

Underwater Vision Profiler (UVP) and an Acoustic Doppler Current Profiler (ADCP). Because of water requirements, water for the analysis of methane and ethylene concentrations, methane stable carbon isotopes, and C-P lyase enzymatic activity was sampled from a cast made before water for ^{33}P incubations was collected, except at Tarhun 8. At the Dagon stations, water for methane and C-P lyase experiments was collected at night, while water for ^{33}P incubations was always recovered at 5:00 Eastern European Time (EET; local time). At Tarhun 8, all the water was sampled from the same cast which was back on deck at 17:00 EET.

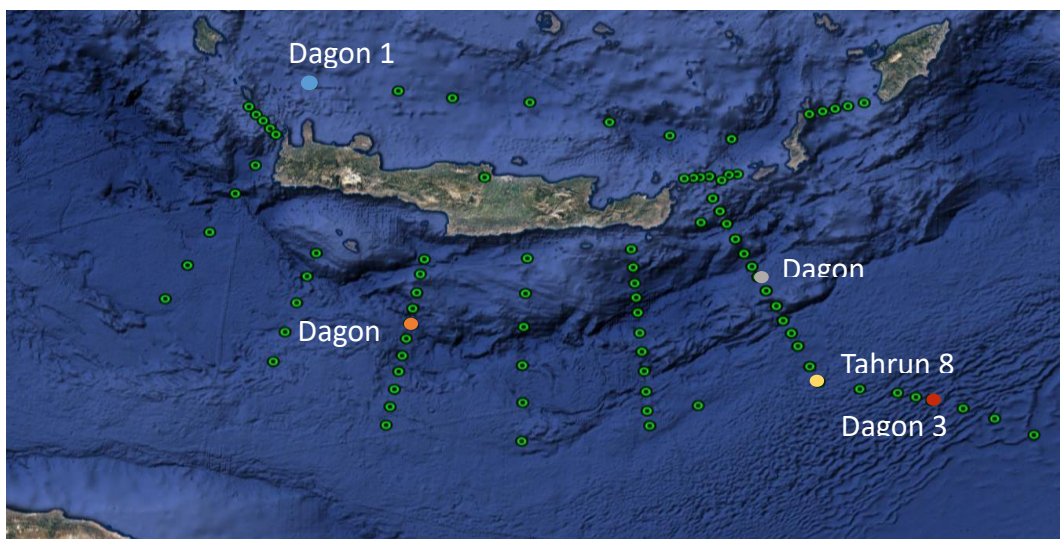


Figure 2.1: Cruise track (green dots) and stations sampled (colored dots) during PERLE 2 cruise.

2.2.2 Methane sampling and measurements

To measure dissolved methane concentrations as well as methane isotopes we used 70 mL and 500 mL combusted glass bottles respectively. We sampled directly from the Niskin bottles into the bottom of the serum bottles using silicon tubing after ensuring that no bubbles were trapped in the tube. We overflowed the bottles with twice their volumes and poisoned them with 0.1 mL of saturated mercuric chloride solution (7% w/v). We then sealed the bottles with polytetrafluoroethylene-lined stoppers and aluminum collars, inverted them a few times and stored them in the dark at ambient temperature until analysis in the lab. We collected samples at 11 depths throughout the water column with a higher resolution near the surface. Even though sampling depths varied between stations there were typically 5 m, 20 m, 30 m, 50 m, 75 m, 100 m, 175 m, 300 m, 500 m, 1000 m and a near bottom sample. We measured dissolved methane and ethylene concentrations using gas chromatography (Agilent 7980A; Agilent, USA) equipped with a flame ionization detector (FID), a gas stripping and a cryo-trap as previously described (Wilson et al., 2017). Briefly, the water samples were first under positive pressure supplied by ultra-high purity helium gas, purified downstream by passage through an additional liquid nitrogen cryo-trap, from the sample bottles to a purge chamber fitted with a porous frit. Then, the dissolved gases were stripped from ~70 mL seawater with ultrahigh purity helium for 10 min and

concentrated onto a sample loop (Porapak Q® trap; 80–100 sieve mesh size) cooled in liquid nitrogen. The sample loop was then heated allowing the gases to enter the gas chromatograph analytical column (30 m × 0.32 mm GS-CarbonPLOT capillary column; J&W Scientific) heated at 30°C for 15 minutes. We calibrated the FID daily by injecting different sized loops (0.1, 0.25, 0.35, 1.0 and 2.0 mL) of a methane standard containing $20.15 \pm 1\%$ ppmv of CH₄ in pure nitrogen gas (Scott-Marrin). All peaks corresponding to CH₄ were manually selected and integrated on the Agilent ChemStation software. The limits of detection and quantification (LODs and LOQs respectively) for methane were 8.3 pmol and 25.4 pmol. LODs and LOQs were calculated based on the root mean square error (RMSE) of the gas standard calibration, where $LOD = 3.29 \text{ RMSE}$, and $LOQ = 10 \text{ RMSE}$ (Bernal & Guo, 2014).

2.2.3 Methane and ethylene supersaturation and air-sea gas exchange

Seawater methane concentrations expected from air-sea equilibrium were calculated based on the Bunsen solubility model of Wiesenburg and Guinasso (1979) using a mean atmospheric CH₄ concentration of 1942.75 ppb as measured in December 2018 at the ENEA Station for Climate Observation (Lampedusa, Italy). Atmospheric CH₄ data were obtained from the National Oceanic and Atmospheric Administration Earth System Research Laboratory (NOAA-ESRL) Global Monitoring Division. The extent to which the measured methane concentrations in seawater (C_{meas}) deviated from the predicted air-sea equilibrium concentrations (C_{eq}) was expressed as the molar differences ($C_{\text{meas}} - C_{\text{eq}}$) or ΔCH_4 . This deviation was also expressed as the percent saturation level ($C_{\text{meas}} \times 100/C_{\text{eq}}$), where 100% corresponds to a null ΔCH_4 . Air-sea flux estimates (F) of methane were calculated with the equation:

$$F = k (C_{\text{meas}} - C_{\text{eq}})$$

where k is the gas transfer coefficient calculated using the gas exchange wind speed relationship by Wanninkhof (2014) and ($C_{\text{meas}} - C_{\text{eq}}$) is the average ΔCH_4 concentration in the surface mixed layer. The k coefficient was calculated using a gas-specific Schmidt number. A seawater CH₄ Schmidt number was calculated as described by Wanninkhof (2014). Wind speed from the ship's anemometer was scaled to 10 m above the sea surface. The average daily wind speed was used to estimate k . The depth of the surface mixed layer depth (MLD) was calculated using a temperature threshold criterion of -0.2°C from a reference depth of 10 m (d'Ortenzio et al., 2005). Positive air-sea fluxes indicated that the ocean is a net source of methane or ethylene to the atmosphere.

2.2.4 C-P lyase activity sampling and measurements

Seawater for C-P lyase activity was sampled on the same casts/depths as samples for gas concentrations. We sampled biological triplicates directly from the Niskin bottles using acid washed 125 mL polycarbonate bottles after rinsing three times. We then adjusted the volume in the bottles to a volume of 120 mL and we spiked 25 μL of 50 μM n-DPPH into each sample to a

final concentration of 10.4 nM of substrate. Finally, samples from depths above 300 m were put in deck incubators adjusted to the appropriate depth irradiances and constantly flushed with surface seawater, while samples from below 300 m were kept in dark incubators set on 16°C for depths between 300 and 500 m and 14°C for samples recovered from below 500 m. As we were sampling at night, we took precautions to keep samples in the dark until they went into the incubators. After 24h, we filtered the samples through 0.2 µm Sterivex filters to remove bacteria. The samples were then passed at a rate of 10 mL/min through SPE cartridges (Bond Elute ENV, 1g) previously activated with methanol. SPE cartridges are washed with 3 column volumes of MQ water to reduce salts and refrigerated until sample extraction. Back in the lab, we rinsed the columns with 10 mL of MQ water and then extracted the samples using 6 mL of methanol. Samples were dried under vacuum at 35°C for 5 h, resuspended into 50 µL of a 1:1 mixture of acetonitrile (AcN) in water then transferred into HPLC vials. To separate n-DDPh from the product n-DP, we used an HPLC (Agilent 1200 Series, Agilent, Santa Clara, CA) with a C-18 column (2.1 x 100 mm, 3 µm; Supelco Ascentis® C18) eluted with a linear gradient from 10% AcN:90% aqueous 20 mM Pi adjusted to pH of 4.2, to 85% AcN:15% aqueous 20 mM Pi over 14 min at a flow rate of 0.3 mL min⁻¹. Products were detected by a fluorescence detector set to $\lambda_{(ex)}$ 341 and $\lambda_{(em)}$ 528 nm. To control for the degradation of n-DP we incubated a sample from 10 m at Dagon 4 amended with 25 µL of 50 µM n-DP and a sample of filtered seawater amended with 25 µL of 50 µM n-DPPh. Those controls were processed the same way as the samples.

2.2.5 ³³P incubations sampling and treatments

To measure phosphonate production rates, we conducted incubations using radiolabeled (³³P) orthophosphoric acid (hereafter ³³Pi; American Radiolabeled Chemicals, Inc, USA) using samples recovered from 5 and 50 m following a protocol slightly modified from Ebling et al. (Ebling et al., 2021; Van Mooy et al., 2015). We used 1 L polycarbonate (PC) bottles to sample from the rosette. Then, we used this water to rinse six acid-washed 60 mL PC bottles and to fill them with 50 mL of seawater. Among the 6 bottles, 3 were spiked with 5 mL of a 10% formaldehyde solution in order to arrest any biological activity (Sohm & Capone, 2010). All samples were then spiked with 100 µL of a ³³Pi solution to a final activity of 25 µCi. In order to measure the specific activity in the bottles, we took 100 µL aliquots out of one biological replicate, put it into a 1.5 mL cryovial and stored it at -80°C. Finally, the 5 m and the 50 m samples were placed for 3 h and 4.5 h respectively in deck incubators flushed with surface seawater and shaded to mimic light at those depths. It should be noted that samples from Dagon 1 were lost because we spiked 5 µCi and incubated them for an hour. Consequently, not enough tracer made it to the cell to see a phosphonate signal. Therefore, for the rest of the stations, we spiked each samples as described above. After the incubation period, 5 mL from of each bottle were filtered on a 25 mm, 0.2 µm PC membrane filter (Millipore). These filters were reserved to measure phosphate uptake rates. The rest of the sample (45 mL) was filtered separately and reserved to measure phosphonate production rates. At the end of the filtration, each filter was

rinsed with ~10 mL of 0.2 μm filtered seawater, put into 1.5 mL cryovials and stored at -80°C until counting or further processing back in the lab.

2.2.6 ^{33}P phosphonate separation

Filters for phosphonate production rate measurements were treated with 1 mL of MQ and brought through 3 freeze (LN_2)-thaw (boiling water) cycles, each for few minutes in order to lyse the cells and release low molecular weight (LMW) phosphonates. Aliquots (700 μL) were transferred into new cryovials, dried under vacuum at 30°C overnight and resuspended into 400 μL MQ. To make sure that the sample was injected consistently during the IC, we added 25 μL of a 750 μM MPn solution and monitored the retention time and area of the MPn peak on the conductimeter. We manually injected three 100 μL aliquots of the sample serially on a preparative ion exchange chromatograph (Thermo Scientific Dionex ISC-2100) equipped with an IonPac AS18 column (Thermo Scientific Dionex) and a fraction collector. To separate phosphonates from phosphate we eluted the column at 1 mL min^{-1} with 23 mM aqueous KOH for 15 minutes followed by a linear gradient to 90 mM aqueous KOH over 10 min followed by a final wash for 10 min with 90 mM aqueous KOH. As phosphonates typically elute between 2 and 10 minutes (Ebling et al., 2021), the first 10 minutes of the run were collected and dried under vacuum at 60°C , then resuspended in 100 μL of MQ water and transferred it into an HPLC vial. To make sure that the sample was injected consistently during the HPLC run, we added 0.5 μL of a 6 mM adenosine monophosphate (AMP) solution and monitored the retention time and area of the AMP peak on the UV detector set to $\lambda=259\text{ nm}$ which correspond to the absorption maximum of AMP (Fischer, 1995). Phosphonates were separated using an HPLC (Agilent 1200, USA) equipped with two Primesep SB columns (250x2.1 mm, 5 μm particle size; SIELC Technologies, USA) in line, a UV detector and a fraction collector. Phosphonates were eluted isocratically at 0.2 mL min^{-1} with 10% AcN:90% aqueous ammonium formate (7 mM adjusted to $\text{pH} = 3$ with formic acid). We made serial, replicate 20 μL injections of the sample and collected fractions every two minutes. Fractions were transferred into scintillation vials, diluted with 10 mL of counting solution (Ultima Gold, Perkin Elmer) and counted each fractions using a liquid scintillation counter (LSC; Perkin Elmer). Separations of samples were alternated with separations of a corresponding killed controls. Filters collected to measure Pi uptake rates and total activities were transferred directly to scintillation vials, 10 mL of counting solution was added and the samples counted as described above.

2.2.7 ³³Pi uptake and turnover rates calculations

To calculate Pi uptake rates and Pi turnover, in situ Pi concentrations were measured on board following the protocols developed by Zhang and Chi (J.-Z. Zhang & Chi, 2002) with a detection limit of 2 nM and a quantification limit of 8 nM. Briefly, this method uses the same principle as the molybdenum blue method (Murphy & Riley, 1962) but here an autoanalyzer is linked to a 1 m liquid waveguide capillary cell (LWCC; World Precision Instruments, Sarasota, FL, USA). The LWCC was coupled to a spectrophotometer (USB 4000 VIS-NIR ; Ocean Optics, Dunedin, FL, USA) set to $\lambda = 710$ nm instead of the recommended $\lambda = 880$ nm due to the limitation of the LWCC (J.-Z. Zhang & Chi, 2002).

First, we subtracted the mean activity of the killed control uptake filters (A_{kill}) from the mean activity of the live sample uptake filters (A_{live}):

$$A_{real} = A_{live} - A_{kill}.$$

Then, we divided this A_{real} by the total activity (TA) added to the sample so we obtain the total Pi uptake ($F_{Pi\ uptake}$):

$$F_{Pi\ uptake} = A_{real}/TA.$$

We then calculated Pi turnover rates ($R_{Pi\ turnover}$ in h^{-1}) by dividing the total $F_{Pi\ uptake}$ by the incubation time ($t = 3$ h or 4.5 h):

$$R_{Pi\ turnover} = F_{Pi\ uptake}/t.$$

Finally, by multiplying the $R_{Pi\ turnover}$ by the measured in situ Pi concentrations ($[Pi]$), we calculated Pi uptake rates ($R_{Pi\ uptake}$ in $nM\ h^{-1}$):

$$R_{Pi\ uptake} = R_{Pi\ turnover} \times [Pi].$$

Errors are calculated using error propagation rules and the standard deviations between replicates.

2.2.8 ³³P phosphonate fraction and production rate calculations

To calculate the percentage of Pi uptake dedicated to the production of phosphonates (% Phn) and phosphonate production rates, samples were separated using IC followed by HPLC. In the killed control radiochromatograms, we obtained a peak with a retention time of 47 min which was attributed to ³³Pi (Chapter 5, Figures 5.12, 5.13 and 5.15). Live samples also have residual ³³Pi. Since the amount of residual ³³Pi varied between killed controls, we can assume this variation also occurs for live samples. However, activity at 47 min in the live samples could also be due to the presence of organic phosphorus compounds, so we do not know the value of residual ³³Pi in each sample. Therefore, we calculated a % Phn with no correction applied (% Phn uncorrected) and a % Phn for which activity at the retention time of Pi was excluded (% Phn; Pi corrected). This approach sets upper and lower boundaries on % Phn. To calculate the

%Phn we first calculated the background value (B) by averaging the ^{33}P activity in the killed controls ($A_{\text{kill fraction}}$) on each side of the Pi peak:

$B = \sum A_{\text{kill fraction}} / N_{\text{fractions}}$ where $N_{\text{fractions}}$ represents the number of fraction used for the integration.

We then subtracted the ^{33}P activity in all the live samples fractions ($A_{\text{live fraction}}$) by this background value:

$$A_{\text{real fraction}} = A_{\text{live fraction}} - B.$$

We then integrated all the ^{33}P fraction activities across the run ($A_{\text{Phn tot}}$) for the % Phn uncorrected value and excluded the signal between 45 min and 55 min corresponding of the elution of Pi (Figure 5, blue boxes) for the Pi corrected % Phn:

$$A_{\text{Phn tot}} = \sum A_{\text{real fraction}} / N_{\text{fractions}}.$$

Finally, we divided the integrated value by the total Pi uptake:

$$\% \text{ Phn} = A_{\text{Phn tot}} / F_{\text{Pi uptake}}.$$

We then multiplied % Phn by the Pi uptake rates to calculate the respective phosphonate production rate ($R_{\text{Phn prod}}$ in $\text{pmol L}^{-1} \text{ h}^{-1}$):

$$R_{\text{Phn prod}} = \% \text{ Phn} \times R_{\text{Pi uptake}}.$$

Standard errors for the % Phn were calculated using

$$\sigma = \sqrt{\frac{\sum (SD)^2}{N}}$$

where SD is the standard deviation between each replicate for each 2 minutes

fraction ($n = 2$ or 3) and N is the number of fractions used for the integration. Errors for the Phn production rates were calculated using error propagation rules using the errors of $F_{\text{Pi uptake}}$, $R_{\text{Pi uptake}}$ and % Phn.

Time between the incubations and the sample processing varied so we applied a decay correction. This was done calculating the number of days between the incubations and the radioactive counting and using a ^{33}P half-life of $t_{1/2} = 25.34$ d.

2.3 RESULTS

2.3.1 Dissolved methane concentrations and stable carbon isotope values

Methane was mostly oversaturated with respect to the atmosphere from surface waters to ~ 2500 m (Figure 2). Methane concentrations were between 2.5–5 nM which is similar to published values (Lamontagne et al., 1973) and gave methane saturation values ranging from 96% to 200%. Methane saturation values were generally highest in the upper 300–500 m and decreased with depth. At Dagon 1, methane saturation was the highest of all stations and remained relatively high throughout the upper 1000 m. At Dagon 3, we measured a high methane saturation near 2500 m suggesting a deep source of methane.

Across all sites sampled, the surface mixed layer had an average methane saturation value of 166% (SE = 3%). This corresponded to an average excess methane concentration (ΔCH_4) of 1.6 nM (SE = 0.1). This ΔCH_4 value yields an average methane air-sea flux of $22 \mu\text{mol m}^{-2} \text{ d}^{-1}$ (SE =

5 $\mu\text{mol m}^{-2} \text{d}^{-1}$) for this region. If we assume that the methane air-sea flux is balanced by autochthonous methane production, production values average 295 pM d^{-1} (SE = 113 pM d^{-1}) in the MLD (Table 2.1).

Table 2.1: Mixed layer depth (MLD), wind speed at 10 m above sea level, methane fluxes, production, excess and supersaturation values based on the MLD. Uncertainties correspond to the standard deviations between the values measured at each depth within the mixed layer. MLD is calculated as the depth where we measure a change in -0.2°C seawater temperature as preconized by d’Ortenzio et al. (2015).

Station	MLD (m)	Ship wind speed U_{10} (m s^{-1})	CH_4 flux ($\mu\text{mol m}^{-2} \text{d}^{-1}$)	CH_4 MLD production (pM d^{-1})	MLD excess CH_4 (nM)	MLD CH_4 saturation (%)
Dagon 1	302	19.2 \pm 2.7	38 \pm 6	126 \pm 18	1.9 \pm 0.3	177 \pm 11
Dagon 2	167	20 \pm 3.9	30 \pm 5	179 \pm 30	1.4 \pm 0.2	155 \pm 9
Dagon 3	61	8.7 \pm 2.9	6.9 \pm 0.00	113 \pm 0	1.7 \pm 0.00	168 \pm 0
Dagon 4	21	13 \pm 5.0	15.1 \pm 0.3	717 \pm 16	1.62 \pm 0.04	166 \pm 1
Tarhun 8	58	15.2 \pm 2.8	20 \pm 2	342 \pm 36	1.6 \pm 0.2	164 \pm 7

The $\delta^{13}\text{C}$ values of methane were measured at stations Dagon 2, Dagon 3, and Dagon 4 (Figure 2.2). At the three stations, the changes in $\delta^{13}\text{C}$ methane values with depth were similar. $\delta^{13}\text{C}$ values in the upper 250 m were lighter (-46.9 ‰ and -43.6 ‰), increased (became heavier) at mid-depths between 200–500 m, and decreased again below 500 m. At the deepest depths, Dagon 2 and 3 $\delta^{13}\text{CH}_4$ values were similar to the surface ones with $\delta^{13}\text{CH}_4 = -43.7$ ‰ and -45.8 ‰ respectively whereas Dagon 4 had lighter $\delta^{13}\text{CH}_4$ values ($\delta^{13}\text{CH}_4 = -39.4$ ‰; Figure 2).

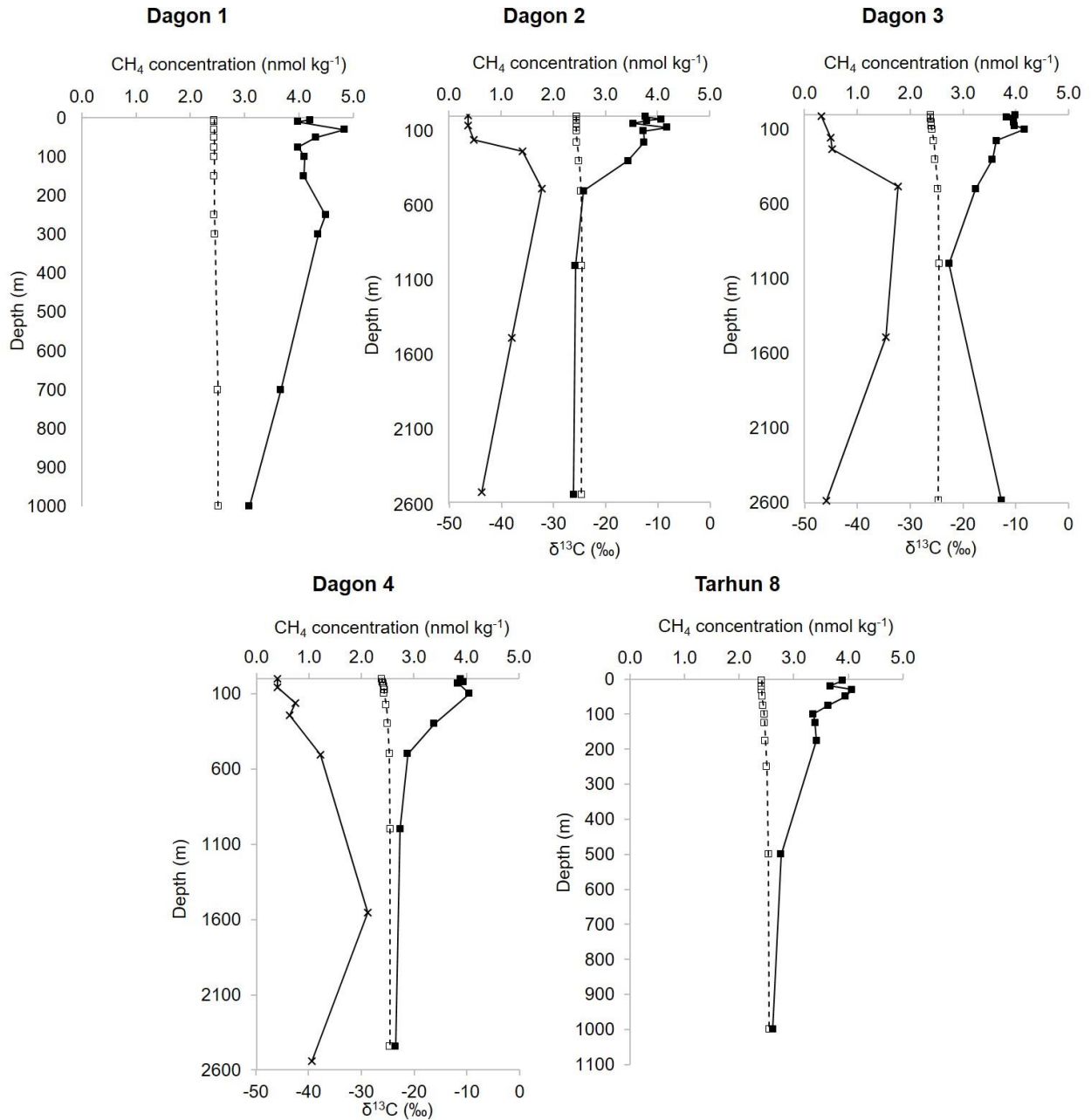


Figure 2.2: Vertical profiles of dissolved CH₄ concentrations measured in situ (black squares), and equilibrium values expected from a 1940 ppb atmosphere (empty squares) and CH₄ stable isotope values measured at Dagon 2-4 (crosses).

2.3.2 C-P lyase activity

The C-P lyase activity profiles in the euphotic zone at Dagon 1 and Tarhun 8 were both characterized by sharp changes with depth with a maximum value of 25 $\mu\text{M d}^{-1}$ at Dagon 1 (Figure 2.3 inset). Both stations also had a local maximum deeper in the water column which occurred at 250 m and 500 m at Dagon 1 and Tarhun 8 respectively. In contrast, C-P lyase activities at Dagon 2, 3 and 4 were more constant and varied between 2 $\mu\text{M d}^{-1}$ and 5 $\mu\text{M d}^{-1}$ through the euphotic zone (Figure 2.3 inset). Below 100-150 m, activities stayed low and constant between 3 $\mu\text{M d}^{-1}$ and 1.7 $\mu\text{M d}^{-1}$ (Figure 2.3). The very large error bars obtained at some depths and some stations are due to the leakage of some columns during the filtrations which caused a loss of sample in one of the replicate.

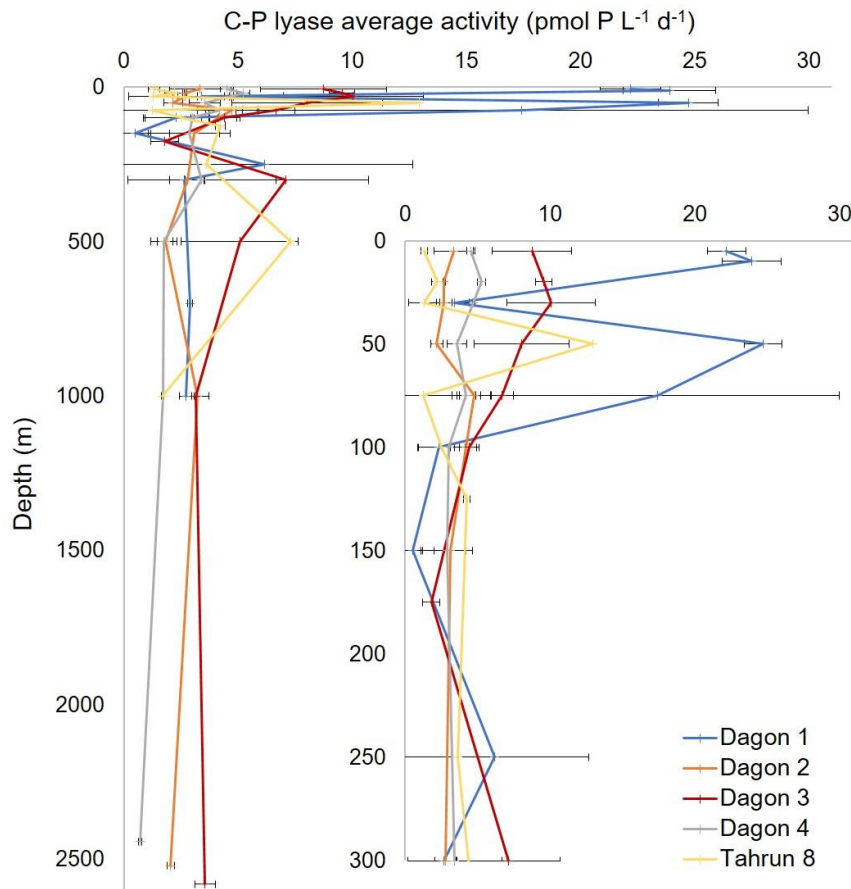


Figure 2.3: Vertical profiles of C-P lyase activity at the different stations and for the top 300 m (inset). Errors are calculated using the standard deviation between the triplicate C-P lyase activities

2.3.3 Pi concentrations, uptake and turnover rate

Pi concentrations in the upper 50 m were highest (22-25 nM) north of Crete (Dagon 1). At the other sampling sites, all south of Crete, Pi concentrations were lower (5-12 nM) (Table 2.2).

Pi turnover rates measured with ^{33}Pi were always higher in the surface than at 50 m and varied between 0.17 h^{-1} and 0.27 h^{-1} . At 50 m they varied less and were between 0.14 h^{-1} and 0.18 h^{-1} . The maximum turnover rates at 5 m and 50 m were both observed at Dagon 3, which is also the station with the lowest Pi concentrations. The minimum turnover rates were observed where Pi was relatively high i.e. 12 nM at 5 m at Tarhun 8 and 10 nM at 50 m at Dagon 4 (Table 2.2). The Pi uptake rates were also higher in the surface varying between 1.34 nM h^{-1} and 2.0 nM h^{-1} . At 50 m rates varied between 0.92 nM h^{-1} and 1.91 nM h^{-1} . Interestingly, in spite of having the highest turnover rates, the lowest Pi uptake rates were observed at Dagon 3 (Table 2.2).

Table 2.2: Pi concentrations, turnover rates and uptake rates

Station	Depth	Pi concentrations (nM)	Pi turnover rate (h^{-1})	Pi uptake rate (nM h^{-1})
Dagon 2	5	10	0.19 ± 0.01	1.9 ± 0.1
	50	11	0.173 ± 0.004	1.91 ± 0.05
Dagon 3	5	5	0.27 ± 0.02	1.34 ± 0.09
	50	5	0.184 ± 0.005	0.92 ± 0.03
Dagon 4	5	9	0.191 ± 0.003	1.72 ± 0.03
	50	10	0.14 ± 0.02	1.4 ± 0.2
Tarhun 8	5	12	0.169 ± 0.009	2.0 ± 0.1
	50	11	0.153 ± 0.003	1.68 ± 0.03

2.3.4 Phosphonate production rates

Phosphonate production rates were calculated by integrating the radiochromatograms after background correction normalized to Pi uptake. Due to the presence of ^{33}Pi in the chromatogram we calculated two phosphonate production values, one with a Pi correction and one without. These values place upper and lower boundaries on phosphonate production rates respectively. The corrections for Pi were variable and represented between 78% and 28% of the total ^{33}P activity of the samples.

Across all stations, phosphonates represent <1% of the Pi uptake. In general, the highest % Phn production and phosphonate production rates were measured in the surface where % Phn varied between 0.30% - 0.72% (uncorrected) and 0.22% - 0.40% (Pi corrected). At 50 m values were between 0.28% - 0.64% (uncorrected) and 0.08% - 0.35% (Pi corrected) (Table 2.3).

Table 2.3: Percentage of phosphate uptake dedicated to the production of phosphonate (% Phn) and associated phosphonate production rates.

Station	Depth (m)	% Phn uncorr.	% Phn Pi corr.	Phn production rate uncorr. (pM h ⁻¹)	Phn production rate Pi corr. (pM h ⁻¹)
Dagon 2	5	0.68 ± 0.02	0.34 ± 0.01	13 ± 1	6.5 ± 0.5
	50	0.61 ± 0.01	0.31 ± 0.01	9.8 ± 0.3	6.0 ± 0.2
Dagon 3	5	0.30 ± 0.01	0.166 ± 0.005	4.0 ± 0.3	2.2 ± 0.2
	50	0.486 ± 0.009	0.349 ± 0.007	4.5 ± 0.2	3.2 ± 0.1
Dagon 4	5	0.69 ± 0.02	0.40 ± 0.02	11.8 ± 0.4	6.9 ± 0.3
	50	0.28 ± 0.02	0.08 ± 0.01	3.8 ± 0.7	1.1 ± 0.3
Tarhun 8	5	0.72 ± 0.02	0.222 ± 0.008	14.5 ± 0.9	3.1 ± 0.3
	50	0.64*	0.14*	10.7 ± 0.2	2.37 ± 0.05

* n = 1

2.3.5 Phosphonate diversity across stations

The phosphonate fraction (0-10 min) from ion chromatography was further separated on Primesep SB columns. The resulting chromatograms all contained three distinct groups of labeled phosphonates. The first phosphonate group has a retention time between 5 and 15 min, the second group between 33 and 39 min and the third group between 39 and 53 min (Figure 2.4). The incorporation of AMP (detected using a UV detector) spike into the ³³P labeled samples, allowed for peaks in ³³P activity to be directly compared to the retention time of phosphonates in our standard mixture (Chapter 4). Consequently, we suspect that the first ³³P labeled fraction corresponds to N-containing phosphonates similarly to 2-AEPn or other weak bases not retained by our column (Chapters 4 and 5). The second group has a retention time similar to 2-HEP (~ 40 min) and might correspond to hydroxyalkylphosphonates. The third group elutes at retention times typical of MPn and ethylphosphonic acid (EtPn) which might mean that alkylphosphonates are present in the sample. As MPn and EtPn elute later than the peak in ³³P activity, it is likely that some extended hydroxyalkylphosphonates might be present as well. The peak from the third phosphonate group tails to 60 min in most samples, which indicates that a suite of hydroxyalkylphosphonates might be present in low concentrations. Interestingly, we did not find any significant peak matching the retention time of phosphite (64 min) which suggests that either phosphite is not a major component of reduced P in our POM samples or that due to matrix effects, the retention time of phosphite was shifted in these samples.

The relative abundance of each phosphonate group identified varied between stations and between depths especially for the first group of phosphonates. Tarhun 8 is the only station where the radiochromatograms do not change with depth (Figure 2.4).

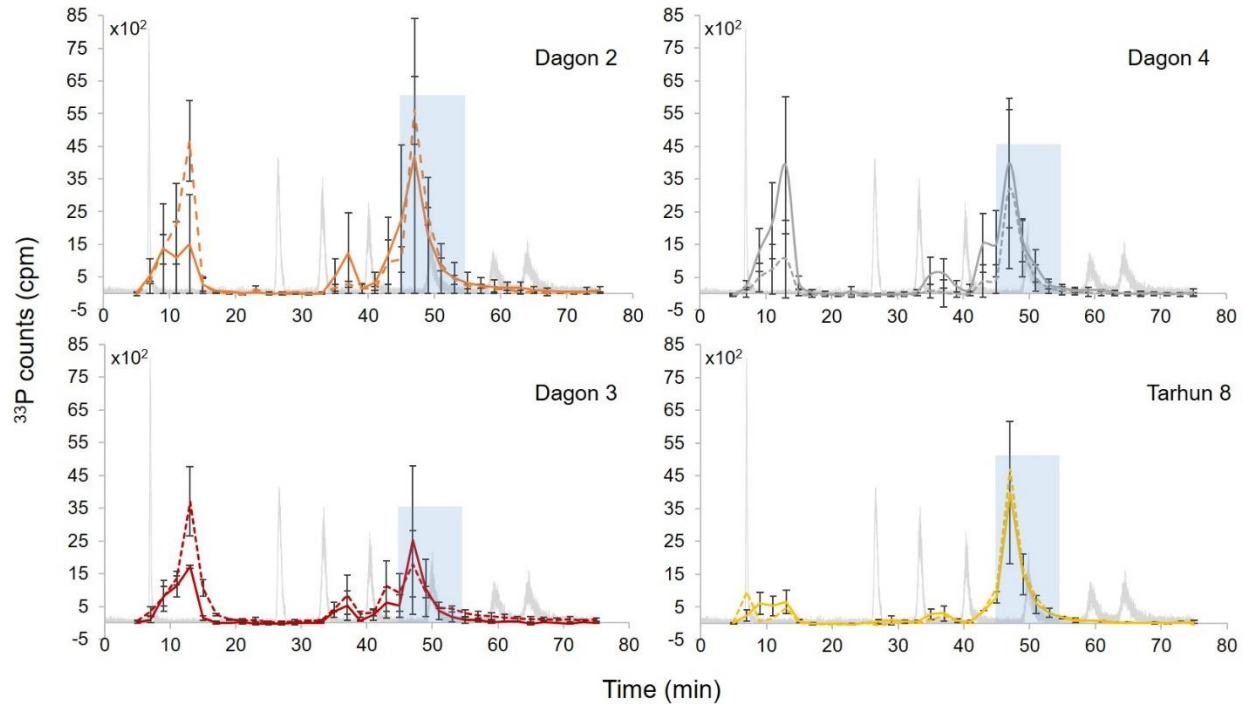


Figure 2.4: ^{33}P radiochromatograms of the phosphonate fraction previously isolated using ion chromatography for samples at 5 m (solid line) and 50 m (dashed line) at each station. The blue shaded areas highlight the 10 min of the radiochromatograms excluded when calculating the Pi corrected % Phn. The sum of the ^{33}P signal is used to calculate the % of Pi dedicated to the production and phosphonate and phosphonate production rates. The standard mixture ICPMS chromatogram is shown in light grey. Standards elute in the order of 2-aminoethylphosphonic acid (6.8 min), AMP (26.4 min), glucose-6-phosphate (33.0 min), 2-HEP (40.0 min), MPn (49.5 min), EtPn (59.3 min) and phosphite (64.1 min). Errors bars correspond to the standard deviation between the replicates (n= 2 or 3 and n= 1 for Tarhun 8 at 50 m).

2.4 DISCUSSION

2.4.1 Phosphonate consumption and production in the EMS compare to other oligotrophic areas

2.4.1.1 Phosphonate consumption

The EMS is a place where P chronically limits growth and where genes associated with P-stress response are highly abundant (Thingstad, 2005; Van Wambeke et al., 2002). Between the NPSG, NASG and the EMS, the EMS has the highest relative abundance of C-P lyase genes (Sosa et al., 2019). Although higher relative gene abundance does not necessarily translate into higher rates,

we expected higher rates of phosphonate consumption via the C-P lyase, and perhaps greater accumulation of methane and ethylene leading to a greater saturation in surface waters due to P-limitation.

During our cruise, average ΔCH_4 values in the mixed layer varied between 1.9 nM at Dagon 1 and 1.4 nM at Dagon 2 which correspond to saturation values of 177 % and 155 % respectively (mean value 166 %). Methane air-sea fluxes varied between 38 $\mu\text{mol m}^{-2} \text{d}^{-1}$ (Dagon 1) and 6.9 $\mu\text{mol m}^{-2} \text{d}^{-1}$ (Dagon 3) with a regional/cruise average of 21.7 $\mu\text{mol m}^{-2} \text{d}^{-1}$ (Table 2.1). Those results are much higher than typical values obtained at Station ALOHA (105% and 1.5 $\mu\text{mol m}^{-2} \text{d}^{-1}$; Bates et al., 1996; Tilbrook & Karl, 1995; Wilson et al., 2017) or in May in the Sargasso Sea where the highest CH_4 saturation measured was 119 % and the highest methane air-sea flux calculated 1.8 $\mu\text{mol m}^{-2} \text{d}^{-1}$ (Holmes et al., 2000; Sosa et al., 2020). Even at a similar wind speed, the EMS methane flux was 3-4 times higher than in the NASG. Moreover, we find a negative correlation between methane saturation values and P_i concentrations for the top 300 m ($R = -0.63$, $p < 0.01$). These results support the hypothesis that, if all the surface methane comes from MPn consumption via C-P lyase, there is more C-P lyase activity when P_i concentrations are lower and that there is a higher C-P lyase activity in the EMS than in the NASG and NPSG due to the increase in P-limitation.

C-P lyase activities in the EMS varied between 0.4 and 24.5 pM d^{-1} (mean value of 5.2 pM d^{-1}). There are few C-P lyase activity measurements made using n-DDPh and all were made at Station ALOHA where P_i concentrations are about an order of magnitude higher than EMS P_i concentrations. At Station ALOHA ($\text{P}_i \sim 100 \text{ nM}$), C-P lyase activities made in January between 5 and 300 m varied between 0.6 pM d^{-1} and 5.1 pM d^{-1} . Rates were higher in March ($\text{P}_i \sim 60 \text{ nM}$) and varied between 16.1 and 252.2 pM d^{-1} and higher still in November ($\text{P}_i \sim 30 \text{ nM}$) where they varied between 324.7 pM d^{-1} and 776.7 pM d^{-1} after the period of intense stratification (Granzow et al., 2021). Thus, our values are closer to the January conditions despite the much lower P_i concentrations and the higher relative abundance of C-P lyase genes in the EMS (Sosa et al., 2019). If we assume a negative correlation between P_i and C-P lyase activity due to the fact that C-P lyase is used to access P from phosphonates, is inhibited by P_i (Imazu et al., 1998; Repeta et al., 2016) and seems most active when P_i is low (Chen et al., 1990; Sosa et al., 2019), our C-P lyase data suggest that C-P lyase is less active in the EMS than at Station ALOHA for similar P_i concentrations.

2.4.1.2 Phosphonate production

Previous studies in the NASG have noted a positive correlation between phosphonate production rates and P_i uptake rates as well as the % Phn and P_i concentrations (Ebling et al., 2021; Van Mooy et al., 2015). Moreover, we noted an increase in the relative abundance of putative phosphonate producers with depth which may be due to an increase in P_i concentrations (Chapter 3). As the EMS is chronically P-limited even when P_i concentrations are relatively high (due to a

persistent N/P ratio >16), we hypothesized that phosphonate production should be less important in the EMS than in the NASG (Ebling et al., 2021; Van Mooy et al., 2015).

Across all stations, Pi concentrations were around 10 nM except for Dagon 3 that had concentrations of 5 nM (Table 2.2). From the ^{33}P incubations, we were able to calculate the Pi uptake rate and the Pi turnover rate for our samples. The Pi turnover rates varied between 0.14 h^{-1} (Dagon 4; 50m) and 0.27 h^{-1} (Dagon 3; 5m), which means that Pi residence times varied between 7.3 and 3.7 h respectively. Our results are in agreement with the EMS Pi residence time of 2-4 h measured in May with lower Pi concentrations (<2 nM) (Thingstad, 2005). Regardless of the station, residence times at 50 m were always higher than at 5 m as previously observed by others in the Sargasso Sea (Ebling et al., 2021; McLaughlin et al., 2013). Similar residence times were observed in the NASG where Pi concentrations were 2 to 5 times lower (Ebling et al., 2021). Pi uptake rates were from 0.92 and 2.0 nM h^{-1} , which was also similar to previous measurements in the EMS ($1.3\text{-}1.6\text{ nM h}^{-1}$; Thingstad, 2005).

Previous studies have assumed that ion chromatography separates phosphonates from phosphate and phosphate esters and used the first 10 minutes corresponding to the phosphonate fraction to calculate phosphonate production rates (Ebling et al., 2021; Van Mooy et al., 2015) and we followed this approach for our measurements just adding a correction for the presence of Pi in our runs (Chapter 5). However, based on Chapter 5, we now know that Pi and ester phosphate elutes in the first ten minutes of the ion chromatography run. Similarly, it is not unlikely that some phosphonates might elute after those ten minutes. Thus, we recognize that the estimates given hereafter might be an approximation of phosphonate production rates and that further work is needed to assess the veracity of our estimates. However, due to the consistency in the approach, we can still compare them with previous measurements. Regardless of the correction, the % of reduced Pi is always <1% of the total Pi uptake, with a surface average of 0.60% (uncorrected) and 0.28% (Pi corrected) and deep averages of 0.50% and 0.22% respectively. Those values are lower than the % of reduced Pi obtained in the NASG, which varied between 0.6% to 1.6% for Pi concentrations around 3 nM (Ebling et al., 2021). The averaged phosphonate production rates we calculated were 11 pM h^{-1} and 5 pM h^{-1} at 5 m and 7 pM h^{-1} and 3 pM h^{-1} at 50 m, whereas phosphonate production rates in the NASG often were above 10 pM h^{-1} . This implies that in the EMS, phosphonate production is lower either because there are fewer cells capable of producing phosphonates or that cells allocate less of P towards phosphonate production. Moreover, for the uncorrected values, we calculated a positive correlation between Pi uptake rates and % Phn ($R = 0.71$, $p = 0.05$) as well as phosphonate production rates ($R = 0.88$, $p < 0.01$). There is also a strong correlation between Pi concentrations and uncorrected % Phn ($R = 0.57$, $p = 0.14$) and phosphonate production rates ($R = 0.71$, $p = 0.05$). Those positive correlations are consistent with the hypothesis that microorganisms with access to higher Pi

concentrations are able to produce more phosphonate and suggest that low Pi environment limit phosphonate production and/or select against microorganisms that produce phosphonates.

2.4.2 Comparison between phosphonate consumption proxies

The use of a fluorescent phosphonate tracer specific to C-P lyase activity (Granzow et al., 2021) allowed us to make an independent measure of phosphonate degradation by the C-P lyase enzyme that can be compared to methane fluxes. When comparing the vertical profiles of dissolved methane concentrations and the C-P lyase activity, the proxies show similar trends: Dagon 1 is the station where C-P lyase activity and % saturation/fluxes are the highest and where the maxima are very prominent, Dagon 2 and Tarhun 8 have the lowest values in the surface, Dagon 4 does not vary much throughout the profile, and at Tarhun 8 the two maxima co-occur. Despite those similarities, the two datasets led us to a contradictory conclusion regarding of the C-P lyase activity in the EMS. Indeed, on one hand, the dissolved gas saturations indicate that C-P lyase is more active in the EMS than in other oligotrophic regions but on the other hand, the C-P lyase measurements using suggest lower rates of activity compare to the NPSG.

There are several possible explanations for the low activities measured for C-P lyase:

- 1) The amount of n-DDPh added as part of the assay was not sufficient to saturate the C-P lyase, and therefore measured rates were substrate limited. At Station ALOHA, in the NPSG, C-P lyase activities were measured using a concentration of n-DDPh of 5 nM, which was shown to be sufficient to saturate C-P lyase (Granzow et al., 2021). As we expected a higher C-P lyase activity in the EMS, we used a concentration of n-DDPh of 10 nM. However, we did not conduct kinetic experiments and it is likely that the substrate concentration used was not sufficient to saturate the system, yielding low C-P lyase activities.
- 2) The incubation time was too short. In the EMS, March is a transition period between winter and spring conditions and surface waters are warming up but not yet stratified. Microbial production is relatively low. Thus, it is possible that microbial production was slow and that, compare to methane (and ethylene) that can accumulate over a longer time period (~ 1 month), our incubation time of 24h was too short to capture a similar accumulation of n-DP in our incubations.
- 3) The sampling time matters. Water for our C-P lyase incubations was always sampled at night for the Dagon stations and late afternoon at the station Tarhun 8. These times may have not coincided with the times when phosphonate consumers are the most active. Interestingly, at Station ALOHA, the lowest C-P lyase activities measured so far were obtained with water collected at night and might not only be a reflection of the high Pi concentrations. Thus, n-DDPh incubations over the diel cycle are recommended.
- 4) n-DDPh hydrolysis is slower than MPn (or 2-HEP) via C-P lyase pathway causing a faster production of methane (or ethylene) than n-DP. This is very likely as n-DDPh is a complex molecule compare to MPn (or 2-HEP) which are some of the simplest

phosphonates. This probably influences enzyme affinities and efficiencies leading to lower apparent rates of phosphonate consumption by C-P lyase.

To confirm or refute those hypotheses, kinetic measurements should be conducted in order to determine the proper amount of n-DDPh to add. The effect of sampling time over the diel cycle should be investigated as well as the hydrolysis rates of MPn, 2-HEP compared to n-DDPh. Finally, as relative gene abundance and rates are two distinct metrics and are not necessarily correlated, we cannot rule out the possibility that our measured C-P lyase activities using n-DDPh are accurate and that the high saturation of methane and ethylene is caused by additional sources of methane in surface waters of the EMS and repeating these measurements in the EMS and elsewhere to compare those two approaches is necessary.

2.4.3 Comparison between phosphonate production and consumption via C-P lyase

To compare the rates of phosphonate production and consumption, we averaged phosphonate production rates using the values at 5 m and 50 m. Averaged phosphonate production rates are lower at Dagon 3 which is in agreement with the lowest Pi concentrations measured at that station. Maximum values are obtained at Tarhun 8 for the uncorrected rate and at Dagon 4 for the Pi corrected rate (Table 2.4).

Table 2.4: MLD methane production and calculated phosphonate degradation, C-P lyase activity and averaged uncorrected and Pi corrected phosphonate production rates all in $\mu\text{M d}^{-1}$. MLD methane production was calculated using the respective fluxes and the MLD (Table 2.1), MLD C-P lyase activity was calculated by averaging C-P lyase activities values (in $\mu\text{M d}^{-1}$) obtained in the ML and the phosphonate production rates (in $\mu\text{M d}^{-1}$) were calculated using the average between phosphonate production rates at 5 m and 50 m.

Station	CH ₄ production	Phn degradation (CH ₄)	C-P lyase activity	Phn production uncorr.	Phn production Pi corr.
Dagon 1	126 ± 18	378 ± 54	11 ± 5	N/A	N/A
Dagon 2	179 ± 30	537 ± 90	3.3 ± 0.8	274 ± 12	150 ± 6
Dagon 3	113 ± 0	339 ± 0	9 ± 2	102 ± 4	65 ± 3
Dagon 4	717 ± 16	2151 ± 48	4.9 ± 2	283 ± 10	166 ± 7
Tarhun 8	342 ± 36	1026 ± 108	4 ± 1	302 ± 11	66 ± 4

To calculate the phosphonate consumption rates using the methane air-sea flux data, we made the following assumptions: 1) all methane produced in the surface mixed layer is supported entirely by the degradation of MPn through the C-P lyase, 2) MPn is 1/3 of the phosphonate HMWDOM pool (Repeta et al., 2016) and 3) HMWDOM is the main source of phosphonate for the C-P lyase pathway. With these assumptions, we calculated MPn degradation rates between 113 and 717 $\mu\text{M d}^{-1}$ (Table 2.4) and total phosphonate consumption rates via C-P lyase between 0.34 and 2.15 nM d^{-1} . Using the C-P lyase assay, we measured phosphonate consumption rates in

the MLD between 11 and 3.3 $\mu\text{M d}^{-1}$. Those two estimates can be compared to the measured in situ phosphonate production rates and we find two different scenarios for the EMS at this time of year: using the methane values as a proxy, phosphonate production can only support about 25 % (12% for the Pi corrected value) of phosphonate consumption which supports the idea of a decoupling in time and/or space, and using the C-P lyase assay, only 3-5% of phosphonate consumption is needed to sustain phosphonate consumption suggesting that phosphonates might accumulate in the EMS at this time of year.

From our results and from previous studies (Ebling et al., 2021; Van Mooy et al., 2015) it seems that phosphonate production is positively correlated with Pi concentrations. Thus, phosphonates are more likely to be produced at depth, or in coastal waters. In contrast, phosphonate consumption via C-P lyase seems to be negatively correlated with Pi concentrations (Granzow et al., 2021; Sosa et al., 2019) which is supported by our methane measurements and the obtained negative correlation between methane saturation and Pi concentrations. This suggests that the EMS should be a place where phosphonate consumption is more intense than in other oligotrophic gyres. Therefore, we conclude that the EMS is a place of low phosphonate production and intense phosphonate consumption due to the chronic P-limitation and we hypothesize that this unbalance is even more pronounced in the summer when Pi concentrations drop below 1 nM. We speculate that most of the phosphonates consumed in the EMS were not produced locally and could have been transported to the EMS as it has been suggested for DOP in the North Atlantic (Reynolds et al., 2014) or that they have been produced when Pi concentrations were higher and the P demand lower i.e, in winter. This implies that the phosphonate cycle, in this region, is probably decoupled in time and/or space.

2.5 CONCLUSION

Our results are consistent with the hypothesis that MPn/2-HEP cycling is an important source of methane/ethylene in Pi depleted ocean regions as predicted by the enrichment of C-P lyase bacteria in the Mediterranean Sea (Sosa et al. 2019). The high values of methane and ethylene saturation measured in surface waters compared to other oligotrophic regions are consistent with the chronic P-limitation in the EMS and the idea of a more intense DOP – including phosphonates – remineralization by microorganisms. However, the current data and analysis cannot reject the possibility that other sources of methane and ethylene, in addition to MPn and 2-HEP hydrolysis, are present in the EMS and that phosphonate production and consumption are balanced. Additionally, our data indicate that phosphonate production in the EMS is lower than in the NASG and that phosphonate production seems to be positively influenced by Pi bioavailability.

It should be noted that C-P lyase activities using the fluorescent assay seemed to be abnormally low compared to the results obtained in the NPSG especially when considering the higher

methane and ethylene saturation values in the EMS compare to the NPSG. We hypothesize that these low values are driven by sampling or incubation conditions and recommend that those be investigated. We also recommend additional studies using simultaneous measurements of C-P lyase activity and Phn production rates to better understand how they relate to each other.

We propose that the elevated methane/ethylene oversaturation in the EMS corresponds to enhanced MPn/2-HEP degradation by C-P lyase. It is not clear if the calculated phosphonate production rates can support the measured methane flux. This could point to a spatial and/or temporal decoupling of phosphonate sources and sinks. The decoupling can be explained by the bacterial degradation of phosphonates that have locally accumulated in the DOM pool or transported from Pi rich waters into oligotrophic waters.

Our study provides the first ethylene profile as well as the first measurements for C-P lyase activity and phosphonate production rates in the EMS, a key region for phosphonate cycling due to the chronic P-limitation. Moreover, this is the first time that phosphonate production and consumption via C-P lyase have been simultaneously measured which raises some interesting questions on the coupling or decoupling of the phosphonate cycle in oligotrophic regions. More experiments are needed to confirm or refute some of our hypotheses.

2.6 REFERENCES

- Bates, T. S., Kelly, K. C., Johnson, J. E., & Gammon, R. H. (1996). A reevaluation of the open ocean source of methane to the atmosphere. *Journal of Geophysical Research*, 101(D3), 6953–6961.
- Bernal, E., & Guo, X. (2014). Limit of detection and limit of quantification determination in gas chromatography. *Advances in Gas Chromatography*, 3(1), 57–63.
- Bonnet, S., Grosso, O., & Moutin, T. (2011). Planktonic dinitrogen fixation along a longitudinal gradient across the Mediterranean Sea during the stratified period (BOUM cruise). *Biogeosciences*, 8(8), 2257–2267.
- Chen, C. M., Ye, Q. Z., Zhu, Z. M., Wanner, B. L., & Walsh, C. T. (1990). Molecular biology of carbon-phosphorus bond cleavage. *The Journal of Biological Chemistry*, 265(8), 4461–4471.
- d’Ortenzio, F., Iudicone, D., de Boyer Montegut, C., Testor, P., Antoine, D., Marullo, S., ... Madec, G. (2005). Seasonal variability of the mixed layer depth in the Mediterranean Sea as derived from in situ profiles. *Geophysical Research Letters*, 32(12).
- de Duve, C. (1991). *Blueprint for a Cell — The Nature and Origin of Life* (p. 275). Burlington, NC, USA: Neil Patterson Publishers.
- del Valle, D. A., & Karl, D. M. (2014). Aerobic production of methane from dissolved water-column methylphosphonate and sinking particles in the North Pacific Subtropical Gyre. *Aquatic Microbial Ecology: International Journal*, 73(2), 93–105.
- Duhamel, S., Björkman, K. M., Van Wambeke, F., Moutin, T., & Karl, D. M. (2011). Characterization of alkaline phosphatase activity in the North and South Pacific Subtropical Gyres: Implications for phosphorus cycling. *Limnology and Oceanography*, 56(4), 1244–1254.
- Dyhrman, S. T., Ammerman, J. W., & Van Mooy, B. A. S. (2007). Microbes and the Marine Phosphorus Cycle. *Oceanography*, 20(2), 110–116.
- Ebling, A.E., Acker, M., Repeta, D. J. and Van Mooy, B. A. S. (2021). Phosphorus Redox Cycling in the Western North Atlantic. *Limnology and Oceanography*.
- Fischer, J. H. (1995). Specific detection of nucleotides, creatine phosphate, and their derivatives from tissue samples in a simple, isocratic, recycling, low-volume system. 8, 254–264.
- Granzow, B. N., Sosa, O. A., Gonnelli, M., Santinelli, C., Karl, D. M., & Repeta, D. J. (2021). A sensitive fluorescent assay for measuring carbon-phosphorus lyase activity in aquatic systems. *Limnology and Oceanography, Methods / ASLO*, (lom3.10418). doi:10.1002/lom3.10418
- Holmes, M. E., Sansone, F. J., Rust, T. M., & Popp, B. N. (2000). Methane production, consumption, and air-sea exchange in the open ocean: An Evaluation based on carbon isotopic ratios. *Global Biogeochemical Cycles*, 14(1), 1–10.

