Colony formation in *Phaeocystis antarctica*: connecting molecular mechanisms with iron biogeochemistry

Sara J. Bender[a,b], Dawn M. Moran[a], Matthew R. McIlvin[a], Hong Zheng[c], John P. McCrow[c], Jonathan Badger[c,f], Giacomo R. DiTullio[e], Andrew E. Allen[c,d], Mak A. Saito[a,*]

[a]Marine Chemistry and Geochemistry Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543 USA

[b]Current address: Gordon and Betty Moore Foundation, Palo Alto, California 94304 USA

[c]Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, California 92037 USA

[d]Integrative Oceanography Division, Scripps Institution of Oceanography, UC San Diego, La Jolla, California 92037 USA

[e]College of Charleston, Charleston South Carolina 29412, USA

[f]Current address: Center for Cancer Research, Bethesda, Maryland 20892, USA

[*]Correspondence to M. Saito (msaito@whoi.edu)

Accepted at Biogeosciences

July 31, 2018

Minor Technical Corrections

Abstract.

Phaeocystis antarctica is an important phytoplankter of the Ross Sea where it dominates the early season bloom after sea ice retreat and is a major contributor to carbon export. The factors that influence *Phaeocystis* colony formation and the resultant Ross Sea bloom initiation have been of great scientific interest, yet there is little known about the underlying mechanisms responsible for these phenomena. Here, we present laboratory and field studies on *Phaeocystis antarctica* grown under multiple iron conditions using a coupled proteomic and transcriptomic approach. *P. antarctica* had a lower iron limitation threshold than a Ross Sea diatom *Chaetoceros* sp., and at increased iron nutrition (>120 pM Fe') a shift from flagellate cells to a majority of colonial cells in *P. antarctica* was observed, implying a role for iron as a trigger for colony formation. Proteome analysis revealed an extensive and coordinated shift in proteome structure linked to iron availability and life cycle transitions with 327 and 436 of proteins measured were significantly different between low and high iron in strains 1871 and 1374, respectively. The enzymes flavodoxin and plastocyanin that can functionally replace iron metalloenzymes were observed at low iron treatments consistent with cellular iron sparing strategies, with plastocyanin having a larger dynamic range. The numerous isoforms of the putative iron-starvation induced protein ISIP group (ISIP2A and ISIP3) had abundance patterns coincided with that of either low or high iron (and coincident flagellate or the colonial cell types in strain 1871), implying that there may be specific iron acquisition systems for each life cycle type. The proteome analysis also revealed numerous structural proteins associated with each cell type: within flagellate cells actin and tubulin from flagella and haptonema structures as well as a suite of calcium-binding proteins with EF domains were observed. In the colony-dominated samples a variety of structural proteins were observed that are also often found in multicellular organisms including spondins, lectins, fibrillins,

and glycoproteins with von Willebrand domains. A large number of proteins of unknown function were identified that became abundant at either high and low iron availability. These results were compared to the first metaproteomic analysis of a Ross Sea *Phaeocystis* bloom to connect the mechanistic information to the *in situ* ecology and biogeochemistry. Proteins associated with both flagellate and colonial cells were observed in the bloom sample consistent with the need for both cell types within a growing bloom. Bacterial iron storage and B_{12} biosynthesis proteins were also observed consistent with chemical synergies within the colony microbiome to cope with the biogeochemical conditions. Together these responses reveal a complex, highly coordinated effort by *P. antarctica* to regulate its phenotype at the molecular level in response to iron and provide a window into the biology, ecology, and biogeochemistry of this group.

1. Introduction

The genus *Phaeocystis* is a cosmopolitan marine phytoplankton group that plays a key role in global carbon and sulfur cycles (Hamm et al., 1999; Matrai et al., 1995; Rousseau et al., 2007; Schoemann et al., 2005; Smith et al., 1991; Solomon et al., 2003; Thingstad and Billen, 1994; Verity et al., 2007). Because of their large cell concentrations during bloom formation, *Phaeocystis* have a significant impact on the ocean biogeochemistry through carbon fixation (Arrigo et al., 1999; Hamm et al., 1999; Matrai et al., 1995; Rousseau et al., 2007; Schoemann et al., 2005; Smith et al., 1991; Solomon et al., 2003; Thingstad and Billen, 1994; Verity et al., 2007), the release of large concentrations of organic carbon upon grazing/viral lysis (Alderkamp et al., 2007; Hamm et al., 1999; Lagerheim, 1896; Verity et al., 2007), and export as aggregates out of the photic zone (DiTullio et al., 2000). Through the production of dimethylsulfide (DMS), they also directly connect ocean and atmospheric processes and carbon and sulfur cycling (Smith et al., 2003).

Some *Phaeocystis* species, including *Phaeocystis antarctica*, undergo multiple morphotypes and can occur as flagellated single-cells or in gelatinous colonies consisting of thousands of non-motile cells (Fig. 1). Microscopic and chemical analyses have found that *Phaeocystis* colonies are filled with a mucilaginous matrix surrounded by a thin, but strong, hydrophobic skin (Hamm, 2000; Hamm et al., 1999). Once formed, cells typically associate with this outer layer of the colony (Smith et al., 2003). Colony formation involves the exudation of (muco)polysaccharides and carbohydrate-rich dissolved organic matter, as well as amino sugars and amino acids; it is estimated that approximately 50 – 80% of *Phaeocystis* carbon is allocated to this extracellular matrix (Hamm et al., 1999; Matrai et al., 1995; Rousseau et al., 2007; Solomon et al., 2003; Thingstad and Billen, 1994). Thus, not only does the colony increase the size of *Phaeocystis* by several orders of magnitude, but the extracellular matrix material also constitutes the majority of measured algal (carbon) biomass (Rousseau et al., 1990). The colonial form of *Phaeocystis* has been suggested as a defense mechanism against grazers (Hamm et al., 1999), a means to sequester micronutrients such as iron and manganese (Noble et al., 2013; Schoemann et al., 2001), as a means of protection from pathogens (Hamm, 2000; Jacobsen et al., 2007), and as a microbiome vitamin B_{12} source (Bertrand et al., 2007). Colony formation of *Phaeocystis* species, including *P. antarctica* and *P. globosa*, has been linked to numerous physiological triggers including the synergistic effects of iron and irradiance (Feng et al., 2010), grazer-induced chemical cues (Long et al., 2007), phosphate concentrations (Riegman et al., 1992), and the presence of different nitrogen species (Riegman and van Boekel, 1996; Smith et al., 2003).

The Ross Sea is one of the most productive regions of the Southern Ocean (Arrigo et al., 1999; 1998; Feng et al., 2010; Garcia et al., 2009; Sedwick and DiTullio, 1997), and the latter is an important contributor to the cycling of carbon in the oceans (Lovenduski et al., 2008; Sarmiento

et al., 1998). In the early spring when the sea ice retreats and polynyas form, phytoplankton blooms and regional phytoplankton productivity are fed by the residual winter iron inventory and perhaps iron-rich sea ice melt (Noble et al., 2013; Sedwick and DiTullio, 1997); blooms have also been linked to changes in irradiance and mixed layer depth (Arrigo et al., 1999; Coale et al., 2003; Martin et al., 1990; Sedwick and DiTullio, 1997; Sedwick et al., 2000). In the Ross Sea Polynya (RSP), *P. antarctica* colonial cells form almost mono-specific blooms until the austral summer season begins, comprising > 98% of cell abundance at the peak of the bloom (Smith et al., 2003). Although diatom abundance dominates in the summer, the RSP typically harbors the co-existence of flagellated single cells of *P. antarctica* along with diatoms (Garrison et al., 2003). During blooms *P. antarctica* can draw down more than twice as much carbon relative to phosphate as diatoms and contribute to rapid carbon export, leaving a lasting biogeochemical imprint on surrounding waters (Arrigo et al., 1999; 2000; DiTullio et al., 2000; Dunbar et al., 1998). Recent *in vitro* iron addition experiments provide evidence that iron nutrition influences *P. antarctica* growth in this region, with increasing *P. antarctica* biomass after iron addition (Bertrand et al., 2007; Feng et al., 2010). Moreover, laboratory experiments with *P. antarctica* have observed a high cellular iron requirement and variable use of strong organic iron complexes (Sedwick et al., 2007; Strzepek et al., 2011; Luxem et al., 2017).

The multiphasic lifecycle of *P. antarctica* in the Ross Sea gives it a spectrum of nutrient drawdown phenotypes and trophic interactions, dependent on the presence of flagellated versus colonial cells (Smith et al., 2003). Given its prominence during early spring sea ice retreat, it has been hypothesized that the triggers of colony formation for *Phaeocystis* cells are also the triggers of the spring phytoplankton bloom. Yet experimental and molecular analyses of potential environmental triggers and how they manifest in changes in cellular morphology have remained

113 elusive. Little is known about the mechanisms responsible for colony formation in *P. antarctica*,
114 nor how these mechanisms respond to an environmental stimulus such as iron, both of which
115 appear to be integral to the ecology and biogeochemistry of *P. antarctica.*

116 **2. Materials and methods**

117 **2.1 Culture experiments**

118 Two strains of *Phaeocystis antarctica* (treated with Provasoli's antibiotics), CCMP 1871 and
119 CCMP 1374 (Provasoli-Guillard National Center for Culture of Marine Phytoplankton), and a
120 Ross Sea centric diatom isolate *Chaetoceros* sp. RS-19 (collected by M. Dennett at 76.5° S, 177.1°
121 W in December 1997 and isolated by D. Moran) were grown in F/2 media with a trace metal stock
122 (minus $FeCl_3$) according to Sunda and Huntsman (Sunda and Huntsman, 2003; 1995), using a
123 modified 10 μM EDTA concentration, and an oligotrophic seawater base. Strains were chosen
124 because they were culturable representatives from two distinct regions in the Southern Ocean.

125 Semi-continuous batch cultures were grown at 4 °C under 200 μmol photons m^{-2} s^{-1}
126 continuous light. Each strain was acclimated to the six iron growth condition concentrations for at
127 least three transfers prior to proteome and growth rate experiments (> 9 generations per transfer
128 for > 27 generations). The concentration of dissolved inorganic iron within each treatment was 2
129 pM, 41 pM, 120 pM, 740 pM, 1200 pM, and 3900 pM Fe' as set by the metal buffer EDTA (where
130 $Fe'/Fe_{Total} = 0.039$) (Sunda and Huntsman, 2003). During the experiment, cultures were maintained
131 in 250 mL polycarbonate bottles; and, subsamples were collected every 1-2 days in 5 mL 13x100
132 mm borosilicate tubes to measure relative fluorescence units (RFUs) and cell counts in the
133 treatments. Mid-to-late exponential phase cultures were harvested for transcriptome and proteome
134 analysis and cell size was measured for both strains; cell pellets were stored at -80 °C (see
135 Supplementary Information for additional methods). Cell counts were conducted using a Palmer-

Maloney counting chamber and a Zeiss Axio Plan microscope on 400x magnification; cell numbers were used to determine the final growth rate of each strain/treatment. During mid-to-late exponential phase (time-of-harvest), cell size was determined for both strains (n=20 cells were counted for each strain), calculated using the Zeiss 4.8.2 software and a calibrated scale bar. The number of cells in colonies (versus as single cells) was determined for strain 1871 only. Briefly, counts (number of cells associated with colonies versus unassociated) were averaged from 10 fields of view at five distinct time points (50 fields of view total).

2.2 Protein extraction, digestion, and mass spectrometry analyses

Proteins from cell pellets (one pellet per treatment, two strains and six iron treatments for a total of 12 proteomes) was extracted using the detergent B-PER (Thermo Scientific), quantified, purified by immobilization within an acrylamide tube gel, trypsin digested, alkylated and reduced, and analyzed by liquid chromatography-mass spectrometry (LC-MS) using a Michrom Advance HPLC with a reverse phase C18 column (0.3 x 10 mm ID, 3 µm particle size, 200 Å pore size, SGE Protocol C18G; flow rate of 1 µL min^{-1}, nonlinear 210 min gradient from 5% to 95% buffer B, where A was 0.1% formic acid in water and B was 0.1% formic acid in acetonitrile, all solvents were Fisher Optima grade) coupled to a Thermo Scientific Q-Exactive Orbitrap mass spectrometer with a Michrom Advance CaptiveSpray source. The mass spectrometer was set to perform MS/MS on the top 15 ions using data-dependent settings (dynamic exclusion 30 s, excluding unassigned and singly charged ions), and ions were monitored over a range of 380-2000 m/z (see Supplementary Information for detailed protocol). Peptide-to-spectrum matching was conducted using the SEQUEST algorithm within Proteome Discoverer 1.4 (Thermo Scientific) using the translated transcriptomes for *P. antarctica* strain 1871 and strain 1374 (Fig 2., see below).

159 Normalized spectral counts were generated from Scaffold 4.0 (Proteome Software Inc.), with a
160 protein false discovery rate (FDR) of 1.0%, a minimum peptide score of 2, and a peptide
161 probability threshold of 95%. Spectral counts refer to the number of peptide-to-spectrum matches
162 that are attributed to each predicted protein from the transcriptome analysis, and the Scaffold
163 normalization scheme involves a small correction normalizing the total number of spectra counts
164 across all samples to correct for run-to-run variability and improve comparisons between
165 treatments. The R package "FactoMineR" (Lê et al., 2008) was used for the PCA analysis; for
166 heatmaps, the package "gplots" was used (Warnes et al., 2009). Proteomic samples taken from
167 each laboratory condition were not pooled downstream as part of the analyses; replicates shown
168 for each treatment are technical replicates.
169
170 **2.3 RNA extraction, Illumina sequencing, and annotation**
171 For *P. antarctica* cultures total RNA was isolated from cell pellets (one pellet per treatment, two
172 strains and three iron concentrations for a total of six transcriptomes) following the TRIzol Reagent
173 (Life Technologies, manufacturer's protocol). RNeasy Mini kit (Qiagen) was used for RNA
174 cleanup, and DNase I (Qiagen) treatment was applied to remove genomic DNA. Libraries, from
175 polyA enrichment mRNA, were constructed using a TruSeq RNA Sample Preparation Kit V2
176 (IlluminaTM), following the manufacturer's TruSeq RNA Sample Preparation Guide. Sequencing
177 was performed using the Illumina HiSeq platform. Downstream, reads were trimmed for quality
178 and filtered. CLC Assembly Cell (CLCbio) was used to assemble contigs, open reading frames
179 (ORFs) were predicted from the assembled contigs using FragGeneScan (Rho et al., 2010), and
180 additional rRNA sequences were removed. The remaining ORFs were annotated de novo via
181 KEGG, KO, KOG, Pfam, and TigrFam assignments. Taxonomic classification was assigned to

each ORF and the Lineage Probability Index (LPI, as calculated in (Podell and Gaasterland, 2007). ORFs classified as Haptophytes were retained for downstream analyses. Analysis of sequence counts ("ASC") was used to assign normalized fold change and determine which ORFs were significantly differentially expressed in pairwise comparisons between treatments. The ASC approach offers a robust analysis of differential gene expression data for non-replicated samples (Wu et al., 2010).

For metatranscriptomes, RNA was extracted from frozen cell pellets using the TRIzol reagent manufacturer's protocol (Thermo Fisher Scientific) (see Supplementary Information for additional details on metatranscriptome processing).

