Influence of cobalamin scarcity on diatom molecular physiology and identification of a cobalamin acquisition protein

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1073/pnas.1201731109">http://dx.doi.org/10.1073/pnas.1201731109</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>National Academy of Sciences</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Mon Nov 26 23:56:38 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/74655">http://hdl.handle.net/1721.1/74655</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td></td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Influence of cobalamin scarcity on diatom molecular physiology and identification of a cobalamin acquisition protein

Erin M. Bertranegra, Andrew E. Allenb, Christopher L. Dupontc, Trina M. Norden-Krichmar2, Jing Bai2, Ruben E. Valasc, and Mak A. Saitol

aMarine Chemistry and Geochemistry Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543; bMarine Chemistry and Geochemistry Joint Program in Chemical Oceanography, Woods Hole, MA 02543; and cMicrobial and Environmental Genomics, J. Craig Venter Institute, San Diego, CA 92121

Diatoms are responsible for ~40% of marine primary production and are key players in global carbon cycling. There is mounting evidence that diatom growth is influenced by cobalamin (vitamin B12) availability. This cobalt-containing micronutrient is only produced by some bacteria and archaea but is required by many diatoms and other eukaryotic phytoplankton. Despite its potential importance, little is known about mechanisms of cobalamin acquisition in diatoms or the impact of cobalamin scarcity on diatom molecular physiology. Proteomic profiling and RNA-sequencing transcriptomic analysis of the cultured diatoms Phaeodactylum tricornutum and Thalassiosira pseudonana revealed three distinct strategies used by diatoms to cope with low cobalamin: increased cobalamin acquisition machinery, decreased cobalamin demand, and management of reduced methionine synthase activity through changes in folate and S-adenosyl methionine metabolism. One previously uncharacterized protein, cobalamin acquisition protein 1 (CBA1), was up to 160-fold more abundant under low cobalamin availability in both diatoms. Autologous overexpression of CBA1 revealed association with the outside of the cell and likely endoplasmic reticulum localization. Cobalamin uptake rates were elevated in strains overexpressing CBA1, directly linking this protein to cobalamin acquisition. CBA1 is unlike characterized cobalamin acquisition proteins and is the only currently identified algal protein known to be implicated in cobalamin uptake. The abundance and widespread distribution of transcripts encoding CBA1 in environmental samples suggests that cobalamin is an important nutritional factor for phytoplankton. Future study of CBA1 and other molecular signatures of cobalamin scarcity identified here will yield insight into the evolution of cobalamin utilization and facilitate monitoring of cobalamin starvation in oceanic diatom communities.
Methionine is also required for dimethylsulfonium propionate production, which is used by a subset of diatoms possibly as a cryoprotectant, osmolyte (25), or antioxidant (26), and is the precursor to the climatically important gas dimethylsulfide (27). In addition, impaired methionine synthase activity prevents efficient folate recycling, which has important implications for nucleic acid biosynthesis (14, 28).

Here, we used MS-based proteomics and RNA-sequencing (RNA-seq) transcriptomics to identify diatom gene products involved in cobalamin metabolism and to assess the overall impact of cobalamin scarcity on diatom molecular physiology. This study has yielded the identification of a unique protein involved in cobalamin acquisition and characterized multiple molecular level responses to cobalamin scarcity in diatoms.

Fig. 1. Effect of low cobalamin and iron availability on growth and protein expression in two diatoms. Cell density over time for *T. pseudonana* (A) and *P. tricornutum* (B) grown under four different nutrient regimes: low vitamin B<sub>12</sub>, low Fe, low vitamin B<sub>12</sub> with low Fe, and replete. Values shown are means of triplicate cultures, and error bars are 1 SD. Arrows indicate where samples for proteomic and transcriptomic analyses were taken. Low cobalamin availability had a much larger impact on *T. pseudonana* growth than on the growth of *P. tricornutum*, likely attributable to *P. tricornutum*’s use of MetE as an alternative to the vitamin B<sub>12</sub>-requiring MetH. Because of the low iron concentrations used, iron limitation had a more severe impact on growth than low vitamin B<sub>12</sub> in this the experiment. (C–F) Limitation scenarios were verified by resupplies of cultures with cobalamin and iron. Fluorescence over time is shown as means of single measurements of triplicate cultures, with error bars representing 1 SD. Each culture in A and B (at arrows) was split in four and resupplied with nothing (control), vitamin B<sub>12</sub>, iron, or vitamin B<sub>12</sub> and Fe together. (C) Growth in low vitamin B<sub>12</sub> cultures of *T. pseudonana* was only rescued by the addition of both vitamin B<sub>12</sub> and iron together; fluorescence in +Fe cultures and +vitamin B<sub>12</sub> and Fe cultures was significantly different (Student’s paired t test, *P* = 0.0004; time = 8 d). (E) Iron addition alone rescued growth in low vitamin B<sub>12</sub> and iron *P. tricornutum*, and coaddition of vitamin B<sub>12</sub> and Fe further enhanced growth; fluorescence in +Fe cultures and + vitamin B<sub>12</sub> and Fe cultures was significantly different (Student’s paired t test, *P* = 0.0007; time = 4 d). Shotgun MS analyses of *T. pseudonana* (G) in the low vitamin B<sub>12</sub> vs. replete treatment and *P. tricornutum* (H) in the low vitamin B<sub>12</sub> with low iron treatment vs. low iron treatment reveal vitamin B<sub>12</sub>-responsive proteins. Each point is an identified protein, with the mean of its technical triplicate abundance scores in one treatment plotted against the mean of abundance scores in another treatment. The solid line is 1:1 abundance, and the dashed lines denote 2:1 and 1:2 abundances. Protein CBA1, Tp11697 and Pt 48322, is highlighted in pink, and MetE is highlighted in blue. Proteins in green are considered differentially abundant (Fisher exact test, *P* < 0.01).
Results and Discussion

