Phytotransferrin endocytosis mediates a direct cell surface-to-chloroplast iron trafficking axis in marine diatoms

Jernej Turnšek[1,2,3,4,5,6], John K. Brunson[6,7], Thomas J. Deerinck[8], Miroslav Oborník[9,10], Aleš Horák[9,10], Vincent A. Bielinski[11], Andrew E. Allen[4,6]*

[1]Biological and Biomedical Sciences, The Graduate School of Arts and Sciences, Harvard University, Cambridge, MA 02138, USA; [2]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA; [3]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA; [4]Integrative Oceanography Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093, USA; [5]Center for Research in Biological Systems, University of California San Diego, La Jolla, CA 92093, USA; [6]Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, CA 92037, USA; [7]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093, USA; [8]National Center for Microscopy and Imaging Research, University of California San Diego, La Jolla, CA 92093, USA; [9]Biology Centre CAS, Institute of Parasitology, Branišovská 31, 370 05 České Budějovice, Czech Republic; [10]University of South Bohemia, Faculty of Science, Branišovská 31, 370 05 České Budějovice, Czech Republic; [11]Synthetic Biology and Bioenergy, J. Craig Venter Institute, La Jolla, CA 92037, USA

*For correspondence: aallen@ucsd.edu (A.E. Allen)

Abstract

Iron is a biochemically critical metal cofactor in enzymes involved in photosynthesis, respiration, nitrate assimilation, nitrogen fixation and reactive oxygen species defense. Marine microeukaryotes have evolved a phytotransferrin-based iron uptake system to cope with iron scarcity, a major factor limiting primary productivity in the global ocean. Diatom phytotransferrin is internalized via endocytosis, however proteins downstream of this environmentally ubiquitous iron receptor are unknown. We applied engineered ascorbate peroxidase APEX2-based subcellular proteomics to catalog proximal proteins of phytotransferrin in the model diatom *Phaeodactylum tricornutum*. Proteins encoded by poorly characterized iron-sensitive genes were identified including three that are expressed from a chromosomal gene cluster. Two of them showed unambiguous colocalization with phytotransferrin adjacent to the chloroplast. Further phylogenetic, domain, and biochemical analyses suggest their involvement in intracellular iron

processing. Proximity proteomics holds enormous potential to glean new insights into iron acquisition pathways and beyond in these evolutionarily, ecologically and biotechnologically important microalgae.

Keywords

diatoms, iron biogeochemistry, phytotransferrin, endocytosis, metal trafficking, chloroplast, proximity proteomics, APEX2

Glossary

APEX	engineered ascorbate peroxidase (APX)
CREG	cellular repressor of E1A-stimulated genes
DAB	3,3'-diaminobenzidine
EYFP	enhanced yellow fluorescent protein
FMN	flavin mononucleotide
Fe'	dissolved iron pool (all unchelated iron species)
MS	mass spectrometry
HNLC	high-nutrient, low-chlorophyll
ISIP	iron starvation-induced protein
PBS	phosphate-buffered saline
pTF	phytotransferrin
RT	room temperature
TEM	transmission electron microscopy
TM	transmembrane
TMT	tandem mass tag
UTR	untranslated region
V-ATPase	vacuolar-type H^+-ATPase
WT	wild type

Introduction

Iron (Fe) likely played an important role in the origin of life (***Bonfio et al., 2017***; ***Jin et al., 2018***; ***Kitadai et al., 2019***; ***Nitschke et al., 2013***) and is fundamental in extant metabolisms acting as a cofactor in enzymes involved in DNA synthesis, cellular respiration, nitrogen fixation and photosynthesis (***Crichton, 2016***). Early anoxic oceans were rich in readily bioavailable ferrous (Fe(II)) iron, but with the rise of oxygenic photosynthesis and the Great Oxygenation Event (GOE) in the Paleoproterozoic ~2.3 billion years ago, followed by the Neoproterozoic Oxygenation Event (NOE) ~1.5 billion years later, most of it oxidized into insoluble ferric (Fe(III)) minerals which are not bioavailable (***Camacho et al., 2017***; ***Knoll et al., 2017***; ***Och and Shields-Zhou, 2012***). Conceivably, these large global shifts in ocean chemistry could have had a major role in driving the evolution of novel Fe uptake mechanisms.

Fe availability limits primary productivity in high nutrient, low chlorophyll (HNLC) regions which cover ~25% of the modern ocean habitat (***Boyd and Ellwood, 2010***) as demonstrated by numerous large-scale Fe fertilization experiments (***Martin et al., 1994***; ***de Baar et al., 2005***) and natural Fe upwelling events (***Arydna et al., 2019***) invariably resulting in diatom-dominated phytoplankton blooms. Dissolved Fe(III) in marine environments is primarily found complexed to organic ligands (***Hutchins and Boyd, 2016***; ***Tagliabue et al., 2018***) such as bacterially-produced siderophores (***Boiteau et al., 2016***) and hemes (***Hogle et al., 2014***), but low (pM) amounts of unchelated labile Fe(III), Fe', serve as a crucial source of iron for eukaryotic phytoplankton, particularly diatoms (***Morel et al., 2008***). Diatoms, which are responsible for 45% of global primary production (***Benoiston et al., 2017***), have convergently evolved phytotransferrin (pTF), which serves as the basis of a high-affinity nonreductive carbonate ion-coordinated ferric iron binding and acquisition pathway (***McQuaid et al., 2018***; ***Morrissey et al., 2015***). Phytotransferrin sequences have a broad taxonomic distribution and are abundant in marine environmental genomic datasets (***Bertrand et al., 2015***; ***Marchetti et al., 2012***).

Phytotransferrins are estimated to have emerged concurrently with the NOE exemplifying the link between large environmental changes and molecular innovation (**McQuaid et al., 2018**). pTF (Ensembl ID: Phatr3_J54465) from the model diatom *Phaeodactylum tricornutum* (**Bowler et al., 2008**) was first identified as an iron-sensitive transcript and named ISIP2a (iron starvation induced protein 2a) (**Allen et al., 2008**), although its expression levels remain relatively high in iron-replete conditions as well (**Smith et al., 2016**). pTF localizes to the cell surface and intracellular puncta, presumably endosomal vesicles (**McQuaid et al., 2018**), which is further supported by the presence of an endocytosis motif in the C-terminus of the protein (**Lommer et al., 2012**). In diatom pTF, dissolved ferric iron is coordinated synergistically with carbonate (CO_3^{2-}) anion. Thus, decline in seawater carbonate concentrations due to ocean acidification caused by elevated atmospheric CO_2 may negatively impact this prevalent iron uptake system in diatoms and other marine phytoplankton (**McQuaid et al., 2018**). Diatom pTF is phylogenetically related to Fe-assimilation domain-containing homologs, such as the recently characterized phytotransferrin from the marine picoalga *Ostreococcus tauri* (**Scheiber et al., 2019**); and represent functional and evolutionary analogs of transferrins (**McQuaid et al., 2018**), iron delivery proteins found in multicellular organisms that employ the same mode of binding (**Cheng et al., 2004**).

While accessing and binding dilute dissolved ferric iron on the cell surface is important, subsequent internalization and delivery to target sites within complex cellular milieu are also critical. Fe(III) is highly insoluble and conversion between Fe(III) and Fe(II) can lead to toxic reactive oxygen species causing damage to proteins, lipids, and nucleic acids (**Cheng et al, 2004**). Precise control of iron internalization and intracellular trafficking in either of its redox states is thus crucial for maintenance of cellular homeostasis (**Philpott and Jadhav, 2019**; **Wang and Pantopoulos, 2011**). Human iron-laden transferrin (Tf) bound to transferrin receptor is internalized via clathrin-mediated endocytosis. Endosome acidification leads to carbonate protonation and Tf conformational change resulting in iron release. Fe(III) is then reduced by six-transmembrane epithelial antigen of prostate 3 (STEAP3), exported to the cytoplasm through divalent metal transporter 1 (DMT1) or ZRT/IRT-like protein 8/14 (ZIP8/14), and

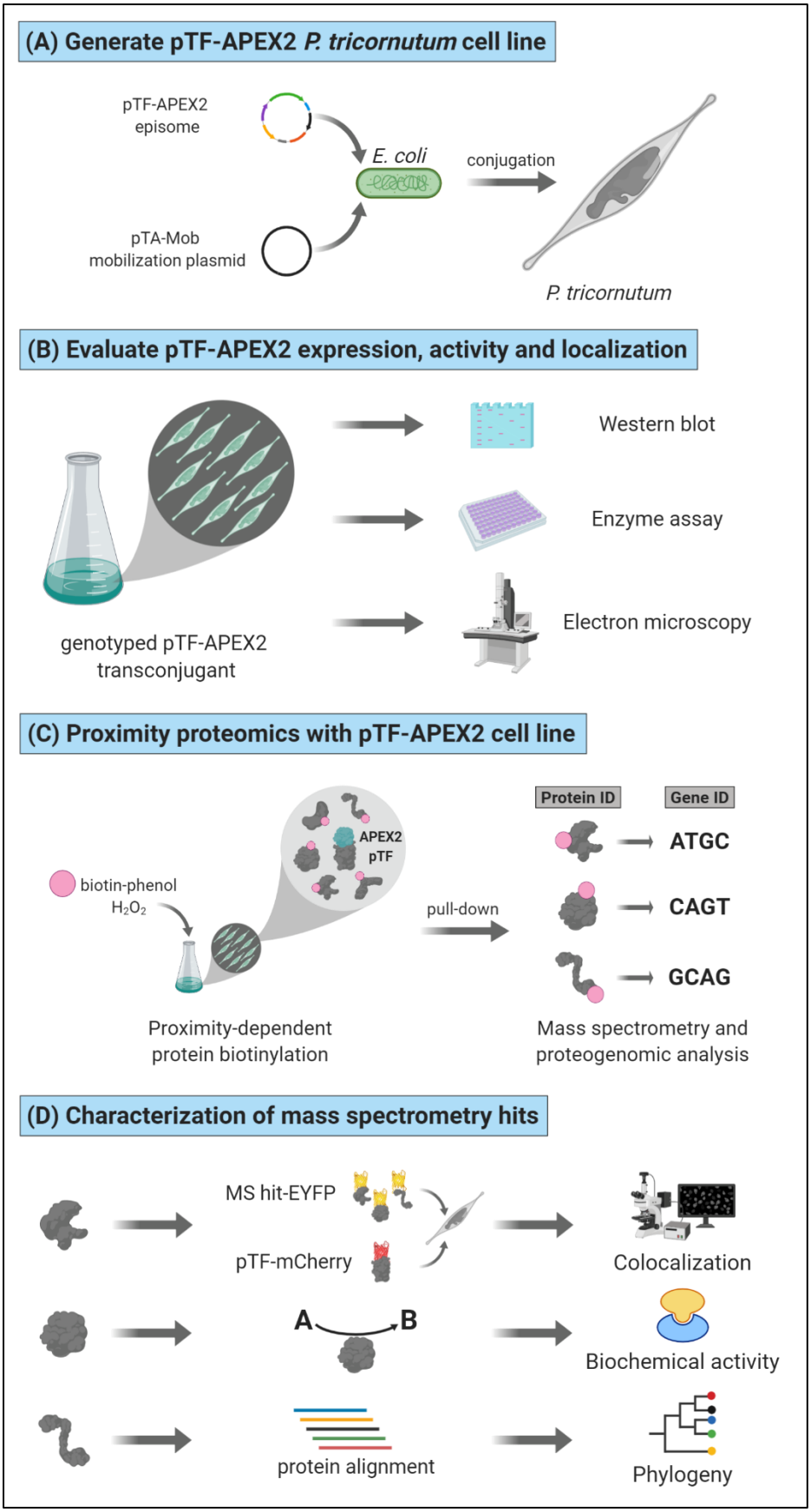
offloaded to iron chaperones for distribution to cellular iron sinks (**Bogdan et al., 2016**; **Eckenroth et al., 2011**; **Philpott and Jadhav, 2019**; **Wang and Pantopoulos, 2011**). In contrast, proteins which conduct key endosomal processes and biochemical transformations and mediate subsequent intracellular allocation of internalized Fe(III) downstream of diatom pTF are unknown.

To investigate the proximal proteomic neighborhood of diatom pTF and thus the degree of mechanistic resemblance between transferrin and phytotransferrin pathways, we turned to APEX2, an engineered heme-containing ascorbate peroxidase (**Lam et al., 2015**) that functions as a dual probe for electron microscopy (**Martell et al., 2017**) and proximity proteomics (**Hung et al., 2016**). When genetically fused to a protein of interest, the ~27 kDa APEX2 enzyme permits spatially resolved proteomic mapping by oxidizing biotin-phenol to short lived (<1 ms) phenoxyl radicals which can covalently react with electron-rich amino acids (tyrosine, but likely also tryptophane, histidine, and cysteine) on the surface of nearby endogenous proteins. Tagged proteins can then be isolated by purification with streptavidin beads and analyzed using mass spectrometry (MS). The “biotinylation radius” in APEX2 experiments should be seen as a “contour map”, a “probability gradient”, where the likelihood of tagging decreases with distance away from APEX2-tagged protein of interest (**Hung et al., 2016**; **Lam et al., 2015**). Target hits can therefore include both strong and stable as well as weak and transient direct interactors, and also vicinal, juxtaposed proteins not interacting with the bait (**Lundberg and Borner, 2019**). Replacing biotin-phenol with diaminobendizine (DAB) enables high-resolution electron microscopy experiments (**Martell et al., 2017**; **Martell et al., 2012**) making APEX2 a powerful bifunctional probe. APEX2 has thus far been used to investigate a variety of cellular compartments and processes including the mitochondrial nucleoid-associated proteome (**Han et al., 2017**), post-Golgi vesicle trafficking (**Otsuka et al., 2019**), stress granules (**Markmiller et al., 2017**), protein occupancy at defined genomic loci (**Myers et al., 2018**), chromatin interactions (**Qiu et al., 2019**), lipid droplets (**Bersuker et al., 2018**), Golgi-localized proteolysis (**Hwang et al., 2016**), subcellular localization of RNAs (**Fazal et al., 2019**), the ciliary membrane-associated proteome (**Kohli et al., 2017**), GPCR signaling (**Paek et al., 2017**), bacterial

pathogen inclusion membranes (***Dickinson et al., 2019***), and finally also endocytosis (***Del Olmo et al., 2019***). These studies, conducted in diverse model systems including mammalian cell culture, fruit fly, yeast and bacteria, testify to the breadth of APEX2 applications, and associated insights and advances that have been achieved in a relatively short time since the technique was first described (***Lam et al., 2015***).