2.4 Ross Sea *Phaeocystis* bloom: sample collection and protein extraction and analysis

The meta 'omics samples were collected in the Ross Sea (170.76° E, 76.82° S) during the CORSACS expedition (Controls on Ross Sea Algal Community Structure) on December 30, 2005 (near pigment station 137; http://www.bco-dmo.org/dataset-deployment/453377) (Saito et al., 2010; Sedwick et al., 2011). Surface water was concentrated via a plankton net tow (20 μm mesh), gently decanted of extra seawater, then split into multiple replicate cryovials and frozen in RNAlater at -80 °C for metatranscriptome and metaproteome analysis. The pore size of the net tow would have preferentially captured the colony form of *Phaeocystis*, although filtration with small pore size membrane filters was particularly challenging during this time period due to the abundance of *Phaeocystis* colonies and the clogging effect of their mucilage. Moreover, the physical process of deploying the net tow appears to have entrained some smaller cells including the *Phaeocystis* flagellate cells by adsorption to partially broken colonies and associated mucilage as observed in the metaproteome results. Two of these replicate bloom samples were frozen for

proteome analysis. A third replicate sample from this field site was extracted for metatranscriptome analysis as described above.

Proteins were extracted, digested, and purified following the lab methods above, and then identified first on a Thermo Q-Exactive Orbitrap mass spectrometer using a Michrom Advance CaptiveSpray source, then samples were subsequently re-run on a two-dimensional chromatographic nanoflow system for increased metaproteomic depth on a Thermo Fusion Orbitrap mass spectrometer (see supplemental materials for further details). Proteins were then identified within the mass spectra using three databases (Fig. 2): the translated transcriptome database for both *Phaeocystis* strains (Database #1), a Ross Sea metatranscriptome generated in parallel from this metaproteome sample (Database #2; this transcriptome is a combination of eukaryotic and prokaryotic communities derived from total RNA and poly(A) enriched RNA sequencing), and a compilation of five bacterial metagenomes from the Amundsen Sea polynya (Database #3) (Delmont et al., 2014), using SEQUEST within Proteome Discoverer 1.4 (Thermo Scientific) (Eng et al., 1994) and collated with normalized spectral counts in Scaffold 4.0 (Proteome Software Inc.) (see Supplementary Information for additional details).

2.5 Data availability

Phaeocystis antarctica RNA sequence data reported in this paper have been deposited in the NCBI sequence read archive under BioProject accession no. PRJNA339150, BioSample accession nos. SAMN05580299 – SAMN05580303. Ross Sea metatranscriptomes have been deposited under BioProject accession no. PRJNA339151, BioSample accession nos. SAMN05580312 – SAMN05580313. Proteomic data from the lab and field components was submitted to the Pride database (Project Name: *Phaeocystis antarctica* CCMP 1871 and CCMP 1374, Ross Sea

Phaeocystis bloom, LC-MSMS; Project accession: PXD005341; Project DOI: 10.6019/PXD005341).

3. Results and discussion

3.1 Physiological response to iron availability: Growth limitation and colony formation

The two strains of *P. antarctica* (1374 and 1871 hereon) were acclimated to six iron concentrations to capture the metabolic response under different iron regimes (Fig. 3*a* and *b*). A biphasic response in *P. antarctica* strain 1871 was observed; cultures exhibited a clear single-cell versus colony response to low and high iron, respectively, that were observed by microscopy and were readily apparent by naked eye due to the millimeter size of the colonies. The three low iron treatments (2 pM, 41 pM, and 120 pM Fe') cultures contained only single, flagellated cells, whereas the three higher iron treatments (740 pM, 1200 pM, and 3900 pM Fe') had a majority of colonial cells, based on detailed microscopy counts shown in Fig. 3*c*. This influence of iron on colony abundance was observed in an additional experiment, where colonial cells were again absent at the lowest three iron concentrations and were present at the three higher concentrations (Fig. S10). The presence of both colony and flagellate cells is expected in actively growing populations since reproduction can involve returning to the flagellate life cycle stage (Rousseau et al., 1994). Single cells and colonies were not counted in experiments with strain 1374, as these experiments were conducted prior to those of 1871 and the iron-induced colony formation observations therein. However, strain 1374 was observed to become "clumpy" at high iron. This clumping observation may reflect the loss of a specific factor needed for the colony completion lost during long-term maintenance in culture. This interpretation is consistent with the overall similar structural protein expression patterns observed in both strains described below. Strzepek et al. also observed co-

varying of iron concentration and colony formation in some strains of *P. antarctica* (Strzepek et al., 2011).

The two strains of *P. antarctica* were able to maintain growth rates for all but the lowest of iron concentrations used here, similar to prior studies of *P. antarctica* strain AA1 that observed no effect of scarce iron on growth rates (Strzepek et al., 2011). Parallel experiments with polar diatoms such as *Chaetoceros* (Fig. 3*d*) observed growth limitation at moderate iron abundances using an identical media composition, indicating 1) that *P. antarctica* has an impressive capability for tolerating low iron compared to *Chaetoceros* and other diatoms (e.g. a Ross Sea *Pseudo-nitzschia* sp. isolate, data not shown), and 2) demonstrating an absence of iron contamination in these experiments. Growth rates for 1871 were significantly different between the 2 pM Fe' treatment and all other treatments (student's t-test with Bonferroni correction, $p < 0.05$; Fig. 3*a*); there were no significant differences among growth rates for strain 1374 (Fig. 3*b*). Cell size (including both flagellate and colonial cells) decreased with lower iron concentration, a trend that was statistically significant (ANOVA with TukeyHSD test, $p < 0.05$) for both strains when cell sizes from each high iron treatment (740 pM, 1200 pM, and 3900 pM Fe') were compared to cell sizes from each low iron treatment (2 pM, 41 pM, and 120 pM Fe') (Fig. 3*e* and *f*).

3.2 Molecular response to low and high iron concentrations

Global proteomics enabled by peptide-to-spectra matching to transcriptome analyses, revealed a clear statistically significant molecular transition across the iron gradient for each strain (Fig. 4). The global proteome consisted of 536 proteins identified in strain 1871 and 1085 proteins identified in strain 1374 (Table 1; Supplementary Data 1), after summing unique proteins across the six iron treatments. There were 55 proteins identified in strain 1871 and 64 proteins in strain

1374 (Fig. 4) that drove the statistical separation of proteomes across iron treatments using principle component analysis (PCA, Axis 1 PCA correlation coefficient ≥ 0.9 or ≤ -0.9). Axis one accounted for 49% variance for 1871 and 36% variance for 1374. Moreover, using a Fisher Test (P-value ≤ 0.05), 327 proteins (strain 1871) and 436 proteins (strain 1374) of those proteins detected were identified as significantly different in relative protein abundance between representative "low" (41 pM Fe') and "high" (3900 pM Fe') iron treatments. This significant change in the proteome composition paralleled observations of a shift from flagellate to colonial cells. Iron-starvation responses and iron metabolism were detected within the high and low iron PCA protein subsets, including iron-starvation induced proteins (ISIPs), flavodoxin, and plastocyanin, demonstrating a multi-faceted cellular response to iron scarcity (Fig. 5). Surprisingly, there was also a highly pronounced signal in the proteome that appeared to reflect the structural changes occurring in *P. antarctica*. These structural proteins included multiple proteins with protein family (PFam) domains suggestive of extracellular function, adhesion, and/or ligand binding, including putative glycoprotein domains (for example, spondin) that were present in the high iron PCA subset in both strains (Fig. 5); the appearance of these proteins also corresponded to the occurrence of colonies in strain 1871 (Fig. 1). Similarly, a distinct suite of proteins was more abundant in the low iron PCA subset (Fig. 5), including proteins relating to cell signaling (for example, calmodulin/EF-hand, PHD zinc ring finger). A number of proteins with unknown function were also detected in the PCA subsets: 71% unknown for strain 1871 and 42% unknown for strain 1374 of a total of 311 proteins annotated as hypothetical proteins (Supplementary Dataset 1). Outside of the PCA analyses, additional iron and adhesion-related proteins were identified that demonstrated a similar expression profile to the PCA subset (Supplementary Fig. 1).

297 Identification and characterization of proteins and transcripts induced by iron scarcity are
298 valuable in improving an understanding of the adaptive biochemical function of these complex
299 phytoplankton as well as for their potential utility for development as environmental stress
300 biomarkers (Roche et al., 1996; Saito et al., 2014). The enzymes flavodoxin and plastocyanin,
301 which require no metal and copper, respectively and that functionally replace iron metalloenzymes
302 counterparts ferredoxin and cytochrome c6, had isoforms that increased in concentrations at the
303 lower iron treatments consistent with cellular iron sparing strategies (Fig. 6, Supplementary Fig.
304 2) (Peers and Price, 2006; Whitney et al., 2011; Zurbriggen et al., 2008). In strain 1374 however,
305 there was an increase in both of these iron-sparing systems at the highest iron concentration (Fig.
306 6d and 6f, Supplementary Table 1). While during both experiments, cells were growing
307 exponentially at the time of harvest, those of strain 1374 were as much as 7.6 fold denser in cell
308 number than those of strain 1874 (based on cell counts from treatments specifically used for
309 transcriptome analyses), and as a result the denser 1374 strain might have also experienced iron
310 stress even at this highest iron concentration as the high biomass depleted iron within the medium.
311 Of these two iron sparing enzymes, plastocyanin appeared to show a clearer increase in abundance
312 at lower environmental iron concentrations (Fig 6c and 6f). In contrast, some flavodoxin isoforms
313 could be interpreted as being constitutive, two of the three isoforms were still present in reasonable
314 spectral counts at higher iron concentrations (Figs. 6a and 6d). Prior measurements during a Ross
315 Sea colonial *P. antarctica* spring bloom in 1998 were consistent with this interpretation, with
316 ferredoxin concentrations were below detection and flavodoxin present (Maucher and DiTullio,
317 2003). A constitutive flavodoxin could help explain *P. antarctica's* ability to tolerate all but the
318 lowest iron treatment observed in the physiological experiments (Fig 3a and 3b), and implies that

319 the careful selection of isoforms, or better, the inclusion of all isoforms of a protein biomarker of
320 interest may be valuable in interpreting complex field results.

321 There were also numerous isoforms of the iron-starvation induced proteins (ISIP) group
322 identified within the proteome of each *P. antarctica* strain: specifically 9 ISIP2A's and 3 ISIP3's
323 in strain 1871 and 3 ISIP2A's and 4 ISIP3's in strain 1374 (Supplementary Fig. 1; Supplementary
324 Table 1). These ISIPs were identified based on their transcriptome response to iron stress in
325 diatoms and most recently have been implicated in a diatom cell surface iron concentrating
326 mechanism (Allen et al., 2008; Morrissey et al., 2015). Interestingly in this *P. antarctica*
327 experiment, these ISIPs exhibited both "high" or "low" iron responses, where specific isoforms
328 were more abundant only under one of those respective conditions (Fig. 6). Given the
329 metamorphosis of *P. antarctica* between flagellate and colonial cell types observed by microscopy
330 and the proteome across the gradient in iron concentrations, we hypothesize that this diversity of
331 iron stress responses in the ISIP proteins may reflect the complexity associated with *P. antarctica*'s
332 life cycle. As the abundant winter iron and sloughed basal sea ice reserves are depleted, newly
333 formed colonial cells will inevitably find themselves in the iron-depleted environments that have
334 been characterized in the Ross Sea almost immediately upon bloom formation due to iron's small
335 dissolved inventory (Bertrand et al., 2015; Sedwick et al., 2011). As a result, *P. antarctica* may
336 have distinct iron stress protein isoforms associated specifically with the colonial cell type (such
337 as the high iron/colonial ISIP proteins, Figs. 5 and 6) in order to acquire scarce iron during blooms,
338 in addition to a distinct suite of iron stress proteins produced within the flagellate cells (low
339 iron/flagellate ISIP proteins, also Figs. 5 and 6). Given the rapid depletion of iron during Ross Sea
340 blooms, it is also conceivable that these iron acquisition proteins are constitutively expressed
341 within the colony morphotype, rather than being connected to an iron-sensing and regulatory

response system. Future short-term iron perturbation studies that would complement the steady-state experiments presented here could further investigate this hypothesis. The multiplicity of ISIP proteins produced within each strain also is consistent with the observation that both *P. antarctica* strains maintained high growth rates even at the lower 41 and 120 pM Fe' concentrations, compared to the diatom *Chaetoceros sp.* whose growth rate is less than 50% of maximal growth in similar media (Fig. 3).

3.3 Correspondence between RNA and protein biomolecules

Many of the RNA transcripts of iron-related genes trended with their corresponding proteins: 60% of the iron-related gene transcripts reflected the proteomic response in strain 1871, whereas there was a 30% correspondence between iron-related transcripts and proteins in strain 1374 (Supplementary Fig. 1). In total, 47% of expressed proteins in strain 1871 and 26% of proteins in strain 1374 shared expression patterns with associated transcripts (Fig. 7), consistent with recent studies of proteome-transcriptome comparisons that showed limited coordination between inventories of each type of biomolecule (Dyhrman, 2012). As mentioned above, while both experiments were in exponential growth at the time of harvest, strain 1374 was 7.6 fold denser in cell number than those of strain 1874 at that time. Hence, this decrease in transcript-proteome coherence in strain 1374 may be related to harvesting in late-log growth phase, and reflects the challenge of trying to conduct comparisons of these biomolecules that function on different cellular timescales.

Examination of the transcriptome revealed a significant increase in transcripts for tonB-like transporters, which can be associated with cross-membrane nutrient transport (e.g. for iron siderophores complexes or vitamin B_{12} (Bertrand et al., 2007; 2013; Morris et al., 2010) under high

iron for strain 1871; and, significantly greater transcript abundances for a putative flavodoxin for strain 1374 under low iron consistent with its substitution for ferredoxin due to iron scarcity (Roche et al., 1996).

3.4 Observation of an iron-induced switch from single cells to colonies

The strong connection of iron availability to putative structural components of *P. antarctica* observed here served as an ideal opportunity to examine the genes and proteins involved in morphological and life cycle transitions and colony construction in this phytoplankter that can otherwise be experimentally difficult to trigger in isolation. *Phaeocystis* colonies have captured the interest of scientists for more than a century (Hamm et al., 1999), yet next to nothing is known about the molecular basis of their construction. Colonies have been considered a collection of loosely connected cells embedded within a gel matrix, and hence described as "balls of jelly" or "bags of water" (Hamm et al., 1999; Lagerheim, 1896; Verity et al., 2007). Results here suggest significant transformations in the cellular proteome that corresponded to solitary and colonial morphological stages, for example, involving structural proteins and proteins known to be post-translationally modified such as glycoproteins or those containing glycoprotein-binding motifs. To our knowledge, such an extensive proteome remodeling has yet to be observed for another colonial organism, or with the influence of any other environmental stimuli in the genus *Phaeocystis*. As a result the details of this response, while fascinating, are challenging to interpret due to their novelty.