We used quantitative proteomic and transcriptomic methodologies to investigate the response of two distantly related diatoms, 

T. pseudonana and P. tricornutum, to low cobalamin availability. 

T. pseudonana has an absolute cobalamin requirement because its genome encodes only the vitamin B12-dependent methionine synthase MetH. P. tricornutum has a flexible cobalamin demand because its genome encodes both MetE and MetH. To facilitate the identification of gene products that specifically respond to low cobalamin supply in these diatoms, we used a multifactorial experimental design to evaluate their response to low cobalamin under conditions of both high and low iron availability (Fig. 1).

Gene products displaying consistent responses under low cobalamin, regardless of iron nutritional status, were identified as specifically involved in cellular responses to cobalamin scarcity. The low iron treatment provided a control against which to judge whether gene product abundance changes were more general to stress or low growth rate, or whether they displayed a specific response to low cobalamin. Iron was chosen over other nutrients for this factorial design because simultaneous iron and cobalamin limitation has been documented in field communities of diatoms (6, 29). Therefore, any efforts to develop molecular markers for low cobalamin availability that can potentially be implemented in environmental samples must consider the impact of iron starvation on those markers.

The growth of diatoms under these limitation scenarios is described in Fig. 1 A and B. Proteome and transcriptome analyses were conducted on single (protein) or duplicate (transcript) cultures from each condition, which were harvested at the time points denoted with (Fig. 1 A and B, arrows). Later time points for the iron-limited cultures were chosen to facilitate the harvesting of additional biomass. Several major patterns emerged from these analyses. First, we measured a large difference in the extent of proteome and transcriptome rearrangement under cobalamin starvation that was dependent on the cobalamin requirements of each diatom (Fig. 1 G and H and Table 1). Second, we identified a previously uncharacterized protein [P. tricornutum 48322 and its ortholog in T. pseudonana 11697, hereafter referred to as cobalamin acquisition protein 1 (CBA1)] that was more abundant under low cobalamin availability in both diatoms (Fig. 1). Third, we identified changes in gene products involved in folate and methionine cycling, supporting the hypothesis that cobalamin starvation has implications for cellular one-carbon metabolism.

Verification of Nutrient Limitation Scenarios. Cobalamin and iron resupply experiments confirmed that these diatom cultures were starved for nutrients as intended, with iron rescuing growth of both low iron cultures and cobalamin rescuing growth only in the low cobalamin cultures of the cobalamin-requiring diatom, T. pseudonana (Fig. 1 C–F and SI Appendix, Fig. S1). Growth in the combined low cobalamin and low iron T. pseudonana culture was only restored upon the addition of both cobalamin and iron together, demonstrating that this culture was simultaneously limited by the availability of both nutrients (Fig. 1C). In contrast, growth in low cobalamin and low iron P. tricornutum cultures was rescued by iron addition alone and was further enhanced by the coaddition of cobalamin and iron (Fig. 1E). These differing responses were expected and are likely attributable to the different requirements for cobalamin in these diatom species.

Cobalamin Scarcity-Induced Proteome and Transcriptome Rearrangement. To detect changes in abundant proteins under environmentally relevant conditions, we used shotgun proteomic methods that allow for a high degree of reproducibility (30) (SI Appendix, Fig. S2) as opposed to those that are optimized for deep proteome coverage. In sum, 764 T. pseudonana proteins were detected from a total of 4,955 unique peptides with a 0.19% peptide false discovery rate. These proteins and peptides are listed with their putative function in Dataset S1. A total of 859 P. tricornutum proteins were detected from 5,172 unique peptides with a 0.22% peptide false discovery rate (Dataset S1). Of the total mass spectra acquired for T. pseudonana, 46% were assigned to peptides found in the genomic databases, whereas 52% of P. tricornutum spectra were assigned, corresponding to the identification of 6% and 8% of the genome-predicted protein models, respectively.

In the cobalamin-requiring diatom T. pseudonana, 19% of detected proteins were significantly differentially abundant under low cobalamin availability compared with the replete control (Fig. 1C and Table 1). This suggests that the diatom conducts a significant rearrangement of cellular function when grown under cobalamin limitation. Although some of these changing

| Table 1. Pairwise comparisons of growth rate, cell yield, protein abundance changes, and transcript abundance changes between low cobalamin vs. replete growth, low cobalamin with low iron vs. low iron growth, and low iron vs. replete growth in two diatoms |
|---------------------------------|-----------------|------------------|-----------------|
|                                 | Low vitamin B12 vs. replete | Low vitamin B12 and Fe vs. low Fe | Low Fe vs. replete |
| % proteins differentially abundant | T. pseudonana | 19 | 18 | 30 |
|                                 | P. tricornutum | 5 | 1 | 20 |
| % transcripts differentially abundant | T. pseudonana | 26 | 5 | 25 |
|                                 | P. tricornutum | 6 | 2 | 16 |
| Fold cell yield decrease | T. pseudonana | 1.8 ± 0.1 | 1.0 ± 0.1 | 3.4 ± 0.1 |
|                                 | P. tricornutum | 1.1 ± 0.1 | 1.6 ± 0.2 | 4.1 ± 0.4 |
| Fold growth rate decrease | T. pseudonana | 1.2 ± 0.1 | 1.2 ± 0.1 | 2.0 ± 0.1 |
|                                 | P. tricornutum | 1.0 ± 0.1 | 1.3 ± 0.1 | 2.8 ± 0.3 |