Here, we present further evidence for *P. tricornutum* phytotransferrin pTF endocytosis, demonstrate correct subcellular localization and activity of its APEX2 fusion, identify almost 40 proximal proteins in proximity proteomics experiments conducted in quintuplicate, and, finally, provide initial characterization, using dual fluorophore protein tagging fluorescence microscopy, phylogenetic, and biochemical analyses, for three proteins (pTF.CREGr, pTF.CatCh1, pTF.ap1) believed to act in association with endocytosed pTF downstream of iron binding at the cell surface. Bacterially-expressed recombinant pTF.CREGr displays flavin reductase activity *in vitro* suggesting this protein may be involved in intracellular ferric, Fe(III), iron reduction, while bioinformatic interrogation of pTF.CatCh1 and its localization indicate it may be a chloroplast-associated metallochaperone.

The overall outline of our work is shown in ***Figure 1***. To the best of our knowledge, this study represents the first application of APEX2 in a marine microbial model system. We conclude with a vision board summarizing outstanding questions in diatom cell biology that could immediately benefit from this and related chemical biology approaches.



175

Figure 1. Functional APEX2-enabled proteogenomics with pTF (phytotransferrin/ISIP2a/Phatr3_J54465) in *Phaeodactylum tricornutum*. (A) pTF-APEX2 encoding episome is introduced into *P. tricornutum* cells using bacterial conjugation. (B) Resulting transconjugants are genotyped and evaluated for fusion protein expression. APEX2 activity and fusion localization are then confirmed with an enzymatic assay and electron microscopy, respectively. (C) In the proximity proteomics experiment, pTF-APEX2 expressing cell line is supplemented with biotin-phenol and hydrogen peroxide, reaction quenched and cells lysed. Cell lysate is then subjected to streptavidin pull-down, proteins analyzed with mass spectrometry (MS), peptides mapped to a *P. tricornutum* proteome database, and corresponding genes identified. (D) Interesting MS hits are further evaluated experimentally (e.g., for colocalization with the bait protein (i.e. pTF) and/or for predicted biochemical activity) as well as bioinformatically. Created with BioRender.com.

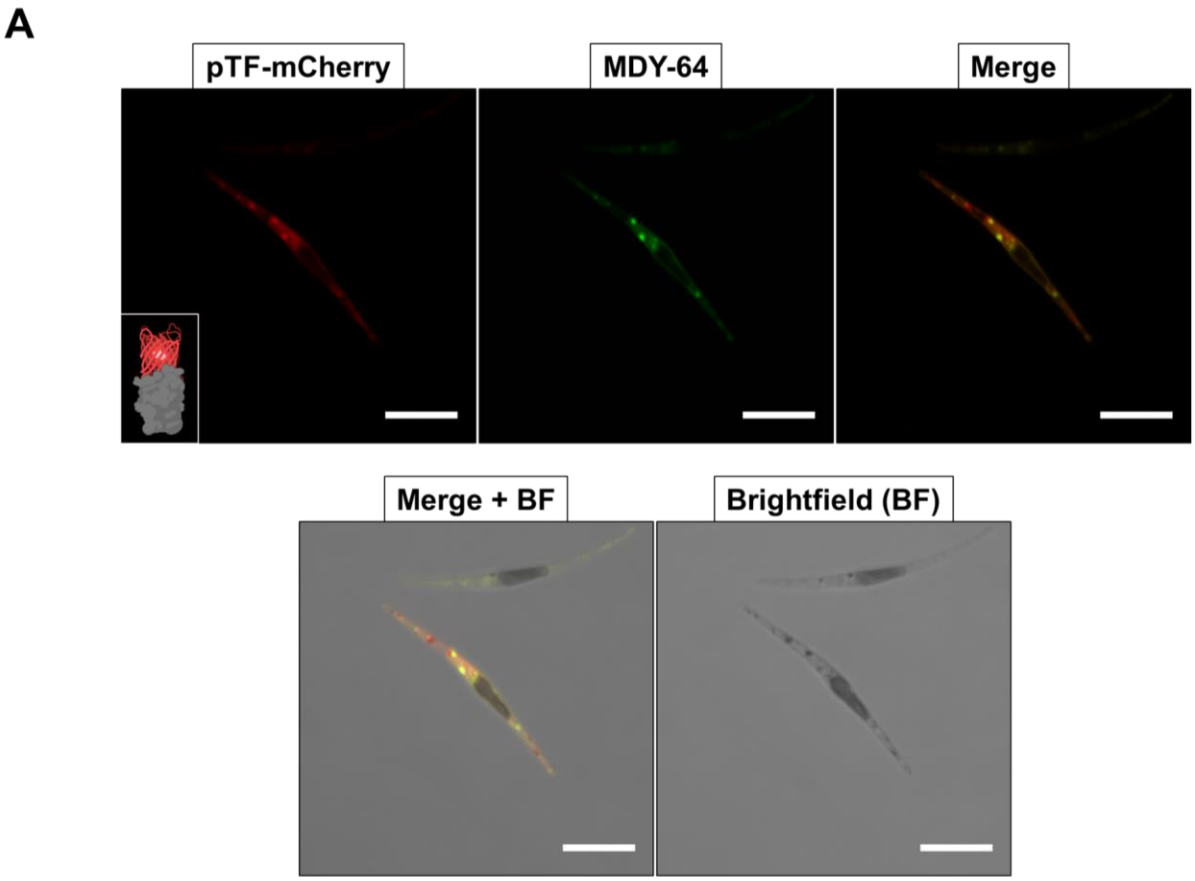
Results

Phytotransferrin (pTF) is localized to endosome-like vesicles

To further examine pTF localization and possible occurrence within intracellular vesicles, a pTF-mCherry encoding episome was conjugated (***Karas et al., 2015***; ***Diner et al., 2016***) into *ΔpTF P. tricornutum* cells (***McQuaid et al., 2018***). After labeling the fluorescent transconjugant cell line with 100 µM of the membrane dye MDY-64 for 10 min at room temperature, vesicles with colocalized mCherry and MDY-64 signal were observed (***Figure 2A***).

Phytotransferrin-APEX2 fusion proteins are enzymatically active *in vivo*

To generate pTF-APEX2 expressing diatom cell lines, an episome encoding pTF (Phatr3_J54465) with C-terminal APEX2 was assembled and conjugated into WT *P. tricornutum* (***Figure 2B***). Considering the predicted pTF domains (***Figure 2—figure supplement 1A***), this likely resulted in APEX2 facing the cytosol both at the cell surface and once internalized into vesicles. Western blot with pTF-specific antibodies (***McQuaid et al., 2018***) confirmed the ~84.7 kDa fusion protein expression in 4 out of 5 tested cell lines. The protein was largely present in the insoluble cell lysate fraction (***Figure 2C***), possibly further indicative of its membrane localization. Amplex UltraRed, a highly sensitive APEX2 substrate (***Hung et al., 2016***), was used to assay live pTF-APEX2 expressing cells. A resorufin (reaction product) signal up to 4- and 40-fold above WT background was observed in experiments performed at room temperature (data not shown) and on ice (***Figure 2D***; ***Supplementary file 2***), respectively, indicating active APEX2 with incorporated heme. Resorufin was also directly visualized by confocal microscopy (***Figure 2—figure supplement 1B***) and a strong cytosolic signal not tightly localized to the expected site of origin, similar to previous reports (***Martell et al., 2012***), was observed. APEX2 was active only in the presence of both Amplex UltraRed and hydrogen peroxide (***Figure 2—figure supplement 1C***) implying that endogenous H_2O_2 levels are not sufficient to drive APEX2-catalyzed reactions and that the overall cell surface and intracellular milieu in *P. tricornutum* is permissive to the APEX2 catalytic cycle.

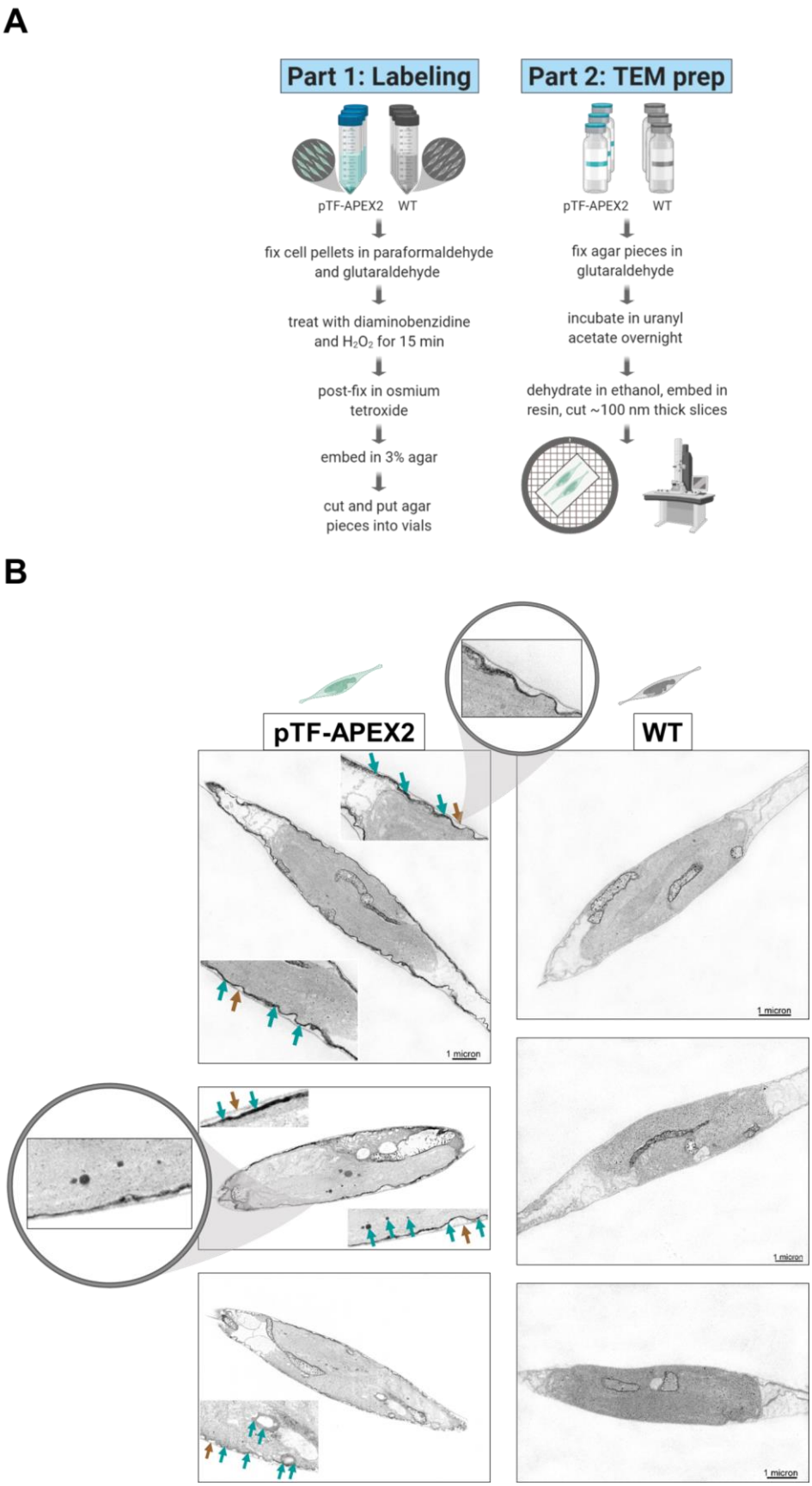


218

Figure 2. Association of pTF with intracellular vesicles and demonstration of APEX2 activity in *Phaeodactylum tricornutum*. (A) pTF-mCherry colocalizes with the membrane dye MDY-64 on the cell surface and within intracellular vesicles. The fusion protein was expressed under the *pTF* promoter and terminator in a *ΔpTF* genetic background. Cells were stained with 100 µM MDY-64 for 10 min at room temperature. Scale bar is 10 µm. Protein fusion scheme created with BioRender.com. (B) Top: *pTF-APEX2* flanked by nitrate-inducible *NR* (nitrate reductase) promoter, terminator and UTRs. Linker between pTF and APEX2: KGSGSTSGSG. Bottom: Schematic of the pTF-APEX2 expressing episome. Tetracycline (Tet^R) and ampicillin (Amp^R) resistance genes for *E. coli* selection, bacterial replication origin, yeast centromere (*CEN6-ARSH4-HIS3*) for episome maintenance, phleomycin/zeocin (*ShBle*) resistance gene for *P. tricornutum* selection, origin of conjugative transfer (*oriT*). Plasmid map created with BioRender.com. (C) Anti-pTF Western blot confirming fusion protein expression (~84.7 kDa) (lanes s1–s4; left: soluble cell lysate fraction, right: insoluble cell lysate fraction) in WT *P. tricornutum* background. Teal circle: APEX2, grey oval: pTF. (D) Left: pTF-APEX2 expressing, but not WT, cells convert APEX2 substrate Amplex UltraRed (50 µM) into a colored product resorufin in the presence of 2 mM H_2O_2. Right: >40-fold higher resorufin signal was observed in supernatants from pTF-APEX2 expressing cells than those from WT cells. Triplicate cultures from one WT and one pTF-APEX2 expressing cell line (strain "s2"). Standard deviation is shown. *P. tricornutum* cartoons created with BioRender.com.

pTF-APEX2 is localized to the cell surface and intracellular vesicles

To confirm pTF-APEX2 is localized to the cell surface and intracellular vesicles similar to mCherry fusions, pTF-APEX2 expressing cells (strain "s2", ***Figure 2***) were treated with 25 mM 3,3'-diaminobenzidine (DAB) in the presence of 3 mM H_2O_2 for 15 min on ice (***Figure 3A***). This reaction leads to DAB polymerization and local precipitation around APEX2 that can be stained with osmium tetroxide and visualized with transmission electron microscopy. Cells were embedded in a 3% agar matrix to prevent losses during numerous washing steps (***Figure 3—figure supplement 1A***). Tightly localized signal was observed on the cell membrane and in intracellular vesicles indicating that pTF-APEX2 fusion trafficked to the correct subcellular sites (***Figure 3B***). Additionally, we observed mitochondrial signal in WT and pTF-APEX2 cell lines subjected to the DAB reaction (***Figure 3B***; ***Figure 3—figure supplement 1B***). Analysis of the *P. tricornutum* proteome revealed 8 peroxidases with Arg38, His42, His163 and Asp208; catalytic residues that are conserved across all identified ascorbate peroxidases including soybean APX (***Raven et al., 2003***) and its engineered version APEX2 (***Figure 3—figure supplement 1C***). Three of these contain proline in place of alanine at position 134 relative to APEX2, a substrate-binding loop mutation rendering APEX2 highly active (***Martell et al., 2015***), and another putative peroxidase is predicted to localize to mitochondria (***Figure 3—figure supplements 1D, 1E***). APEX2 was inactivated in iron-limiting conditions (40 nM total Fe) and could be reactivated by supplementing fixed cells with 10 µM hemin chloride for 3 hrs (data not shown), but this substantially increased background in the Amplex UltraRed assay. Therefore, subsequent proximity labeling experiments were carried out on cells growing in iron-replete conditions.