A putative spondin protein exhibited one of the largest responses between low and high iron in both strains with a greater than 20-fold increase in relative protein abundance and normalized 11-fold change in transcript abundance in strain 1871, and a greater than 9-fold

388 increase in relative protein abundance and 3-fold change in transcript abundance in strain 1374
389 (Fig. 5*a* and Supplementary Data 1). Spondin proteins are known to be glycosylated, and to be a
390 component of the extracellular matrix (ECM) environment, which may enable multicellularity in
391 metazoans through cell adhesion, and have been found to help coordinate nerve cell development
392 through adhesion and repulsion (Michel et al., 2010; Tzarfati-Majar et al., 2011). Despite this large
393 variation in protein abundance, the function of spondins in eukaryotic phytoplankton, including
394 *Phaeocystis* remains largely unknown. Given their responsiveness to iron availability and
395 associated enrichment in colony rich cultures, these proteins could potentially contribute to ECM-
396 related adhesion of cells, to each other or the colony skin, or even perhaps to the mucilage interior.
397 Additional glycoproteins that exhibited a strong iron response in both strains include those
398 containing von Willebrand factor domains (for example, protein families PF13519, PF00092), and
399 fibrillin and lectin (Fig. 5 and Supplementary Fig. 1). In biomineralizing organisms, such as corals,
400 glycoproteins with von Willebrand domains are hypothesized to play a role in the formation of the
401 extracellular organic matrix through adhesion (Drake et al., 2013; Hayward et al., 2011) laying the
402 scaffolding for calcification. Orthologs of the von Willebrand proteins that contain these domains
403 have also been characterized in humans and have protein-binding capabilities, which are important
404 for coagulation (Ewenstein, 1997). These dynamic von Willebrand proteins appear to contribute
405 to the cell aggregation and colony formation of *P. antarctica* colonies.
406 The suite of structural and modified proteins described above demonstrates a means
407 through which *P. antarctica*'s colonial morphotype could be constructed, and this dataset provides
408 rare molecular evidence for the proteome reconstruction needed to switch between single
409 organisms to a multicellular colony. The evolution of multicellularity in Eukaryotes is an area of
410 significant interest that has mostly focused on model organisms with colonial forms such as

411 Choanoflagellates and *Volvox* (Abedin and King, 2010). Genomic studies of the former identified
412 the presence of protein families involved in cell interactions within metazoans, including C-type
413 lectins, cadherins, and fibrinogen (King et al., 2003). In other lineages of microalgae that form
414 colonial structures, such as *Volvox carteri*, there is supporting evidence for glycoproteins cross-
415 linking within the extracellular matrix of colonies (Hallmann, 2003), as well as serving other
416 important functional roles in cell-cell attachment during colony formation (for example, colony
417 formation in the cyanobacteria *Microcystis aerginosa*) and as an integral component of cell walls
418 (for example, the diatoms *Navicula pelliculosa* and *Craspedostauros australis*) (Chiovitti et al.,
419 2003; Kröger et al., 1994; Zilliges et al., 2008). In this study, environmental isolates of *P.*
420 *antarctica* displayed consistent trends in similar protein families (for example, lectins, fibrillins,
421 and glycoproteins), and they were produced at higher levels under elevated iron conditions when
422 strain 1871 was primarily in colonial form. Given *P. antarctica*'s environmental importance and
423 an ability to control the transition between flagellates and colony cell types through iron
424 availability, *P. antarctica* may serve as a useful model for studying multicellularity in nature and
425 in the context of environmental change.

426 In contrast to these putative colonial structural proteins, there were canonical cytoskeletal
427 proteins such as actin and tubulin observed in *P. antarctica* cultures grown under low iron
428 conditions (Supplementary Fig. 1). These proteins were likely associated with the flagella and the
429 haptonema, a shorter organelle containing nine microtubules that is characteristic of Haptophytes
430 (Zingone et al., 1999), found in the solitary *Phaeocystis* cell, and similar to other eukaryotic
431 flagellar systems such as *Chlamydomonas* (Watanabe et al., 2004). Additionally, a suite of proteins
432 with calcium-binding domains (EF-hand protein families) was identified in greater relative
433 abundance under low iron growth conditions in both strains (Fig. 5; Supplementary Fig. 1 and

Supplementary Data 1). In diatoms, calcium-signaling mechanisms have been directly linked with how cells respond to bioavailable iron, as well as stress responses (Allen et al., 2008; Vardi, 2008). Calcium (and magnesium) ions also play an integral role in the ability for extracellular mucus to gel (van Boekel, 1992). The greater abundance of putative calcium-binding proteins under low iron conditions suggests an important role for intracellular calcium, either in its involvement in flagellate motility and/or having a role in inhibiting the cells' abilities to form colonies while under iron limitation. This use of calcium signaling is notable given that calcium is a major constituent of seawater (0.01 mol L^{-1}), implying a need for efflux and exclusion of calcium from the cytoplasm.

3.5 *Phaeocystis antarctica* strain-specific responses

Phaeocystis antarctica is believed to have speciated from warm-water ancestors, and populations within the Antarctic are mixed via the rapid Antarctic Circumpolar Current (ACC, 1-2 years) circulation with the Ross Sea and Weddell Sea, which entrains strains nearly simultaneously (Lange et al., 2002). Moreover, high genetic diversity has been observed across a large number of *P. antarctica* isolates and even within isolates co-isolated from a bloom (Gäbler-Schwarz et al., 2015). Given the differences in geographic location of the isolates used in this study, there may be some differences regarding adaptation and ecological role between them. In the Ross Sea, *P. antarctica* dominates, and cells exhibit seasonal variability between flagellated states (early Spring, late summer) and colonial stage (late Spring/early summer) (Smith et al., 2003). In contrast, in the Western Antarctic Peninsula, near the Weddell Sea where strain 1871 was isolated from (Palmer station), *P. antarctica* is outnumbered by diatoms and cryptomonads in terms of algal biomass, and colonies are generally rare (Ducklow et al., 2007). While global proteomic and

transcriptomic analyses revealed differences among strains (Supplementary Data 1), both strains had responses that overwhelmingly supported a concerted effort towards structural changes under high iron versus low iron, consistent with the minor phylogenetic differences previously reported for *P. antarctica* isolates due to rapid ACC circulation (Lange et al., 2002).

3.6 Examination of a *Phaeocystis* bloom metaproteome from the Ross Sea

The detailed laboratory studies above can be compared to a first metaproteomic analysis of a Ross Sea *Phaeocystis antarctica* bloom to provide an examination of the *in situ* ecology and biogeochemical and their underlying biochemical signatures. Due to the newness of metaproteomic eukaryotic phytoplankton research, some methodological detail has been incorporated into this section. For field analysis a net tow sample was collected north of Ross Island (Fig. 8) on December 30th 2005, in which *Phaeocystis* colonies were visually dominant. Temporal changes in the bloom composition have been described for this summer expedition and an austral spring expedition later that year (NBP06-01 and NBP06-08, respectively), and a shift was observed from a *P. antarctica* dominated ecosystem to a mixture of *P. antarctica* and diatoms (Smith et al., 2013). Surface pigment distributions showed the sampling region to be within a particularly intense bloom dominated by *Phaeocystis* as observed by abundant 19'-hexanoyloxyfucoxanthin pigment (Fig. 8), reaching concentrations of 1096 ng L^{-1} and total chlorophyll *a* concentrations of 1860 ng L^{-1} on the sampling day. CHEMTAX analysis of these HPLC pigments found that *P. antarctica* populations accounted for approximately 88% of surface water total chlorophyll at this time. Fucoxanthin pigment, characteristic of diatoms, was lower here (136 ng L^{-1}) compared to samples from the western Ross Sea (Fig. 8), consistent with prior Ross Sea observations. Repeated sampling near the sampling region (~77.5^{o}S) two weeks after

taking the metaproteome sample found lower overall chlorophyll *a* levels (Smith et al., 2013), consistent with bloom decay. Iron measured very near this location (76.82° S, 170.76° E also on December 30, 2005), found a surface dissolved iron concentration of 170pM (6m depth) and an acid-labile particulate iron concentration of 1590 pM (Sedwick et al., 2011), consistent with iron depletion in seawater following drawdown of the accumulated winter iron supply and incorporation of iron into biological particulate material (Noble et al., 2013; Sedwick et al., 2000).

The metaproteome analyses of the Ross Sea sample were conducted by bottom-up mass spectrometry analysis of tryptic peptides using initially a 1-dimension and subsequently a deeper 2-dimension chromatographic methodology (1D and 2D hereon), followed by peptide-to-spectrum matching of putative peptide masses and their fragment ions to predicted peptides from translated DNA sequences. While this approach is common for model organisms and has been successfully applied to primarily prokaryotic components of natural communities (Morris et al., 2010; Ram et al., 2005; Sowell et al., 2008; Williams et al., 2012), there continue to be challenges in metaproteomics analyses of diverse communities particularly when including an extensive eukaryotic component such as is present in the Ross Sea phytoplankton bloom. VerBerkmoes et al. (2005) demonstrated the feasibility of using mass spectrometry metaproteomic analysis for the detection of eukaryotic proteins in a complex sample matrix. To address these issues, we utilized three sequences databases for peptide-to-spectrum matching (see Methods and Supplementary Information Table S2). Analysis of both unique (tryptic) peptides and identified proteins are provided here, where unique peptides are particularly valuable in metaproteome interpretation as a basal unit of protein diversity that can be definitively compared across the three sequence databases (Saito et al., 2015).

The combined *P. antarctica* strain transcriptome database (Database #1) generated the largest number of protein and unique peptide identifications 1545 and 3816 in 2D, (912 and 2103 in 1D), respectively (Table 2, Fig. 9*a*). This strong relative performance of the strain database was surprising, and likely reflects the depth of the *P. antarctica* isolate transcriptomes and resultant translation into greater metaproteomic depth. Approximately sixty percent of field identifications mapped to strain 1374 (57%); a broad synthesis of all proteomes based on KOG annotations also indicated that the metaproteomes appeared most similar to the Ross Sea strain 1374 (Supplementary Fig. 3). The Ross Sea metatranscriptome database (Database #2) resulted in 1475 proteins and 3210 unique peptides in 2D analyses (859 proteins and 1520 unique peptides in 1D) distributed across a large number of taxa, with 324 of those proteins associated with *P. antarctica.* The Antarctic bacterial metagenome database (Database #3) produced 102 proteins and 237 unique peptides in 2D (98 proteins and 186 peptides in 1D) that mapped to bacteria likely associated with the phytoplankton communities, given the use of a net that would not otherwise capture free-living bacteria. The low number of bacterial protein and peptides identifications could reflect their small abundance or limited metagenomic coverage. Due to the extensive diversity present, there was overlap between the peptide identifications from each database for the 5885 total unique peptides in2D (3193 in 1D) : 1222 (in 2D; 544 in 1D) *P. antarctica* peptides were shared between the *Phaeocystis* strain and Ross Sea metatranscriptome databases, 158 (in 2D; 69 in 1D) bacterial peptides were in common between the Ross Sea metatranscriptome and the bacterial metagenomic databases, followed by very small numbers shared between bacterial metagenome and the *Phaeocystis* strains database searches (8 peptides in both 1D and 2D), and all three databases (7 and 4 peptides in 1D and 2D, respectively), likely due to a small fraction of tryptic peptides shared between diverse organisms (Saito et al., 2015).

525 This multi-database approach and the relatively low overlap illustrates the necessity of
526 employing diverse sequence databases that target distinct components of the biological
527 community, as well as the value in coupling metatranscriptomic and metagenomic sequence
528 databases to metaproteomic functional analysis to capture the extent of natural diversity. This is
529 evident in the taxon group analysis, where the metatranscriptome has a large representation of
530 Dinophyta and diatoms and only a small contribution from Haptophyta that include *Phaeocystis*,
531 likely due to the large genome sizes and transcription rates, particularly of dinoflagellates, and
532 perhaps due to interferences of *Phaeocystis* RNA extraction due to the copious mucilage present
533 (Fig. 9*b*). In contrast the metaproteome derived from the metatranscriptome database is dominated
534 by Haptophyta and Dinophyta, with minor contributions from other groups (Fig. 9*d*), reflecting
535 the dominant organismal composition seen in the pigment analyses (Fig. 8). Due to a coarse net
536 mesh size much larger than a typical bacterial cell, the bacterial community captured by these
537 metatranscriptome and metaproteome analyses most likely reflects the microbiome associated with
538 larger phytoplankton and protists, particularly within the abundant *P. antarctica* colonies.
539 Database #2 and #3 result in 211 and 102 bacterial protein identifications (in 2D; 148 and 100 in
540 1D), respectively, including representatives from *Oceanospirillacea*e, *Rhodobacteraceae*,
541 *Cryomorphaceae*, *Flavobacteria*, and *Gamma proteobacteria* (Fig. 9*c* and *d*). The lower number
542 of bacterial identifications could be due to low bacterial biomass in the net tow sample relative to
543 phytoplankton biomass and/or limited metagenomics coverage.

544 Together this Ross Sea bloom metaproteome-metatranscriptome analysis provides a
545 window into the complex interactions of this community with its chemical environment.
546 *Phaeocystis antarctica* proteins were abundant in the sample with over 450 (in 2D; 300 in 1D)
547 proteins identified, yet interestingly, we identified proteins associated with both high and low iron

548 treatments, including those corresponding to flagellate and colonial life stages identified in the
549 culture experiments (Fig. 10 and Supplementary Fig. 1). This presence of both life cycle stages of
550 *Phaeocystis* could be interpreted as evidence of an actively growing bloom, with growing
551 flagellate cells coalescing to form new colonies, as well as a standing stock of colonial cells. As
552 mentioned earlier, division and growth of *P. antarctica* colonies is believed to require transitioning
553 back through the flagellate life cycle stage (Rosseau et al., 1994), hence a mixed population of
554 flagellate and colonial stages would be expected of a growing population, consistent with our
555 laboratory observations (Fig. 3*c*).

556 The presence of well-known iron-sparing proteins such as plastocyanin (Fig. 10) was
557 consistent with the depleted dissolved iron concentration (170 pM) in nearby surface waters that
558 are closest to the 120 pM Fe' of the low iron treatments (Peers and Price, 2006; Sedwick et al.,
559 2011), as well as incubation experiments on the same expedition initiated three days prior that
560 demonstrated iron limitation of *P. antarctica* (and iron-B_{12} colimitation of diatom) populations
561 (Bertrand et al., 2007). Notably, the actual Fe' of the Ross Sea was likely considerably lower than
562 this due to the presence of strong organic iron complexes (Boye et al., 2001). Strzepek et al. found
563 evidence for growth of *P. antarctica* and some polar diatoms on strong organic iron complexes at
564 somewhat reduced growth rates in their culture experiments, implying a high-affinity iron
565 acquisition system such as a ferric reductase, although the molecular components of such a system
566 have yet to be identified in *P. antarctica* (Strzepek et al., 2011). As described above, it is likely
567 that both flagellate and colonial cell types have a need to manifest iron stress responses (e.g.
568 distinct ISIP proteins found in the flagellate and colonial dominated cultures, Figs. 5 and 6), and
569 that those distinct responses may be based on the extensive physical differences between life cycle
570 phenotypes. The low contribution of chain-forming diatoms to this metaproteome sample was

consistent with the higher sensitivity of some Ross Sea diatom strains to iron stress such as *Chaetoceros* (Fig. 3*d*) and the low iron availability. Careful examination of targeted mass spectrometry results (precursor and fragment ion analysis) for select iron proteins identified in culture studies showed consistently high quality chromatograms within the field sample, demonstrating a capability to measure these potential peptide biomarkers within complex environmental samples in future field studies characterizing bloom and biogeochemical dynamics (Fig. 11 and Supplementary Figs. 4-10).