- Horsman, G. P., & Zechel, D. L. (2017). Phosphonate Biochemistry. *Chemical Reviews*, 117(8), 5704–5783.
- Ibello, V., Cantoni, C., Cozzi, S., & Civitarese, G. (2010). First basin-wide experimental results on N₂fixation in the open Mediterranean Sea: NITROGEN FIXATION IN THE MEDITERRANEAN. *Geophysical Research Letters*, 37(3). doi:10.1029/2009gl041635
- Imazu, K., Tanaka, S., Kuroda, A., Anbe, Y., Kato, J., & Ohtake, H. (1998). Enhanced utilization of phosphonate and phosphite by *Klebsiella aerogenes*. *Applied and Environmental Microbiology*, 64(10), 3754–3758.
- Karl, D. M., Beversdorf, L., Björkman, K. M., Church, M. J., Martinez, A., & Delong, E. F. (2008). Aerobic production of methane in the sea. *Nature Geoscience*, 1(7), 473–478.
- Karl, D. M., & Björkman, K. M. (2015). Chapter 5 - Dynamics of Dissolved Organic Phosphorus. In D. A. Hansell & C. A. Carlson (Eds.), *Biogeochemistry of Marine Dissolved Organic Matter (Second Edition)* (pp. 233–334). Boston: Academic Press.
- Kolowitz, L. C., Ingall, E. D., & Benner, R. (2001). Composition and cycling of marine organic phosphorus. *Limnology and Oceanography*, 46(2), 309–320.
- Krom, M. D., Emeis, K. C., & Van Cappellen, P. (2010). Why is the Eastern Mediterranean phosphorus limited? *Progress in Oceanography*, 85(3–4), 236–244.
- Krom, M. D., Kress, N., Brenner, S., & Gordon, L. I. (1991). Phosphorus Limitation of Primary Productivity in the Eastern Mediterranean-Sea. *Limnology and Oceanography*, 36(3), 424–432.
- Krom, M. D., Woodward, E. M. S., Herut, B., Kress, N., Carbo, P., Mantoura, R. F. C., ... Zodiatis, G. (2005). Nutrient cycling in the south east Levantine basin of the eastern Mediterranean: Results from a phosphorus starved system. *Deep-Sea Research. Part II, Topical Studies in Oceanography*, 52(22–23), 2879–2896.
- Lamontagne, R. A., Swinnerton, J. W., Linnenbom, V. J., & Smith, W. D. (1973). Methane concentrations in various marine environments. *Journal of Geophysical Research*, 78(24), 5317–5324.
- Lomas, M. W., Burke, A. L., Lomas, D. A., Bell, D. W., Shen, C., Dyrman, S. T., & Ammerman, J. W. (2010). Sargasso Sea phosphorus biogeochemistry: an important role for dissolved organic phosphorus (DOP). *Biogeosciences*, 7, 695–710.
- Mather, R. L., Reynolds, S. E., Wolff, G. A., Williams, R. G., Torres-Valdes, S., Woodward, E. M. S., ... Achterberg, E. P. (2008). Phosphorus cycling in the North and South Atlantic Ocean subtropical gyres. *Nature Geoscience*, 1(7), 439–443.
- McLaughlin, K., Sohm, J. A., Cutter, G. A., Lomas, M. W., & Paytan, A. (2013). Phosphorus cycling in the Sargasso Sea: Investigation using the oxygen isotopic composition of phosphate,

enzyme-labeled fluorescence, and turnover times. *Global Biogeochemical Cycles*, 27(2), 375–387.

Murphy, J., & Riley, J. P. (1962). A modified single solution method for the determination of phosphate in natural waters. *Analytica Chimica Acta*, 27, 31–36.

Popendorf, K. J., Tanaka, T., Pujo-Pay, M., Lagaria, A., Courties, C., Conan, P., ... Van Mooy, B. A. S. (2011). Gradients in intact polar diacylglycerolipids across the Mediterranean Sea are related to phosphate availability. *Biogeosciences*, 8(12), 3733–3745.

Rahav, E., Herut, B., Levi, A., Mulholland, M. R., & Berman-Frank, I. (2013). Springtime contribution of dinitrogen fixation to primary production across the Mediterranean Sea. *Ocean Science*, 9(3), 489–498.

Repeta, D. J., Ferrón, S., Sosa, O. A., Johnson, C. G., Repeta, L. D., Acker, M., ... Karl, D. M. (2016). Marine methane paradox explained by bacterial degradation of dissolved organic matter. *Nature Geoscience*, 9(12), 884–887.

Reynolds, S., Mahaffey, C., Roussenov, V., & Williams, R. G. (2014). Evidence for production and lateral transport of dissolved organic phosphorus in the eastern subtropical North Atlantic. *Global Biogeochemical Cycles*, 28(8), 805–824.

Seidel, H. M., Freeman, S., Seto, H., & Knowles, J. R. (1988). Phosphonate biosynthesis: isolation of the enzyme responsible for the formation of a carbon-phosphorus bond. *Nature*, 335(6189), 457–458.

Sohm, J. A., & Capone, D. G. (2010). Zonal differences in phosphorus pools, turnover and deficiency across the tropical North Atlantic Ocean: NORTH ATLANTIC P POOLS AND CYCLING. *Global Biogeochemical Cycles*, 24(2). doi:10.1029/2008gb003414

Sosa, O. A., Burrell, T. J., Wilson, S. T., Foreman, R. K., Karl, D. M., & Repeta, D. J. (2020). Phosphonate cycling supports methane and ethylene supersaturation in the phosphate-depleted western North Atlantic Ocean. *Limnology and Oceanography*.

Sosa, O. A., Repeta, D. J., DeLong, E. F., Ashkezari, M. D., & Karl, D. M. (2019). Phosphate-limited ocean regions select for bacterial populations enriched in the carbon-phosphorus lyase pathway for phosphonate degradation. *Environmental Microbiology*, 21(7), 2402–2414.

Thingstad, T. F. et al. (2005). Nature of phosphorus limitation in the ultraoligotrophic eastern Mediterranean. *Science*, 309(2005), 1068–1071.

Tilbrook, B. D., & Karl, D. M. (1995). Methane sources, distributions and sinks from California coastal waters to the oligotrophic North Pacific gyre. *Marine Chemistry*, 49(1), 51–64.

Van Mooy, B. A. S., Krupke, A., Dyhrman, S. T., Fredricks, H. F., Frischkorn, K. R., Ossolinski, J. E., ... Sylva, S. P. (2015). Phosphorus cycling. Major role of planktonic phosphate reduction in the marine phosphorus redox cycle. *Science*, 348(6236), 783–785.

- Van Wambeke, F., Christaki, U., Giannakourou, A., Moutin, T., & Souvemerzoglou, K. (2002). Longitudinal and vertical trends of bacterial limitation by phosphorus and carbon in the Mediterranean Sea. *Microbial Ecology*, 43(1), 119–133.
- Wanninkhof, R. (2014). Relationship between wind speed and gas exchange over the ocean revisited: Gas exchange and wind speed over the ocean. *Limnology and Oceanography, Methods / ASLO*, 12(6), 351–362.
- White, A. K., & Metcalf, W. W. (2007). Microbial Metabolism of Reduced Phosphorus Compounds. *Annual Review of Microbiology*, 61(1), 379–400.
- Wiesenburg, D. A., & Guinasso, N. L. (1979). Equilibrium solubilities of methane, carbon monoxide, and hydrogen in water and sea water. *Journal of Chemical and Engineering Data*, 24(4), 356–360.
- Wilson, S. T., Ferrón, S., & Karl, D. M. (2017). Interannual variability of methane and nitrous oxide in the north pacific subtropical gyre: Methane and nitrous oxide in the NPSG. *Geophysical Research Letters*, 44(19), 9885–9892.
- Zhang, J.-Z., & Chi, J. (2002). Automated analysis of nanomolar concentrations of phosphate in natural waters with liquid waveguide. *Environmental Science & Technology*, 36(5), 1048–1053.

CHAPTER 3. PHOSPHONATE PRODUCTION BY MARINE MICROBES: SOURCES AND POTENTIAL FUNCTION

This paper has been submitted to the bioRxiv preprint server:
<https://www.biorxiv.org/content/10.1101/2020.11.04.368217v1>

Marianne Acker^{1,2,†}, Shane L. Hogle^{3,4†}, Paul M. Berube³, Thomas Hackl³, Ramunas Stepanauskas⁶, Sallie W. Chisholm^{3,5} and Daniel J. Repeta^{2,*}

¹ MIT-WHOI Joint Program in Oceanography/Applied Ocean Science and Engineering, Woods Hole, Massachusetts, USA

² Department of Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

³ Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁴ Department of Biology, University of Turku, Turku 20500, Finland

⁵ Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02139, USA

⁶ Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine, 04544, USA

† These authors contributed equally to this work

Corresponding author : Correspondence and requests should be addressed to DJR
(drepeta@whoi.edu)

3.1 PREFACE

This work is the results of an extensive and close collaboration with Shane L. Hogle who performed and interpreted all the bioinformatic analyses. My intellectual, experimental and editorial contribution are specified in details hereafter.

Experimentally, I grew *Prochlorococcus* SB and MIT9301 cultures to assess whether the presence/absence of the phosphonate biosynthetic pathway in those strains genomes caused the presence/absence of phosphonate in the cells respectively. I also separated the macromolecular fractions of *Prochlorococcus* SB to prove that phosphonates were associated with the protein fraction in this strain. As an additional proof, I purified *Prochlorococcus* SB following a different protocol. For each of the cultures and the experiments conducted after, I acquired the ³¹P-NMR spectra to monitor the presence of phosphonates and interpreted them. As a result of those analyses, I wrote the paragraphs about phosphonates in *Prochlorococcus* cultures (3.4.3), phosphonates as a phosphorus storage (3.4.4), the macromolecular form of phosphonate in *Prochlorococcus* SB (3.4.5) the concluding paragraph (3.5) as well as extensive parts of the abstract, the introduction, and the potential role of phosphonate associated with proteins (3.4.6). I also edited the remaining paragraphs. Additionally, I designed Figures 3.4 and 3.6, the Supplementary Figure S3.6 and generated Supplementary Tables S3.3 and S3.4. As a lead on this project, I did not only contribute intellectually to the experiments listed above but also suggested that we look, using bioinformatics, at 1) the relationships between the abundance of phosphonate producers and environmental parameters (within the same location or between different ocean basins) 2) genes surrounding the phosphonate biosynthetic pathway in the *Prochlorococcus* SB genome to get more context and be able to speculate about phosphonate function in this strain as well as link those information with phosphonate macromolecular form.

3.2 ABSTRACT

Phosphonates, organic compounds with a C-P bond, constitute 20-25% of phosphorus in high molecular weight dissolved organic matter and are a significant phosphorus source for marine microbes. However, little is known about phosphonate sources, biological function, or biogeochemical cycling. Here, we determine the biogeographic distribution and prevalence of phosphonate biosynthesis potential using thousands of genomes and metagenomes from the upper 250 meters of the global ocean. Potential phosphonate producers are taxonomically diverse, occur in widely distributed and abundant marine lineages (including SAR11 and *Prochlorococcus*) and their abundance increases with depth. Within those lineages, phosphonate biosynthesis and catabolism pathways are mutually exclusive, indicating functional niche partitioning of organic phosphorus cycling in the marine microbiome. Surprisingly, one strain of *Prochlorococcus* (SB) can allocate more than 40% of its cellular P-quota towards phosphonate production. Chemical analyses and genomic evidence suggest that phosphonates in this strain are incorporated into surface layer glycoproteins that may act to reduce mortality from grazing or viral infection. Although phosphonate production is a low-frequency trait in *Prochlorococcus*

populations (~ 5% of genomes), experimentally derived production rates suggest that *Prochlorococcus* could produce a significant fraction of the total phosphonate in the oligotrophic surface ocean. These results underscore the global biogeochemical impact of even relatively rare functional traits in abundant groups like *Prochlorococcus* and SAR11.

3.3 INTRODUCTION

In nutrient-impoverished mid-ocean gyres, microbial demand for phosphorus (P) is often so high that concentrations of inorganic phosphate are drawn down to sub nanomolar levels. Under these conditions, up to half of microbial P demand is met through the uptake and metabolism of P-containing dissolved organic matter (Karl & Björkman, 2015). Dissolved organic phosphorus (DOP) is a complex, poorly characterized mixture of high and low molecular weight (HMW and LMW) phosphate and phosphonate esters. Phosphate esters are common in nucleic acids and lipids, and are synthesized by all marine microbes. The presence of phosphate esters in DOP is easily explained. In contrast, phosphonates, reduced P compounds with a stable, covalent C-P bond (Horsman & Zechel, 2017), are a poorly-understood component of marine DOP, but nevertheless constitute 20-25% of the P in HMWDOP (Kolowitz et al., 2001).

All phosphonate production pathways are initially catalyzed through the same steps involving the enzymes PepM, Ppd, and Pdh (Horsman & Zechel, 2017) (Figure 3.1A). Despite these shared catalytic steps, the chemical diversity of phosphonates is extensive and includes small bioactive metabolites as well as macromolecules such as lipids, polysaccharides, and proteins. The functional roles and macromolecular forms of phosphonates in marine microbes are unknown. In contrast, phosphonate degradation in the marine environment is better understood (Martinez et al., 2010; Repeta et al., 2016; Sosa et al., 2019a) (Figure 3.1B), and is correlated with phosphate availability (Coleman & Chisholm, 2010; Feingersch et al., 2012; Sosa et al., 2019b).

Despite the high abundance of phosphonates in DOP, only two bacterioplankton species (members of bacteria and archaea) have been experimentally confirmed as phosphonate producers: *Trichodesmium erythraeum*, a nitrogen-fixing cyanobacterium (Dyhrman et al., 2009), and *Nitrosopumilus maritimus* from the Marine Group I (MGI) Thaumarchaeota (Metcalf et al., 2012). *Trichodesmium* has a relatively restricted geographic range (Breitbarth et al., 2007) and global abundance, while MGI Thaumarchaeota are abundant in the mesopelagic, but comparatively rare in sunlit surface waters (Santoro et al., 2019). Moreover, *Candidatus Nitrosopelagicus brevis*, the only characterized pelagic representative of MGI Thaumarchaeota, lacks phosphonate biosynthesis genes and is significantly more abundant than *Nitrosopumilus* in the open ocean (Santoro et al., 2015). It is unlikely that phosphonate production by *Trichodesmium* and Thaumarchaeota alone is enough to support the large and ubiquitous inventory of phosphonates observed in the ocean euphotic zone. Metagenomic surveys of PepM have estimated that between 8-16% of all marine microbes in the surface ocean may be capable

of producing phosphonates (Metcalf et al., 2012; X. Yu et al., 2013), but the taxonomic composition and ecological niches of phosphonate producers remain unclear.

Much about the biology, ecology, and biogeochemistry of marine phosphonates remains to be discovered. Recent expansions of marine genomic data (Berube et al., 2018; Biller et al., 2018; Karsenti et al., 2011; Pachiadaki et al., 2019), as well as advances in chemical analyses, makes it possible to systematically investigate which organisms are producing the enigmatic phosphonate pool in the oceans. Here we combine laboratory studies and chemical analyses with comparative genomic and metagenomic analyses to investigate the prevalence, taxonomic distribution, and potential function of phosphonate biosynthesis in the surface ocean. We find that phosphonate biosynthesis genes are found in a wide variety of marine microbes including members of the two most abundant groups in the surface ocean: *Prochlorococcus* and SAR11. We experimentally demonstrate that a *Prochlorococcus* strain produces phosphonates almost exclusively in the HMW protein fraction and, surprisingly, these phosphonoproteins account for over 40% of total cellular phosphorus.

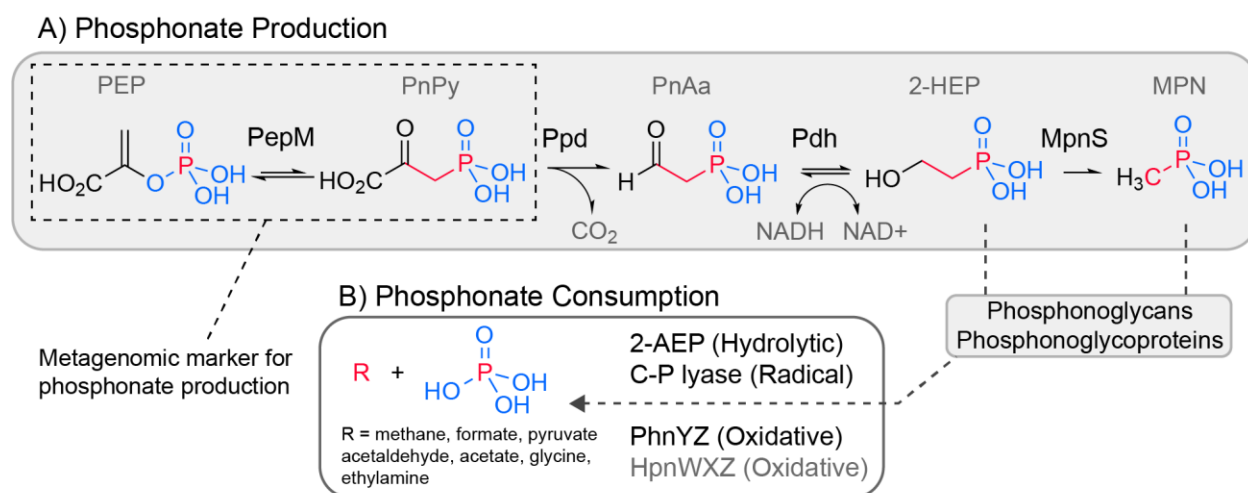


Figure 3.1: Phosphonates production and consumption in the marine environment

A) Phosphonate biosynthesis includes three key steps; (1) isomerization of phosphoenolpyruvate (PEP) to phosphonopyruvate (PnPy) via phosphoenolpyruvate mutase (PepM), (2) decarboxylation of PnPy to phosphonoacetaldehyde (PnAa) via phosphonopyruvate decarboxylase (Ppd), and (3) dehydrogenation of PnAa to 2-hydroxyethylphosphonic acid (2-HEP) via phosphonoacetaldehyde dehydrogenase (Pdh). 2-HEP can be further transformed by methylphosphonate synthase (MpnS), to produce methylphosphonate (MPN), a substrate for aerobic marine methane production (Repeta et al., 2016). Phosphonates in the marine environment are likely used to decorate macromolecules like glycans and glycoproteins (Horsman & Zechel, 2017; Metcalf et al., 2012; Rice et al., 2019). B) Phosphonate degradation via cleavage of the C-P bond proceeds through at least three mechanisms: hydrolytic (2AEP, named for the representative 2-aminoethylphosphate degradation pathway via PhnWX), radical (C-P lyase), and oxidative (PhnYZ). The proposed HpnWXZ pathway is likely oxidative proceeding via the oxidative deamination of an aminophosphonate similar in structure to 2-aminoethylphosphate (Lahiri et al., 2006). We find the HpnWXZ pathway to be rare in marine genomes.

3.4 RESULTS AND DISCUSSION

3.4.1 The taxonomic distribution of phosphonate producers and consumers

How widespread is the ability to produce and consume phosphonates among diverse marine microbes? Prior efforts to address this question have relied on mapping metagenomic reads to key marker genes such as the PepM found in phosphonate producers (X. Yu et al., 2013) or the C-P lyase multi-enzymatic cluster and other catabolic pathway proteins found in phosphonate consumers (Feingersch et al., 2012; Martinez et al., 2010; Sosa et al., 2019a; Sosa et al., 2019b). Here we expand on prior read recruitment methods and use a genome-resolved approach with thousands of randomly sampled single-cell amplified genomes (SAGs) from the Global Ocean Reference Genome (GORG-Tropics) dataset (Pachiadaki et al., 2019). We found that the taxonomic richness of phosphonate producers is higher than for consumers (Supplementary Note S3.1, Figure 3.2A). However, the taxonomic evenness of producers was low, with over 60% of them assigned to SAR11 clades. In contrast, phosphonate consumers were more evenly distributed, but over 90% of SAGs came from just four taxonomic orders: Cyanobacteria, Pelagibacterales, Rhodobacterales and HIMB59 (Figure 3.2B). SAR11 are small aquatic chemoheterotrophic Alphaproteobacteria estimated to constitute up to half of the total plankton cells in the surface ocean (R. M. Morris et al., 2002), and *Prochlorococcus* are unicellular photosynthetic picocyanobacteria that numerically dominate the euphotic zone of subtropical and tropical oligotrophic areas (Biller et al., 2015). As the two most numerically abundant and cosmopolitan marine groups, these two groups alone are likely important phosphonate producers across subtropical/tropical surface ocean.

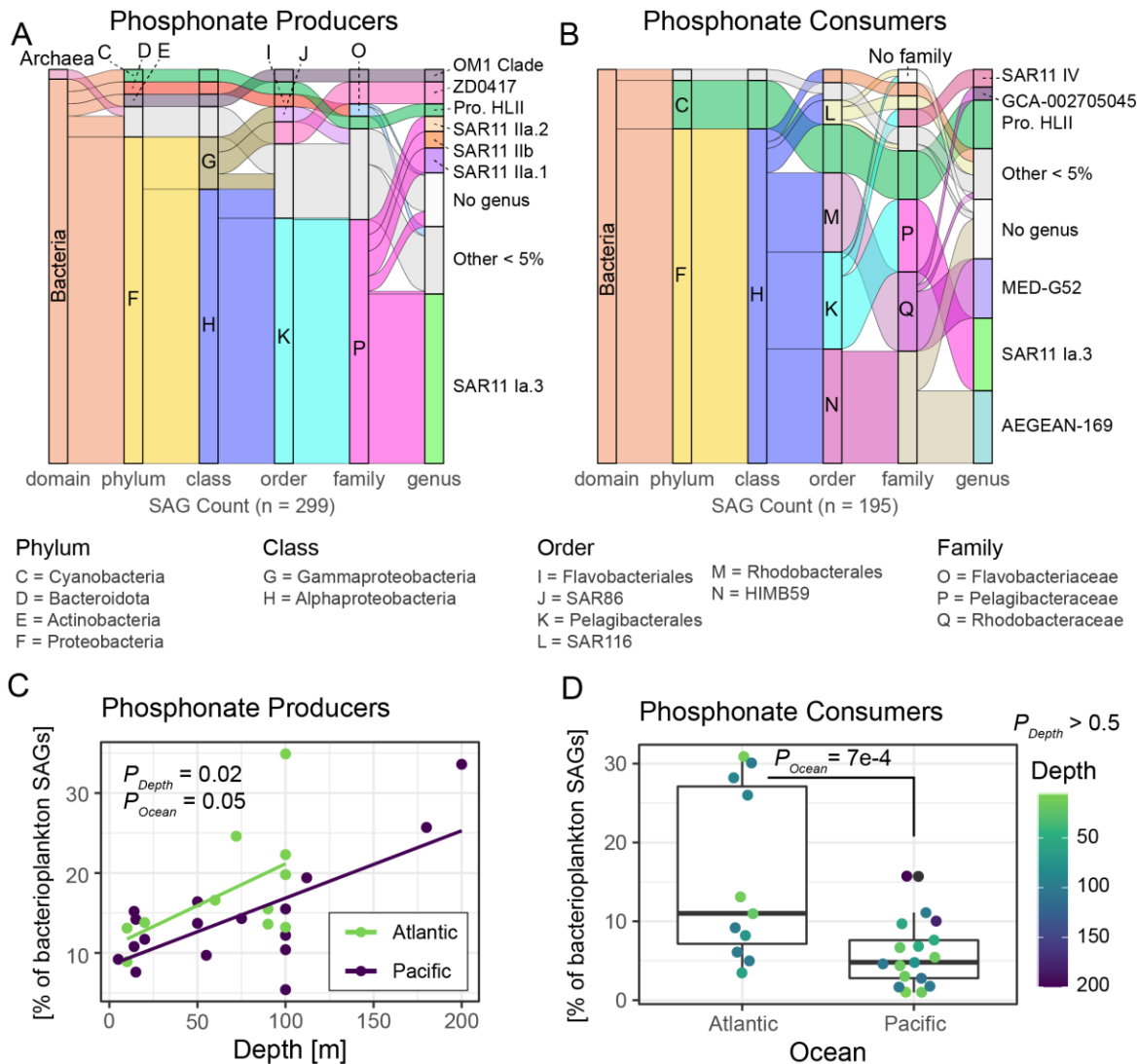


Figure 3.2: Genomic potential for phosphonate production and consumption in nearly 13000 single-cell genomes from the tropical ocean

Flow diagrams represent the taxonomic hierarchy of A) phosphonate producers and B) consumers from the GORG-Tropics database (Pachiadaki et al., 2019). Flows of the same color spanning multiple taxonomic ranks indicate that the higher ranks are entirely subsumed by the lowest rank. For example, all cyanobacterial genomes were from the *Prochlorococcus* genus. Otherwise each partition at each taxonomic level is colored uniquely. For the quantitative comparison of GORG-Tropics sample BATS248 was randomly subsampled to the median depth of the other 27 samples. Letters at each taxonomic rank refer to key marine archaea/bacterial groups (key left) and bar size is proportional to relative abundance. C) The proportion of phosphonate producers significantly increases with depth in the 28 GORG-Tropics samples. Each point is a sample (median SAGs = 241). Producers are slightly more common in the Atlantic than the Pacific [Beta-binomial regression; Depth - Est=0.0036, Err=0.0014, t=2.604, P=0.02; Ocean - Est=-0.30, Err=0.14, t=-2.08, P=0.05; link=logit; log L = -78.179, df=4, resid df=24]. D) Phosphonate consumers are significantly more abundant in the Atlantic than the Pacific [Beta-binomial regression; Ocean - Est=-1.04, Err=0.27, t=-3.81, P=7e-4; link=logit; log L = -80.425, df=4, resid df=24]. Proportions [%] are total producers or consumers divided by the total number GORG assemblies and are corrected using the estimated sequence recovery from assemblies (see methods).

3.4.2 The abundance of phosphonate producing and consuming microbes in the oceans

We leveraged the GORG-Tropics reference database to quantify the proportion of cells in the surface ocean that are phosphonate consumers or producers. GORG-Tropics utilized a randomized cell selection strategy for generating SAGs – thus, minimizing issues related to functional or taxonomic biases more common to targeted genome sequencing. We estimate that, globally, 15% of all bacterioplankton in the upper 100 meters are potential phosphonate producers in the GORG-tropics database. Compared to bacteria, phosphonate producing archaea were rare in the euphotic zone (< 3% of all GORG producers) and phosphonate consuming archaea were absent (Figure 3.2A and B). Planktonic archaea comprised only 14% of GORG-Tropics with 80% of archaea from Marine Group II and 20% from Marine Group I. However, 37% of Marine Group I SAGs contained phosphonate biosynthesis genes. All phosphonate producing MGI archaea were isolated from below 100 meters and included both *Nitrosopumilus* (20%) and *Nitrosopelagicus* (80%). Thus, while archaea may be important phosphonate producers at the base of the euphotic zone and mesopelagic, bacterial producers dominate in the sunlit ocean. Indeed, SAR11 (order Pelagibacterales) is by far the most abundant producer. 18% of all SAR11 SAGs have phosphonate biosynthesis potential and constitute over 60% of all potential phosphonate producer SAGs (Figure 3.2A). Most phosphonate producers are from the surface SAR11 clades 1a.3, IIa.1 and IIa.2, and the mesopelagic IIb clade (Figure 3.2A, Supplementary Figure S3.1). The potential for methylphosphonate production is rare among the SAGs (< 1% of genomes contain MpnS) but 80% of the rare occurrences were in SAR11 genomes. *Prochlorococcus* constitutes 6% of all phosphonate producers (Figure 3.2A) and 3% of all SAGs in the dataset which is significantly less than the 20% estimated from metagenome recruitment in other studies (X. Yu et al., 2013). Most *Prochlorococcus* phosphonate producers are from the surface high-light (HL) clades (Supplementary Figure S3.1). The percentage of bacterioplankton that can produce (15%) and consume (10%) phosphonates is similar across the global surface ocean. On average there are three times more SAR11 phosphonate producers (18%) than consumers (5%), while for *Prochlorococcus* the trend is reversed – 6% producers versus 16% consumers (Supplementary Figure S3.2). SAR11 has been implicated as a major contributor to marine surface water methane supersaturation due to the metabolism of methylphosphonate (Carini et al., 2014). Our results imply that SAR11 has an even more important global role as a phosphonate producer.

We identified two major trends for producers and consumers in GORG-Tropics: 1) the abundance of phosphonate producers significantly increases with depth and is slightly higher in the Atlantic Ocean compared with the Pacific (Figure 3.2C) and 2) consistent with past reports (Feingersch et al., 2012; Sosa et al., 2019b) phosphonate consumers are significantly more abundant in the Atlantic Ocean (Figure 2D). The same trends broadly held for SAR11 and *Prochlorococcus* consumers (Supplementary Figure S3.2). These trends motivated us to look for relationships between phosphonate producers and environmental variables using a metagenomic

dataset encompassing nearly 700 metagenomes from bioGEOTRACES, Tara Oceans, and two long-term ocean time-series sites (Biller et al., 2018; Karsenti et al., 2011; Schlitzer et al., 2018). We determined the relative abundance of the PepM in the upper 250 meters of the global ocean using read recruitment to PepM and conserved marker genes (see methods). We estimate that the median proportion of phosphonate producers in the global ocean is 6%, 10%, and 15% of *Prochlorococcus*, SAR11, and bacterioplankton genomes, respectively (Figure 3.3A), which agrees well with observations from GORG-Tropics. The discrepancy for SAR11 is likely explained by the low estimated sensitivity (60%) for our method to SAR11 (Supplementary Note S3.5). Assuming we missed 40% of SAR11 PepM sequences, the corrected fraction of SAR11 is 17% which, again, agrees well with results from GORG. There was no statistically significant difference in producer abundance between ocean basins (Supplementary Figure S3.3A), and no significant time-averaged difference between the Hawaii Ocean Time-series (HOT) and the Bermuda Atlantic Time Series (BATS) (Supplementary Figure S4), two long- running time series representative of the N. Pacific and N. Atlantic subtropical gyres (Karl & Lukas, 1996; Steinberg et al., 2001). However, phosphonate producers have a significant seasonal dependence at the surface at BATS where producer abundance peaks during winter (November through the following March), which coincides with the peak of wind-driven deep mixing (Supplementary Figure S3.4, Supplementary Note S3.2). Overall, the global median abundance for phosphonate producers is modest (15% of total sequenced genomes) and largely stable across the global subtropical surface ocean.

Are there particular ocean features that may select for the ability of a microbe to produce phosphonate? We used machine learning approaches (Wright & Ziegler, 2015) (Supplementary Note S3.6) and parametric regression (B. D. Martin et al., 2020) (Supplementary Table S3.1) to identify biotic and chemical/physical factors driving the distributions of *Prochlorococcus*, SAR11, and total bacterioplankton phosphonate producers (Supplementary Figure S3). Generally, we find that the greatest amount of phosphonate producer variation is explained by multiple depth-dependent biotic factors (Figure 3.3B, Supplementary Figure S3.3 and S3.5, Supplementary Note S3). In surface samples shallower than 100 m, a median of 15% of bacterioplankton can produce phosphonates, while below 100 m that proportion steadily increases to nearly 30-40% at 200 m. This 100 m break coincides with the nutrient-driven “genomic transition zone” i.e. a zone where bacteria and archaea tend to have larger genomes with higher GC content, and proteins with higher N/C ratios (Mende et al., 2017), and implies that nutrient availability is an important factor setting the relative abundance of phosphonate producers. Overall it appears that a modest proportion of highly abundant oligotrophic “surface” bacterial clades including SAR11 Ia.3 and HL *Prochlorococcus* are capable of producing phosphonates in the upper 50 to 100 m and contribute to an endemic phosphonate pool there. Below 100 m, the genetic potential for phosphonate production shifts to other groups including SAR86, SAR11 Iib, the OM1 clade, MGI Archaea and the ZD0417 clade (Milici et al., 2017). It may be that deep producers are sometimes uplifted into the upper 100 m of the euphotic zone

during seasonal mixing events or ephemeral upwelling events like cyclonic mesoscale eddies (Barone et al., 2019; Nelson et al., 2014).

Our genomic and metagenomic results offer a number of predictions relevant to the biogeochemistry of marine phosphonates. First, the vertical distribution of phosphonate producers in the upper 250 meters is not even, with a distinct and rapid increase in producers beginning at 100 m depth. Bulk phosphonate production rates are similarly stratified (Van Mooy et al., 2015) which implies that deeper phosphonate pools may be important for the total phosphonate balance in the sunlit surface ocean. Second, unlike for phosphonate consumers (Sosa et al., 2019b), phosphate concentration does not appear to have a strong selective effect on the biogeography of phosphonate producers in surface waters. However, this might be due to our statistical approach not having sufficient power to detect a true difference between ocean regions since the mean proportions we compare are small with relatively high variability. Further investigation into ocean basin differences in phosphonate production potential is warranted. Finally, the median proportion of bacteria and archaea that can produce phosphonates in the upper 100 meters of the surface ocean appears to be relatively small (15% of bacterioplankton cells). Since phosphonates appear to be relatively labile and rapidly cycled (Repeta et al., 2016), this implies that cellular production rates would need to be quite high to account for the large global inventory of phosphonates in marine dissolved organic matter.

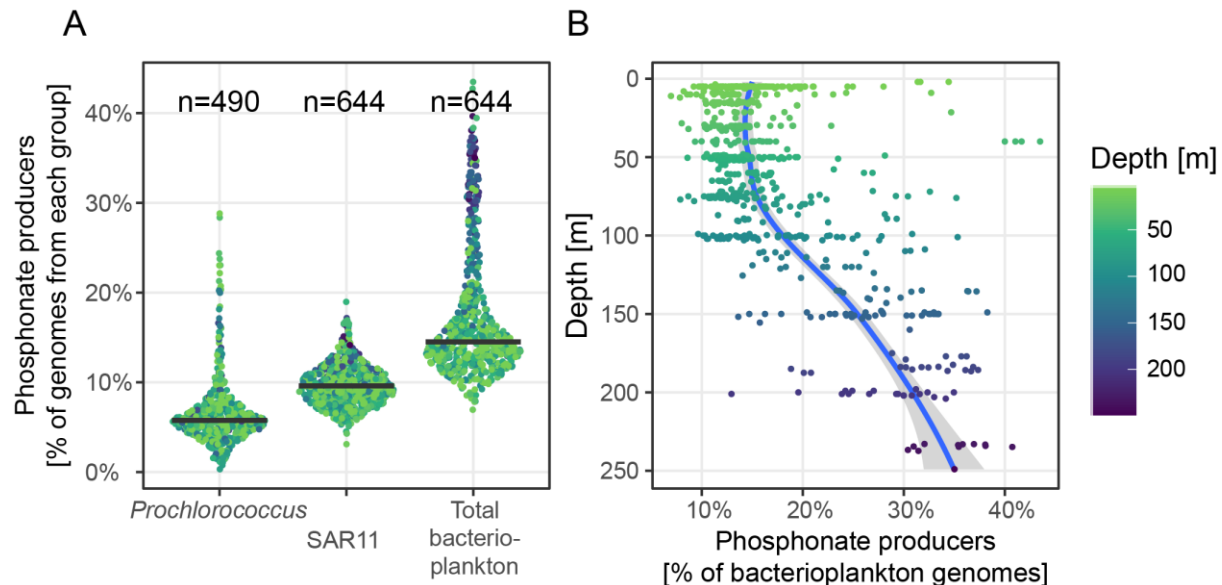


Figure 3.3: Phosphonate biosynthesis genes in surface (< 300 meters) ocean metagenomes

A) Phosphonate producers as the percentage of all bacterioplankton genomes estimated from combined BioGEOTRACES and Tara Oceans samples for *Prochlorococcus*, SAR11, and all Bacteria and Archaea combined. Black line is the median value for n metagenome samples for each taxonomic group. B) The relationship between depth and percent phosphonate producers in the global ocean. The blue line is a simple Loess regression fit to the data. Norm. PepM is the normalized fraction genomes with PepM and is estimated as the length-normalized abundance of PepM divided by length-normalized abundance of taxon-specific marker genes.

3.4.3 Phosphonate production in *Prochlorococcus* SB

To better understand how much cellular P is allocated to phosphonate production and with which macromolecules they are associated, we studied phosphonate production by an abundant, widespread, and experimentally tractable marine microbe. We found only one cultured isolate that meets these criteria: *Prochlorococcus* SB. Since there is as yet no genetic system for knockout mutants in *Prochlorococcus*, we used *Prochlorococcus* MIT9301, a closely related strain lacking the phosphonate biosynthesis cluster, as a control and analyzed cell pellets from each strain using ^{31}P nuclear magnetic resonance spectroscopy (^{31}P -NMR). As expected, the ^{31}P -NMR spectrum of MIT9301 cells shows that all P is allocated to phosphate and pyrophosphate esters (-12 to 12 ppm; Figure 3.4B). By contrast, ^{31}P -NMR spectrum of *Prochlorococcus* SB cells displays strong signals between both -10 and 12 ppm from phosphate and pyrophosphate esters, and between 18 and 27 ppm from phosphonates (Quin & Williams, 2004) (Figure 3.4B). Integration of the phosphate and phosphonate ester regions of the NMR spectrum yields a cellular ratio of Phosphonate/Phosphate of 0.72. (Figure 3.4B; Supplementary Table S3.3) and under nutrient replete conditions *Prochlorococcus* SB cells allocate ~ 40% of their cellular P to phosphonate production.

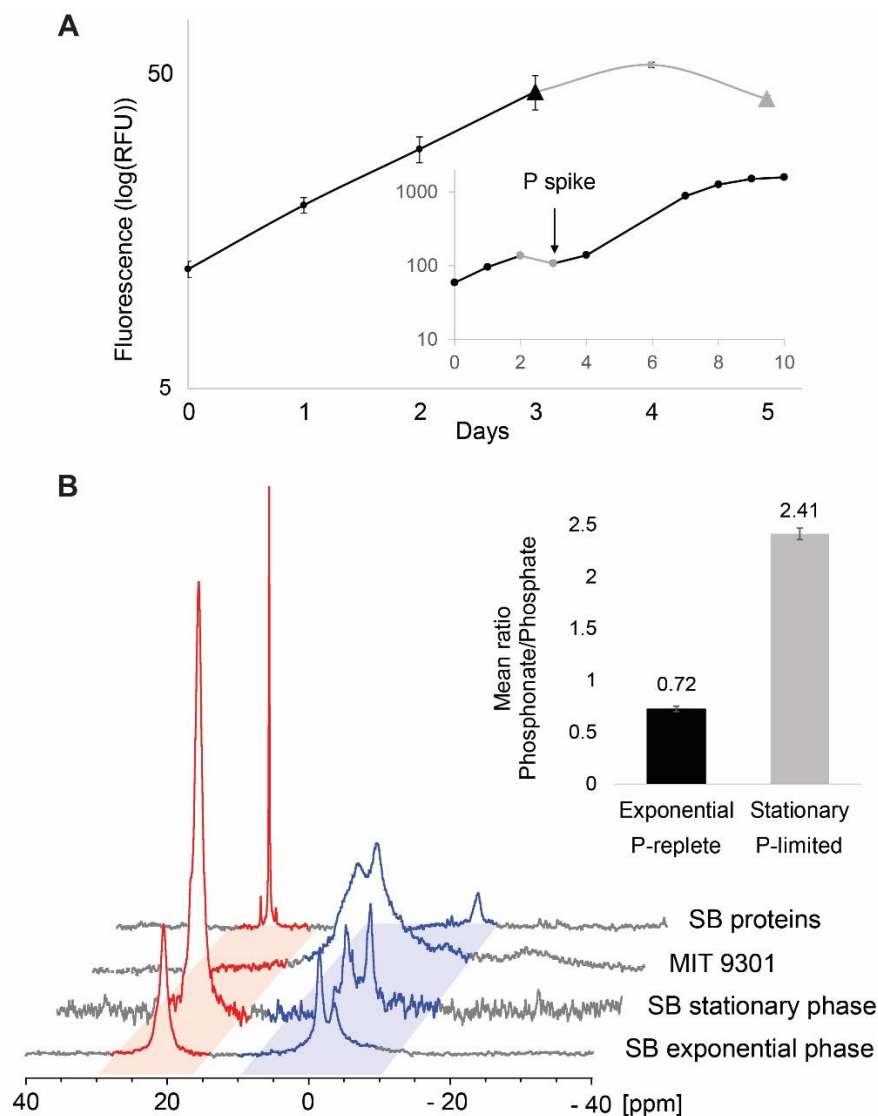


Figure 3.4: Phosphonate production by *Prochlorococcus* cultures

A) *Prochlorococcus* SB growth curve showing exponential phase (black) and stationary phase (grey) due to P-limitation. Triangle data points correspond to the days cultures were harvested in the different growth phases. Error bars are calculated based on the standard deviation between the biological duplicates. The inset represents the *Prochlorococcus* SB growth curve in medium with N/P = 350/1 in which inorganic phosphate was added on Day 3 to reach N/P = 16/1. B) ^{31}P -NMR spectra of *Prochlorococcus* SB whole cells harvested in exponential phase (P-replete) and in stationary phase (P-limited), the negative control *Prochlorococcus* MIT9301 harvested in exponential phase (P-replete) and the insoluble protein fraction of *Prochlorococcus* SB harvested in exponential growth phase. The phosphonate and phosphate regions of the spectra are indicated in red and blue respectively. While *Prochlorococcus* SB produces phosphonate and doubles its relative phosphonate content in P-limited stationary phase, the negative control, MIT9301 only produces phosphates. The histogram (inset) displays the mean Phosphonate/Phosphate ratios for *Prochlorococcus* SB cells harvested in exponential (black) and stationary (grey) phase calculated by integrating the phosphonate and phosphate peaks in *Prochlorococcus* SB whole cell ^{31}P -NMR spectra obtained for the duplicates in each growth phase. Error bars correspond to the standard deviation of the biological replicate phosphonate/phosphate ratio values with $n=2$.

3.4.4 Phosphonates and P storage

The physiological and ecological roles of phosphonates are poorly understood. However, the presence of PepM among diverse genome-streamlined bacteria and the high relative abundance of cellular phosphonates in *Prochlorococcus* SB both suggest that phosphonates serve an important function for some microbes inhabiting oligotrophic marine waters. They may serve as an intra-cellular P-storage reservoir (Metcalf & van der Donk, 2009), for example. In highly stratified oligotrophic waters, nutrient supply to phytoplankton is episodic; driven by mixing events that bring nutrient-rich waters from below the surface into the euphotic zone. To synchronize their need for nutrients with an episodic supply, some microbes take up excess P during periods of high nutrient concentrations and sequester them internally (P. Martin et al., 2014; Temperton et al., 2011). When nutrient concentrations fall, internal P-stores are metabolized to release inorganic phosphate. However, the *Prochlorococcus* SB genome lacks phosphonate degradation pathways including C-P lyase, phosphonate hydrolytic pathways (Horsman & Zechel, 2017), and PhnYZ an oxidative pathway recently discovered in *Prochlorococcus* (Sosa et al., 2019a). Recognizing that *Prochlorococcus* SB might use an uncharacterized pathway to repurpose P from phosphonates, we tested the P-storage hypothesis by comparing phosphonate production in cultures grown under P-starved and P-replete conditions (Figure 3.4A), expecting that if phosphonates were used for P-storage, P-limitation would reduce the allocation of P to phosphonates. However, we found that *Prochlorococcus* SB allocates significantly more P to phosphonates relative to phosphates upon entering P-limited stationary phase growth (Phosphonate/Phosphate = 2.4) than during P-replete exponential growth (Phosphonate/Phosphate = 0.72; Figure 3.4B). Cellular phosphorus to carbon (C/P) was relatively stable across exponential and stationary phases (145 and 131 respectively) and changes in Phosphonate/Phosphate were driven by a decrease in phosphate ester content during P-starvation, with continued production of phosphonates (Supplementary Figure S3.6, Supplementary Table S3.4). As *Prochlorococcus* SB becomes increasingly P starved, the regulation of the cellular phosphate and phosphonate pools becomes decoupled from exponential growth conditions. While reallocation of P away from labile phosphates is one mechanism by which *Prochlorococcus* adapts to P-starvation (Van Mooy et al., 2006), *Prochlorococcus* SB appears to be less able to regulate phosphonate production or repurpose P in phosphonates towards other cellular functions. Both scenarios are inconsistent with phosphonates as a P-storage reservoir in *Prochlorococcus* SB, and P locked into phosphonates is not internally recycled to sustain growth during periods of P-limitation.

This conclusion is reinforced by our comparative genomic analysis. Indeed, in addition to biosynthesis genes, we searched for all known phosphonate degradation pathways (Figure 3.1) in the GORG-Tropics and MARMICRODB (Becker et al., 2019) genome datasets and we found that less than 1% of all genomes encoding at least one marker gene implicated in phosphonate metabolism are both producers and consumers (Figure 3.5). This mutual exclusivity occurs at the species/strain level - for example within HL *Prochlorococcus* - which implies strong functional

differentiation between even closely related marine microbes. This suggests fine-scale niche partitioning between phosphonate producers and consumers in the environment and may reflect functional incompatibility or ecological/evolutionary tradeoffs between biosynthesis and catabolism. It also implies that if phosphonates were a widespread storage strategy, most phosphonate producers would have no way to reclaim P from phosphonates during times of need.

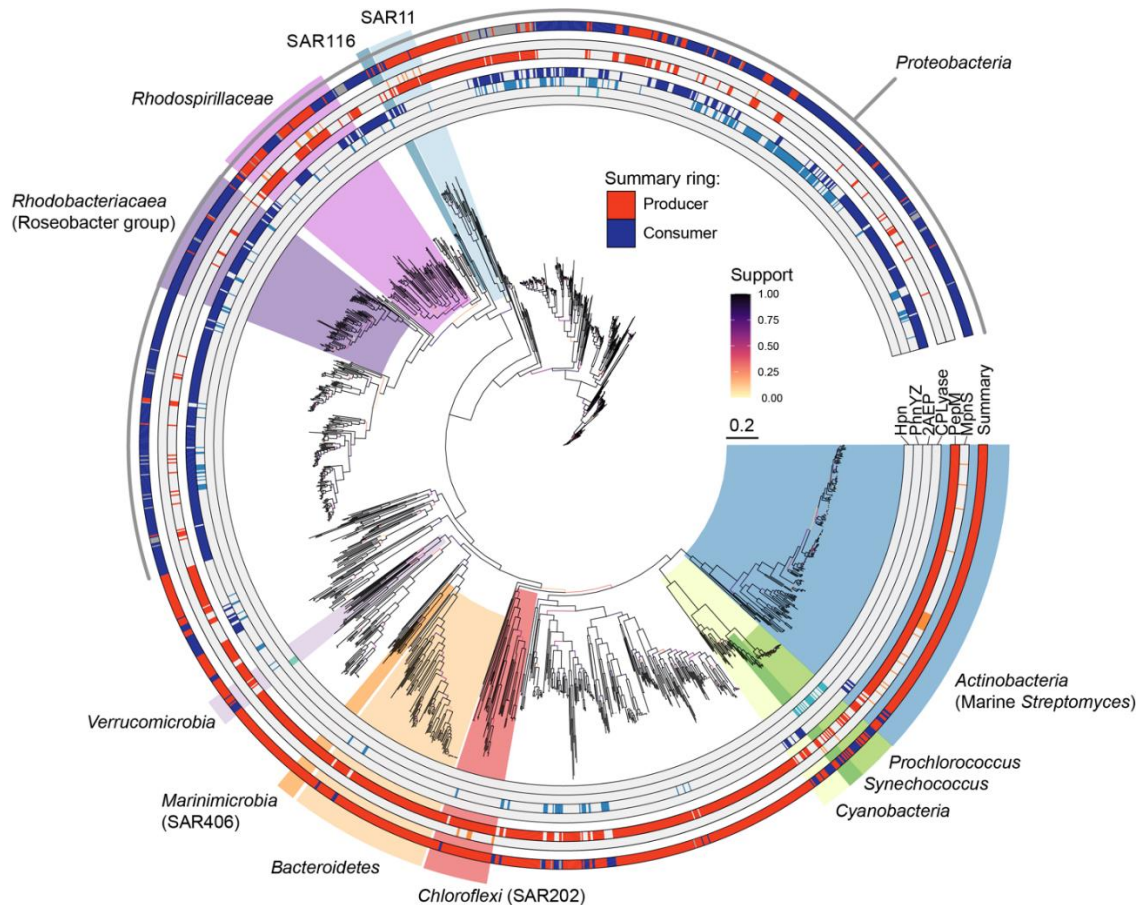


Figure 3.5: Distribution of phosphonate biosynthesis and utilization pathways in bacterial genomes

Phylogeny is constructed from 120 concatenated, single-copy marker genes from 1890 bacterial genomes from MARMICRODB containing either a phosphonate catabolism or phosphonate biosynthesis pathway. Scale bar is 0.2 amino acid substitution and the tree is unrooted. Monophyletic taxonomic groups with marine representatives are highlighted. The presence of four different phosphonate catabolic pathways (Hpn: Phosphonate catabolism via HpnWXZ, PhnYZ: phosphite or methylphosphonate catabolism via PhnYZ (Sosa et al., 2019a), 2AEP: 2-aminoethylphosphonate catabolism via phosphonoacetate, CPLyase: multisubunit C-P lyase system) is displayed in blue on an inner ring while phosphonate/methylphosphonate biosynthesis pathways (PepM and MpnS) are shown in red on the middle ring. The outer summary ring indicates whether a genome contains at least one catabolic pathway (blue), a phosphonate biosynthesis pathway (red), or both (dark grey, < 0.5% of all genomes). For simplification, the MpnS category also includes the functionally related enzyme HepDI (Born et al., 2017).

3.4.5 Macromolecular forms of phosphonates

Past studies have linked the biochemical function of phosphonates to their macromolecular form. For example, incorporation of phosphonates into membrane lipids or capsular polysaccharides may protect cells against phospholipase activity or inhibit phage attachment (Coyne et al., 2000; Hilderbrand, 2018; Wilkinson & Holmes, 1979). Therefore, we next asked what macromolecular form do phosphonates take in *Prochlorococcus* SB? Phosphonate biosynthesis genes in most *Prochlorococcus* and SAR11 genomes are surrounded with genes coding for glycosylating enzymes predicted to be involved in the biosynthesis of large extracellular polysaccharide structures, i.e. bacterial capsules (Supplementary Figure S3.7, Supplementary Table S3.2). Many marine phytoplankton and bacteria produce extracellular layers of polysaccharides to facilitate aggregation, for defense against predation, and to manufacture biofilms (Decho & Gutierrez, 2017). Phosphonates have been shown to be constituents of the large reservoir of dissolved polysaccharides that accumulate in the surface ocean (Repeta et al., 2016), and we expected to find phosphonates in *Prochlorococcus* SB to be associated with polysaccharide macromolecules. However, upon fractionation of *Prochlorococcus* organic matter into major biochemical classes, we found that phosphonates were recovered in the protein fraction (Figure 3.4B) specifically within the methanol/acetone insoluble HMW protein fraction, which includes membrane-associated proteins.

We reconcile these results by proposing that, in *Prochlorococcus* SB, phosphonates are integrated into glycan polymer chains that are then post-translationally attached to HMW proteins, likely membrane-anchored proteins. Indeed, one of the most common post-translational modifications of bacterial proteins is O-linked glycosylation (Iwashkiw et al., 2013), where glycan polymers are covalently bound to serine or threonine side-chains of larger protein complexes. The phosphonate gene clusters in *Prochlorococcus* SB, SAR11 strain HTCC7217, and SAR11 RS40 genomes are adjacent to glycan assembly enzymes and lipid carriers predicted to be involved in the biosynthesis of capsules (Supplementary Figure S3.7). However, in some bacteria the glycan building blocks for capsule biosynthesis are also routed towards post translational O-linked glycosylation of membrane-bound lipoproteins (Lees-Miller et al., 2013). *Prochlorococcus* SB genome contains an O-oligosaccharyltransferase protein directly upstream of the PepM, Ppd, and Pdh cluster. This protein family was first characterized in the attachment of glycans to liposaccharides, but O-oligosaccharyltransferases have more recently been demonstrated to transfer preassembled glycan chains onto protein substrates (Faridmoayer et al., 2007; Power et al., 2006). O-oligosaccharyltransferase genes are also present near PepM in *Pelagibacter* strains HTCC7217 and RS40 (Supplementary Figure S3.7), and 15% of GORG-tropics SAGs contain PepM plus an O-oligosaccharyltransferase-like domain colocalized within the same 10kbp genome segment. This proportion increases to 50% if we relax the condition that the genes must occur on the same genomic contig. The genomic evidence implies that

phosphonates could be a common moiety involved in the post-translational O-linked glycosylation of proteins in the ocean.

3.4.6 Functional roles of phosphonylated glycoproteins

Why would an oligotrophic-adapted organism like *Prochlorococcus* use scarce P to produce large amounts of phosphonate-containing glycoproteins? Phosphonates are often incorporated into cell-surface structures (Coyne et al., 2000; Rice et al., 2019) because they are highly resistant to hydrolysis and can inhibit the activity of some hydrolytic enzymes by mimicking carboxylic acids and phosphate esters (White & Metcalf, 2007). For planktonic bacteria in the ocean, surface-expressed structures provide protection against enzymatic attack, UV-radiation, phages (Seed, 2015), and protozoan grazers (Jürgens & Matz, 2002). Modifying cell-surface structures with phosphonates may play a role in reducing vulnerability to grazers and phage, two major drivers of mortality for *Prochlorococcus* (Partensky et al., 1999). Given that surface expressed proteins are some of the most abundant proteins associated with the cell (Fagan & Fairweather, 2014), the large proportion of cellular P devoted to phosphonates in *Prochlorococcus* SB (Figure 3.4B) is also consistent with surface protein modification.

Natural selection generally promotes high turnover and low population frequency of genes involved in phage -and predator- interactions (Cordero & Polz, 2014) (Supplementary Note S4). Horizontal gene exchange is a key mutational mechanism maintaining genes at low frequencies within microbial populations (Cordero & Polz, 2014), and cell surface modification traits are highly enriched in horizontal gene exchange networks (Nakamura et al., 2004). Similarly, we find that patterns of phosphonate biosynthesis gene flow in marine microbial communities do not follow a tree-like pattern and are better explained by horizontal exchange and/or gene loss. Generally, there is a poor correlation between the species tree, inferred from conserved marker genes, and the PepM tree. The topological distance between the species phylogeny and the PepM phylogeny approaches values for random simulated tree sets (Supplementary Table S3.5). In *Prochlorococcus*, nearly all phosphonate biosynthesis gene cassettes are located within “genomic islands” (Supplementary Figure S3.8), which contain the majority of laterally transferred genes in *Prochlorococcus* (Coleman et al., 2006; Kettler et al., 2007). Indeed, there is also a transposase, an important molecular mechanism of lateral gene transfer in bacteria (Frost et al., 2005), four genes upstream of PepM in the *Prochlorococcus* SB genome. Taken together, we speculate that the low frequency of phosphonate producers in bacterial populations and the evidence for horizontal transfer is consistent with a role for phosphonates within cell surface structures – potentially as a defense against phage or grazers.

3.5 CONCLUSION: BIOGEOCHEMICAL IMPLICATIONS - A MICROBIAL PHOSPHONATE LOOP WITHIN THE MARINE PHOSPHORUS CYCLE

Above 50 meters we expect that chronic, long-term P scarcity drives the cost of phosphonate synthesis to outweigh its fitness benefit for most of the bacterioplankton. This is consistent with the numerical dominance of small, genome-streamlined cells, the low gene frequency of the phosphonate biosynthesis pathway, the low measured concentrations of particulate phosphonates, and the low rates of phosphorus reduction relative to rates of community phosphate uptake in the upper euphotic zone (Sannigrahi et al., 2006; Van Mooy et al., 2015). At the nutrient-driven genomic transition zone and near the deep chlorophyll maximum, organic matter remineralization rates are relatively high and the bacterioplankton community is in close proximity to episodic phosphate inputs from below. Phosphate concentrations begin to rise, and the fitness benefit of phosphonate biosynthesis as a defense against viral lysis and grazing begins to outweigh its cost (Figure 3.6), sustaining the high frequency of phosphonate producers we found in this depth range. We expect that phosphonates will be most abundant in particulate matter collected near the genomic transition zone, and that higher rates of phosphonate production would lead to higher rates of phosphonate degradation, consistent with the higher C-P lyase gene copy numbers and higher C-P lyase activity measured in microbes inhabiting the deep chlorophyll maximum (Granzow et al., 2020; Sosa et al., 2019b). If phosphonates are used to modify cell-surface glycoproteins as we suggest, then after cell death the labile protein component likely turns over quickly, but the inherently slow degradation of structural polysaccharides would lead to the accumulation of phosphonoglycans as high molecular weight dissolved organic matter (Repeta, 2014) explaining their abundance in marine DOP (Figure 3.6A).