260

Figure 3. pTF-APEX2 localizes to the cell membrane and intracellular vesicles. (A) Summary of the transmission electron microscopy protocol. Briefly: WT and pTF-APEX2 expressing cells were fixed, treated with diaminobenzidine (DAB), post-fixed with osmium tetroxide, embedded in agar, negatively stained with uranyl acetate, dehydrated, embedded in resin, and visualized. Created with BioRender.com. (B) Expected cell surface and intracellular pTF-APEX2 localization. Teal and brown arrows point to APEX2-induced signal and cell wall, respectively. Zoomed in is the cell periphery clearly showing cell membrane, and not cell wall, is occupied by the fusion protein (top left image). pTF-APEX2-containing vesicles of various sizes were seen (bottom two images on the left; acquired with backscatter scanning electron microscopy). Zoomed in are cell membrane and vesicles with pTF-APEX2 (middle left image). Mitochondrial signal in both WT and pTF-APEX2 expressing cells is likely due to endogenous (mitochondrial) APEX2-like peroxidases. Scale bar is 1 µm. *P. tricornutum* cartoons created with BioRender.com.

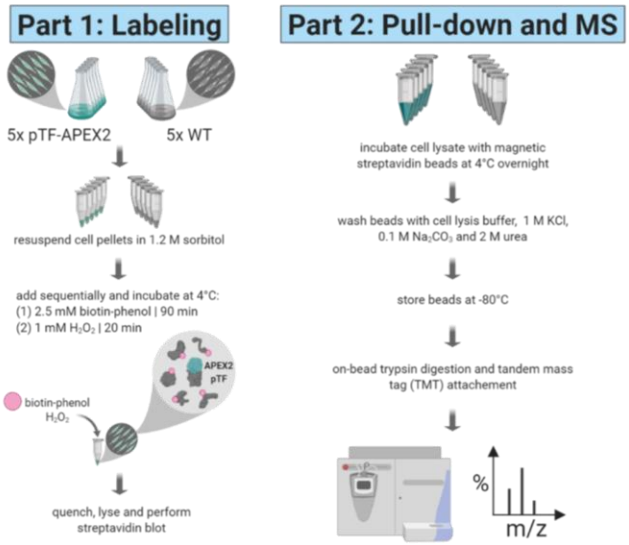
Identification of the proximal phytotransferrin proteome with biotin-phenol labeling

To identify proteins proximal to pTF, quintuplicate cultures from one WT and one pTF-APEX2 expressing cell line (strain "s2", ***Figure 2***) were grown to mid- to late-exponential phase and supplemented with biotin-phenol and hydrogen peroxide, following the APEX2 proximity labeling protocol developed for yeast (***Hwang and Espenshade, 2016***; ***Hwang et al., 2016***). Incubation with hydrogen peroxide was extended to 20 min to mirror Amplex UltraRed assay and DAB reaction conditions (***Figure 4A***). Increasing the biotin-phenol concentration from 0.5 mM, usually used in mammalian cells, to 2.5 mM and exposing cells to osmotic stress with 1.2 M sorbitol (both are critical for efficient labeling in yeast) was necessary to detect enrichment of biotinylated proteins in experimental samples (***Figure 4B***). This result confirmed our hypothesis that the lack of heavy silicification in *P. tricornutum* (***Francius et al., 2009***; ***Tesson et al., 2009***), making its cell wall composition and cell membrane permeability likely similar to those in yeast, would permit labeling. Streptavidin pull-downs were then performed with clarified cell lysates followed by tandem mass tag (TMT)-based quantitative proteomics. WT and pTF-APEX2 proteomic replicates, with the exception of one WT and one pTF-APEX2 condition, formed two distinct clusters (***Figure 4—figure supplement 1***; ***Supplementary file 3***) indicating minimal technical variability. 38 statistically significant proteins ($P \leq 0.05$) with APEX2/WT ratios of at least 1.5 were identified (***Figure 4B***). These ratios were obtained from average total peptide counts across quintuplicates (***Supplementary file 4***). Predicted endogenous biotinylated proteins were also detected and had APEX2/WT ratios close to 1, thus serving as an intrinsic pull-down control (***Supplementary file 1—Table S1***; ***Supplementary file 4***). Some background enrichment in WT cells was likely due to endogenous APEX2-like peroxidases and one would thus expect mitochondrial proteins with APEX2/WT ratios close to 1 to be present in our MS dataset. Indeed, at least three were detected: mitochondrial chaperonin CPN60 (Phatr3_J24820, UniProt ID: B7FQ72, APEX2/WT = 0.99), mitochondrial import receptor subunit TOM70 (Phatr3_J47492, UniProt ID: B7G3J4, APEX2/WT = 0.91), and acetyl-CoA dehydrogenase (Phatr3_J11014, UniProt ID: B7FTR6, APEX2/WT = 1.23) (***Supplementary file 4***). 14 proteins were detected

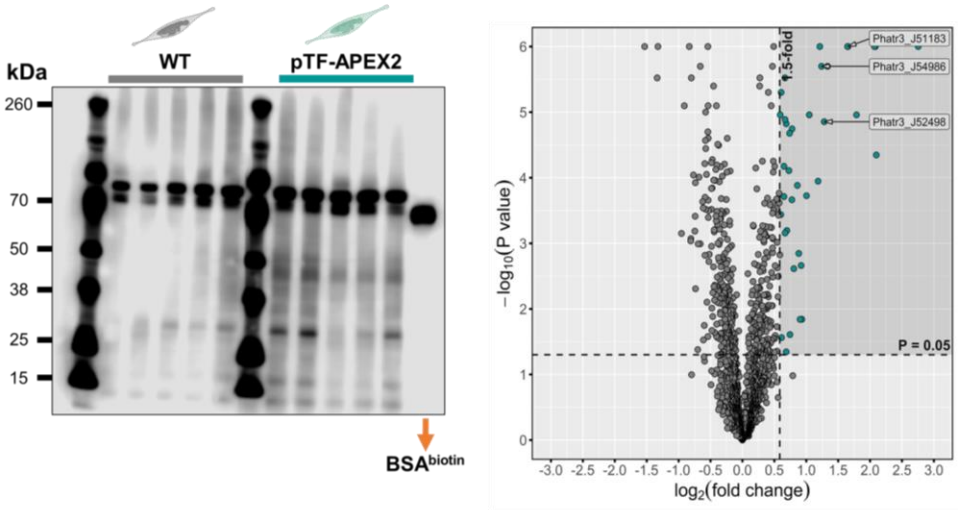
with an APEX2/WT ratio of at least 2. Of these 14 proteins, 9 are known to be transcriptionally sensitive to iron availability; this is also the case for an additional 5 proteins with APEX2/WT ratio of at least 1.5 (***Supplementary file 4***) (***Smith et al., 2016***). We note that pTF was present, but not enriched, in our pTF-APEX2 proteomics samples (***Supplementary file 4***). One possible explanation for this result is that pTF does not have surface exposed amino acid residues that are permissive to biotinylation by APEX2-generated phenoxyl radicals, something only pTF structure would reveal; and so its presence in all of our pull-down samples could be due to unspecific binding to streptavidin beads.

Some of the most apparent biologically interesting hits are summarized in ***Figure 4C***. Considering iron is involved in all the major photosynthetic complexes (***Rochaix, 2011***), we asked if some of the proteins are perhaps predicted to be localized to the chloroplast. Indeed, five are predicted to be chloroplastic, albeit with low confidence, by ASAFind (***Gruber et al., 2015***), including two (Phatr3_J51183 and Phatr3_J54986) that are part of a gene cluster on chromosome 20 and two (Phatr3_J41423 and Phatr3_J55031) that are known to associate with the *P. tricornutum* chloroplast (***Allen et al., 2012***; ***Kazamia et al., 2019***; ***Figure 4C***). Further details on how ASAFind works and which proteins are given a low confidence chloroplastic prediction are presented in ***Gruber et al., 2015*** and Materials and Methods, respectively. We elaborate on the potential role for these proteins in intracellular iron trafficking in Discussion.

A



B



C

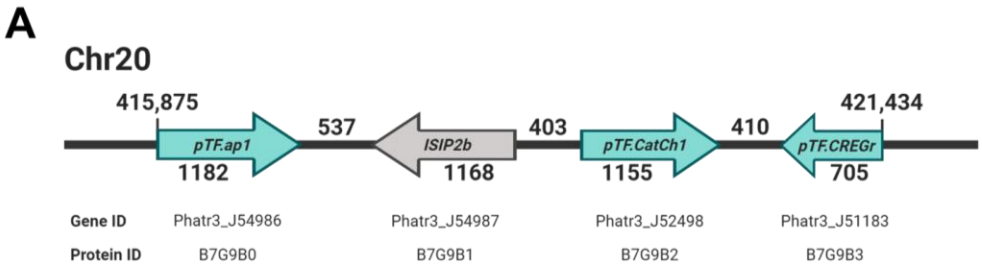
UniProt ID	APEX2/WT ratio	Gene ID	Iron-sensitive transcript	Chloroplast localization/ confidence	Annotation
B7GE67	6.75	Phatr3_J41423	yes	yes/low*	fructose-biphosphate aldolase FBAC5
B7GA90	4.20	Phatr3_J55031	yes	yes/low**	iron starvation induced protein (ISIP) 1
B5Y488	3.46	Phatr3_J46929	yes	no	siderophore-bound iron assimilation protein
B7G9B3	**3.15**	**Phatr3_J51183**	**yes**	**yes/low**	**CREG1-like protein**
B5Y3Y4	3.13	Phatr3_J54656	yes	no	heat shock protein (HSP)
B7G9B2	**2.43**	**Phatr3_J52498**	**yes**	**no**	**cell surface protein**
B7G9B0	**2.36**	**Phatr3_J54986**	**yes**	**yes/low**	**cell surface protein**
B7G195	2.32	Phatr3_J46448	yes	no	heat shock protein (HSP)
B7G9L9	1.82	Phatr3_J30139	yes	no	small GTPase
B7GBX2	1.71	Phatr3_J23497	yes	no	ATPase
B7FNT2	1.62	Phatr3_J42574	no	yes/low	α-carbonic anhydrase
B7GDH4	1.59	Phatr3_J41172	no	no	calreticulin/mobilferrin
B7G5Y2	1.57	Phatr3_J5651	no	no	14-3-3-like protein
B7FR50	1.52	Phatr3_J43251	no	no	small GTPase
B7G162	1.51	Phatr3_J27923	no	no	vacuolar-type H+-ATPase subunit A

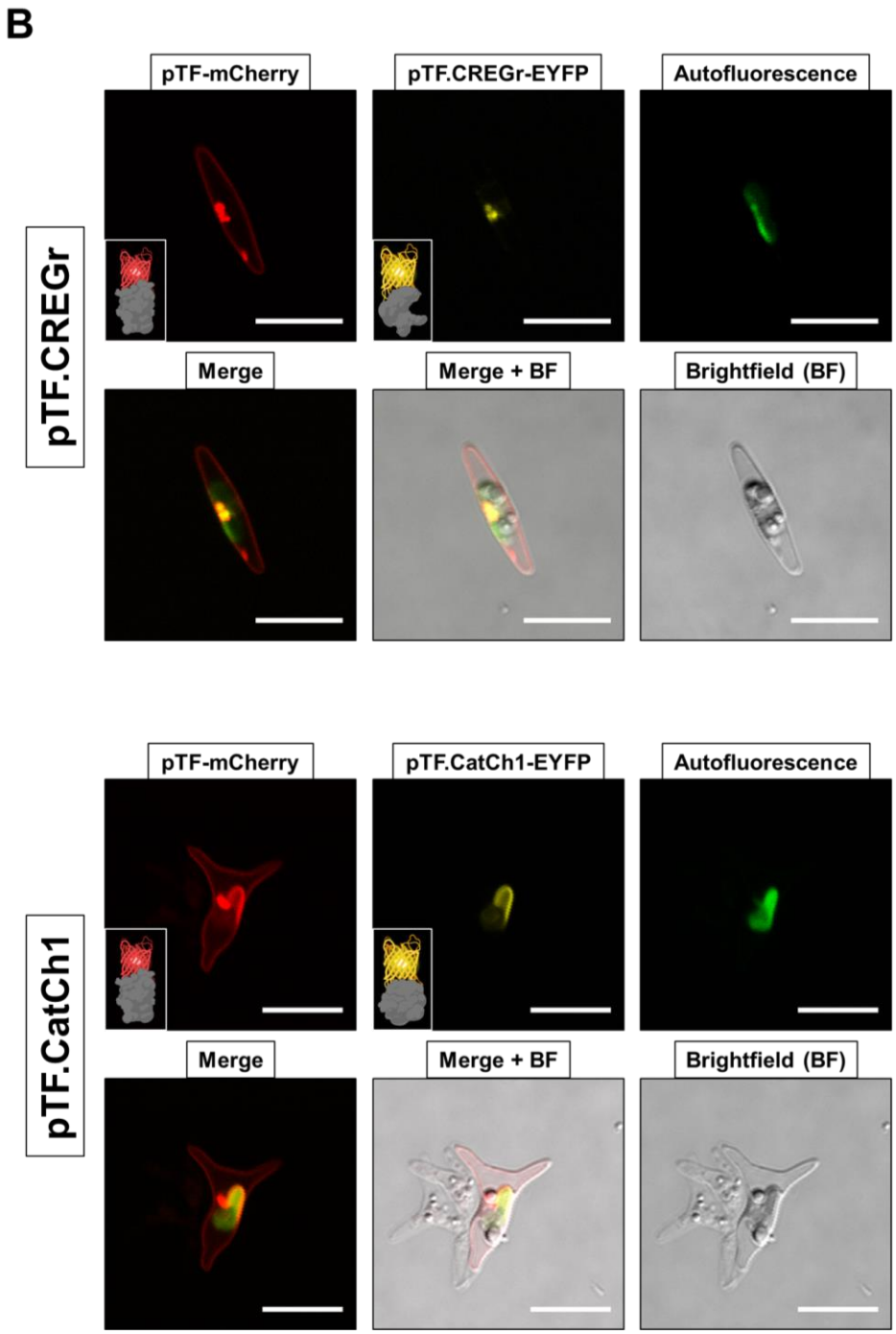
325

Figure 4. Proximity proteomics with APEX2 reveals candidate proteins involved in phytotransferrin (pTF) endocytosis in *Phaeodactylum tricornutum.* (A) Summary of the proximity proteomics experiment. Briefly: WT and pTF-APEX2 expressing cells were chilled on ice, pelleted, treated with 1.2 M sorbitol, supplemented with 2.5 mM biotin-phenol and 1 mM H_2O_2. The labeling reaction was quenched, cells were lysed, and evaluated for biotin enrichment. Supernatants were then subjected to streptavidin pull-down followed by quantitative mass spectrometry using tandem mass tags (TMT). Created with BioRender.com. (B) Left: Enrichment of biotinylated proteins over WT background was observed with streptavidin blot; ~66.5 kDa biotinylated BSA control, equal loading. The most prominent band in all samples is likely the endogenous biotin-containing propionyl-CoA carboxylase (Phatr3_J51245, ~72.5 kDa). Right: Vulcano plot of quantitative MS data highlighting proteins with APEX2/WT ratio of at least 1.5 and P-value ≤ 0.05 (shaded area with teal data points). Proteins from a known iron-sensitive gene cluster located on chromosome 20 are highlighted. *P. tricornutum* cartoons created with BioRender.com. (C) 14/38 (10 shown here) MS hits are proteins with iron-sensitive transcripts. Proteins we tagged with EYFP and co-expressed with pTF are bolded. Chloroplast localization and the associated prediction confidence were determined with SignalP 4.1 and ASAFind (version 1.1.7). *Experimentally shown to localize to the pyrenoid—a RuBisCO-containing suborganelle— in the interior of the *P. tricornutum* chloroplast (***Allen et al., 2012***). **Experimentally shown to be localized adjacently to the chloroplast (***Kazamia et al., 2019***).