The metaproteome analyses also captured relevant functional elements of the bacterial microbiome associated with the eukaryotic community, based on the bacterial proteins identified in both the bacterial databases and the Ross Sea metatranscriptome (Fig. 9*c* and 9*d*). For example, the SAR92 clade of proteorhodopsin-containing heterotrophic bacteria was present (Stingl et al., 2007), and expressed both the iron storage protein bacterioferritin and TonB receptors, the latter of which are involved in siderophore and B_{12} transport. In addition, the Fur iron regulon, iron-requiring ribonucleotide reductase, as well as the vitamin related CobN cobalamin biosynthesis protein, B_{12}-requiring methyl-malonyl CoA, and thiamine ABC transporter were observed from several heterotrophic bacteria species including *Oceanospirillacea*e, *Rhodobacteraceae*, and *Cryomorphaceae* (Supplementary Data 2) (Bertrand et al., 2015; Murray and Grzymski, 2007). These results imply that heterotrophic bacteria known to be associated with the *Phaeocystis* colonies, such as SAR92 and *Oceanospirillaceae*, were also likely responding to micronutrients by concentrating and storing iron, and through biosynthesis of B_{12}. In doing so this bacterial microbiome could have been harboring an “internal” source of the micronutrients, fostering a mutualism with *Phaeocystis* colonies in exchange for a carbon source and consistent with the high particulate iron measured during this station (Sedwick et al., 2011). Together this could create a

competitive advantage for *P. antarctica* relative to the iron and B_{12}-stressed diatoms for early season bloom formation, as previously hypothesized and observed in the Ross Sea in enrichment studies (Bertrand et al., 2007). Although diatoms were less prominent in the dataset, several diatom proteins identified were indicative of the potential for iron stress (e.g., plastocyanin and ISIP3; Supplementary Data 2); however, the diatom CBA1 cobalamin acquisition protein was not identified in the metatranscriptome, and hence would not be detected in the metaproteome using the current methods, but could be targeted in future studies from this dataset.

4. Conclusions

Phaeocystis antarctica is a major contributor to Southern Ocean primary productivity, yet arguably is one of the least well understood of key marine phytoplankton species. The multiple life cycle stages of *P. antarctica* add to its ecological and biochemical complexity. Here we have undertaken a detailed combined physiological and proteomic analysis enabled by transcriptomic sequencing under varying conditions of iron nutrition, and compared these to an initial study of the metaproteome of a Ross Sea *Phaeocystis* bloom. These results demonstrate that *P. antarctica* has evolved to utilize elaborate capabilities to confront the widespread iron scarcity that occurs in the Ross Sea and Southern Ocean, including iron metalloenzyme sparing systems and the deployment of transport and other systems that appear to be unique to the flagellate and colonial morphotypes. To our surprise, increasing iron abundance triggered colony formation in one strain in this study, and visual and proteomic evidence implied the second strain was also attempting to do so. Prior studies have invoked light irradiance and mixed layer depth as key factors in colony production and the concurrent Ross Sea *P. antarctica* bloom initiation (Arrigo et al., 1999), and hence there may be other factors that could have this effect as well. These results also provide

preliminary insight into the cellular restructuring processes that occurs upon cellular metamorphosis between life cycle stages in *P. antarctica,* as well as identifying numerous dynamic proteins of unknown function for future study. Finally, this study demonstrates the potential for the application of coupled transcriptomic and proteomic biomarker methodologies in studying the ecology of microbial interactions (including iron and B_{12}) and their influence on biogeochemistry in complex polar ecosystems such as the Ross Sea. The improved molecular and biochemical understanding of *P. antarctica* and its response to iron provided here are valuable in the design of future experiments and targeted metaproteomic assays to examine natural populations and to improve understanding of environmental factors that influence the annual bloom formation of an important coastal ecosystem of the Southern Ocean.

Acknowledgements

Support for this study was provided by an Investigator grant to M. Saito from the Gordon and Betty Moore Foundation (GBMF3782), and National Science Foundation grants NSF-PLR 0732665, OCE-1435056, OCE-1220484, the WHOI Coastal Ocean Institute, and a CINAR Postdoctoral Scholar Fellowship provided to S. Bender through the Woods Hole Oceanographic Institution. Support was provided to A. Allen through NSF awards ANT-0732822, ANT-1043671, OCE-1136477, and Gordon and Betty Moore Foundation grant GBMF3828. Additional support was provided to GRD through NSF award, OPP-0338097. We are indebted to Roberta Marinelli for her leadership and vision. We would also like to thank Emily Lorch for her assistance with culturing, Julie Rose for generously sharing a net tow field sample, and Andreas Krupke for manuscript feedback.

Author Contributions

S.J.B. contributed to data analysis and writing; D.M.M. conducted the laboratory experiments and (meta)proteome extractions; M.R.M. conducted the mass spectrometry sample preparation and processing; H.Z. conducted RNA extractions; J.P.M. and J.B. contributed to transcriptome sequence analyses; G.R.D. contributed to field measurements and manuscript edits; A.E.A. contributed to the experimental design, data analysis, and writing; M.A.S. contributed to experimental design, data analysis, and writing.

Financial Conflicts: The authors have no financial conflicts involving the research presented in this manuscript.

Table 1. Comparison of the total number of proteins and spectra measured in the proteome for each strain/treatment along with the number of differentially expressed transcripts between select conditions for *P. antarctica* strain 1871 and strain 1374. Proteins were identified using a 1% FDR (false discovery rate) threshold, a peptide threshold of 95%, and a minimum of 2 unique peptides per protein. The total number of peptide-to-spectrum matches (PSMs) is given for the total of each strain in parentheses. A threshold of 3 spectral counts in at least one of the treatments was selected for inclusion in the comparative analysis.

Strain	Treatment (Fe' pM)	Proteins Identified (PSMs)
1871	2	204
	41	214
	120	234
	740	226
	1200	251
	3900	258
	Total	536 (28887)
1374	2	581
	41	613
	120	600
	740	654
	1200	623
	3900	527
	Total	1085 (72087)

Table 2. Comparison of the total number of proteins, peptides, and spectra measured in the Ross Sea metaproteome net tow sample using three databases for peptide-to-spectrum matching (see Table S2). Results from 2-dimension and 1-dimension (1D in parentheses) analyses are shown.

Peptide-to-spectrum - matching database	Total proteins	Total Unique Peptides	Total spectra matched	Decoy FDR[+] Percent (peptide level)
1) Phaeocystis strains transcriptomes*	1545 (912)	3816 (2103)	14088 (8226)	0.6 (0.17)
2) Ross Sea metatranscriptome**	1474 (859)	3210 (1520)	10154 (4725)	0.1 (0.7)
3) Antarctic bacterial metagenomes***	102 (92)	237 (186)	530 (440)	3.6 (2.3)

[+]FDR refers to false discovery rate of a reversed peptide database

* Metaproteome annotated using the laboratory-generated transcriptomes for strain 1871 and strain 1374 (database #1).

** Metaproteome annotated using the metatranscriptome generated from sample split of original Ross Sea sample (database #2).

*** Bacterial metaproteome annotated using bacterial metagenomes from Delmont et al., 2014 (database #3).

Figure Legends

Figure 1. Micrographs of (a) a single *Phaeocystis* in cell culture, and (b) *Phaeocystis* colonies in a Ross Sea bloom.

Figure 2. Experimental workflow used in this study. Culture and field samples (top), transcriptome analyses (2nd row), sequence database construction for proteomics (3rd row), and proteomic and metaproteomic analyses (bottom row).

Figure 3. The effect of iron concentration on colony formation and cell physiology in two strains of *P. antarctica* – 1871 and 1374. Growth rates collected from acclimated culture stocks prior to the start of the experiments (*a*, strain 1871; *b*, strain 1374), calculated using relative fluorescence units from three transfers of acclimated cultures (error bars indicate SD, n=3). Accompanying gray bars represent growth rates calculated based on cell counts made during the course of the proteome-harvest experiments (n=1). (*c*) The number of *P. antarctica* 1871 free-living cells (gray bars) compared to cells associated with colonies (black bars) showed a shift to a majority of colonial cells when Fe' $\geq$ 740 pM. (*d*) Growth rate of Ross Sea diatom isolate *Chaetoceros* sp. strain RS-19 in the same media compositions (n=1), demonstrated a higher sensitivity to iron scarcity and a lack of iron contamination in the media. Cell size for strain 1871 (*e*; black circles) and strain 1374 (*f*; white circles); error bars represent SD of n=20 cell measurements per treatment.

Figure 4. Principle Component Analysis (PCA) of the measured proteomes for each iron condition for strain 1871 and strain 1374 and corresponding line graphs highlighting the proteins

694 driving the PCA separation (PCA analyses: ≥ 0.9 or ≤ -0.9). (*a* and *d*) Iron treatments (pM Fe')
695 are highlighted by color (2, black; 41, red; 120, orange; 740, green; 1200, purple; 3900, blue) and
696 large ellipses indicate confidence ellipses calculated using the R package, FactorMineR. Each
697 small, solid circle represents a technical replicate per iron treatment (n=3); colored, open squares
698 represent the mean of the iron treatment (empirical variance divided by the number of
699 observations). Proteins with Eigen values ≥ 0.9 or ≤ -0.9 are plotted in graphs b and c for strain
700 1871 and *e* and *f* for strain 1374 to highlight the subset of proteins driving the variance in
701 Dimension 1. Individual protein spectral counts normalized to total spectral counts for all
702 treatments for a given protein, written as "normalized relative protein abundance" are plotted on
703 the y-axis. The six iron treatments (pM Fe') are plotted from low to high (left to right) on the x-
704 axis.
705
706 **Figure 5.** Heatmaps highlighting the relative protein abundance for the six treatments for *P.*
707 *antarctica* strain 1871 (*a*) and strain 1374 (*b*). The darker green color indicates a greater relative
708 abundance compared to the purple treatments. The "shared abundance patterns" column features
709 a check-mark when a shared response to changes in iron availability between the relative protein
710 abundance and the transcript abundance was observed (for example, both transcripts and proteins
711 have a higher abundance under high iron compared to low iron growth [or] both transcripts and
712 proteins have a higher abundance under low iron compared to high iron growth). The "field
713 presence" column indicates whether or not that protein was detected in the field metaproteome
714 (annotated using Database #1). Protein annotations are based on KEGG, KOG, and PFam
715 descriptions. Annotations in red are associated with iron metabolism and those in blue, cell
716 adhesion/structure.

Figure 6. Examination of iron stress response proteins in *P. antarctica* strain 1871 (top) and 1374 (bottom). Relative protein abundance is shown as normalized spectral counts, where spectral counts have been normalized across experiment treatments for each strain, but not to the maximum of each protein as used in prior figures to allow comparison of abundance for similar isoforms. Error bars indicate the standard deviation of technical triplicate analyses.

Figure 7. Scatterplots of relative transcript abundance (y-axis) and relative protein abundance (x-axis) for *P. antarctica* strain 1871 (a) and strain 1374 (b) for a high iron treatment (3900 pM Fe') relative to a low iron treatment (41 pM Fe'). Gray circles represent instances where transcript abundance was not significantly different between conditions ($P \geq 0.99$). Quadrants where relative protein and transcript abundances agree (upper right, lower left) and disagree (upper left, lower right) are noted, as are select genes exhibiting the greatest relative protein abundance and/or transcript abundance under a given treatment.

Figure 8. Location of the metaproteome sample and pigment data from a Ross Sea *Phaeocystis* bloom net tow sample. (*a*) Station map of NBP06-01 (December 27, 2005 to January 23, 2006) and the metaproteome sample was taken on December 30th by net tow location (red circle). (*b*) 19'-hexanoyloxyfucoxanthin ("19'-Hex") pigment is associated with *Phaeocystis*, while (c) peridinin and (d) fucoxanthin pigments are typically associated with dinoflagellates and diatoms, respectively (although dinoflagellates living heterotrophically can be lacking in pigment). Comparisons of the spring and summer expeditions (NBP06-08 and NBP06-01, respectively),

observed a shift from being dominated by *P. antarctica* to being a mixture of *P. antarctica* and diatoms. See Smith et al., 2013 for further details (Smith et al., 2013).

Figure 9. (a) Venn diagram of the attribution of the5885 total unique peptides identified in the metaproteome sample to three DNA/RNA sequence databases (Supplementary Table 2). (*b*) Taxon group composition of genes identified by metatranscriptome analyses (combining Total RNA and PolyA RNA fractions). (c) Taxon group composition of proteins identified by the bacterial metagenomic database (Database #3). (d) Taxon group composition of proteins identified by metatranscriptome database (Database #2).

Figure 10. Putative biomarkers identified in the *Phaeocystis* metaproteome annotated using the field metatranscriptome (error bars represent SD of replicate samples; n=2; 1D dataset used). Green bars indicate putative "low iron" biomarkers; red bars indicate putative "high iron" biomarkers, and correspond to the life cycle stages observed (Fig. 3).

Fig. 11. Example spectra and chromatograms of fragment ions for two peptides corresponding to a *P. antarctica* flavodoxin identified from the Ross Sea metaproteome sample (peptide sequences found within Database #1, 1871, contig_31444_1_606_+, 1374 contig_202625_47_661_+; and, Database #2 contig_175060_39_653_+). Peptide fragmentation spectra are shown in (a) and (c) and example chromatograms of MS1 intensities as well as with +1 and +2 mass addition for isotopic distributions is shown (b) and (d), demonstrating the utility of these iron stress biomarkers in field samples.

762

763 **Figure 1.**
764
765
766

(a)

(b)

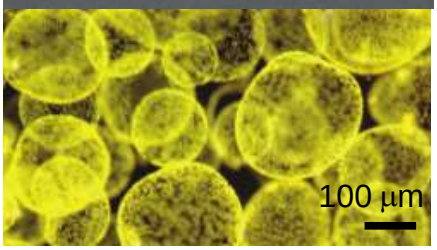


Figure 2.