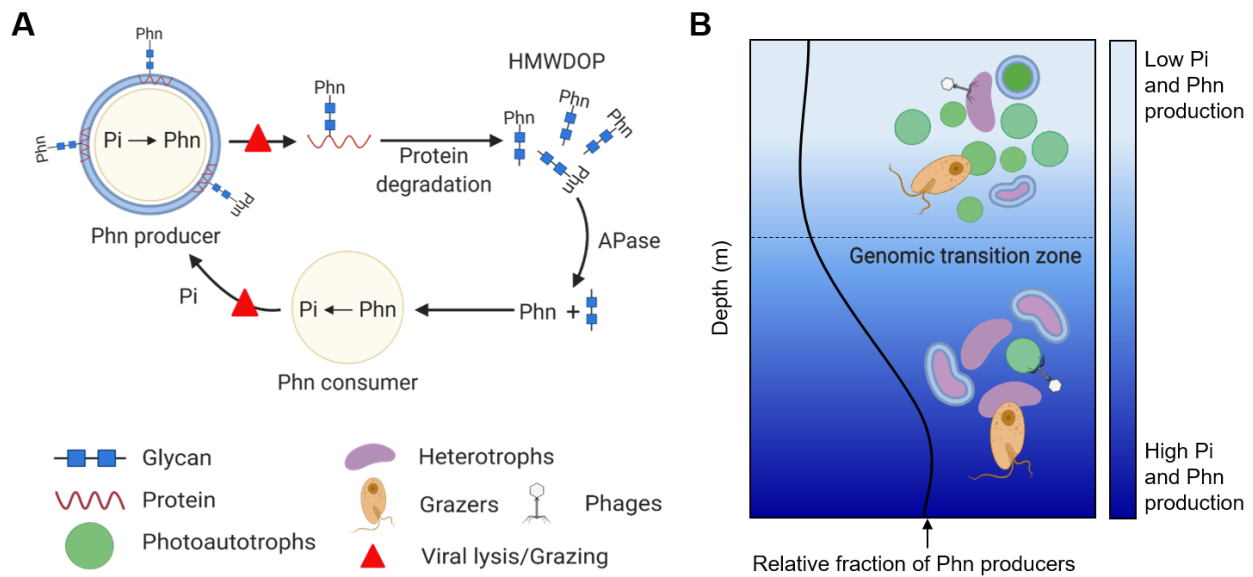


Figure 3.6: Microbial cycling of phosphonates in the upper ocean – an hypothesis

A) Microbes with PepM produce cell surface-layer phosphonoglycoproteins to reduce mortality from grazing or viral infection. Upon death of the cell phosphonoglycoproteins are released into seawater where heterotrophic microbes quickly remineralize proteins leaving phosphonoglycans to accumulate as high molecular weight dissolved organic phosphorus (HMWDOP). Phosphonates are hydrolyzed from glycans by alkaline phosphatases (APase) family enzymes and further hydrolyzed into inorganic phosphate (Pi) by C-P lyase or other hydrolytic pathways. Recycled Pi can then be used to produce new phosphonates. B) In the surface ocean, Pi is scarce and often limiting. Producing phosphonate as a mortality defense may be costly in terms of resource allocation. Therefore, the relative fraction of phosphonate producers in the microbial community is low. As nutrient availability increases through the genomic transition zone (Mende et al., 2017) the benefit/cost ratio of phosphonate production increases and phosphonate producers are relatively more abundant.

3.6 MATERIAL AND METHODS

3.6.1 Genomic data sources

We used two collections of genomes for the comparative genomics analysis: MARMICRODB (Becker et al., 2019) (<https://zenodo.org/record/3520509>) and the Global Ocean Reference Genomes (GORG) Tropics dataset (Pachiadaki et al., 2019). MARMICRODB contains over 18000 archaeal, bacterial, eukaryotic, and viral genomes from predominantly the marine environment, but also terrestrial and host-associated systems. GORG Tropics consists of approximately 13000 single-cell genomes sequenced from 28 samples across the tropical surface ocean. Roughly 6000 genomes from GORG Tropics come from a single sample (GORG BATS) from the Sargasso Sea. For quantitative analysis of the complete GORG dataset we randomly subsampled GORG BATS (Sample SWC-09) genomes to the median genomes per sample (N=241) from the remainder of GORG Tropics. We also used 195 surface and Deep Chlorophyll

Maximum metagenomes from Tara Oceans project (Karsenti et al., 2011) and 480 metagenomes from bioGEOTRACES, HOT, and BATS (Biller et al., 2018). Genome and metagenome quality control and exclusion criteria were performed as described earlier (Becker et al., 2019).

3.6.2 Homology searches

Phosphonate biosynthesis and catabolism proteins were identified by homology to a collection of hidden Markov models using HMMERv3.1b2 (<http://hmmer.org/>) and the trusted cutoffs of each individual model. PepM sequences were initially identified using PF13714 (PEP_mutase) and then aligned to the Tigrfam (Haft et al., 2013) Hidden Markov Model (HMM) TIGR02320 using hmmlalign. Authentic PepM sequences were identified by strict presence of conserved active site motif “EDK(X)5NS” or if they lacked at most two active site residues but were located adjacent (within 5000 nucleotides upstream/downstream) to a gene coding for Ppd (TIGR03297) or MpnS (Chen et al., 2006; Cioni et al., 2014; Ju et al., 2015; X. Yu et al., 2013). The PF13714 model identified many sequences lacking the PepM active site motif, and with significant similarity to the related isocitrate lyase superfamily PF00463. Sequences passing the PF13714 bitscore cutoffs, but lacking the “EDK(X)5NS” motif were classified as members of the isocitrate lyase superfamily. Sequences from authentic MpnS and the related HepDI/HepDII proteins were identified using custom HMMs built from alignments of curated sequences identified as containing essential catalytic residues (Born et al., 2017). These alignments were constructed using MAFFT v7.273 (Katoh & Standley, 2013) in E-INS-i iterative refinement mode. The alignment was trimmed to peptide positions 317-792 to isolate the core catalytic residues and used to construct HMMs. MpnS, HepDI, and HepDII sequences in MARMICRODB and GORG-tropics were identified using the resulting HMMs (evalue < 0.05), and the resulting sequences were aligned using MAFFT. The alignments were manually inspected to remove sequences lacking the essential catalytic residues and were iteratively realigned after poorly-aligned sequences were removed.

For phosphonate catabolism proteins, gene neighborhoods were inspected to confirm the presence of multiple co-occurring genes from the PhnYZ (Sosa et al., 2019a), C-P lyase, 2-aminoethylphosphonate, and phosphonoacetaldehyde phosphonatase catabolism pathways. Specifically, valid C-P lyase clusters required at least five other genes from phnC, phnD, phnE, phnF, phnG, phnH, phnI, phnJ, phnK, phnL, phnM, phnN, phnO, phnP to occur within uninterrupted (ie no scaffold/contig breaks) 16000 nucleotide sliding windows. Valid phosphite clusters required two of ptxD, phnY, and phnZ. Valid 2-aminoethylphosphonate clusters required co-occurring phnX and phnW or phnW, phnA, and phnY. Valid phosphonoacetaldehyde phosphonatase clusters required phnZ, hpnW, and hpnZ. All catabolic pathways other than C-P lyase required genes to co-occur within uninterrupted 10000 nucleotide sliding windows.

We also identified two sets of ten highly conserved, single-copy families from *Prochlorococcus* and SAR11 lineages for use in normalization with metagenome profiling. These gene families

had the ten lowest dN/dS ratios of the entire pangenome, that is the highest purifying/stabilizing selection, for *Prochlorococcus* or SAR11, and should thus be highly specific and sensitive for metagenomic quantification. *Prochlorococcus* and SAR11 pangenomes were characterized and defined using PanX (Ding et al., 2018). To profile metagenomes for total bacterioplankton, we used fetchMG from the mOTUs2 tool (Milanese et al., 2019). Briefly, we identified ten COG families (COG0012, COG0016, COG0018, COG0172, COG0215, COG0495, COG0525, COG0533, COG0541, COG0552) from all bacterial and archaeal genomes in MARMICRODB. To reduce redundancy, we clustered protein sequences from each COG at 90% sequence identity using MMseqs2 v4e23d (Steinegger & Söding, 2017) and used the 90% clustered sequence representatives in metagenome search databases.

The HMM protein sequence profiles, results from these searches, and descriptions of the sequence families are available from <https://github.com/slhogle/phosphonates>.

3.6.3 Multiple sequence alignments, phylogenetic inference, and topological comparison

Authentic PepM sequences containing the “EDK(X)5NS” motif were aligned to TIGR02320 using hmalign and trimmed using trimAl v1.4.rev22 (Capella-Gutiérrez et al., 2009) with the automated -gappyout option, and alignments were inspected manually to ensure the veracity of the “EDK(X)5NS” motif. The multi-phylum genome phylogenies were created from genome assemblies with confirmed phosphonate biosynthesis pathway or at least one confirmed phosphonate catabolic pathway. We used the GTDB-Tk v1.3.0 pipeline (Chaumeil et al., 2019) with default settings and the GTDB R05-RS95 database (Parks et al., 2018) to identify conserved proteins (120 bacterial proteins/122 archaeal proteins) and generate concatenated multi-protein alignments. We filtered alignment columns using the bacterial and archaeal alignment masks from (<http://gtdb.ecogenomic.org/downloads>). We then removed columns represented by fewer than 50% of all taxa and/or columns with no single amino acid residue occurring at a frequency greater than 25%. We further trimmed the alignments using trimAl with the automated -gappyout option to trim columns based on their gap distribution. The multi-phylum genome phylogenies and the PepM phylogeny were inferred using FastTree v2.1.10 (Price et al., 2010) under the GAMMA model of rate heterogeneity and the WAG+ substitution model (Whelan & Goldman, 2001). Support values were determined using 100 non-parametric bootstrap replicates. Both the PepM tree and genome tree were left unrooted. Phylogenies and associated data were visualized using ggtree (G. Yu et al., 2017). Detailed phylogenies of *Prochlorococcus* and *Pelagibacterales*/SAR11 in the supplementary materials were constructed as described earlier (Becker et al., 2019).

Comparisons between the topology of the PepM tree and the genome phylogeny were performed using ETE3 v3.1.1 (Huerta-Cepas et al., 2016a). We pruned both the PepM and genome phylogenies so that they contained the same number taxa and no duplication events in the PepM

tree, which resulted in 999 leaves for bacteria and 16 leaves for archaea. We then compared the topologies of the PepM tree, the genome tree, and ten simulated random trees using the Robinson-Foulds symmetric distance (Robinson & Foulds, 1981) and the fraction of edge similarity. For Robinson-Foulds distances, edges were treated as unpolarized splits rather than “true” clades because all trees were compared as unrooted. The Robinson-Foulds distance simply counts the number of branch partitions (nodes) that appear in one tree but not the other. Therefore, the maximum possible Robinson-Foulds distances of an n-taxa unrooted tree is $2(n-3)$. To compute a normalized distance, we simply divided the observed Robinson-Foulds distance by the maximum distance of the two-way comparison between trees. Thus, the normalized Robinson-Foulds distance is a value from 0 to 1, which can be interpreted as the fraction of nodes/splits missing in the query tree compared with the reference tree.

3.6.4 Gene enrichment analysis

We annotated the 16 *Prochlorococcus* and 22 SAR11/*Pelagibacterales* genomes containing verified PepM sequences with eggNOG 4.5.1 (Huerta-Cepas et al., 2016b) using eggNOG- Mapper v1.0.3-3-g3e22728 (Huerta-Cepas et al., 2017). We then collected the resulting KEGG orthology annotations (Kanehisa & Goto, 2000) from all genes and tested for enrichment of modules and pathways in the KEGG hierarchy in genes within 10000 nucleotides upstream/downstream of the PepM sequence start and stop coordinates respectively. Significant enrichment of KEGG categories were determined using the hypergeometric test (Rivals, Personnaz, Taing, & Potier, 2006) implemented in clusterProfiler v3.8 (G. Yu et al., 2012). We excluded genomes from the analysis that had contig/scaffold breaks within 10000 nucleotides upstream/downstream of the PepM sequence.

3.6.5 Identification of PepM sequences within *Prochlorococcus* genomic islands

We predicted genomic islands using Hidden Markov Models trained from conserved gene synteny patterns in closed *Prochlorococcus* genomes. Briefly, we assume genomic islands to be contiguous stretches of DNA enriched in flexible genes families (i.e. genes found in a subset of all genomes). We then used previously described genomic islands in six *Prochlorococcus* genomes (Avrani et al., 2011; Coleman et al., 2006; Dufresne et al., 2008) to define core, flexible, and inconclusive gene “states” in four distinct hidden markov models, each with two hidden states (island or non-island). We then used empirical gene family frequencies as a proxy for the core/flexible/inconclusive state of each gene and the gene order on each scaffold as input

to the Viterbi algorithm for predicting the hidden island state of each gene in all other *Prochlorococcus* genomes.

3.6.6 Estimation of PepM prevalence in *Prochlorococcus* and SAR11 genomes

Because many of the *Prochlorococcus* and SAR11 genomes analyzed here are incomplete (SAGs or MAGs), we attempted to estimate the ‘true’ proportion of genomes with PepM while correcting for genome incompleteness. Briefly, we used the estimated genome completeness from checkM v1.0.11 (Parks et al., 2015), which is based on the presence of core marker genes, to estimate the number of missing bases per taxonomic group (*Prochlorococcus* or SAR11) and then use this scale the relative abundance of potential phosphonate producers per group. We estimated the corrected prevalence of phosphonate producers as:

$$\left(\sum_{i=1}^n p / \sum_{i=1}^n g \right) \times \sum_{i=1}^n (g/c) \times \bar{p}^{-1} \times n^{-1}$$

where for each clade, p is the length of the phosphonate biosynthesis/degradation operon in base pairs, g is the length of the genome assembly in base pairs, c is the completeness estimate from CheckM, \bar{p} is the average length of all phosphonate biosynthesis/degradation operons from each clade, and n is the total number of assemblies from each clade.

3.6.7 Metagenome read classification

We created Diamond v0.9.22.123 (Buchfink et al., 2015) search databases for each of the three ten-gene marker gene sets identified for *Prochlorococcus*, SAR11, and all bacterioplankton using reference genomes from MARMICRODB. We also created Diamond databases for only *Prochlorococcus* PepM, only SAR11 PepM, and all bacterial and archaeal PepM sequences identified with the “EDK(X)5NS” catalytic motif from MARMICRODB only. We identified the marker gene sets using a two-tiered search strategy using Diamond in default fast mode. We first searched the metagenomes against reduced marker sets clustered at 70% amino acid identity by MMseqs2. We pulled all reads with hits to these reduced marker sets, then searched the reads against the entire MARMICRODB, and finally retained read mappings with a best scoring match to a MARMICRODB protein from the original marker family. We searched metagenomic reads against the PepM databases using Diamond (mode --more-sensitive) and score cutoffs of 55% amino acid identity and an E-value of 1e-5. We determined these cutoffs empirically to be those that produced the highest F Score (harmonic mean of precision and recall) from mock metagenomes simulated from GORG-tropics (Supplementary Note S3.5, Supplementary Figure S3.9). Since the E-value is a function of database size it is important to note that the significance of this cutoff is specific to the reference databases here. In the case of ambiguous alignments (i.e., identical best alignment scores to sequences from different taxonomic groups) we classified

reads using a probability-weighted random sampling of all taxonomic groups matching the best hits. We derived the probability distributions for random samplings from the taxonomic classification uniquely mapped PepM reads.

To normalize PepM reads across metagenomes we calculated the length normalized counts (RPKM) of the 10 single copy marker genes from *Prochlorococcus*, SAR11, and all bacteria and archaea in each metagenome as reads per kilobase marker and then estimated the number of ‘genome equivalents’ in each metagenome as the median of the 10 marker families. We also calculated the length normalized abundance of PepM from *Prochlorococcus*, SAR11, and all bacteria and archaea as reads per kilobase from each metagenome. We then divided the length normalized abundance of PepM by the estimated number of genome equivalents for each taxonomic group to estimate the fraction of genomes with phosphonate biosynthesis potential. To ensure robust estimates of normalized abundance we excluded samples where the median marker gene coverage was less than 100X for all taxonomic groups.

3.6.8 Biotic and abiotic data associated with metagenomes

The biotic and abiotic variables used in this study were obtained and preprocessed as described in detail here: <https://doi.org/10.5281/zenodo.3689249> and here: <https://doi.org/10.5281/zenodo.3786232>. We obtained phosphate concentrations from the GEOTRACES Intermediate Data Product IDP2017 version 3 (Schlitzer et al., 2018), specifically from sections GA02 (Rijkenberg et al., 2014; Salt et al., 2015), GA03, GA10 (Wyatt et al., 2014), and GP13. We obtained dissolved phosphate concentrations from the Tara Oceans project (Speich et al., 2017) (<https://doi.pangaea.de/10.1594/PANGAEA.875579>). Modeled climatological dissolved organic phosphorus and other variables were obtained from the MIT Darwin model (v0.1_llc90, <http://darwinproject.mit.edu/>) from the Simons Collaborative Marine Atlas project (CMAP) <https://simonscmap.com/> using pycmap v0.1.2 (<https://doi.org/10.5281/zenodo.3561147>). Taxonomic profiles derived from metagenomic reads were generated as described earlier (Becker et al., 2019) using the MARMICRODB database. As a first order estimate of *Prochlorococcus* and SAR11 ecotype relative abundance we divided the number of reads mapping to each ecotype by the number of reads mapping in total the *Prochlorococcus* genus or the family *Pelagibacterales* for SAR11. Some proportion of reads mapping to highly conserved (core) regions can only be reliably classified at the genus or family level thus our ecotype relative abundance estimates using total read counts are underestimates.

3.6.9 Random Forest Regression

We trained one random forest model ($n_{\text{trees}}=1000$) for each taxonomic group; *Prochlorococcus*, SAR11, and bacteria plus archaea. Each model was trained using up to 44 abiotic/biotic variables including trace metal and macronutrient data from GEOTRACES and Tara Oceans, modeled climatological means from the MIT Darwin model (<http://darwinproject.mit.edu/>), and ecotype

relative abundances. For the outcome variable (normalized PepM relative abundance) we used a modified splitting rule for tree construction that maximized the log-likelihood of the beta distribution on the interval [0,1] (Weinhold et al., 2019). Although our relative abundance measure is not theoretically restricted to the unit interval (values greater than one could exist, e.g. if the PepM copies per genome greatly exceeded one), in practice PepM relative abundance was always bounded [0,1] in our datasets. We performed random forest regression using nested ten-fold cross-validation to prevent data leakage to the validation phase (Teschendorff, 2019). We reserved 20% of the data for estimating final model performance. The remaining 80% of training data was split into 10 resampled partitions, each with analysis and assessment partitions, to tune and estimate the performance of preprocessing, supervised feature selection, hyperparameter tuning steps. We tuned hyperparameters (mtry, min.node.size), by maximizing the coefficient of determination from correlation (model R^2) and minimizing the root mean square error (RMSE). We used CAR Scores (Zuber & Strimmer, 2011) for recursive feature elimination to retain only the top 50th percentile of informative variables. We included the feature elimination step to reduce the computation costs and runtime of the feature importance step (see below). Random forest regression was implemented with the package Ranger (Wright & Ziegler, 2015) and the Beta Forest algorithm (Weinhold et al., 2019). We determined predictor variable rankings on the final model from the cross validation step using the Boruta heuristic (Kursa et al., 2010). This step allowed us to identify all predictor variables that consistently performed better than chance and to compare the importance of each variable to a reference importance level, i.e. random data.

3.6.10 Beta-Binomial Regression and Generalized Additive Models

We used the R package corncob (B. D. Martin et al., 2020) and modeled PepM relative abundance directly from PepM read counts and the median read counts to marker gene sets as “successes/total” which is appropriate for the beta-binomial probability distribution. We used the log-odds link function for both relative abundance and the overdispersion parameter. For each taxonomic group (SAR11/*Prochlorococcus*/bacteria and archaea) we modeled only the top five most important biotic/abiotic variables identified in the Random Forest regression and variable importance steps. We did not include additional model terms because multiple collinearities between covariates and the additional model complexity prevented model convergence in most cases. We estimated the probability for each biotic/abiotic covariate being informative to the overall model by using bootstrapped likelihood ratio tests (N=1000). We estimated 95% confidence intervals for model coefficients and standard errors using 1000 random draws from the beta-binomial distribution.

We estimated seasonal effects in time-series metagenomes using Generalized Additive Mixed Models and Linear Mixed-Effect Models implemented through the mgcv v1.8-26 (Wood, 2017) nlme v3.1-148 libraries in R v3.6.2. To decompose any potential seasonal effects, we fit a cyclic spline term to a variable for the day of the year, which we use as a proxy for season, and we fit a

global trend term to the cumulative time since sampling onset. We considered a seasonal effect present if the model term for “day of the year” was statistically significant ($p < 0.05$).

3.6.11 *Prochlorococcus* cultures under P-replete and P-deficient conditions

To investigate the linkage of PepM with phosphonate production, we grew two HLII strains of *Prochlorococcus*: *Prochlorococcus* SB, which has PepM, and the closely related *Prochlororoccus* MIT9301, which lacks the PepM gene sequence. Both strains were grown axenically, under constant light ($30 \mu\text{mol quanta m}^{-2} \text{s}^{-1}$) in artificial seawater medium AMP1 prepared as described before (Moore et al., 2007), but using $3.75 \mu\text{M}$ TAPS as a buffer instead of 1mM HEPES. After growing *Prochlorococcus* SB in the regular medium i.e with a phosphate concentration of $50 \mu\text{M}$ (N/P = 16/1) to assess whether or not this strain could produce phosphonate, we decreased the phosphate concentration to $2.28 \mu\text{M}$ (N/P = 350/1). Although the phosphate concentration was lower, *Prochlorococcus* SB cells were not P-limited in exponential phase growth. However, stationary phase was induced by P-starvation, as demonstrated by the initiation of further growth after adding phosphate to the culture (Figure 3.4A inset). For all experiments, biological duplicates were grown to ensure reproducibility. Cultures axenicity was assessed by flowcytometry and by confirming a lack of turbidity for at least 30 days after inoculation with three test broths: ProAC (J. J. Morris et al., 2008), MPTB (Saito et al., 2002) and ProMM (Pro99 medium (Moore et al., 2007) supplemented with 1xVa vitamin mix (Waterbury & Willey, 1988) and 0.05% w/v each of pyruvate, acetate, lactate and glycerol. ProMM is the 100% seawater based version of the PLAG medium (J. J. Morris et al., 2008). All the glassware and polycarbonate bottles (1 L for the blank, 2 L and 20 L for the cultures) were cleaned by soaking overnight in 2% detergent (micro), rinsed 6 times with deionized water, soaked overnight in 1 M HCl and rinsed 6 times with ultra-high purity water.

3.6.12 Cell harvest and treatment

Prochlorococcus SB cultures grown in high N/P medium were harvested twice: ~6 L were harvested during exponential growth and the remainder (14 L) harvested two days after the onset of stationary phase growth. To ensure that stationary phase cells were limited by phosphorus, 25 mL of culture was amended with phosphate. Fluorescence increased in the phosphate-amended cultures, reaching levels regularly observed in *Prochlorococcus* HLII cultures (Figure 3.4A inset). Cells were separated from the growth medium by centrifugation ($15,970 \text{ rcf}$ for 30 minutes at 4°C) and the growth medium was saved for other analyses. Cell pellets were transferred into 50 mL falcon tubes, suspended in Turk Island mix (Moore et al., 2007) to rinse the cells of external nutrients and centrifuged ($6,523 \text{ rcf}$ for 15 minutes at 15°C). This time, we discarded the supernatant and repeated the operation two more times. After, we flash froze the cell pellets in liquid nitrogen and stored them at -20°C until NMR analyses. We also measured *Prochlorococcus* cell abundance by flow cytometry. Samples were prepared and processed as previously described (Malmstrom et al., 2010; Zinser et al., 2006) a d run on an Guava 12HT

flow cytometer (Luminex Corp., Austin , TX, USA). Cells were excited with a blue 488 nm laser analyzed for chlorophyll fluorescence (692/40nm), SYBR Green I stained DNA fluorescence content (530/40nm), and size (forward scatter). Samples were ran stained with 1x SYBR Green I (Invitrogen, Grand Island, NY) and unstained then incubated for 60 min in the dark prior to running. All flow cytometry files were analyzed using Guavacyte.

3.6.13 Nuclear Magnetic Resonance

NMR spectra were acquired at 25°C on a 400 MHz Bruker AVANCE DPX spectrometer using a 5mm inverse broadband probe and running TOPSPIN 1.3. ³¹P shifts are reported relative to external 85% phosphoric acid at 0 ppm. For the proton-decoupled ³¹P-NMR spectra, we used ‘zgdc30’ with WALTZ16 decoupling and sweep width of 80 ppm, a 3 seconds relaxation delay, 100K scans and 20Hz line broadening. *Prochlorococcus* SB and MIT9301 whole cells were packed into a 5mm BMS tube (Shigemi Inc.) with magnetic susceptibility of the glass inserts matching D₂O.

³¹P-NMR spectra of *Prochlorococcus* SB protein fraction was acquired at 25°C on a 400 MHz Bruker Ascend 400 equipped with a Sample CASE. The ³¹P-NMR spectra were acquired using the program ‘zgpg30’ with a sweep width of 80 ppm, a relaxation delay of 2 seconds, a 15Hz line broadening and for 13K scans. ³¹P chemical shifts are reported relative to external phosphoric acid at 0 ppm.

3.6.14 Elemental composition of *Prochlorococcus* SB and MIT 9301

Elemental C/N/P ratios were measured at the University of Hawai’i nutrient facility according to the protocols employed by the Hawaii Ocean Time series program (<http://hahana.soest.hawaii.edu/hot/protocols/protocols.html#>). Briefly, cell pellets from ~ 900 mL of culture were transferred to combusted glass vials, dried, and powdered. C and N were measured on subsamples using a PE-2400 Carbon/Nitrogen analyzer calibrated with acetanilide standards. Cellular P was measured by the molybdenum blue method (Murphy & Riley, 1962; Strickland & Parsons, 1972) after first combusting cell pellets at 450°C for 3h, and dissolving the residue in 10 mL of 0.5 M HCl.

3.6.15 Separation of cellular macromolecular components

To fractionate *Prochlorococcus* organic matter into different classes of major biochemicals, we followed the protocols of Karl et al. (Karl et al., 1981). Cells from 3 L of culture were centrifuged and the isolated cell pellet extracted with 5 mL of cold 5% trichloroacetic acid (TCA) for 1h. The mixture was centrifuged (12,100 rcf for 30 minutes at 4°C), the supernate decanted, and the TCA insoluble material washed twice with 5% cold TCA. The TCA fractions were combined, evaporated to dryness. Lipids were recovered from the TCA insoluble material

by extraction (3x, room temperature, 20 minutes each) with 95% ethanol (5 mL). Residual ethanol was evaporated, and RNA within the dry TCA/95% ethanol insoluble material hydrolyzed at 37°C for 1 hour with 2.5 mL of 1 M NaOH. The hydrolysis was quenched by immersing the sample tube in an ice bath for 15 minutes, after which the sample was acidified to pH 1 by adding 2.5 mL 1 M HCl and 0.5 mL of 50% TCA. The mixture was allowed to sit for 15 min to precipitate proteins and DNA, then centrifuged for 30 minutes at 12,100 rcf. After collecting the supernatant (containing RNA), we rinsed and centrifuged the pellet 2x with 5 mL of 5% TCA and 2x with 5 mL of 95% ethanol. To hydrolyze DNA, we added 5 mL of 5% TCA to the insoluble material and immersed the tube in boiling water for 30 minutes. After centrifuging the tube and collecting the DNA-containing supernatant, we rinsed and centrifuged the pellet 2x with 5 mL of ice-cold 5% TCA and 2x with ice-cold 95% ethanol as before. Finally, we extracted the protein from the remaining cell pellet with 5 mL of 1 M NaOH (37°C, 18 h). A small amount of insoluble debris remained after the protein extraction. This was removed by centrifugation (12,100 rcf; 30 minutes) and followed by syringe filtration of the supernatant.

3.6.16 Protein extraction and precipitation

Peptides and denatured “soluble” proteins were fractionated from native HMW “insoluble” proteins using the protocol described by Hutchins et al. (Hutchins et al., 2015). The cell pellet of 0.5 L of culture was lysed (15 minutes, RT) with 1 mL of 1% SDS extraction buffer (1% SDS, 0.1 M Tris/HCl pH 7.5, 10 mM EDTA) then heated at 95°C for 10 minutes. Samples were allowed to cool to RT and were then agitated at 350 rpm for 1 hour. The resulting suspension was centrifuged (20 minutes, 14,100 rcf), and the supernatant decanted. After transferring the supernatant containing the proteins, we concentrated the proteins from the supernate by membrane centrifugation using 5 K molecular weight cutoff Vivaspin units of 6 mL (Sartorius Stedim, Goettingen, Germany). The retentate (~300 µL) was recovered, and 1 mL of cold 50/50 methanol/acetone solution (acidified with HCl to a final concentration of 0.5 mM) added. After sitting for 3 days at -20°C, insoluble proteins were pelleted by centrifuge (14,100 rcf for 30 minutes at 4°C) and the supernatant decanted.

3.6.17 Dataset/code availability

The datasets and computer code supporting the findings in this study are available from:

<https://github.com/slhogle/phosphonates>.

The entire MARMICRODB dataset including a comprehensive description, raw protein fasta files, Kaiju v1.6.0 (Menzel et al., 2016) formatted databases, scripts and instructions for how to use the resource is available from <https://doi.org/10.5281/zenodo.3520509>.

GEOTRACES chemical data was processed and matched to metagenome samples using code/methods available from <https://doi.org/10.5281/zenodo.3689249>.

Tara Oceans chemical and hydrographic data was processed and matched to metagenome samples using code/methods available from <https://doi.org/10.5281/zenodo.3786232>.

The list of *Prochlorococcus* core PFAM and TIGRFAM families, a compiled HMMERv3 hidden Markov model database, and a CheckM formatted marker list file is available from <https://doi.org/10.5281/zenodo.3719132>.

3.7 ACKNOWLEDGMENTS

We thank A. Coe, K. Dooley and C. Bliem for their invaluable help with *Prochlorococcus* cultivation; E. M. Grabowski, K. M. Bjorkman and D. M. Karl for their measurements of particulate carbon and phosphorus; C. G. Johnson for his assistance with the NMR. This work was supported in part by grants from the National Science Foundation (OCE-1153588, and DBI-0424599 to S.W.C.; OCE-1335810 and OIA-1826734 to R.S. and OCE-1634080 to D.J.R.), the Gordon and Betty Moore Foundation (#6000 to D.J.R) and the Simons Foundation (Life Sciences Project Award IDs 337262, 647135, S.W.C.; 510023, R.S.; SCOPE Award ID 329108, S.W.C. and D.J.R.). This paper is a contribution from the Simons Collaboration on Ocean Processes and Ecology (SCOPE).

3.8 REFERENCES

- Avrani, S., Wurtzel, O., Sharon, I., Sorek, R., & Lindell, D. (2011). Genomic island variability facilitates *Prochlorococcus*--virus coexistence. *Nature*, *474*(7353), 604–608.
- Barone, B., Coenen, A. R., Beckett, S. J., McGillicuddy, D. J., Weitz, J. S., & Karl, D. M. (2019). The ecological and biogeochemical state of the North Pacific Subtropical Gyre is linked to sea surface height. *Journal of Marine Research*, *77*(2), 215–245.
- Becker, J. W., Hogle, S. L., Rosendo, K., & Chisholm, S. W. (2019). Co-culture and biogeography of *Prochlorococcus* and SAR11. *The ISME Journal*, *13*(6), 1506–1519.
- Berube, P. M., Biller, S. J., Hackl, T., Hogle, S. L., Satinsky, B. M., Becker, J. W., ... Chisholm, S. W. (2018). Single cell genomes of *Prochlorococcus*, *Synechococcus*, and sympatric microbes from diverse marine environments. *Scientific Data*, *5*, 180154.
- Biller, S. J., Berube, P. M., Dooley, K., Williams, M., Satinsky, B. M., Hackl, T., ... Chisholm, S. W. (2018). Marine microbial metagenomes sampled across space and time. *Scientific Data*, *5*(1), 180176.
- Biller, S. J., Berube, P. M., Lindell, D., & Chisholm, S. W. (2015). *Prochlorococcus*: The structure and function of collective diversity. *Nature Reviews. Microbiology*, *13*(1), 13–27.
- Born, D. A., Ulrich, E. C., Ju, K.-S., Peck, S. C., van der Donk, W. A., & Drennan, C. L. (2017). Structural basis for methylphosphonate biosynthesis. *Science*, *358*(6368), 1336–1339.
- Breitbarth, E., Oschlies, A., & LaRoche, J. (2007). *Physiological constraints on the global distribution of Trichodesmium? effect of temperature on diazotrophy*. Retrieved from <https://hal.archives-ouvertes.fr/hal-00297595/>
- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60.
- Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973.
- Carini, P., White, A. E., Campbell, E. O., & Giovannoni, S. J. (2014). Methane production by phosphate-starved SAR11 chemoheterotrophic marine bacteria. *Nature Communications*, *5*, 1–7.
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*. doi:10.1093/bioinformatics/btz848
- Chen, C. C. H., Han, Y., Niu, W., Kulakova, A. N., Howard, A., Quinn, J. P., ... Herzberg, O. (2006). Structure and kinetics of phosphonopyruvate hydrolase from *Variovorax* sp. Pal2: new

- insight into the divergence of catalysis within the PEP mutase/isocitrate lyase superfamily. *Biochemistry*, 45(38), 11491–11504.
- Cioni, J. P., Doroghazi, J. R., Ju, K.-S., Yu, X., Evans, B. S., Lee, J., & Metcalf, W. W. (2014). Cyanohydrin phosphonate natural product from *Streptomyces regensis*. *Journal of Natural Products*, 77(2), 243–249.
- Coleman, M. L., & Chisholm, S. W. (2010). Ecosystem-specific selection pressures revealed through comparative population genomics. *Proceedings of the National Academy of Sciences of the United States of America*, 107(43), 18634–18639.
- Coleman, M. L., Sullivan, M. B., Martiny, A. C., Steglich, C., Barry, K., Delong, E. F., & Chisholm, S. W. (2006). Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science*, 311(5768), 1768–1770.
- Cordero, O. X., & Polz, M. F. (2014). Explaining microbial genomic diversity in light of evolutionary ecology. *Nature Reviews. Microbiology*, 12(4), 263–273.
- Coyne, M. J., Kalka-Moll, W., Tzianabos, A. O., Kasper, D. L., & Comstock, L. E. (2000). *Bacteroides fragilis* NCTC9343 produces at least three distinct capsular polysaccharides: cloning, characterization, and reassignment of polysaccharide B and C biosynthesis loci. *Infection and Immunity*, 68(11), 6176–6181.
- Decho, A. W., & Gutierrez, T. (2017). Microbial Extracellular Polymeric Substances (EPSs) in Ocean Systems. *Frontiers in Microbiology*, 8, 922.
- Ding, W., Baumdicker, F., & Neher, R. A. (2018). panX: pan-genome analysis and exploration. *Nucleic Acids Research*, 46(1), e5.
- Dufresne, A., Ostrowski, M., Scanlan, D. J., Garczarek, L., Mazard, S., Palenik, B. P., ... Partensky, F. (2008). Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biology*, 9(5), R90.
- Dyrhman, S. T., Benitez-Nelson, C. R., Orchard, E. D., Haley, S. T., & Pellechia, P. J. (2009). A microbial source of phosphonates in oligotrophic marine systems. *Nature Geoscience*, 2(10), 696–699.
- Fagan, R. P., & Fairweather, N. F. (2014). Biogenesis and functions of bacterial S-layers. *Nature Reviews. Microbiology*, 12(3), 211–222.
- Faridmoayer, A., Fentabil, M. A., Mills, D. C., Klassen, J. S., & Feldman, M. F. (2007). Functional characterization of bacterial oligosaccharyltransferases involved in O-linked protein glycosylation. *Journal of Bacteriology*, 189(22), 8088–8098.
- Feingersch, R., Philosof, A., Mejuch, T., Glaser, F., Alalouf, O., Shoham, Y., & Béjà, O. (2012). Potential for phosphite and phosphonate utilization by *Prochlorococcus*. *The ISME Journal*, 6(4), 827–834.

- Frost, L. S., Leplae, R., Summers, A. O., & Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nature Reviews. Microbiology*, 3(9), 722–732.
- Granzow, B. N., Sosa, O. A., Gonnelli, M., Santinelli, C., Karl, D. M., and Repeta, D. J. (in review). A sensitive fluorescent assay for measuring carbon-phosphorus lyase activity in aquatic systems. *Limnology and Oceanography, Methods / ASLO*.
- Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., & Beck, E. (2013). TIGRFAMs and Genome Properties in 2013. *Nucleic Acids Research*, 41(Database issue), D387-95.
- Hilderbrand. (2018). *The Role of Phosphonates in Living Systems*. Taylor & Francis.
- Horsman, G. P., & Zechel, D. L. (2017). Phosphonate Biochemistry. *Chemical Reviews*, 117(8), 5704–5783.
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, 34(8), 2115–2122.
- Huerta-Cepas, J., Serra, F., & Bork, P. (2016a). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution*, 33(6), 1635–1638.
- Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., ... Bork, P. (2016b). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*, 44(D1), D286-93.
- Hutchins, D. A., Walworth, N. G., Webb, E. A., Saito, M. A., Moran, D., McIlvin, M. R., ... Fu, F.-X. (2015). Irreversibly increased nitrogen fixation in *Trichodesmium* experimentally adapted to elevated carbon dioxide. *Nature Communications*, 6, 8155.
- Iwashkiw, J. A., Vozza, N. F., Kinsella, R. L., & Feldman, M. F. (2013). Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation. *Molecular Microbiology*, 89(1), 14–28.
- Ju, K.-S., Gao, J., Doroghazi, J. R., Wang, K.-K. A., Thibodeaux, C. J., Li, S., ... Metcalf, W. W. (2015). Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *Proceedings of the National Academy of Sciences of the United States of America*, 112(39), 12175–12180.
- Jürgens, K., & Matz, C. (2002). Predation as a shaping force for the phenotypic and genotypic composition of planktonic bacteria. *Antonie van Leeuwenhoek*, 81(1–4), 413–434.
- Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1), 27–30.

- Karl, D. M., & Lukas, R. (1996). The Hawaii Ocean Time-series (HOT) program: Background, rationale and field implementation. *Deep-Sea Research. Part II, Topical Studies in Oceanography*. Retrieved from <https://www.sciencedirect.com/science/article/pii/0967064596000057>
- Karl, D. M., Winn, C. D., & Wong, D. C. L. (1981). RNA synthesis as a measure of microbial growth in aquatic environments. II. Field applications. *Marine Biology*, *64*(1), 13–21.
- Karl, David M., & Björkman, K. M. (2015). Chapter 5 - Dynamics of Dissolved Organic Phosphorus. In D. A. Hansell & C. A. Carlson (Eds.), *Biogeochemistry of Marine Dissolved Organic Matter (Second Edition)* (pp. 233–334). Boston: Academic Press.
- Karsenti, E., Acinas, S. G., Bork, P., Bowler, C., De Vargas, C., Raes, J., ... Tara Oceans Consortium. (2011). A holistic approach to marine eco-systems biology. *PLoS Biology*, *9*(10), e1001177.
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.
- Kettler, G. C., Martiny, A. C., Huang, K., Zucker, J., Coleman, M. L., Rodrigue, S., ... Chisholm, S. W. (2007). Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genetics*, *3*(12), e231.
- Kolowitz, L. C., Ingall, E. D., & Benner, R. (2001). Composition and cycling of marine organic phosphorus. *Limnology and Oceanography*, *46*(2), 309–320.
- Kursa, M. B., Rudnicki, W. R., & Others. (2010). Feature selection with the Boruta package. *Journal of Statistical Software*, *36*(11), 1–13.
- Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., & Allen, K. N. (2006). Diversification of function in the haloacid dehalogenase enzyme superfamily: The role of the cap domain in hydrolytic phosphorus-carbon bond cleavage. *Bioorganic Chemistry*, *34*(6), 394–409.
- Lees-Miller, R. G., Iwashkiw, J. A., Scott, N. E., Seper, A., Vinogradov, E., Schild, S., & Feldman, M. F. (2013). A common pathway for O-linked protein-glycosylation and synthesis of capsule in *A. cinetobacter baumannii*. *Molecular Microbiology*, *89*(5), 816–830.
- Malmstrom, R. R., Coe, A., Kettler, G. C., Martiny, A. C., Frias-Lopez, J., Zinser, E. R., & Chisholm, S. W. (2010). Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *The ISME Journal*, *4*(10), 1252–1264.
- Martin, B. D., Witten, D., & Willis, A. D. (2020). Modeling microbial abundances and dysbiosis with beta-binomial regression. *The Annals of Applied Statistics*, *14*(1), 94–115.
- Martin, P., Dyhrman, S. T., Lomas, M. W., Poulton, N. J., & Van Mooy, B. A. S. (2014). Accumulation and enhanced cycling of polyphosphate by Sargasso Sea plankton in response to

low phosphorus. *Proceedings of the National Academy of Sciences of the United States of America*, 111(22), 8089–8094.

Martinez, A., Tyson, G. W., & DeLong, E. F. (2010). Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. *Environmental Microbiology*, 12(1), 222–238.

Mende, D. R., Bryant, J. A., Aylward, F. O., Eppley, J. M., Nielsen, T., Karl, D. M., & DeLong, E. F. (2017). Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nature Microbiology*, 2(10), 1367–1373.

Menzel, P., Ng, K. L., & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*, 7, 11257.

Metcalf, W. W., Griffin, B. M., Cicchillo, R. M., Gao, J., Janga, S. C., Cooke, H. A., ... van der Donk, W. A. (2012). Synthesis of methylphosphonic acid by marine microbes: a source for methane in the aerobic ocean. *Science*, 337(6098), 1104–1107.

Metcalf, W. W., & van der Donk, W. A. (2009). Biosynthesis of Phosphonic and Phosphinic Acid Natural Products. *Annual Review of Biochemistry*, 78(1), 65–94.

Milanese, A., Mende, D. R., Paoli, L., Salazar, G., Ruscheweyh, H.-J., Cuenca, M., ... Sunagawa, S. (2019). Microbial abundance, activity and population genomic profiling with mOTUs2. *Nature Communications*, 10(1), 1014.

Milici, M., Vital, M., Tomasch, J., Badewien, T. H., Giebel, H.-A., Plumeier, I., ... Simon, M. (2017). Diversity and community composition of particle-associated and free-living bacteria in mesopelagic and bathypelagic Southern Ocean water masses: Evidence of dispersal limitation in the Bransfield Strait: Bacteria in the deep Southern Ocean. *Limnology and Oceanography*, 62(3), 1080–1095.

Moore, L. R., Coe, A., Zinser, E. R., Saito, M. A., Sullivan, M. B., Lindell, D., ... Chisholm, S. W. (2007). Culturing the marine cyanobacterium *Prochlorococcus*: *Prochlorococcus* culturing. *Limnology and Oceanography, Methods / ASLO*, 5(10), 353–362.

Morris, J. J., Kirkegaard, R., Szul, M. J., Johnson, Z. I., & Zinser, E. R. (2008). Facilitation of robust growth of *Prochlorococcus* colonies and dilute liquid cultures by “helper” heterotrophic bacteria. *Applied and Environmental Microbiology*, 74(14), 4530–4534.

Morris, R. M., Rappé, M. S., Connon, S. A., Vergin, K. L., Siebold, W. A., Carlson, C. A., & Giovannoni, S. J. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature*, 420(6917), 806–810.

Murphy, J., & Riley, J. P. (1962). A modified single solution method for the determination of phosphate in natural waters. *Analytica Chimica Acta*, 27, 31–36.

- Nakamura, Y., Itoh, T., Matsuda, H., & Gojobori, T. (2004). Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nature Genetics*, *36*(7), 760–766.
- Nelson, C. E., Carlson, C. A., Ewart, C. S., & Halewood, E. R. (2014). Community differentiation and population enrichment of Sargasso Sea bacterioplankton in the euphotic zone of a mesoscale mode-water eddy. *Environmental Microbiology*, *16*(3), 871–887.
- Pachiadaki, M. G., Brown, J. M., Brown, J., Bezuidt, O., Berube, P. M., Biller, S. J., ... Stepanauskas, R. (2019). Charting the Complexity of the Marine Microbiome through Single-Cell Genomics. *Cell*, *179*(7), 1623-1635.e11.
- Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*, *36*(10), 996–1004.
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055.
- Partensky, F., Hess, W. R., & Vaultot, D. (1999). *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiology and Molecular Biology Reviews: MMBR*, *63*(1), 106–127.
- Power, P. M., Seib, K. L., & Jennings, M. P. (2006). Pilin glycosylation in *Neisseria meningitidis* occurs by a similar pathway to wzy-dependent O-antigen biosynthesis in *Escherichia coli*. *Biochemical and Biophysical Research Communications*, *347*(4), 904–908.
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. *PloS One*, *5*(3), e9490.
- Quin, L. D., & Williams, A. J. (2004). Practical interpretation of P-31 Nmr Spectra and Computer Assisted Structure Verification (p. 132).
- Repeta, D. J. (2014). Chemical Characterization and Cycling of Dissolved Organic Matter (pp. 21–63).
- Repeta, D. J., Ferrón, S., Sosa, O. A., Johnson, C. G., Repeta, L. D., Acker, M., ... Karl, D. M. (2016). Marine methane paradox explained by bacterial degradation of dissolved organic matter. *Nature Geoscience*, *9*(12), 884–887.
- Rice, K., Batul, K., Whiteside, J., Kelso, J., Papinski, M., Schmidt, E., ... Horsman, G. P. (2019). The predominance of nucleotidyl activation in bacterial phosphonate biosynthesis. *Nature Communications*, *10*(1), 3698.
- Rijkenberg, M. J. A., Middag, R., Laan, P., Gerringa, L. J. A., van Aken, H. M., Schoemann, V., ... de Baar, H. J. W. (2014). The distribution of dissolved iron in the West Atlantic Ocean. *PloS One*, *9*(6), e101323.

- Rivals, I., Personnaz, L., Taing, L., & Potier, M.-C. (2006). Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics*, *23*(4), 401–407.
- Robinson, D. F., & Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*, *53*(1), 131–147.
- Saito, M. A., Moffett, J. W., Chisholm, S. W., & Waterbury, J. B. (2002). Cobalt limitation and uptake in *Prochlorococcus*. *Limnology and Oceanography*, *47*(6), 1629–1636.
- Salt, L. A., Heuven, S., Claus, M. E., Jones, E. M., & Baar, H. J. W. (2015). *Rapid acidification of mode and intermediate waters in the southwestern Atlantic Ocean*. Retrieved from <https://hal.archives-ouvertes.fr/hal-01251672/>
- Sannigrahi, P., Ingall, E. D., & Benner, R. (2006). Nature and dynamics of phosphorus-containing components of marine dissolved and particulate organic matter. *Geochimica et Cosmochimica Acta*, *70*(23), 5868–5882.
- Santoro, A. E., Dupont, C. L., Richter, R. A., Craig, M. T., Carini, P., McIlvin, M. R., ... Saito, M. A. (2015). Genomic and proteomic characterization of “*Candidatus Nitrosopelagicus brevis*”: An ammonia-oxidizing archaeon from the open ocean. *Proceedings of the National Academy of Sciences*, *112*(4), 1173–1178.
- Santoro, A. E., Richter, R. A., & Dupont, C. L. (2019). Planktonic Marine Archaea. *Annual Review of Marine Science*, *11*, 131–158.
- Schlitzer, R., Anderson, R. F., Dodas, E. M., Lohan, M., Geibert, W., Tagliabue, A., ... Zurbrick, C. (2018). The GEOTRACES Intermediate Data Product 2017. *Chemical Geology*, *493*, 210–223.
- Seed, K. D. (2015). Battling Phages: How Bacteria Defend against Viral Attack. *PLoS Pathogens*, *11*(6), e1004847.
- Sosa, O. A., Casey, J. R., & Karl, D. M. (2019a). Methylphosphonate Oxidation in *Prochlorococcus* Strain MIT9301 Supports Phosphate Acquisition, Formate Excretion, and Carbon Assimilation into Purines. *Applied and Environmental Microbiology*, *85*(13). doi:10.1128/AEM.00289-19
- Sosa, O. A., Repeta, D. J., DeLong, E. F., Ashkezari, M. D., & Karl, D. M. (2019b). Phosphate-limited ocean regions select for bacterial populations enriched in the carbon-phosphorus lyase pathway for phosphonate degradation. *Environmental Microbiology*, *21*(7), 2402–2414.
- Speich, S., Chaffron, S., Ardyna, M., Pesant, S., Tara Oceans Consortium, C., & Tara Oceans Expedition, P. (2017). Environmental context of all samples from the Tara Oceans Expedition (2009-2013), about the water column features at the sampling location. PANGAEA.
- Steinberg, D. K., Carlson, C. A., Bates, N. R., Johnson, R. J., Michaels, A. F., & Knap, A. H. (2001). Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-

- scale look at ocean biology and biogeochemistry. *Deep-Sea Research. Part II, Topical Studies in Oceanography*, 48(8), 1405–1447.
- Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11), 1026–1028.
- Strickland, J. D. H., & Parsons, T. R. (1972). *A practical handbook of seawater analysis*. Fisheries Research Board of Canada.
- Temperton, B., Gilbert, J. A., Quinn, J. P., & McGrath, J. W. (2011). Novel analysis of oceanic surface water metagenomes suggests importance of polyphosphate metabolism in oligotrophic environments. *PloS One*, 6(1), e16499.
- Teschendorff, A. E. (2019). Avoiding common pitfalls in machine learning omic data science. *Nature Materials*, 18(5), 422–427.
- Van Mooy, B. A. S., Krupke, A., Dyhrman, S. T., Fredricks, H. F., Frischkorn, K. R., Ossolinski, J. E., ... Sylva, S. P. (2015). Phosphorus cycling. Major role of planktonic phosphate reduction in the marine phosphorus redox cycle. *Science*, 348(6236), 783–785.
- Van Mooy, Benjamin A. S., Rocap, G., Fredricks, H. F., Evans, C. T., & Devol, A. H. (2006). Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23), 8607–8612.
- Waterbury, J. B., & Willey, J. M. (1988). [6] Isolation and growth of marine planktonic cyanobacteria. In *Methods in Enzymology* (Vol. 167, pp. 100–105). Academic Press.
- Weinhold, L., Schmid, M., Mitchell, R., Maloney, K. O., Wright, M. N., & Berger, M. (2019). A Random Forest Approach for Bounded Outcome Variables. *Journal of Computational and Graphical Statistics*, 1–20.
- Whelan, S., & Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution*, 18(5), 691–699.
- White, A. K., & Metcalf, W. W. (2007). Microbial Metabolism of Reduced Phosphorus Compounds. *Annual Review of Microbiology*, 61(1), 379–400.
- Wilkinson, B. J., & Holmes, K. M. (1979). Staphylococcus aureus cell surface: capsule as a barrier to bacteriophage adsorption. *Infection and Immunity*, 23(2), 549–552.
- Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R*, Second Edition. CRC Press.
- Wright, M. N., & Ziegler, A. (2015). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Retrieved from <http://arxiv.org/abs/1508.04409>

- Wyatt, N. J., Milne, A., Woodward, E. M. S., Rees, A. P., Browning, T. J., Bouman, H. A., ... Lohan, M. C. (2014). Biogeochemical cycling of dissolved zinc along the GEOTRACES South Atlantic transect GA10 at 40 S. *Global Biogeochemical Cycles*, 28(1), 44–56.
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. (2017). ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution / British Ecological Society*, 8(1), 28–36.
- Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omics: A Journal of Integrative Biology*, 16(5), 284–287.
- Yu, X., Doroghazi, J. R., Janga, S. C., Zhang, J. K., Circello, B., Griffin, B. M., ... Metcalf, W. W. (2013). Diversity and abundance of phosphonate biosynthetic genes in nature. *Proceedings of the National Academy of Sciences*, 110(51), 20759–20764.
- Zinser, E. R., Coe, A., Johnson, Z. I., Martiny, A. C., Fuller, N. J., Scanlan, D. J., & Chisholm, S. W. (2006). *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. *Applied and Environmental Microbiology*, 72(1), 723–732.
- Zuber, V., & Strimmer, K. (2011). High-Dimensional Regression and Variable Selection Using CAR Scores. *Statistical Applications in Genetics and Molecular Biology*, 10(1). doi:10.2202/1544-6115.1730

3.9 SUPPLEMENTARY MATERIAL

3.9.1 Supplementary Notes

Supplementary Note S3.1: Taxonomic breadth of marine phosphonate producers and consumers

Phosphonate producers in the GORG dataset are more diverse than consumers. We identified phosphonate producers in 30 unique families and 11 different phyla (n domain=2; n phylum=11; n class=12; n order=25; n family=30; n genus=40; total genomes=6720). Producers are from diverse groups including the Alphaproteobacteria, Gammaproteobacteria, Flavobacteriales, Bacteroidia, Actinobacteria, Archaea, and Cyanobacteria, and also uncultured groups like SAR202 and the candidate phyla Marinamargulisbacteria and Marinisomatia (Figure 2). While for phosphonate consumers, we identified 13 unique families and 5 different phyla (n domain=1; n phylum=5; n class=5; n order=10; n family=13; n genus=17; total genomes=6720). However, the distribution of taxonomic ranks for phosphonate producers was less even than for consumers. Over 60% of phosphonate producers were assigned to the SAR11 group, mostly clade Ia.3, while the remaining 40% were evenly distributed amongst low frequency families/orders. The ZD0417 (6%) clade, the OM1 clade (4%) and High-light *Prochlorococcus* (4%) were the next most

abundant producer groups. In contrast, 30% of phosphonate consumers are from the Aegean-169 clade (including *Alphaproteobacterium* HIMB59) (Alonso-Sáez & Gasol, 2007), 21% from the *Rhodobacteraceae*, 18% from SAR11 clade Ia.3, 8% from SAR11 clade IV, and 12% were High-light *Prochlorococcus*.

In MARMICRODB we found phosphonate producers in most of the same taxonomic groups as the GORG database. Unlike GORG, MARMICRODB also includes many genomes isolated from the deep ocean (> 500 meters) and from marine sediments. We found abundant production potential in many deep ocean planktonic groups such as Chloroflexi (including the SAR202 clade) and Marinimicrobia (SAR406). We also found PepM and the Ppd expressed in 24/186 marine transcriptomes from the MMETSP (Keeling et al., 2014) mostly from dinoflagellates as well as in the abundant and cosmopolitan marine Prasinophyte, *Ostreococcus*. Relative abundances estimated from MARMICRODB alone should be viewed with caution since MARMICRODB has a biased composition. However, we note that fewer than 1.5% of the archaeal genomes in MARMICRODB were phosphonate producers (n=18/1256) which is in strong agreement with the distributions in GORG-tropics. Like GORG-tropics, no archaeal genomes from MARMICRODB contained phosphonate catabolism pathways.

Supplementary Note S3.2: Seasonal dependence of phosphonate producers at BATS

We examined the abundance of phosphonate producers in two long-standing ocean time-series: The Bermuda Atlantic Time-series Study in the North Atlantic Ocean (Sargasso Sea) and the Hawaii Ocean Time-series near Hawaii in the North Pacific Ocean. Both study sites are in oligotrophic ocean gyres and have comparable biological productivity and carbon export (Neuer et al., 2002). *Prochlorococcus* and SAR11 are the most abundant planktonic cells in the upper 300 meters at both sites (Giovannoni & Vergin, 2012). However, the two study sites have different seasonal patterns. BATS experience strong convective mixing during the winter months, and HOT is more stably stratified, although HOT also displays seasonal productivity cycles (Karl & Church, 2014). The two study sites also differ in their biogeochemical inventories. Inorganic phosphate concentrations are lower at BATS than HOT which is likely due to the effect of iron supply on dinitrogen fixation rates in each ocean basin (Wu et al., 2000). Indeed, atmospheric iron fluxes are higher at BATS compared with HOT (Boyle et al., 2005; Sedwick et al., 2005). Differences in dissolved phosphorus inventories between HOT and BATS have been previously linked to gene content variation between *Prochlorococcus* populations. *Prochlorococcus* populations in the Sargasso sea are more likely to have additional P acquisition genes such as alkaline phosphatase (Coleman & Chisholm, 2010; Martiny et al., 2006). We detected phosphonate producers at both HOT and BATS at similar abundances as in the rest of the global ocean. For SAR11 and total bacterioplankton these abundances generally increased with depth and were highest in the deep mixed layer below the subsurface deep chlorophyll maximum layer (> 200 meters). *Prochlorococcus* phosphonate producers were at highest abundance in the surface layer (approximately 10 meters) at HOT and BATS and decreased in abundance with depth. At HOT approximately a constant 6-9% of *Prochlorococcus* genomes,

10% of SAR11 genomes, and 10% of total bacterioplankton genomes had a phosphonate production gene cluster in the surface layer. At BATS we detected a significant seasonal effect on the abundance of all phosphonate producers in the surface layer that temporally corresponds with the recurring deepening of the mixed layer during winter months (Supplementary Figure S3.4C). For example, *Prochlorococcus* phosphonate producers are most abundant from February to May then decline to negligible levels in the summer at the onset of peak water column restratification. The seasonality of SAR11 producers and the total community producers was tightly coupled, suggesting that peak producer abundance occurs during periods with the deepest mixed layer (November through the following March). This pattern follows the seasonal succession of ecotypes at BATS where SAR11 clades IIa and IIb are most abundant in the upper 200 meters during the winter (Vergin et al., 2013). SAR11 clades IIa and IIb were also the most abundant producers in GORG-tropics and GORG-BATS after the surface clade Ia, and SAR11 IIb abundance showed a strong positive correlation with phosphonate producers in the global dataset.