Proteins encoded by an iron-sensitive gene cluster on chromosome 20 colocalize with pTF

We focused on three proteins among the top MS hits—Phatr3 IDs: J51183 (hereafter: pTF.CREGr; CREG-like protein with likely Reductive function), J52498 (hereafter: pTF.CatCh1; Chloroplast-AssociaTed CHaperone 1) and J54986 (hereafter: pTF.ap1; pTF-Associated Protein 1)—that are expressed from a previously identified iron- (***Allen et al., 2008***) and silicon-sensitive gene cluster (***Sapriel et al., 2009***) on chromosome 20 (***Figure 5A***). One additional protein (Phatr3_J54987 known also as ISIP2b), which is also transcriptionally sensitive to iron and silicon and that we did not detect in our proximity proteomics experiment, is also encoded by this uncharacterized locus. We note that *pTF* is not co-located with these genes, but instead lies on chromosome 7 (genomic location 1,000,053–1,001,833, forward strand). All three genes exhibit a transcriptional profile similar to *pTF* (***Figure 5—figure supplement 1***), two proteins (pTF.CREGr and pTF.ap1) are predicted to localize to the chloroplast (with low confidence as determined by ASAFind), and two (pTF.CatCh1 and pTF.ap1) contain a C-terminal transmembrane domain (***Figure 4C***; ***Supplementary file 1—Table S2***). It has been hypothesized that organellar pH and proteome biochemistry co-evolved (***Brett et al., 2006***); we note the isoelectric points of all three proteins are close to 5 overlapping the expected endosomal pH (***Supplementary file 1—Table S2***). To test whether any of them colocalize with pTF, co-expression episomes were assembled and conjugated into WT *P. tricornutum* cells which resulted in diatom cell lines expressing pTF-mCherry and MS hit-EYFP fusion proteins. Detailed assembly strategy and cell line verification results are presented in ***Figure 5—figure supplement 2A*** and ***Figure 5—figure supplement 2B***, respectively. pTF.CREGr and pTF.CatCh1 colocalized with pTF in the chloroplast vicinity or on the chloroplast margin, respectively (***Figure 5B***). Colocalization of pTF.ap1 with pTF close to the chloroplast was also evident, though somewhat less precise (***Figure 5—figure supplement 3***). Imaging conditions were optimized with mCherry and Venus expressing cell lines for minimal cross-channel bleed-through (***Figure 5—figure supplement 2C***).

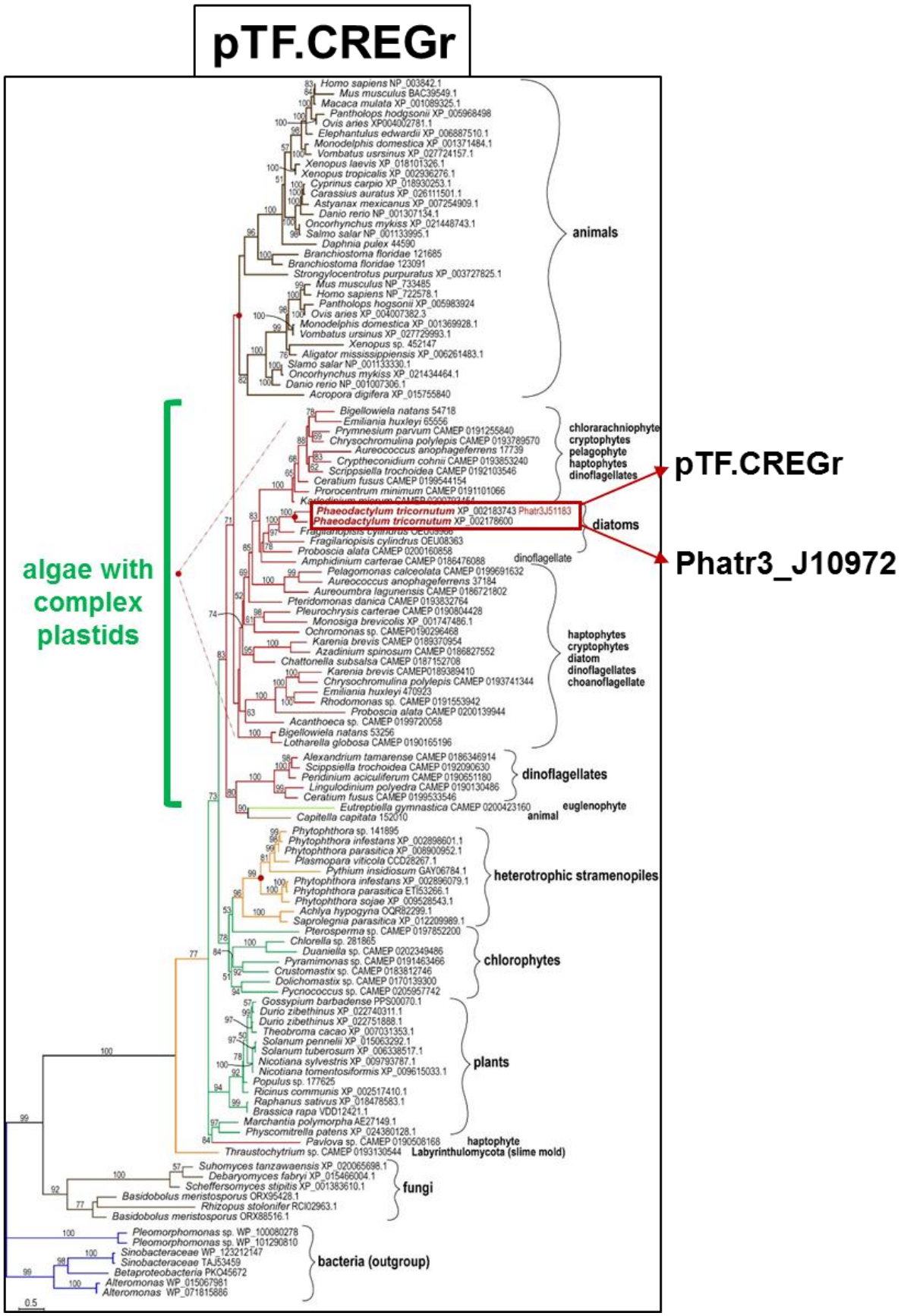




372

373 **Figure 5.** Proteins from a known iron-sensitive locus in *Phaeodactylum tricornutum* colocalize with pTF
374 and chloroplasts. (A) Genes corresponding to a subset of proteomic hits are clustered together on
375 chromosome 20. With exception of ISIP2b, which we did not detect in our proximity proteomics
376 experiment, all proteins were co-expressed with pTF. Corresponding Phatr3 gene IDs and UniProt protein
377 IDs are noted. Numbers indicate base pairs. Created with BioRender.com. (B) Two proteins—pTF.CREGr
378 and pTF.CatCh1—show clear, yet distinct, colocalization with pTF and the chloroplast periphery.
379 pTF.CREGr-EYFP was consistently punctate whereas pTF.CatCh1-EYFP lined the chloroplast margin.
380 pTF-mCherry punctum was almost exclusively positioned next to this “chloroplast lining” pattern. Scale
381 bar is 10 µm. *P. tricornutum* is pleiomorphic which explains why different cell morphologies were
382 observed in pTF.CREGr-EYFP (fusiform morphotype) and pTF.CatCh1-EYFP (triradiate morphotype) cell
383 lines. Protein fusion schemes created with BioRender.com.

384 **Phylogenetic, domain, and biochemical characterization of pTF.CREGr and**
385 **pTF.CatCh1 support a putative role in diatom iron metabolism**
386 To shed light on possible functions of the two poorly annotated proteins that colocalized
387 with pTF and to examine their occurrence in other diatoms and marine phytoplankton
388 beyond *Phaeodactylum tricornutum*, phylogenetic analysis was performed. pTF.CREGr
389 was found conserved across the tree of life (***Figure 6***), but diatom proteins can be seen
390 in a crown group with other complex plastid-containing algae such as cryptophytes,
391 haptophytes, pelagophytes, chlorarachniophytes and dinoflagellates (***Archibald, 2009***;
392 ***Oborník and Füssy, 2018***) suggesting their function may be distinct from those in the
393 sister animal group.
394
395 Interestingly, pTF.CREGr and its paralog Phatr3_J10972 contain a predicted flavin
396 mononucleotide (FMN)-binding domain (***Figure 7A***) hinting at a possibility that
397 pTF.CREGr is involved in oxidation-reduction reactions as opposed to having a more
398 canonical role associated with CREG1 homologs in multicellular organisms (***Ghobrial***
399 ***et al., 2018***). To address potential enzymatic activity, various truncations of His6-tagged
400 pTF.CREGr were expressed in *Escherichia coli* and purified using Co^{2+} immobilized
401 metal affinity chromatography (IMAC) to test solubility (***Figure 7—figure supplement***
402 ***1A***). A construct encoding pTF.CREGr with a 31 amino acid N-terminal deletion leading
403 to a soluble protein was selected for larger-scale purification. Despite slight precipitation
404 and degradation during the purification process, the Co^{2+}-bound fraction exhibited
405 modest yet reproducible flavin reductase activity, as measured by the enzyme-facilitated
406 oxidation of NADPH in the presence of the potential flavin substrates riboflavin and
407 flavin mononucleotide (FMN) (***Figure 7B***; ***Figure 7—figure supplement 1B***;
408 ***Supplementary file 5***). Comparable consumption of NADPH was not seen when either
409 flavins or pTF.CREGr were omitted from the reaction (***Figure 7—figure supplement 2***;
410 ***Supplementary file 5***), suggesting that pTF.CREGr is capable of reducing exogenous
411 flavins.



412

Figure 6. pTF.CREGr is a ubiquitous protein conserved across the tree of life. pTF.CREGr homolog search was performed with the National Center for Biotechnology Information (NCBI) and the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) databases (***Caron et al., 2016***; ***Keeling et al., 2014***). Notably, a clear division between the diatom-containing crown group (algae with complex plastids) and animals can be seen suggesting distinct roles for this protein in multicellular versus unicellular organisms. Red dots indicate predicted gene duplication events. pTF.CREGr clusters with its paralog, Phatr3_J10972. Scale bar: 0.5 substitutions per position.