The percentage of proteins changing in abundance was calculated from the total number of identified proteins and those that had significantly different abundance between the two treatments compared (Fisher exact test, P < 0.01). The percentage of differentially abundant transcripts was calculated from the number of transcripts mapped to genomic locations that had log2-fold change RPMK values greater than 1 or less than −1 between the two treatments. Fold cell yield and growth rate decreases were calculated by determining the fold change between the maximum cell density or cell-specific growth rate in each treatment and are given as means of biological triplicates ± 1 SD. Growth rates are cell-specific and were calculated from the following time periods: T. pseudonana high iron, days 2–4; T. pseudonana low iron, days 3–5; P. tricornutum high iron, days 3–6; and P. tricornutum low iron, days 5–7.
proteins are likely responding to the accompanying growth rate depression, there are many that display different behavior under cobalamin vs. iron limitation (Dataset S1) and have putative functions suggesting they are directly related to vitamin B\textsubscript{12} metabolism (Table 2 and SI Appendix, Table S1). Even though iron limitation induced in this study had a much more severe impact on growth rate than low cobalamin did, changes induced in the *T. pseudonana* proteome by cobalamin starvation were nearly as large as those induced by iron limitation (Table 1 and SI Appendix, Fig. S3). In contrast, *P. tricornutum*, which can accomplish methionine synthesis without the use of the vitamin, displayed a relatively minor proteome change in response to cobalamin scarcity (Table 1 and SI Appendix, Fig. S3). Protein abundance changes under the combined low vitamin B\textsubscript{12} and low iron treatment vs. treatment with low iron alone showed a similar pattern. *T. pseudonana*, even under severe iron limitation, rearranged its protein complement significantly to manage cobalamin scarcity, whereas *P. tricornutum* changed the abundance of less than 1% of the proteins in its detected proteome (Fig. 1H and Table 1). The small change detected between these two treatments in *P. tricornutum* reflects both the minimal metabolic rearrangement induced in these cells and the efficacy of the proteomic analyses applied here.

RNA-seq transcriptomic analyses revealed trends in diatom molecular physiology that were broadly coherent with those observed via proteomics; a similar percentage of the measured transcriptome and proteome changed as a result of each starvation scenario (Table 1). Given the deep coverage of the diatom genomes obtained via these RNA-seq analyses (10,404 genes with mapped transcripts in *P. tricornutum* and 11,778 genes with mapped transcripts in *T. pseudonana*; Dataset S2) and the coherence in the proteome and transcriptome datasets, these data suggest that the cobalamin-requiring diatom *T. pseudonana* conducts a significant rearrangement of its molecular physiology under cobalamin starvation. The diatom with a flexible cobalamin demand, *P. tricornutum*, changed a much smaller proportion of transcript abundances in response to cobalamin scarcity compared with the cobalamin-requiring *T. pseudonana*, also consistent with changes observed in the proteome.

Despite similar global trends in percentage of gene products changing under these conditions, both diatoms showed little coherence in the abundance patterns of specific proteins and transcripts observed in this study (Fig. 2 A and B). This is consistent with the limited available previous results, suggesting that there is substantial decoupling between transcript and protein abundance patterns in eukaryotic algae (31). Notable exceptions were certain nutrient-specific gene products that showed coordinated increased protein and transcript abundance in response to phosphorus deprivation previously (31, 32) and to cobalamin scarcity in this study, as discussed in subsequent sections (Fig. 2).