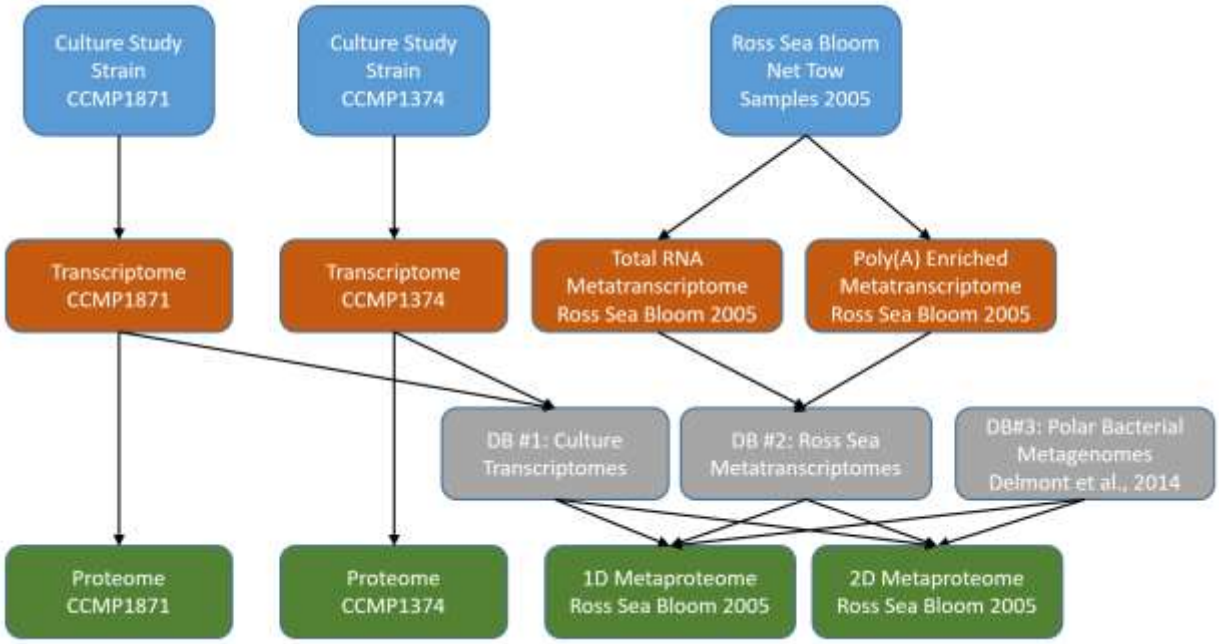
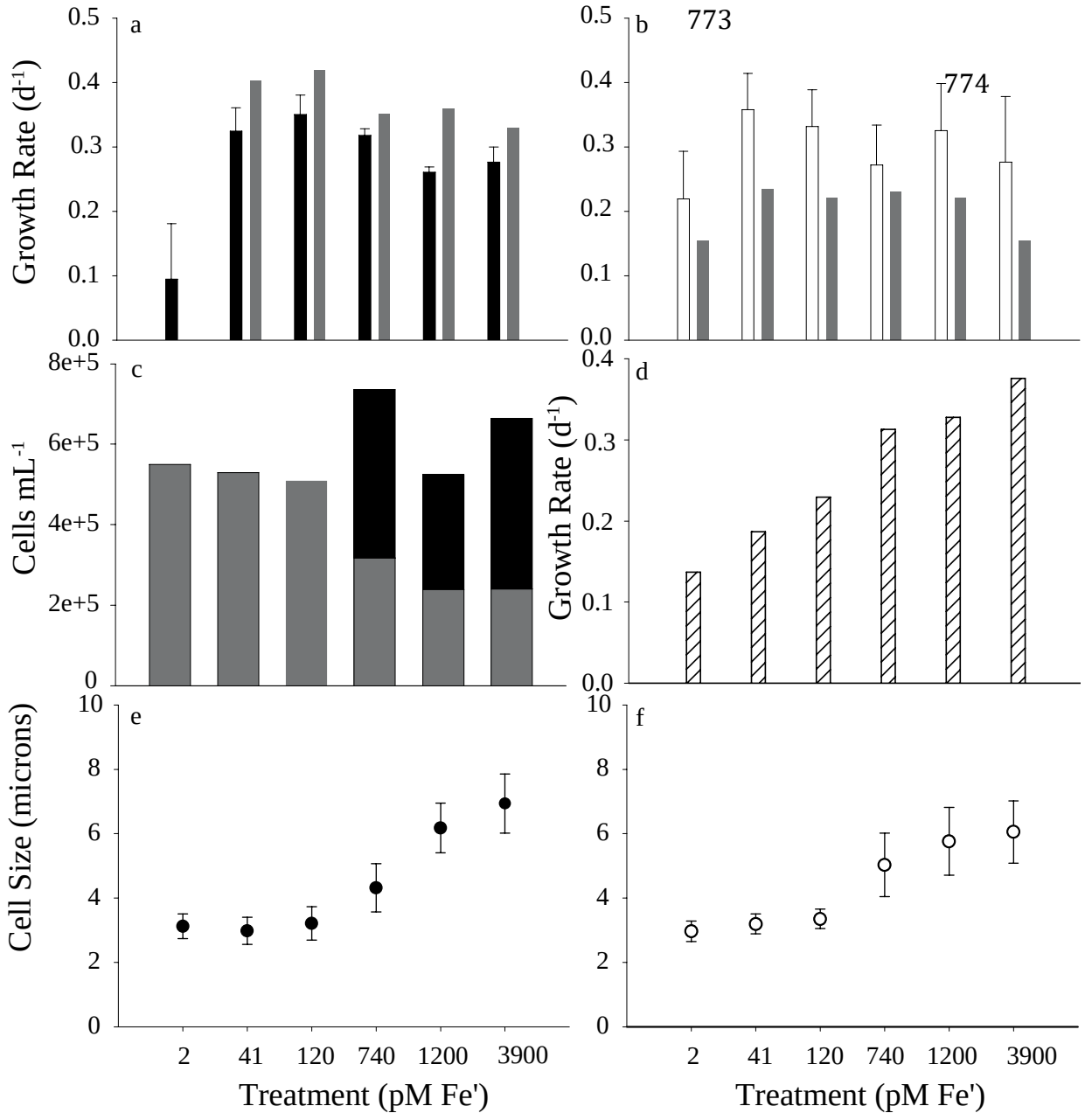

Culture Study Strain CCMP1871
Culture Study Strain CCMP1374
Ross Sea Bloom Net Tow Samples 2005
Transcriptome CCMP1871
Transcriptome CCMP1374
Total RNA Metatranscriptome Ross Sea Bloom 2005
Poly(A) Enriched Metatranscriptome Ross Sea Bloom 2005
DB #1: Culture Transcriptomes
DB #2: Ross Sea Metatranscriptomes
DB#3: Polar Bacterial Metagenomes Delmont et al., 2014
Proteome CCMP1871
Proteome CCMP1374
1D Metaproteome Ross Sea Bloom 2005
2D Metaproteome Ross Sea Bloom 2005

Figure 3.


a
b
c
d
e
f
Growth Rate (d-1)
Cells mL-1
Cell Size (microns)
0.5
0.4
0.3
0.2
0.1
0.0
8e+5
6e+5
4e+5
2e+5
0
10
8
6
4
2
2
41
120
740
1200
3900
Treatment (pM Fe')

775 **Figure 4.**

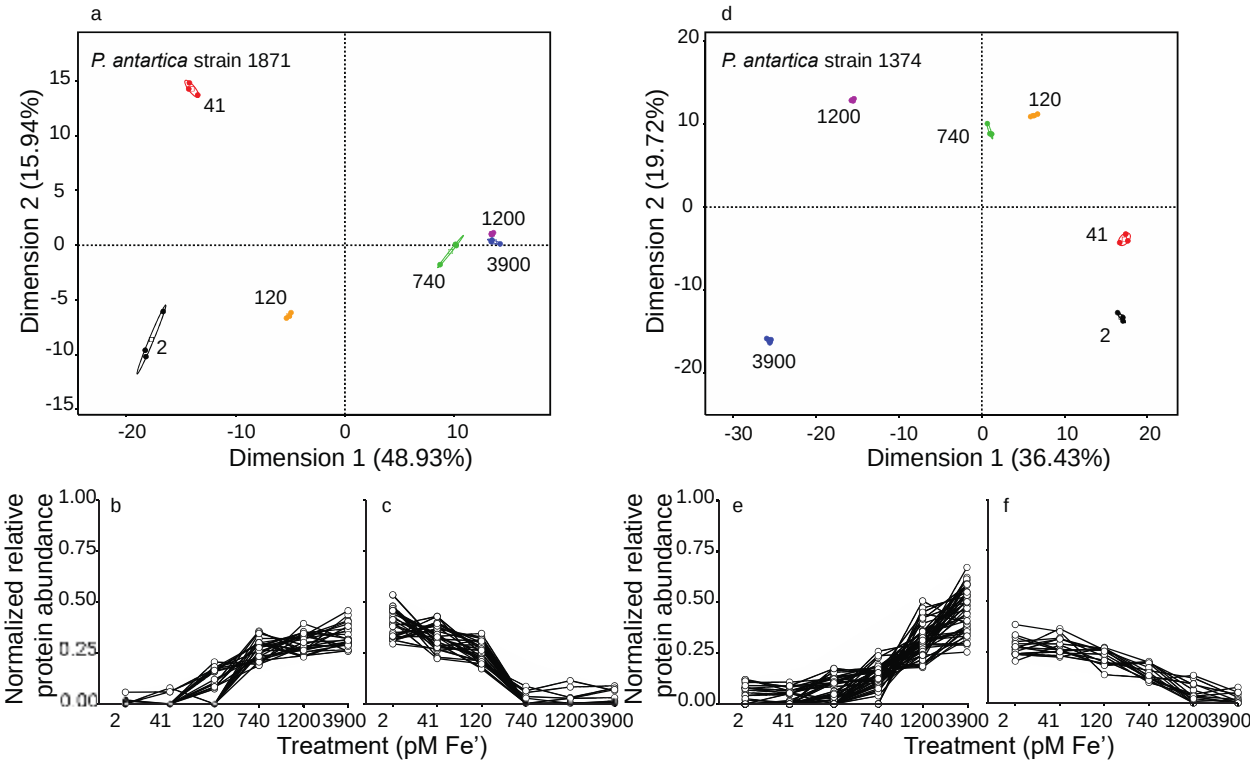
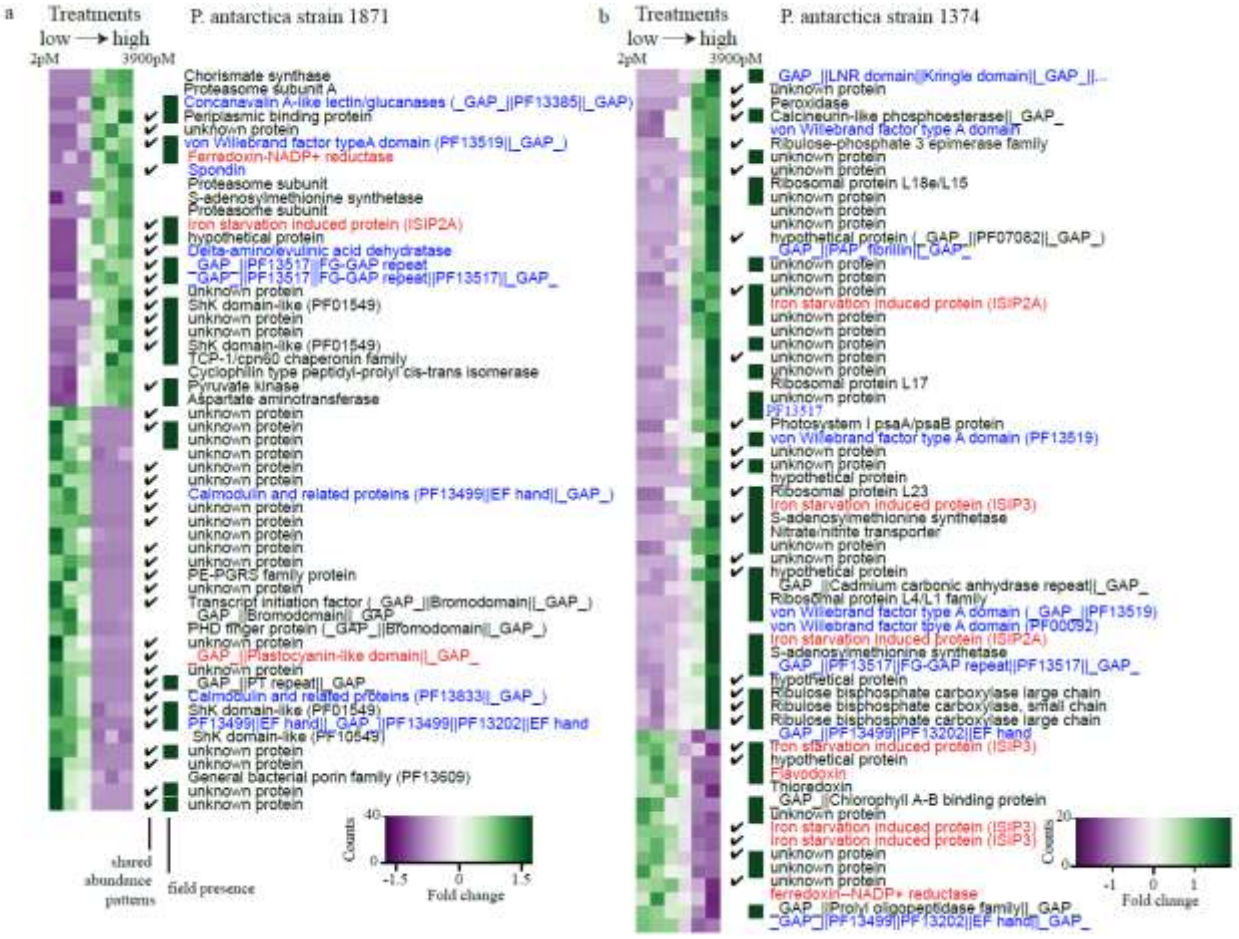
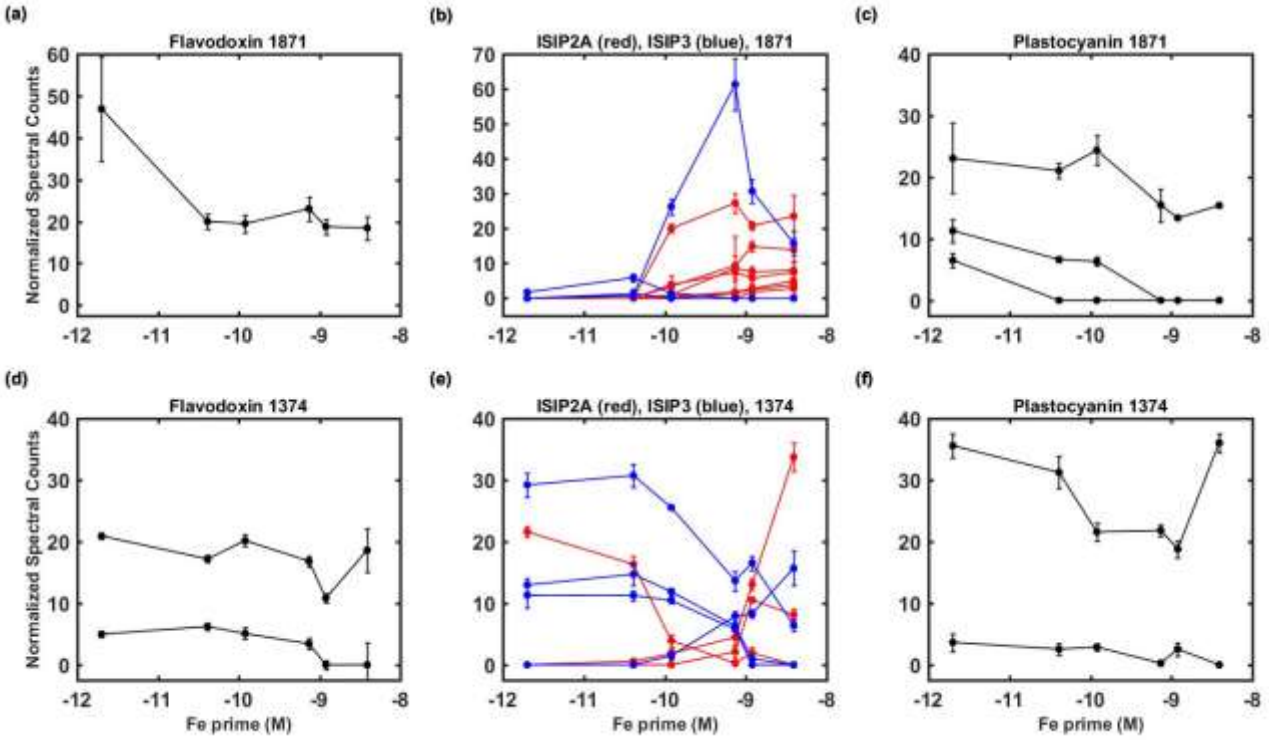

a
P. antartica strain 1871
41
1200
3900
740
120
2
Dimension 2 (15.94%)
Dimension 1 (48.93%)
d
P. antartica strain 1374
1200
120
740
41
2
3900
Dimension 2 (19.72%)
Dimension 1 (36.43%)
b
c
e
f
Normalized relative protein abundance
Treatment (pM Fe')
2 41 120 740 1200 3900

Figure 5.