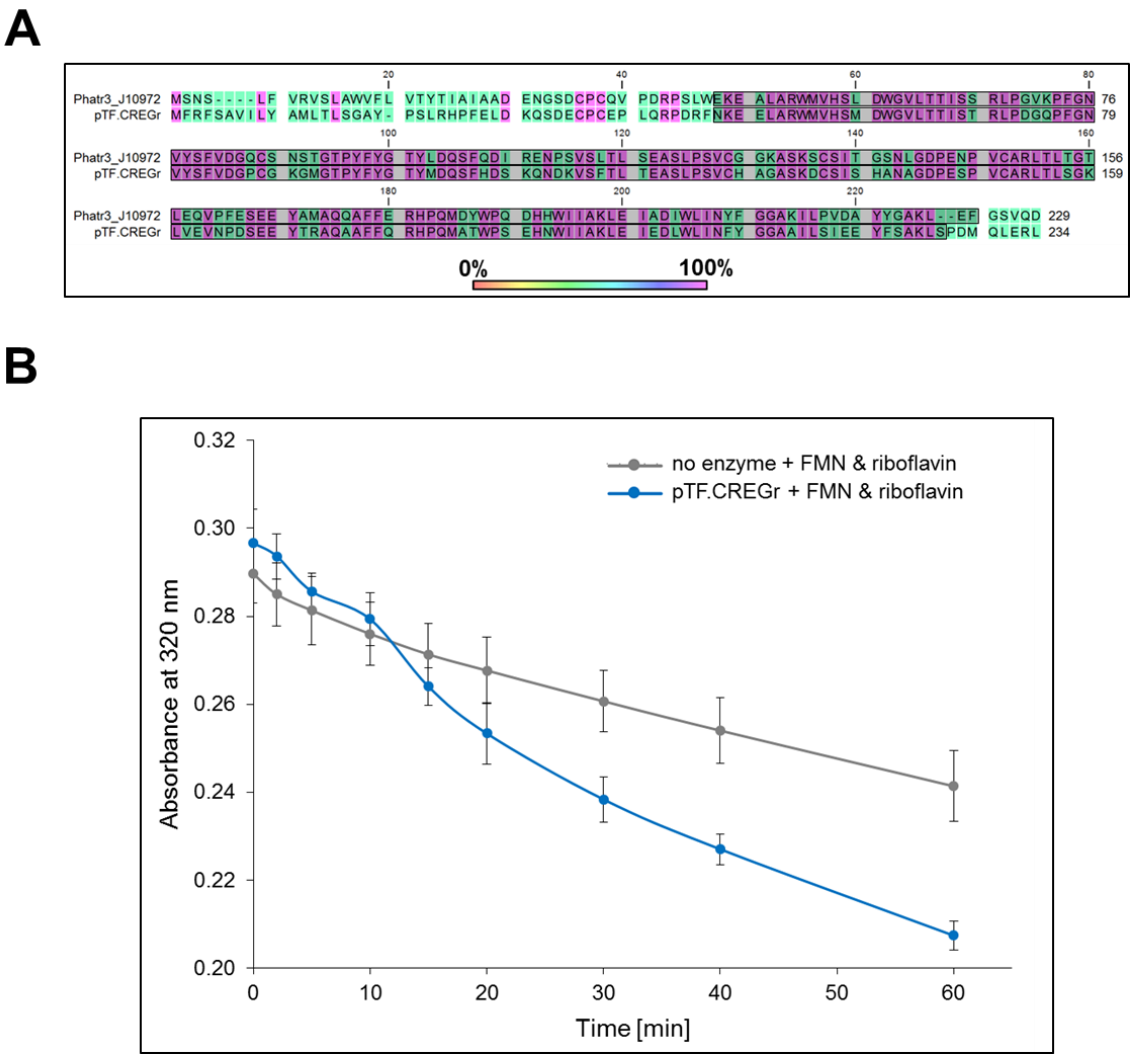
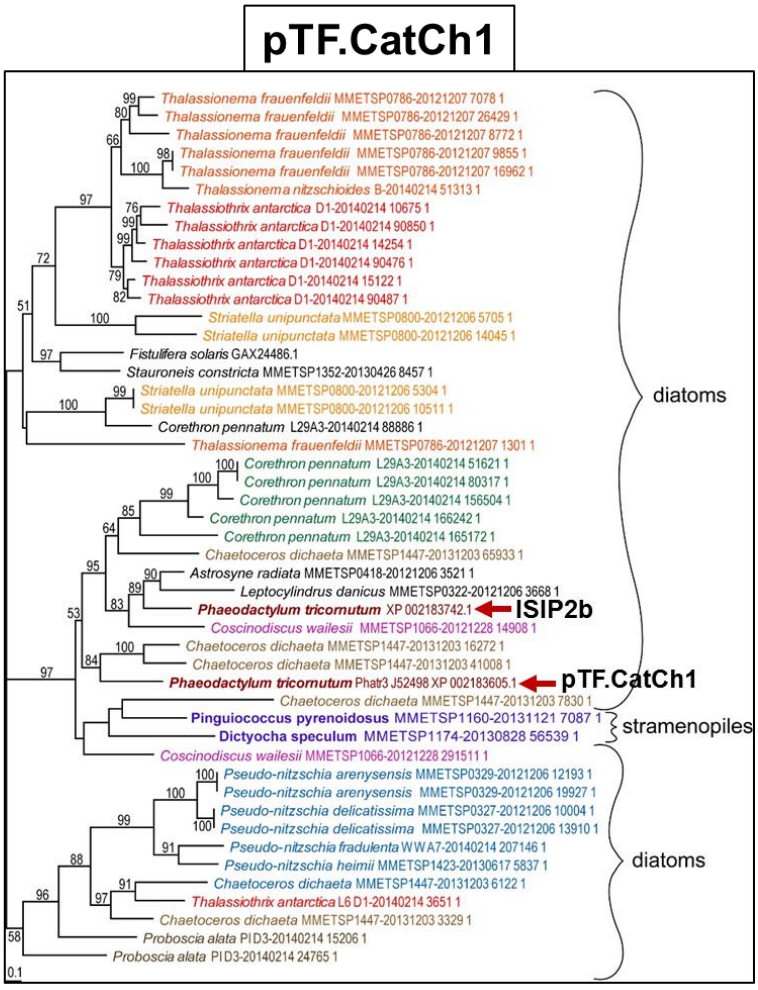


Figure 7. (A) pTF.CREGr and its paralog Phatr3_J10972 contain a putative FMN-binding domain (shaded in gray). Considering the evolutionary link between human CREG1 and oxidoreductases, and the observed colocalization of its diatom homolog with pTF, pTF.CREGr may be a ferric reductase. (B) pTF.CREGr can reduce exogenous flavins. Comparison of NADPH oxidation (measured by decrease in absorbance at 320 nm) by recombinant pTF.CREGr supplemented with flavins (flavin mononucleotide (FMN) and riboflavin) versus a flavin-only, no enzyme treatment. Assays were started by the addition of enzyme or water blank and carried out in triplicate for each treatment. Background 320 nm absorbance of water blanks (approx. 0.186) was subtracted from each replicate. Standard deviation is shown.

In contrast, pTF.CatCh1 homologs were identified predominately in diatoms (***Figure 8A***). This protein is paralogous to ISIP2b which lies in the same chromosome 20 gene cluster. They share 3 CXC and 1 CXXC motifs indicative of metal-binding proteins (***Figure 8B***). One of the CXC motifs and the CXXC motif are conserved across the vast majority of homologs in our phylogenetic analysis (***Figure 8—figure supplement 1A***) and the latter is predicted to be located in the disordered C-terminal domain in pTF.CatCh1 (***Figure 8—figure supplement 1B***) possibly indicating redox-controlled coupling of metal coordination and order-disorder transitions (***Erdős et al, 2019***). The pTF.CatCh1 region between amino acid residues 27 and 226 contains 20 cysteines, further indicative of possible metal binding (***Poole, 2014***), and the predicted transmembrane domain (296–315) is flanked by a polyserine and a short arginine-rich stretch. These observations are synthesized in the putative domain organization schematic shown in ***Figure 8—figure supplement 1C***. Phylogenetic characterization was performed with pTF.ap1 as well and numerous homologs with at least 5 highly conserved motifs were identified in diatoms and other marine microeukaryotes, including chlorophytes, cryptophytes and haptophytes (***Figure 8—figure supplements 2A, 2B***). These pTF.ap1 homologs contain 5 100% conserved cysteines indicating this protein may, similarly to pTF.CatCh1, also coordinate metal ions, possibly iron (2 conserved cysteines are shown in ***Figure 8—figure supplement 2B***).

A



B

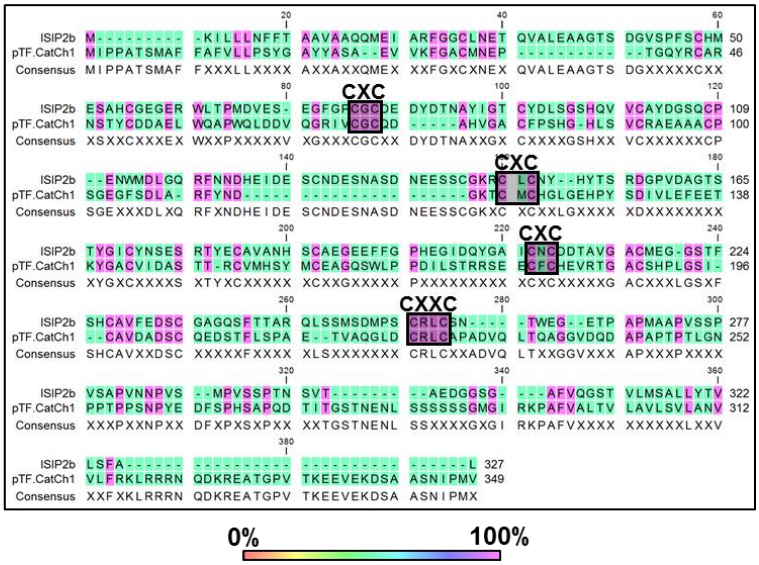


Figure 8. pTF.CatCh1 may be a diatom metallochaperone. (A) pTF.CatCh1 homolog search was performed with NCBI and marine microbial eukaryote (MMETSP) databases and numerous diatom proteins were identified. Notably, pTF.CatCh1 is paralogous to ISIP2b. Scale bar: 0.1 substitutions per position. (B) pTF.CatCh1 and ISIP2b share 3 CXC and 1 CXXC motifs, known divalent metal coordination motifs in proteins such as metallochaperones.

Discussion

ISIP2a (iron starvation induced protein 2a) was previously identified in *P. tricornutum* as a marker for iron (Fe) limitation (**Allen et al., 2008**), subsequently shown to be involved in Fe acquisition (**Morrissey et al., 2015**) and is widespread in ocean phytoplankton communities (**Bertrand et al., 2015**; **Carradec et al., 2018**). In a breakthrough study, **McQuaid et al., 2018** demonstrated that ISIP2a, is a type of transferrin, phytotransferrin (pTF), that is crucial for acquisition of dissolved Fe; a critical micronutrient for cellular biochemistry. **McQuaid et al., 2018** showed that pTF is essential for high affinity iron uptake in diatoms and that carbonate and Fe interact synergistically to control the iron uptake rate. The finding that carbonate ions are required for the activity of a key diatom iron transport system is noteworthy and suggests that ocean acidification might inhibit iron uptake. The occurrence of transferrin in unicellular marine diatoms raises significant evolutionary and functional questions concerning its origin and role within marine phytoplankton. In the current work, we used phytotransferrin (pTF) as a bait to implement APEX2-enabled proximity proteomics in a model marine diatom *P. tricornutum*, and the resulting data provides several key advances regarding the function of pTF and its putative downstream interaction partners.

McQuaid et al., 2018, observed pTF in intracellular puncta, showed that clathrin-mediated endocytosis inhibitor reduces iron (Fe) uptake rates, visualized vesicles after adding iron to an iron-limited WT *P. tricornutum* culture stained with the membrane dye FM1-43, but our MDY-64 labeling data provide the first direct evidence that diatom pTF is indeed associated with membranous compartments. Further studies are needed to elucidate pTF dynamics upon internalization—especially after iron addition to iron-deplete cells—and its trafficking in the endomembrane system (**Naslavsky and Kaplan, 2018**). This may include a combination of time-lapse microscopy using pTF-YFP expressing *P. tricornutum* cells stained with the FM1-43 dye, pulse-chase experiments, biophysical techniques, and mathematical modeling (**Mayle et al., 2012**).

Enzymatically active APEX2 in *P. tricornutum* was largely expected as heme-containing horseradish peroxidase (HRP) had been successfully expressed in diatoms (***Sheppard et al., 2012***). However, it was less clear what the contribution of endogenous peroxidases would be to the background signal in various APEX2 assays. In our Amplex UltraRed assay experiments, an order of magnitude higher signal-to-noise ratio was detected when live cells were reacted on ice as opposed to on room temperature, suggesting endogenous *P. tricornutum* peroxidases, but not exogenous APEX2, are largely inhibited on ice.

Identifying Amplex UltraRed assay conditions which minimized WT background was promising, but not a guarantee that some endogenous enzymes would not be able to process other APEX2 substrates, such as those crucial for electron microscopy (diaminobenzidine) and proximity proteomics (biotin-phenol). Indeed, mitochondrial background signal was detected in our electron microscopy experiments both in WT and pTF-APEX2 expressing cells. Nevertheless, C-terminal tagging of pTF with APEX2 resulted in the expected cell surface and vesicular fusion protein localization away from mitochondria indicating no spatial overlap between endogenous APEX2-like enzymes and pTF-APEX2. Some off-target biotinylation in our proximity labeling experiments was likely due to background mitochondrial peroxidase activity as indicated by the presence of mitochondrial proteins in our MS data. Their APEX2/WT ratios were close to 1 which provides us with more confidence that those with elevated ratios above 1.5 were enriched due to APEX2 activity. Many of these have biologically relevant annotations for proteins we might expect to play a role in endosomal Fe assimilation, which further supports this notion. Additional efforts to express pTF tagged with N-terminal APEX2, which should lead to intravesicular proximity labeling according to our pTF orientation model, as well as optimizing conditions for APEX2 reactivation in iron-deplete conditions with hemin chloride might lead to identification of additional proteins. Proteomic maps with improved spatial and temporal resolution may emerge from experiments conducted with synchronized cell lines sampled at different timepoints throughout diel cycle, as iron metabolism genes, including *pTF*, are highly expressed at night (***Huysman et al., 2014***; ***Lundberg and Borner, 2019***; ***Smith et al., 2016***). We are particularly excited by the

recent demonstration of using TurboID—an engineered biotin ligase—for proximity proteomics in plants: *Arabidopsis thaliana* (**Kim et al., 2019**; **Mair et al., 2019**) and *Nicotiana benthamiana* (**Zhang et al., 2019**). Due to its independency of Fe and considering the presence of putative APEX2-like peroxidases in *P. tricornutum* causing unwanted mitochondrial background, TurboID (**Branon et al., 2018**) could represent a valuable proximity proteomics tool in this model diatom. These proposed next steps, improvements, and alternative experimental approaches should cross-validate the proteins identified in our experiments as well as lead to new ones.

The largely uncharacterized gene cluster on chromosome 20 we focused on is transcriptionally upregulated in iron-deplete (**Allen et al., 2008**; **Smith et al., 2016**) and silicic acid-replete (**Sapriel et al., 2009**) conditions corroborating a known, albeit understudied, link between iron and silicon metabolism in diatoms (**Brzezinski et al., 2015**; **Durkin et al., 2012**; **Hutchins et al., 1998**; **Leynaert et al., 2004**). Operon-like gene clusters encoding secondary metabolite pathway components are common in plants (**Boycheva et al., 2014**; **Nützmann et al., 2017**) and a tightly localized ~7 kbp gene cluster encoding enzymes for domoic acid production has recently been characterized in a cosmopolitan diatom *Pseudo-nitzschia multiseries* (**Brunson et al., 2018**). We propose a term “nutritional gene cluster” to describe loci that are sensitive to fluctuations in environmental nutrient status such as the studied ~5.6 kbp one in *P. tricornutum*. Our phylogenetic analyses suggest all three proteins encoded by this locus are broadly present in diatoms and other (marine) microeukaryotes while pTF.CREGr is evolutionarily conserved across the tree of life. Together with the rest of our proteomic data they provide an emerging view of possible molecular parallels and differences between metazoan transferrin and marine microbial phytotransferrin pathways.

In transferrin-mediated Fe uptake in human cells, endosome acidification induces structural rearrangements and causes protonation of coordinated carbonate, facilitating iron release (**Cheng et al., 2004**). This event is concurrent with or followed by iron reduction by endosomal ferric reductase STEAP3 and export via an Fe(II)-specific divalent metal transporter DMT1 (**Oghami et al., 2005**). Two putative proton pumps are

present among our MS hits: Phatr3_J23497—predicted as a P-type ATPase—and Phatr3_J27923—a vacuolar-type H^+-ATPase subunit A. Multisubunit V-ATPases regulate pH homeostasis in virtually all eukaryotes and are known to function as endosome acidifying protein complexes (***Finbow and Harrison, 1997***; ***Maxson and Grinstein, 2014***; ***Merkulova et al., 2015***). This makes Phatr3_J27923 a candidate protein involved in acidification of phytotransferrin-rich endosomes in *P. tricornutum*. Notably, V-ATPase transcripts have been overrepresented in diatoms following iron enrichment (***Marchetti et al., 2012***), possibly to account for an increase in the number of endosomes associated with pTF-mediated Fe acquisition. V-ATPase has also been predicted to be involved in silica deposition vesicle (SDV) acidification in diatoms (***Vartanian et al., 2009***; ***Hildebrand et al., 2018***). Such endosome acidification, perhaps via V-ATPase, may thus facilitate phytotransferrin-bound Fe(III) release, possibly for processing by an iron reducing enzyme that performs a function similar to STEAP3.