Supplementary Note S3.3: Biotic and abiotic covariates shaping the global distribution of phosphonate producers

For *Prochlorococcus* producers we find that a small, but significant, amount of variation is explained by inorganic phosphate concentrations, and the modeled climatological mean of dissolved organic phosphorus (Supplementary Figure S3.3 and S3.5). The models suggest that on average *Prochlorococcus* phosphonate producers are most abundant where inorganic phosphate and dissolved organic phosphorus concentrations are highest. For example, the North Atlantic and Mediterranean had the lowest median abundance of *Prochlorococcus* producers. In the deeply sequenced N. Atlantic sample GORG-BATS we detected no *Prochlorococcus* phosphonate producer genomes and significantly more *Prochlorococcus* phosphonate consumer genomes than in the rest of the global samples. This pattern mirrors the overall depletion of inorganic phosphate relative to other nutrients in the N. Atlantic (Wu et al., 2000). We interpret this as reflecting potential selection against the phosphonate biosynthesis trait in *Prochlorococcus* populations from consistently P-limited regions, such as the tropical N. Atlantic, due to the additional phosphorus cost of phosphonate biosynthesis. The largest proportion of variation for *Prochlorococcus* phosphonate producers is explained by the total abundance of *Prochlorococcus*. That is, *Prochlorococcus* phosphonate producers are more abundant when *Prochlorococcus* constitutes a smaller fraction of the total microbial community. It is unclear what ecological mechanisms underlie this pattern, and it may simply be a consequence of the heteroskedasticity of compositional data, i.e. higher variances as the denominator in the ratio of (PepM/Core genes) approaches zero. The relative abundance of SAR11 phosphonate producers and the bacterioplankton producers were strongly positively correlated, consistent with SAR11 being the dominant phosphonate producer in GORG-tropics and GORG-BATS. Thus, the relative abundances of SAR11 producers and total producers were shaped by many of the same factors. A small but significant

amount of variation in SAR11 phosphonate producers was explained by increasing climatological mean nitrate concentrations and dissolved aluminum concentrations. This reflects that the highest average abundance of SAR11 producers occurred in the North Atlantic and Mediterranean Sea, regions which have higher than average dissolved aluminum concentrations (Orians & Bruland, 1986) and climatological mean nitrate concentrations in the upper 100 meters. Generally, we find that the greatest amount of phosphonate producer variation is explained by multiple depth-dependent biotic factors implying that increasing depth is a unifying master parameter (Figure 3.2, Supplementary Figure S3.2, S3.3 and S3.5, main text). Total bacterial/archaeal and SAR11 phosphonate producers increase alongside depth-associated biotic factors including low light *Prochlorococcus* and the abundance of mesopelagic groups like Archaea (Karner et al., 2001) and SAR11 Iib (Nelson et al., 2014). Phosphonate producers are negatively associated with known surface clades like SAR11 IV and V and abiotic factors like dissolved manganese concentrations (manganese being a surface enriched element (van Hulst et al., 2017)). These modeling results offer a prediction that bulk phosphonate concentration is greatest in the lower reaches of the euphotic zone and upper reaches of the mesopelagic. Future chemical studies targeting the 100-300 meters depth range will hopefully provide greater understanding of phosphonate cycling in the upper ocean.

Supplementary Note S3.4: Evidence for phosphorylated surface layer structures by Prochlorococcus SB

The phosphonate biosynthesis gene cluster contains genes predicted to be involved in cell wall/membrane/envelope biogenesis and extracellular polysaccharide biosynthetic processes (Supplementary Table S3.2), suggesting that *Prochlorococcus* SB may produce extracellular saccharides modified with phosphonate groups. We also found that genes coding for glycosylation reactions were statistically more abundant within ± 5 Kbp of the phosphonate biosynthesis cluster than in the remainder of the complete genomes from *Prochlorococcus* SB and *Pelagibacter* sp. HTCC7217 and RS40 (Supplementary Table S3.2). Many of these genes encode glycosylation enzymes involved in cell membrane/envelope biogenesis (Supplementary Figure S3.7). However, our chemical data shows that phosphonates are associated with a high molecular weight protein fraction. We propose that in *Prochlorococcus* SB phosphonates serve as molecular decorations of glycan chains bound to high molecular membrane-anchored proteins. In glycoproteins sugar polymer chains are covalently attached to amino acid side-chains of a parent protein through post-translational modification.

Glycoproteins were once thought to be exclusive to eukaryotes, but recent efforts indicate that they are ubiquitous in bacteria and archaea (Varki et al., 2016). In bacteria glycans are attached to specific target proteins via both N- and O-linked glycosylation (Nothaft & Szymanski, 2010). Protein substrates for glycosylation reactions include the motor protein flagellins, pilins found in bacterial pilus structures (Nothaft & Szymanski, 2010), and S-layers which are two-dimensional paracrystalline structures that cover the outer surface of bacteria and archaea (Fagan & Fairweather, 2014). Protein glycosylation occurs via two distinct mechanisms: the direct

attachment of glycan monomers or small chains to proteins via glycosyltransferases (common in flagellins and adhesins) and the pre assembly of glycan onto lipid carriers followed by transfer en masse to a protein acceptor (Hug & Feldman, 2011). Recently, it has been discovered that many bacteria also possess general O-linked glycosylation systems capable of promiscuous modification of numerous different types of membrane associated proteins (Fletcher et al., 2009; Ku et al., 2009; Vik et al., 2009).

In *Prochlorococcus* SB, we find genes with homologies to both S-layer biosynthesis and the general O-linked glycosylation system. S-layer biosynthetic clusters encode genes for both biosynthesis and export of lipopolysaccharide anchors (Fagan & Fairweather, 2014) and the biosynthesis and export of an extracellular glycoprotein containing the critical Ca^{2+} - binding hemolysin domain (Bharat et al., 2017; von K ugelgen et al., 2020). In the *Prochlorococcus* SB genome between 1.15 - 1.17 Mbp we find genes encoding putative surface antigens of the filamentous hemagglutinin-glycoprotein family, Ca^{2+} - binding hemolysin domain proteins, and a type-I secretion system ABC transporter, and other antigen-polysaccharide repeats (Supplementary Table S3.2). Most compelling, however, is the presence of general O-linked glycosylation reaction enzymes adjacent to the *Prochlorococcus* SB, *Pelagibacter* HTCC7217 and *Pelagibacter* RS40 phosphonate biosynthesis clusters (Supplementary Figure S3.7). In *Prochlorococcus* SB we find homologs to known genes involved in extracellular capsule synthesis/export, glycan chain extension reactions, and glycosyl transferase reactions. The presence of these capsular biosynthesis enzymes is consistent with the significant cross-talk between lipopolysaccharide biosynthesis and glycoprotein biosynthesis (Hug & Feldman, 2011) and the shared glycan substrates and enzymes between both pathways (Lees-Miller et al., 2013). Most importantly we identify an initiating glycotransferase (PgIC) which links the first glycan subunit to a lipid carrier, a flippase (Wzx) which translocates the assembled glycan-bound lipid carrier to the periplasmic face, and an O-oligosaccharyltransferase (PgIL) which is the critical enzyme moving the assembled glycan chain to the final acceptor protein (Iwashkiw et al., 2013; Lees-Miller et al., 2013). The striking mutual exclusivity we observe between phosphonate production and consumption genes in marine genomes may also be related to the molecular mechanisms and subcellular location of the machinery from both pathways. For example, both the catalytic domains of C-P lyase (Hove-Jensen et al., 2014) and the assembly of glycans (Varki et al., 2016) occur on the cytoplasmic side of the bacterial inner membrane. A cell simultaneously expressing both these systems might cannibalize its own phosphonates. Taken together our findings strongly suggest that *Prochlorococcus* SB has the genomic potential to produce surface-expressed phosphoglycoproteins.

We also note that the low frequency of phosphonate biosynthesis components in *Prochlorococcus*, and SAR11 populations is consistent with the evolutionary mechanism of negative frequency dependent selection (Cordero & Polz, 2014). Negative frequency dependent selection favors the presence of rare traits in a population, and negative frequency dependent patterns are well documented for microbial surface structures including lipopolysaccharides, S-layers, O-antigens, and capsular polysaccharides (Wildschutte et al., 2010). Rare variation in these surface structure traits, for example slight structural modification, can facilitate bacterial evasion from phage and host immune responses (Needham & Trent, 2013; Wildschutte et al., 2004). However, once a variant becomes too abundant within a population then predators, phage or the immune system may adapt to the variant, neutralizing any conferred advantage. Thus, it is theorized that natural selection in these genes results in rapid gain and loss, which ultimately makes these genes and their variants rare in the wider population. The apparent negative-frequency dependence of phosphonate producers within *Prochlorococcus* populations suggests a role for the gene in mediating predator or phage interactions.

Supplementary Note S3.5: Robust quantification of PepM and marker genes using short metagenomic reads

Metagenomic classification of diverse functional genes can present methodological challenges because the amount of discriminative taxonomic signal in these genes may be marginal or confined to specific regions of the gene. For example, PepM appears to be horizontally transferred both within closely related microbial groups like *Prochlorococcus* and between diverse microbial lineages). In some cases, it may be difficult to taxonomically assign a short metagenome read derived from a PepM gene because the read may map equally well to a PepM sequence from multiple taxonomic lineages. We quantified the skill with which we could classify PepM sequences derived from SAR11, *Prochlorococcus*, and all bacteria and archaea genomes using mock metagenomes of known composition.

We simulated mock metagenomes from 200 randomly sampled single cell template genomes from the approximately 13000 genomes from GORG-tropics. The 200 template genomes were sampled from all available taxonomic orders but were restricted to those genomes from each order that fell into the top 10% highest estimated completeness from checkM. Genomes estimated to be below 30% completeness were excluded. These criteria resulted in 74 SAR11 and 48 *Prochlorococcus* template genomes. We simulated 150 bp paired-end Illumina sequencing reads mock metagenomes using randomreads.sh from BMap v37.90 (<https://sourceforge.net/projects/bbmap/>) with 3x coverage, an insert distribution of 155-320 bp and parameters snprate=0.005, insrate=0.0005, delrate=0.0005, and subrate=0.0005. We next included all identified PepM sequences from GORG-tropics to represent the full diversity of PepM sequences in the surface ocean. As a decoy case we also included all GORG genes from the isocitrate-lyase superfamily, which have superficial similarity to PepM but lack the essential conserved active site motif “EDK(X)5NS.” We then cut these authentic and decoy PepM gene sequences with 75 bp upstream and downstream flanking DNA sequence from their parent

genomes and simulated sequencing reads using the same parameters for the full genomes but with a sequencing depth of 30x. This step produced sequencing reads derived from 25, 96, and 14 authentic PepM sequences from *Prochlorococcus*, SAR11, and all remaining bacteria and archaea, respectively. It also generated reads from 49 and 134 isocitrate lyase decoy sequences from bacteria and archaea and SAR11, respectively. The final resulting simulated metagenomic reads were preprocessed and merged as the Tara ocean and bioGEOTRACES metagenomes. Because we knew the precise quantity of PepM-derived reads in the simulated metagenome we could then iteratively calculate the true positive rate, true negative rate, positive predictive value, and the F1 score (harmonic mean of the true positive rate and positive predictive value) for different combinations of sequence similarity cutoffs. In this case the positive predictive value is the probability that a metagenomic read classified as PepM is derived from a PepM sequence, the true positive rate is the percentage of true PepM reads correctly identified as such, and the true negative rate is the percentage of false PepM reads correctly identified. We selected the percent identity and E value classification cutoffs that produced the highest F1 score for PepM. Ultimately, we selected an E value cutoff of $1e-5$ and a percent identity cutoff of 55%, which resulted in a true positive rate of 95% and a true negative rate of 99% for classifying any bacterioplankton PepM (Supplementary Figure S9). For *Prochlorococcus* PepM and SAR11 PepM sequences, respectively, our classification criteria produced a true positive rate of 96%/60% and a true negative rate of 97%/98%. These performance metrics suggest that in all cases we are unlikely to classify superficially related reads as a true PepM sequence. For *Prochlorococcus* and combined bacteria/archaea it appears that we correctly capture 95-96% of the total PepM sequence diversity in the mock metagenomes. The true positive rate for SAR11 was significantly lower (60%) which suggests that 40% of SAR11 PepM sequences are missed and probably assigned only to the bacteria/archaea domain level. This result is consistent with the high sequence similarity between a handful of PepM genes from SAR11 and other diverse bacterial groups, and suggests that our metagenomic database probably underestimates the contribution of SAR11 to the total inventory of PepM genes in the ocean. It probably also accounts for some of the significant positive correlation between SAR11 PepM abundance and combined bacteria/archaea PepM abundance (Supplementary Figure S3.5). Overall, these classification metrics indicate that we are unlikely to misclassify PepM reads from marine metagenomes and suggest our relative abundance estimates are robust.

Supplementary Note S3.6: Random forest regression

We used random forest regression, a nonparametric machine learning approach, to identify biotic and abiotic factors structuring the distributions of phosphonate producers. We selected random forest regression because the dataset has nearly 45 abiotic/biotic variables, many of which are highly correlated, and random forest is capable of handling many predictor variables and is robust to covariate co-linearities. During the cross-validation and tuning phase of the random forest model generation, we performed feature selection heuristics (see methods) to find the top 50th percentile of informative variables for *Prochlorococcus*, SAR11, and combined

bacteria/archaea. We used a forest size of 1000 trees for all models. For *Prochlorococcus* we ultimately selected 10 model covariates, a minimal node size of 2, and 7 randomly sampled variables at each split resulting in a coefficient of determination (R^2_{Pro}) of 0.74 and Root Mean Square Error (RMSE_{Pro}) of 0.04. For SAR11 we selected 15 selected model covariates, a minimal node size of 2, and 6 randomly sampled variables at each split resulting in a coefficient of determination (R^2_{SAR11}) of 0.69 and Root Mean Square Error ($\text{RMSE}_{\text{SAR11}}$) of 0.03. For the combined bacteria/archaea we selected 21 selected model covariates, a minimal node size of 2, and 8 randomly sampled variables at each split resulting in a coefficient of determination (R^2_{combined}) of 0.94 and Root Mean Square Error ($\text{RMSE}_{\text{combined}}$) of 0.02. We then assessed whether these top 50th percentile variables were more informative to the model than including random information, using the Boruta heuristic (Kursa et al., 2010). The Boruta algorithm is a robust, statistically grounded feature ranking method that aims to identify all variables in a dataset that contain useful information for model predictive outcome. This approach is useful when the variables driving the performance model are of interest and not necessarily the ultimate predictive outcome itself. Boruta performs feature ranking by comparing all variables with randomized versions of themselves in an iterative framework. In addition to variable ranking based on relative scores, Boruta, therefore, provides a measure of the significance of each variable relative to noise.

For *Prochlorococcus* and SAR11 there was an appreciable amount of variation in producers that the models could not capture ($R^2_{\text{Pro}}= 0.74$, $R^2_{\text{SAR11}}= 0.69$). There were 10 and 15 covariates that showed better predictive utility than randomly shuffled data for *Prochlorococcus* and SAR11, respectively (Figure S3.3). However, most predictive utility was concentrated into the top 5 variables, while the importance of the other lower ranking variables was in most cases only marginally higher than noise. The five most highly predictive variables for *Prochlorococcus* phosphonate producers (Supplementary Figure S3.3 and S3.5, Supplementary Table S3.1) included inorganic phosphate concentrations, the modeled climatological mean dissolved organic phosphorus concentrations, the subsurface chlorophyll maximum layer type, and the relative abundance of the *Prochlorococcus* LLIV clade. The subsurface chlorophyll maximum layer types 1-4 were characterized by deepening SCML depths (~ 20 to 115 meters) while types 5 and 6 were characterized by high fluorescence throughout the upper 75 meters. The top five variables for SAR11 were the relative abundance of total phosphonate producers in the sample, the climatological mean nitrate concentrations, SAR11 IV relative abundance, SAR11 clade IIb relative abundance, and dissolved aluminum concentrations (Supplementary Figure S3.3 and S3.5, Supplementary Table S3.1). The integrated bacteria/archaea phosphonate producer model performed best ($R^2_{\text{combined}} = 0.94$) and the top five variables (SAR11 clade IIb relative abundance, SAR11 PepM relative abundance, archaea relative abundance, SAR11 clade IV relative abundance, Depth) were strongly informative for the random forest prediction when compared to noise. We also confirmed these factors for all groups using parametric beta-binomial regression (Supplementary Table S1) (B. D. Martin et al., 2020).

3.9.2 Supplementary Tables

Table S3.1: Model coefficients from beta-binomial regression of PepM abundance

Taxonomic group	Term	Estimate	95% CI [lower]	95% CI [upper]	Std. Error	t value	Bootstrapped LRT P value	LRT P value	Term Significance
<i>Prochlorococcus</i>	<i>Prochlorococcus</i> [%]	-0.011	-0.015	-0.007	0.003	-4.433	0.004	4.05E-34	***
<i>Prochlorococcus</i>	Pro LLIV [%]	0.011	0.002	0.018	0.006	1.771	0.077	6.11E-02	.
<i>Prochlorococcus</i>	DOP Clim. avg [umol kg-1]	1.342	0.360	2.385	0.567	2.366	0.009	4.53E-04	*
<i>Prochlorococcus</i>	PO4 [umol kg-1]	-0.014	-0.197	0.167	0.126	-0.107	0.745	7.13E-01	NS
<i>Prochlorococcus</i>	DCM type 2	-0.164	-0.304	-0.043	0.077	-2.129	0.003	6.41E-22	*
<i>Prochlorococcus</i>	DCM type 3	-0.561	-0.710	-0.419	0.085	-6.567	0.003	6.41E-22	***
<i>Prochlorococcus</i>	DCM type 4	-0.250	-0.380	-0.125	0.076	-3.277	0.003	6.41E-22	**
<i>Prochlorococcus</i>	DCM type 5	-0.721	-0.970	-0.473	0.134	-5.387	0.003	6.41E-22	***
<i>Prochlorococcus</i>	DCM type 6	-0.084	-0.210	0.037	0.074	-1.136	0.003	6.41E-22	NS
SAR11	Total PepM [per genome]	0.021	0.018	0.024	0.002	11.502	0.001	6.94E-29	***
SAR11	NO3 Clim. avg [umol kg-1]	0.004	0.001	0.006	0.001	2.640	0.120	2.95E-04	**
SAR11	SAR11 IV [%]	-0.014	-0.021	-0.007	0.004	-3.908	0.062	3.34E-05	***
SAR11	SAR11 IIb [%]	-0.005	-0.007	-0.001	0.002	-2.498	0.003	5.99E-05	*
SAR11	Dissolved Aluminum [nmol kg-1]	0.003	0.001	0.004	0.001	3.370	0.003	1.92E-04	***
All bacterioplankton	SAR11 IIb [%]	0.0114	0.0088	0.0137	0.0017	6.835	< 0.001	6.08E-13	***
All bacterioplankton	SAR11 PepM [per genome]	0.0452	0.0394	0.0506	0.0035	12.901	< 0.001	8.93E-34	***
All bacterioplankton	Archaea [%]	0.0198	0.0165	0.0223	0.0022	9.083	< 0.001	1.70E-21	***
All bacterioplankton	SAR11 IV [%]	-0.0148	0.0209	-0.0099	0.0035	-4.237	< 0.001	2.31E-08	***
All bacterioplankton	Pro LLIV [%]	0.0047	0.0031	0.0066	0.0016	2.902	< 0.001	1.88E-07	**
All bacterioplankton	Depth [m]	0.0006	0.0004	0.0008	0.0001	4.117	< 0.001	1.88E-07	***

Coefficient estimates are of unit log-odd change in the probability of sampling a PepM read due to a one unit covariate change while controlling for the other covariates. For the categorical variable DCM type the coefficient estimate is from the reference level DCM type 1. 95% confidence intervals are simulated from 1000 random draws from the beta-binomial distribution parameterized from the covariate data. Each covariate is tested for statistical significance using a parametric bootstrapped likelihood ratio test (N=1000) and a classic likelihood ratio (LRT) test using a canonical Chi-squared distribution. Term significance: *** $P < 0.001$, ** $0.001 < P < 0.01$, * $0.01 < P < 0.05$, . $0.05 < P < 0.1$, NS $P > 0.1$. *Prochlorococcus* and Archaea are percentage of all successfully mapped reads to MARMICRODB. Clades LLIV, SAR11 IV, SAR11 IIb are percentage of reads mapped to *Prochlorococcus* and SAR11, respectively. Clim. avg, climatological average from the MIT Darwin model; DCM depth range, deep chlorophyll maximum layer depth category – 1=shallow, 6=deep.

Table S3.2: Functional enrichment of polysaccharide biosynthesis genes within phosphonate biosynthesis gene clusters

Genome	Kegg Pathway	Gene Ratio	Background Ratio	p.adjust	qvalue	KO	Gene function Description
SB	map00520	6/16	24/754	7×10 ⁻⁵	6×10 ⁻⁵	K02377, K02377	Nad-dependent epimerase dehydratase
SB	map00520	6/16	24/754	7×10 ⁻⁵	6×10 ⁻⁵	K01784	udp-glucose 4-epimerase
SB	map00520	6/16	24/754	7×10 ⁻⁵	6×10 ⁻⁵	K01710, K08679	Nucleotide sugar epimerase
SB	map00520	6/16	24/754	7×10 ⁻⁵	6×10 ⁻⁵	K01711	Gdp-mannose 4,6-dehydratase
SB	map00520	6/16	24/754	7×10 ⁻⁵	6×10 ⁻⁵	K00971, K16011	Mannose-1-phosphate guanylyltransferase
SB	map00520	6/16	24/754	7×10 ⁻⁵	6×10 ⁻⁵	K00012	Udp-glucose 6-dehydrogenase

RS40	map00520	6/28	29/755	1×10 ⁻²	1×10 ⁻²	K01784	epimerase dehydratase
RS40	map00520	6/28	29/755	1×10 ⁻²	1×10 ⁻²	K13010	DegT/DnrJ/EryC1/StrS aminotransferase family
RS40	map00520	6/28	29/755	1×10 ⁻²	1×10 ⁻²	K01654, K18430	synthase
RS40	map00520	6/28	29/755	1×10 ⁻²	1×10 ⁻²	K01791, K18429	UDP-N-acetylglucosamine 2-epimerase
RS40	map00520	6/28	29/755	1×10 ⁻²	1×10 ⁻²	K00966, K16881	Glucose-1-phosphate cytidyltransferase
RS40	map00520	6/28	29/755	1×10 ⁻²	1×10 ⁻²	K01654, K18430	synthase

Subset of results from a Universal Enrichment Analysis implemented with the bioconductor package clusterProfiler. The remainder of the results for *Prochlorococcus* and SAR11 are available from <https://github.com/slhogle/phosphonates>. SB, *Prochlorococcus* sp. SB; RS40, Candidatus Pelagibacter sp. RS40; map00520, Amino sugar and nucleotide sugar metabolism; KO, Kegg Orthology identifier; p.adjust, Bounded False Discovery Rate (FDR) adjusted probability for hypergeometric test statistic under the null hypothesis; qvalue, direct estimate of the FDR associated with p.adjust; Gene Ratio, k/n where n is the total number of genes within 10000 nucleotides upstream/downstream of the PepM gene and k is the number of genes in that region annotated to the

Kegg Pathway; Background ratio, M/N where N is the total number of genes with a Kegg Orthology assignment and M is the number of Kegg Orthology assigned genes annotated to the Kegg Pathway.

Table S3.3: *Prochlorococcus* SB phosphonate (Phn) and phosphate (Ph) cellular proportions

Harvest stage	#	Phn proportion	Ph proportion
<i>Prochlorococcus</i> SB	1	0.41	0.59
P-replete	2	0.43	0.57
(exponential phase)	mean	0.42	0.58
<i>Prochlorococcus</i> SB	1	0.70	0.30
P-limited	2	0.71	0.29
(stationary phase)	mean	0.71	0.29

³¹P-NMR spectra of each conditions and each replicates were manually integrated and those values used to calculate the proportion of phosphonate and phosphate in the cells.

Table S3.4: *Prochlorococcus* SB phosphorus (P), phosphonate (Phn) and phosphate (Ph) to carbon ratios

Harvest stage	#	P/C	Phn/C	Ph/C
<i>Prochlorococcus</i> SB	1	0.009 ± 0.001	0.0036 ± 0.0005	0.0052 ± 0.0007
P-replete	2	0.0057 ± 0.0005	0.0024 ± 0.0002	0.0033 ± 0.0003
(exponential phase)	mean	0.007 ± 0.001	0.0030 ± 0.0004	0.0042 ± 0.0007
<i>Prochlorococcus</i> SB	1	0.008 ± 0.001	0.0059 ± 0.0009	0.0025 ± 0.0004
P-limited	2	0.0070 ± 0.0002	0.0050 ± 0.0002	0.00203 ± 0.00007
(stationary phase)	mean	0.0076 ± 0.0005	0.0054 ± 0.0003	0.0047 ± 0.0002

The proportions of phosphonate and phosphate in the cells and the P/C ratio are used to calculate the Phn/C and C/Ph ratios. Errors associated the P/C ratio are calculated using the errors associated with the %C and %P in the

cells and as: $\Delta \frac{P}{C} = \frac{P}{C} \times \sqrt{\left(\frac{\Delta \%C}{\%C}\right)^2 + \left(\frac{\Delta \%P}{\%P}\right)^2}$. Errors associated with Phn/C and Ph/C ratios are calculated as: $\Delta \frac{Phn}{C} =$

$\frac{Phn}{C} \times \sqrt{\left(\frac{\Delta P/C}{P/C}\right)^2}$. Errors associated with the mean values are calculated using the standard error of the mean.

Table S3.5: Comparison of genome, PepM, and random tree topologies

Tree	RF	Max RF	Norm RF	Eff tree size	Ref in Src	Src in Ref
Gphylo vs Gphylo	0	1984	0	999	1	1
PepM vs Gphylo	1372	1926	0.71	999	0.64	0.66
Rtrees vs Gphylo (n=15)	1708 ± 18	1988	0.86 ± 0.01	999	0.570 ± 0.005	0.570 ± 0.005

Gphylo vs Gphylo	0	26	0	16	1	1
PepM vs Gphylo	22	26	0.85	16	0.61	0.61
Rtrees vs Gphylo (n=15)	21.8 ± 1.9	26	0.84 ± 0.08	16	0.61 ± 0.04	0.61 ± 0.04

When compared to the concatenated genome phylogeny (Gphylo), 10 trees simulated at random (Rtrees) have similar Robinson-Foulds distances to the PepM phylogeny (PepM) indicating that the PepM evolutionary history corresponds poorly to that of the core genome in both archaea and bacteria. The upper half of the table compared the bacterial genome phylogeny while the lower half compares the archaeal genome phylogeny. Column descriptions: Tree, tree comparison performed; RF, calculated Robinson-Foulds metric; Max RF, maximum possible RF distance for the comparison; Norm RF, for normalized comparison between different tree comparisons (RF/Max RF); Eff tree size, effective tree size used for the calculation after pruning unshared leaves; Ref in Src, normalized number of splits from Gphylo found in either PepM or Rtrees; Src in Ref, normalized number of splits from either PepM or Rtrees found in Gphylo.

3.9.3 Supplementary Figures

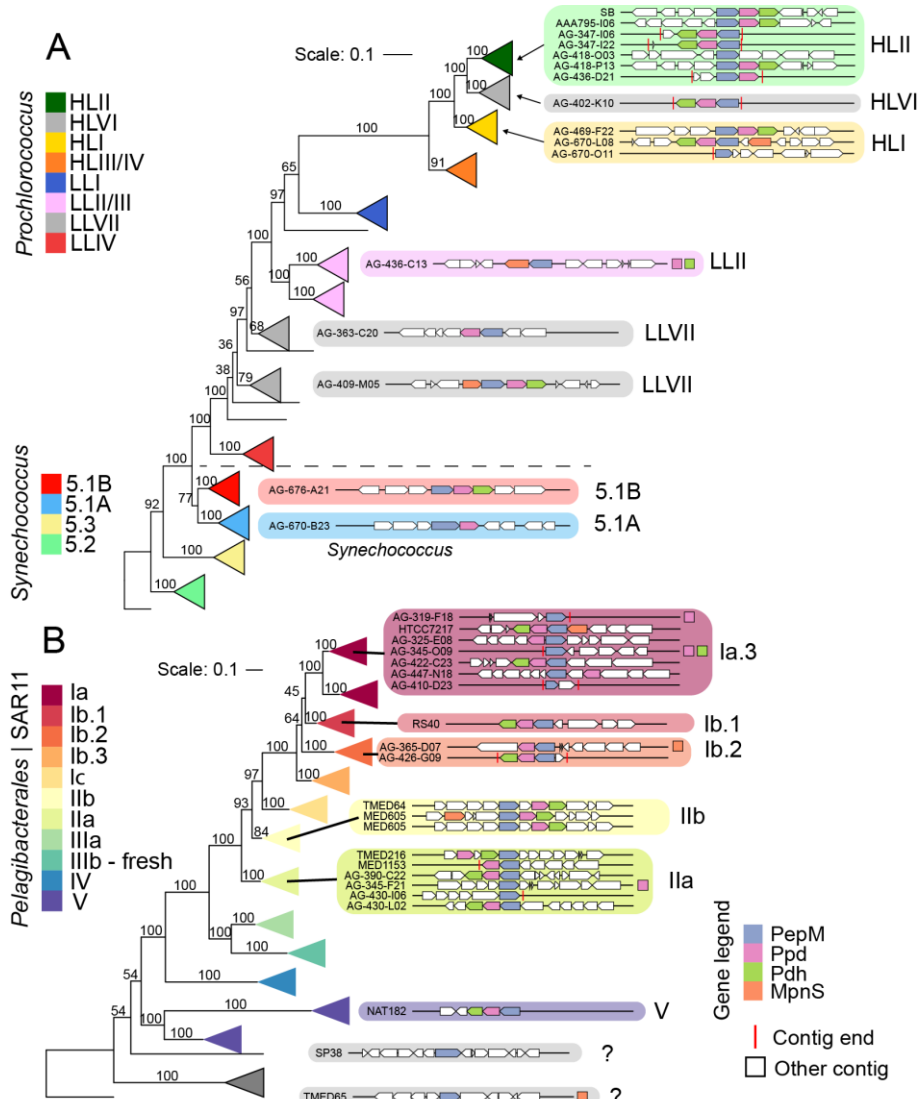


Figure S3.1: A) *Prochlorococcus* and B) SAR11 phosphonate producers

Phylogenies are constructed from 120 concatenated, single-copy marker genes from *Prochlorococcus* (112 genomes) and *Synechococcus* (23 genomes) and SAR11 (92 genomes). Scale bar is 0.1 amino acid substitutions. The cyanobacteria tree is rooted at *Synechococcus* sp. WH5701 and the SAR11 tree is rooted at TMED13, a metagenome-assembled genome from Tara Oceans distantly related to ‘Candidatus Pelagibacter ubique.’ Bootstrap values are based on 250 resamplings. Monophyletic ecotypes are represented by colored cartoon triangles. Inclusion of new genomes in this study resulted in a polyphyletic origin for the original SAR11 clade Ib, therefore we have partitioned clade Ib into 3 monophyletic subclades (Ib.1, Ib.2 and Ib.3). Phosphonate biosynthetic gene cluster diagrams for all genomes from each ecotype are displayed to the right. Vertical red lines represent contig breaks in the gene diagrams and filled squares next to the diagram indicate that the gene was detected on a separate noncontiguous contig. For simplification, MpnS category also includes the functionally related enzyme HepDI.

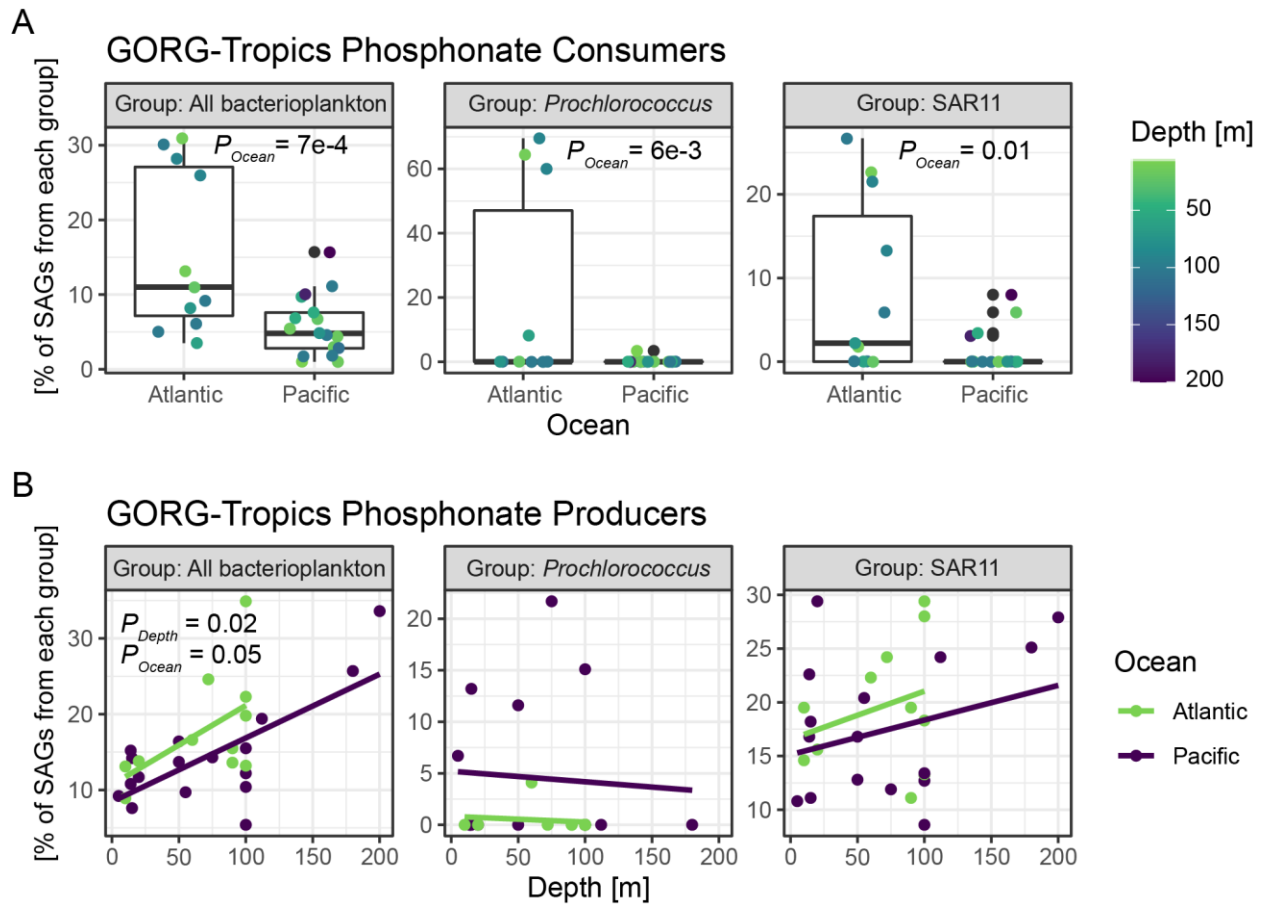


Figure S3.2: Phosphonate consumers and producers in 28 GORG-Tropics samples

Same as in Figure 2.2 C,D but includes *Prochlorococcus* and SAR11. Proportions [%] are total producers or consumers divided by the total number GORG assemblies and are corrected using the estimated sequence recovery from assemblies (see methods). P values for covariates in beta-binomial regression are displayed if statistically significant ($P \leq 0.05$). Beta-binomial regression all consumers; Ocean - Est=-1.04, Err=0.27, $t=-3.81$, $P=7e-4$; link=logit; log L = -80.425, df=4, resid df=24. Beta-binomial regression *Prochlorococcus* consumers; Ocean - Est=-3.50, Err=1.15, $t=-3.04$, $P=6e-3$; link=logit; log L = -20.181, df=4, resid df=21. Beta-binomial regression SAR11 consumers; Ocean - Est=-1.73, Err=0.65, $t=-2.66$, $P=0.01$; link=logit; log L = -41.916, df=4, resid df=24. Beta-binomial regression all producers; Depth - Est=0.0036, Err=0.0014, $t=2.604$, $P=0.02$; Ocean - Est=-0.30, Err=0.14, $t=-2.08$, $P=0.05$; link=logit; log L=-78.179, df=4, resid df=24.

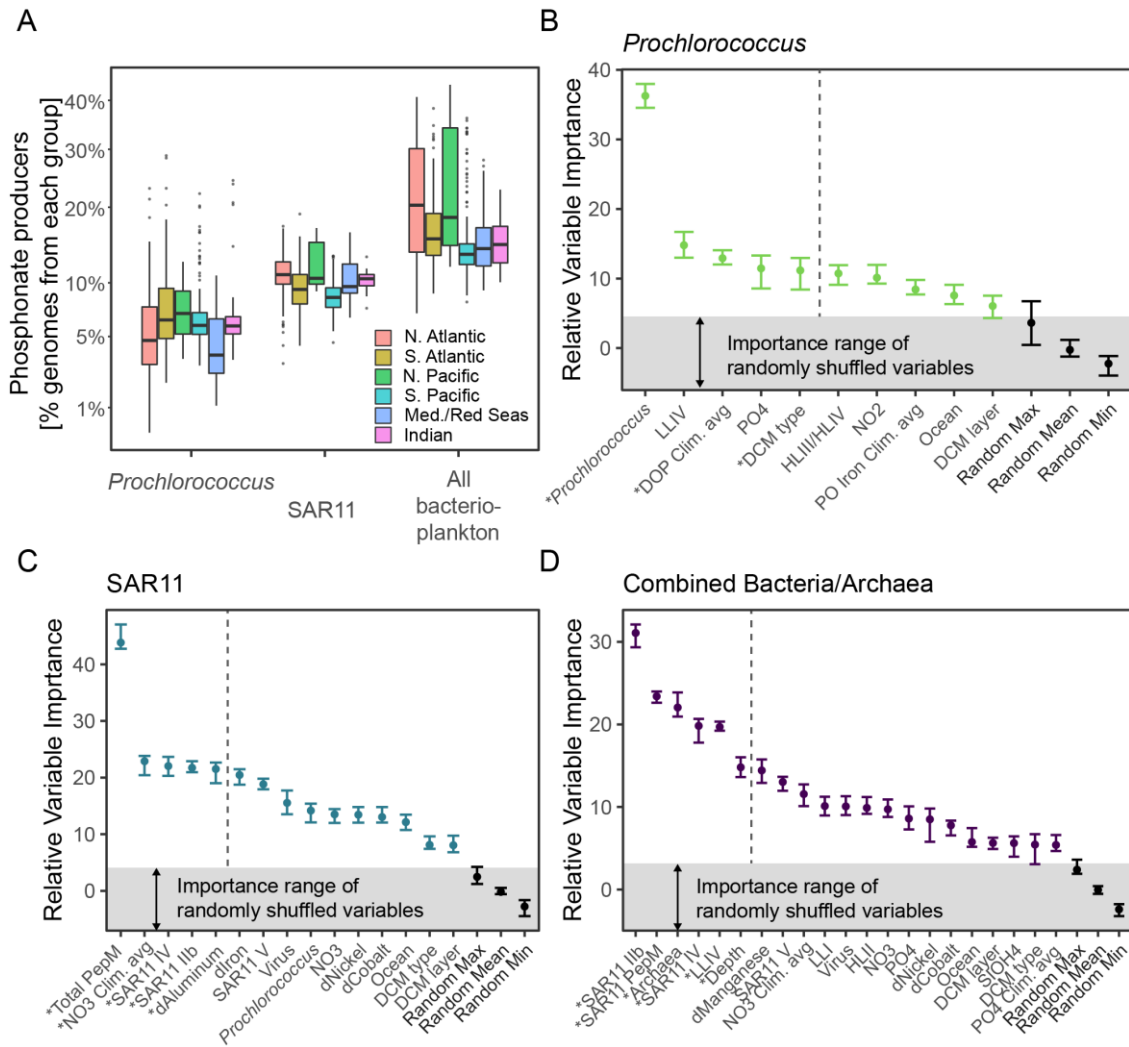


Figure S3.3: Environmental factors driving the relative abundance of PepM in *Prochlorococcus*, SAR11, and combined Bacteria/Archaea

A) Same as in Figure 3A but binned by ocean region. There is no statistically significant difference between ocean regions for *Prochlorococcus*, SAR11, or combined Bacteria/Archaea PepM relative abundance. Black line is the median value, the box shows the 25th and 75th percentiles, and the whiskers extend to 1.5 of the interquartile range for n metagenome samples in each group. Individual points are outliers, and the vertical axis is square-root transformed. B-D) Random forest variable importance measures for each taxonomic group. Point ranges show median and maximum/minimum for 100 random permutations of the variable importance algorithm. Only informative variables for predicting PepM distributions are displayed. Randomized controls in black represent the distribution of the median, minimum, and maximum importance of randomly shuffled variables at each algorithm iteration and can be interpreted as the baseline variable importance expected simply due to chance. Asterisks indicate whether a variable was also statistically significant in a parametric beta-binomial regression (Supplementary Table S1). The top five predictive variables are demarcated by a dashed line. *Prochlorococcus*, Archaea, and viruses are units percentage of all successfully mapped reads to MARMICRODB. Clades HLII, LLI, LLIV, HLIII/HLIV, SAR11 IV, SAR11 IIb, SAR11 V, are percentage of reads mapped to *Prochlorococcus* and SAR11, respectively. Clim. avg, climatological average from the MIT Darwin model; DCM depth range, deep chlorophyll maximum layer depth category – 1=shallow, 6=deep. PepM units are the percentage of *Prochlorococcus*, SAR11, or all genomes with the trait.

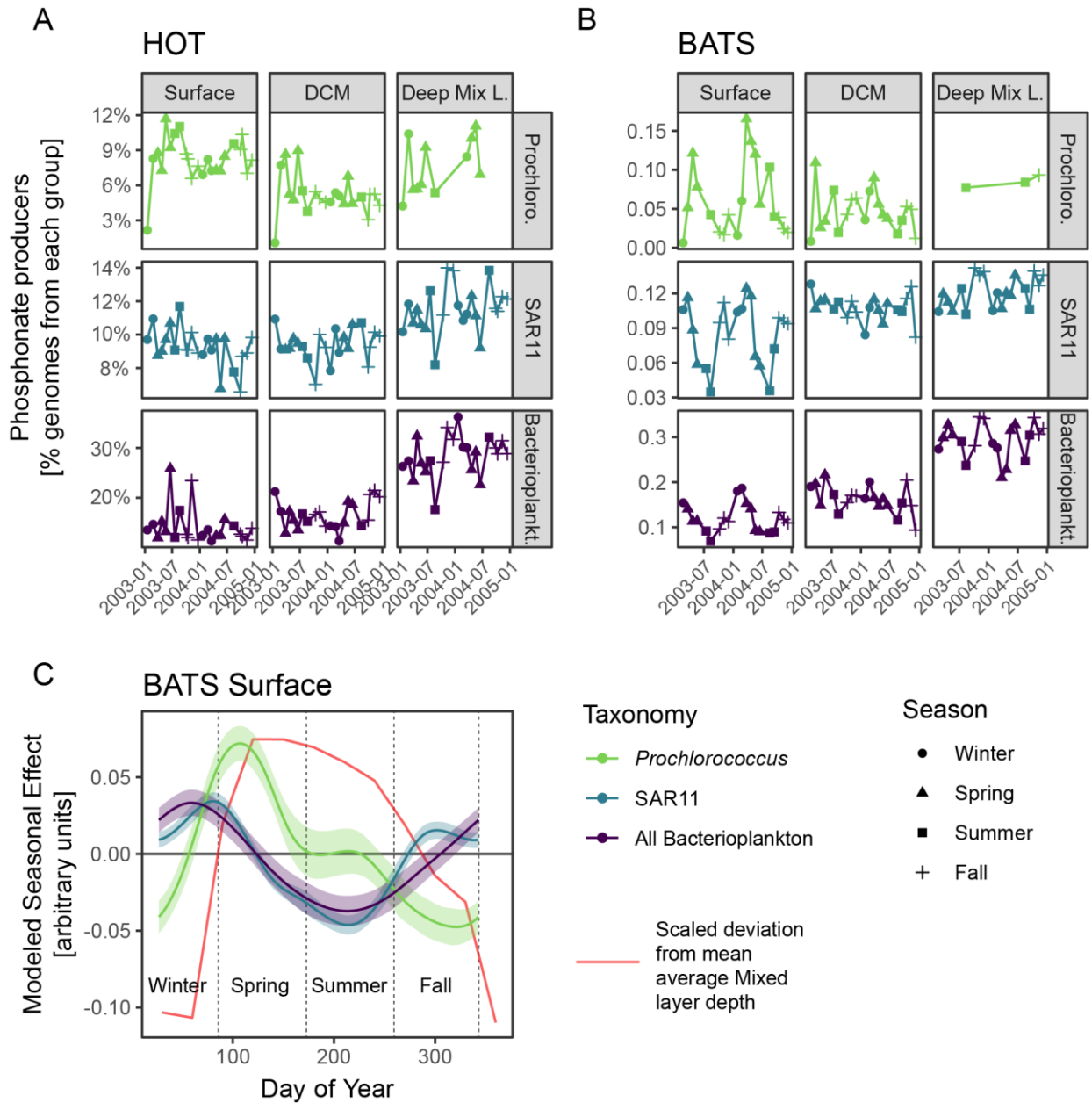


Figure S3.4: Time series of PepM relative abundance at the Hawaii Ocean Time-series (HOT) and the Bermuda Atlantic Time Series (BATS)

PepM relative abundance shown by depth range (horizontal grid axis, DCM = deep chlorophyll maximum layer, Deep Mix L. = deep mixed layer, 1% photosynthetically active radiation) and by taxonomic grouping (vertical grid at A) HOT and B) BATS. Horizontal axis shows time and points are shaped by sampling season. C) The modeled seasonal effect at BATS surface on the relative abundance of PepM of each taxonomic grouping. [GAM SAR11, family=Gamma, link=identity: $s(\text{dayOfYear})$, edf=5.4, Ref. df=8, $F=9.8$, $P=7e-5$. GAM *Prochlorococcus*, family=Gamma, link=identity: $s(\text{dayOfYearDCM})$, edf=3.8, Ref. df=8, $F=4.5$, $P=0.001$, GAM Bacterioplankton, family=Gamma, link=identity: $s(\text{dayOfYearDCM})$, edf=3.2, Ref. df=8, $F=5.0$, $P=2e-4$]. The red line shows the scaled deviation from the annual mean of the mixed layer depth at BATS for the time range in B). For example, a negative deviation indicates that the mixed layer depth is deeper than the annual mean.

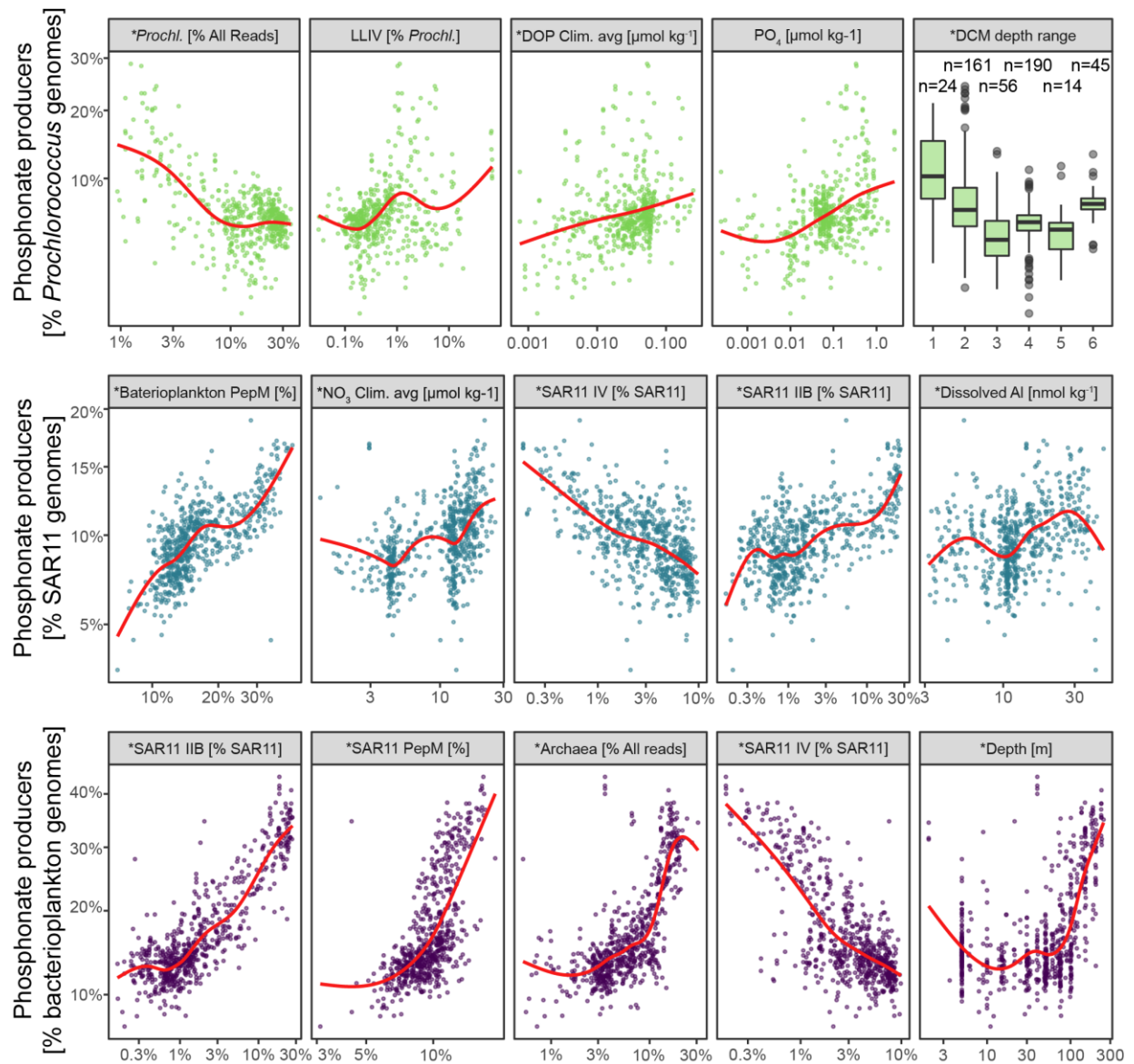


Figure S3.52: Relationship between *Prochlorococcus*, SAR11, and combined Bacteria/Archaea PepM relative abundance and the top five biotic/abiotic covariates for each taxonomic group

Phosphonate producer relative abundance (vertical axis) vs environmental features (horizontal axis) for the top five most informative variables from random forest regressions of each taxonomic group (Figure 3). Each point is a metagenomic observation. Climatological averages are from the MIT Darwin model while the other biotic/abiotic variables are in situ measurements from GEOTRACES and Tara Oceans. Red lines are local polynomial regression fits to the data. Asterisks indicate whether a variable was also statistically significant in a parametric beta-binomial regression (Supplementary Table S1). *Prochlorococcus* and Archaea are percentage of all successfully mapped reads to MARMICRODB. Clades LLIV, SAR11 IV, SAR11 IIB are percentage of reads mapped to *Prochlorococcus* and SAR11, respectively. Clim. avg, climatological average from the MIT Darwin model; DCM depth range, deep chlorophyll maximum layer depth category – 1=shallow, 6=deep. PepM units are the percentage of *Prochlorococcus*, SAR11, or all genomes with the trait.

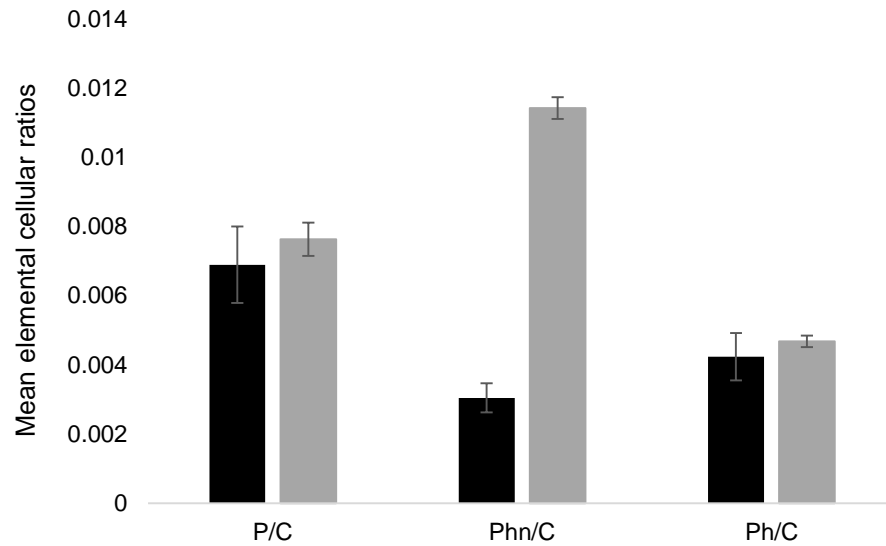


Figure S3.63: Variations of *Prochlorococcus* SB P/C, Phn/C and Ph/C ratios between in exponential (P-replete; black) and stationary phase (P-limited; grey)

When *Prochlorococcus* SB cells become P-limited, their P/C ratio increases slightly whereas their Phn/C ratio increases and their Ph/C ratio decreases indicating an increase in phosphonate content and a decrease in phosphate (Supplementary Table S4). As before, errors associated with the mean values are calculated using the standard error of the mean.

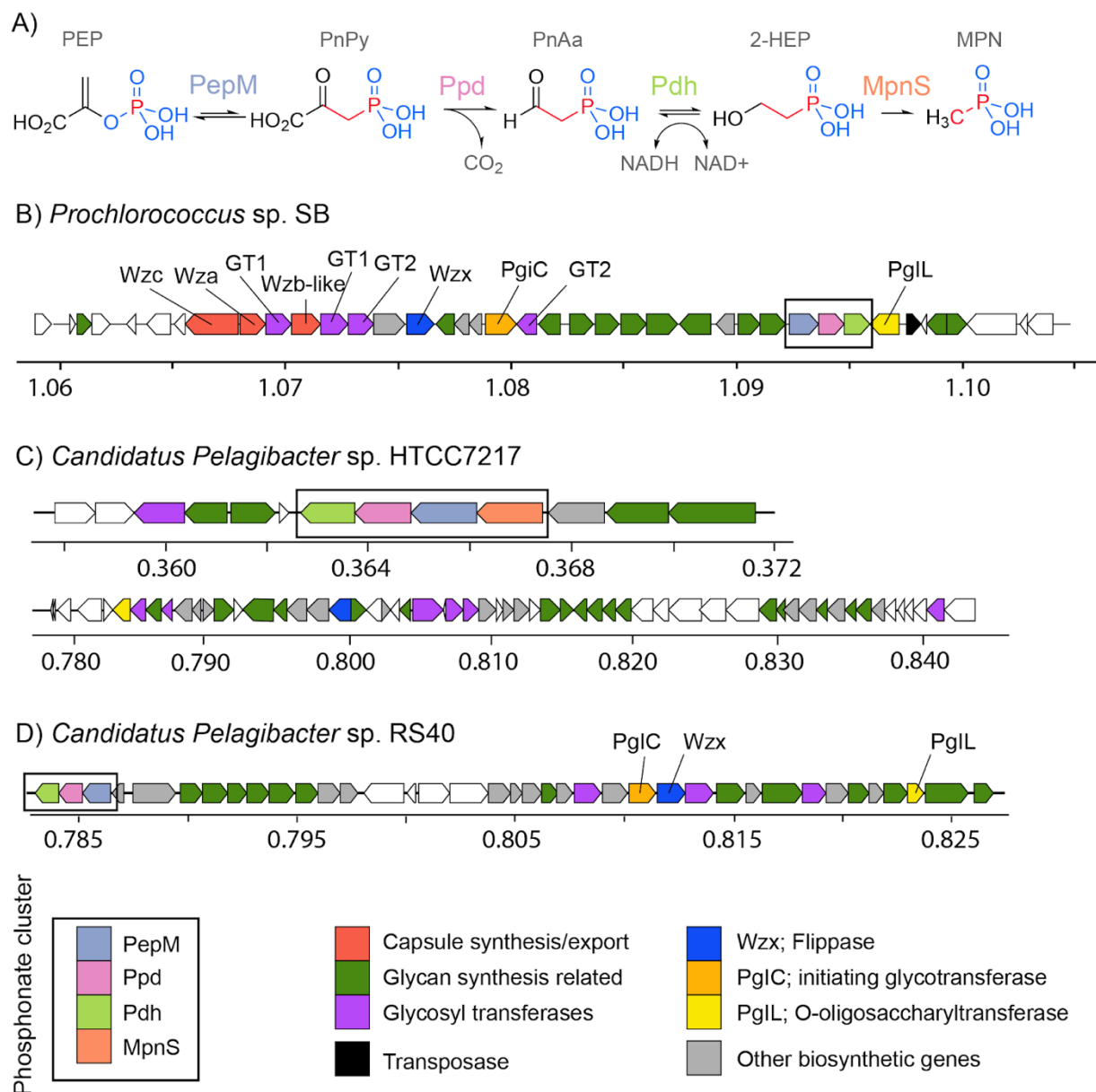


Figure S3.74: Polysaccharide and phosphonate biosynthesis in *Prochlorococcus* and SAR11 isolate genomes

A) same reaction diagram as in Figure 1 with chemical intermediates (black) and catalyzing enzymes (color). B-D) Putative phosphonoglycoprotein biosynthesis gene clusters in *Prochlorococcus* SB, *Candidatus Pelagibacter* sp. HTCC7217, and *Candidatus Pelagibacter* sp. RS40 genomes annotated by antiSMASH (Blin et al., 2019) and eggNOG mapper (Huerta-Cepas et al., 2017). Genome coordinates are $\times 10^6$ base pairs. Phosphonate biosynthetic genes are colored by enzyme as in A) and are shown in black boxes. Genes involved in modular biosynthesis of sugars are shown in red, green, and purple. PgiC is the putative initiating glycotransferase, which links the first glycan subunit to a lipid carrier, Wzx is a flippase which translocates the assembled glycan-bound lipid carrier to the periplasmic face, and PglL is an O-oligosaccharyltransferase which is the critical enzyme moving the assembled glycan chain to the final acceptor protein. Other biosynthetic genes (grey) have functional predictions not directly related to glycan biosynthesis.