**Identification of a Previously Undescribed Cobalamin Acquisition Protein.** In this shotgun proteomic analysis, the protein that showed the largest response to cobalamin starvation in *T. pseudonana* was a previously uncharacterized hypothetical protein. Here, we identify this protein as a cobalamin acquisition protein, as described below, and refer to it as CBA1 (Figs. 1G and 3). A protein homologous to the *T. pseudonana* CBA1 was detected in the *P. tricornutum* global proteome, also only under low cobalamin availability, suggesting that this protein may be playing a similar role in both diatoms and that it is likely involved in cobalamin metabolism (Figs. 1H and 3D). These shotgun proteomic results were confirmed through absolute protein quantification via a more sensitive and quantitative technique, selected reaction monitoring (SRM) MS, revealing that the concentration of CBA1 protein was between 10- and 160-fold higher under low vitamin B\textsubscript{12} availability in *P. tricornutum* (Fig. 2C). These SRM assays were developed by choosing two tryptic peptides diagnostic of CBA1 and designing specific MS detection assays for
This method involves the use of stable, isotopically labeled versions of two diagnostic tryptic peptides (Pt48322_1 and Pt48322_2) that were used as internal standards. Each behaved linearly over four orders of magnitude and allowed for absolute quantification of each of these peptides, both generated from the CBA1 amino acid sequence, in *P. tricornutum* peptide samples. However, one peptide, Pt48322_2, was measured at consistently higher abundance in *P. tricornutum* than the other diagnostic peptide, Pt48322_1 (Fig. 3C). Careful examination of each, as previously described (30) (SI Appendix, Table S2), showed that there was both protein- and transcript-based quantitative information are displayed. The fold change (log2) between the protein abundance (spectral counting score) and the transcript abundance (RPKM value) in the cobalamin-starved and replete treatments is shown on the y axis, and the fold change (log2) between the protein abundance (spectral counting score) and the transcript abundance (RPKM value) in the cobalamin-starved and replete treatments is shown on the x axis. For the protein data, any null values were replaced with a spectral counting score of 0.33, the lowest measurable value in our experiments, to facilitate the computation. Generally, coherence between the proteome and transcriptome responses is limited to specific proteins that display enhanced abundance under cobalamin starvation in both the transcript and protein pools. These include CBA1, MetE, ThiC, and cytosolic SHMT, as noted by color and identified in the key. (C) Heat map displays select *T. pseudonana* transcript responses to cobalamin and iron starvation. Fold change RPKM values are shown for the low iron vs. replete, low vitamin B12 with low Fe vs. low Fe, low vitamin B12 with low Fe vs. replete, and low vitamin B12 vs. replete treatments, with up-regulation denoted in red and down-regulation denoted in green. The genes were selected by high-to-low ordering of the log2-transformed fold change RPKM values and sorted by the comparison between low vitamin B12 vs. replete treatments. Gene products highlighted in A (ThiC, CBA1) are also highlighted in C.
of nucleic acid sequences amplified from cobalamin-limited *P. tricornutum* RNA extracts revealed that this variability was attributable to SNPs within allelic copies of the CBA1 coding sequences of this diatom diatom genome. Diatoms possess two copies of each chromosome, and sequence analysis revealed that there are minutely different versions of the gene encoding CBA1 on each of these copies. These slightly different genes produce CBA1 protein with sequences that differ by three amino acid residues. The diagnostic tryptic peptide target Pt48322_1 contains one of these variable amino acid residues, and hence is encoded by one of the two homologous copies, whereas the other peptide target, Pt48322_2, does not contain a variable site and is encoded by both copies (details are provided in *SI Appendix*, Fig. S1). The abundance of these peptides was linearly correlated across all samples (\( r^2 = 0.999; \text{SI Appendix, Fig. S5} \)), and Pt48322_2, the peptide encoded by both allelic copies, was more abundant (Fig. 2C). This is therefore an example of canonical gene expression in a diploid genome, where allelic copies display similar expression patterns.

RNA-seq analysis revealed that CBA1 transcript abundance patterns were similar to those observed for the corresponding proteins via the two proteomic approaches described above. Much higher CBA1 transcript abundance (4- to 30-fold) was observed under low cobalamin availability in both *T. pseudonana* and *P. tricornutum* (*Fig. 3B*). Together, these analyses reveal that CBA1 protein and transcripts display coordinated behavior under cobalamin scarcity (*Fig. 2*).

CBA1 has a clear N-terminal signal peptide sequence for secretion (Cello and SignalP- predicted (33, 34)) and no transmembrane domains. It contains a partial conserved domain that is weakly similar to the periplasmic component of a bacterial iron exportation [Cello and SignalP-predicted (33, 34)] and no transmembrane domains. It contains a partial conserved domain that is weakly similar to the periplasmic component of a bacterial iron hydroxamate ATP-binding cassette (ABC) transport system (FepB; N-terminal end is truncated, Pt48322 BLASTp search E-value 1.33e-4), but the protein otherwise lacks characterized domains. There appear to be homologous versions of CBA1 encoded in all currently sequenced diatom genomes as well as in those from other members of the stramenopile lineage, *Ectocarpus siliculosus* and *Aureococcus anophagefferens*. Outside of the stramenopile lineage, we find no sequences that clearly encode CBA1 within available algal genome (*SI Appendix, Table S3*) or transcriptome datasets, including extensive dinoflagellate transcriptome sequencing. However, because these sequences are not highly conserved and show some similarities to a class of bacterial proteins, rigorous analysis of the functional similarity of these proteins between algal lineages will depend on the future identification of residues essential for CBA1’s functionality.

**Overexpression of CBA1, Subcellular Localization, and Phenotypic Characterization.** We examined the subcellular localization of CBA1 through overexpression of the *P. tricornutum* isofrom (Pt48322) in the native host as an YFP fusion construct. Epi-fluorescent microscopy showed that the YFP signal was associated with the outer membrane and also with an intracellular organelle-like structure adjacent to the chloroplast (*Fig. 4A*, white arrow). YFP fluorescence was not observed throughout the cytoplasm or within the mitochondria, chloroplast, peroxisome, or nucleus. Intracellular localization around, but not within, the chloroplast was verified using confocal microscopy (*Fig. 4B*) and is similar to fluorescence microscopy of other proteins localized to the diatom endoplasmic reticulum (ER), which envelopes the chloroplast in red lineage algae (35). The most well-characterized eukaryotic protein secretion pathway begins with translocation of proteins with signal peptides across the ER membrane into the lumen, where they are retained for processing and folding and are subject to quality control. After ER processing, these proteins are exported to the Golgi network and then relocated to the outside of the cell via vesicle transport (36). The likely ER lumen localization detected here for CBA1 may thus be attributable to processing of this protein before secretion. This is consistent with the strong prediction for a signal peptide (\( D = 0.86; \text{SignalP 3.0} \)) and observed additional targeting to the periphery of the cell.