Human cellular repressor of E1A-stimulated genes 1 (CREG1) is a homodimer with a loop occluding the putative flavin mononucleotide (FMN)-binding pocket at the dimer interface, typical of evolutionarily related FMN-binding split barrel fold oxidoreductases such as those found in bacteria and yeast (***Sacher et al., 2005***). In humans, CREG1 is involved in embryonic development, growth, differentiation, and senescence as part of endosomal-lysosomal system, though its exact function and role are debated (***Ghobrial et al., 2018***). CREG1-like protein pTF.CREGr (Phatr3_J51183) was proposed to be central for sustained growth of iron-limited *P. tricornutum* cells (***Allen et al., 2008***) or to serve as a cell surface iron receptor (***Lommer et al., 2012***). CREG transcripts were more commonly found in diatom transcriptomes from the Southern Ocean as opposed to non-Southern Ocean regions suggesting their importance in coping with iron limitation in this major HNLC zone (***Moreno et al., 2018***). This may also explain why many more diatom homologs are not present in our phylogenetic tree as marine microbial eukaryote transcriptomes from MMETSP (***Caron et al., 2016***; ***Keeling et al., 2014***) were not obtained from iron-limited cultures. Evolutionary link of CREG proteins to oxidoreductases and phylogenetic classification of pTF.CREGr away from

multicellular organisms prompted us to investigate whether this protein could be a ferric reductase. Importantly, none of the four canonical predicted ferric reductases (FRE1–4, Phatr3_J IDs: 54486, 46928, 54940, 54409) that may have a role in cell surface iron reduction processes (***Allen et al., 2008***) were present in our MS data. In contrast, our data demonstrating punctate pTF.CREGr colocalization with pTF in the vicinity of chloroplast periphery suggest that this protein has no cell surface-associated role, but rather acts intracellularly, perhaps in an endosomal-lysosomal system. This localization is consistent with the observation that proteins with low confidence ASAFind chloroplast prediction (such as pTF.CREGr) can be associated with “blob-like structures” (BLBs) adjacent to, but not inside, the chloroplast (***Gruber et al., 2015***; ***Kilian and Kroth, 2005***). Biochemical characterization of pTF.CREGr revealed that the protein is able to reduce free flavins in an NADPH-dependent manner. In bacteria, flavin reductases can also function physiologically as ferric reductases, promoting the transfer of electrons to ferric Fe(III) iron using reduced flavins as an intermediate (***Fontecave et al., 1994***; **Schröder *et al., 2003***). While the true physiological role of pTF.CREGr is unclear at present, it is possible that the flavin reductase activity of the enzyme drives ferric iron reduction as a part of the phytotransferrin (pTF)-dependent iron uptake system in diatoms. We do, however, express concern if placing pTF.CREGr inside a vesicle is justifiable considering likely, but not certain, APEX2 orientation away from the vesicle interior in our proximity labeling experiment and the known inability for phenoxyl radicals created by APEX2 to penetrate endomembranes (***Rhee et al., 2013***). Considering pTF orientation has not been experimentally validated, pTF.CREGr may have been labeled while being recruited to the vesicle, and diatom membranes may be permeable to phenoxyl radical, pTF.CREGr localization to the vesicle interior remains plausible.

In summary, pTF.CREGr may be a ferric reductase critical for endosome-associated iron reduction step facilitating further intracellular iron allocation and assimilation. Intracellular Fe(III) reduction with pTF.CREGr would also be consistent with the nonreductive cell surface Fe(III) uptake step proposed for pTF (***McQuaid et al., 2018***; ***Morrissey et al., 2015***). It is unclear at this point how would ferrous, Fe(II), iron be exported from the endosome for cellular use. One possible candidate, although not

present in our MS data, is ferroportin (Fpn), the only known ferrous iron exporter in mammalian cells (**Ward and Kaplan, 2012**). Ferroportin homologs in *Arabidopsis thaliana* have been shown to be involved in iron and cobalt homeostasis both intracellularly (vacuole) and on the cell surface (**Morrissey et al., 2009**). The *P. tricornutum* genome contains a putative ferroportin gene, *Phatr3_J54495*. It is plausible that such "phytoferroportins" in single-celled marine microeukaryotes such as diatoms also function intracellularly, perhaps in Fe(II)-rich endosomes.

No functional annotations could be predicted for pTF.CatCh1 in Pfam (**El-Gebali al., 2019**). This protein contains a cysteine-rich domain and a disordered C-terminal domain split in two by a predicted transmembrane region. The transmembrane (TM) region is preceded by a flexible polyserine stretch and is followed by a short arginine-rich motif. Such positively charged amino acid tracts in the vicinity of TM regions have been shown to be orientation determinants in outer chloroplast membrane proteins (**May and Soll, 1998**). Disordered domains are known mediators of protein-protein interactions and play an important role in the formation of protein complexes (**Uversky, 2015**; **Uversky, 2016**). It is therefore plausible pTF.CatCh1 interacts with pTF with its C-terminal disordered region. The cysteine-rich domain contains two highly conserved motifs—CXC and CXXC—which have been observed in a variety of metalloregulatory and metal sensing proteins (**O'Halloran, 1993**) as well as metallothioneins (**Romero-Isart, 2002**) and metallochaperones, particularly those involved in copper binding (**Blaby-Hass et al., 2014**; **O'Halloran and Culotta, 2002**; **Robinson and Winge, 2012**). Two additional CXC motifs are conserved in its paralog ISIP2b. Metallochaperones are known to engage in protein-protein interactions (**Rosenzweig, 2002**) and are important in preventing metal-induced toxicity caused by Fenton reactions, especially common with iron (**Valko et al., 2005**). Taken together, these observations indicate that pTF.CatCh1 might be a metallochaperone localized to the chloroplast margin as suggested by our fluorescence microscopy data, possibly binding iron Fe(II) downstream of the reduction step and subsequently directing Fe traffic to the chloroplast interior. The observed colocalization of pTF.CatCh1 with pTF at the chloroplast periphery may alternatively indicate that Fe(III), as opposed to Fe(II), is offloaded directly from pTF to pTF.CatCh1

which would imply the existence of an additional intracellular Fe(III) reduction pathway. It is unclear which proteins could mediate the subsequent ferrous, Fe(II), iron import into the chloroplast. We note that transcripts for two putative organellar cation diffusion facilitators (CDFs), divalent metal ion transporters, are enriched under iron limitation in *P. tricornutum* (**Allen et al., 2008**), so they may fulfill this role.

Chloroplasts are the most iron-rich system in plant cells (**López-Millán et al., 2016**). Iron—in particular iron-sulfur clusters (**Balk and Schaedler, 2014**)—serves as an essential cofactor for the photosynthetic electron transfer chain, catalytic processes such as chlorophyll biosynthesis, and chloroplast protein import (**Marchand et al., 2018**; **Soll and Schleiff, 2004**). A complex vesicle-based system possibly in perpetual dynamic exchange with cytosolic vesicles is present in chloroplasts (**Lindquist and Aronsson, 2018**), an endoplasmic reticulum (ER) to Golgi to chloroplast protein trafficking pathway exists (**Villarejo et al., 2005**; **Radhamony and Theg, 2006**), and further hypotheses for such specialized vesicle-mediated plastid targeting of proteins have been established (**Baslam et al., 2016**). Thus, while very little is known about intracellular iron trafficking in diatoms or other algae with complex plastids, a direct link between endocytic internalization of phytotransferrin-bound iron and its offloading in or adjacent to chloroplasts as suggested by our results seems plausible. Analysis of our mass spectrometry hits that were not co-expressed with pTF reveals four additional proteins—ISIP1 (Phatr3_J55031), FBP (Phatr3_ J46929), 14-3-3 protein (Phatr3_J5651), and FBAC5 (Phatr3_J41423)—that support the idea of a direct cell surface-to-chloroplast iron trafficking axis in *P. tricornutum*.

ISIP1 (iron starvation-induced protein 1; Phatr3_J55031) is induced by iron limitation in *P. tricornutum* (**Allen et al., 2008**) and other marine microalgae (**Marchetti et al., 2012**) and was the second most enriched protein in our MS dataset. It is an environmentally ubiquitous protein that has been shown to have a role in iron uptake (**Kazamia et al., 2018**) and more recently to colocalize with pTF adjacent to the chloroplast in what has been referred to as an “iron processing” compartment (**Kazamia et al., 2019**). Metal homeostasis by such discrete membranous compartments is widespread in nature

671 (**Blaby-Haas and Merchant, 2014**). Copper- and zinc-storing compartments,
672 cuprosomes and zincosomes, respectively, are present in the model chlorophyte
673 microalga *Chlamydomonas reinhardtii* (**Aron et al., 2015**; **Hong-Hermesdorf et al.,**
674 **2014**; **Merchant, 2019**), and lipid-bound iron-accumulating ferrosome compartments
675 exist in diverse bacteria (**Komeili, 2018**). Phatr3_J46929—a siderophore ferrichrome-
676 binding protein (FBP) (**Coale et al., 2019**)—was another highly enriched protein in our
677 dataset indicating the internalization pathways for inorganic (i.e. phytotransferrin-bound)
678 and organic (i.e. siderophore-bound) iron in *Phaeodactylum tricornutum* may be
679 physically coupled intracellularly and similarly directed to iron-accumulating vesicles
680 close to the chloroplast.
681
682 Preproteins in *Arabodopsis thaliana* are targeted to the chloroplast—specifically to Toc
683 GTPases—by the cytosolic guidance complex consisting of a protein dimer comprised
684 of HSP70 (heat shock protein-70) and 14-3-3 (**Bölter, 2018**; **Soll and Schleiff, 2004**).
685 14-3-3 proteins are widespread regulatory proteins usually acting through protein-
686 protein interactions (**Soll and Schleiff, 2004**). They recognize phosphorylated serines
687 and threonines in preprotein transit peptides and serve as cytosolic chaperones
688 preventing premature chloroplast-specific activity of bound preproteins (**May and Soll,**
689 **2000**). One of our MS hits—Phatr3_J5651—is a 14-3-3 protein and it is plausible
690 pTF.CatCh1 localization to the chloroplast margin depends on this chloroplast targeting
691 pathway. Two small heat shock proteins are also among our highly enriched proteins—
692 Phatr3_J54656 and Phatr3_J46468—however they are of the HSP20 variety and it is
693 more likely they are involved in iron stress response more broadly (**Carra et al., 2019**)
694 as opposed to interacting with Phatr3_J5651 specifically.
695
696 Finally, the most enriched protein in our MS dataset was Phatr3_J41423, iron-sensitive
697 and iron-independent class I fructose-biophosphate aldolase 5 (FBAC5), which was
698 shown to localize to the pyrenoid in the interior of the *P. tricornutum* chloroplast and was
699 proposed to link Calvin-Benson cycle activity with the CO_2 concentrating mechanism
700 (**Allen et al., 2012**). FBAC5 is one of the most reliable markers for diatom iron stress as
701 demonstrated by laboratory culture (**Cohen et al., 2018**; **Lommer et al., 2012**),

microcosm incubation (**Cohen et al., 2017**; **Cohen et al., 2017**), and field studies (**Bertrand et al., 2015**), however its exact role, despite some existing hypotheses, remains elusive. FBAs are known moonlighting proteins (**Jeffrey, 2018**) and have been shown to have a role in endosome acidification via direct interaction with V-ATPase (**Merkulova et al., 2011**) which offers a tantalizing possibility for ancillary FBAC5 function in *P. tricornutum*.

Additional notable proteins with APEX2/WT ratio of at least 1.5 include a Rab family small GTPase Sec4 (Phatr3_J30139) which has been observed to localize between chloroplasts and nuclei during cell division in *P. tricornutum* (**Tanaka et al., 2015**), Phatr3_J41172, annotated as Ca^{2+}-binding protein calreticulin, and an α-carbonic anhydrase Phatr3_J42574 (predicted to be chloroplastic with low confidence).

Interestingly, rat calreticulin is homologous to an iron-binding protein mobilferrin (**Conrad et al., 1993**). This protein has been shown to associate with transferrin-containing vesicles and to act as an intermediate between the transferrin-bound iron and the incorporation of iron into hemoglobin in erythroleukemia cells (**Conrad et al., 1996**). It has also been identified as a constituent of a paraferritin protein complex with ferrireductase activity in rats (**Umbreit et al., 2002**; **Umbreit et al., 1996**). This may suggest that Phatr3_J41172, despite calreticulin annotation, has a role in intracellular iron transport, perhaps to direct iron to mitochondria for incorporation into heme.

Considering carbonic anhydrases catalyze rapid conversion of HCO_3^- to CO_2 (**Matsuda et al., 2017**), the presence of an α-carbonic anhydrase Phatr3_J42574 in our data raises an intriguing possibility: could it be that pTF-bound carbonate not only coordinates iron, but is itself a cargo? It would be conceivable that a protonated carbonate released from pTF gets metabolized by a carbonic anhydrase with the resulting CO_2 permeating through the chloroplast and increasing CO_2 levels around RuBisCO for efficient CO_2 fixation. Such direct molecular bridge between iron acquisition and CO_2 concentrating mechanism (CCM) would further explain commonly

732 observed prevalence of diatoms in phytoplankton blooms in times of oceanic iron
733 repletion.

734

735 ## Conclusion and Outlook

736

737 We here implemented the bifunctional genetically encoded probe APEX2 for use in
738 diatoms which will complement existing efforts to comprehensively map diatom
739 organellar proteomes (***Schober et al., 2019***). Direct imaging of APEX2-tagged
740 phytotransferrin in *Phaeodactylum tricornutum* revealed its tight cell membrane and
741 vesicular localization consistent with fluorescent microscopy results (***McQuaid et al.,***
742 ***2018***; ***Morrissey et al., 2015***; ***this study***). APEX2-based imaging is thus
743 complementary to super-resolution microscopy in diatoms (***Gröger et al., 2016***) and
744 represents a basis for electron tomography applications (***Sengupta et al., 2019***). In the
745 proximity proteomics experiment, several proteins with a predicted role in iron
746 metabolism and protein transport were identified, and we showed that two of them—
747 pTF.CREGr and pTF.CatCh1—colocalize with pTF adjacent to the chloroplast
748 suggesting a direct cell surface-to-chloroplast iron trafficking axis. Initial biochemical
749 data for pTF.CREGr support its hypothesized role as a ferric reductase, similar to
750 bacterial flavin reductases. *In vivo* experiments involving either split fluorescent proteins
751 (***Kudla and Bock, 2016***) or fluorescence resonance energy transfer (FRET) (***Marshall***
752 ***et al., 2012***) will determine if they directly interact with or are only vicinal to pTF.
753 Differential fluorophore tagging of these two and other identified proteins coupled with
754 time-lapse microscopy should shed light on their localization dynamics relative to pTF.
755 Finally, genome editing using either TALENs or CRISPR/Cas9 (***Kroth et al., 2018***)
756 coupled with physiological measurements of the resulting knockout cell lines under
757 various environmental conditions will also be crucial to advance the work herein.