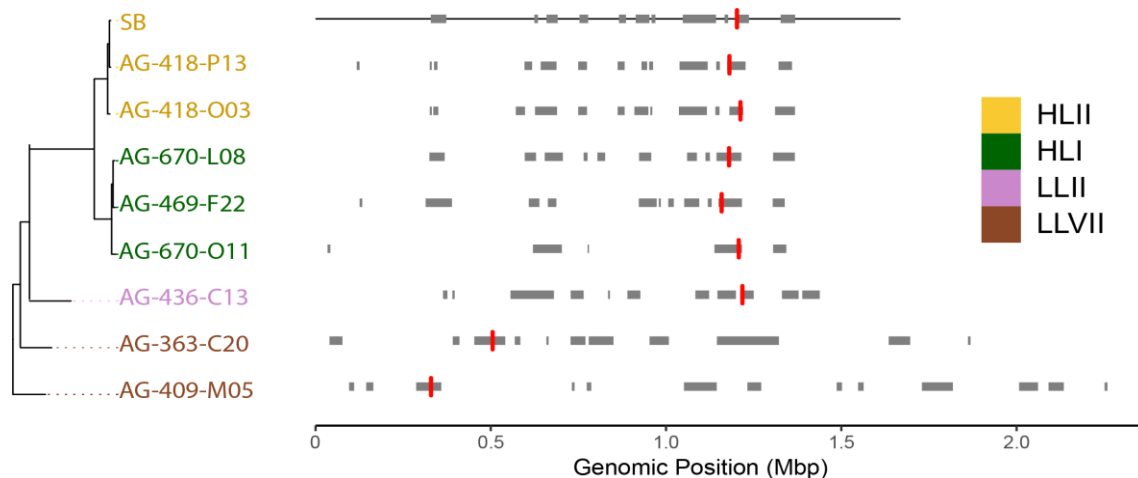


Figure S3.85: Distribution of phosphonate biosynthesis gene clusters within *Prochlorococcus* genomic islands

Includes a subset of all *Prochlorococcus* genomes with i) high enough completeness and low enough contig fragmentation for island prediction (see methods) and ii) that contain phosphonate biosynthesis genes. Genomes are ordered by phylogeny (left) and colored by ecotype/clade (right). Grey bars denote the location of predicted genomic islands and red vertical lines denote the location of phosphonate biosynthesis clusters within predicted genomic islands. *Prochlorococcus* SB genome assembly length is shown as a black horizontal line. The assembly lengths of single cell genomes are omitted due to incompleteness and contig fragmentation. Single cell genome contigs and genomic islands are arbitrarily ordered to *Prochlorococcus* SB.

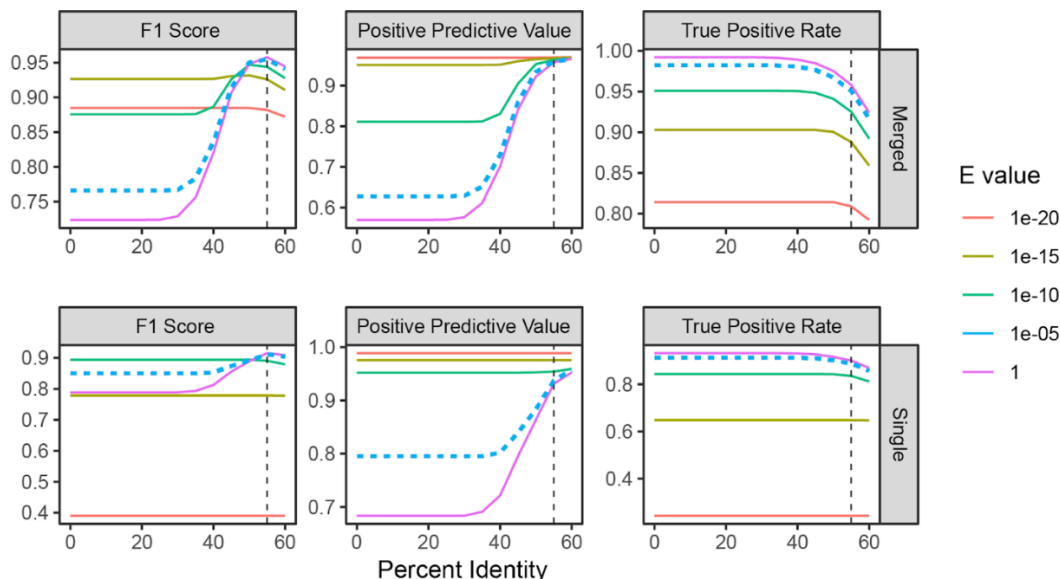


Figure S3.96: Empirical determination of similarity cutoffs for identifying PepM-derived metagenomic short sequencing reads

F1 score, Positive Predictive Value, and True Positive Rate for classifying PepM from mock metagenomes simulated from 150 bp Illumina paired-end reads. Results are shown both for merged read pairs (top), and the best scoring read from an unmerged read pair (bottom). The True Positive Rate measures the proportion of “true” simulated PepM reads correctly identified by a E value/Percent Identity cutoff combination while Positive Predictive Value measures the proportion of “true” simulated PepM reads that are correctly identified out of all reads passing the E value/Percent Identity cutoff combination. The F1 score is the harmonic mean of precision and sensitivity. Cutoffs were empirically chosen to maximize the F1 score resulting in an E value threshold of 1e-5 (dashed blue line) and a Percent Identity cutoff of 55% (dashed vertical line).

3.9.4 Supplementary references

- Alonso-Sáez, L., & Gasol, J. M. (2007). Seasonal variations in the contributions of different bacterial groups to the uptake of low-molecular-weight compounds in Northwestern Mediterranean coastal waters. *Applied and Environmental Microbiology*, 73(11), 3528–3535.
- Bharat, T. A. M., Kureisaite-Ciziene, D., Hardy, G. G., Yu, E. W., Devant, J. M., Hagen, W. J. H., ... Löwe, J. (2017). Structure of the hexagonal surface layer on *Caulobacter crescentus* cells. *Nature Microbiology*, 2, 17059.
- Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., ... Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*, 47(W1), W81–W87.
- Boyle, E. A., Bergquist, B. A., Kayser, R. A., & Mahowald, N. (2005). Iron, manganese, and lead at Hawaii Ocean Time-series station ALOHA: Temporal variability and an intermediate water hydrothermal plume. *Geochimica et Cosmochimica Acta*, 69(4), 933–952.
- Coleman, M. L., & Chisholm, S. W. (2010). Ecosystem-specific selection pressures revealed through comparative population genomics. *Proceedings of the National Academy of Sciences of the United States of America*, 107(43), 18634–18639.
- Cordero, O. X., & Polz, M. F. (2014). Explaining microbial genomic diversity in light of evolutionary ecology. *Nature Reviews. Microbiology*, 12(4), 263–273.
- Fagan, R. P., & Fairweather, N. F. (2014). Biogenesis and functions of bacterial S-layers. *Nature Reviews. Microbiology*, 12(3), 211–222.
- Fletcher, C. M., Coyne, M. J., Villa, O. F., Chatzidaki-Livanis, M., & Comstock, L. E. (2009). A general O-glycosylation system important to the physiology of a major human intestinal symbiont. *Cell*, 137(2), 321–331.
- Giovannoni, S. J., & Vergin, K. L. (2012). Seasonality in ocean microbial communities. *Science*, 335(6069), 671–676.
- Hove-Jensen, B., Zechel, D. L., & Jochimsen, B. (2014). Utilization of glyphosate as phosphate source: biochemistry and genetics of bacterial carbon-phosphorus lyase. *Microbiology and Molecular Biology Reviews: MMBR*, 78(1), 176–197.
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, 34(8), 2115–2122.
- Hug, I., & Feldman, M. F. (2011). Analogies and homologies in lipopolysaccharide and glycoprotein biosynthesis in bacteria. *Glycobiology*, 21(2), 138–151.

- Iwashkiw, J. A., Vozza, N. F., Kinsella, R. L., & Feldman, M. F. (2013). Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation. *Molecular Microbiology*, *89*(1), 14–28.
- Karl, David M., & Church, M. J. (2014). Microbial oceanography and the Hawaii Ocean Time-series programme. *Nature Reviews. Microbiology*, *12*(10), 699–713.
- Karner, M. B., DeLong, E. F., & Karl, D. M. (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature*, *409*(6819), 507–510.
- Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., ... Worden, A. Z. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biology*, *12*(6), e1001889.
- Ku, S. C., Schulz, B. L., Power, P. M., & Jennings, M. P. (2009). The pilin O-glycosylation pathway of pathogenic *Neisseria* is a general system that glycosylates AniA, an outer membrane nitrite reductase. *Biochemical and Biophysical Research Communications*, *378*(1), 84–89.
- Kursa, M. B., Rudnicki, W. R., & Others. (2010). Feature selection with the Boruta package. *Journal of Statistical Software*, *36*(11), 1–13.
- Lees-Miller, R. G., Iwashkiw, J. A., Scott, N. E., Seper, A., Vinogradov, E., Schild, S., & Feldman, M. F. (2013). A common pathway for O-linked protein-glycosylation and synthesis of capsule in *A. cinetobacter baumannii*. *Molecular Microbiology*, *89*(5), 816–830.
- Martin, B. D., Witten, D., & Willis, A. D. (2020). Modeling microbial abundances and dysbiosis with beta-binomial regression. *The Annals of Applied Statistics*, *14*(1), 94–115.
- Martiny, A. C., Coleman, M. L., & Chisholm, S. W. (2006). Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(33), 12552–12557.
- Needham, B. D., & Trent, M. S. (2013). Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. *Nature Reviews. Microbiology*, *11*(7), 467–481.
- Nelson, C. E., Carlson, C. A., Ewart, C. S., & Halewood, E. R. (2014). Community differentiation and population enrichment of Sargasso Sea bacterioplankton in the euphotic zone of a mesoscale mode-water eddy. *Environmental Microbiology*, *16*(3), 871–887.
- Neuer, S., Davenport, R., Freudenthal, T., Wefer, G., Llinás, O., Rueda, M.-J., ... Karl, D. M. (2002). Differences in the biological carbon pump at three subtropical ocean sites. *Geophysical Research Letters*, *29*(18), 32-1-32–34.

- Nothaft, H., & Szymanski, C. M. (2010). Protein glycosylation in bacteria: sweeter than ever. *Nature Reviews. Microbiology*, 8(11), 765–778.
- Orians, K. J., & Bruland, K. W. (1986). The biogeochemistry of aluminum in the Pacific Ocean. *Earth and Planetary Science Letters*, 78(4), 397–410.
- Rivals, I., Personnaz, L., Taing, L., & Potier, M.-C. (2006). Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics*, 23(4), 401–407.
- Robinson, D. F., & Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*, 53(1), 131–147.
- Sedwick, P. N., Church, T. M., Bowie, A. R., Marsay, C. M., Ussher, S. J., Achilles, K. M., ... McGillicuddy, D. J. (2005). Iron in the Sargasso Sea (Bermuda Atlantic Time-series Study region) during summer: Eolian imprint, spatiotemporal variability, and ecological implications. *Global Biogeochemical Cycles*, 19(4). doi:10.1029/2004GB002445
- van Hulst, M., Middag, R., Dutay, J.-C., de Baar, H., Roy-Barman, M., Gehlen, M., ... Sterl, A. (2017). *Manganese in the west Atlantic Ocean in the context of the first global ocean circulation model of manganese* [Data set]. doi:10.5194/bg-14-1123-2017
- Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., ... Seeberger, P. H. (Eds.). (2016). *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
- Vergin, K. L., Beszteri, B., Monier, A., Thrash, J. C., Temperton, B., Treusch, A. H., ... 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(7), 1322–1332.
- Vik, A., Aas, F. E., Anonsen, J. H., Bilsborough, S., Schneider, A., Egge-Jacobsen, W., & Koomey, M. (2009). Broad spectrum O-linked protein glycosylation in the human pathogen *Neisseria gonorrhoeae*. *Proceedings of the National Academy of Sciences of the United States of America*, 106(11), 4447–4452.
- von K ugelgen, A., Tang, H., Hardy, G. G., Kureisaite-Ciziene, D., Brun, Y. V., Stansfeld, P. J., ... Bharat, T. A. M. (2020). In Situ Structure of an Intact Lipopolysaccharide-Bound Bacterial Surface Layer. *Cell*, 180(2), 348-358.e15.
- Wildschutte, H., Preheim, S. P., Hernandez, Y., & Polz, M. F. (2010). O-antigen diversity and lateral transfer of the wbe region among *Vibrio splendidus* isolates. *Environmental Microbiology*, 12(11), 2977–2987.

Wildschutte, H., Wolfe, D. M., Tamewitz, A., & Lawrence, J. G. (2004). Protozoan predation, diversifying selection, and the evolution of antigenic diversity in *Salmonella*. *Proceedings of the National Academy of Sciences of the United States of America*, 101(29), 10644–10649.

Wu, J., Sunda, W., Boyle, E. A., & Karl, D. M. (2000). Phosphate depletion in the western North Atlantic Ocean. *Science*, 289(5480), 759–762.

**CHAPTER 4. DEVELOPMENT AND OPTIMIZATION OF
A METHOD FOR THE IDENTIFICATION OF
ORGANOPHOSPHORUS COMPOUNDS USING LIQUID
CHROMATOGRAPHY COUPLED WITH INDUCTIVELY
COUPLED PLASMA MASS SPECTROMETRY AND
ELECTROSPRAY IONIZATION MASS SPECTROMETRY
(LC-ICPMS-ESIMS)**

4.1 INTRODUCTION

The chemical characterization of organophosphorus in the ocean currently relies on two different methods. The first method is a colorimetric assay that relies on the formation of a blue complex between inorganic P or PO_4^{3-} and molybdenum (molybdenum blue method) (Murphy & Riley, 1962). The molybdenum blue method is a highly sensitive and convenient way to measure the concentration of phosphate in environmental samples, however, due to the acidic conditions necessary to obtain the blue complex, some phosphate esters, which are part of the organic P pool, are hydrolyzed to inorganic phosphate. In recognition of this, phosphate measured by the molybdenum blue method is therefore noted as “soluble reactive phosphorus” or SRP. The molybdenum blue method has a limit of detection (LOD) of ~30 nM (Murphy & Riley, 1962). This LOD has been decreased to 1 nM using MAGIC (Karl & Tien, 1992) or even 0.5 nM using Liquid Waveguide Capillary Cell (LWCC ; (Zhang & Chi, 2002)) but some oligotrophic waters have SRP concentrations below those limits which prevents the observation of some P dynamics (Martiny et al., 2019)

Organic P (P_{org}) is measured as the difference in inorganic P concentration before (SRP) and after oxidation of organic matter (total phosphorus or P_{tot}) by UV light, high temperature combustion, or strong chemical oxidants (Murphy & Riley, 1962). Thus: $[\text{P}_{\text{org}}] = [\text{P}_{\text{total}}] - [\text{SRP}]$. Some organophosphorus compounds are highly resistant to oxidation and may not have the same efficiency of conversion to phosphate as the model compounds used to optimize organic matter oxidation. Combined, these effects may lead to an overestimation of the inorganic P pool and an underestimation of the total P pool which ultimately leads to an underestimation of the organic P pool, but in different proportion depending on the oxidation method used. Some efforts have been made recently to address the total oxidation issue (Foreman et al., 2019).

The second method used to characterize organic P is ^{31}P nuclear magnetic resonance (NMR), which allows the identification of the different P functional groups such as phosphate mono- and di-esters, phosphonates and polyphosphates. ^{31}P NMR has shown that marine organophosphorus is mainly made of phosphate mono- and di-esters, phosphonate esters and polyphosphates. For both the dissolved and the particulate pool, these functional groups have been identified (Cade-Menun et al., 2005; Young & Ingall, 2010). However, the very complex mixture of organophosphorus compounds, the similar chemical shifts for compounds with the same functional groups and the low abundance of each individual compounds, leads to broad peaks in ^{31}P NMR spectra that prevent the use of other NMR experiments such as 2D NMR to characterize the breadth of most organophosphorus compounds.

LMW dissolved organic phosphorus (LMWDOP) is the largest pool of organic P in surface waters and are used by the marine microbial community to meet their P demands. This pool cycles the fastest. Thus, being able to characterize this pool is of great interest when studying the marine biogeochemical P cycle and so far, limitations originated from the methods available

affect greatly our capacity at understanding it. LMWDOP compounds can be most readily characterized by mass spectrometry (MS) coupled with a separation method such as gas chromatography (GC) or liquid chromatography (LC). This approach has been employed to measure the presence and the concentration of organophosphorus pollutants such as herbicides/pesticides (Liu et al., 2005; Sinha, Vasudev, & Rao, 2012; Wollersen & Musshoff, 2007), chemical warfare agents (Mawhinney et al., 2007) and their degradation products, and drugs (Ghassabian et al., 2012), as well as household and industrial products (Castro et al., 2020). However, all of those analyses were for structurally resolved organophosphorus compounds with known masses i.e. in a targeted manner and have been marginally applied to marine samples (Johnson et al., 2017; Johnson et al., 2020; Kujawinski et al., 2017). LC-MS using electrospray ionization (ESI) is heavily used in marine sciences and is now well established in the field of metabolomics i.e. the study of small molecules involved in metabolism as substrate, intermediary or product. Metabolomics has helped identify a large number of key metabolites and using a targeted approach measure their concentration and distribution whereas the untargeted approach is often used as a comparison tool to study the impact of environmental variables such as P concentrations (Kujawinski et al., 2017) or light (Boysen et al., 2020). These methods have led to the characterization of a number of new nitrogen (Widner et al., 2020) and sulfur (Durham et al., 2019) metabolites but few P containing metabolites. There are several reasons for this. First most P-containing metabolites are not retained or only very poorly retained by solid phase extraction (SPE) techniques used to recover organic matter from seawater (Johnson et al., 2017) which prevents these compounds from being pre-concentrated before LC-ESIMS analysis. Second, P containing compounds are often small and highly polar and thus they are not retained on typical reverse phase columns such as C18 which are commonly used in LC-MS. Third, phosphorus has a high ionization potential which leads to a poor ionization efficiency in the mass spectrometer and a relatively high limit of detection which is especially an issue in ICPMS which rely only on P detection and not on other easily ionized atoms present in the molecule (Bandura et al., 2002; Johnson et al., 2017). Last but not least, unlike nitrogen or sulfur or even iron, P does not have any stable isotope. This limits our capability to find P containing metabolites using mass search algorithms that uses natural abundance isotope ratios to target elements containing a specific heteroatom (S, Br, Fe, Cu, etc.) (Baumeister et al., 2018; Boiteau & Repeta, 2015; Boiteau et al., 2016; Durham et al., 2019).

To fill the gap in our knowledge of the molecular composition of LMW P containing compounds in marine DOM and POM, we developed an element targeted-approach similar to what has been done for trace metals (Boiteau & Repeta, 2015). For this we used LC-MS but used two different source of ionization: inductively coupled plasma mass spectrometry (ICPMS) and electrospray ionization mass spectrometry (ESIMS). ICP is a highly destructive form of ionization which fully breaks a molecule into its constituent elements. The coupling of ICPMS with LC allows for the quantitation of different organic P compounds, and tells the analyst where they elute in the chromatogram, but it provides no information on the molecular characteristics of the compound.

In contrast, the soft ionization offered by the ESI allows the determination of the exact mass of intact molecular ions, which is key to generating an elemental formula. In complex environmental samples, many compounds co-elute at the same retention time and to identify organic P compounds, the molecular ion of the P-containing compound must be identified within the suite of co-eluting molecular ions. Using tandem mass spectrometry, co-eluting molecular ions can be selected and further fragmented in an ion trap and the fragmentation pattern used to help identify or further characterize the compound. Therefore, after separation of P containing compounds by LC, ICPMS data is used to obtain the retention times of the P containing compounds and ESIMS is used to extract all the molecular masses detected at the retention time of interest. By using an online elemental formula generator to which we apply different filters e.g. presence of a P atom we can generate elemental formulae for each P-containing masses detected. We can then match this formula to a P-containing molecule and confirm the identity of this molecule using the fragmentation pattern obtained with the Orbitrap.

To develop this method, we started by optimizing the separation of a mixture of organic phosphonate and phosphate using a reverse-phase column functionalized with anion exchange sites. Then, we optimized the ICPMS conditions by changing the collision gas, the collision gas flow and the dwell time. Finally, we optimized the ESIMS parameters by switching to negative mode and varying the voltage as well as the source's gas flow. We also optimized the Orbitrap parameters and created a new method in order to only acquire fragmentation information for P-containing compounds.

4.2 PHOSPHONATE SEPARATION BY LIQUID CHROMATOGRAPHY

4.2.1 Choice of the column and the standard mixture

Organophosphorus compounds, including phosphonates, are highly polar and for the most part not retained by reverse-phase columns such as C18. However, phosphonates and a majority of phosphate esters have relatively low pKa (~ 2 for phosphonates) and are therefore negatively charged at pH > 3. Based on this, Primesep SB (Sielc Technology, USA), an amine-functionalized reverse-phase column was used to increase the retention of the P-compounds through ionic interactions between the column with the phosphate or phosphonate anion. This column was previously shown to be efficient at retaining and separating methylphosphonic acid (MPn) from ethylphosphonic acid (EtPn) and 2-hydroxyethylphosphonic acid (2-HEP) from 2-hydroxypropylphosphonic acid (2-HPP); phosphonates likely to be found in marine samples (Repeta et al., 2016).

As our primary goal was to identify and characterize LMW P compounds that we expect to be in marine samples, we chose a suite of low molecular weight, highly polar P compounds with different functional groups as model compounds to develop the chromatographic separations.

Thus, we selected adenosine monophosphate and glucose-6-phosphate as representatives for nucleic acids and sugar phosphates as well as 2-aminoethylphosphonic acid (2-AEPn), the first phosphonate discovered in nature (Horiguchi & Kandatsu, 1959), 2-hydroxyethylphosphonic acid (2-HEP) and methylphosphonic acid (MPn) (both previously identified in HMWDOM), ethylphosphonic acid (EtPn) and phosphorous acid (phosphite), the reduced form of inorganic phosphate which is of great interest for marine biogeochemistry since abundant and common microbes have the potential to use phosphorus acid as a source of P (Feingersch et al., 2012; Figueroa & Coates, 2017; Polyviou et al., 2015). 2-AEPn, AMP, G6P, MPn, EtPn and phosphorus acid were purchased from Sigma Aldrich and 2-HEP was purchased from Manchester Organics (UK). Because all the model compounds are hygroscopic, they were kept them in a glass dessicator filled with Drierite, sealed with grease and kept under vacuum.

4.2.2 Optimization of the mobile phase

Based on the notes available on the SIELC website 2-HEP can be separated from 2-hydroxypropylphosphonic acid (HPP) using 20-40% acetonitrile (AcN) and 10-15 mM aqueous ammonium formate (AmFm) at pH 3. For the separation of MPn, EtPn and ethyl methyl phosphonic and isopropyl methylphosphonic acid esters, mobile phases at pH 3-4 using 5-40% AcN and 25-50 mM ammonium acetate were used. Similar conditions were used to separate sugar phosphates and nucleotide monophosphates. Therefore, I chose to work at pH = 3, with acetonitrile and used the aqueous formic acid/AmFm as a buffer. It should be noted that all the tests were performed using one Primesep SB column (2.1 x 250 mm, 5 μ m, 100 Å) equilibrated for 2 hours with the mobile phase at a 400 μ L/min flow rate.

The initial trial used a mixture of 30% AcN with 15 mM AmFm at pH= 3. As shown on Figure 4.1, this mobile phase leads to the elution of the P standards in < 10 min with MPn and EtPn co-eluting at 7.5 min.

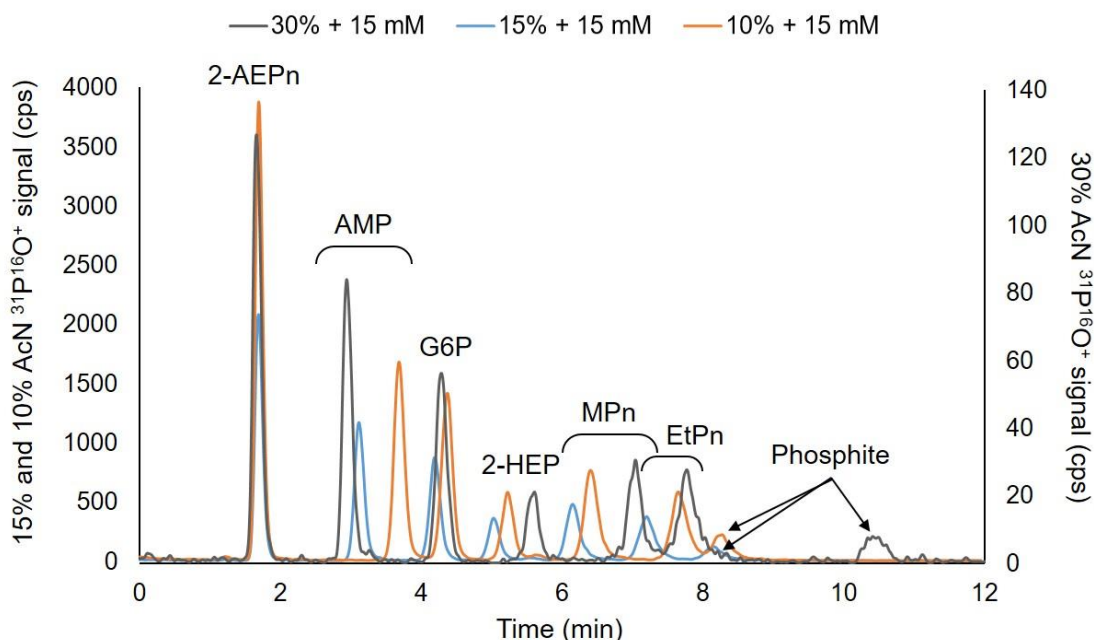


Figure 4.1: Chromatogram of the P standards mixture (750 pmol each) using a mobile phase with 30% AcN (grey trace), 15% AcN (blue trace) and 10% AcN (orange trace) in 15 mM aqueous AmFm at a pH 3.

In order to increase the retention of compounds and separate MPn from EtPn, we decreased the acetonitrile stepwise from 30% to 15% to 10%. Surprisingly, decreasing the AcN content did not change the retention times of the compounds very much except for phosphite (Table 4.1). This is due to the fact that the standards have very similar polarities, so they are not affected by the AcN content of the mobile phase. Their retention on the Primesep SB is primarily due to their ionic interactions with the column.

Table 4.1: Influence of the mobile phase AcN content on the retention times (in seconds) of the P-standards. A positive value indicates that the compounds are more retained compared to 30 % AcN and a negative value indicates that they are less retained. All mobile phases were made with a concentration of 15 mM AmFm.

Mobile phase	2-AEPn		AMP		G6P		2-HEP		MPn		EtPn		Phosphite	
	RT	Shift	RT	Shift	RT	Shift	RT	Shift	RT	Shift	RT	Shift	RT	Shift
30%	100		176		257		336		422		466		630	
15%	102	2	188	12	252	-5	303	-33	370	-52	434	-32	492	-138
10%	102	2	222	46	264	7	315	-21	386	-36	461	-5	498	-132

Decreasing the AcN content had an important effect on the ICPMS sensitivity. At 30% AcN, the signal/noise for 2-AEPn is about three times lower than when using 15% and 10% AcN (Table 4.2) due to the introduction of more organics. There is no significant difference in the sensitivity whether we used 15% or 10% AcN, but due to the overall better retention of the P standards using 10% AcN, we chose to use that percentage moving forward.

Table 4.2: Influence of the different mobile phase compositions on the sensitivity

Mobile phase mixtures	2-AEPn signal (cps)	Average noise (cps)	S/N	Peak width (sec)
30% AcN + 15 mM AmFm	126.69	0.40	318	19.5
15% AcN + 15 mM AmFm	2083.30	1.86	1119	34.5
10% AcN + 15 mM AmFm	3878.99	3.84	1011	32.5
10% AcN + 30 mM AmFm	2226.31	29.12	76	20.5
10% AcN + 10 mM AmFm	3022.38	10.88	278	43.5
10% AcN + 8 mM AmFm	3299.99	3.59	919	35.0
10% AcN + 7 mM AmFm	2938.73	7.84	374	35.5

As previously shown, decreasing the percentage of AcN does not increase the retention of the P standards. Since those compounds are negatively charged at pH 3, they interact with the ammonium sites of the column packing material. Thus, reducing the competition for those sites by reducing the AmFm buffer concentrations (also negatively charged) should increase the retention of the P standards. To verify this, we increased the concentration of AmFm from 15 mM to 30 mM.

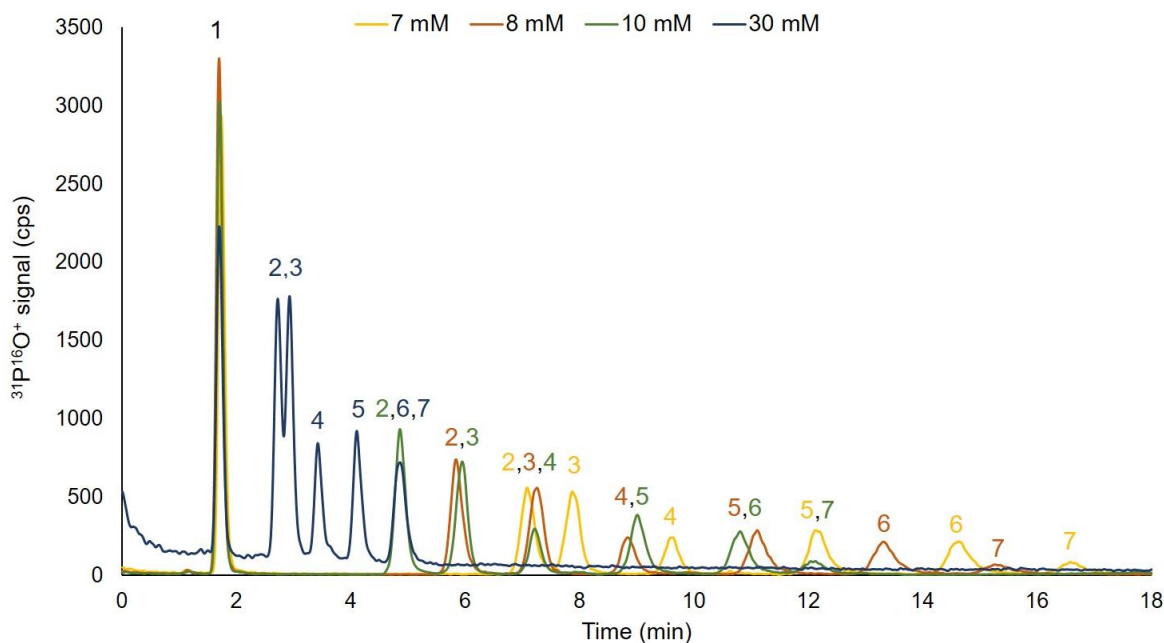


Figure 4.2: Chromatogram of the P standards mixture (750 pmol) using a mobile phase with 10 % AcN in water and a concentration of AmFm of 30 mM (orange), 10 mM (red), 8 mM (dark yellow) and 7 mM (yellow) at a pH 3. Each number represent a compound with 1: 2-AEPn, 2: AMP, 3: G6P, 4: 2-HEP, 5: MPn, 6: EtPn and 7: phosphite.

As expected, the P standards that had eluted in 10 minutes now elute in 5 min (Figure 4.2). The peaks are not well separated. We then decreased the concentration of AmFm first to 10 mM, then to 8 mM and finally to 7 mM. Decreasing the AmFm content increased the retention of the P

standards (Figure 4.2, Table 4.3). 2-AEPn is the only standard not affected by this change. 2-AEP has a pK_{a1} of 2.41 ($pK_{a2} = 7.0$) and at pH 3 this compound is neutral. Therefore, 2-AEPn is not strongly retained by the ionic interactions like the other compounds, and as it has a LMW and is polar it is not well retained by the column. As 7 mM AmFm leads to good retention of our P-compounds and because there is a benefit at keeping the buffer concentration as low as possible since AmFm has a deleterious effect on the negative-ionization in ESIMS (Wu et al., 2004) (see Section 4.4), we chose a mobile phase containing 10% AcN with 7 mM AmFm acidified to pH 3 with formic acid.

Table 4.3: Influence of the mobile phase AmFm concentration on the retention times (in seconds) of the different standards. A positive value indicates that the compounds are more retained than in the 10 % AcN with 15 mM AmFm mobile phase and a negative value indicates that they are less retained.

Mobile phase	2-AEPn		AMP		G6P		2-HEP		MPn		EtPn		Phosphite	
	RT	Shift	RT	Shift	RT	Shift	RT	Shift	RT	Shift	RT	Shift	RT	Shift
30 mM	102	0	164	-58	176	-88	206	-109	246	-140	292	-169	N/A	N/A
15 mM	102		222		264		315		386		461		498	
10 mM	102	0	292	70	356	92	433	118	540	154	648	187	723	225
8 mM	102	0	350	128	436	172	531	216	666	280	799	338	916	418
7 mM	104	2	425	203	472	208	578	263	728	342	878	417	995	497

4.3 DETECTION OF PHOSPHORUS USING INDUCTIVELY COUPLED PLASMA MASS SPECTROMETRY (ICPMS)

4.3.1 Instrumentation and initial conditions

To develop this method, we used an iCap Qc (Thermo Fisher Scientific, USA) equipped with a perfluoroalkoxy (PFA) nebulizer delivering a flow between 35 and 15 $\mu\text{L}/\text{min}$ to a quartz cyclonic spray chamber. The instrument has a quartz injector of 2 mm ID and we used a quartz torch optimized for use with organic solvents. The sampling interface is made of a nickel and copper sampling cone, a nickel skimmer cone and an extraction lens. The collision cell, or QCell, is a quadrupole in which argon (Ar) is introduced along with either helium (He) or oxygen (O_2). Helium is used in kinetic energy discrimination (KED) mode to remove undesirable polyatomic interferences while O_2 is used when running with the collision cell technology (CCT) mode i.e. without KED barrier. Because the mobile phase contains acetonitrile, formic acid and ammonium formate which can decrease the sensitivity of the instrument, O_2 was introduced in the plasma at a flow rate of 25 mL/min to facilitate oxidation to CO_2 . The maximum flow rate sprayed into the instrument was set to be 50 $\mu\text{L}/\text{min}$ (Boiteau & Repeta, 2015). As the flow rate for the HPLC separation exceeded the optimum flow of the ICPMS, we split the HPLC flow post-column so that only 50 $\mu\text{L}/\text{min}$ was directed into the ICPMS instrument.

Before each run, the instrument was tuned using a mass calibration solution containing 35 µg/L of Be, 15 µg/L of Cu and Ni, 10 µg/L of Al, Ga and Mg, 8 µg/L of Co, Li and Sc, 6 µg/L of Ag and Mn, 5 µg/L of Sr, 4 µg/L of Ba and Ti, 3 µg/L of Bi, Ce, Cs, Ho, In, Rh, Ta, Tb, U and Y in 2.5% (v/v) nitric acid (Inorganic Ventures, USA) in STD mode using the source autotune High Matrix 2. When using CCT mode with O₂, we also tuned in CCT mode using the source autotune CCT Line 2. When using KED initially, I performed performance reports as well, using Tune B solution (Inorganic Ventures, USA). Once every few months, the mass calibration solution was used to perform a mass calibration and a detector setup.

Initially, analyses were performed using the KED mode with He as a collision gas. The He flow was set each time by the instrument software to optimize discrimination and sensitivity while executing the tuning protocol, and varied between 3.5 and 4.5 mL/min. In order to monitor the presence of organophosphorus compounds, we monitored the ion ³¹P⁺ with a dwell time of 0.02 s. Using those conditions, we injected the standard mixture prepared at concentrations ranging from 300 µM to 50 µM and obtained the chromatogram shown in Figure 4.3 for a concentration of 150 µM. Based on the 20 µL injection volume and the split causing a loss of 75% of the flow, these concentrations correspond to a quantity of 15 nmol to 0.25 nmol of each standard compound reaching the spray chamber.

4.3.2 Optimization of ICPMS detection sensitivity

4.3.2.1 Use of oxygen as a collision gas

The main issue encountered when quantifying P using the ICPMS is that m/z 31, mass at which P ions are detected, has several polyatomic interferences: ¹⁵N¹⁶O⁺, ¹⁴N¹⁶O¹H⁺ and ¹²C¹H₃¹⁶O⁺ (Bandura et al., 2002). Those interferences are likely to be important considering the high abundance of C, N and O in the mobile phase and sample. This leads to an increase of the noise i.e. a decrease in sensitivity. For example, the signal/noise (S/N) for 2-AEPn in the 150 µM standard mixture is 4.4. Defining our limit of detection (LOD) as a S/N = 1.3, the detection limits are ~150 pmol for 2-AEPn, ~250 pmol for AMP and G6P, ~375 pmol for 2-HEP, MPn and EtPn and ~500 pmol for phosphite. Using O₂ instead of He in the collision cell yields a polyatomic P species ⁴⁷PO⁺ that can be detected at m/z 47, which does not have isobaric interferences. Thus, even though the signal is an order of magnitude lower, the signal/noise increases by approximately an order of magnitude and limits of detection are ~20 pmol for 2-AEPn, ~35 pmol for AMP and G6P, ~50 pmol for 2-HEP, MPn and EtPn and ~70 pmol for phosphite.

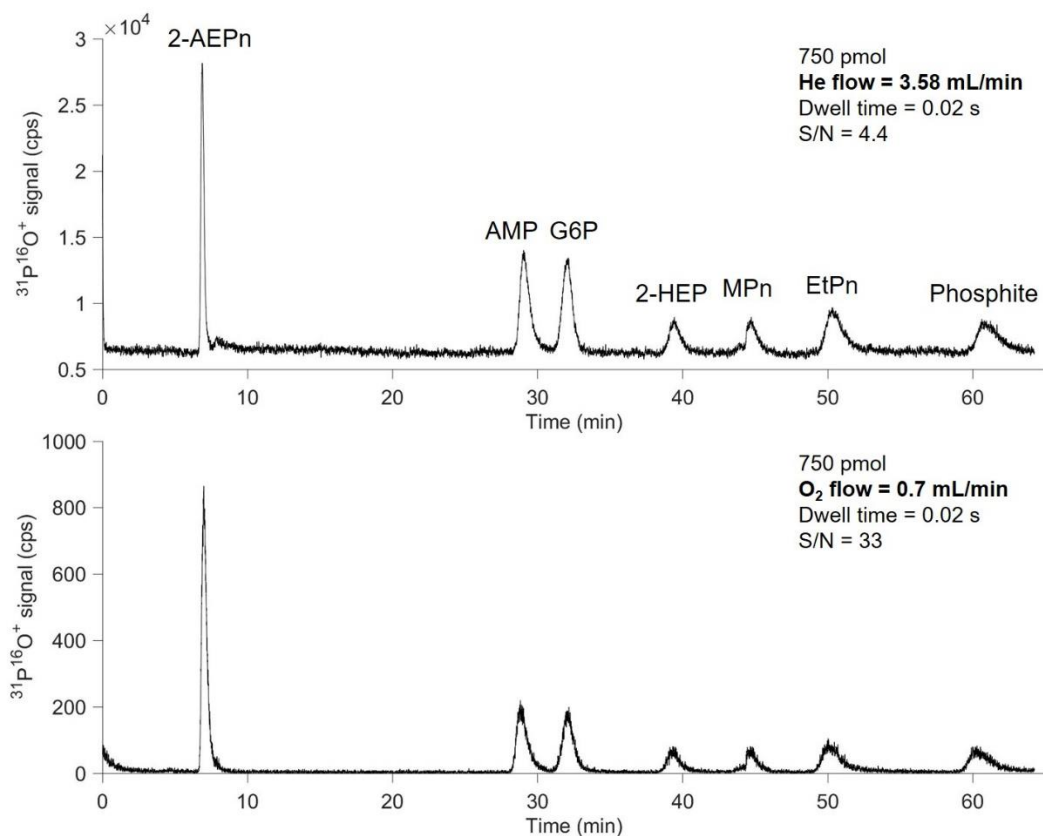


Figure 4.3: Influence of the introduction of O₂ instead of He in the collision cell on the P signal sensitivity for the standard mixture (750 pmol) with a dwell time of 0.02 s and collision gas flow set by the instrument.

4.3.2.2 Optimization of dwell time

The dwell time is the amount of time during which the ICPMS accumulates signal for each mass chosen. This time can vary between few milliseconds to seconds. When looking at multiple elements, it is recommended to set the dwell time in the low range (typically 100s ms) so multiple elements can be analyzed simultaneously with a reasonable sensitivity for each. However, when acquiring data for only one mass, increasing the dwell time is equivalent to obtaining more signal, which leads to better sensitivity. As we are here only interested in the $^{31}\text{P}^{16}\text{O}^+$ mass, we increased the dwell time from 20 ms to 500 ms. As shown on Figure 4.4 below, this led to a much smoother signal and an increase of signal/noise from 33 to 82 for 2-AEPn when injecting 750 pmol.

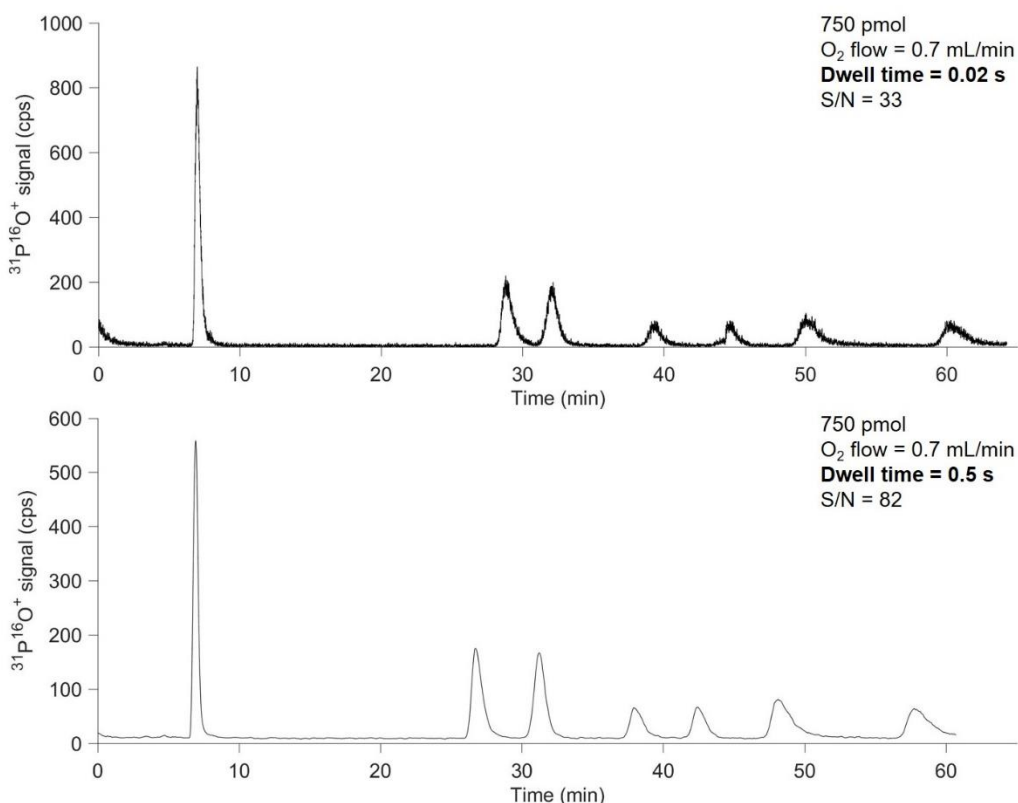


Figure 4.4: Influence of the dwell time on $^{31}P^{16}O^+$ standard mixture signal (750 pmol) and a collision cell O_2 flow of 0.7 mL/min. The elution order for the different P-standards is the same as indicated in Figure 4.3.

4.3.2.3 Optimization of the oxygen flow

Because P is absent from the mass calibration tuning solution, the O_2 flow is not optimized for P, but is arbitrarily set to a flow of 0.7-0.8 mL/min. However, it has been previously shown that the $^{31}P^{16}O^+$ signal is higher at a lower O_2 flow of 0.2 mL/min (Bandura et al., 2002). Recognizing that this study did not report the signal/noise and was using an older version of the instrument, we manually varied the O_2 flow while 1 μ M P, 10 μ M P or MQ blank solutions were infused into the spectrometer. We prepared the P standard solution by diluting a 1001 mg/L P standard solution for ICP (Sigma Aldrich) into 10 mL of MQ water. The samples were infused for 5 min in order to allow the O_2 level to equilibrate in the collision cell. We performed the experiment by increasing the O_2 flow from 0.0 mL/min to 0.775 mL/min (the instrument tuned value) by increments of 0.05 mL/min. To verify that the equilibration time was sufficient to avoid residual O_2 concentrations, the O_2 flow was decreased step-wise from 0.775 mL/min back to 0.0 mL/min. As shown on Figure 4.5, the highest $^{31}P^{16}O^+$ signal was obtained at an O_2 flow of 0.2 mL/min, but the best S/N was obtained at an O_2 flow of 0.65 mL/min.

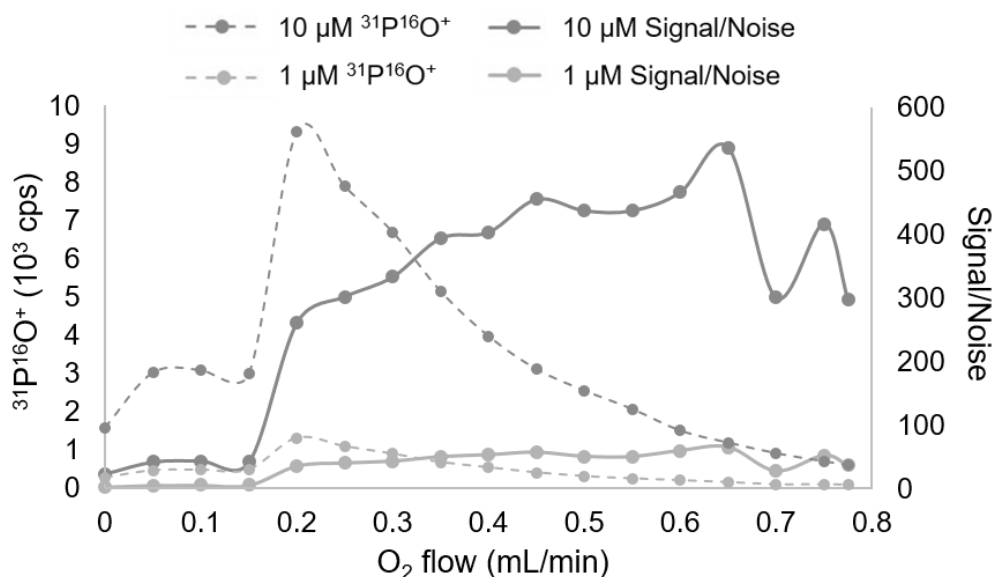


Figure 4.5: Collision cell O₂ flow optimization for ³¹P¹⁶O⁺ signal (dashed lines) and ³¹P¹⁶O⁺ signal/noise (full lines) using standard solutions of 1 μM P (light grey) and 10 μM P (dark grey)

To confirm that the O₂ flow rate was also optimized for HPLC, we analyzed different concentrations of the P-standard mixture using a 10% AcN mobile phase with 7 mM AmFm acidified to pH 3 in isocratic mode at a flow rate of 200 μL/min. We used O₂ flows of 0.7 mL/min (instrument tuned value), 0.65 mL/min, 0.6 mL/min, 0.45 mL/min and 0.2 mL/min (Figure 4.6). As before, we calculated the signal/noise ratio using the intensity of the signal for 2-AEPn and the highest or average noise between 800 and 1000 seconds for 0.02 s and 0.5 s dwell times respectively. The results were similar to the direct infusion experiments illustrated in Figure 4.5 except that the highest S/N is obtained for an O₂ flow of 0.6 mL/min instead of 0.65 mL/min. Changing the O₂ flow from 0.7 mL/min to 0.6 mL/min led to an increase of S/N for 2-AEP from 82 to 252 (~ 3x, compared to ~ 2x for direct infusion measured in Figure 4.5), and we used an O₂ flow of 0.6 mL/min moving forward. These analyses show the importance of the O₂ flow rate on S/N and sensitivity, and it is recommended to perform similar experiments when using different instrumentation or conditions.

4.3.2.4 Summary

Table 4.4 summarizes the impact of ICPMS instrument parameters on S/N for 2-AEP. When using KED mode with a dwell time of 0.02 s, the S/N was 4.4 with an LOD of ~150 pmol. Using the CCT mode and introducing O₂ as a collision gas at a flow rate of 0.7 mL/min, the S/N increased by a factor of 7.5 (Table 4) which corresponds to a LOD of 20 pmol. Increasing the dwell time to 0.5 s, increased the S/N by a factor of 2.5 with a LOD of ~8 pmol. Optimizing the O₂ flow in the collision cell, increased the S/N by a factor of 3 which yielded a LOD of ~2.5 pmol (Table 4.4).

Table 4.4: Summary of the signal/noise ratio obtained for the different tested parameters and the associated increase factors.

Mode	Dwell	Collision gas	Gas flow	S/N	Increase factor	
KED	0.02	He	3.58	4.4		
CCT	0.02	O ₂	0.7	33	7.5	
CCT	0.5	O ₂	0.7	82	2.5	
CCT	0.5	O ₂	0.2	137	1.7	57.3 ²
CCT	0.5	O ₂	0.65	186	1.4	3.1 ¹
CCT	0.5	O ₂	0.45	215	1.2	
CCT	0.5	O ₂	0.6	252	1.2	

¹ S/N improvement when changing the O₂ flow from 0.7 to 0.6 mL/min

² Overall S/N improvement from the initial to the optimized conditions

In the method development, the LOD was calculated for 2-AEPn, but as the retention time increases, the peaks become broader which leads to a lower signal and a higher LOD. Phosphite, which has the greatest retention time of the standards, has the largest and smallest peak and a LOD of 10 pmol, compared to 2.5 pmol for 2-AEPn, under the optimized conditions (Figure 4.6, top panel).

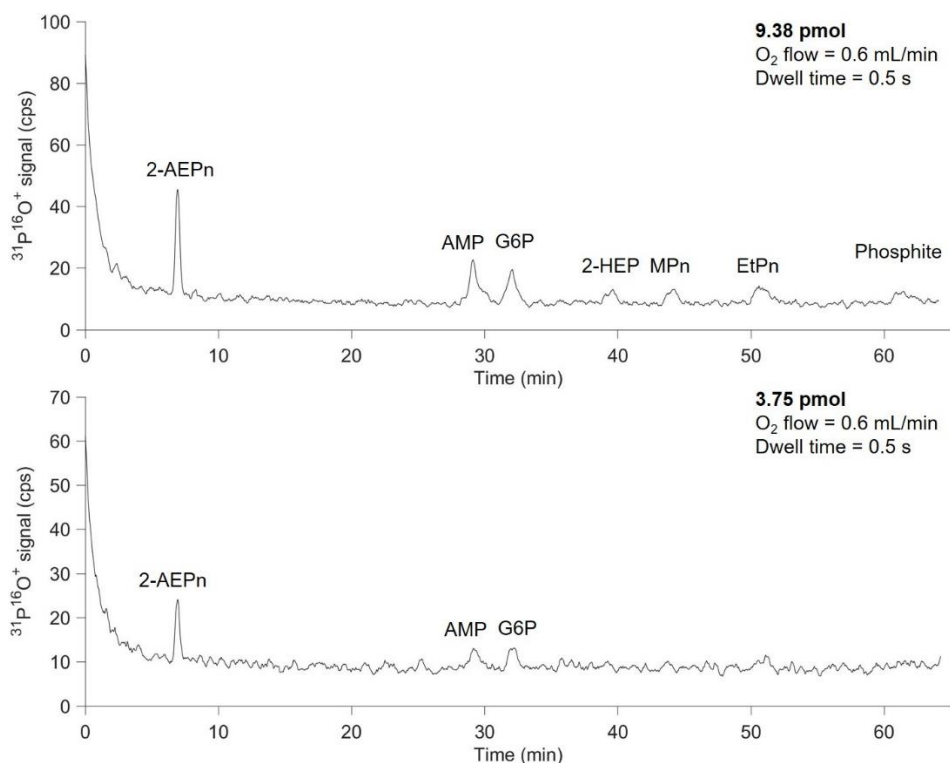


Figure 4.67: Changes in LOD with retention time. Chromatograms of mixtures where 9.38 pmol (top) or 3.75 pmol (bottom) of standard is injected and acquired using the optimized conditions i.e. CCT mode, dwell time = 0.5 s, O₂ flow = 0.6 mL/min

4.4 DETECTION OF ORGANOPHOSPHORUS COMPOUNDS BY ELECTROSPRAY IONIZATION MASS SPECTROMETRY (ESIMS)

4.4.1 Instrumentation and initial conditions

To identify molecular ions associated with organophosphorus compounds, the HPLC was coupled to an Orbitrap Fusion (Thermo Scientific, USA) mass spectrometer. Separations were performed using two Primesep SB (2.1x250 mm, 5 μm , 100 \AA) columns connected in series and eluted with, 90/10 (v/v) water/AcN with 7 mM AmFM acidified to pH 3 with formic acid at a flow rate of 200 $\mu\text{L}/\text{min}$ with 100% of the eluent was directed into the mass spectrometer through a heated electrospray ionization (H-ESI) inlet. The distance between the inlet capillary and the sampling cone can be adjusted (graduations from 1 to 3, arbitrary units). For a 200 $\mu\text{L}/\text{min}$ flow rate, the best position was determined by visually monitoring the total ion current (TIC) and its stability while the inlet capillary was manipulated. Based on this, analyses were performed at position 2.

Initially, ions were collected in positive mode using the instrument recommended method parameters (Table 4.5, Column 1). However, under those conditions phosphite did not yield measurable positive ions, and could not be detected. In negative mode, organophosphorus compounds are easily charged, especially inorganic species like phosphate and phosphite. Another advantage is that there are fewer background interferences when using negative mode over positive mode.

4.4.2 ESIMS parameters optimization

To switch to negative mode, we used the parameters recommended by a Thermo Fisher Scientific technical note for the quantification of glyphosate and aminomethylphosphonic acid, widely used phosphonates in agriculture, in water (<https://apps.lab.thermofisher.com/App/3140/an491-analysis-glyphosate-ampa-environmental-water-by-ion-chromatography-electrospray-tandem-mass-spectrometry-icesimsms>) (Table 4.5, Column 2). Under these conditions the detection of MPn and EtPn was improved by a factor of 1.4 and 2.2 respectively compared to positive mode, but most importantly we were able to detect phosphite.

To test the effect of auxiliary gas flow on sensitivity, we increased this flow from 2 to 5 (arbitrary units) which led to a slight improvement in the detection of all compounds (Table 4.5, Column 3). Increasing the sheath gas flow to 45 (arbitrary units, Table 4.5, Column 4) led to further improvement in the S/N for all compounds. Decreasing the sheath gas flow to 30 decreased the S/N of 2-AEPn but slightly increased the S/N for the other compounds. However, the S/N was uniformly lower than the values obtained for a gas flow setting of 45. Thus, further parameters were optimized using a 45 setting. We then increased the capillary voltage to 3100V

(Table 4.5, column 6) which did not affect the signal for 2-AEPn, but improved the S/N by factors between 1.1 to 2.2 for other compounds in the standard mixture. We then increased the vaporizer temperature to 300°C. This led to an increase in S/N by a factors ranging from 1.2 (for EtPn) to 2.4 (for 2-AEPn). A further increase of temperature increased the S/N (data not shown) further, but a very high vaporization temperature can potentially oxidize compounds, so we made the choice to use a vaporizer temperature of 300°C moving forward. Decreasing the ion tube temperature to 350°C decreased the S/N by almost a factor of 2, so for all analyses the ion tube temperature was set to 400°C.

Table 4.5: Electrospray parameters optimization for the detection of the P standards and improvement in the S/N value for each. Bold values show the changed parameter compared to the previous conditions. Values in red font mark a decrease in S/N whereas values in green font mark an increase. Values in black font indicate no change.

ESI parameters	1	2	3	4	5	6	7	8	9*	10	11
Positive (V)	3500								3500		
Negative (V)		3000	3000	3000	3000	3100	3100	3100	3100	2500	2500
Sheath gas (arb)	12	40	40	45	30	45	45	45	45	35	35
Aux gas (arb)	6	2	5	5	5	5	5	5	5	7	7
Sweep gas (arb)	2	1	1	1	1	1	1	1	1	0	0
Ion transfer tube temp (°C)	280	400	400	400	400	400	400	350	350	400	400
Vaporizer tube temp (°C)	80	275	275	275	275	275	300	300	300	275	300
Standards S/N											
2-AEPn	64.5	3.8	8.5	10.0	6.3	6.2	15.1	6.6	8.7	30.2	19.6
2-HEP	2.8	1.0	1.3	3.2	1.7	3.7	5.4	2.9	1.2	15.0	10.2
MPn	0.6	0.8	1.1	2.3	1.5	3.3	3.8	2.2	1.4	11.2	7.6
EtPn	0.4	0.9	1.4	2.8	1.6	4.1	4.9	2.6	1.5	13.7	9.0
Phosphite	ND	0.08	0.2	0.3	0.2	0.7	0.8	0.5	0.1	2.5	1.4

ND: not detected

Using the conditions listed in Table 4.5, Column 7, the influence of the mass range, Orbitrap resolution and the injection times on sensitivity was investigated (Table 4.6).