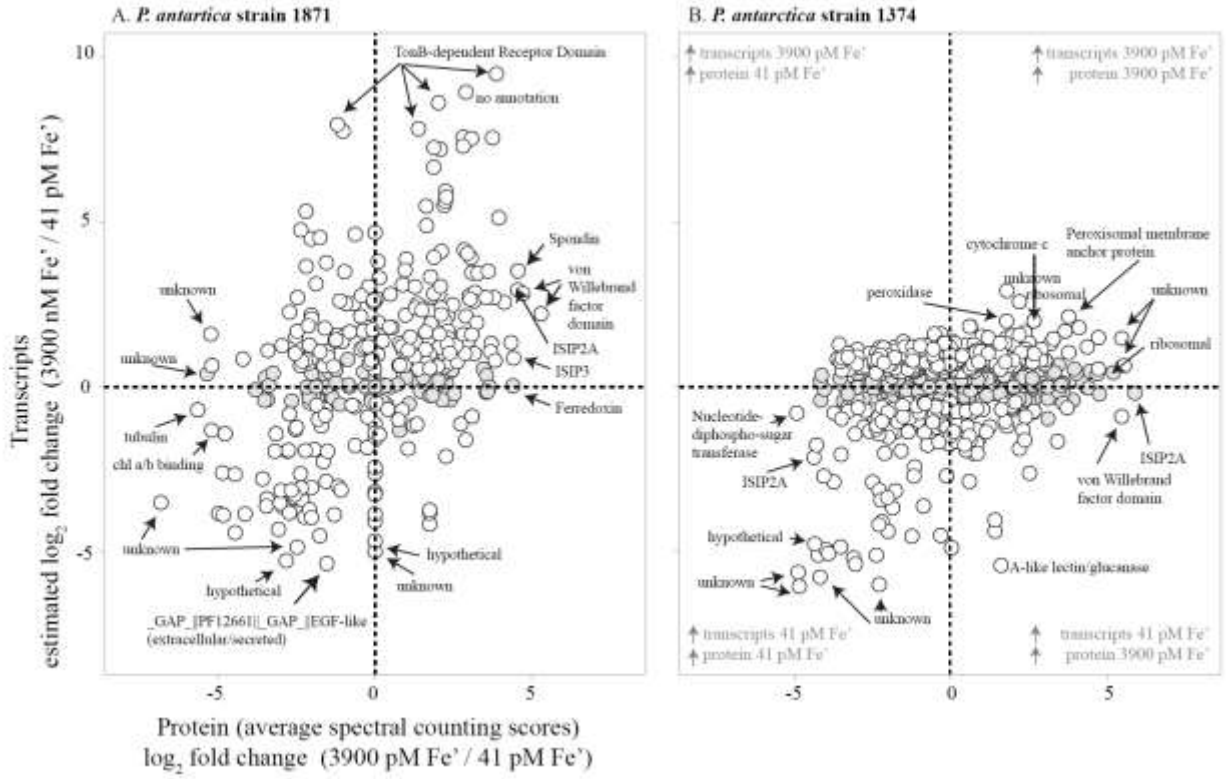

a
Treatments
low → high
2pM
3900pM
P. antarctica strain 1871
Chorismate synthase
Proteasome subunit A
Concanavalin A-like lectin/glucanases (_GAP_||PF13385||_GAP_)
Periplasmic binding protein
unknown protein
von Willebrand factor typeA domain (PF13519||_GAP_)
Ferredoxin-NADP+ reductase
Spondin
Proteasome subunit
S-adenosylmethionine synthetase
Proteasome subunit
Iron starvation induced protein (ISIP2A)
hypothetical protein
Delta-aminolevulinic acid dehydratase
GAP||PF13517||FG-GAP repeat
GAP||PF13517||FG-GAP repeat||PF13517||_GAP_
Unknown protein
ShK domain-like (PF01549)
unknown protein
unknown protein
ShK domain-like (PF01549)
TCP-1/cpn60 chaperonin family
Cyclophilin type peptidyl-prolyl cis-trans isomerase
Pyruvate kinase
Aspartate aminotransferase
unknown protein
unknown protein
unknown protein
unknown protein
unknown protein
unknown protein
Calmodulin and related proteins (PF13499||EF hand||_GAP_)
unknown protein
unknown protein
unknown protein
unknown protein
unknown protein
PE-PGRS family protein
unknown protein
Transcript initiation factor (_GAP_||Bromodomain||_GAP_)
GAP||Bromodomain||_GAP_
PHD finger protein (_GAP_||Bromodomain||_GAP_)
unknown protein
GAP||Plastocyanin-like domain||_GAP_
unknown protein
GAP||PT repeat||_GAP_
Calmodulin and related proteins (PF13833||_GAP_)
ShK domain-like (PF01549)
PF13499||EF hand||_GAP_||PF13499||PF13202||EF hand
ShK domain-like (PF10549)
unknown protein
unknown protein
General bacterial porin family (PF13609)
unknown protein
unknown protein
shared abundance patterns
field presence
Counts
40
0
-1.5
0
1.5
Fold change
b
Treatments
low → high
2pM
3900pM
P. antarctica strain 1374
GAP||LNR domain||Kringle domain||_GAP_||...
unknown protein
Peroxidase
Calcineurin-like phosphoesterase||_GAP_
von Willebrand factor type A domain
Ribulose-phosphate 3 epimerase family
unknown protein
unknown protein
Ribosomal protein L18e/L15
unknown protein
unknown protein
unknown protein
hypothetical protein (_GAP_||PF07082||_GAP_)
GAP||PAP_fibrillin||_GAP_
unknown protein
unknown protein
unknown protein
Iron starvation induced protein (ISIP2A)
unknown protein
unknown protein
unknown protein
unknown protein
unknown protein
Ribosomal protein L17
unknown protein
PF13517
Photosystem I psaA/psaB protein
von Willebrand factor type A domain (PF13519)
unknown protein
unknown protein
hypothetical protein
Ribosomal protein L23
Iron starvation induced protein (ISIP3)
S-adenosylmethionine synthetase
Nitrate/nitrite transporter
unknown protein
unknown protein
hypothetical protein
GAP||Cadmium carbonic anhydrase repeat||_GAP_
Ribosomal protein L4/L1 family
von Willebrand factor type A domain (_GAP_||PF13519)
von Willebrand factor type A domain (PF00092)
Iron starvation induced protein (ISIP2A)
S-adenosylmethionine synthetase
GAP||PF13517||FG-GAP repeat||PF13517||_GAP_
hypothetical protein
Ribulose bisphosphate carboxylase large chain
Ribulose bisphosphate carboxylase, small chain
Ribulose bisphosphate carboxylase large chain
GAP||PF13499||PF13202||EF hand
Iron starvation induced protein (ISIP3)
hypothetical protein
Flavodoxin
Thioredoxin
GAP||Chlorophyll A-B binding protein
Unknown protein
Iron starvation induced protein (ISIP3)
Iron starvation induced protein (ISIP3)
unknown protein
unknown protein
unknown protein
ferredoxin--NADP+ reductase
GAP||Prolyl oligopeptidase family||_GAP_
GAP||PF13499||PF13202||EF hand||_GAP_
Counts
20
0
-1
0
1
Fold change

780 **Figure 6.**
781



782

Figure 7.

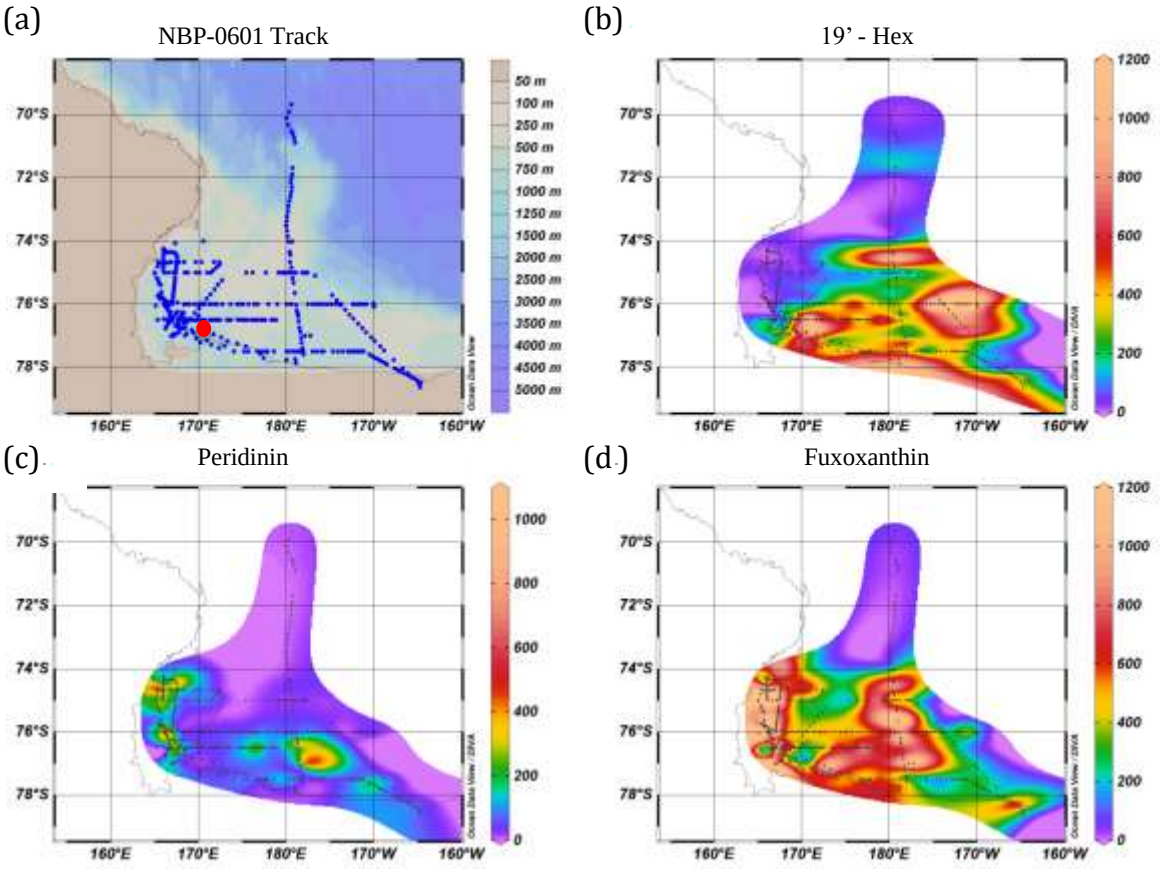

A. P. antartica strain 1871
B. P. antarctica strain 1374
TonB-dependent Receptor Domain
no annotation
Spondin
von Willebrand factor domain
ISIP2A
ISIP3
Ferredoxin
unknown
unknown
tubulin
chl a/b binding
unknown
hypothetical
hypothetical
unknown
GAP[PF12661]|_GAP_[EGF-like (extracellular/secreted)
transcripts 3900 pM Fe'
protein 41 pM Fe'
transcripts 3900 pM Fe'
protein 3900 pM Fe'
cytochrome c
Peroxisomal membrane anchor protein
unknown ribosomal
peroxidase
unknown
ribosomal
Nucleotide-diphospho-sugar transferase
ISIP2A
ISIP2A
von Willebrand factor domain
hypothetical
unknown
unknown
A-like lectin/glucanase
transcripts 41 pM Fe'
protein 41 pM Fe'
transcripts 41 pM Fe'
protein 3900 pM Fe'
Transcripts
estimated log2 fold change (3900 nM Fe' / 41 pM Fe')
Protein (average spectral counting scores)
log2 fold change (3900 pM Fe' / 41 pM Fe')