758

759 Taken altogether, our data enable proposition of a model connecting cell surface ferric
760 iron binding, internalization and intracellular trafficking in *Phaeodactylum tricornutum*
761 (**Figure 9**) with many outstanding questions ripe for additional study.

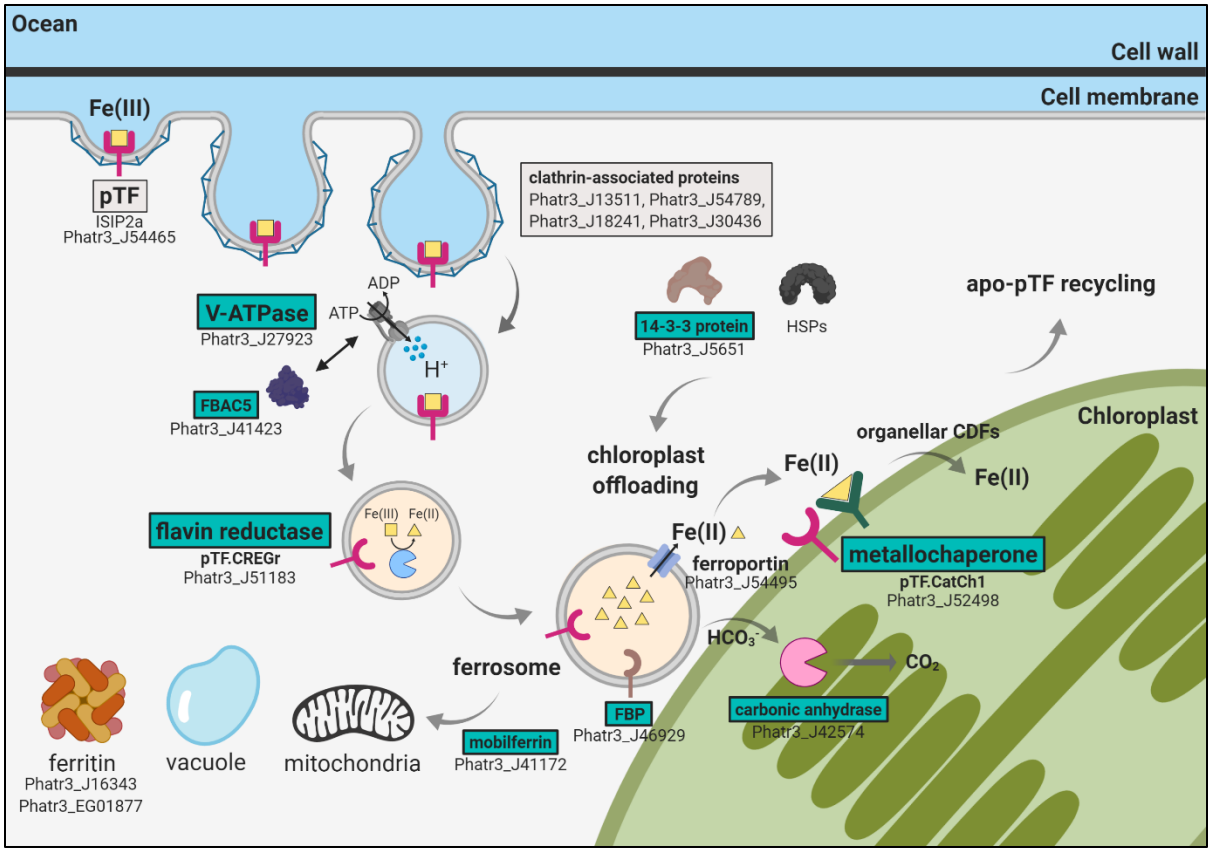


Figure 9. Emerging view of inorganic iron allocation pathways in *Phaeodactylum tricornutum.* We propose a hypothetical model for clathrin-mediated internalization of phytotransferrin (pTF)-bound ferric iron, its trafficking and offloading to the chloroplast. Data from this proteomics study (teal panels) and a previous transcriptomics study (grey panels)—which identified additional endocytosis- and trafficking-related iron-sensitive transcripts—are synthesized. One phytotransferrin-Fe(III) complex is shown for clarity. Detailed chloroplast membrane layers are omitted (diatom chloroplasts contain four membranes). In this model, pTF undergoes a structural change upon endosome acidification mediated by a V-ATPase releasing Fe(III) which gets reduced by a ferric reductase, a role which may be served by the flavin reductase pTF.CREGr. Reduced iron may then be exported through an unknown endosomal ferrous transporter (perhaps ferroportin) and offloaded to a chloroplastic metallochaperone (perhaps pTF.CatCh1), possibly located in the outer chloroplast membrane, and from there shuttled into the chloroplast interior (perhaps by cation diffusion facilitators). Our colocalization data alternatively suggest pTF may be recruited to the chloroplast margin together with pTF.CatCh1 which would imply an additional Fe(III) reduction pathway. 14-3-3 proteins such as the identified Phatr3_J5651 could be involved in the offloading process. The whole pathway might intersect with that for internalizing siderophore-bound iron as indicated by the presence of a ferrichrome-binding protein (FBP) in our data. Finally, pTF-coordinated carbonate may itself be a cargo. Its endosomal protonation would provide bicarbonate acting as a substrate for the uncharacterized carbonic anhydrase Phatr3_J42574. Created with BioRender.com.

As our understanding of intracellular iron trafficking is relatively incomplete even in more established systems (***Philpott and Jadhav, 2019***) and our knowledge of iron transport into the chloroplast rudimentary (***Blaby-Haas and Merchant, 2012***), this work represents an important advance beyond adding a new tool to the molecular toolbox of this slowly maturing model system.

Diatoms more generally are suitable to address a range of questions related to evolutionary cell biology, biotechnology and nanotechnology (***Figure 9—figure supplement 1***). Elaborately patterned silica cell walls in diatoms have intrigued scientist for hundreds of years (***Anonymous, 1703***). Cell wall synthesis happens in a silica deposition vesicle and numerous proteins involved in this process have been identified (***Hildebrand et al., 2018***) providing a rich pool of baits for APEX2 tagging. Elucidating new proteins and protein-protein interactions in this pathway has the potential to inspire novel nanotechnologies based on biologically directed mineral growth (***Kröger and Brunner, 2014***). Similarly elusive is the composition of diatom pyrenoids—chloroplastic RuBisCO-containing suborganelles at the core of CO_2 concentrating mechanism (***Tsuyi et al., 2017***). Understanding their composition and biogenesis will inform existing efforts to increase photosynthetic efficiency of important food crops (***Orr et al., 2017***).

In conclusion, APEX2 and related molecular approaches such as TurboID together with emerging computational protein-protein interaction prediction methods (***Cong et al., 2019***) will be instrumental in advancing diatom cell biology and hold potential for establishment of cell-wide subcellular proteomic and interactomic maps, a vision that is already being realized for a human cell (***Go et al., 2019***). Considering enormous marine microbiome diversity (***de Vargas et al., 2015***) and the advent of new model marine microeukaryotes (***Faktorová et al., 2019***; ***Waller et al., 2018***) we also believe these tools are well-positioned to impact evolutionary cell biology more broadly. This will have important implications for understanding life on length scales from atomic to planetary as well as on temporal scales ranging from emergence of life billions of years ago to understanding the present impact of human activity on climate change.

Materials and Methods

Vector construction

pJT_NR_pTF-AP2: Gibson Assembly (***Gibson et al., 2009***) was performed with three DNA fragments: (1) linearized pPt-BR322-1 episome backbone (Addgene, Cambridge, MA, plasmid "pPtPBR1", Catalog #80388) opened ~280 bp downstream of the tetracycline resistance gene, (2) amplicon with nitrate reductase gene (*NR*) promoter, 5' *NR* UTR and *pTF*, and (3) gBlocks Gene Fragment (Integrated DNA Technologies (IDT), Coralville, IA) with linker sequence, *APEX2* codon optimized using IDT Codon Optimization Tool (selecting "*Thalassiosira pseudonana*" from the "Organism" drop-down menu), 3' *NR* UTR and *NR* terminator.

pJT_native_pTF-mCherry: Gibson Assembly was performed with three DNA amplicons: (1) pPt-BR322-1 episome backbone split into two fragments and (2) expression cassette including *pTF*, *mCherry* and *pTF* promoter and terminator.

pJT_pTF-mCherry_MS hit-EYFP: pJT_native_pTF-mCherry was split into two fragments using PCR stitching (keeping ampicillin resistance gene split). MS hit genes were amplified using cDNA from iron-starved WT *P. tricornutum* cells and assembled with *EYFP* and flavodoxin (*Phatr_J23658*) promoter and terminator through two PCR stitching rounds into a single fragment. Gibson Assembly was then performed to combine all three final amplicons. Detailed assembly scheme is presented in ***Figure 5—figure supplement 2A***.

pJT_Δ31_pTF.CREGr-His6: Cloning of gene fragments into the *E. coli* protein expression vector PtpBAD-CTHF was performed as described previously (***Brunson et al., 2018***). Briefly, PtpBAD-CTHF was linearized by digestion with XhoI (New England Biolabs (NEB), Ipswich, MA, Catalog #R0146S) and the resulting DNA was column purified. Truncated Δ18_pTF.CREGr gene was obtained by PCR from *P. tricornutum* gDNA with PrimeSTAR polymerase (Takara Bio, Kusatsu, Japan) and primer set

JT31/JT32 to incorporate the appropriate Gibson assembly overhangs and remove the putative N-terminal signal peptide (amino acids 1–18). Insertion of truncated pTF.CREGr into linearized PtpBAD-CTHF was performed using Gibson Assembly Master Mix (NEB, Catalog #E2611S), 1 uL of the reaction mixture was transformed via heat shock into chemically competent NEB 5-alpha cells (NEB, Catalog #C2988J), and cells incubated on lysogeny broth (LB)-Tet10 plates overnight at 37°C. Transformants were screened by colony PCR using the primer set JT37/JT38 and Sapphire polymerase (Takara Bio), and positive clones were selected for outgrowth. Isolated plasmids were sequence-validated by Sanger sequencing (Eurofins, Luxembourg, Luxembourg). A sequence-validated clone was designated as PtpBAD-Δ18_pTF.CREGr-CTHF and transformed into chemically competent BL21 *E. coli* cells (NEB, Catalog #C2530H) which were plated on LB agar with tetracycline (10 µg/mL). The resulting transformants were used for subsequent Δ18_pTF.CREGr expression experiments. Following unsuccessful expression testing of the Δ18 N-terminal truncation construct, an additional set of N-terminal truncations (Δ26, Δ31, Δ39, Δ43) was generated by PCR using PrimeStar polymerase, primer sets JT33/JT32, JT34/JT32, JT35/JT32, JT36/JT32, and PtpBAD-Δ18_pTF.CREGr-CTHF as a template. Assembly, colony PCR screening, plasmid isolation, sequencing, and transformation into chemically competent BL21 *E. coli* cells of all additional expression vectors, including PtpBAD-Δ31_pTF.CREGr-CTHF (=pJT_Δ31_pTF.CREGr-His6), was performed as above. An additional N-terminal His6 Δ39 construct was built (PtpBAD-NTH-Δ39_pTF.CREGr) using the vector NTH-PtpBAD (constructed similarly to PtpBAD-CTHF, see ***Brunson et al., 2018*** and ***Savitsky et al., 2010***), but was not pursued beyond initial expression testing.

pTF (*Phatr3_J54465*) in all episomes was in its native form (3 exons, 2 introns). Molecular cloning primers are listed in ***Supplementary file 1—Table S3.*** Vector details, further amplicon information, and fusion protein sequences are catalogued in ***Supplementary file 1—Table S4***. PrimeSTAR GXL DNA Polymerase (Takara Bio) is recommended for amplifying (parts of) diatom episomes.

Related resource: Turnšek J. and Gholami P. 2017. Guidelines for highly efficient construction of diatom episomes using Gibson Assembly. *protocols.io*. DOI: dx.doi.org/10.17504/protocols.io.jy7cpzn

Diatom culturing, conjugation and genotyping

Culturing: Sequenced *Phaeodactylum tricornutum* strain CCMP632 (synonymous to CCMP2561 and CCAP 1055/1; NCMA, East Boothbay, ME) was used throughout the study and grown at 18°C and 300 µmol quanta m^{-2} s^{-1} in a 10 h:14 h dark:light cycle in biotin-free L1 medium prepared by mixing 1 L Aquil salts, 2 mL nitrate and phosphate (NP) nutrient stock, 1 mL trace metal stock, and 1 mL thiamine hydrochloride and cyanocobalamin (TC) stock unless otherwise noted. Aquil salts: 0.5 L anhydrous salts (0.5 L Milli-Q, 24.5 g NaCl, 4.09 g Na_2SO_4, 0.7 g KCl, 0.2 g $NaHCO_3$, 0.1 g KBr, 900 µL 33.3 mg/mL H_3BO_3 stock, 300 µL 10 mg/mL NaF stock) and 0.5 L hydrous salts (0.5 L Milli-Q, 11.1 g $MgCl_2$ x $6H_2O$, 1.54 g $CaCl_2$ x $2H_2O$, 100 µL 170 mg/mL $SrCl_2$ x $6H_2O$ stock) were prepared separately, combined and filter sterilized (0.2 µm). NP nutrient stock: 37.5 g $NaNO_3$ and 2.5 g NaH_2PO_4 were dissolved in 100 mL Milli-Q, filter sterilized (0.2 µm) and stored at 4°C. Trace metal stock (for 1 L 1000x stock): 3.15 g $FeCl_3$ x $6H_2O$, 4.36 g Na_2EDTA x $2H_2O$, 0.25 mL 9.8 g/L $CuSO_4$ x $5H_2O$, 3.0 mL 6.3 g/L Na_2MoO_4 x $2H_2O$, 1.0 mL 22 g/L $ZnSO_4$ x $7H_2O$, 1.0 mL 10 g/L $CoCl_2$ x $6H_2O$, 1.0 mL 180 g/L $MnCl_2$ x $4H_2O$, 1.0 mL 1.3 g/L H_2SeO_3, 1.0 mL 2.7 g/L $NiSO_4$ x $6H_2O$, 1.0 mL 1.84 g/L Na_3VO_4, 1.0 mL 1.94 g/L K_2CrO_4, and Milli-Q up to 1 L were combined, filter sterilized (0.2 µm), and kept at 4°C. TC stock: 20 mg thiamine hydrochloride and 0.1 mL 1 g/L cyanocobalamin were mixed in 100 mL Milli-Q. The solution was stored at 4°C. *ΔpTF P. tricornutum* strain (maintained in the Allen lab; ***McQuaid et al., 2018***) was additionally supplemented with 200 µg/mL nourseothricin (GoldBio, Saint Louis, MO, Catalog #N-500-1). All transconjugant *P. tricornutum* strains were supplemented with 50 or 100 µg/mL phleomycin (InvivoGen, San Diego, CA, Catalog #ant-ph-10p).