Table 4.6: Influence of Orbitrap resolution, injection times and mass range on the S/N of the different P standards. As a reference, the S/N for the parameters used in Column 7 in the table above (Table 5) are reported. Bold values show the changed parameter compared to the previous conditions

Orbitrap parameters	7	7.1	7.2	7.3	7.4
Orbitrap resolution	120000	240000	240000	120000	120000
Scan range	70-150	70-150	70-150	70-2000	70-150
RF lens %	60	60	60	60	60
AGC Target	2.00E+05	2.00E+05	2.00E+05	2.00E+05	2.00E+05
Max injection time (ms)	200	200	50	200	50
Isolation window	1.6	1.6	1.6	1.6	1.6
Activation type	HCD	HCD	HCD	HCD	HCD
HCD collision energy	35	35	35	35	35
Detector type	Orbitrap	Orbitrap	Orbitrap	Orbitrap	Orbitrap
Orbitrap resolution	30000	60000	60000	30000	30000
AGC target	5.00E+04	5.00E+04	5.00E+04	5.00E+04	5.00E+04
Microscans	1	1	1	1	1
Max injection time (ms)	54	54	118	54	54
Standards S/N					
AEPn	6.6	5.7	9.7	12.5	13.5
2-HEP	2.9	2.2	2.6	2.3	3.6
MPn	2.2	1.8	2.8	1.7	3.0
EtPn	2.6	2.4	3.4	2.4	4.0
Phosphite	0.5	0.3	0.4	0.07	0.5

Increasing the Orbitrap mass resolution slightly decreased the S/N for all P standards which is to be expected as mass resolution is dependent on the amount of signal received (Table 4.6, Column 7.1). Counterintuitively, the S/N increased when using a shorter injection time (Table 4.6, Column 7.2). This phenomenon was also observed when operating at a lower resolution (Table 4.6, Column 7.4). However, similar to what we observed with ICPMS sensitivity when using a shorter dwell time, the signal is noisier, therefore an injection time of 200 ms was used. Increasing the mass range also led to a decrease in S/N especially for phosphite. Because some P-containing metabolites have a very low mass (<150 m/z) where the sensitivity is the lowest due to the presence of background ions, samples were analyzed with a low mass (m/z = 70-150) range and again with a higher m/z (150-1000) range to monitor the loss of signal using the full mass range (75-1000).

Using Column 7 (Tables 4.5 and 4.6) conditions, the influence of decreasing the voltage was investigated. Decreasing the voltage to 2500V, a big improvement in sensitivity was measured. With this voltage, the sheath and auxiliary flows were adjusted and optima were 35 arb and 7 arb respectively (Table 4.5, Column 10). Using 2500V led to an improvement of a factor of 2 to 3 compared to the parameters settings used previously with 3100V (Table 4.5, Column 7) which correspond to an overall improvement of a factor of 8 for 2-AEPn, ~15 for 2-HEP, MPn and EtPn and 31 for phosphite compared to the initial negative mode conditions (Table 4.5, Column

2). The vaporizer temperature was set to 275°C. Indeed, decreasing the vaporizer temperature to 275°C increased the signal compared to 300°C (Table 4.5, Column 11) and the parameters from Table 4.5, Column 10 i.e. negative mode, 2500V, sheath gas of 35 arb, auxiliary gas of 7 arb, ion transfer tube temperature of 400°C and vaporizer temperature of 275°C were chosen for future work. All Orbitrap parameters previously optimized were kept identical i.e. resolution of 120000, m/z scan range of 70-150, injection time of 200 ms.

4.4.3 MS² of organophosphorus compounds using precursor-product ions

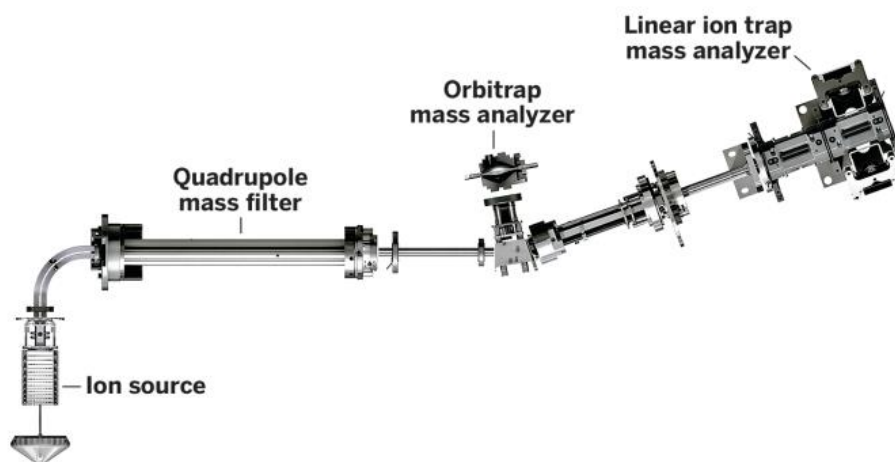


Figure 4.7: Ion path of the Orbitrap Fusion (<https://cen.acs.org/articles/91/i27/Thermo-Fisher-Scientifics-Orbitrap-Fusion.html>). Ions are generated at the ion source and separated by the quadrupole mass filter. Ions selected by the mass filter can be directed to the high resolution Orbitrap or to the lower resolution ion trap mass analyzers.

Initially, the instrument was operated such that all ions within the mass range selected that passed through the quadrupole mass filter were detected, within the mass range selected, by the Orbitrap mass analyzer (Figure 4.7). The Orbitrap mass analyzer detects the mass of molecular ions (MS¹). Ions that have intensity above a certain threshold value were further fragmented (MS²). For example, AMP (M = 347.063 g/mol), the corresponding molecular ion [M-H]⁻ will be detected in negative mode at m/z 346.054 and further fragmented to obtain an MS² spectrum with product ions at m/z 211.000 (C₅H₈O₇P⁻); 150.979 (C₃H₄O₅P⁻); 134.046 (C₅H₄N₅⁻); 96.969 (H₂PO₄⁻) and 78.958 (PO₃⁻) (Figure 5.2). However, most ions selected for MS² are not from P containing compounds and P compounds present in low abundance might be missed. To address this, we designed an analytical protocol guided by precursor-product ions. Product ions are ions emanating from the fragmentation of precursor (usually molecular) ions in the mass analyzers (Figure 4.7) and that are diagnostic for the compounds of interest (in this case, organophosphorus compounds). For example, all phosphonates have common m/z 78.958 (PO₃⁻) and 80.974 (H₂PO₃⁻) product ions and all phosphates have m/z 78.958 (PO₃⁻) and 96.969 (H₂PO₄⁻) product ions. When molecular ions are fragmented, any molecular ions containing one of the chosen

product ions detected by the ion trap will trigger the Orbitrap mass analyzer (Figure 4.7) to acquire MS² data for that molecular ion. In the case of AMP, the fragmentation of its molecular ion in the ion trap will lead to the presence of the product ions at m/z 78.958 (PO₃⁻) and 96.969 (H₂PO₄⁻) which will trigger the Orbitrap for a high resolution MS² acquisition. However, all organic compounds fragmented by the ion trap that do not contain the product ion of interest will not have a high resolution MS². Thus, this approach allows both the collection of fewer MS² spectra and the collection of MS² for molecular ions of low intensity but of interest for the analysis i.e. presence of a P containing fragment. The goals here are to first to reduce the number of target ions and second to specifically target P containing ions for collection of high resolution MS² spectra.

As we are interested in characterizing LMW phosphonates and phosphates, we started by selecting 3 product ions: m/z 78.958 (PO₃⁻), m/z 80.974 (H₂PO₃⁻), m/z 96.969 (H₂PO₄⁻) in the targeted mass list with a mass tolerance of ± 0.5 m/z and an intensity threshold at 5.0x10². To reduce the analysis time in the ion trap, we set the number of microscans per scans to 1 with an injection time of 50 ms and chose a rapid ion trap scan rate. The isolation was performed by the quadrupole with a window of 1.6 m/z. The precursor ion was activated by HCD (higher collision energy dissociation) with an energy fixed to 40%. The Orbitrap HCD collision energy was set for 35%, with a resolution of 30000, and an injection time of 54 ms with 1 microscan per scan.

This experiment significantly reduced the number of ions selected for MS² analyses. For example, when the standard mixture of organophosphorus compounds was analyzed over 150 Orbitrap MS² were acquired compared to 9 with product ion scanning. An analysis of the standard mixture showed that the product ion m/z 96.969 (H₂PO₄⁻) included ions with a m/z 96.959, which corresponds to HSO₄⁻. Further, the product ion m/z = 80.974 (H₂PO₃⁻) was present at a very low intensity. Thus, we removed these two products from the precursor scan mass list and only used m/z 78.958 (PO₃⁻). This allowed for a reduction of ion trap the mass scanning range to 75 to 84 and led to an increase in the sensitivity.

4.5 CONCLUSION AND RECOMMENDATIONS FOR FUTURE METHOD IMPROVEMENT

The Primesep SB column retains polar, low molecular weight phosphonates and phosphates. The organic solvent content of the mobile phase does not have a large effect on the retention times of P-containing analytes, however the concentration of the ammonium formate buffer has a significant effect on retention time, with lower buffer concentrations yielding better separation of the standard mixture of compounds used here. In this Chapter, the optimization of the mobile phase was performed in isocratic mode using 10% AcN, 7 mM of AmFm at pH 3 (formic acid) but for environmental samples, different conditions and gradients might yield better separations.

For example, there is some evidence that substitution of acetic acid for formic acid in the HPLC mobile phase will enhance detection limits in ESIMS (Wu et al., 2004). Likewise, increasing the content of AcN in the mobile phase should also be investigated as a higher percentage of organic solvent may facilitate desolvation in the electrospray interface.

Optimizing mass spectral acquisition parameters increased the sensitivity of detection by a factor of ~50 for the ICPMS and a factor of between 8 and 33 for the ESIMS. Detection limits on the ICPMS were 40 pmol for phosphite and 10 pmol for 2-AEPn. Detection limits on the ESIMS were ~900 pmol for phosphite and 90 pmol for 2-AEPn. Moreover, by introducing the precursor scans, the number of molecular ion candidates potentially attributed to organophosphorus compounds after alignment of the ICPMS and ESIMS traces was greatly reduced, facilitating the identification of P containing compounds. As an alternative to the precursor-product ion approach, an algorithm using the alignment between the ICPMS and the ESIMS datasets as well as other factors such as peak shape in order to find P-containing masses should be investigated.

4.6 REFERENCES

- Bandura, D. R., Baranov, V. I., & Tanner, S. D. (2002). Detection of ultratrace phosphorus and sulfur by quadrupole ICPMS with dynamic reaction cell. *Analytical Chemistry*, *74*(7), 1497–1502.
- Baumeister, T. U. H., Ueberschaar, N., Schmidt-Heck, W., Mohr, J. F., Deicke, M., Wichard, T., ... Pohnert, G. (2018). DeltaMS: a tool to track isotopologues in GC- and LC-MS data. *Metabolomics: Official Journal of the Metabolomic Society*, *14*(4), 41.
- Boiteau, R. M., & Repeta, D. J. (2015). An extended siderophore suite from *Synechococcus* sp. PCC 7002 revealed by LC-ICPMS-ESIMS. *Metallomics: Integrated Biometal Science*, *7*(5), 877–884.
- Boiteau, R. M., Till, C. P., Ruacho, A., Bundy, R. M., Hawco, N. J., McKenna, A. M., ... Repeta, D. J. (2016). Structural Characterization of Natural Nickel and Copper Binding Ligands along the US GEOTRACES Eastern Pacific Zonal Transect. *Frontiers in Marine Science*, *3*, 243.
- Boysen, A. K., Carlson, L. T., Durham, B. P., Groussman, R. D., Aylward, F. O., Ribalet, F., ... Ingalls, A. E. (2020). Diel Oscillations of Particulate Metabolites Reflect Synchronized

Microbial Activity in the North Pacific Subtropical Gyre (p. 2020.05.09.086173).
doi:10.1101/2020.05.09.086173

Cade-Menun, B. J., Benitez-Nelson, C. R., Pellechia, P., & Paytan, A. (2005). Refining ³¹P nuclear magnetic resonance spectroscopy for marine particulate samples: Storage conditions and extraction recovery. *Marine Chemistry*, *97*(3), 293–306.

Castro, V., Montes, R., Quintana, J. B., Rodil, R., & Cela, R. (2020). Determination of 18 organophosphorus flame retardants/plasticizers in mussel samples by matrix solid-phase dispersion combined to liquid chromatography-tandem mass spectrometry. *Talanta*, *208*, 120470.

Durham, B. P., Boysen, A. K., Carlson, L. T., Groussman, R. D., Heal, K. R., Cain, K. R., ... Armbrust, E. V. (2019). Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. *Nature Microbiology*, *4*(10), 1706–1715.

Feingersch, R., Philosof, A., Mejuch, T., Glaser, F., Alalouf, O., Shoham, Y., & Béjà, O. (2012). Potential for phosphite and phosphonate utilization by *Prochlorococcus*. *The ISME Journal*, *6*(4), 827–834.

Figueroa, I. A., & Coates, J. D. (2017). Microbial Phosphite Oxidation and Its Potential Role in the Global Phosphorus and Carbon Cycles (Vol. 98, pp. 93–117). Elsevier Ltd.

Foreman, R. K., Björkman, K. M., Carlson, C. A., Opalk, K., & Karl, D. M. (2019). Improved ultraviolet photo-oxidation system yields estimates for deep-sea dissolved organic nitrogen and phosphorus. *Limnology and Oceanography, Methods / ASLO*, *17*(4), 277–291.

Ghassabian, S., Wright, L. A., Dejager, A. D., & Smith, M. T. (2012). Development and validation of a sensitive solid-phase-extraction (SPE) method using high-performance liquid chromatography/tandem mass spectrometry (LC-MS/MS) for determination of risedronate concentrations in human plasma. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, *881–882*, 34–41.

Horiguchi, M., & Kandatsu, M. (1959). Isolation of 2-aminoethane phosphonic acid from rumen protozoa. *Nature*, *184*(Suppl 12), 901–902.

Johnson, W. M., Kido Soule, M. C., & Kujawinski, E. B. (2017). Extraction efficiency and quantification of dissolved metabolites in targeted marine metabolomics: Matrix effects in marine metabolomics. *Limnology and Oceanography, Methods / ASLO*, *15*(4), 417–428.

Johnson, W. M., Longnecker, K., Kido Soule, M. C., Arnold, W. A., Bhatia, M. P., Hallam, S. J., ... Kujawinski, E. B. (2020). Metabolite composition of sinking particles differs from surface

suspended particles across a latitudinal transect in the South Atlantic: Sinking particle metabolite composition. *Limnology and Oceanography*, 65(1), 111–127.

Karl, D. M., & Tien, G. (1992). MAGIC: A sensitive and precise method for measuring dissolved phosphorus in aquatic environments. *Limnology and Oceanography*, 37(1), 105–116.

Kujawinski, E. B., Longnecker, K., Alexander, H., Dyhrman, S. T., Fiore, C. L., Haley, S. T., & Johnson, W. M. (2017). Phosphorus availability regulates intracellular nucleotides in marine eukaryotic phytoplankton: Phosphorus availability regulates intracellular nucleotides. *Limnology and Oceanography Letters*, 2(4), 119–129.

Liu, M., Hashi, Y., Song, Y., & Lin, J.-M. (2005). Simultaneous determination of carbamate and organophosphorus pesticides in fruits and vegetables by liquid chromatography–mass spectrometry. *Journal of Chromatography. A*, 1097(1), 183–187.

Martiny, A. C., Lomas, M. W., Fu, W., Boyd, P. W., Chen, Y.-L. L., Cutter, G. A., ... Moore, J. K. (2019). Biogeochemical controls of surface ocean phosphate. *Science Advances*, 5(8), eaax0341.

Mawhinney, D. B., Hamelin, E. I., Fraser, R., Silva, S. S., Pavlopoulos, A. J., & Kobelski, R. J. (2007). The determination of organophosphonate nerve agent metabolites in human urine by hydrophilic interaction liquid chromatography tandem mass spectrometry. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, 852(1–2), 235–243.

Murphy, J., & Riley, J. P. (1962). A modified single solution method for the determination of phosphate in natural waters. *Analytica Chimica Acta*, 27, 31–36.

Polyviou, D., Hitchcock, A., Baylay, A. J., Moore, C. M., & Bibby, T. S. (2015). Phosphite utilization by the globally important marine diazotroph *Trichodesmium*. *Environmental Microbiology Reports*, 7(6), 824–830.

Repeta, D. J., Ferrón, S., Sosa, O. A., Johnson, C. G., Repeta, L. D., Acker, M., ... Karl, D. M. (2016). Marine methane paradox explained by bacterial degradation of dissolved organic matter. *Nature Geoscience*, 9(12), 884–887.

Sinha, S. N., Vasudev, K., & Rao, M. V. V. (2012). Quantification of organophosphate insecticides and herbicides in vegetable samples using the “Quick Easy Cheap Effective Rugged and Safe”(QuEChERS) method and a high-performance liquid chromatography--electrospray ionisation--mass spectrometry (LC--MS/MS) technique. *Food Chemistry*, 132(3), 1574–1584.

Widner, B., C. Kido Soule, M., Ferrer-Gonzalez, F. X., Ann Moran, M., & Kujawinski, E. B. (2020). Superior and Novel Detection of Small, Polar Metabolites in Saline Samples Using Pre-Extraction Benzoyl Chloride Derivatization and Ultra-High Performance Liquid

Chromatography Tandem Mass Spectrometry (UHPLC MS/MS).

doi:10.26434/chemrxiv.12915488.v1

Wollersen, H., & Musshoff, F. (2007). Chromatographic Techniques for Measuring Organophosphorus Pesticides. In *Chromatographic Methods in Clinical Chemistry and Toxicology* (pp. 139–169). Chichester, UK: John Wiley & Sons, Ltd.

Wu, Z., Gao, W., Phelps, M. A., Wu, D., Miller, D. D., & Dalton, J. T. (2004). Favorable effects of weak acids on negative-ion electrospray ionization mass spectrometry. *Analytical Chemistry*, 76(3), 839–847.

Young, C. L., & Ingall, E. D. (2010). Marine Dissolved Organic Phosphorus Composition: Insights from Samples Recovered Using Combined Electrodialysis/Reverse Osmosis. *Aquatic Geochemistry*, 16(4), 563–574.

Zhang, J.-Z., & Chi, J. (2002). Automated analysis of nanomolar concentrations of phosphate in natural waters with liquid waveguide. *Environmental Science & Technology*, 36(5), 1048–1053.

**CHAPTER 5. IDENTIFICATION OF
ORGANOPHOSPHORUS COMPOUNDS IN MARINE
SAMPLES USING LC-ICPMS-ESIMS**

5.1 INTRODUCTION

As previously covered in Chapter 4, one of the biggest limitations to fully understand P biogeochemical cycling is the lack of appropriate methods. Thanks to the “molybdenum blue method” (Karl & Tien, 1992; Murphy & Riley, 1962; Zhang & Chi, 2002) and the recent efforts made to improve the oxidation of TDP (Foreman et al., 2019) standing stocks of SRP and DOP can be measured quite precisely, which provides precious information on dissolved organic phosphorus (DOP) distribution and has allowed for the identification of P-(co)limited regions. With ^{31}P NMR, P functional groups present in DOP led to the discovery of phosphonates and the acknowledgment of a P redox cycle.

In addition, cycling rates and bioavailability of the P pools using ^{33}P -labeled Pi or adenosine triphosphate (ATP) have been studied. ^{33}P Pi has been used to measure Pi uptake rates in several oligotrophic regions (Björkman et al., 2018; Orchard et al., 2010; Thingstad et al., 2005) as well as phosphonate production rates in the North Atlantic (Ebling et al., 2021; Van Mooy et al., 2015) and more recently in the Eastern Mediterranean Sea (Chapter 2). ^{33}P -ATP has been used as a tool to estimate DOP uptake rates (Björkman et al., 2018; Orchard et al., 2010) and as a way to estimate the biologically available P pools (Björkman & Karl, 2003; Karl & Bossard, 1985). Enzymatic assays, especially for alkaline phosphatase, have also been used to help understand the dynamics of P demand as well as the capacity of microbes to use phosphoesters (Karl & Björkman, 2015). Moreover, several efforts have been made to study the chemical composition and molecular concentrations of the marine organic P pool. Those include polyphosphates (Martin & Van Mooy, 2013; Solórzano & Strickland, 1968), nucleic acids (Karl & Bailiff, 1989), nucleotides (Björkman & Karl, 2001) and monoesters (Strickland & Solórzano, 1966).

However, most of the DOP pool chemical composition remains unknown especially on a molecular level, which prevents a full understanding of the extent of the bioavailability of DOP as well as the exchange and transformation between the different P pools. With these issues in mind we developed an element targeted method optimized for P designed to separate P-containing compounds using liquid chromatography (LC), detect and quantify P-containing compounds using inductively coupled plasma mass spectrometry (ICPMS) and finally to characterize them using electrospray ionization in tandem mass spectrometry (ESIMS/MS). Results were presented in Chapter 4. Due to the absence of a stable isotope for ^{31}P , we could not rely on search algorithms previously used in other element targeted methods, and had to develop a precursor-product ion method to integrate ICP and ESIMS data. This method was previously presented in detail in Chapter 4 as well.

Here, we applied the LC-ICPMS/ESIMS method to cultured isolates of *Prochlorococcus* SB and MIT9313. We were able to identify nucleotides and several other P-containing compounds expected in particulate organic matter (POM), which proved the ability of the method to detect P-compounds leading to their identification. We then applied this method to LMWDOM sample extracted from spent *Prochlorococcus* SB culture media by SPE and were also able to identify very diverse P-compounds showing the broad application of this method. Finally, we worked with a seawater POM sample that had been previously fractionated and purified using ion chromatography like in Chapter 2 and were able to fully separate the P-compounds present in select fractions and identify a majority of them. Our results are presented hereafter.

5.2 MATERIAL AND METHODS

5.2.1 Solvents, standards and columns

Mobile phases were prepared using MQ water, distilled Optima Grade (Fisher Scientific) acetonitrile (AcN) and methanol (MeOH), ammonium formate (AmFm) and formic acid. Phosphorus standards: 2-aminoethylphosphonate (2-AEPn), adenosine monophosphate (AMP), glucose-6-phosphate (G6P), methylphosphonic acid (MPn), ethylphosphonic acid (EtPn) and phosphorus acid were purchased from Sigma Aldrich and 2-hydroxyethylphosphonic acid (2-HEP) was purchased from Manchester Organics (UK). Two different types of HPLC columns: a Zorbax SB-C18 (0.5 x 150 mm, 5 μ m; Agilent, USA) and the Primesep SB (2.1 x 250 mm, 5 μ m, 100 Å or 0.5 x 250 mm, 5 μ m, 100 Å; SIELC, USA) packed with a C18 functionalized with quaternary amine groups were used to separate organic phosphorus compounds from POM and DOM samples. While the Zorbax column separated compounds based on polar interactions, the Primesep SB column separations were based on polar and anion-exchange interactions.

5.2.2 Sample preparation

5.2.2.1 Organic phosphorus in *Prochlorococcus* MIT9313

A 5 L culture of *Prochlorococcus* MIT9313 was grown axenically under constant low light ($\sim 11 \mu\text{mol photon m}^{-2} \text{s}^{-1}$) in Sargasso seawater based Pro99 medium (Moore et al., 2007), but substituting 3.75 μM TAPS as a buffer instead of 1 mM HEPES. To harvest the cells, the cultures were centrifuged, the spent medium discarded, the cell pellet rinsed with filtered seawater, and the pellets frozen at -20°C until analysis. Right before analysis, the pellets were thawed, centrifuged and the seawater supernatant removed. Five milliliters of MQ water were added and the suspension was flash frozen in liquid nitrogen followed by thawing with boiling water. This was repeated 5 times. The samples were centrifuged and the supernatant removed. The supernate was filtered using 0.22 μm syringe filters (PES) and then put into 6 mL Vivaspin tubes with a 5000 molecular weight cutoff. The residue in the Vivaspin tubes was rinsed with 1

mL MQ water, centrifuged and the filtrates were combined then dried at 40°C under vacuum. The extract was resuspended in 1 mL of MQ water and stored in the freezer at -20°C. For the analysis, the sample was concentrated to a volume of 500 µL and 6.4 µL were injected.

5.2.2.2 Organic phosphorus in *Prochlorococcus* SB POM and LMWDOM

A 20 L culture of *Prochlorococcus* SB was grown axenically, under constant light (30 µmol photon m⁻² s⁻¹) in artificial seawater medium AMP1 prepared as described by (Moore et al., 2007), but using 3.75 µM TAPS as a buffer instead of 1 mM HEPES. The cells were harvested by centrifugation, the cell pellets rinsed with filtered Turk Island mix i.e. non amended ASW and frozen at -20°C. Meanwhile, the spent medium was filtered with a 0.22 µm cartridge filter (previously cleaned by rinsing with 1M HCl) and collected in an acid clean 20 L carboy. HMWDOM was recovered from the spent media using a stirred cell fitted with a 1 kDa membrane filter. The filtrate (nominally < 1 kD fraction) was pumped through two ENV+ SPE columns (1 g; previously activated with 12 mL of MeOH followed by 12 mL of MQ) connected in series. The columns were desalted using 12 mL MQ water and LMW hydrophobic DOM eluted with 12 mL MeOH. The MeOH wash of the two columns were reduced in volume to ~ 400 µL in a vacuufuge at 35°C for 2h and stored in the freezer until further analysis. For the analysis, the sample was concentrated to a final volume of 40 µL and 16 µL were injected.

The cell pellet from the equivalent of 2 L of *Prochlorococcus* SB culture was thawed, resuspended in 25 mL of MQ water and split into four, 15 mL falcon tubes. The cells were extracted using a sonicator (30 sec on/off repeated 7 times) and the sample centrifuged to remove the cell debris. The supernatant was collected and transferred to 6 mL Vivaspin tubes with a 5000 MW cutoff. The Vivaspin tubes were centrifuged, the filtrates combined and reduced in volume to ~ 500 µL at 40°C under vacuum then stored in the freezer until further analysis. For the analysis, the sample was concentrated to a final volume of 100 µL and 6.4 µL were injected.

5.2.2.3 North Atlantic POM

POM samples were collected by filtering 8 L of North Atlantic (38° 30.479' N; 68° 0.117' W) surface (5 m) seawater onto 0.22 µm Durapore filters. The filters were flash frozen and stored in a -20°C freezer before analysis. Back in the lab, the filters were placed in microvials to which 1 mL of MQ water was added. The vials were flash frozen in liquid nitrogen then thawed by immersion in boiling water. In total, the samples were subjected to three freeze/thaw cycles to extract intracellular P. The supernatant was recovered and phosphonates separated from phosphates using a preparatory ion exchange chromatograph (Thermo Dionex ISC-2100 equipped with an autosampler and a fraction collector) fitted with an IonPac AS18 column (4 x 250 mm) eluted step-wise from 23 mM KOH (0-15 min) to 90 mM (15-25 min) and held at 90 mM for 10 minutes at a flow rate of 1 mL min⁻¹. One-minute fractions were collected for the ten

first minutes of each run. Each one minute fraction was dried then dissolved in 100 μL MQ water and 20 μL were injected on the ICPMS and the ESIMS.

The protocol followed a method developed by Van Mooy et al. (2015) and modified by Ebling et al. (2021) that is designed to separate ^{33}P -labeled reduced P-compounds from ^{33}P -labeled phosphates in order to measure phosphonate production rates. To verify that the fractions collected in the first 10 minutes only containing reduced P-compounds, all fractions were analyzed by ICPMS. Fractions F4 (3-4 min), F6 (5-6 min) and F7 (6-7 min) which displayed ^{33}P activity were also analyzed by ESIMS (Ebling et al., 2021). Thus, only those fractions are discussed hereafter.

5.3.3 Liquid chromatography instrumentation conditions

Separations were performed using an Ultimate 3000 (Dionex, USA) HPLC. Depending on the sample analyzed, we used different LC conditions. For the North Atlantic POM samples, we used two Primesep SB columns (2.1 x 250 mm) connected in series and eluted at a flow rate of 200 $\mu\text{L}/\text{min}$ with 10% AcN, 90% aqueous AmFm (7 mM) acidified to pH 3 with formic acid. For *Prochlorococcus* POM samples, separations were effected using 3 Primesep SB columns (0.5 x 250 mm) connected in series eluted at 13 $\mu\text{L}/\text{min}$ using a water/AcN gradient as detailed in Table 5.1. For the *Prochlorococcus* SB LMWDOM SPE extract, we used a Zorbax SB-C18 column (0.5 x 150 mm) with a water/MeOH gradient (Table 5.1) flow rate of 40 $\mu\text{L}/\text{min}$.

Table 5.1: Summary of the liquid chromatography conditions used for each sample analyzed

	North Atlantic POM	<i>Prochlorococcus</i> SB and MIT9313 POM	<i>Prochlorococcus</i> SB LMWDOM
Column	2 Primesep SB (2.1 x 250 mm)	3 Primesep SB (0.5 x 250 mm)	Zorbax SB-C18 (0.5 x 150 mm)
Solvent A	10% AcN + 90% 7mM aqueous AmFm; pH 3	Aqueous 7 mM AmFm; pH 3	Aqueous 40 mM AmFm
Solvent B	N/A	50% AcN + 50% aqueous 7mM AmFm; pH 3	100% MeOH + 40 mM AmFm
Flow rate ($\mu\text{L}/\text{min}$)	200	13	40
Elution	Isocratic	0-20 min : 100% A 20-30 min: 10% B 30-40 min: 20% B 40-60 min: 20% B 60-68 min: 100% B 68-100 min: 100% B 100-140 min: 100% A	0-5 min: 5% B 5-65 min: 95% B 65-75 min : 95% B 75-105 min : 5% B

5.3.4 ICPMS instrumentation and conditions

The ICPMS was an iCap Qc (Thermo Fisher Scientific, USA) equipped with a perfluoroalkoxy (PFA) nebulizer delivering a flow between 15–35 $\mu\text{L}/\text{min}$ to a quartz cyclonic spray chamber. The instrument was equipped with a 2 mm ID quartz injector and a quartz torch optimized for use with organic solvents. The sampling interface includes nickel sampling and skimmer cones. The collision cell, or QCell, is a quadrupole in which argon (Ar) was introduced. Measurements were made in CCT mode using an O_2 collision gas for which the flow was adjusted as detailed later. O_2 was introduced into the plasma at a flow rate of 25 mL/min to facilitate oxidation of organics to CO_2 and avoid deposition on the sampling cone (Boiteau & Repeta, 2015). The maximum flow from the HPLC injected into the ICPMS was 50 $\mu\text{L}/\text{min}$ (Boiteau & Repeta, 2015). Due to the limited flow rate, the flow was split post-column for some analyses and only a fraction of the analyte was directed into the ICPMS.

For the North Atlantic POM fractions and the *Prochlorococcus* SB LMWDOM sample, we used an O_2 collision gas flow of 0.75 mL/min and dwell times of 0.02 s and 0.05 s respectively. For the *Prochlorococcus* MIT9313 and SB POM samples, we used a O_2 flow of 0.6 mL/min and a dwell time of 0.5 s based on our optimization work (Chapter 4).

5.3.5 ESIMS instrumentation and conditions

The ESIMS is an Orbitrap Fusion (Thermo Scientific, USA) mass spectrometer. The HPLC eluent was directed into the mass spectrometer through a heated electrospray ionization (H-ESI) inlet. The distance between the capillary and inlet can be adjusted (graduations from 1 to 3, arbitrary units) and was changed for each flow rate used. For flow rates of 10 and 40 $\mu\text{L}/\text{min}$ we positioned the capillary in position 1 while for the 200 $\mu\text{L}/\text{min}$ flow rate the capillary was set to position 2. Because we improved the method in parallel with sample analyses we adjusted the source, Ion Trap and Orbitrap parameters between samples. The settings used for each sample type are summarized in Table 5.2. Other parameters were fixed across the analyses. All samples were run in negative mode with an ionization voltage of 2500V. Orbitrap resolution was 120000 with an AGC Target of 2.0×10^5 and an injection time of 50 ms. For the ion trap, an isolation window of 1.6 m/z, an HCD collision energy of 40% with an AGC Target of 5.0×10^4 and an injection time of 50 ms were used. For the targeted precursor mass, we set the intensity threshold to 50 with a mass tolerance of 0.5 m/z. For the MS^2 Orbitrap parameters we used a 1.6 m/z isolation window, an Orbitrap resolution of 30000 with an AGC Target of 5.0×10^4 and an injection time of 54 ms.

Table 5.25: ESIMS(-) source, Orbitrap and Ion Trap parameters adjusted for each samples

Source parameters	North Atlantic POM	<i>Prochlorococcus</i> SB and MIT9313 POM	<i>Prochlorococcus</i> SB LMWDOM
Sheath gas (arb)	35	15	25
Aux gas (arb)	7	5	5
Ion tube temp ($^{\circ}\text{C}$)	400	275	275
Vaporizer tube temp ($^{\circ}\text{C}$)	300	30	75
MS^1 Orbitrap parameters			
Mass range	75-150/150-1000	75-150/150-1000	75-500
Dynamic exclusion	No	Yes	Yes
MS^2 Ion Trap parameters			
Targeted precursors	96.964/80.97/78.958	78.958	78.958
Scan range mode	Normal	Mass range: 75-84	Mass range: 75-84
MS^2 Orbitrap parameters			
HCD Collision Energy (%)	35	35	45

To be able to determine the mass of the P-containing compounds and acquire their fragmentation pattern or MS^2 , the precursor-product method described in Chapter 4 was used. Briefly, the fragmentation of phosphonates and phosphates (here after called P-containing compounds) yields a common fragment at m/z 78.958 which corresponds to the ion PO_3^- . Thus, the detection of this product ion is used to detect the presence of a P-containing compound in the chromatogram and

triggers the acquisition of a high resolution MS² of the precursor ion. The mass of the precursor ion is then listed and an elemental formula found for that mass.

5.3.6 P-compounds characterization and putative annotation

To determine the elemental formula of the precursor molecular ions, an elemental formula generator available online (<https://www.chemcalc.org/mf-finder>) was used with the following filters: carbon number was allowed to vary between 0 and 100, hydrogen between 1 and 100, nitrogen between 0 and 20, oxygen between 0 and 20, phosphorus between 1 and 2, and sulfur between 0 and 2. If no elemental formula was found, the option of having 1 halogen (fluorine, chlorine, bromine or iodine), or alkaline salt (sodium, magnesium, potassium, calcium) atom was added. Because our standard masses deviated from the monoisotopic mass by a Δ ppm varying between ± 0.6 and 6.3 ppm, we narrowed our search window for elemental formulae to a 10 ppm range. We then ran the possible elemental formulae against the PubChem online database (<https://pubchem.ncbi.nlm.nih.gov/>) to only keep elemental formulae yielding known P-compounds. In most cases, only one elemental formula corresponded to P-compounds.

Using the elemental formula(e) and the precursor ion fragmentation pattern (eq. to MS² spectra), we then putatively annotated the P-containing molecule. To confirm our putative annotations, we then used the MetFrag online tool (<https://msbi.ipb-halle.de/MetFragBeta/>). MetFrag is an *in silico* fragmentation tool which uses the precursor mass to find candidates within a database and the overlap between the experimental MS² spectrum and the MS² spectrum computationally generated for each candidate to attribute a score and rank those candidates. Various parameters can be changed including the adduct type of precursor ion, the candidate search ppm and the relative mass deviation for MS² peaks. Additional candidate filters can be applied and scoring term improved. Here, we looked for candidates within the PubChem database, we chose an adduct type of [M-H]⁻, a 10 ppm window for the candidate search and a relative mass deviation of 5 ppm. We also chose a scoring term that would include spectral similarity and exact spectral similarity with the MoNA database. Additionally, we applied a substructure exclusion filter for the substructure phosphoric acid OP(=O)(O)O when the peak at m/z 96.968 corresponding to the fragment H₂PO₄⁻ was absent in the experimental MS² spectrum. Based on the score obtained for our phosphonate standards (2.4 and 1.7-1.9 with and without exact spectral similarity match respectively) we considered that a score >2.4 or 1.7 to be good and set a cutoff at 2.2 or 1.5 depending if the exact spectral similarity was considered in the score calculation. Our putative annotations almost always fell within the top 5 scores and always scored above the cutoffs. We report the MetFrag rank and score as (#rank, score) for all our attributions hereafter. Using MetFrag allowed us to move from a Level 3 of confidence which corresponds to the

characterization of the molecule to a Level 2 of confidence which corresponds to its putative annotation according to published guidelines (Sumner et al., 2007).

5.3 RESULTS AND DISCUSSION

5.3.1 *Prochlorococcus* SB and MIT9313 POM

We analyzed the samples using 3 Primesep SB columns (0.5 x 250 mm) in series and a water/AcN gradient described earlier. The ICPMS data was acquired using O₂ as a collision gas in CCT mode with a flow of 0.6 mL/min and a dwell time of 0.05 s. The ESIMS data was acquired using the experiment with dynamic exclusion, with $m/z = 78.958$ as a product ion (detected in the ion trap in a $m/z = 75-84$ range at an intensity threshold of 500) and with Orbitrap mass ranges of $m/z = 75-150$ and $m/z = 150-1000$.

5.3.1.1 ICPMS data and interpretation

The ⁴⁷PO⁺ ICPMS chromatograms of *Prochlorococcus* MIT9313 (dark green) and SB (light green) POM sample show multiple P peaks along the gradient (Figure 5.1). In both chromatograms, the largest peak is at 11.0 min, which is slightly before the elution of 2-AEPn. This common peak encompasses all the P-containing compounds that are not retained by the column due to their high polarity and basic character with relatively high pKas similarly to 2-AEPn that is positively charged at pH = 3 (pKa₁ = 2.45 (+1), pKa₂ = 7.0 (0) and pKa₃ = 10.8 (-1)). In both chromatograms, this peak is followed by a smaller peak at 15.4 min. Because those P-containing compounds elute in 100% water (buffered with 7 mM AmFm), they probably have a slightly more acidic character but similar polarity. The next peak in the *Prochlorococcus* MIT9313 is much broader and elutes between 30 and 37 min at the time where the AcN content increases to 10%. These compounds are probably less basic or neutral at pH = 3, similar to AMP. Indeed, AMP (pKa₁ = 0.9 (+1), pKa₂ = 3.8 (0) and pKa₃ = 6.1 (-1)) elutes at 42.3 min despite being mostly in its neutral form at the mobile phase pH. A sharp peak elutes (RT = 42.6 min) in the *Prochlorococcus* MIT9313 sample and is most likely AMP. There is also a large peak for the *Prochlorococcus* SB sample that elutes at 45.4 min. Finally, the chromatogram for each sample has a sharp peak at 48.5 min (*Prochlorococcus* MIT9313) and 57 min (*Prochlorococcus* SB). For *Prochlorococcus* MIT9313 this peak elutes before G6P (RT = 51.1 min) and is within the isocratic hold at 10% AcN. Because G6P is acidic (pKa = 1.49) and polar, the P-containing compound(s) eluting at 48.5 min is either be more polar or more acidic than G6P. For the

Prochlorococcus SB sample this peak also elutes during the hold at 10% AcN but between G6P and 2-HEP (RT = 56.5 min).

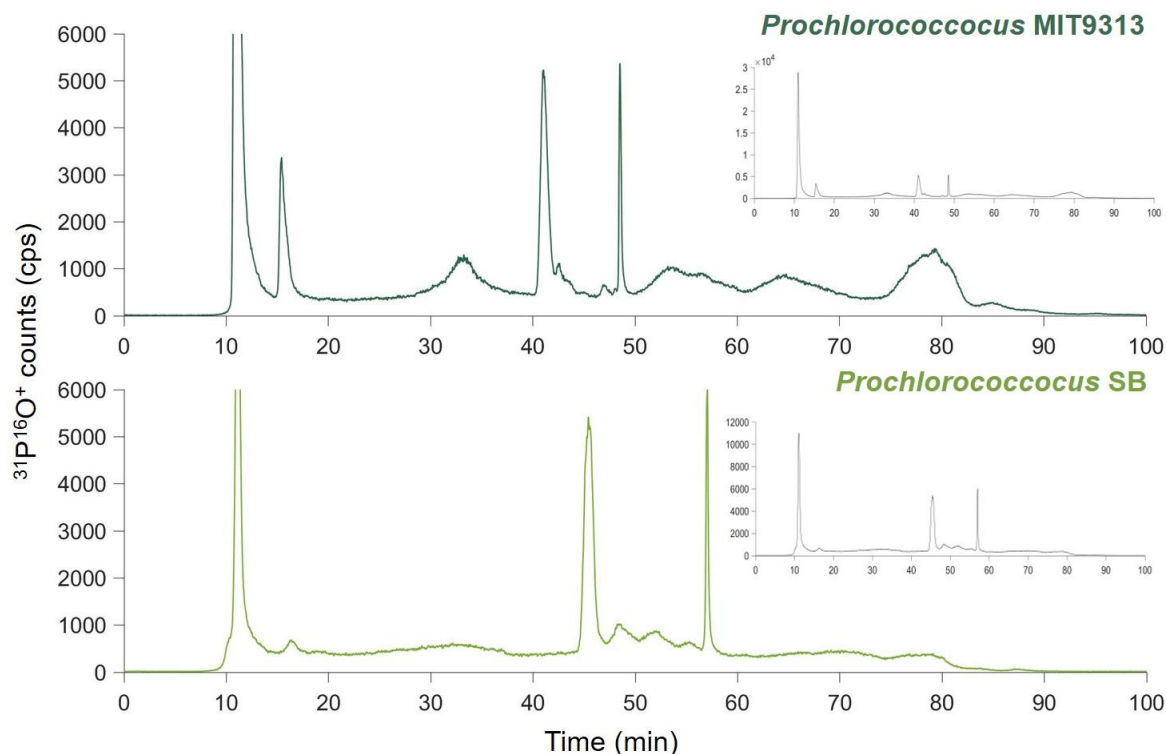


Figure 5.18: *Prochlorococcus* MIT9313 (top panel) and SB (bottom panel) POM sample ICPMS traces. The insets represent the scaled ICPMS traces for each sample. The samples were extracted by freeze/thaw cycles and 3 Primesep SB (0.5 x 250 mm) and a water/AcN gradient described in Table 1 were used.

Due to the reverse-phase and the ion exchange character of this column, it is hard to predict the elution order of specific compounds. For example, 2-HEP (pKa = 2.65), MPn (pKa = 2.12), EtPn (pKa = 2.43) and phosphite (pKa = 2.07) all elute after G6P despite being less acidic. MPn elutes before EtPn despite being more acidic since EtPn is more nonpolar. Based on the chromatograms, it is clear that multiple P-containing compounds were extracted and some peaks include several co-eluting compounds.

5.3.1.2 ESIMS data and interpretation

ESIMS data was acquired using the source parameters described in Table 5.2 and the masses of P-containing molecular ions were found using the precursor-product ion detection method targeting the product ion PO_3^- (m/z 78.9586) within the $m/z = 75-84$ mass window. The samples were run first in the low mass range ($m/z = 70-150$) and then in the high mass range ($m/z = 150-$

1000). In all the following tables and figures, we used dark and light green when displaying data for both *Prochlorococcus* MIT9313 and SB respectively.

For both samples, numerous masses (>20) were collected in the low mass range, but only 2 of them were found to correspond to P-containing compounds: m/z 78.958 (PO_3^-) and MPn (m/z 94.989) which was added as an internal standard. In the high mass range 70 and 25 ions were listed has P-containing compounds for the *Prochlorococcus* MIT9313 and *Prochlorococcus* SB respectively. All contained the product ion m/z 78.958.

Ions corresponding to nucleotide monophosphates were detected in both samples except for cytidine monophosphate (CMP) which was only detected in *Prochlorococcus* MIT9313. For *Prochlorococcus* SB we looked manually for the CMP mass to obtain its extracted ion chromatogram (EIC) and its retention time (Table 5.3). This absence can be explained by the fact that the *Prochlorococcus* SB POM sample is less concentrated than the MIT9313 POM sample (~ 2 times according to the ICPMS traces), that CMP seems to be ionized less easily than other nucleotides (CMP signal intensity is of 4.69×10^4 compare to 2.95×10^5 for GMP), and/or to the fact that *Prochlorococcus* SB is a high light strain whereas MIT9313 is part of the low light adapted clade. High light strains are known for their very small genome and low GC content whereas low light strains have a larger genome with a higher GC content (Biller et al., 2015). This was confirmed by comparing the intensities for the GMP signal in the *Prochlorococcus* SB and MIT9313 samples that are 2.95×10^5 and 3.65×10^6 respectively.

Table 5.36: *Prochlorococcus* MIT9313 (dark green) and SB (light green) nucleotide ESIMS(-) data

Nucleotide	Formula	[M-H] ⁻ monoisotopic mass calc.	[M-H] ⁻ monoisotopic mass found	RT (min)	Main fragments
Adenosine monophosphate (AMP)	C ₁₀ H ₁₄ N ₅ O ₇ P	346.0553	346.0541 346.0538	47.40 51.98	211.000 150.979 134.046 96.969 78.958
Guanosine monophosphate (GMP)	C ₁₀ H ₁₄ N ₅ O ₈ P	362.0502	362.0490 362.0484	79.66 80.41	211.000 150.041 96.969 78.958
Cytidine monophosphate (CMP)	C ₉ H ₁₄ N ₃ O ₈ P	322.0440	322.0433 322.0428	17.51 17.98*	211.000 150.979 138.979 96.969 78.958
Thymidine monophosphate (TMP)	C ₁₀ H ₁₅ N ₂ O ₈ P	321.0488	321.0477 321.0477	89.473 88.845	195.005 176.995 125.025 96.969 78.958
Uridine monophosphate (UMP)	C ₉ H ₁₃ N ₂ O ₉ P	323.0280	323.0269 323.0268	76.36 76.09	211.000 192.989 150.979 138.979 111.019 96.969 78.958
Xanthosine monophosphate (XMP)	C ₁₀ H ₁₃ N ₄ O ₉ P	363.0342	363.0331 363.0328	84.83 84.93	211.000 151.025 96.969 78.958

All nucleotides have the fragments m/z 78.958, 96.969 and 211.000, which correspond to the fragments a, b and c below (Figure 5.2). The other fragments, d and e shown for AMP, are related to the nucleobase. Our putative annotations (Table 5.3) were confirmed by using MetFrag as the nucleotides were always the first ranked candidate and the scores varied between 3.157 (XMP) and (UMP).

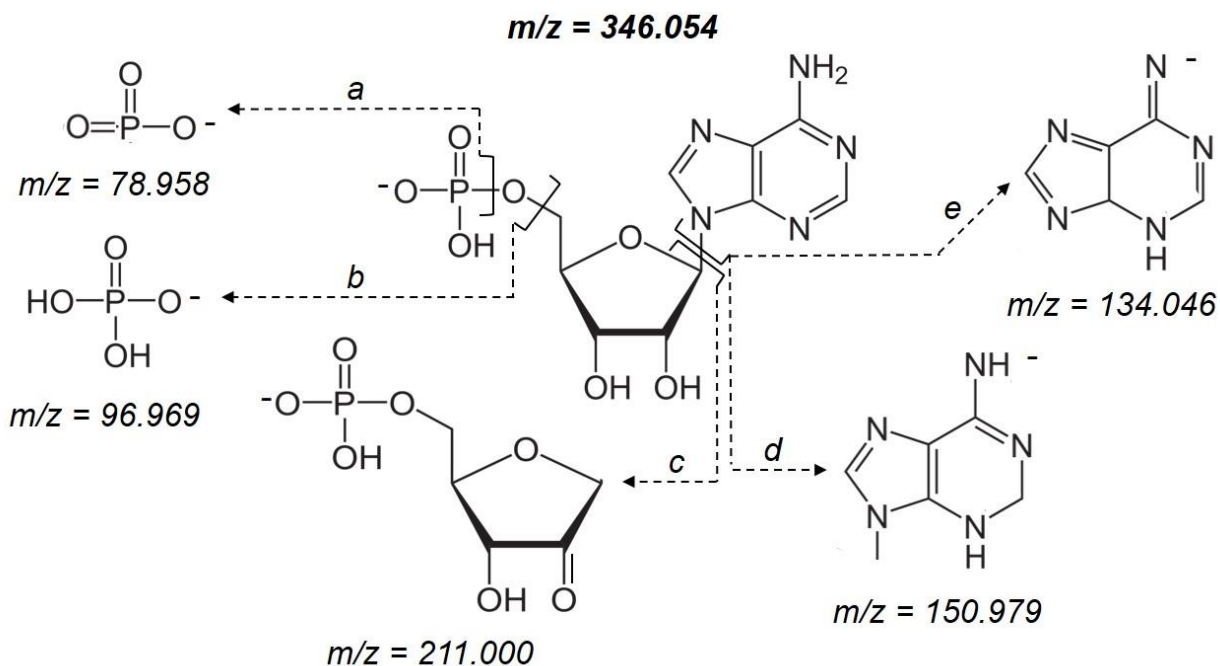


Figure 5.2: AMP fractionation pattern

Additionally, other P-containing precursor ions were detected and their high resolution MS^2 acquired which allowed for their identification. Some were unique to one sample and some were found in both (Table 5.4). Those were involved in glycolysis, related to lipids or involved in other key metabolic processes and their annotation is presented hereafter.

Table 5.47: Putatively identified compounds present in the *Prochlorococcus* MIT9313 (dark green) and SB (light green) POM samples based on their mass and their fragmentation pattern after injection onto the ESIMS(-)

Molecular ion mass	RT (min)	Main fragments	Putative formula	Δ ppm	Putative molecule
184.9844 184.9843	61.62 58.35	96.969 78.958	$C_3H_6O_7P^-$	-3.86	3-phosphoglyceric acid
245.0055	59.214 61.589	165.039 160.841 96.969 78.958	$C_5H_{10}O_9P^-$	-3.04	5-phosphoarabinonic acid
259.0215	44-58	138.979 96.969 78.958	$C_6H_{12}O_9P^-$	-1.52	Glucose-6-phosphate ¹
273.0008	61.589	254.989 138.979 113.023 96.969 78.958	$C_6H_{10}O_{10}P^-$	-1.31	D-glucuronic-1-phosphate
275.0163 275.0160	61.56 76.29	195.050 129.018 96.969 78.958	$C_6H_{12}O_{10}P^-$	-1.85	6-phosphogluconic acid
316.0437	86.670	218.067 175.061 96.969 78.958	$C_8H_{15}NO_{10}P^-$	1.18	Phosphono-2-acetyl-2-amino-3,4,5,6-tetrahydroxyhexanoate
330.0593 330.0593	61.353 58.24	195.005 176.995 134.046 96.969 78.958	$C_{10}H_{13}N_5O_6P^-$	-3.17	Deoxyadenosine-5-monophosphate
333.0580	42-58	315.048 273.934 237.090 152.995 96.969 78.958	$C_9H_{18}O_{11}P^-$	-2.02	Glycerophosphoinositol
375.0684	54.486	259.061 213.016 195.005 96.969 78.958	$C_{11}H_{20}O_{12}P^-$	-2.23	L-alpha-lysophosphatidylinositol

Molecular ion mass	RT (min)	Main fragments	Putative formula	Δ ppm	Putative molecule
402.0089	70.240	383.998 272.955 158.924 110.035 78.958	$C_9H_{14}N_3O_{11}P_2^-$	-3.62	Cytidine diphosphate
426.0201	76.560	408.009 328.045 272.956 173.497 158.924 134.046 96.969 78.958	$C_{10}H_{14}N_5O_{10}P_2^-$	-3.50	Adenosine diphosphate
430.0407	71.608	322.043 279.037 211.000 186.955 138.955 96.969 78.958	$C_{11}H_{18}N_3O_{11}P_2^-$	-2.23	Cytosine-phosphoryl-1-hydroxyethyl-phosphonic acid
662.0993	51.311	426.021 408.010 346.054 328.044 272.956 254.945 192.989 174.979 158.924 134.046 96.969 78.958	$C_{21}H_{26}N_7O_{14}P_2^-$	-3.02	NAD
708.1046 708.1042	47.604 51.35	408.010 346.054 328.044 272.956 158.924 134.046 96.969 78.958	$C_{22}H_{28}N_7O_{16}P_2^-$	-3.07	NADH formate

742.0651	79.0651	540.051 408.010 346.054 328.044 272.956 254.945 192.989 174.979 158.924 134.046 96.969 78.958	$C_{21}H_{27}N_7O_{17}P_3^-$	-1.85	NADP
----------	---------	--	------------------------------	-------	------

¹ or other isomers

In both samples, at a retention time ~60 min, a molecular ion with a mass at m/z 184.9844 is detected. Within our 10 ppm window, only one molecular formula was found for this mass: $C_3H_7O_7P$. Except the two common P fragments at m/z 78.958 and 96.969 i.e. phosphite and Pi respectively, there are no other major fragments. Thus, we putatively annotated this molecule to phosphoglyceric acid and most likely 3-phosphoglyceric acid (#2, 1.899).

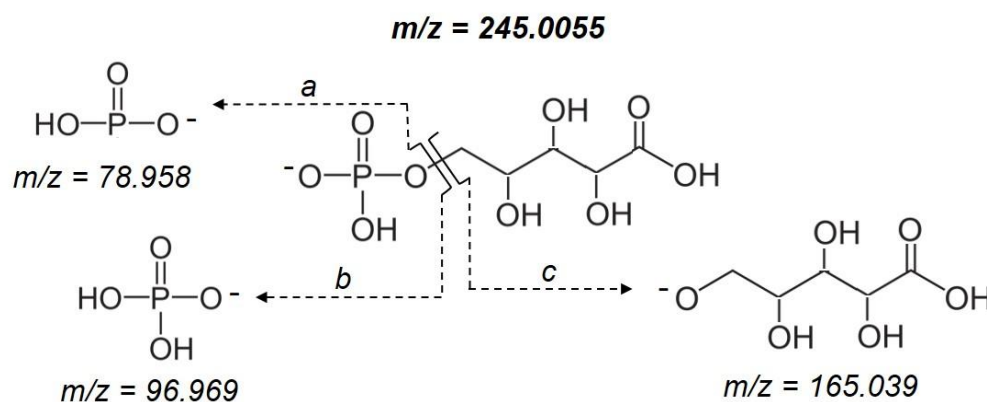


Figure 5.39: 5-phosphoarabinonic acid fragmentation pattern

The *Prochlorococcus* MIT9313 sample has a mass at m/z 245.0055 and only one elemental formula was found within 10 ppm: $C_5H_{11}O_9P$. In addition to the phosphates common P-fragments (a and b; Figure 5.3), a fragment at m/z 165.039 $[M - HPO_3]^-$ was obtained (c; Figure 5.3). Thus, we putatively annotated this molecule to 5-phosphoarabinonic acid (#1, 2.0)

The *Prochlorococcus* MIT9313 sample has a mass at m/z 259.0215 and only one elemental formula was found within 10 ppm: $C_6H_{13}O_9P$. This elemental formula corresponds to several isomers that are likely to be present in a cell: glucose-6-phosphate or mannose-6-phosphate (#2, 3.731) and β -D-fructose-6-phosphate (#1, 3.847). For this mass, the fragmentation contains only

three meaningful fragments at m/z 78.958, 96.969 and 138.979 (a,b and c; Figure 5.4) which could belong to any of those compounds. However, the EIC peak in our sample is much broader than the peak for glucose-6-phosphate and we suspect that several isomers of hexose or pentose phosphate might be eluting within this peak.

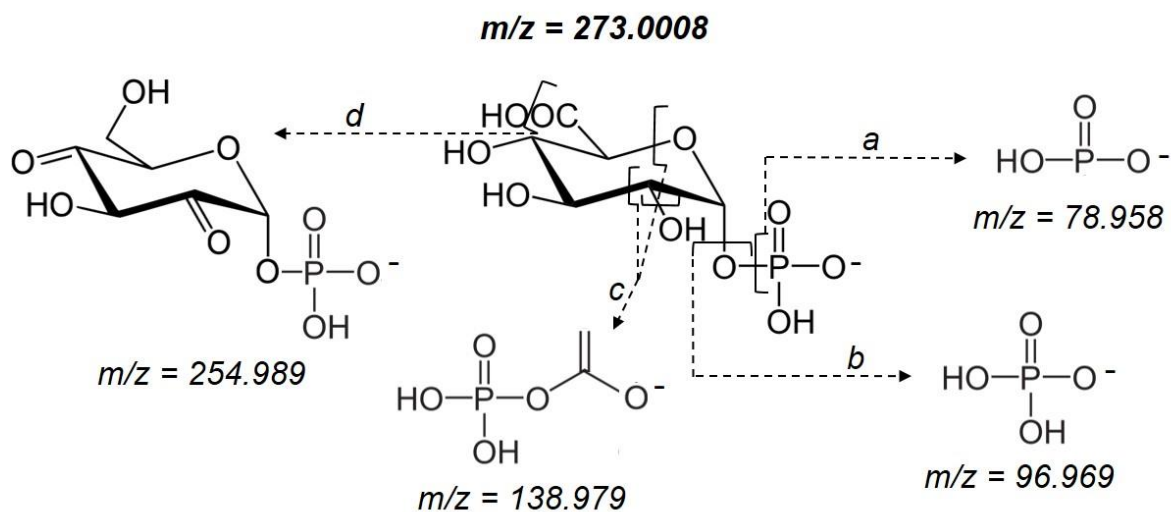


Figure 5.4: D-Glucuronic-1-phosphate fragmentation pattern

The *Prochlorococcus* MIT9313 sample has a mass at m/z 273.0008 and only one elemental formula was found within 10 ppm: $C_6H_{11}O_{10}P$ (Figure 5.4). Due to very similar retention times, the mass at m/z 273.0008 could be the keto form of 6-phosphogluconic acid formed during the ionization process. However, their fragmentation patterns are distinct. We found fragments formed during the fragmentation of D-glucuronic acid at m/z 162.838 and 113.023 as well as a fragment at m/z 138.979. Thus, we putatively annotated this molecule as D-glucuronic-1-phosphate (#3, 1.954).