788 **Figure 8.**

789

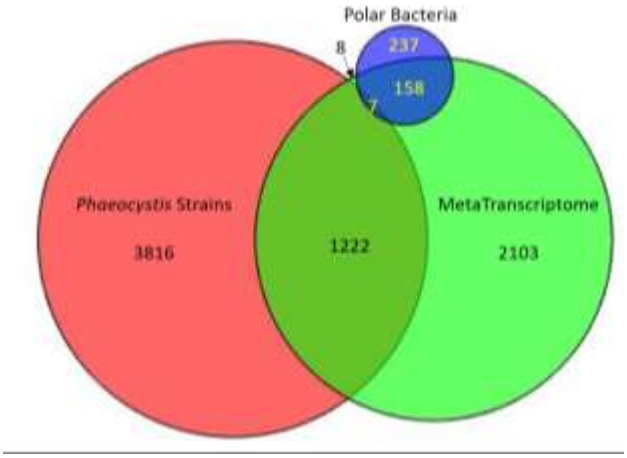
790

791
792
793

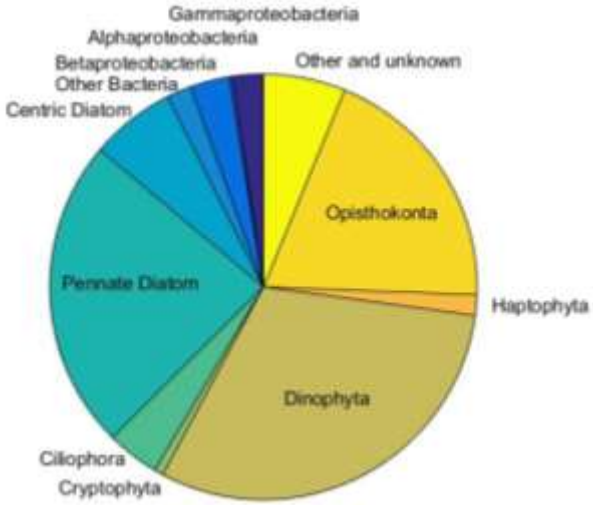


794 **Figure 9.**
795

(a) Unique Peptides by Database

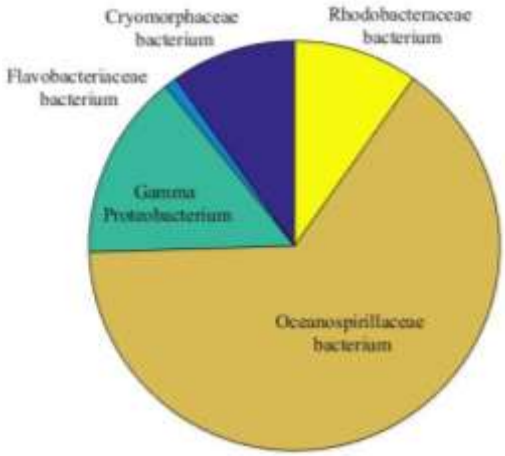


(b) Genes Identified by Metatranscriptome

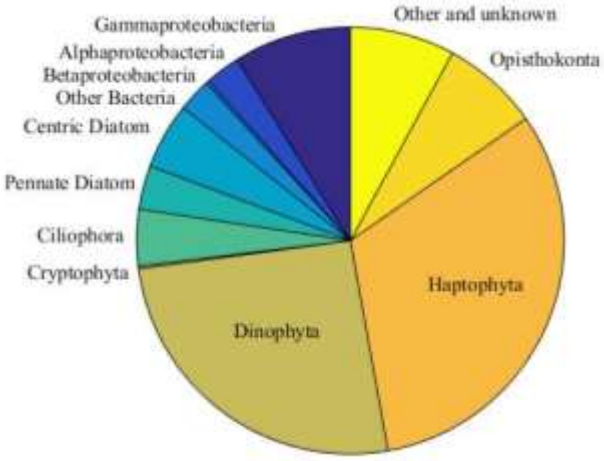


796

(c) Proteins by Bacterial Metagenome



(d) Proteins Identified by Metatranscriptome



797 **Figure 10.**

798

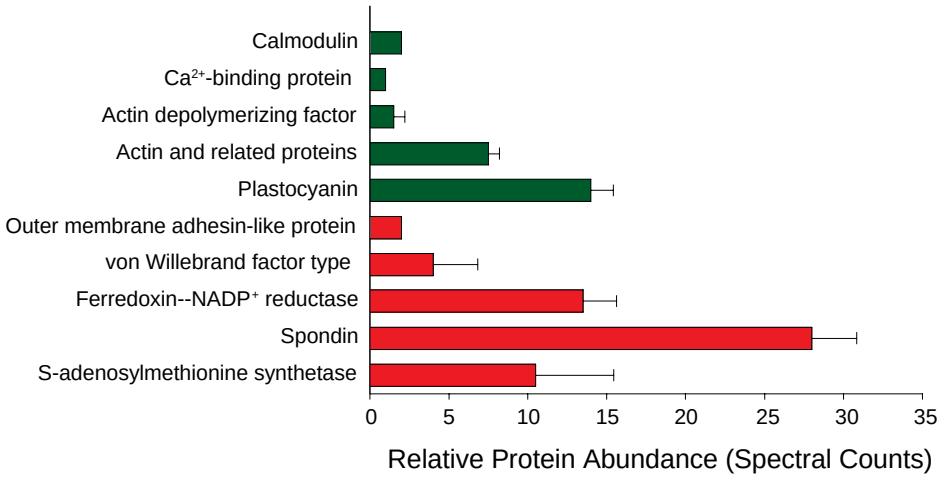
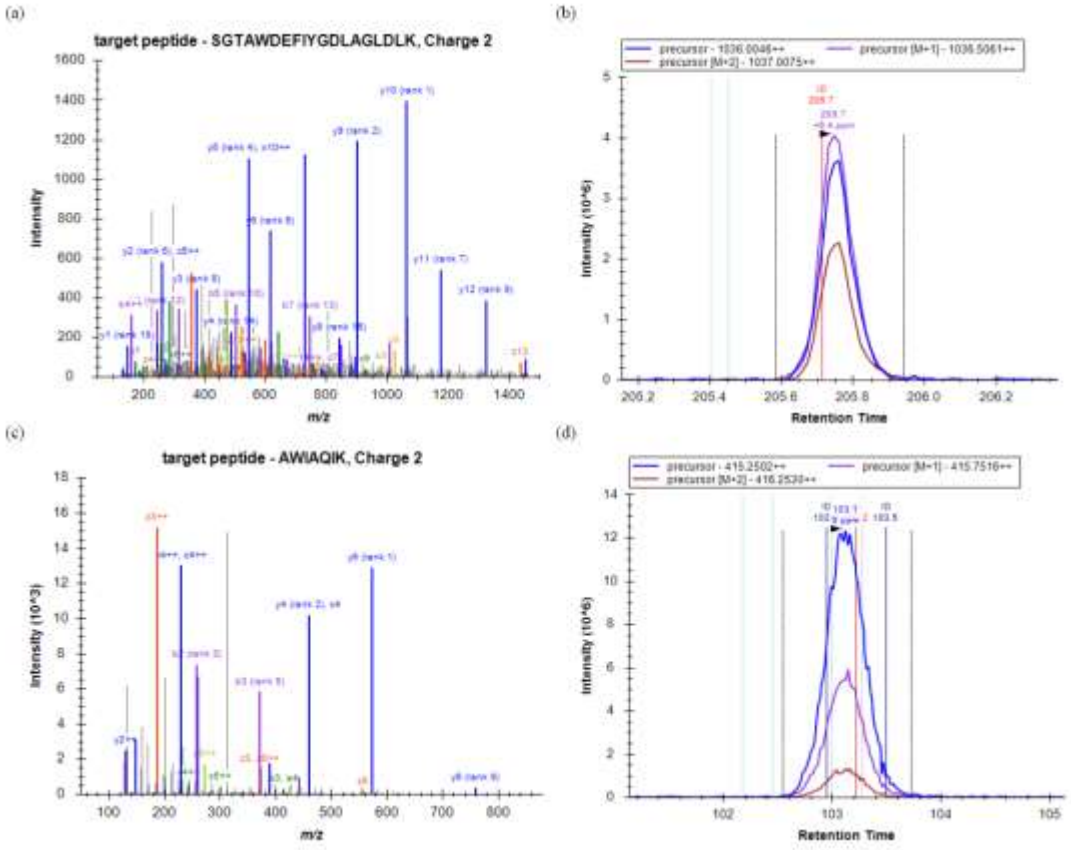


Figure 11.


(a)
target peptide - SGTAWDEFIYGDLAGLDLK, Charge 2
Intensity
m/z
(b)
precursor - 1036.0046++
precursor [M+1] - 1036.5061++
precursor [M+2] - 1037.0075++
Intensity (10^5)
Retention Time
(c)
target peptide - AWIAQIK, Charge 2
Intensity (10^3)
m/z
(d)
precursor - 415.2502++
precursor [M+1] - 415.7516++
precursor [M+2] - 416.2530++
Intensity (10^6)
Retention Time

References

Abedin, M. and King, N.: Diverse and evolutionary paths to cell adhesion, Trends in Cell Biology, 20(12), 734–742, 2010.

Alderkamp, A. C., Buma, A. G. J. and Van Rijssel, M.: The carbohydrates of *Phaeocystis* and their degradation in the microbial food web, Biogeochemistry, 83, 99–118, 2007.

Allen, A. E., LaRoche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P. J., Finazzi, G., Fernie, A. R. and Bowler, C.: Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation, vol. 105, pp. 10438–10443. 2008.

Arrigo, K. R., R, D. G., Dunbar, R. B., Robinson, D. H., vanWoert, M. L., Worthen, D. L. and Lizotte, M. P.: Phytoplankton taxonomic variability in nutrient utilization and primary production in the Ross Sea, Journal of Geophysical Research, 105(C4), 8827–8846, doi:10.1029/1998JC000289, 2000.

Arrigo, K. R., Robinson, D. H., Worthen, D. L., Dunbar, R. B., R, D. G., vanWoert, M. L. and Lizotte, M. P.: Phytoplankton community structure and the drawdown of nutrients and CO_2 in the Southern Ocean, Science, 283, 365–367, doi:10.1126/science.283.5400.365, 1999.

Arrigo, K. R., Worthen, D., Schnell, A. and Lizotte, M. P.: Primary production in Southern Ocean waters, Journal of Geophysical Research, 103(C8), 15587–15600, 1998.

Bertrand, E. M., McCrow, J. P., Moustafa, A., Zheng, H., McQuaid, J. B., Delmont, T. O., Post, A. F., Sipler, R. E., Spackeen, J. L., Xu, K., Bronk, D. A., Hutchins, D. A. and Allen, A. E.: Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge, Proceedings of the National Academy of Sciences, 112(32), 9938–9943, doi:10.1073/pnas.1501615112, 2015.

Bertrand, E. M., Moran, D. M., McIlvin, M. R., Hoffman, J. M., Allen, A. E. and Saito, M. A.: Methionine synthase interreplacement in diatom cultures and communities: Implications for the persistence of B_{12} use by eukaryotic phytoplankton, Limnology and Oceanography, 58(4), 1431–1450, doi:10.4319/lo.2013.58.4.1431, 2013.

Bertrand, E. M., Saito, M. A., Lee, P. A., Dunbar, R. B., Sedwick, P. N. and R, D. G.: Iron limitation of a springtime bacterial and phytoplankton community in the Ross Sea: Implications for Vitamin B_{12} nutrition, Frontiers in Microbiology, 2, 1–12, doi:10.3389/fmicb.2011.00160, 2011.

Bertrand, E. M., Saito, M. A., Rose, J. M., Riesselman, C. R., Lohan, M. C., Noble, A. E., Lee, P. A. and R, D. G.: Vitamin B_{12} and iron colimitation of phytoplankton growth in the Ross Sea, Limnology and Oceanography, 52(3), 1079–1093, 2007.

Boye, M., van den Berg, C., de Jong, J., Leach, H., Croot, P. and de Baar, H. J. W.: Organic complexation of iron in the Southern Ocean, Deep-Sea Research Part I, 48(6), 1477–1497, 2001.

Chiovitti, A., Bacic, A., Burke, J. and Wetherbee, R.: Heterogeneous xylose-rich glycans are

associated with extracellular glycoproteins from the biofouling diatom *Craspedostauros australis* (Bacillariophyceae), European Journal of Phycology, 38(4), 351–360, doi:10.1080/09670260310001612637, 2003.

Coale, K. H., Wang, X., Tanner, S. J. and Johnson, K. S.: Phytoplankton growth and biological response to iron and zinc addition in the Ross Sea and Antarctic Circumpolar Current along 170°W, Deep-Sea Research Part II, 50, 635–653, 2003.

Delmont, T. O., Hammar, K. M., Ducklow, H. W., Yager, P. L. and Post, A. F.: *Phaeocystis antarctica* blooms strongly influence bacterial community structures in the Amundsen Sea polynya, Frontiers in Microbiology, 5, 646, 2014.

DiTullio, G. R., Grebmeier, J. M., Arrigo, K. R., Lizotte, M. P., Robinson, D. H., Leventer, A., Barry, J. P., vanWoert, M. L. and dunbar, R. B.: Rapid and early export of *Phaeocystis antarctica* blooms in the Ross Sea, Antarctica, Nature, 404, 595–598, doi:10.1038/35007061, 2000.

Drake, J. L., Mass, T., Haramaty, L., Zelzion, E., Bhattacharya, D. and Falkowski, P. G.: Proteomic analysis of skeletal organic matrix from the stony coral *Stylophora pistillata*, Proceedings of the National Academy of Sciences, 110(10), 3788–3793, 2013.

Ducklow, H. W., Baker, K., Martinson, D. G., Quetin, L. B., Ross, R. M., Smith, R. C., Stammerjohn, S. E., Vernet, M. and Fraser, W.: Marine pelagic ecosystems: the West Antarctic Peninsula, Philosophical Transactions of the Royal Society B, 362, 67–94, doi:10.1098/rstb.2006.1955, 2007.

Dunbar, R. B., Leventer, A. R. and Mucciarone, D. A.: Water column sediment fluxes in the Ross Sea, Antarctica: Atmospheric and sea ice forcing, Journal of Geophysical Research, 103(C13), 30741–30759, doi:10.1029/1998JC900001, 1998.

Dyhrman, S. T.: Identifying reference genes with stable expression from high throughput sequence data,, 1–10, doi:10.3389/fmicb.2012.00385/abstract, 2012.

Eng, J. K., McCormack, A. L. and Yates, J. R.: An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, J Am Soc Mass Spectrom, 5(11), 976–989, doi:10.1016/1044-0305(94)80016-2, 1994.

Ewenstein, B. M.: Von Willebrand's disease, Annu. Rev. Med., 48(1), 525–542, doi:10.1146/annurev.med.48.1.525, 1997.

Feng, Y., Hare, C. E., Rose, J. M., Handy, S. M., DiTullio, G. R., Lee, P. A., Smith, W. O., Jr, Peloquin, J., Tozzi, S., Sun, J., Zhang, Y., Dunbar, R. B., Long, M. C., Sohst, B., Lohan, M. and Hutchins, D. A.: Interactive effects of iron, irradiance and CO_2 on Ross Sea phytoplankton, Deep-Sea Research Part I, 57, 368–383, doi:10.1016/j.dsr.2009.10.013, 2010.

Gäbler-Schwarz and L. K. Medlin and F. Leese. A puzzle with many pieces: the genetic structure and diversity of *Phaeocystis antarctica Karsten* (Prymnesiophyta). European Journal of Phycology, 50 (1): 112-124. 2015. https://doi.org/10.1080/09670262.2014.998295

876 Garcia, N. S., Sedwick, P. N. and DiTullio., G.R.: Influence of irradiance and iron on the growth
877 of colonial *Phaeocystis antarctica*: implications for seasonal bloom dynamics in the Ross Sea,
878 Antarctica, Aquatic Microbial Ecology, 57, 203–220, 2009.

879 Hallmann, A.: Extracellular matrix and sex-inducing pheromone in *Volvox*, in International
880 Review of Cytology, vol. 227, pp. 131–182, International Review of Cytology. 2003.

881 Hamm, C. E.: Architecture, ecology and biogeochemistry of *Phaeocystis* colonies, Journal of Sea
882 Research, 43, 307–315, 2000.

883 Hamm, C. E., Simson, D. A., Merkel, R. and Smetacek, V.: Colonies of *Phaeocystis globosa* are
884 protected by a thin but tough skin, Marine Ecology Progress Series, 187, 101–111, 1999.

885 Hayward, D. C., Hetherington, S., Behm, C. A., Grasso, L. C., Forêt, S., Miller, D. J. and Ball, E.
886 E.: Differential gene expression at coral settlement and metamorphosis - A subtractive
887 hybridization study, PLoS ONE, 6(10), e26411, doi:10.1371/journal.pone.0026411, 2011.

888 Jacobsen, A., Larsen, A., Martínez-Martínez, J., Verity, P. G. and Frischer, M. E.: Susceptibility
889 of colonies and colonial cells of *Phaeocystis pouchetii* (Haptophyta) to viral infection, Aquatic
890 Microbial Ecology, 48, 105–112, 2007.

891 King, N., Hittinger, C. T. and Carroll, S. B.: Evolution of key cell signaling and adhesion protein
892 families predates animal origins, Science, 301(5631), 361–363, doi:10.1126/science.1083853,
893 2003.

894 Kröger, N., Bergsdorf, C. and Sumper, M.: A new calcium binding glycoprotein family
895 constitutes a major diatom cell wall component, The EMBO Journal, 13(19), 4676–4683, 1994.

896 Lagerheim, G.: Ueber Phaeocystis poucheti (Har.) Lagerh., eine Plankton-Flagellate, Oeivers af
897 Vet Akad Foerhandl, 4, 277–288, 1896.

898 Lange, M., Chen, Y.-Q. and Medlin, L. K.: Molecular genetic delineation of *Phaeocystis* species
899 (Prymnesiophyceae) using coding and non-coding regions of nuclear and plastid genomes,
900 European Journal of Phycology, 37(1), 77–92, doi:10.1017/S0967026201003481, 2002.

901 Lê, S., Josse, J. and Husson, F.: FactoMineR: an R package for multivariate analysis, Journal of
902 Statistical Software, 25(1), 2008.

903 Long, J. D., Smalley, G. W., Barsby, T., Anderson, J. T. and Hay, M. E.: Chemical cues induce
904 consumer-specific defenses in a bloom-forming marine phytoplankton, Proceedings of the
905 National Academy of Sciences, 104(25), 10512–10517 [online] Available from:
906 http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4224216, 2007.

907 Lovenduski, N. S., Gruber, N. and Doney, S. C.: Toward a mechanistic understanding of the
908 decadal trends in the Southern Ocean carbon sink, Global Biogeochemical Cycles, 22(GB3016),
909 doi:10.1029/2007GB003139, 2008.

910 Lubbers, G., Gieskes, W., Del Castilho, P., Salomons, W., and Bril, J.: Manganese accumulation

911 in the high pH microenvironment of *Phaeocystis sp.*(Haptophyceae) colonies from the North
912 Sea, Marine Ecology Progress Series, 1990. 285-293, 1990.

913 Luxem, K. E., Ellwood, M. J., and Strzepek, R. F.: Intraspecific variability in *Phaeocystis*
914 *antarctica*'s response to iron and light stress, PLoS ONE, 12, e0179751, 2017.