We characterized the phenotypic response of this overexpression line of *P. tricornutum* by measuring cobalamin uptake rates in two cell lines overexpressing this protein (CBA1-OE1 and CBA1-OE2) and comparing them with uptake rates in the wild type (wt) and a line overexpressing an unrelated protein, Urease (Urease-OE1). To repress native CBA1 expression, uptake rates were measured in cultures grown in cobalamin-replete conditions. In the transgenic diatoms, the promoter for a light harvesting complex protein (FcpB), which is highly expressed under exponential growth, controls CBA1 overexpression. Overexpression of CBA1 enhanced cell-specific, radiolabeled cobalamin uptake rates in exponentially growing *P. tricornutum* cells two- to threefold (*Fig. 4C*). This enhanced uptake rate directly implicate CBA1 in cobalamin acquisition and, along with its localization to the outside of the cell, suggests that CBA1 may bind cobalamin and, through association with additional unidentified proteins, participate in its transport into the cell. This finding is significant in that CBA1 is, to our knowledge, the sole identified protein in any marine eukaryotic microbe to be directly linked to cobalamin acquisition.

CBA1 in Environmental Sequencing Datasets. We identified transcripts that likely encode CBA1 in cDNA libraries generated from natural phytoplankton communities (*Fig. 5 and SI Appendix, Fig. S7*). CBA1 transcripts were detected in each library examined. These libraries were generated from diverse marine locations, including sea ice and water column samples from the Ross Sea of the Southern Ocean and water column samples from the North Pacific Ocean, Monterey Bay, and Puget Sound, suggesting that
CBA1 is widely distributed within the marine environment. These nucleic acid sequences are displayed as a phylogenetic tree that uses CBA1 sequences from available genomes of cultured organisms to construct a reference tree onto which metatranscriptomic sequences are placed (37) (Fig. 5). Although CBA1 is not among the most abundant diatom transcripts detected, its expression level is comparable to that of transcripts encoding other essential well-characterized proteins (SI Appendix, Fig. S7). This suggests that CBA1 and the micronutrient cobalamin play important roles in the molecular physiology and ecology of natural diatom communities. In metagenomic data, DNA sequences most similar to CBA1 were detected only in the larger size fraction (>3 μm) samples, suggesting that this protein is restricted to larger phytoplankton. The majority of environmentally derived transcripts encoding CBA1 are most similar to those encoded by cultured diatoms (Fig. 5), suggesting that diatoms are likely the major reservoir of this protein in the marine environment.

**Other Molecular Responses to Low Cobalamin Availability.** Identification of CBA1 and its abundance patterns in culture suggests that diatoms adjust their molecular physiology to increase capacity for cobalamin acquisition in the face of cobalamin deprivation. We can consider other cobalamin-sensitive transcripts and proteins to identify additional molecular responses to low vitamin availability.

Included in the small pool of *P. tricornutum* gene products changing under cobalamin starvation is the cobalamin-independent methionine synthase MetE (28056), which was much more abundant under low cobalamin availability (Figs. 1D and 2 and Table 2). This suggests that *P. tricornutum* expresses MetE when cobalamin is scarce, consistent with transcript abundance patterns observed previously in this diatom (21). RNA-seq results revealed that an adjacent two-component histidine kinase sensor appears to be coregulated with *metE*, and thus may play a role in the *P. tricornutum* response to cobalamin starvation (Fig. 6). The cobalamin-dependent methionine synthase MetH was not detected in our proteome study, possibly attributable to low abundance, which is expected because MetH has much higher catalytical activity compared with MetE (24). MetH (Pt23399, Tp 693) transcripts were detected here via RNA-seq and did not show significant changes in abundance as a function of cobalamin availability in *P. tricornutum* but were more abundant under low cobalamin in *T. pseudonana* (Fig. 2C and Dataset S2).

![Fig. 5](image-url). Transcripts encoding CBA1 are expressed in diverse marine environments. Phylogenetic tree containing CBA1 sequences from 454 metatranscriptomic (cDNA) libraries from the Ross Sea of the Southern Ocean, Monterey Bay, Puget Sound, and North Pacific Ocean. Reference sequences from *P. tricornutum*, *F. cylindrus*, *T. pseudonana*, *A. anophagefferens*, and *E. siliculosus* genomes were used to construct these trees (37) and are shown in black. CBA1-like sequences from environmental samples are shown in color, as described in the key. CBA1 transcripts were detectable in diverse marine environments, suggesting that cobalamin acquisition is an important component of diatom molecular physiology.