In both samples, a mass at m/z 275.016 was found and only one elemental formula could be assigned within the 10 ppm range: $C_6H_{13}O_{10}P$. This mass had fragments at m/z 177.039 [$M-HPO_4-2H$] $^-$; 129.018 [$M-HPO_4-COOH-6H$] $^-$ matching spectra for 6-phosphogluconic acid available on the HMDB spectral database (https://hmdb.ca/spectra/ms_ms/5187). Thus, we putatively annotated this molecule as 6-phosphogluconic acid (#4, 1.881).

The *Prochlorococcus* MIT9313 sample displayed a peak with a mass at m/z 316.0437. Two elemental formulae were possible in the 10 ppm range: $C_8H_{16}NO_{10}P$ (Δ ppm = 1.08) and $C_9H_{12}N_5O_6P$ (Δ ppm = -3.15). We used the fragmentation pattern to discriminate between the two. The main fragment is at m/z 218.067 and can be matched with $C_8H_{12}NO_6^-$ (Δ ppm = 2.47) which is [$C_8H_{15}NO_{10}P-H_3PO_4$] $^-$ but also with $C_9H_8N_5O_2^-$ (Δ ppm = -3.67) which is [$C_9H_{11}N_5O_6P-H_3PO_4$] $^-$. Using MetFrag, we found that [(3R,4R,5S)-5-(6-aminopurin-9-yl)-4-hydroxy-

tetrahydrofuran-3-yl] dihydrogen phosphate is the best match for this mass and fragmentation pattern (#1, 1.816) but that one of the major MS² peak at m/z 175.061 is not matched. Additionally, this molecule contains adenine and we should observe adenine related fragments (typically at m/z 150.979 and 134.046). For the second best candidate however, the MS² peak at m/z 175.061 was matched and it was the only molecule with a score higher than 1.7 for which it was the case. Thus, we characterized that molecule as C₈H₁₆NO₁₀P and putatively annotated it as phosphono-2-acetyl-2-amino-3,4,5,6-tetrahydroxyhexanoate (#2, 1.777).

The *Prochlorococcus* MIT9313 sample has a mass at m/z 330.0593 for which two elemental formulae were found within the 10 ppm range: C₉H₁₈NO₁₀P (Δ ppm = 0.89) and C₁₀H₁₄N₅O₆P (Δ ppm = -3.17). We relied on the fragmentation pattern to discriminate between the two. The fragmentation pattern contains a fragment at m/z 195.005, which is common to thymidine-5-monophosphate (Table 5.3) and corresponds to the fragment phosphate plus deoxypentose. Additionally, a fragment at m/z 134.046 corresponds to the adenine base observed in adenosine-5-monophosphate. Thus, we putatively annotated this molecule as deoxyadenosine-5-monophosphate (#1, 3.387)

The *Prochlorococcus* MIT9313 sample has a mass at m/z 333.0580 and only one elemental formula could be assigned within the 10 ppm range: C₉H₁₉O₁₁P (Δ ppm = -2.02). The fragmentation pattern obtained for this mass contained fragments at m/z 152.995 [M-C₆H₁₁O₆]⁻ which is glycerophosphate, 273.934 and 237.090, which are phosphoinositol fragments. Thus we putatively annotated that molecule as glycerophosphoinositol (#1, 2.0)

The *Prochlorococcus* MIT9313 sample has a mass at m/z 375.0684 and only one elemental formula could be assigned within the 10 ppm range: C₁₁H₂₁O₁₂P. The fragmentation pattern showed fragments at m/z 259.061 and 195.005 which were attributed to phosphoinositol and lysophosphatidyl respectively. Thus, we putatively annotated this molecule as lysophosphatidylinositol most likely in its L- α form as this is the biologically active form (#1, 1969).

The *Prochlorococcus* SB sample has masses at m/z 402.0089 (Figure 5.5) and 426.0201 and the only elemental formulae found within the 10 ppm range were C₉H₁₅N₃O₁₁P₂ and C₁₀H₁₅N₅O₁₀P₂ respectively. Both fragmentation patterns contained a fragment at m/z 158.924 (c), which corresponds to diphosphate. The mass at m/z 402.0089 had a fragment at m/z 110.035 (e) which corresponds to cytosine and the second mass one at m/z 134.046 which corresponds to adenine.

Thus, we putatively annotated these molecules as cytidine diphosphate (#1, 3.294) and adenosine diphosphate (#1, 3.184) respectively.

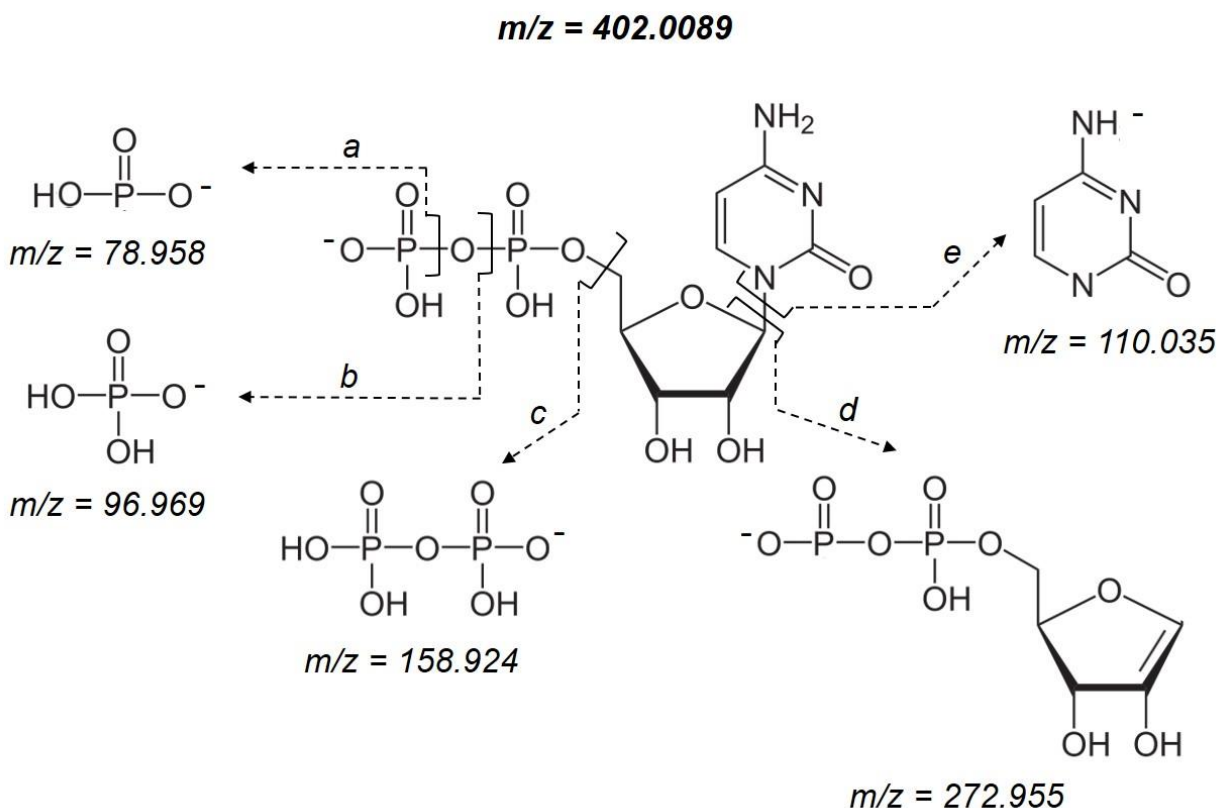


Figure 5.510: Cytidine diphosphate fragmentation pattern

The *Prochlorococcus* SB sample has a mass at m/z 430.0407 and only one elemental formula could be assigned within the 10 ppm range: $C_{11}H_{19}N_3O_{11}P_2$ ($\Delta ppm = -3.62$). The fragmentation pattern contained fragments associated to cytidine monophosphate at m/z 322.043, 211.000 and 138.955. Moreover, the $^{13}C/^{12}C$ isotopic molecular peak ratio indicates that this molecule contains 11 carbons, in agreement with our assignment. It should be noted that no fragment at m/z 158.924 indicative of a diphosphate (Figure 5.5) was observed. Thus, we putatively annotated this molecule as cytosine-phosphoryl-1-hydroxyethyl-phosphonic acid (#2, 1.611).

The *Prochlorococcus* SB sample has masses at m/z 662.0993 and 742.0651. Those two masses showed very similar fragmentation patterns and the mass difference between the two (79.966 Da) which corresponds to HPO_3^- . This led us to conclude that those molecules were identical except for the addition of a new PO_3^- moiety. The fragmentation pattern of the mass at m/z 662.0993 showed the presence of fragments at m/z 346.054, 211.000, 134.046 typically found associated with adenosine monophosphate as well as at m/z 408.010, 272.956 previously found for adenosine diphosphate. Moreover, the fragmentation pattern contained a fragment at m/z

158.924 indicative of a diphosphate (Figure 5.5). The $^{13}\text{C}/^{12}\text{C}$ isotopic molecular peak ratio indicates that the molecule contains 21 carbons. Based on these data, we putatively annotated the molecule with a mass at m/z 662.0993 as nicotinamide adenine dinucleotide (NAD; #1, 2.518) which led to the further putative annotation of the mass at m/z 742.0651 as nicotinamide adenine dinucleotide phosphate (NADP; #1, 2.0).

Both samples had a mass at m/z 708.1046. The fragmentation pattern showed intense fragments at m/z 346.054 and 134.046 typical of AMP (Table 5.3) as well as fragments at m/z 408.010, 272.956 and 158.924 typical of ADP and a fragment at m/z 328.043 observed in NAD (Table 5.4). This mass also corresponds to $[\text{NAD}+\text{HCOOH}]^-$. Thus, we characterized that molecule as $\text{C}_{22}\text{H}_{29}\text{N}_7\text{O}_{16}\text{P}_2$ and putatively annotated it as NADH formate. It should be noted that NADH formate is not found as a candidate by MetFrag despite its existence on the PubChem database and with a mass within the 10 ppm window ($\text{C}_{22}\text{H}_{29}\text{N}_7\text{O}_{16}\text{P}_2$, $\Delta\text{ppm} = -3.07$). In consequence, we were unable to attribute a score to this annotation but based on the high MS^2 spectra similarities are confident in our annotation.

Finally, the *Prochlorococcus* MIT9313 sample had a mass at m/z 693.1930. The EIC peak for this mass had the same retention time and fragmentation pattern as AMP. Moreover, this mass is equal to 2 times the mass of AMP. Thus, we concluded that this mass is due to the presence of an AMP dimer formed in the instrument. However, AMP is the only nucleotide for which we observe the phenomenon.

5.3.1.3 Conclusion on the *Prochlorococcus* POM samples

Using the detection of the product mass at m/z 78.958 corresponding to PO_3^- in the ion trap, 66 ions for the *Prochlorococcus* MIT9313 sample and 25 ions for the *Prochlorococcus* SB sample were detected. In those, some ions had the same retention time and the same fragmentation pattern despite having different masses. This was either due to the presence of the ^{13}C form of the molecule or were due to the presence of compounds co-eluting with the P-compound of interest and with a mass within the 1.6 m/z isolation window that we set. Among the 66 and 25 ions detected, we estimate that ~ 35 and ~ 20 individual compounds were detected for *Prochlorococcus* MIT9313 and for *Prochlorococcus* SB respectively. Of these, we were able to putatively annotated 17 and 14 P-containing compounds respectively.

Interestingly, we only found one putative annotation for a phosphonate in *Prochlorococcus* SB biomass despite the large relative abundance of phosphonate shown by ^{31}P NMR (Chapter 3). This is most likely due to the fact that phosphonates are linked to proteins which are not present in our sample.

5.3.2 *Prochlorococcus* SB LMWDOM

5.3.2.1 ICPMS data and interpretation

The $^{47}\text{PO}^+$ chromatogram contains numerous distinct P peaks eluting between 1 min and 37 min (Figure 5.6). The major peak has a very short retention time, and elutes immediately after the dead volume of the column (RT = 1.86 min vs 0.6 min). This peak is preceded by a small shoulder and followed by two small peaks at 2.67 min and 3.09 min. At 12 min, when reaching 25% MeOH, the background increases and we observe very small features seem to be present starting at 15 min until 19.23 min which correspond to the retention time of the second taller peak. After that we can distinguish multiple peaks of various intensities between at 20.14 min and 30.01 min. After that and until the end of the run, we discern 6 small peaks between 31.94 min and 36.12 min.

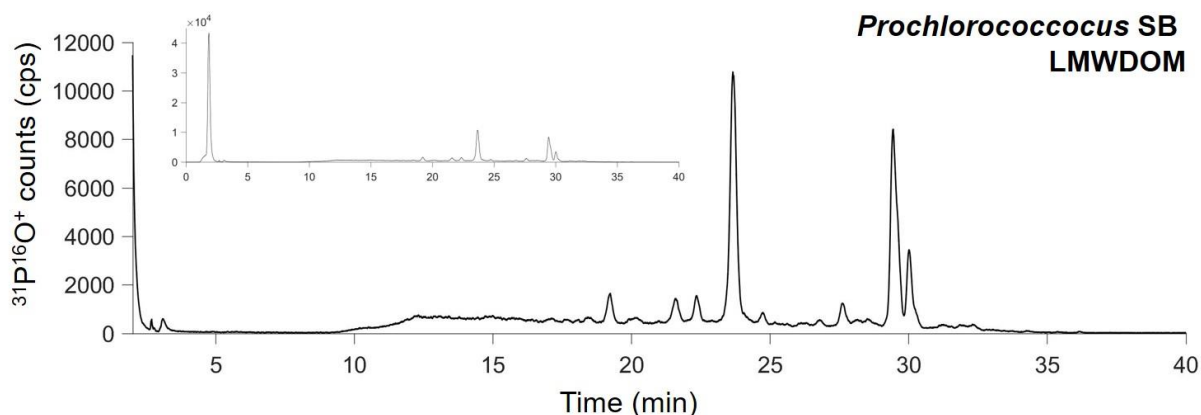


Figure 5.611: *Prochlorococcus* SB LMWDOM sample ICPMS trace. We extracted the sample by SPE (ENV+) using MeOH and used a Zorbax C18 column (0.5 x 150 mm) with a water/MEOH gradient described in Table 5.1.

Because this column (C18) retains nonpolar compounds, LMW phosphonates separated by the Primesep SB column are not retained and elute < 2 minutes. AMP is the only standard for which we observed a peak distinct from the other standards (RT = 1.65 min). Moreover, the retention times should be negatively correlated with polarity and we expect to find lower molecular weight, O-rich compounds at the beginning of the run and higher molecular weight, C-rich compounds -most likely from lipids - later in the run.

5.3.4.2 ESIMS data and interpretation

Table 5.58: Identification of compounds present in the *Prochlorococcus* SB LMWDOM sample based on their mass and their fragmentation pattern after injection onto the ESIMS(-)

Molecular ion mass	RT (min)	Main fragments	Putative formula	Δ ppm	Putative molecule
80.9743 [*]	1.38	No MS2	H ₂ PO ₃ ⁻	1.78	Phosphite
171.0063	1.21	101.9436 96.969 91.7946 78.958	C ₃ H ₈ O ₆ P ⁻	2.63	Glycerophosphoric acid
193.0986	19.03	175.088 149.060 136.052 121.064 107.049 78.958	C ₈ H ₁₈ O ₃ P ⁻	-3.92	Octylphosphonic acid
221.1295	40.17	193.098 177.127 163.076 149.060 136.052 121.065 78.958	C ₁₀ H ₂₂ O ₃ P ⁻	-5.23	Decylphosphonic acid
249.0324	27.34	154.990 108.021 93.034 78.958	C ₁₂ H ₁₀ O ₄ P ⁻	2.93	(2-hydroxyphenoxy)-phenyl-phosphinic acid
285.1243	48.56	270.128 189.127 187.112 137.097 93.034 78.958	C ₁₄ H ₂₂ O ₄ P ⁻	-4.46	2-(2,6-diisopropylphenoxy)ethylphosphonic acid ¹
321.2203	56.68	209.095 96.969 78.958	C ₁₆ H ₃₄ O ₄ P ⁻	2.58	Dioctylphosphate ¹

* extracted mass ion chromatogram ¹ or another isomer

The only elemental formula we found within our 10 ppm range for the mass at m/z 171.0063 was $C_3H_9O_6P$ ($\Delta ppm = 2.63$). The fragmentation pattern and only confirms the presence of a phosphate moiety due to the fragment at m/z 96.969. We putatively annotated this molecule as glycerophosphoric acid (#2, 3.391).

Only one elemental formula, $C_8H_{19}O_3P$, was found within the 10 ppm range for m/z 193.0997 ($\Delta ppm = -3.92$). No fragment at m/z 96.969 was observed which implied that the molecule is a phosphonate. A fragment at 175.089 $[M-H_2O]^-$ as well as intense fragments at m/z 136.052 ($C_8H_8O_2^-$), 121.064 ($C_8H_9O^-$) and 107.049 ($C_7H_7O^-$) which were previously found as key fragments in octylphosphonic acid (<https://mona.fiehnlab.ucdavis.edu/spectra/display/LU029051>). Additional, fragments were observed at m/z 106.989 ($C_2H_4O_3P^-$) and 93.9813 ($CH_3O_3P^-$). Thus, we putatively annotated this molecule as octylphosphonic acid (#3, 1.910). It should be noted that the two best hits in MetFrag correspond to phosphite esters and that we were not able to discriminate between the two functional group. Octylphosphonic acid is a manufactured compound used as an anti-corrosive agent (Felhősi et al., 2002).

A mass at m/z 221.1295 was found. This mass has a mass difference of 28.031 (C_2H_4) with octylphosphonic acid and the only elemental formula found within the 10 ppm range was: $C_{10}H_{23}O_3P$ ($\Delta ppm = -5.23$). This mass had a very similar fragmentation patterns with a large number of common fragments to $C_8H_{19}O_3P$, which we putatively annotated as octylphosphonic acid, such as fragments at m/z 193.098, 136.0516, 121.064 and 149.096. Additional fragments at m/z 177.127 ($C_{12}H_{17}O^-$), 165.0906 ($C_{10}H_{13}O_2^-$) and 150.067 ($C_9H_{10}O_2^-$) were found and confirmed that the compound has a longer alkane chain than octylphosphonic acid. Thus, we putatively annotated this molecule as decylphosphonic acid (#2, 1.917). As before, the phosphite ester was the best hit and we were not able to discriminate between the two. Similarly, decylphosphonic acid is used as a coating agent against corrosion (Felhősi et al., 2002).

The mass at m/z 249.0324 was found to correspond to only one elemental formula within our 10 ppm range: $C_{12}H_{11}O_4P$. This elemental formula corresponded to different isomers of diphenyl phosphate and the fragmentation pattern was used to discriminate between those. The presence of fragments at m/z 93.034 (b) and 154.990 (d) associated with the absence of the fragment at m/z 96.969 typically observed in phosphates, ruled out phenylphenol phosphate as a potential candidate (Figure 5.7). The presence of a fragment at m/z 108.21 (c) led us towards the presence of a phenoxy group. The relatively low intensity of the MS^2 peak at m/z 78.958 seemed to indicate that the molecule is a phosphonodiester. On MetFrag, the best hit was phenyl(phenylperoxy)phosphinic acid. As the peroxide bond is chemically unstable it seemed unlikely that we would find it in our sample. The second hit was (2-hydroxyphenoxy)-phenylphosphinic acid and this molecule could explain the high relative intensity of the MS^2 peak at m/z 93.034. Thus, we putatively annotated this molecule as (2-hydroxyphenoxy)-phenyl-

phosphinic acid (#2, 1.691). To our knowledge, there is no mention of this compound as a natural product.

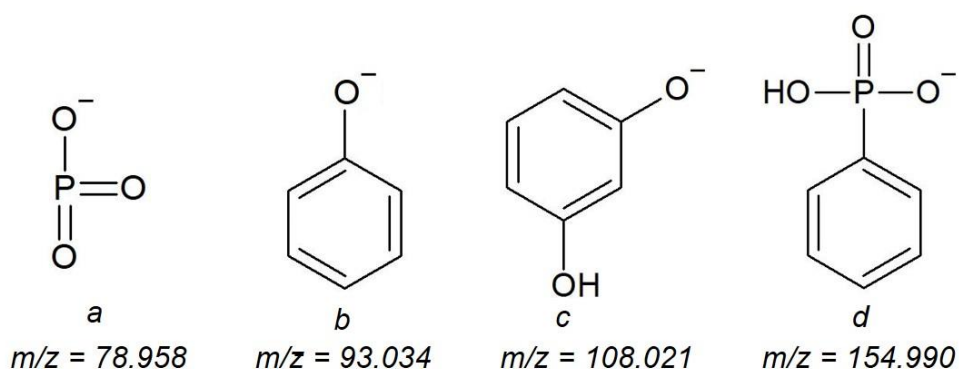


Figure 5.712: Chemical structures of the m/z fragments obtained from the mass m/z 249.0324 fragmentation pattern

A mass at m/z 285.1243 was detected and we only found one elemental formula within the 10 ppm range: C₁₄H₂₃O₄P. The MS₂ spectrum contains numerous peaks and is challenging to interpret but we found some information to rule out several isomers found for this elemental formula. First, the absence of a fragment at m/z 96.969 suggests that this compound is not a phosphate and most likely a phosphonate. We also observe a fragments at m/z 93.034 characteristic of a phenol (Figure 5.7). Additionally, no repeating unit characteristic of relatively long alkane chains were observed, so we hypothesized that there are several carbon groups in addition to the benzene ring. For the rest, we used MetFrag scoring to putatively annotate this molecule. The best hit was obtained for 2-(2,6-diisopropylphenoxy)ethylphosphonic acid which is consistent with the lack of alkane repeating units that would probably be observed for the candidates ranking below. Thus, we putatively annotated this molecule as 2-(2,6-diisopropylphenoxy)ethylphosphonic acid but recognize that other isomers are possible. Records of this molecule were found associated with patents of sedation drugs (EPO Patent No. 2281565:A2, 2011) and there are no records of this molecule as a natural product.

A mass at m/z 321.2203 was found and the only elemental formula within the 10 ppm range was: $C_{16}H_{35}O_4P$. This time, a fragment at m/z 96.969 was observed which indicates that this compound is a phosphate ester. However, this fragment is small and the third most intense fragment was found at m/z 209.0946 corresponding to $C_8H_{18}O_4P^-$ which is octylphosphate or one of its isomer. This suggests that this compound is a symmetrical phosphodiester with 2 C8 groups on each side. Thus, we putatively annotated this molecule as dioctylphosphate (#34, 1.565) but due to the low ranking and the lack of more informative fragments we recognize that other isomers are very possible. There is no record of this compound being a natural product and is compound is used as a metal protector (Chirkunov et al., 2013) or as a lubricant additive according to the EPA.

5.3.4.3 Conclusion on *Prochlorococcus* SB LMWDOM

Interestingly, the *Prochlorococcus* SB LMWDOM sample did not have many common peaks with the extracted biomass. This suggests that *Prochlorococcus* SB produces secondary P-metabolites that are excreted in the growth medium and further suggest that in the marine environment, DOM and POM might have different P chemical composition. Most of the compounds we were able to identify are phosphonates that have not previously been described as natural products, but are rather plastic/metal additives or used as herbicides. Thus, it is unclear if those compounds are contaminants or were produced by *Prochlorococcus* SB. We did not detect phosphonates in the POM sample, which is consistent with our result that phosphonates in *Prochlorococcus* SB are incorporated into proteins.

We were able to find phosphite by its EIC. This would also support the observation that organisms are able to grow on phosphite as a source of P (Martínez, Osburne, Sharma, DeLong, & Chisholm, 2012; Metcalf & Wolfe, 1998; Polyviou, Hitchcock, Baylay, Moore, & Bibby, 2015) and several genomic surveys that show phosphite utilization genes are present in the marine environment (Feingersch et al., 2012; Martinez et al., 2010; Polyviou et al., 2015).

5.3.1 North Atlantic POM

5.3.1.1 ICPMS data and interpretation

Select fractions (F4 (3-4 min), F6 (5-6 min), and F7 (6-7 min)) from the fractionation of North Atlantic POM by ion chromatography were separated on Primesep SB to characterize the suite of phosphonates in the sample. Each fraction had peaks in the ^{31}P -chromatogram that elute within three time windows: 6-10 min, 35-45 min, and > 60 min (Figure 5.8; Table 5.6). Each fraction is different and had unique peaks. To identify the P-containing compounds eluting in each fraction, samples were analyzed by ESIMS using the product-precursor ion method with PO_3^- (m/z 78.958) as the targeted product ion.

Table 5.69: Retention times of the ICPMS peaks for the 3 fractions of interest

	F4	F6	F7
Group 1	6.9	7.3	6.9
	7.3	8.4	8.1
	7.6		9.6
Group 2	36.7		
	41.7	41.6	41.7
	43.7	43.5	43.6
	45.4	45.3	
Group 3	57.8	68.0	68.1
	67.7		
Additional peaks	32.6	32.3	
	51.8	48.5	

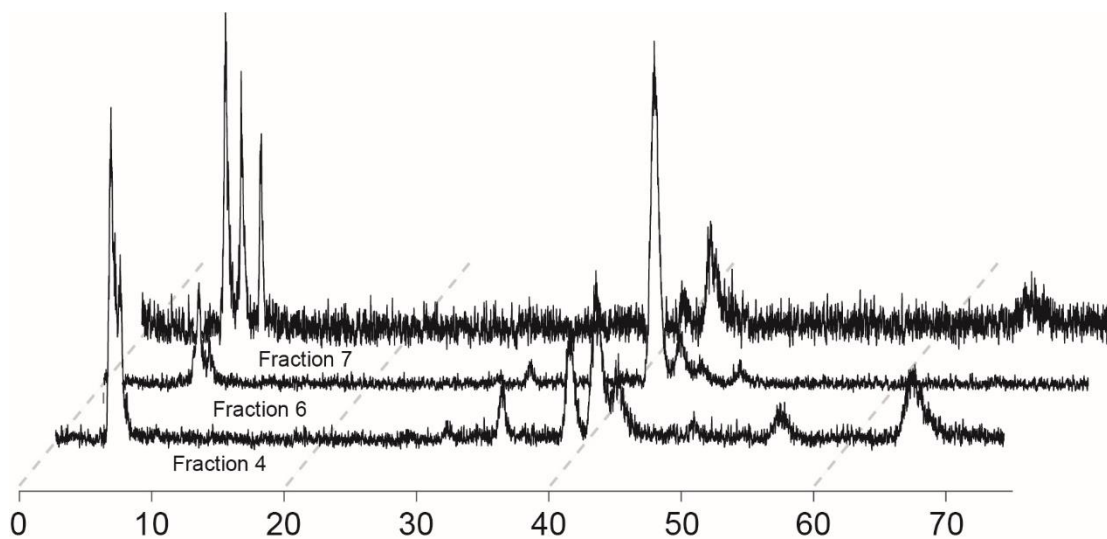


Figure 5.813: Biofest Fractions 4, 6 and 7 (3-4 min, 5-6 min and 6-7 min fraction collection on the IC) ICPMS traces

5.3.1.2 Identification of P-containing compounds in F4

Three precursor ions containing the product ion PO_3^- (m/z 78.958) eluted between 6-10 min in F4. The first peak, at 6.45 min, had a mass at m/z 170.0205. For this mass, the only elemental formula within our 10 ppm window was $\text{C}_3\text{H}_{10}\text{NO}_5\text{P}$. No fragment was observed at m/z 96.969 so we looked for phosphonates with this elemental formula and found several amino phosphonates isomers. However, the MS^2 spectrum only contains an ion for the PO_3^- fragment (78.958) and does not allow us to fully discriminate between the possible isomers. Using MetFrag, we found the best hit for 2-Hydroxy-3-(hydroxyamino)propyl]phosphonic acid (#1, 1.9122) so we made that putative annotation but recognize that another isomer is also likely. Based on the retention time, this elemental formula makes sense as it contains a nitrogen and our N-containing phosphonate standard, 2-AEPn have a similar retention. This compound should correspond to the peak observed at 6.9 min on the ICPMS trace (Figure 5.8, Table 5.6).

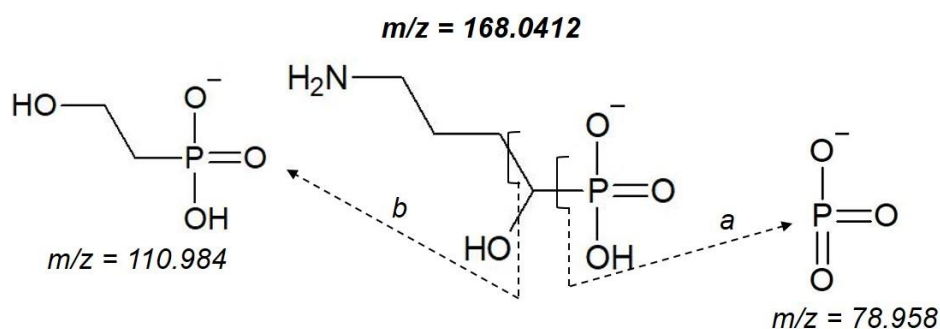


Figure 5.914: 4-amino-1-hydroxy-butylphosphonic acid fragmentation pattern

The second peak, eluting at 6.52 min had a mass at m/z 168.0412. The only elemental formula within the 10 ppm range is $\text{C}_4\text{H}_{12}\text{NO}_4\text{P}$ ($\Delta\text{ppm} = -8.15$). This formula corresponds to several isomers but, as before, the absence of H_2PO_4^- fragment allows us to eliminate phosphates and conclude that this compound is most likely an amino phosphonate. We find a fragment at m/z 110.984 (b) corresponding to $\text{CH}_4\text{O}_4\text{P}^-$, which indicates that the 4th oxygen is 1 carbon away from the P group (Figure 5.9). On MetFrag, the best hit found was for 4-amino-1-hydroxy-butylphosphonic acid (#1, 1.986) so we made that putative annotation (Figure 5.9) but recognize that another isomer is also likely.

The last peak in this group elutes at 6.93 min and has a mass at m/z 242.0775 (Figure 5.10). This mass corresponds to one elemental formula within the 10 ppm range: $\text{C}_7\text{H}_{18}\text{NO}_6\text{P}$ ($\Delta\text{ppm} = -7.64$). This time, the fragmentation pattern contains a H_2PO_4^- fragment (b) typical of phosphoesters but of very low intensity, which indicates that the compound is most likely a phosphodiester. The fragmentation pattern suggests that the N atom is at the other extremity of the molecule than the O-containing part due to the presence of the fragments at m/z 152.994 (c, $\text{C}_3\text{H}_6\text{O}_5\text{P}^-$), 171.004 (d, $\text{C}_3\text{H}_8\text{O}_6\text{P}^-$) and 168.041 (e, $\text{C}_4\text{H}_{11}\text{NO}_4\text{P}^-$). Moreover, those fragments

were found for glycerophosphate and confirmed the presence of a glycerol group in that molecule (<https://mona.fiehnlab.ucdavis.edu/spectra/display/MoNA017590>). On MetFrag, the best hit found was 2,3-dihydroxypropyl-3-(methylamino)propyl hydrogen phosphate (#1, 1.842) which corresponds to the characteristics described above so we made that putative annotation (Figure 5.10) but recognize that other isomers are also likely.

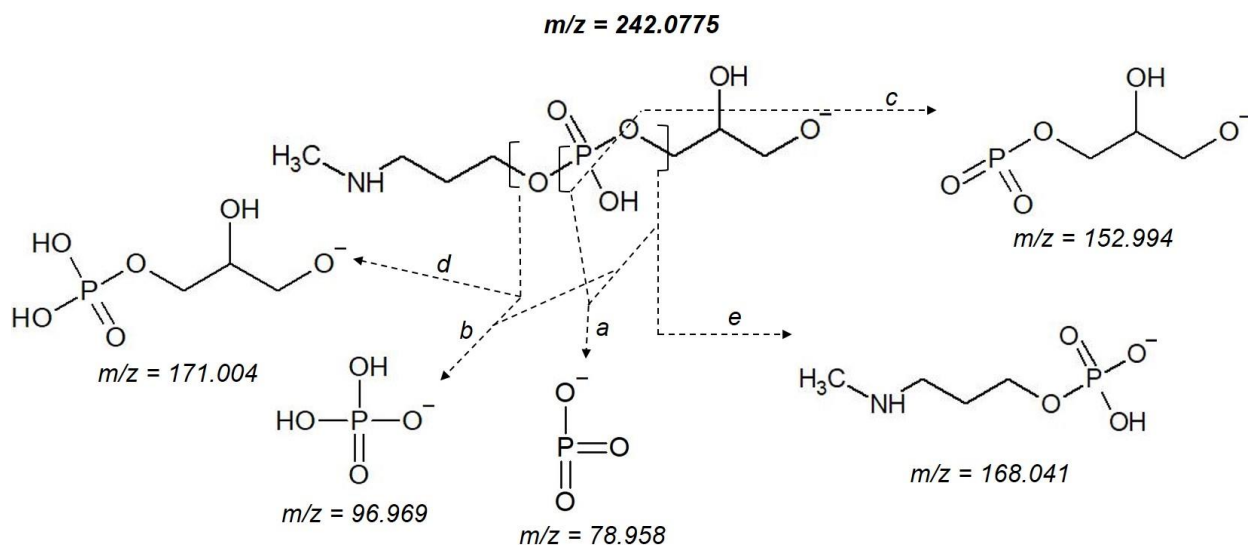


Figure 5.1015: 2,3-dihydroxypropyl-3-(methylamino)propyl hydrogen phosphate fragmentation pattern

The next peak in F4 elutes at 33.0 min and has an m/z of 245.0409. This mass corresponds to one elemental formula within our 10 ppm range: $C_6H_{15}O_8P$ ($\Delta ppm = -7.06$). For this mass, identical fragments to the compound with m/z 242.0775 and glycerophosphate were found i.e. m/z 96.968 (b, Figure 5.10), 152.994 (c, Figure 5.10) and 171.005 (d, Figure 5.10) but the fragment at m/z 96.968 was of low intensity suggesting the presence of a diester. Thus, we putatively annotated this molecule as glycerophosphoglycerol (#9, 1.875). Other isomers were ranked better but the relatively high intensity observed for the MS^2 peak at $m/z = 152.994$ led us to speculate that we have a symmetrical diester.

In the group of peaks eluting between 35-45 min we were only able to characterize Pi eluting at 46.10 min. Finally, was part of the third group of peaks eluting > 60 min, we were able to identify phosphite at 75.41 min.

For F4, the precursor-product method gave us five masses corresponding to P-containing compounds and we were able to identify them all as well as phosphate. Thus we identified 6 compounds out of the 11 peaks (Figure 5.11). This is surprising as the peaks at 41.7 and 45.4 min on the ICPMS trace are relatively intense. Because we were able to detect P-containing compounds in that time frame for F6 and F7 (see after), we know that this is not due to a methodology issue. Thus, we argue that our issue to find compounds for all the P-ICPMS peaks

in F4 is due to the presence of abundant S-containing compounds that have fragments at $m/z = 80.964$ which is within the $\pm 0.5 m/z$ detection range around the precursor $m/z = 80.974$ that we selected to detect phosphite. This is why, moving forward with the samples discussed above, we decided to not target that precursor anymore and only focus the common fragment at $m/z 78.958$.

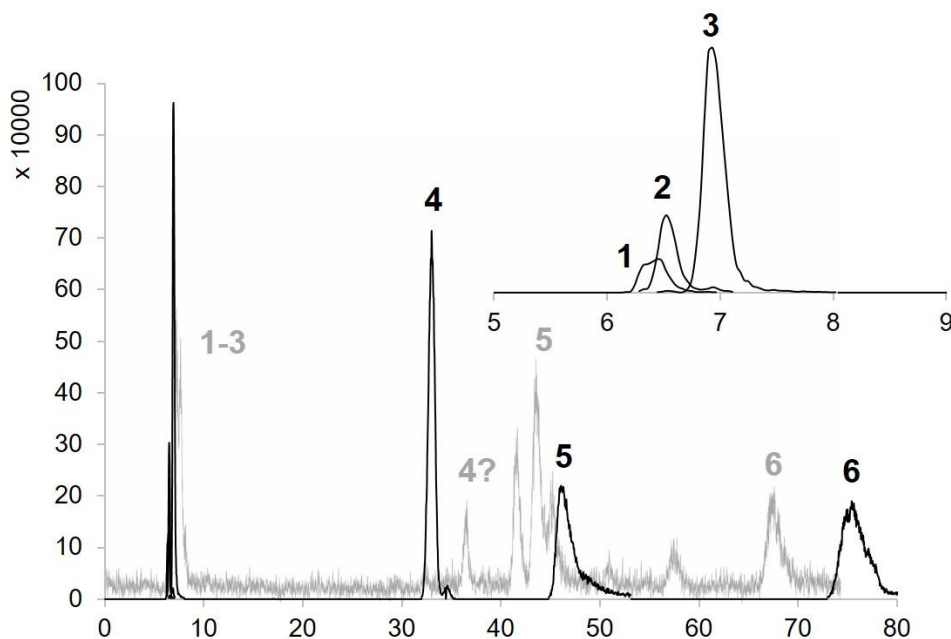


Figure 5.1116: F4 ICPMS trace (grey) and ESIMS reconstructed trace (black) using the EIC for each identified compound. Based on our molecular identifications and the elution order found in the ESIMS data, the compounds eluting are 1: 2-Hydroxy-3-(hydroxyamino)propyl]phosphonic acid (6.9, 6.45 min); 2: 4-amino-1-hydroxy-butylphosphonic acid (7.3, 6.52); 3: 2,3-dihydroxypropyl-3-(methylamino)propyl hydrogen phosphate (7.6, 6.93); 4: glycerophosphoglycerol (36.7, 33.04); 5: Pi (43.7, 46.10) and 6: phosphite (67.7, 75.41)

Despite the shift in retention time between the ICPMS and the ESIMS data, we were able to pair the peaks on the ICPMS trace with the compounds identified with the ESIMS data. For the glycerophosphoglycerol and Pi peaks (4 and 5, Figure 5.11) those attributions were harder to make because we did not detect the other compounds in that group and could not rely on the elution order as we did for the first peak group. However, since Pi has the same retention time in F4 and F7, we assumed that the matrix effect for Pi between those two fractions is similar and we decided to pair Pi with the peak eluting at 43.7 min on the ICPMS trace as we determined that Pi eluted at 43.5 min in F7. The peaks are all shifted differently and those attributions remain to be confirmed.

5.3.1.3 Identification of P-containing compounds in F6

In F6, the first mass elutes at 6.47 min and with m/z 168.0412 which corresponds to $C_4H_{12}NO_4P$ ($\Delta ppm = -8.15$) and which was also found in F4. The ion has the same fragmentation pattern (Figure 5.9) and was putatively annotated as 4-amino-1-hydroxy-butylphosphonic acid. The second peak to elute has a mass at m/z 242.0775 (RT = 6.95 min) which corresponds to $C_7H_{18}NO_6P$ ($\Delta ppm = -7.64$) and was also present in F4 (Figure 5.10). This molecule was putatively annotated as 2,3-dihydroxypropyl-3-(methylamino)propyl hydrogen phosphate. Thus, these two compounds are common to F4 and F6 (Table 5.6). Based on the elution order, we find that for both fractions, 4-amino-1-hydroxy-butylphosphonic acid elutes at 7.3 min on the ICPMS which means that 2,3-dihydroxypropyl-3-(methylamino)propyl hydrogen phosphate has a slightly a different retention time (7.6 min in F4 and 8.1 min in F6) between the two fractions. This could be explained by a shift due to the sample matrix between the two samples or the presence of different isomers in each fractions (same mass but different configuration leading to a different retention time).

The ICPMS trace for F6 indicates the presence of at least three P-containing compounds between 35-45 min. The first ion to elute in that group (RT = 40.15 min) is unique to F6 and has a mass at m/z 171.0045. The fragmentation pattern of the precursor ion is the same as glycerophosphoric acid described previously (Table 5.4). The second mass to elute has a mass at m/z 240.0618 (RT = 44.26 min). The only elemental formula corresponding to that mass is $C_7H_{16}NO_6P$ ($\Delta ppm = -7.91$) and here again, the absence of a fragment at m/z 96.969 indicates that this compound is a phosphonate. The fragmentation pattern has an intense fragment at m/z 110.984 indicating the presence of a $CH_4O_4P^-$ fragment and a fragment at m/z 168.041 - which could correspond to 4-amino-1-hydroxy-butylphosphonic acid, molecule found in F4 - due to the presence of a $C_4H_{11}NO_4P^-$ (Figure 5.9). On MetFrag, the best hit was obtained for [2-azaniumyl-3-hydroxy-2-(hydroxymethyl)propoxy]-[(2R,3S)-3-methyloxiran-2-yl]phosphinate but this compound is classified as non-live. The next phosphonate we found with an O placed 1 carbon away from the P group and with a score >1.5 was (1,1-diethoxy-2-formamido-ethyl)phosphonic acid (#6, 1.548) and the two diethoxy branches could explain the high relative intensity of the MS^2 peak at m/z 110.984. Thus, we putatively annotated this molecule as (1,1-diethoxy-2-formamido-ethyl)phosphonic acid. Finally, we also find Pi (RT = 46.62 min) phosphite (RT = 75.45) in F6 by manually looking for their exact masses.

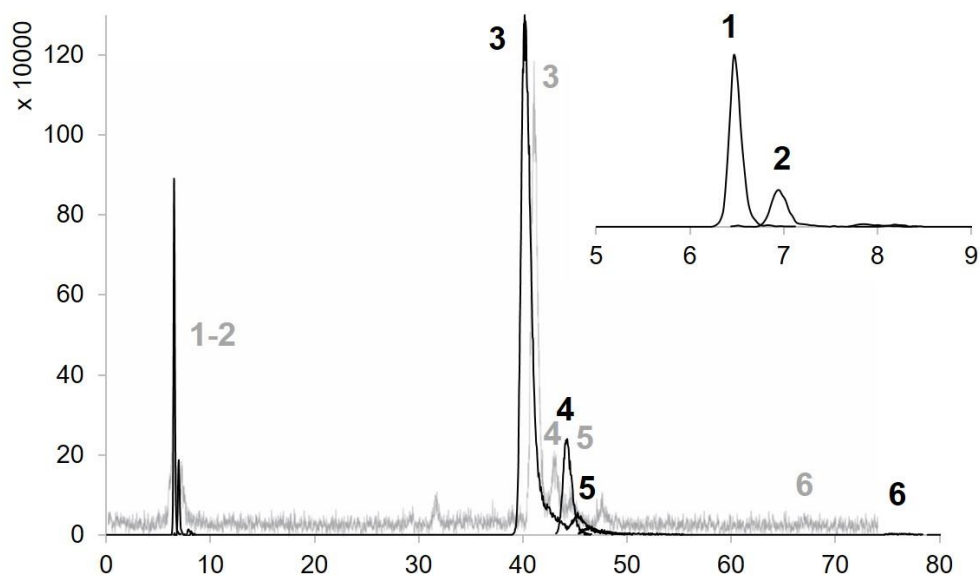


Figure 5.1217: F6 ICPMS trace (grey) and ESIMS reconstructed trace (black) using the EIC for each identified compound. Based on our molecular identifications and the elution order found in the ESIMS data, the compounds eluting are 1: 4-amino-1-hydroxy-butylphosphonic acid (7.3, 6.47); 2: 2,3-dihydroxypropyl-3-(methylamino)propyl hydrogen phosphate (8.4, 6.93); 3: glycerophosphoric acid (41.6, 40.15); 4: (1,1-diethoxy-2-formamido-ethyl)phosphonic acid (43.5, 44.26); 5: Pi (45.3, 46.10) and 6: phosphite (68.0, 75.45)

For F6, we detected 4 compounds using the precursor-product method and were also able to find Pi and phosphite by looking directly for their masses. According to the ICPMS trace (Figure 5.8), two more P-containing compounds were present but were not detected using our precursor-product method. We speculate that this is due to their relative low abundances in the sample.

As in F4, the peaks between the ICPMS and the ESIMS traces are shifted in different manners most likely due an inconsistency in the mobile phase but based on the relative intensity of the peaks and the elution order, we were able to pair up all the peaks and compounds (Figure 5.12).

5.3.1.4 Identification of P-containing compounds in F7

The first ion eluting in F7 has a retention time of 6.14 min and has a mass at m/z 214.0462. For this mass, only one elemental formula was found within the 10 ppm range and is $C_5H_{14}NO_6P$. This mass fragmentation pattern does not contain a fragment at m/z 96.969 suggesting that this compound is a phosphonate. We found fragments at m/z 140.010 (b, Figure 5.13) and 152.994 (c, Figure 5.13). Thus, we putatively annotated this molecule as (2-amino-1-hydroxyethyl)-(2,3-dihydroxypropoxy)phosphinic acid (#2, 1.954).

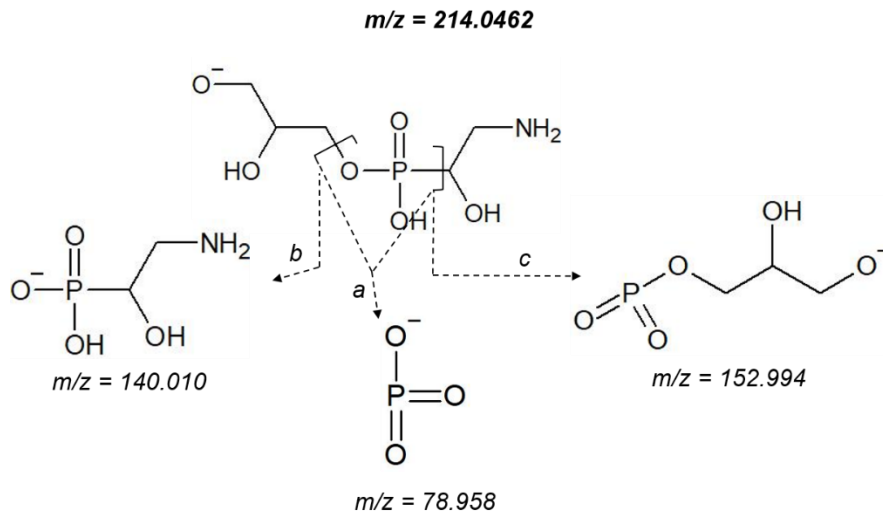


Figure 5.13: (2-amino-1-hydroxyethyl)-(2,3-dihydroxypropoxy)phosphinic acid fragmentation pattern

The second mass to elute in this group is at m/z 140.0102 (RT = 6.45 min). The only elemental formula within the 10 ppm range is $C_2H_8NO_4P$ ($\Delta ppm = -7.64$). No fragment at m/z 96.969 ($H_2PO_4^-$) was observed which implies that the compound eluting is a phosphonate. The rest of the fragmentation pattern did not allow us to discriminate between the different amino phosphonate isomers. On MetFrag, the best hit found was 1-amino-2-hydroxy-ethylphosphonic acid (#1, 1.922) so we made that putative annotation but recognize that another isomer is also possible. The third mass to elute in this group is at m/z 168.0412 which is also detected in F4 and F6 and was attributed to the presence of 4-amino-1-hydroxy-butylphosphonic acid (Figure 5.9). We found a fourth mass corresponding to the peak in the ICPMS trace at 9.6 min. This ion has mass at m/z at 306.0467 (RT = 9.97 min). We found only one elemental formula for this mass within the 10 ppm range: $C_9H_{15}N_3O_7P$. Thanks to the presence of fragment m/z 96.968 (a) we know that this compound is a phosphoester. Moreover, we found fragments at m/z 110.035 (b) and 195.004 (c) corresponding to cytosine and phosphomethyl deoxypentose respectively (Table 5.3). Thus, we putatively annotated this molecule as deoxycytidine monophosphate (#1, 2.346).

In the region eluting between 35-45 min, F7 contains two peaks. The two of them have the same retention times as peaks in F4 and F6 and even though the precursor-product method did not detect any P-containing compounds, we were able to manually confirm that the first peak to elute in this group has a m/z at 240.0618 (RT = 44.22 min) which was putatively annotated as (1,1-diethoxy-2-formamido-ethyl)phosphonic acid, while the second peak (RT = 46.16) was identified as Pi. The last peak at 68.1 min on the ICPMS trace, which has a similar retention time to the last peak in F4, was identified as phosphite (RT = 75.74 min on the ESIMS trace).

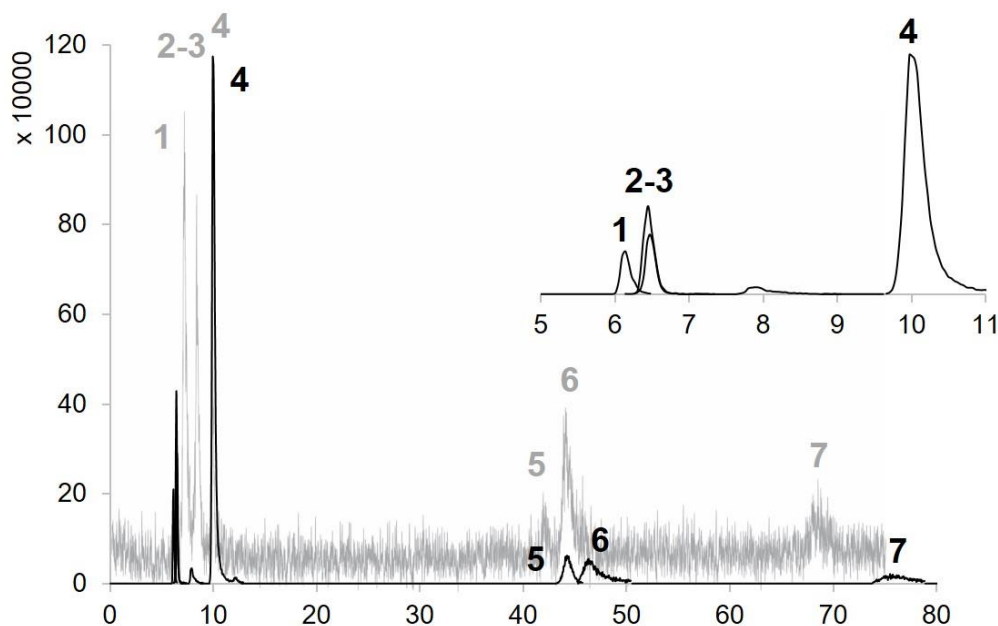


Figure 5.14: F7 ICPMS trace (grey) and ESIMS reconstructed trace (black) using the EIC for each identified compound. Based on our molecular identifications and the elution order found in the ESIMS data, the compounds eluting are 1: (2-amino-1-hydroxyethyl)-(2,3-dihydroxypropoxy)phosphinic acid (6.9, 6.14); 2: 1-amino-2-hydroxyethylphosphonic acid (8.1, 6.45); 3: 4-amino-1-hydroxy-butylphosphonic acid (8.1, 6.47); 4: deoxycytidine monophosphate (9.6, 9.97); 5: (1,1-diethoxy-2-formamido-ethyl)phosphonic acid (41.7, 44.22), 6: Pi (43.6, 46.16) and 7: phosphite (68.1, 75.74)

For F7, we identified all the peaks and even found that the peak at 8.1 min on the ICPMS was a result of the co-elution of two compounds, showing the potential of this method to detect a large number of compounds. Using the relative intensities of the peaks and the elution order we were able to assign compounds for each peak in the ICPMS and ESIMS chromatograms (Figure 5.14).

5.4 CONCLUSIONS

We developed a new element targeted method in order to characterize organophosphorus compounds in marine samples. ICPMS detects all P-containing compounds by monitoring the $^{47}\text{PO}^+$ ion allowing us to determine the retention times and amounts of P-containing compounds. In parallel, ESIMS is a soft ionization technique that detects the molecular ions targets fragmentation based on the presence of a product ion. In our case, we used the detection of the product ion PO_3^- ($m/z = 78.958$) in order to acquire MS^2 only for P-containing ions. This allowed us to find the masses for P-containing compounds and allowed us, through their fragmentation spectra, to identify many of them. We used *Prochlorococcus* biomass as a way to test the method and were able to identify a relatively large number of P-containing compounds, including nucleotide monophosphates, molecule involved in sugar cycles such as glucose-6-phosphate or lipid synthesis with phosphoglyceric acid and glycerophosphoinositols as well as

molecules involved in key metabolic processes such as NAD and NADP. We also tested the method on *Prochlorococcus* SB LMWDOM recovered from spent medium. The ICPMS trace showed the presence of numerous P peaks and we obtained target masses for which high resolution MS² spectra were acquired. Using those MS² we were able to identify a suite of P-compounds. We found phosphomono- and diesters as well as phenyl and alkylphenyl phosphonates. If those compounds are all produced by *Prochlorococcus* SB and are not contaminants, which remains to be investigated, our results show the breath of what one unique strain could produce. Finally, we investigated organophosphorus compounds extracted from suspended POM collected from North Atlantic surface water. These samples had been fractionated by ion chromatography and had a simpler matrix therefore P peaks were well separated but some had a relatively low intensity. Despite the relatively low abundance of those compounds, we were able to detect and identify most of them. In those samples, we detected glycerophosphate but also diverse aminophosphonates which are natural products. Some are found in bacteria sphingophosphonolipids (Jayasimhulu et al., 2007) and aminophosphonates have been found associated with proteins (Quin, 1964; Shelburne & Quin, 1967). In those samples as well as the LMWDOM sample, we detected phosphite. It is unlikely that this phosphite derives from the degradation of phosphonate during sample preparation or storage. Indeed, other fractions of the North Atlantic POM sample, which also contain phosphonates, did not display peaks around the retention times at which phosphite elutes (data not shown). Finding and being able to detect phosphite is of great interest due to capability of some marine organisms to grow on phosphite and due to the genomic evidence for phosphite transporters in the marine environment. To our knowledge, this is the first measurement of phosphite in POM.

Looking ahead, a chromatographic separation capable of resolving polar compounds such as a HILIC column would be beneficial. Some compounds such as triesters, phosphinates, phosphorothioates (P-S bond) or phosphonamides (P-N bond) might also not be detected if their fragmentation does not yield PO₃⁻ or HPO₄⁻ product ions. This problem could be addressed by new algorithms that better align ICPMS and ESIMS data.

This study represents the first screening of the marine organophosphorus using a dual ICPMS and ESIMS approach. Regardless of the type of sample analyzed, we were able to detect a relatively large number of P-containing compounds and identify some of them showing that this method can be further developed applied to environmental samples. Thus, this method is potentially a powerful tool enabling us to survey the marine organophosphorus pool in an untargeted manner and will allow us to develop new methods focused on specific P-compounds to study their biogeochemical cycle as well as their importance for the overall marine P cycle in the future.

5.5 REFERENCES

- Billler, S. J., Berube, P. M., Lindell, D., & Chisholm, S. W. (2015). Prochlorococcus: The structure and function of collective diversity. *Nature Reviews. Microbiology*, *13*(1), 13–27.
- Björkman, K. M., & Karl, D. M. (2001). A novel method for the measurement of dissolved adenosine and guanosine triphosphate in aquatic habitats: applications to marine microbial ecology. *Journal of Microbiological Methods*, *47*(2), 159–167.
- Bjorkman, K. M., & Karl, D. M. (2003). Bioavailability of dissolved organic phosphorus in the euphotic zone at station ALOHA, North Pacific Subtropical Gyre. *Limnology and Oceanography*, *48*(3), 1049–1057.
- Björkman, Karin M., Duhamel, S., Church, M. J., & Karl, D. M. (2018). Spatial and Temporal Dynamics of Inorganic Phosphate and Adenosine-5'-Triphosphate in the North Pacific Ocean. *Frontiers in Marine Science*, *5*, 235.
- Boiteau, R. M., & Repeta, D. J. (2015). An extended siderophore suite from *Synechococcus* sp. PCC 7002 revealed by LC-ICPMS-ESIMS. *Metallomics: Integrated Biometal Science*, *7*(5), 877–884.
- Chirkunov, A. A., Gorbachev, A. S., Kuznetsov, Y. I., & Andreeva, N. P. (2013). Adsorption of dioctyl phosphate and inhibition of dissolution of low-carbon steel in neutral solution. *Protection of Metals and Physical Chemistry of Surfaces*, *49*(7), 854–858.
- Ebling AE, Acker M, Repeta DJ and Van Mooy BAS. (2021). Phosphorus Redox Cycling in the Western North Atlantic. *Limnology and Oceanography*.
- Feingersch, R., Filosof, A., Mejuch, T., Glaser, F., Alalouf, O., Shoham, Y., & Béjà, O. (2012). Potential for phosphite and phosphonate utilization by *Prochlorococcus*. *The ISME Journal*, *6*(4), 827–834.
- Felhösi, I., Telegdi, J., Pálinkás, G., & Kálmán, E. (2002). Kinetics of self-assembled layer formation on iron. *Electrochimica Acta*, *47*(13), 2335–2340.
- Foreman, R. K., Björkman, K. M., Carlson, C. A., Opalk, K., & Karl, D. M. (2019). Improved ultraviolet photo-oxidation system yields estimates for deep-sea dissolved organic nitrogen and phosphorus. *Limnology and Oceanography, Methods / ASLO*, *17*(4), 277–291.
- Jayasimhulu, K., Hunt, S. M., Kaneshiro, E. S., Watanabe, Y., & Giner, J.-L. (2007). Detection and identification of *Bacteriovorax stolpii* UKi2 sphingophosphonolipid molecular species. *Journal of the American Society for Mass Spectrometry*, *18*(3), 394–403.
- Karl, D. M., & Bailiff, M. D. (1989). The measurement and distribution of dissolved nucleic acids in aquatic environments: Dissolved DNA and RNA. *Limnology and Oceanography*, *34*(3), 543–558.