915 Martin, J. H., Fitzwater, S. E. and Gordon, R. M.: Iron deficiency limits phytoplankton growth in
916 Antarctic waters, Global Biogeochemical Cycles, 4(1), 5–12, doi:10.1029/GB004i001p00005,
917 1990.

918 Matrai, P. A., Vernet, M., Hood, R., Jennings, A., Brody, E. and Saemundsdottir, S.: Light-
919 dependence of carbon and sulfur production by polar clones of the genus *Phaeocystis*, Marine
920 Biology, 124, 157–167, 1995.

921 Maucher, J. M. and G. R. DiTullio (2003). Flavodoxin as a Diagnostic Indicator of Chronic Fe-
922 Limitation in the Ross Sea and New Zealand Sector of the Southern Ocean. Biogeochemistry in
923 the Ross Sea. G. R. DiTullio and R. B. Dunbar. Washington DC, AGU**:** 209-220.

924 Michel, G., Tonon, T., Scornet, D., Cock, J. M. and Kloareg, B.: The cell wall polysaccharide
925 metabolism of the brown alga *Ectocarpus siliculosus*. Insights into the evolution of extracellular
926 matrix polysaccharides in Eukaryotes, New Phytologist, 188(1), 82–97, 2010.

927 Morris, R. M., Nunn, B. L., Frazar, C., Goodlett, D. R., Ting, Y. S. and Rocap, G.: Comparative
928 metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy
929 transduction, The ISME Journal, 4(5), 673–685, doi:10.1038/ismej.2010.4, 2010.

930 Morrissey, J., Sutak, R., Paz-Yepes, J., Tanaka, A., Moustafa, A., Veluchamy, A., Thomas, Y.,
931 Botebol, H., Bouget, F.-Y., McQuaid, J. B., Tirichine, L., Allen, A. E., Lesuisse, E. and Bowler,
932 C.: A novel protein, ubiquitous in marine phytoplankton, concentrates iron at the cell surface and
933 facilitates uptake, Current Biology, 25, 364–371, doi:10.1016/j.cub.2014.12.004, 2015.

934 Murray, A. E. and Grzymski, J. J.: Diversity and genomics of Antarctic marine micro-organisms,
935 Philosophical Transactions of the Royal Society B: Biological Sciences, 362(1488), 2259–2271,
936 doi:10.1098/rstb.2006.1944, 2007.

937 Noble, A. E., Moran, D. M., Allen, A. E. and Saito, M. A.: Dissolved and particulate trace metal
938 micronutrients under the McMurdo Sound seasonal sea ice: basal sea ice communities as a
939 capacitor for iron, Frontiers in Chemistry, 1(25), 1–18, doi:10.3389/fchem.2013.00025, 2013.

940 Peers, G. and Price, N. M.: Copper-containing plastocyanin used for electron transport by an
941 oceanic diatom, Nature, 441(7091), 341–344, doi:10.1038/nature04630, 2006.

942 Podell, S. and Gaasterland, T.: DarkHorse: a method for genome-wide prediction of horizontal
943 gene transfer, Genome Biology, 8, R16, doi:10.1186/gb-2007-8-2-r16, 2007.

944 Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., Blake, R. C., Shah,
945 M., Hettich, R. and Banfield, J.: Community proteomics of a natural microbial biofilm, Science,
946 308(5730), 1915–1920, doi:10.1126/science, 2005.

947 Rho, M., Tang, H. and Ye, Y.: FragGeneScan: predicting genes in short and error-prone reads,
948 Nucleic Acids Research, 38(20), e191–e191, doi:10.1093/nar/gkq747, 2010.

949 Riegman, R. and van Boekel, W.: The ecophysiology of *Phaeocystis globosa*: a review, Journal
950 of Sea Research, 35(4), 235–242, 1996.

951 Riegman, R., Noordeloos, A. A. M. and Cadee, G. C.: *Phaeocystis* blooms and eutrophication of
952 the continental coastal zones of the North Sea, Marine Biology, 112, 479–484,
953 doi:10.1007/BF00356293, 1992.

954 Roche, J. L., Boyd, P. W., McKay, R. M. L. and Geider, R. J.: Flavodoxin as an *in situ* marker
955 for iron stress in phytoplankton, Nature, 382, 802–805, doi:10.1038/382802a0, 1996.

956 Rousseau, V., Vaulot, D., Casotti, R., Cariou, V., Lenz, J., Gunkel, J., and Baumann, M. The
957 Life Cycle of *Phaeocystis* (Prymnesiophyceae): Evidence and Hypotheses. Journal of Marine
958 Systems. 5. 23-39. 1994. 10.1016/0924-7963(94)90014-0.

959 Rousseau, V., Chrétiennot-Dinet, M.-J., Jacobsen, A., Verity, P. G. and Whipple, S.: The life
960 cycle of *Phaeocystis*: state of knowledge and presumptive role in ecology, Biogeochemistry, 83,
961 29–47, doi:10.1007/s10533-007-9085-3, 2007.

962 Rousseau, V., Mathot, S. and Lancelot, C.: Calculating carbon biomass of *Phaeocystis sp.* from
963 microscopic observations, Marine Biology, 107, 305–314, 1990.

964 Saito, M. A., Dorsk, A., Post, A. F., McIlvin, M. R., Rappé, M. S., R, D. G. and Moran, D. M.:
965 Needles in the blue sea: Sub-species specificity in targeted protein biomarker analyses within the
966 vast oceanic microbial metaproteome, Proteomics, 15(20), 3521–3531,
967 doi:10.1002/pmic.201400630, 2015.

968 Saito, M. A., Goepfert, T. J., Noble, A. E., Bertrand, E. M., Sedwick, P. N. and DiTullio, G. R.:
969 A seasonal study of dissolved cobalt in the Ross Sea, Antarctica: micronutrient behavior,
970 absence of scavenging, and relationships with Zn, Cd, and P, Biogeosciences, 7, 4059–4082,
971 doi:10.5194/bg-7-4059-2010, 2010.

972 Saito, M. A., McIlvin, M. R., Moran, D. M., Goepfert, T. J., R, D. G., Post, A. F. and Lamborg,
973 C. H.: Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein
974 biomarkers, Science, 345(6201), 1173–1177, 2014.

975 Sarmiento, J. L., Hughes, T. M. C., Stouffer, R. J. and Manabe, S.: Simulated response of the
976 ocean carbon cycle to anthropogenic climate warming, Nature, 393, 245–249,
977 doi:10.1038/30455, 1998.

978 Schoemann, V., Becquevort, S., Stefels, J., Rousseau, V. and Lancelot, C.: *Phaeocystis* blooms
979 in the global ocean and their controlling mechanisms: a review, Journal of Sea Research, 53, 43–
980 66, 2005.

981 Schoemann, V., Wollast, R., Chou, L. and Lancelot, C.: Effects of photosynthesis on the
982 accumulation of Mn and Fe by *Phaeocystis* colonies, Limnology and Oceanography, 46(5),

983 1065–1076, 2001.

984 Sedwick, P. N. and DiTullio, G. R.: Regulation of algal blooms in Antarctic shelf waters by the
985 release of iron from melting sea ice, Geophys. Res. Lett., 24(20), 2515–2518,
986 doi:10.1029/97GL02596, 1997.

987 Sedwick, P. N., DiTullio, G. R. and Mackey, D. J.: Iron and manganese in the Ross Sea,
988 Antarctica: Seasonal iron limitation in Antarctic shelf waters, Journal of Geophysical Research,
989 105(C5), 11321–11336, doi:10.1029/2000JC000256, 2000.

990 Sedwick, P. N., Garcia, N. S., Riseman, S. F., Marsay, C. M. and DiTullio, G. R.: Evidence for
991 high iron requirements of colonial *Phaeocystis antarctica* at low irradiance, Biogeochemistry,
992 83(1-3), 83–97, doi:10.1007/s10533-007-9081-7, 2007.

993 Sedwick, P. N., Marsay, C. M., Sohst, B. M., Aguilar Islas, A. M., Lohan, M. C., Long, M. C.,
994 Arrigo, K. R., Dunbar, R. B., Saito, M. A., Smith, W. O. and DiTullio, G. R.: Early season
995 depletion of dissolved iron in the Ross Sea polynya: Implications for iron dynamics on the
996 Antarctic continental shelf, Journal of Geophysical Research, 116(C12019), 2011.

997 Smith, W. O., Jr, Codispoti, L. A., Nelson, D. M., Manley, T., Buskey, E. J., Niebauer, H. J. and
998 Cota, G. F.: Importance of *Phaeocystis* blooms in the high-latitude ocean carbon cycle, Nature,
999 352, 514–516, 1991.

1000 Smith, W. O., Jr, Dennett, M. R., Mathot, S. and Caron, D. A.: The temporal dynamics of the
1001 flagellated and colonial stages of *Phaeocystis antarctica* in the Ross Sea, Deep-Sea Research
1002 Part II, 50, 605–617, 2003.

1003 Smith, W. O., Tozzi, S., Long, M. C., Sedwick, P. N., Peloquin, J. A., Dunbar, R. B., Hutchins,
1004 D. A., Kolber, Z. and R, D. G.: Spatial and temporal variations in variable fluorescence in the
1005 Ross Sea (Antarctica): Oceanographic correlates and bloom dynamics, Deep Sea Research I, 79,
1006 141–155, 2013.

1007 Solomon, C. M., Lessard, E. J., Keil, R. G. and Foy, M. S.: Characterization of extracellular
1008 polymers of *Phaeocystis globosa* and *P. antarctica*, Marine Ecology Progress Series, 250, 81–
1009 89, 2003.

1010 Sowell, S. M., Wilhelm, L. J., Norbeck, A. D., Lipton, M. S., Nicora, C. D., Barofsky, D. F., H,
1011 C., Smith, R. D. and Giovanonni, S. J.: Transport functions dominate the SAR11 metaproteome
1012 at low-nutrient extremes in the Sargasso Sea, The ISME Journal, 3(1), 93–105,
1013 doi:10.1038/ismej.2008.83, 2008.

1014 Stingl, U., Desiderio, R. A., Cho, J. C., Vergin, K. L. and Giovannoni, S. J.: The SAR92 Clade:
1015 an Abundant Coastal Clade of Culturable Marine Bacteria Possessing Proteorhodopsin, Applied
1016 and Environmental Microbiology, 73(7), 2290–2296, doi:10.1128/AEM.02559-06, 2007.

1017 Strzepek, R. F., Maldonado, M. T., Hunter, K. A., Frew, R. D. and Boyd, P. W.: Adaptive
1018 strategies by Southern Ocean phytoplankton to lessen iron limitation: Uptake of organically
1019 complexed iron and reduced cellular iron requirements, Limnology and Oceanography, 56(6),

1020 1983–2002, doi:10.4319/lo.2011.56.6.1983, 2011.

1021 Sunda, W. and Huntsman, S.: Effect of pH, light, and temperature on Fe–EDTA chelation and Fe
1022 hydrolysis in seawater, Marine Chemistry, 84(1-2), 35–47, doi:10.1016/S0304-4203(03)00101-4,
1023 2003.

1024 Sunda, W. G. and Huntsman, S. A.: Iron uptake and growth limitation in oceanic and coastal
1025 phytoplankton, Marine Chemistry, 50, 189–206, 1995.

1026 Thingstad, F. and Billen, G.: Microbial degradation of *Phaeocystis* material in the water column,
1027 Journal of Marine Systems, 5, 55–65, doi:10.1016/0924-7963(94)90016-7, 1994.

1028 Tzarfati-Majar, V., Burstyn-Cohen, T. and Klar, A.: F-spondin is a contact-repellent molecule
1029 for embryonic motor neurons, Proceedings of the National Academy of Sciences, 98(8), 4722–
1030 4727, 2011.

1031 van Boekel, W.: *Phaeocystis* colony mucus components and the importance of calcium ions for
1032 colony stability, Marine Ecology Progress Series, 87, 301–305, 1992.

1033 Vardi, A.: Cell signaling in marine diatoms, Communicative & Integrative Biology, 1(2), 134–
1034 136, doi:10.1016/j.cub.2008.05.037, 2008.

1035 VerBerkmoes, N. C., Hervey, W. J., Shah, M., Land, M., Hauser, L., Larimer, F. W., Van
1036 Berkel, G. J., and Goeringer, D. E.: Evaluation of "shotgun" proteomics for identification of
1037 biological threat agents in complex environmental matrixes: experimental simulations,
1038 Analytical chemistry, 77, 923-932, 2005.
1039
1040 Verity, P. G., Brussaard, C. P., Nejstgaard, J. C., van Leeuwe, M. A., Lancelot, C. and Medlin,
1041 L. K.: Current understanding of *Phaeocystis* ecology and biogeochemistry, and perspectives for
1042 future research, Biogeochemistry, 83, 311–330, doi:10.1007/s10533-007-9090-6, 2007.

1043 Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T.,
1044 Maechler, M., Magnusson, A., Moeller, S. and Schwartz, M.: gplots: Various R programming
1045 tools for plotting data. 2009.

1046 Watanabe, Y., Hayashi, M., Yagi, T. and Kamiya, R.: Turnover of actin in *Chlamydomonas*
1047 flagella detected by fluorescence recovery after photobleaching (frap), Cell, 29, 67–72,
1048 doi:10.1247/csf.29.67, 2004.

1049 Whitney, L. P., Lins, J. J., Hughes, M. P., Wells, M. L., Chappell, P. D. and Jenkins, B. D.:
1050 Characterization of putative iron responsive genes as species-specific indicators of iron stress in
1051 *Thalassiosiroid* diatoms, Frontiers in Microbiology, 2, 1–14,
1052 doi:10.3389/fmicb.2011.00234/abstract, 2011.

1053 Williams, T. J., Long, E., Evans, F., DeMaere, M. Z., Lauro, F. M., Raftery, M. J., Ducklow, H.,
1054 Grzymski, J. J., Murray, A. E. and Cavicchioli, R.: A metaproteomic assessment of winter and
1055 summer bacterioplankton from Antarctic Peninsula coastal surface waters, The ISME Journal,
1056 6(10), 1883–1900, doi:10.1038/ismej.2012.28, 2012.

Wu, Z., Jenkins, B. D., Rynearson, T. A., Dyhrman, S. T., Saito, M. A., Mercier, M. and Whitney, L. P.: Empirical bayes analysis of sequencing-based transcriptional profiling without replicates, BMC Bioinformatics, 11, 564, doi:10.1186/1471-2105-11-564, 2010.

Zilliges, Y., Kehr, J. C., Mikkat, S., Bouchier, C., de Marsac, N. T., Borner, T. and Dittmann, E.: An extracellular glycoprotein is implicated in cell-cell contacts in the toxic cyanobacterium *Microcystis aeruginosa* PCC 7806, Journal of Bacteriology, 190(8), 2871–2879, doi:10.1128/JB.01867-07, 2008.

Zingone, A., Chrétiennot-Dinet, M.-J., Lange, M. and Medlin, L.: Morphological and genetic characterization of *Phaeocystis cordata* and *P. jahnii* (prymnesiophyceae), two new species from the Mediterranean Sea, Journal of Phycology, 35(6), 1322–1337, doi:10.1046/j.1529-8817.1999.3561322.x, 1999.

Zurbriggen, M. D., Tognetti, V. B., Fillat, M. F., Hajirezaei, M.-R., Valle, E. M. and Carrillo, N.: Combating stress with flavodoxin: a promising route for crop improvement, Trends in Biotechnology, 26(10), 531–537, 2008.