![Fig. 6](image-url). RNA-seq analysis reveals coregulation of MetE and a two-component sensor. RNA-seq coverage for an 11-kb region of the *P. tricornutum* genome is shown. Individual tracks are shown for each treatment, cobalamin and iron starvation, cobalamin starvation, iron starvation, and the replete control. The x axis shows the position in the genome, and the y axis (gray shading) shows the relative coverage of transcript data. Vertical color lines represent areas in the coverage mapping where there were mismatches of the reads to the reference genome (A = green, C = blue, G = yellow, T = red). The bottom track shows the gene models from the Joint Genome Institute 2.0 genome project. Transcripts mapping to cobalamin-independent methionine synthase (*metE*) are much more abundant under cobalamin scarcity and with low cobalamin and low iron. In addition, a two-component histidine kinase sensor appears to be coregulated with *metE*, and may thus play a role in the *P. tricornutum* response to cobalamin starvation.
Another potential use for cobalamin in diatom cells is as a co-factor for methylmalonyl CoA mutase (MmcM; Pt 51830, Tp 33685). This enzyme’s function remains unclear in diatoms, although it may be important in the citric acid cycle as well as in propionate and fatty acid metabolism (20). MmcM uses adenosylcobalamin as a co-factor, which could be produced via an adenosylcobalamin transferase enzyme encoded in these diatom genomes (CblB; Pt 45992, Tp 263198). MmcM encoding transcripts did not respond to low cobalamin, but those encoding CblB were more abundant under cobalamin scarcity (Fig. 3C and Dataset S2). Because this protein may be responsible for generating adenosylcobalamin, a biologically active form of the vitamin that is not used by MetH, the reason for this response remains unclear and leaves open the possibility that there are additional, unrecognized cobalamin-dependent metabolisms in diatoms.

Another potentially important consequence of low cobalamin availability to diatom cells is methyl folate trapping (14, 38). Under conditions of reduced methionine synthase activity, 5-methyltetrahydrofolate (MeTHF) buildup can prevent efficient folate cycling, and thus have an impact on other folate-dependent metabolisms, such as DNA synthesis (28). We find molecular evidence for this phenomenon in these diatoms, as summarized in Fig. 7. A protein involved in folate one-carbon metabolism, cytosolic serine hydroxymethyltransferase (SHMT) (36, 37), is more abundant under low cobalamin (Figs. 2 and 7 and Dataset S2). SHMT is pyridoxal 5'-phosphate (PLP, vitamin B6)-dependent and catalyzes the interconversion of THF and 5,10-methylene tetrahydrofolate [5,10 MTHF (39)]. The 5,10 MTHF can then be converted irreversibly to MeTHF, which accumulates under low methionine synthase activity and leads to folate trapping in other organisms (28). The increase in cytosolic SHMT abundance under cobalamin scarcity suggests that diatoms may increase their capacity for THF and 5,10 MTHF interconversion under low cobalamin conditions, perhaps in an effort to mediate folate cycling. This is consistent with assertions that SHMT mediates the partitioning of one-carbon units between DNA synthesis and methionine cycling in other organisms (40). Additionally, an enzyme putatively involved in PLP synthesis is much more abundant under low cobalamin in T. pseudonana (Figs. 2 and 7). This increase is consistent with a higher demand for PLP under low cobalamin, potentially for use by the PLP-dependent SHMT enzymes. Taken together, these data suggest that folate, cobalamin, and PLP metabolism are linked in diatoms, as observed previously in metazoans (41).

These results also suggest that AdoMet starvation may be an important consequence of cobalamin deprivation in diatoms (Fig. 7). Methionine adenosyltransferase (MetK), which converts methionine to AdoMet, was more abundant under low cobalamin in both diatoms (Fig. 7 and SI Appendix, Table S1). In addition to many other cellular functions, AdoMet is responsible for reductive repair and remethylation of cobalamin in MetH (42). Increased MetK levels may enhance AdoMet production, leading to more efficient repair of oxidized cobalamin in MetH. It is also possible that enhanced MetK abundance results from attempts to meet cellular AdoMet demand despite methionine scarcity. ThiC, a thiamine biosynthesis protein, which uses an AdoMet-dependent radical reaction to form the nonsulfur-containing branch of thiamine, 4-amino-5-hydroxymethyl-2-methylpyrimidine (43), was more abundant under low cobalamin (Figs. 2 and 7). Because the other proteins involved in thiamine biosynthesis were not observed to be more abundant under cobalamin scarcity, it may be that ThiC is up-regulated in response to AdoMet deprivation (Fig. 7), supporting the notion that AdoMet starvation is an important consequence of low cobalamin availability.

Although some proteins that display cobalamin-responsive abundance patterns have predicted cellular functions as described above, more than half of them play unknown roles (Table 2 and SI Appendix, Table S2). Three proteins of unknown function in T. pseudonana (24346, 23556, and 1896; Table 2) do not have homologs in P. tricornutum, were more abundant under low cobalamin availability, and did not increase in abundance under low iron. These unknown proteins may be involved in the T. pseudonana response to cobalamin scarcity and warrant further evaluation of their potential role in cobalamin acquisition.

**Implications for Cobalamin Biochemistry and Marine Biogeochemistry.** Our results suggest that diatoms use at least three strategies in response to cobalamin scarcity, including efforts to increase cobalamin acquisition machinery, reduce cobalamin demand, and mitigate damage induced by reduced methionine synthase activity.
Fig. 8. Diatoms display three primary responses to cobalamin scarcity. Schematic representation of the three primary responses to cobalamin starvation in two diatoms. Both diatoms enhanced CBA1 production, likely in an effort to enhance cobalamin acquisition. The magnitude of the increase in CBA1 protein and transcripts was larger for T. pseudonana, likely because it has an absolute cobalamin requirement. P. tricornutum enhanced MetE production to reduce cobalamin demand. Both diatoms also appeared to conduct cellular rearrangements to cope with reduced methionine synthase activity, including enhanced cytosolic SHMT, MetK, and radical AdoMet (SAM) enzyme ThiC abundance under low cobalamin availability.