- Karl, D. M., & Björkman, K. M. (2015). Chapter 5 - Dynamics of Dissolved Organic Phosphorus. In D. A. Hansell & C. A. Carlson (Eds.), *Biogeochemistry of Marine Dissolved Organic Matter (Second Edition)* (pp. 233–334). Boston: Academic Press.
- Karl, D. M., & Bossard, P. (1985). Measurement and significance of ATP and adenine nucleotide pool turnover in microbial cells and environmental samples. *Journal of Microbiological Methods*, 3(3), 125–139.
- Karl, D. M., & Tien, G. (1992). MAGIC: A sensitive and precise method for measuring dissolved phosphorus in aquatic environments. *Limnology and Oceanography*, 37(1), 105–116.
- Martin, P., & Van Mooy, B. A. S. (2013). Fluorometric quantification of polyphosphate in environmental plankton samples: extraction protocols, matrix effects, and nucleic acid interference. *Applied and Environmental Microbiology*, 79(1), 273–281.
- Martínez, A., Osburne, M. S., Sharma, A. K., DeLong, E. F., & Chisholm, S. W. (2012). Phosphite utilization by the marine picocyanobacterium *Prochlorococcus* MIT9301. *Environmental Microbiology*, 14(6), 1363–1377.
- Martinez, A., Tyson, G. W., & DeLong, E. F. (2010). Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. *Environmental Microbiology*, 12(1), 222–238.
- Metcalf, W. W., & Wolfe, R. S. (1998). Molecular genetic analysis of phosphite and hypophosphite oxidation by *Pseudomonas stutzeri* WM88. *Journal of Bacteriology*, 180(21), 5547–5558.
- Moore, L. R., Coe, A., Zinser, E. R., Saito, M. A., Sullivan, M. B., Lindell, D., ... Chisholm, S. W. (2007). Culturing the marine cyanobacterium *Prochlorococcus*: *Prochlorococcus* culturing. *Limnology and Oceanography, Methods / ASLO*, 5(10), 353–362.
- Murphy, J., & Riley, J. P. (1962). A modified single solution method for the determination of phosphate in natural waters. *Analytica Chimica Acta*, 27, 31–36.
- Orchard, E. D., Ammerman, J. W., Lomas, M. W., & Dyrman, S. T. (2010). Dissolved inorganic and organic phosphorus uptake in *Trichodesmium* and the microbial community: The

importance of phosphorus ester in the Sargasso Sea. *Limnology and Oceanography*, 55(3), 1390–1399.

Polyviou, D., Hitchcock, A., Baylay, A. J., Moore, C. M., & Bibby, T. S. (2015). Phosphite utilization by the globally important marine diazotroph *Trichodesmium*. *Environmental Microbiology Reports*, 7(6), 824–830.

Quin, L. D. (1964). 2-Aminoethylphosphonic Acid in Insoluble Protein of the Sea Anemone *Metridium dianthus*. *Science*, 144(3622), 1133–1134.

Shelburne, F. A., & Quin, L. D. (1967). Isolation of 2-(methylamino)ethylphosphonic acid from the proteinaceous residue of a sea anemone. *Biochimica et Biophysica Acta*, 148(2), 595–597.

Solórzano, L., & Strickland, J. D. H. (1968). POLYPHOSPHATE IN SEAWATER. *Limnology and Oceanography*, 13(3), 515–518.

Strickland, J. D. H., & Solorzano, L. (1966). Determination of monoesterase hydrolysable phosphate and phosphomonoesterase activity in sea water. *Some Contemporary Studies in Marine Science*, 665–674.

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., ... Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics: Official Journal of the Metabolomic Society*, 3(3), 211–221.

Thingstad, T. F., Krom, M. D., Mantoura, R. F. C., Flaten, G. A. F., Groom, S., Herut, B., ... Zohary, T. (2005). Nature of phosphorus limitation in the ultraoligotrophic eastern Mediterranean. *Science*, 309(5737), 1068–1071.

Van Mooy, B. A. S., Krupke, A., Dyhrman, S. T., Fredricks, H. F., Frischkorn, K. R., Ossolinski, J. E., ... Sylva, S. P. (2015). Phosphorus cycling. Major role of planktonic phosphate reduction in the marine phosphorus redox cycle. *Science*, 348(6236), 783–785.

Wingard, P., Burak, E. S., Gibiansky, E., & Vornov, J. J. (2011). *EPO Patent No. 2281565:A2*. Retrieved from <https://patentimages.storage.googleapis.com/f3/5f/f1/ad357072fac911/EP2281565A2.pdf>

Zhang, J.-Z., & Chi, J. (2002). Automated analysis of nanomolar concentrations of phosphate in natural waters with liquid waveguide. *Environmental Science & Technology*, 36(5), 1048–1053.

**APPENDIX 1. ETHYLENE CONCENTRATIONS AND
SATURATION VALUES IN THE EASTERN
MEDITERRANEAN SEA**

A.1.1 INTRODUCTION

Alike methane that is a by-product of the methylphosphonate (MPn) degradation by C-P lyase, ethylene (C_2H_4) is produced when marine microbes with C-P lyase use 2-hydroxyethylphosphonate (2-HEP) as a source of P (Repeta et al., 2016). Thus ethylene dissolved concentrations and saturation can be used to calculate ethylene air-sea fluxes, estimate 2-HEP degradation rates and further, phosphonate degradation rates by the C-P lyase. MPn and 2-HEP are present in approximately a 1:1 ratio in high molecular weight dissolved organic matter HMWDOM and in seawater incubations enriched with HMWDOM, methane and ethylene were produced in the same ratio (Repeta et al., 2016). An average 1:1 ratio of methane to ethylene saturation was also observed in the North Atlantic Subtropical Gyre (NPSG) (Sosa et al., 2020) and the gas saturations were correlated. We wanted to investigate if the same pattern was observed in the EMS which would indicate that similar processes are at play and that sources of MPn and 2-HEP in the EMS are similar to the ones in the North Atlantic. Thus, we measured ethylene concentrations, calculated ethylene saturation and air-sea fluxes and estimated 2-HEP as well as phosphonate degradation rates.

A.1.2 METHODS

A.1.2.1 Ethylene sampling and measurements

Dissolved ethylene concentrations were obtained from the same bottle than for dissolved methane concentrations and used the same instrumental setup (Chapter 2). All peaks corresponding to C_2H_4 were manually selected and integrated on the Agilent ChemStation software. The limits of detection and quantification (LODs and LOQs respectively) for ethylene were 13.4 pmol and 40.6 pmol. As before, LODs and LOQs were calculated based on the root mean square error (RMSE) of the gas standard calibration, where $LOD = 3.29 \text{ RMSE}$, and $LOQ = 10 \text{ RMSE}$ (Bernal & Guo, 2014).

A.1.2.2 Ethylene saturation and air-sea gas exchange

The equilibrium concentrations of ethylene in seawater were calculated based on the Bunsen solubility model of Breitbarth et al. (Breitbarth et al., 2004) using a mean atmospheric ethylene concentration of 0.4 ppb in the North Atlantic Ocean (Rudolph & Ehhalt, 1981). The extent to which the measured ethylene concentrations in seawater (C_{meas}) deviated from the predicted air-sea equilibrium concentrations (C_{eq}) was expressed as the molar differences ($C_{\text{meas}} - C_{\text{eq}}$) expressed as ΔC_2H_4 , respectively. This deviation was also expressed as the percent saturation

level ($C_{\text{meas}} \times 100/C_{\text{eq}}$), where 100% corresponds to a null ΔC_2H_4 . Air-sea flux estimates (F) of ethylene were calculated with the equation:

$$F=k(C_{\text{meas}} - C_{\text{eq}})$$

where k is the gas transfer coefficient calculated using the gas exchange wind speed relationship by Wanninkhof (2014) and $(C_{\text{meas}} - C_{\text{eq}})$ is the average ΔC_2H_4 concentration in the surface mixed layer. The k coefficient was calculated using a gas-specific Schmidt number. A Schmidt number specific to C_2H_4 was obtained using the diffusion coefficient equations implemented by Johnson (Johnson, 2010). Wind speed from the ship's anemometer was scaled to 10 m above the sea surface. The average daily wind speed was used to estimate k . The depth of the surface mixed layer depth (MLD) was calculated using a temperature threshold criterion of -0.2°C from a reference depth of 10 m (d'Ortenzio et al., 2005). Positive air-sea fluxes indicated that the ocean is a net source of methane or ethylene to the atmosphere.

A.1.3 RESULTS

Similarly to methane, ethylene was largely oversaturated at all the stations sampled in the EMS from surface waters to 2600 m (Figure A.1.1). Highest ethylene concentrations were always measured at depths ≤ 75 m; decreasing below this depth and often reaching a minimum at 500 m. In the surface, saturation values ranged from 1966% to 8513% (excluding the extreme value obtained at Dagon 2) (Table 2.2). We measured the highest ethylene saturation of all stations at the stations Dagon 1 and Tarhun 8 and the lowest at the station Dagon 4.

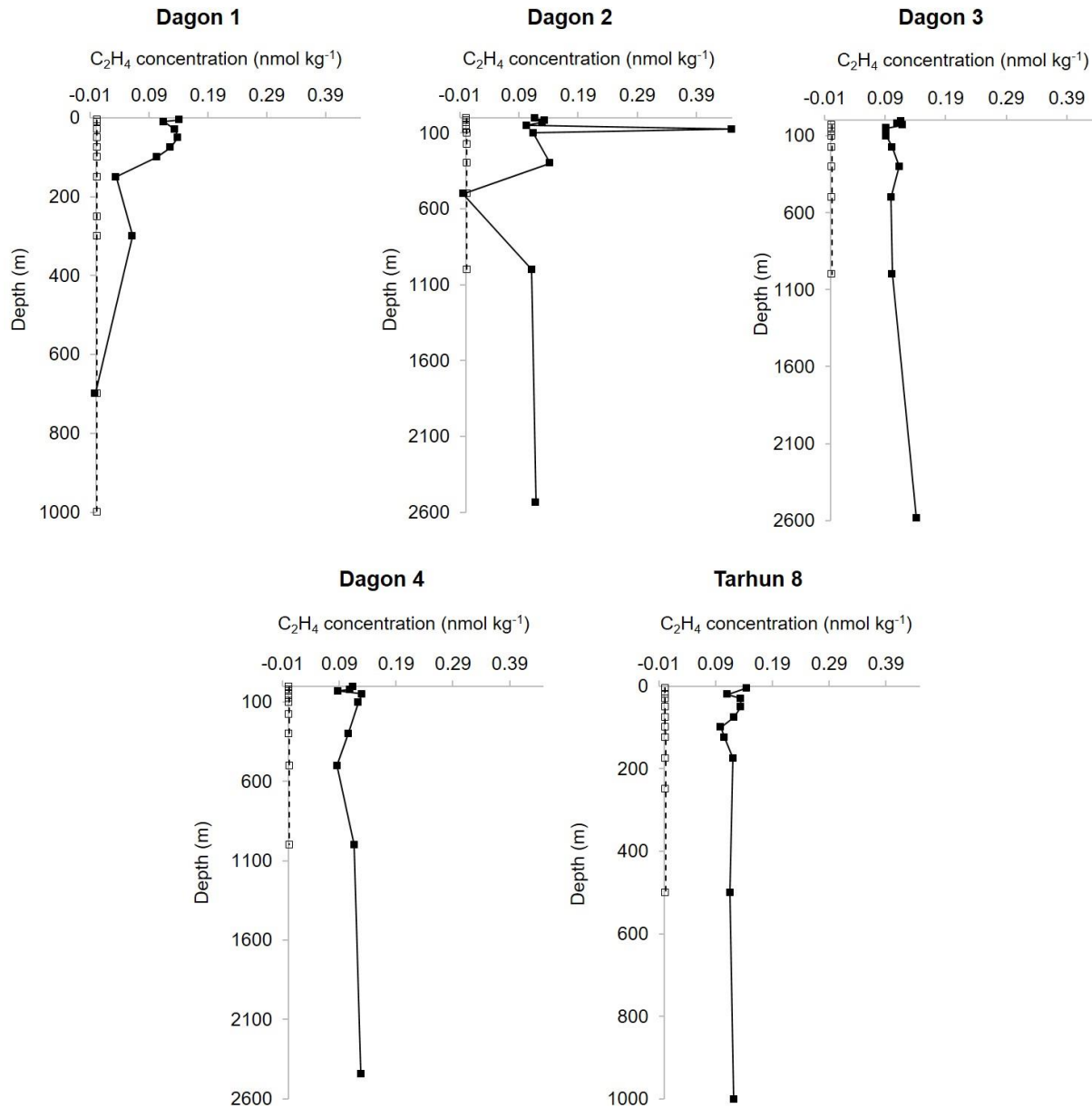


Figure A.1.118: Vertical profiles of dissolved C_2H_4 concentrations (black squares) and values expected at equilibrium with an atmospheric concentration of 400 ppb (empty squares).

Using the same MLD and the same wind speed than for methane (Table 2.1), we calculated ethylene air-sea fluxes. We calculated the highest ethylene flux to the atmosphere at Dagon 2 and the lowest flux at Dagon 3. Calculated values for ethylene production were greatest at Dagon 4, while Dagon 1 and 3 had the lowest values. The methane:ethylene ratio varied between 17.5 and 8.2, with a mean value of 13.6 (Table A.1.1).

Table A.1.1: Ethylene fluxes, production, excess and supersaturation values based on the MLD and MLD methane:ethylene ratio. Uncertainties correspond to the standard deviations between the values measured at each depth within the mixed layer. MLD is calculated as the depth where we measure a change in -0.2°C seawater temperature as preconized by d'Ortenzio et al. (2015).

Station	C_2H_4 flux ($\mu\text{mol m}^{-2} \text{d}^{-1}$)	C_2H_4 MLD production (pM d^{-1})	MLD excess C_2H_4 (nM)	MLD C_2H_4 saturation (%)	MLD excess $\text{CH}_4:\text{C}_2\text{H}_4$ ratio
Dagon 1	2.0 ± 0.8	7 ± 2	0.11 ± 0.04	6188 ± 2277	17.5
Dagon 2	3.5 ± 2.7	21 ± 16	0.2 ± 0.1	9798 ± 7590	8.2
Dagon 3	0.42 ± 0.08	7 ± 1	0.11 ± 0.02	6294 ± 1192	15.3
Dagon 4	1.00 ± 0.03	48 ± 1	0.11 ± 0.00	6632 ± 220	14.7
Tarhun 8	1.6 ± 0.2	27 ± 3	0.13 ± 0.02	7697 ± 853	12.2

A.1.4 DISCUSSION

$\Delta\text{C}_2\text{H}_4$ averaged values for the mixed layer varied between 0.173 nM at Dagon 2 and 0.107 nM at Dagon 3. Due to the extremely low accumulation of ethylene in the atmosphere, saturation values were very high at all stations. Air-sea fluxes of ethylene varied between $3.5 \mu\text{mol m}^{-2} \text{d}^{-1}$ at Dagon 2 due to the abnormally high concentration measured at 75 m and $0.42 \mu\text{mol m}^{-2} \text{d}^{-1}$ at Dagon 3. The corresponding fluxes in the EMS are about 2 times higher than measured in the NASG (Sosa et al., 2020) which again supports the hypothesis that C-P lyase is more active in the EMS than in the NASG. However, the MLD methane:ethylene ratios vary between 8 and 18 (Table A.1.1) which is higher than the average value of 1 observed in the NASG (Sosa et al., 2020) or for incubations of Station ALOHA seawater amended with high molecular weight DOM (Repeta et al., 2016). Moreover, there is no correlation between ethylene saturation and Pi concentrations which suggests that different processes affecting methane and/or ethylene saturations are at play in the EMS compared to the NASG and NPSG.

To measure phosphonate hydrolysis rates by C-P lyase, we measured dissolved methane and ethylene concentrations to calculate their air-sea fluxes values. This approach has been used as proxy to estimate the turnover rate of MPn and 2-HEP (Repeta et al., 2016; Sosa et al., 2020), but assumes that the degradation of those compounds is the primary source of methane and ethylene. In the EMS, the methane and ethylene vertical profiles from the same stations show very similar patterns with maximum and minimum surface concentrations obtained at similar depths (Figures 2.2 and 2.3). Moreover, at Dagon 3, 4 and Tarhun 8 MLD methane and ethylene % saturation as well as MLD methane:ethylene ratios were similar, which suggest that methane and ethylene are affected by the same processes across these stations (Tables 2.1 and A.1.1). However, the MLD methane:ethylene ratios are consistently about ten times higher than the value of 1 observed in the NASG and NPSG (Sosa et al., 2020) suggesting key differences between those oligotrophic regions and the EMS. We find 3 likely explanations:

- 1) Methane production from the degradation of MPn by C-P lyase is not the only source of methane in the EMS waters. Indeed, the EMS is known for the presence of mud

- volcanoes and deep cold seeps which are the major source of methane in deep waters (Charlou et al., 2003) and it is possible that methane from those sources is not fully oxidized, contributing to the large % saturation observed. However, methane profiles decrease to nM concentrations few hundred meters above the volcanoes (Charlou et al., 2003) suggesting rapid oxidation in situ. Additionally, benthic methane is isotopically light ($\delta^{13}\text{C} \sim -60-70 \text{ ‰}$) and values we measured in the EMS upper water column are characteristic of the NASG and NPSG (Holmes et al., 2000). However, we cannot exclude the possibility of methane sources with similar isotopic fractionation effect that might occur in oxic conditions such as methane production by acetate consumption (Bogard et al., 2014) or by cyanobacteria during C fixation (Bižić et al., 2020) that have an unknown isotopic value.
- 2) Phosphonate sources are different. HMWDOM from the NPSG and the NASG has a MPn:2-HEP ratio of ~ 1 which can explain the 1:1 excess ratio in methane and ethylene (Repeta et al., 2016; Sosa et al., 2020). If the HMWDOM MPn:2-HEP is higher in the EMS it could explain why we measure a higher methane % saturation compare to ethylene. Alternatively, HMWDOM might not be the main or the only phosphonate source for C-P lyase which is likely as we were able to detect LMW phosphonates in particulate matter from the North Atlantic (Chapitre 5).
 - 3) Ethylene and methane are being oxidized aerobically at different rates. Ethylene is produced by stressed plants to reduce the activity of methanotrophs (Zhou et al., 2018) and was shown to inhibit methane oxidation in a forest soil (Bu et al., 2019) suggesting that ethylene oxidation is a preferred source of carbon for hydrocarbon oxidizing bacteria. Thus, it is likely that ethylene is oxidized faster than methane allowing the methane to build up leading to a higher proportion of methane compare to ethylene which would increase the methane:ethylene ratio.

To calculate the phosphonate consumption rates using the ethylene air-sea flux data, we made the following assumptions: 1) ethylene produced in the surface mixed layer are supported entirely by the degradation of 2-HEP respectively through the C-P lyase, 2) 2-HEP are each 1/3 of the phosphonate HMWDOM pool (Repeta et al., 2016) and 3) HMWDOM is the main source of phosphonate for the C-P lyase pathway. With these assumptions, we calculated 2-HEP degradation rates are between 7 and 48 pM d^{-1} and total phosphonate consumption rates between 21 and 145 pM d^{-1} . Those estimates can be compared to the measured in situ phosphonate production rates (Table X) and we find that phosphonate production is at steady state with phosphonate consumption. This conclusion is different from the ones using methane as a proxy for phosphonate degradation by C-P lyase or the C-P lyase assay. As we discussed previously, this could be due to the factors influencing the methane:ethylene ratio and we are faced with two different possibilities. If there is another source of methane but ethylene only comes from the degradation of 2-HEP by C-P lyase then those rates are accurate and the phosphonate cycle is at steady state in the EMS. If ethylene is consumed or if there is less ethylene in the EMS, then

degradation rates using methane are appropriate and phosphonate are being consumed more than they are being made and the cycle is not at steady state.

A.1.5 CONCLUSION

Ethylene dissolved concentrations were largely above saturation in the EMS in March. The highest MLD ethylene production was observed in Dagon 4 and the lowest at Dagon 1 and 3. Unlike for methane, we did not find a correlation between ethylene concentrations and Pi nor between methane and ethylene concentrations. Using ethylene as a proxy for phosphonate degradation rates, we found that those are similar to phosphonate production rates measures simultaneously which leads to the conclusion that the phosphonate cycle is at equilibrium. When comparing to phosphonate degradation rates using methane as a proxy or the C-P lyase assay, we drew different conclusions (Chapter 2) and more work is needed to constrain our measurements and be able to compare them.

A.1.6 REFERENCES

- Bernal, E., & Guo, X. (2014). Limit of detection and limit of quantification determination in gas chromatography. *Advances in Gas Chromatography*, 3(1), 57–63.
- Bižić, M., Klintzsch, T., Ionescu, D., Hindiyeh, M. Y., Günthel, M., Muro-Pastor, A. M., ... Grossart, H.-P. (2020). Aquatic and terrestrial cyanobacteria produce methane. *Science Advances*, 6(3), eaax5343.
- Bogard, M. J., del Giorgio, P. A., Boutet, L., Chaves, M. C. G., Prairie, Y. T., Merante, A., & Derry, A. M. (2014). Oxic water column methanogenesis as a major component of aquatic CH₄ fluxes. *Nature Communications*, 5, 5350.
- Breitbarth, E., Mills, M. M., Friedrichs, G., & LaRoche, J. (2004). The Bunsen gas solubility coefficient of ethylene as a function of temperature and salinity and its importance for nitrogen fixation assays: Bunsen coefficient and N₂fixation. *Limnology and Oceanography, Methods / ASLO*, 2(8), 282–288.
- Bu, X., Krause, S. M. B., Gu, X., Tian, J., & Zhou, X. (2019). Ethylene rather than acetylene inhibits soil methane oxidation rates in a subtropical evergreen forest. *Soil Biology & Biochemistry*, 135, 10–12.
- Charlou, J. L., Donval, J. P., Zitter, T., Roy, N., Jean-Baptiste, P., Foucher, J. P., & Woodside, J. (2003). Evidence of methane venting and geochemistry of brines on mud volcanoes of the

eastern Mediterranean Sea. *Deep Sea Research Part I: Oceanographic Research Papers*, 50(8), 941–958.

d’Ortenzio, F., Iudicone, D., de Boyer Montegut, C., Testor, P., Antoine, D., Marullo, S., ... Madec, G. (2005). Seasonal variability of the mixed layer depth in the Mediterranean Sea as derived from in situ profiles. *Geophysical Research Letters*, 32(12). Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005GL022463>

Holmes, M. E., Sansone, F. J., Rust, T. M., & Popp, B. N. (2000). Methane production, consumption, and air-sea exchange in the open ocean: An Evaluation based on carbon isotopic ratios. *Global Biogeochemical Cycles*, 14(1), 1–10.

Johnson, M. T. (2010). A numerical scheme to calculate temperature and salinity dependent air-water transfer velocities for any gas. *Ocean Science Discussions*, 7(1), 251–290.

Repeta, D. J., Ferrón, S., Sosa, O. A., Johnson, C. G., Repeta, L. D., Acker, M., ... Karl, D. M. (2016). Marine methane paradox explained by bacterial degradation of dissolved organic matter. *Nature Geoscience*, 9(12), 884–887.

Rudolph, J., & Ehhalt, D. H. (1981). Measurements of C₂–C₅hydrocarbons over the north Atlantic. *Journal of Geophysical Research*, 86(C12), 11959.

Sosa, O. A., Burrell, T. J., Wilson, S. T., Foreman, R. K., Karl, D. M., & Repeta, D. J. (2020). Phosphonate cycling supports methane and ethylene supersaturation in the phosphate-depleted western North Atlantic Ocean. *Limnology and Oceanography*. Retrieved from <https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.1002/lno.11463>

Zhou, X., Xu, C.-Y., Bai, S. H., Xu, Z., Smaill, S. J., Clinton, P. W., & Chen, C. (2018). Technical note: Manipulating interactions between plant stress responses and soil methane oxidation rates. *Biogeosciences*, 15(13), 4125–4129.

APPENDIX 2. ABSENCE OF PHOSPHONATE IN
***Trichodesmium erythraeum* IMS101 AND**
***Prorocentrum minimum* (CCMP1329)**

A2.1 INTRODUCTION

Phosphonates are organophosphorus esters where the carbon (C) is directly link to the phosphorus (P) atom. This causes the P to be in a reduced form compare to a more common phosphate ester bond (C-O-P) and its oxidation state is +III. Phosphonates are abundant in marine HMWDOM where they represent 20-25% of the P and are an alternative source of P for some marine microbes. However, so far, only three model marine microbes have been shown to produce phosphonates in cultures which limits our understanding of phosphonate biosynthesis and biochemistry. They are: the cyanobacteria *Trichodesmium* IMS101(Dyhrman et al., 2009) and *Prochlorococcus* SB (Chapter 3) and the archaeon *Nitrosopumilus maritimus* (Metcalf et al., 2012). Finding new model microbes able to produce phosphonates is especially important since among all the current identified phosphonate producers, only *Nitrosopumilus maritimus* has been shown to be able to produce methylphosphonate which constitute 1/3 of the phosphonate pool in HMWDOM and which degradation through the C-P lyase yields methane (Repeta et al., 2016)

The goal of this study was to find other organisms producing phosphonates to compare them with *Prochlorococcus* SB in term of P allocated to phosphonate production and their macromolecular form. Moreover, we were also trying to see which types of phosphonates were produced and if they were different from the ones in *Prochlorococcus* SB.

All known phosphonate biosynthetic pathways start with the isomerization of phosphoenolpyruvate (PEP) to phosphonopyruvate (PnPy) by phosphonopyruvate mutase (PepM) (Seidel et al., 1988). This step is followed by the decarboxylation of PnPy to phosphonoacetaldehyde (PnAa) by phosphonopyruvate decarboxylase (Ppd) (Zhang et al., 2003). The gene sequence coding for PepM has been used in multiple studies to investigate phosphonate production potential in marine microbial genomes and metagenomes (Metcalf et al., 2012; Villarreal-Chiu et al., 2012; Yu et al., 2013). More recently, we found that 15% of bacterial+archeal cells were putative phosphonate producers (Chapter 3). Genomic analyses support the potential for phosphonate synthesis, but they are not a definite proof that those organisms are indeed producing phosphonate. Thus analysis of laboratory cultures is needed to confirm phosphonate production and identify which phosphonates and how much phosphonates is being produced.

Because *Trichodesmium* IMS101 is a known phosphonate producer and readily available, we cultured *Trichodesmium* IMS101 to identify which phosphonates were being produced and characterize any macromolecules they might associated with. Metagenomic analysis have also suggested that some marine dinoflagellates cultures may also synthesize phosphonates (Cui et al., 2016). Thus we selected dinoflagellate strains that had genes coding for PepM and Ppd, were available for purchase from the Bigelow culture collection and were axenic. This narrowed down

the list to 3 strains: *Alexandrium tamarense* (CCMP1771), *Amphidinium carterae* (CCMP1314) and *Prorocentrum minimum* (CCMP1329) which is the one we retained for this experiment.

A2.2 MATERIAL AND METHODS

A2.2.1 *Trichodesmium erythraeum* sp. IMS101

Trichodesmium erythraeum sp. IMS101 was grown in RMP growth media prepared with filtered and sterile oligotrophic Sargasso seawater (Webb et al., 2001). The culture was not axenic but had been partially purified by serial sterile transfers for hundreds of generations into RMP media which could have caused the emergence of a strain different from IMS101. The culture was maintained at 26.9°C in a 14:10h day:night light cycle. Approximately 50 mL of dense, healthy filaments were inoculated into 300 mL RMP medium for a final volume of 350 mL. The flask was gently stirred daily. The culture was allowed to acclimate and grow for 5 days prior to sampling and was harvested during day light exposure in exponential growth. Cultures were filtered through a 0.22 µm Durapore filter under gentle vacuum to recover biomass. The filter was immediately frozen at -20°C until analysis.

To extract phosphonates from *Trichodesmium* IMS 101 biomass, filters were cut into pieces and hydrolyzed by sonicating for 1 hr in 1.5 mL of 2N KOH followed by heating at 80°C for 24h. The reaction was quenched by stirring with 1 g of Biorex 50W-X8 cation exchange resin (50-100 mesh, H⁺ form, previously washed 4x with 10 mL of MQ) for ~ 1hr. The sample was centrifuged, the supernate removed and the resin washed 2x with 0.5 mL MQ. The supernate and washes were combined, dried, and the residue dissolved in 800 µL D₂O for NMR analysis.

A2.2.2 *Prorocentrum minimum*

Prorocentrum minimum (CCMP 1329) was grown in a f/2-Si medium (Guillard & Ryther, 1962) prepared in filtered (0.22 µm) Vineyard Sound seawater using sterile protocols. The medium was modified by adding Na₂SeO₃ and reducing the concentration of CuSO₄·5H₂O to a final concentration of 10⁻⁸ M each (Anderson et al., 1994). Cultures were grown at 19°C on a 14:10 hr light:dark cycle (ca. 200 µmol photons·m⁻²·sec⁻¹ irradiance provided by cool white fluorescent bulbs). Growth was monitored by visual cell counts and the culture was harvested in exponential growth. Cells were pelleted by centrifuge and the spent medium decanted. The cells were washed 2x by resuspending them in 20 mL Turk Island Mix or non-amended artificial seawater (Moore et al., 2007) and centrifuged. The cell pellet was then frozen at -20°C until further analysis.

A2.2.3 Nuclear magnetic resonance (NMR)

To assess whether or not the cultures were producing phosphonates, we used ³¹P NMR. NMR spectra were acquired at 25°C on a 400 MHz Bruker AVANCE DPX spectrometer using a 5mm

inverse broadband probe and running TOPSPIN 1.3. Phosphorus shifts are reported relative to external 85% phosphoric acid at 0 ppm. For the proton-decoupled ^{31}P NMR spectrum of whole *Prorocentrum minimum* cells, the frozen cell pellet was thawed and transferred to a 5mm BMS tube (Shigemi Inc.) with a magnetic susceptibility matching D_2O . Samples were analyzed using “zgdc30” with WALTZ16 decoupling, a sweep width of 80 ppm, a 3 seconds relaxation delay, 100K scans and 20Hz line broadening. For the ^{31}P NMR spectrum of hydrolyzed *Trichodesmium erythraeum* IMS101 we used the program “zgdc30” with a sweep width of 80 ppm, a 3 seconds relaxation delay for 30K scans and 15Hz line broadening.

A2.3 RESULTS AND DISCUSSION

A2.3.1 Absence of phosphonate production in *Trichodesmium erythraeum* IMS101

The ^{31}P NMR spectrum of soluble P after hydrolysis of *Trichodesmium erythraeum* IMS101 (Figure A.2.1) displays a sharp peak at 0 ppm from phosphate, and a suite of peaks between 1-3 ppm characteristic of phosphate esters (Quin & Williams, 2004).

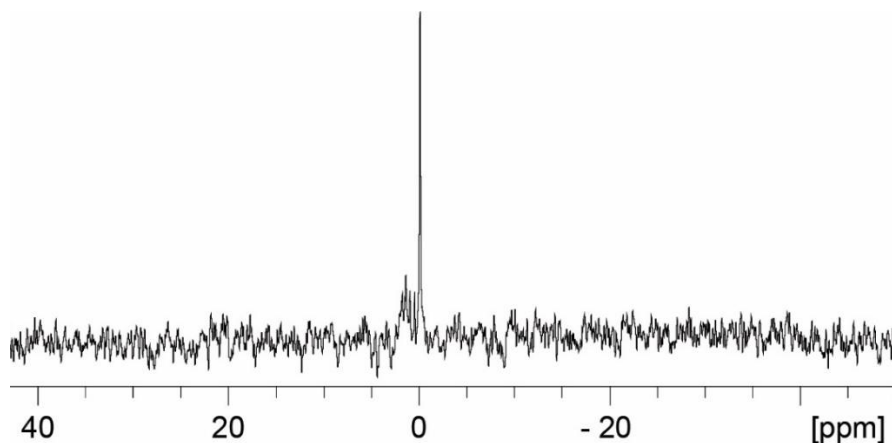


Figure A.2.1: ^{31}P NMR spectrum of soluble phosphorus after 2N KOH hydrolysis of *Trichodesmium erythraeum* IMS101 cells

Hydrolysis does not affect the C-P bound in phosphonate (Horsman & Zechel, 2017) and if phosphonates were produced by *Trichodesmium*, peaks should appear in phosphonate region (15 to 30 ppm) (Quin & Williams, 2004) of the ^{31}P NMR spectrum. Unexpectedly, there were no discernable peaks in this region. Dyhrman et al. (2009) used solid state NMR of whole cells to show that *Trichodesmium erythraeum* IMS101 dedicated ~10% of its cellular P towards phosphonate production. We were not able to confirm that result in this study.

Several hypotheses might explain this difference. First, if *Trichodesmium* synthesizes a complex mixture of phosphonates with individual phosphonates representing < 3-5% of the total

phosphorus in the sample, it is unlikely that we would have been able to distinguish these phosphonates from noise in the spectrum and phosphonate from this strain might have gone undetected. Second, phosphonates in the Dyrman sample may also have originated from epibionts (Hmelo et al., 2012) and not from *Trichodesmium* as suggested by a study where PepM gene sequence was found in the *Trichodesmium* holobiont and not only in the *Trichodesmium* genome (Frischkorn et al., 2017). Third, the microbial community associated with *Trichodesmium erythraeum* IMS101 may have changed since 2009, and finally, *Trichodesmium erythraeum* itself might have evolved over the decade separating the two studies. Even though the culture has been described as uni-algal, the culture might have had a minority of other *Trichodesmium* strains that could have outcompeted the initial culture. It is also likely that if the epibionts have changed, the culture itself evolved in parallel since signaling, cross-feeding and other symbiotic processes might have been altered. It should also be noted that we did not investigate the presence of phosphonate in the spent growth medium (unlike for *Prochlorococcus* SB). However, previous studies reported phosphonates associated with *Trichodesmium* cells (Dyrman et al., 2009; Van Mooy et al., 2015) and it is therefore unlikely that our culture excreted phosphonates into the spent medium.

A2.3.2 Absence of phosphonate production in *Prorocentrum minimum*

As for *Trichodesmium erythraeum* IMS101, the ^{31}P NMR spectrum *Prorocentrum minimum* (Figure A.2.2) only displays one large peak due to the presence of phosphate esters and diesters between 10 and -5 ppm (Quin & Williams, 2004). The absence of discernable peak in the phosphonate region of the spectrum (15 to 30 ppm) suggests that despite having the full phosphonate biosynthetic pathway, *Prorocentrum minimum* did not produce phosphonate in this culture.

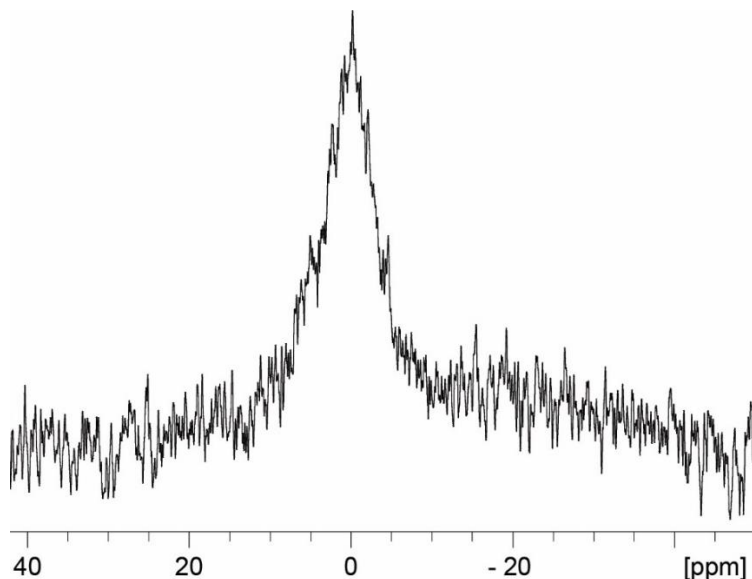


Figure A.2.219: ^{31}P NMR spectrum of *Prorocentrum minimum* whole cells

Dinoflagellates have a very large genome, a highly complex phenotype (Spector, 2012), transcriptionally inactive DNA (Roy & Morse, 2013) a complicated and poorly understood gene expression which involves transcriptional and post-transcriptional regulation (Hackett et al., 2004). Moreover, *Prorocentrum minimum* has been shown to perform substitutional mRNA editing (Lin et al., 2002) which further complicates the link between genomic interpretation and observed phenotype. Because of this, it is likely that even though *Prorocentrum minimum* genome encodes the full phosphonate biosynthetic pathway, active phosphonate production is dormant and only occurs under specific conditions. Phosphonates can be used by organisms as potent metabolites as they can mimic phosphates and carboxylic acids hereby inhibiting the activity of numerous enzymes (Horsman & Zechel, 2017). Therefore, we hypothesize that *Prorocentrum minimum* and probably other dinoflagellates with the full phosphonate biosynthetic pathway (Cui et al., 2016) are more likely to transcribe the genes involved in phosphonate production and actively produce phosphonates during harmful algal blooms, events where various toxins are produced. Thus, it is likely that the axenic culture conditions did not provoke phosphonate production in our culture leading to the absence of phosphonates associated with the cells. As for *Trichodesmium erythraeum* IMS101, we did not investigate if phosphonates were present in the spent medium.

A2.4 REFERENCES

- Anderson, D. M., Kulis, D. M., Doucette, G. J., Gallagher, J. C., & Balech, E. (1994). Biogeography of toxic dinoflagellates in the genus *Alexandrium* from the northeastern United States and Canada. *Marine Biology*, *120*(3), 467–478.
- Cui, Y., Lin, X., Zhang, H., Lin, L., & Lin, S. (2016). PhnW-PhnX Pathway in Dinoflagellates Not Functional to Utilize Extracellular Phosphonates. *Frontiers in Marine Science*, *2*, 120.
- Dyhrman, S. T., Benitez-Nelson, C. R., Orchard, E. D., Haley, S. T., & Pellechia, P. J. (2009). A microbial source of phosphonates in oligotrophic marine systems. *Nature Geoscience*, *2*(10), 696–699.
- Frischkorn, K. R., Rouco, M., Van Mooy, B. A. S., & Dyhrman, S. T. (2017). Epibionts dominate metabolic functional potential of *Trichodesmium* colonies from the oligotrophic ocean. *The ISME Journal*, *11*(9), 2090–2101.
- Guillard, R. R., & Ryther, J. H. (1962). Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (Cleve) Gran. *Canadian Journal of Microbiology*, *8*, 229–239.
- Hackett, J. D., Anderson, D. M., Erdner, D. L., & Bhattacharya, D. (2004). Dinoflagellates: a remarkable evolutionary experiment. *American Journal of Botany*, *91*(10), 1523–1534.
- Hmelo, L. R., Van Mooy, B. A. S., & Mincer, T. J. (2012). Characterization of bacterial epibionts on the cyanobacterium *Trichodesmium*. *Aquatic Microbial Ecology: International Journal*, *67*(1), 1–14.
- Horsman, G. P., & Zechel, D. L. (2017). Phosphonate Biochemistry. *Chemical Reviews*, *117*(8), 5704–5783.
- Lin, S., Zhang, H., Spencer, D. F., Norman, J. E., & Gray, M. W. (2002). Widespread and extensive editing of mitochondrial mRNAs in dinoflagellates. *Journal of Molecular Biology*, *320*(4), 727–739.
- Metcalf, W. W., Griffin, B. M., Cicchillo, R. M., Gao, J., Janga, S. C., Cooke, H. A., ... van der Donk, W. A. (2012). Synthesis of methylphosphonic acid by marine microbes: a source for methane in the aerobic ocean. *Science*, *337*(6098), 1104–1107.
- Moore, L. R., Coe, A., Zinser, E. R., Saito, M. A., Sullivan, M. B., Lindell, D., ... Chisholm, S. W. (2007). Culturing the marine cyanobacterium *Prochlorococcus*: *Prochlorococcus* culturing. *Limnology and Oceanography, Methods / ASLO*, *5*(10), 353–362.
- Quin, L. D., & Williams, A. J. (2004). *Practical interpretation of P-31 Nmr Spectra and Computer Assisted Structure Verification* (p. 132).

- Repeta, D. J., Ferrón, S., Sosa, O. A., Johnson, C. G., Repeta, L. D., Acker, M., ... Karl, D. M. (2016). Marine methane paradox explained by bacterial degradation of dissolved organic matter. *Nature Geoscience*, 9(12), 884–887.
- Roy, S., & Morse, D. (2013). Transcription and Maturation of mRNA in Dinoflagellates. *Microorganisms*, 1(1), 71–99.
- Seidel, H. M., Freeman, S., Seto, H., & Knowles, J. R. (1988). Phosphonate biosynthesis: isolation of the enzyme responsible for the formation of a carbon-phosphorus bond. *Nature*, 335(6189), 457–458.
- Spector, D. L. (2012). *Dinoflagellates*. Academic Press.
- Van Mooy, B. A. S., Krupke, A., Dyhrman, S. T., Fredricks, H. F., Frischkorn, K. R., Ossolinski, J. E., ... Sylva, S. P. (2015). Phosphorus cycling. Major role of planktonic phosphate reduction in the marine phosphorus redox cycle. *Science*, 348(6236), 783–785.
- Villarreal-Chiu, J. F., Quinn, J. P., & McGrath, J. W. (2012). The Genes and Enzymes of Phosphonate Metabolism by Bacteria, and Their Distribution in the Marine Environment. *Frontiers in Microbiology*, 3. doi:10.3389/fmicb.2012.00019
- Webb, E. A., Moffett, J. W., & Waterbury, J. B. (2001). Iron stress in open-ocean cyanobacteria (*Synechococcus*, *Trichodesmium*, and *Crocospaera* spp.): identification of the IdiA protein. *Applied and Environmental Microbiology*, 67(12), 5444–5452.
- Yu, X., Doroghazi, J. R., Janga, S. C., Zhang, J. K., Circello, B., Griffin, B. M., ... Metcalf, W. W. (2013). Diversity and abundance of phosphonate biosynthetic genes in nature. *Proceedings of the National Academy of Sciences*, 110(51), 20759–20764.
- Zhang, G., Dai, J., Lu, Z., & Dunaway-Mariano, D. (2003). The phosphonopyruvate decarboxylase from *Bacteroides fragilis*. *The Journal of Biological Chemistry*, 278(42), 41302–41308.

**APPENDIX 3. EXTRACTION EFFICIENCY OF LOW
MOLECULAR WEIGHT DISSOLVED ORGANIC
PHOSPHORUS (LMWDOP) BY VARIOUS SOLID PHASE
EXTRACTION (SPE) COLUMNS**

A3.1 SUMMARY

As low molecular weight organic phosphorus (LMWDOP) is very dilute in the ocean, a pre-concentration method would considerably aid in identifying and quantifying phosphorus containing metabolites in the ocean. Here we compare four different SPE columns for their efficiency at retaining phosphorus-containing compounds in seawater.

A3.2 INTRODUCTION

Marine dissolved organic matter (DOM) is typically divided into high and low molecular weight DOM (HMWDOM and LMWDOM respectively) (Hansell, D. A., & Carlson, C. A., 2002/2014). HMWDOM is typically isolated using ultra-filtration membranes with a molecular weight cut off of ~1 kDa whereas to concentrate the small and dilute molecules constituting LMWDOM, solid phase extraction (SPE) is used (Kido Soule et al., 2015; Koch et al., 2008; Lechtenfeld et al., 2011; Petras et al., 2017). SPE relies on the greater affinity of an analyte to adsorb onto a solid phase than to remain in solution (Hennion, 1999). However, common SPE extraction methods employed in metabolomics fail to capture most LMWDOP (Johnson et al., 2017; Lechtenfeld et al., 2011) most likely due to the small size, high polarity and hydrophilic character of P containing metabolites. To identify the SPE technique that is able to retain the most LMWDOP, seawater was passed through four different SPE columns and the amount of DOP removed from seawater was measured. For this quantitative approach, we used a colorimetric method i.e. the molybdenum blue method (Murphy & Riley, 1962), and a recently developed DOP oxidation method (Foreman et al., 2019).

There are various ways to oxidize DOP to Pi and be able to measure TDP to obtain DOP concentrations by difference. Oxidation can be done using wet chemical oxidation with persulfate (Menzel & Corwin, 1965) or magnesium nitrate (Cembella et al., 1986), high-temperature combustion (HTC; ref) or ultraviolet (UV) light (Armstrong & Tibbitts, 1968). UV oxidation is one of the widely use method and has been recently optimized by Foreman et al. (Foreman et al., 2019). Their method, using a new photo-oxidation system followed directly by a colorimetric measure of TDP with an auto-analyzer (described in more details below), proved to be efficient towards phosphate esters and phosphonates. Therefore, we used this method to estimate SRP, TDP and DOP concentrations in North Atlantic seawater before and after the extraction of the four selected columns.

A3.3 MATERIAL AND METHODS

A3.3.1 SPE columns

We selected 4 different columns: solute® ENV+ (Supelco), Supelclean™ Envi-Carb™ (Sigma-Aldrich), titanium oxide (TiO₂; Sigma-Aldrich; column packed in-house), and Hybrid-SPE® (Sigma-Aldrich). The ENV+ resin is a polystyrene-divinyl benzene co-polymer designed to retain hydrophobic organic compounds from aqueous media. Envi-Carb is a non-porous activated charcoal substrate with high surface area and a high affinity for both polar and nonpolar organic compounds. TiO₂ has been used extensively to retain phosphorus containing biopolymers such as phosphopeptides in phosphoproteomic from aqueous media (Thingholm et al., 2006) and Hybrid-SPE® is a zirconia-coated substrate commonly used to concentrate phospholipids (Ahmad et al., 2012). Columns (1 g each ENV+, Envi-Carb and TiO₂ (rutile form; <5 µm, >99.9% trace metal basis; Sigma Aldrich), 2 x 0.5g Hybrid-SPE) were activated before running the sample. The ENV+ column was activated using 12 mL of MQ water followed by 12 mL of methanol. The Envi-Carb column was activated with 12 mL pH 2 MQ water. The TiO₂ and Hybrid-SPE columns were activated with 12 mL of MQ water.

A3.3.2 Extraction of DOP from seawater

Twenty liters of 0.22 µm filtered Sargasso Sea surface seawater was pumped through each column at 15-20 mL min⁻¹. Samples for DOP were collected before and after passage through the column. For the Envi-Carb column, the seawater was acidified to pH 3 using HCl before passing through the column. To measure potential contamination introduced during sampling, DOP was measured in 10 L sized samples of MQ water before and after passage through the system. All tubing and bottles were previously acid washed. Water samples DOP analysis and SPE columns were frozen (-20°C freezer) immediately after collection.

A3.3.3 UV oxidation

To oxidize samples for TDP measurements, we used the protocol developed by Foreman et al. (2019). Briefly, the UV light system is a VelaCure12 equipped with a microwaved powered 1800 W mercury lamp (F300S lamp, Heraeus Noblelight) isolated by a quartz window from the samples. Custom quartz glass vials (Suprasil 310; Allen Scientific Glass) of 22 to 25 mL previously washed with 10% HCl were used and filled with 20 mL of sample and 50 µL of 30% hydrogen peroxide (ACS grade, Fisher) and 9 µL of 5 M sulfuric acid (ACS Plus grade, Fisher). Seawater before and after each column were split in two samples to have duplicate measurements and were oxidized for 1.5 h.

A3.3.4 SRP and TDP measurements

Once oxidized in the UV system samples were let to cool then each was transferred into two clean 4 mL sample cups to be analyzed for TDP on the auto-analyzer which gives a duplicate for each oxidized sample for a total of four TDP measurements for one particular sample. For the analyses, we used an auto-analyzer SEAL Analytical AutoAnalyzer III equipped with a high-resolution detector set on 880 nm as preconized by Murphy and Riley (Murphy & Riley, 1962) and prepared the reagents according to Foreman et al. (2019). Standards for the instrument calibration were prepared using a mother solution of 10 mM sodium phosphate used to prepare an intermediary solution of 100 μM . This solution was added to low nutrient seawater (LNSW), to correct for any matrix effect, to obtain 6 standards ranging from 1.004 μM to 0 μM added. Similarly, these samples were oxidized and TDP and SRP measured in duplicates. A reference standard containing 1.0 μM phosphate was run which, together with the concentration calculated for the 0 μM added standard, allowed us to calculate the phosphate concentration in the LNSW. Moreover, as the 0 μM added standard contain some phosphate, we also measure the baseline by injecting MQ water. This baseline is corrected for seawater using the refractive index (-266 for a gain of 400) and those are used to calculate the TDP and SRP concentrations of the various samples. Every 22 sample cups, the baseline is measured to make sure that no drift occurred. This method typically has a limit of quantification (LOQ) of ~ 30 nM (Foreman et al., 2019) which gives a limit of detection (LOD) of ~ 9 nM.

A3.4 RESULTS AND DISCUSSION

A3.4.1 North Atlantic seawater initial concentrations

SRP and TDP concentrations of the North Atlantic seawater sample were 405.8 ± 0.8 nM and 539 ± 4 nM respectively (Table A.3.1). DOP was calculated as 133 ± 4 nM. The SRP concentration was much higher (by about two orders of magnitude) than typical Sargasso Sea surface water, which is often below 10 nM (Cavender-Bares et al., 2001; Lomas et al., 2010). This high SRP concentration is probably due to contamination from the ship's seawater intake or the sampling barrel. However, DOP concentrations are in the range of what would be expected from Sargasso seawater (Cavender-Bares et al., 2001; Lomas et al., 2010; Mather et al., 2008).

A3.4.2 MilliQ blanks

To measure phosphorus contamination from the SPE columns, SRP and TDP of MQ water was measured before and after it went through each column (Table A.3.1). The ENV+ column did not appear to contaminate with either SRP or DOP. The Hybrid SPE column added some SRP, but no DOP was added or removed. The Envi-Carb column added ~ 10 nM SRP and may have removed some DOP. Finally, the TiO_2 column added a significant amount (~ 100 nM) of SRP

and TDP concentration was less than SRP, and no determination could be made about addition or retention of DOP.

Table A.3.1: Summary of SRP, TDP and calculated DOP concentrations of Sargasso seawater and blanks before and after the 4 different SPE columns

	MQ water			Sargasso SW		
	SRP (nM)	TDP (nM)	DOP (nM)	SRP (nM)	TDP (nM)	DOP (nM)
Before columns	1.3 ± 0.4	24 ± 2	23 ± 2	405.8 ± 0.8	539 ± 4	133 ± 4
After columns						
ENV+	1 ± 2	26 ± 2	25 ± 3	423 ± 2	538 ± 1	114 ± 3
Envi-Carb	11 ± 2	27 ± 6	16 ± 6	414.8 ± 0.4	466 ± 2	51 ± 2
TiO ₂	105 ± 3	78 ± 5	-27 ± 6	69 ± 2	186 ± 2	116 ± 3
Hybrid-SPE	15 ± 2	35 ± 2	20 ± 3	401 ± 1	503 ± 3	102 ± 3

A3.4.3 Sargasso seawater

The differences between the phosphorus concentrations of the initial seawater and the seawater after the columns were used to measure how much phosphorus had been retained on the column. When looking at SRP, the TiO₂ column retained 83% whereas other columns didn't retain any and all contaminated the sample with SRP. For DOP, the Envi-Carb was the most efficient and retained 60% of DOP. The Hybrid-SPE had an intermediate efficiency with 24% of DOP retained whereas the TiO₂ and ENV+ columns only retained ~ 10-15% (Table A.3.2). It should be noted, that the protocol using TiO₂ columns for phosphopeptides enrichment recommends a sample dilution in a solution of dihydrobenzoic acid or a mixture of phthalic acid in acetonitrile and trifluoroacetic acid which we did not do here and might have affected the efficiency of the column. Concerning the ENV+, this is not highly surprising as very few P peaks were observed by liquid chromatography coupled with ionization coupled plasma and mass spectrometry (HPLC-ICPMS) for seawater samples previously collected with this column. Therefore, this column is probably not the most interesting one to study untargeted organophosphorus and the most promising column is the Envi-Carb.

Table A.3.110: Summary of the amount of DOP retained by the 4 different SPE columns

Column	ΔSRP (nM)	ΔTDP (nM)	ΔDOP (nM)	DOP retained (%)
ENV+	-17 ± 2	2 ± 4	19 ± 4	14 ± 3
Envi-Carb	-9.0 ± 0.9	73 ± 4	82 ± 4	61 ± 4
TiO ₂	337 ± 3	353 ± 4	17 ± 5	12. ± 4
Hybrid-SPE	5 ± 2	36 ± 5	31 ± 5	24 ± 4

A3.5 CONCLUSION

Being able to pre-concentrate and purify LMWDOP compounds is key in order to identify and quantify P containing metabolites in order to study their biogeochemical cycle as well as their cellular function. Our results show that all column were able to retain some DOP from seawater with various efficiency. The Envi-Carb column performed the best followed by the Hybrid-SPE. The ENV+ and TiO₂ had low but comparable efficiency towards DOP. Our data also shows that the columns need to be cleaned more thoroughly to eliminate SRP contamination. The DOP retained by those columns has not been characterized and should be investigated. Using those columns in line might also present some advantages as they most likely have different affinity for different organophosphorus compounds, enabling the possibility of identifying a relatively important part of the DOP pool.

A3.6 REFERENCES

- Ahmad, S., Kalra, H., Gupta, A., Raut, B., Hussain, A., & Rahman, M. A. (2012). HybridSPE: A novel technique to reduce phospholipid-based matrix effect in LC-ESI-MS Bioanalysis. *Journal of Pharmacy & Bioallied Sciences*, 4(4), 267–275.
- Armstrong, F. A. J., & Tibbitts, S. (1968). Photochemical combustion of organic matter in sea water, for nitrogen, phosphorus and carbon determination. *Journal of the Marine Biological Association of the United Kingdom. Marine Biological Association of the United Kingdom*, 48(1), 143–152.
- Cavender-Bares, K. K., Karl, D. M., & Chisholm, S. W. (2001). Nutrient gradients in the western North Atlantic Ocean: Relationship to microbial community structure and comparison to patterns in the Pacific Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, 48(11), 2373–2395.
- Cembella, A. D., Antia, N. J., & Taylor, F. J. R. (1986). The determination of total phosphorus in seawater by nitrate oxidation of the organic component. *Water Research*, 20(9), 1197–1199.
- Foreman, R. K., Björkman, K. M., Carlson, C. A., Opalk, K., & Karl, D. M. (2019). Improved ultraviolet photo-oxidation system yields estimates for deep-sea dissolved organic nitrogen and phosphorus. *Limnology and Oceanography, Methods / ASLO*, 17(4), 277–291.

- Hansell, D. A., & Carlson, C. A. (Ed.). (2014). *Biogeochemistry of Marine Dissolved Organic Matter, 2nd edition*. Academic Press. (Original work published 2002)
- Hennion, M. C. (1999). Solid-phase extraction: method development, sorbents, and coupling with liquid chromatography. *Journal of Chromatography, A*, 856(1–2), 3–54.
- Johnson, W. M., Kido Soule, M. C., & Kujawinski, E. B. (2017). Extraction efficiency and quantification of dissolved metabolites in targeted marine metabolomics: Matrix effects in marine metabolomics. *Limnology and Oceanography, Methods / ASLO*, 15(4), 417–428.
- Kido Soule, M. C., Longnecker, K., Johnson, W. M., & Kujawinski, E. B. (2015). Environmental metabolomics: Analytical strategies. *Marine Chemistry*, 177, 374–387.
- Koch, B. P., Ludwichowski, K.-U., Kattner, G., Dittmar, T., & Witt, M. (2008). Advanced characterization of marine dissolved organic matter by combining reversed-phase liquid chromatography and FT-ICR-MS. *Marine Chemistry*, 111(3), 233–241.
- Lechtenfeld, O. J., Koch, B. P., Geibert, W., Ludwichowski, K.-U., & Kattner, G. (2011). Inorganics in organics: quantification of organic phosphorus and sulfur and trace element speciation in natural organic matter using HPLC-ICPMS. *Analytical Chemistry*, 83(23), 8968–8974.
- Lomas, M. W., Burke, A. L., Lomas, D. A., Bell, D. W., Shen, C., Dyhrman, S. T., & Ammerman, J. W. (2010). Sargasso Sea phosphorus biogeochemistry: an important role for dissolved organic phosphorus (DOP). *Biogeosciences*, 7, 695–710.
- Mather, R. L., Reynolds, S. E., Wolff, G. A., Williams, R. G., Torres-Valdes, S., Woodward, E. M. S., ... Achterberg, E. P. (2008). Phosphorus cycling in the North and South Atlantic Ocean subtropical gyres. *Nature Geoscience*, 1(7), 439–443.
- Menzel, D. W., & Corwin, N. (1965). THE MEASUREMENT OF TOTAL PHOSPHORUS IN SEAWATER BASED ON THE LIBERATION OF ORGANICALLY BOUND FRACTIONS BY PERSULFATE OXIDATION¹. *Limnology and Oceanography*, 10(2), 280–282.
- Murphy, J., & Riley, J. P. (1962). A modified single solution method for the determination of phosphate in natural waters. *Analytica Chimica Acta*, 27, 31–36.
- Petras, D., Koester, I., Da Silva, R., Stephens, B. M., Haas, A. F., Nelson, C. E., ... Dorrestein, P. C. (2017). High-Resolution Liquid Chromatography Tandem Mass Spectrometry Enables Large Scale Molecular Characterization of Dissolved Organic Matter. *Frontiers in Marine Science*, 4, 405.
- Thingholm, T. E., Jørgensen, T. J. D., Jensen, O. N., & Larsen, M. R. (2006). Highly selective enrichment of phosphorylated peptides using titanium dioxide. *Nature Protocols*, 1(4), 1929–1935.