(Fig. 8). These results implicate enhanced CBA1, MetE, and SHMT abundance as well as altered folate and PLP metabolism in the acclimation of diatom cells to low cobalamin availability and suggest that AdoMet starvation is an important consequence of cobalamin deprivation in diatoms. Future metabolomic mapping of the folate, PLP, homocysteine, and methionine-derived compounds involved in these processes could yield important additional insight into these dependencies.

Because CBA1 is a component of a previously undescribed cobalamin acquisition pathway that differs from the well-characterized bacterial and metazoan mechanisms, its identification may advance efforts to understand the biochemistry and evolution of cobalamin acquisition. Further analyses are required to elucidate the specific role served by CBA1 in cobalamin acquisition and to identify the other proteins involved in this process. Subsequent comparative studies of different cobalamin acquisition pathways may then yield valuable insight into cobalamin biochemistry and utilization throughout time.

Previous work revealed that diatoms and other algae exude a protein into their growth media that binds vitamin Bi2 very strongly (44, 45). In T. pseudonana, this protein was shown to be a distinct, large, multimeric complex (46), but the identity of its components remains unknown. It is possible that CBA1 represents a part of this multimer, although this is unlikely because overexpression of CBA1 without any other components of a complex enhances cobalamin uptake, whereas the exuded vitamin Bi2 binder is thought to restrict vitamin uptake (44, 45). It is also possible that CBA1 interacts with this complex to move cobalamin from the growth media into the cell (47). Future study will be required to elucidate the relationship between CBA1 and the strong exuded vitamin Bi2 binders produced by algae.

Detection of CBA1 transcripts in existing marine environmental datasets (Fig. 5 and SI Appendix, Fig. S7) implies that this protein is abundant and used by natural phytoplankton populations, and that cobalamin acquisition is thus an important component of diatom molecular physiology in the natural environment. This, along with the large-scale proteome and transcriptome rearrangement induced by cobalamin scarcity in diatoms, supports including cobalamin availability and its production and consumption dynamics in efforts to predict and model marine primary productivity. Future studies mapping the abundance of CBA1 and other cobalamin-responsive gene products identified here, along with traditional measures of productivity and community composition, may prove useful for elucidating the major oceanic controls on cobalamin starvation and utilization in phytoplankton communities and could yield insights into the influence of cobalamin availability on marine primary productivity and microbial community composition.

Methods

Axenic cultures of T. pseudonana CCMP 1335 and P. tricornutum CCMP 632 were grown in f/2-type media with a natural seawater base, with modified vitamin, EDTA, and trace metal concentrations (48). For the P. tricornum experiment, cells were acclimated with 5 nM added total iron (10−11 M Fe) and 0.5 pM added vitamin Bi2 for four transfers, allowing at least three doublings per transfer. Cells were then inoculated (3.2% vol/vol) into twelve 2.2-L bottles containing 1.8 L of media (4 treatments, biological triplicates). The media was as described above except for variable iron and vitamin Bi2 concentrations, which were chosen based on previous studies to result in cultures experiencing replete, iron-limited, low cobalamin and low cobalamin with iron limitation conditions (49, 50). The low iron treatment had 2.5 nM Fe\textsubscript{total} (10−11.3 M Fe) and 100 pM added vitamin Bi2, the low vitamin Bi2 treatment had no added vitamin Bi2 and 100 nM Fe\textsubscript{total} (10−9.9 M Fe), and the low vitamin Bi2 and low iron treatment had no added vitamin Bi2 and 2.5 nM Fe\textsubscript{total} (10−11 M Fe), whereas the replete treatment had 100 pM added vitamin Bi2 and 100 nM Fe\textsubscript{total} (10−9.9 M Fe). Samples were taken daily for fluorescence and cell counts, and were harvested for protein and RNA after 6 d for the high iron conditions and after 9 d for the low iron condition (Fig. 1 A and B, arrows). Just after the harvest time point, each remaining culture was split in four and resupplied with either nothing, 100 pM vitamin Bi12, 100 nM Fe\textsubscript{total} or both vitamin Bi2 and Fe, and growth was monitored via in vivo fluorescence.

The T. pseudonana experiment was conducted as above except with different vitamin Bi2 and iron concentrations chosen due to known differences in cobalamin and iron requirements for these diatoms (49–51). Acclimation cultures had 1 pM added vitamin Bi2 and 65 nM added Fe\textsubscript{total} (10−9.8 M Fe). The low iron treatment had 50 nM Fe\textsubscript{total} (10−10 M Fe) and 100 pM added vitamin Bi2, the low vitamin Bi2 treatment had 0.3 pM added vitamin Bi2 and 400 nM Fe\textsubscript{total} (10−9.09 M Fe), and the low vitamin Bi2 and low iron treatment had 0.3 pM added vitamin Bi2 and 50 nM Fe\textsubscript{total} (10−10.0 M Fe), and the replete treatment had 100 pM added vitamin Bi2 and 400 nM Fe\textsubscript{total} (10−9.9 M Fe). Just after the harvest time point, each culture was split in four and resupplied with either nothing, 100 pM vitamin Bi12, 400 nM Fe\textsubscript{total} or both vitamin Bi2 and Fe, and growth was monitored via in vivo fluorescence.

Protein was extracted, digested, and analyzed via shotgun MS as previously described, with minor modifications (30, 31, 52). Briefly, protein digests from one of the three replicates were analyzed in technical triplicate via liquid chromatography MS using a Paradigm MS4 HPLC system (Michrom) with reverse phase chromatography, a Michrom ADVANCE source, and a Thermo Scientific LTQ ion trap mass spectrometer. The mass spectra were searched using SEQUEST (Thermo, Inc.). Database search results were further processed using the PeptideProphet statistical model (53) and spectral counting abundance scoring within Scaffold 3.0 (Proteome Software, Inc.). Proteins discussed as “differentially expressed” in pairwise comparisons were determined by the Fisher exact test (P < 0.01) (54).

Targeted MS was conducted via SRM as previously described (30) for two tryptic peptides found to be unique to CBA1 in P. tricornutum, using a Thermo Advantage Triple Quadrupole Mass Spectrometer (Thermo Scientific). Iso-topically labeled versions of each tryptic peptide (Sigma-Aldrich) (55) were used as internal standards. Method details are given in SI Appendix, Table S2.

RNA from duplicate samples from each treatment was purified, amplified, and then used to prepare SOLID Total RNA-Seq Kit (Life Technologies) libraries according to the instructions of the manufacturer. The raw SOLID sequence data were mapped and aligned against the reference genome, and reads per kilobase of exon model (RPKM) values for each sample were assigned. These RPKM values were pooled for replicates, and differentially abundant transcripts were identified as those that had a log-fold change in RPKM values greater than 1 or less than –1 between the different treatments.

For CBA1 overexpression, full-length P. tricornutum 48322 cDNA was PCR-amplified and cloned into a TOPO pENTR vector (56) and then subjected to Gateway (Invitrogen) recombination with a diatom C-terminal YFP pDONR vector (56), which was transformed into P. tricornutum via particle bombardment (57). For localization, epitope tagged YFP expression lines similar to previously described methods, using 3′Co-labeled cDNA, were monitored via confocal microscopy performed on a Leica TCS SP5 spectral system.

Vitamin Bi2 uptake rate assessments were conducted on wt and overexpression lines similar to previously described methods, using 3′Co-labeled cobalamin monitored via gamma counting (38).

Metatranscriptomic analyses used to construct the tree in Fig. 5 were conducted on multiple samples collected from various marine locations. RNA was purified and amplified linearly, and cDNA was synthesized and prepared...
for sequencing on the 454 platform (Roche Diagnostic) according to manu-
facturer protocols. Orthologs to CBA1 were retrieved from the cDNA se-
quence data by TBLASTn. For comparative abundance analyses (Si Appendix,
Fig. S7), cDNA was prepared from seawater samples for sequencing on the illumina platform. The resulting reads were assembled de novo and compared with available sequence databases for functional and phylogenetic annota-
tion. Diatom ORFs assigned to each domain of interest were identified, and read counts for each ORF were summed across all six libraries. RPKM values were calculated for each ORF and were then summed across ORFs that con-
tained a domain of interest, effectively normalizing transcript abundance to
coding sequence length and abundance of all diatom-annotated transcripts.

ACKNOWLEDGMENTS. We thank Dawn Moran, Abigail Heithoff, Louie Wurch, Michael Hoffman, and Vladimir Bulgin for technical assistance; Abigail Noble, Jeff Hof-
man, and Jeff McQuaid for Antarctic sample collection; and Mauricio Arriagada for contributions to metatranscriptomic analyses. We acknowledge Cathy Drennan, Sonya Dyhrman, Dianne Newman, and Ben Van Mooy for helpful discussions. We are grateful for comments provided by anonymous reviewers. This work was supported by National Science Foundation Awards ANT 0732665, OCE 0752291, and OCE 1031271 and Gordon and Betty Moore Foundation funding (to M.A.S.); by National Science Foundation Graduate Research Fellowship 2007037200 and Environmental Protection Agency STAR Fellowship FE823224 (to E.E.B.); and by National Science Foundation Awards ANT 0732282, ANT 1043671, MCB 1024913, OCE 072799, and OCE 1136477 and Department of Energy Award DE SC0006719 (to A.E.A.).

dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional
synthase is a modular protein with distinct regions for binding homocysteine, methylytetrahydrofolate, cobalamin, and adenosylmethionine. Biochemistry 36:8082–8091.
cobalamin-independent and cobalamin-dependent methionine synthases from Ech-
22. Stefels JP (2000) Physiological aspects of the production and conversion of DMS in
prevent methyl group deficiency in kwashiorkor (methionine deficiency) and an
explanation for folinic acid induced exacerbation of subacute combined degeneration in
27. Saito MA, et al. (2011) Iron conservation by reduction of metalloenzymes inventories in
driven by phosphorus deficiency and recovery in the brown tide-forming alga Auresuccus anophagefferens. PLoS ONE 6:e28949.
and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 10:1–7.
Proteins 64:643–651.
33. Ellgaard L, Molinari M, Helenius A (1999) Setting the standards: Quality control in the
Health Aging 6:39–42.
37. Herbig K, et al. (2002) Cytoplasmic serine hydroxymethyltransferase mediates com-
petition between folate-dependent deoxyribonucleotide and S-adenosylmethionine
intake as primary determinants of homocysteinemia in an elderly population. JAMA
270:2693–2699.
synthase: The structure of a methylcobalamin-binding fragment and implications for