Conjugation: (1) Bacterial donor preparation: Chemically competent pTA-Mob-containing TransforMax™ EPI300™ *E. coli* cells (Lucigen, Middleton, WI, Catalog

#EC300110) were transformed with sequence-verified pPt-BR322-1 episomes. Transformants were selected on gentamycin-, carbenicillin- and tetracycline-containing LB agar plates. 3 mL overnight LB cultures supplemented with antibiotics were inoculated from glycerol stocks. (2) *P. tricornutum* preparation: ~2 x 10^8 *P. tricornutum* cells (in 200 µL) from mid- to late-exponential phase were spread on pre-dried (=plates with lid half open and kept in the laminar flow hood for at least 90 min) ½ L1 1% agar plates with 5% LB and left growing for 1 or 2 days. Plates were additionally supplemented with 200 µg/mL nourseothricin for conjugation of *ΔpTF P. tricornutum* cells. (3) Conjugation: Overnight bacterial cultures were diluted 1:50 in 25 mL LB supplemented with antibiotics, grown at 37°C until OD_{600} 0.8–1, spun down, resuspended in 150 uL Super Optimal broth with Catabolite repression (SOC) medium and spread as evenly as possible on top of a *P. tricornutum* lawn. Plates with donor-*P. tricornutum* co-culture were left for 90 min in dark at 30°C, then for 1 or 2 days at standard growth conditions. (4) Selection: Co-culture lawn was scrapped off with 1 mL fresh L1 medium, transferred to a microcentrifuge tube, and 200 µL spread on pre-dried (see above) ½ L1 1% agar plates with 50 or 100 µg/mL phleomycin. Plates were additionally supplemented with 200 µg/mL nourseothricin for conjugation of *ΔpTF P. tricornutum* cells. Porous adhesive tape was used to seal the plates and transconjugants emerged after ~10 days of incubation under standard growth conditions. Please see ***Karas et al., 2015*** and ***Diner et al., 2016*** for further description of diatom conjugation.

Genotyping: Candidate transconjugant colonies were inoculated in 300 µL L1 medium supplemented with 50 or 100 µg/mL phleomycin (and 200 µg/mL nourseothricin in case of *ΔpTF P. tricornutum* cells) and typically grown for ~1 week. 0.5 µL culture was then genotyped using either Phire Plant Direct PCR Master Mix (Thermo Fischer Scientific, Waltham, MA, Catalog #F160L) or Phire Plant Direct PCR Kit (Thermo Fischer Scientific, Catalog #F130WH). 200 µL of each genotype positive strain was passaged in 30 mL L1 medium supplemented with 50 or 100 µg/mL phleomycin (and 200 µg/mL nourseothricin).

Related resource: Turnšek J. 2017. Simple & rapid genotyping of marine microeukaryotes. *protocols.io*. DOI: dx.doi.org/10.17504/protocols.io.jcdcis6

RNA extraction and cDNA synthesis

Diatom culture: 100 mL *P. tricornutum* culture was grown in iron-limiting conditions for 2 weeks (L1 medium with 7.5 nM total Fe). Cells were then centrifuged, supernatants discarded, pellets flash frozen in liquid nitrogen and stored at -80°C.

RNA extraction: Pellets were resuspended in 800 µL Trizol, equal amount of 100% ethanol, and centrifuged in spin columns. Columns were then washed with 400 µL RNA Wash Buffer followed by on-column DNA digestion with 5 µL DNase I in 75 µL DNA Digestion Buffer for 15 min at RT, washing twice with 400 µL Direct-zol RNA PreWash and once with 700 µL RNA Wash Buffer. RNA was eluted with 50 µL DNase/RNase-Free Water, RNA integrity number (RIN) evaluated with 2200 TapeStation (Agilent Technologies, Santa Clara, CA; measured RIN was 7.0), concentration measured with Qubit 2.0 Fluorometer (Thermo Fischer Scientific; measured concentration was 27.6 ng/µL), and samples stored at -80°C. Direct-zol™ RNA Miniprep Plus RNA extraction kit was used (Zymo Research, Irvine, CA, Catalog #R2070).

cDNA synthesis: 1 µL total RNA was combined with 0.5 µL Oligo(dT)$_{20}$ Primer, 0.5 µL 10 mM dNTP Mix and 3 µL nuclease-free water followed by 5 min and 1 min incubation at 65°C and on ice, respectively. After 5 µL cDNA Synthesis Mix (1 µL 10x Reverse Transcription buffer, 2 µL 25 mM $MgCl_2$, 1 µL 0.1 M DTT, 0.5 µL RNase OUT, 0.5 µL SuperScript III Reverse Transcriptase) was added, samples were incubated 50 min and 5 min at 50°C and 85°C, respectively, then chilled on ice. Finally, 1 µL RNase H was added and sample kept for 20 min at 37°C. cDNA was stored at -20°C until use. cDNA synthesis kit was from Thermo Fischer Scientific (SuperScript™ III First-Strand Synthesis System, Catalog #18080-051).

MDY-64 labeling and imaging

10 mL of a *ΔpTF P. tricornutum* pTF-mCherry expressing strain grown in iron-limiting conditions (L1 medium with 7.5 nM total Fe) was spun down, supernatant discarded, and pellet resuspended in 50 µL phosphate-buffered saline (PBS) (pH 7.4). 0.5 µL 10 mM MDY-64 stock (in DMSO; Thermo Fischer Scientific, Catalog #Y7536) was added, cells incubated for 10 min at RT, pelleted, resuspended in 50 µL fresh PBS (pH 7.4) and imaged. 5 µL cell suspension was placed between a 1.5 mm microscope slide and a cover slip (this setup applies to all imaging experiments in the study). Imaging conditions: Leica TCS SP5 confocal microscope (Leica Microsystems, Wetzlar, Germany), argon laser strength set to 30%, 458 nm laser line at 50%, emission window set to 477–517 nm (MDY-64; MDY-64 excitation and emission maxima are 451 nm and 497 nm, respectively), 514 nm laser line at 50%, emission window set to 620–640 nm (mCherry; mCherry excitation and emission maxima are 587 nm and 610 nm, respectively).

Protein expression analyses

pTF-APEX2 detection: Cell pellets from 8 mL mid- to late-exponential WT or transconjugant cultures were resuspended in 150 µL cell lysis buffer (50 mM Tris-HCl, 200 mM NaCl, 1 mM DTT, 1 mM PMSF, pH 8.5) and sonicated for 5 min (30 sec on, 1 min off) with Bioruptor UCD-200TM. Lysates were centrifuged, total protein content in supernatants measured with the Bradford Assay Kit (Thermo Fischer Scientific, Catalog #23236), and insoluble fractions resuspended in 150 µL cell lysis buffer. 1 µg soluble protein and 0.5 µL resuspended insoluble protein fraction were resolved on NuPage 4-12% Bis-Tris 1.5 mm gels (Thermo Fischer Scientific, Catalog #NP0335BOX), wet transferred to polyvinylidene difluoride (PVDF) membranes (Thermo Fischer Scientific, Catalog #LC2005) and visualized with WesternBreeze™ Chemiluminescent Kit, anti-rabbit (Thermo Fischer Scientific, Catalog #WB7106).

pTF-mCherry detection: Cell pellets from 400 µL mid- to late-exponential WT or transconjugant cultures were resuspended in 50 µL cell lysis buffer (50 mM Tris-HCl, 200 mM NaCl, 1 mM DTT, 1 mM PMSF, pH 8.5) and sonicated 15 min (30 sec on, 1 min off) with Bioruptor UCD-200TM. 1 µL whole cell lysates were resolved on NuPage 4-12% Bis-Tris 1.5 mm gels, wet transferred to PVDF membranes and visualized with WesternBreeze™ Chemiluminescent Kit, anti-rabbit.

pTF antibody details: The immunogen were amino acid residues 32–223 (N-terminal pTF region just downstream of the signal peptide). The antibody was produced in a rabbit. 1:10,000 dilution of 1.14 mg/mL antibody stock was used for background-free results.

Protein ladder: MagicMark™ XP Western Protein Standard (Thermo Fischer Scientific, Catalog #LC5602).

Related resource: Turnšek J. 2017. HA tag enables highly efficient detection of heterologous proteins in *Phaeodactylum tricornutum* (*Pt*) exconjugants. *protocols.io*. DOI: dx.doi.org/10.17504/protocols.io.j7ncrme

Amplex UltraRed assay and resorufin imaging

Amplex UltraRed assay: 5 mL WT or pTF-APEX2 expressing cells in mid- to late-exponential phase were incubated on ice for 5 min, spun down, supernatant discarded, pellet resuspended in 500 µL ice-cold PBS (pH 7.4), and transferred to microcentrifuge tubes. Cells were spun down again, resuspended in 200 µL ice-cold reaction buffer (50 µM Amplex UltraRed (AUR; Thermo Fischer Scientific, Catalog #A36006), 2 mM H_2O_2, in PBS (pH 7.4)) and incubated on ice for 15 min unless otherwise noted. 50 µL supernatant was mixed with 50 µL PBS (pH 7.4) and resorufin fluorescence measured in a black microtiter plate with black bottom using Flexstation 3 microtiter plate reader (Molecular Devices, San Jose, CA; excitation: 544 nm, emission: 590 nm; resorufin excitation and emission maxima are 568 nm and 581 nm, respectively). Horseradish

peroxidase was always included as a positive assay control. Fluorescence was normalized to OD_{750} of experimental *P. tricornutum* cultures. Amplex UltraRed was prepared as 10 mM stock in DMSO and stored in 20 µL aliquots at -20°C. 3% (w/w) H_2O_2 stock (Sigma-Aldrich, Saint Louis, MO, Catalog #323381-25ML) was stored in 100 µL aliquots at -20°C.

Resorufin imaging: WT and pTF-APEX2 expressing cells after performing the assay were imaged with Leica TCS SP5 confocal microscope using the following parameters: argon laser strength at 30%, 514 nm laser line at 50%, resorufin emission window: 575–605 nm, autofluorescence emission window: 700–750 nm.

Transmission electron microscopy (TEM)

Part 1: Labeling

5 mL WT or pTF-APEX2 expressing cells from mid- to late-exponential phase were spun down (4000 rpm, 4°C, 10 min) and fixed in 5 mL ice-cold 2% (w/v) paraformaldehyde (PFA) and 2% (v/v) glutaraldehyde in 0.15 M sodium cacodylate buffer (pH 7.4) for 30 min on ice. Cells were rinsed in 5 mL 0.15 M sodium cacodylate buffer (pH 7.4) five times for 3 min on ice, then once again in 5 mL 0.15 M sodium cacodylate buffer (pH 7.4) with 10 mM glycine for 3 min on ice. Cells were then treated with 25 mM 3,3'-diaminobenzidine (DAB; Sigma-Aldrich, Catalog #D8001) as follows: 5.36 mg DAB was dissolved in 1 mL 0.1 N HCl and sonicated for 45 min. 5 mL of 0.3 M sodium cacodylate buffer (pH 7.4) was added to dissolved DAB, final volume adjusted to 10 ml with ddH_2O, solution filtered through a 0.22 µm syringe filter, and 3 µL 30% (w/w) H_2O_2 added for 3 mM final concentration. Cells were incubated in 5 mL of this reaction buffer for 15 min on ice, rinsed five times for 3 min in 5 mL 0.15 M sodium cacodylate buffer (pH 7.4), post-fixed in 2 mL 1% osmium tetroxide (OsO_4) (Electron Microscopy Sciences, Hatfield, PA, Catalog #19150) in 0.15 M sodium cacodylate buffer (pH 7.4) for 30 min on ice, and rinsed in 5 mL ice-cold ddH_2O five times for 3 min. Cells were then resuspended in 300 µL melted 3% agar, poured onto a glass slide sitting on ice, and cut into small ~3 x 3 x 1 mm pieces which were transferred into scintillation

vials with 10 mL ice-cold ddH_2O. All spin down steps were done at 4000 rpm and 4°C for 1.5 min. All buffers were used ice-cold.

Part 2: TEM prep

Agar blocks were fixed in 10 mL 2% glutaraldehyde in ddH_2O to crosslink agar, rinsed five times for 2 min in 5 mL ice-cold ddH_2O, and left incubating in 5 mL 2% uranyl acetate (Electron Microscopy Sciences, Catalog #22400) overnight at 4°C. Next morning, agar blocks were first dehydrated in the following ethanol series: 20%, 50%, 70%, 90%, 100% (on ice, 10 mL), 100%, 100% (at RT, 10 mL), then infiltrated in 10 mL 50% resin for ~1 h. To prepare 20 mL 100% resin, Durcupan ACM mixture (Electron Microscopy Sciences, Catalog #14040) components were combined as follows: 11.4 g A (epoxy resin), 10 g B (964 hardener), 0.3 g C (964 accelerator) and 0.1 g D (dibutyl phthalate) in this exact order (for 50% resin 1 part 100% resin was combined with 1 part 100% ethanol). Agar blocks were transferred into 10 mL 100% resin for ~4 h, then into fresh 10 mL 100% resin overnight, again into fresh 10 mL 100% resin for 4 h the following morning before being poured into aluminum boats and left to polymerize in 60°C oven for at least 48 h (usually over the weekend). Polymerized resins were detached from aluminum boats, agar blocks dense with cells cut out, and glued to "dummy" blocks by incubation in a 60°C oven for at least 15 min. ~500 nm thick sections were cut with an ultramicrotome (Leica Ultracut), stained with 1% toluidine blue and observed under light microscope to make sure embedded cells were exposed. ~100 nm thick sections were then cut, placed on TEM grids, labeled and saved until imaging. Imaging was performed with JEOL JEM-1200 (Japan Electron Optics Laboratory, Akishima, Tokyo, Japan) transmission electron microscope at 80 keV. Backscatter scanning electron microscopy was performed with Zeiss Merlin (Oberkochen, Germany) scanning electron microscope (SEM) at 2 keV by placing ~80 nm thick sections on a silicon wafer and imaged with inverted contrast which gives a TEM-like image.

Proximity labeling

Part 1: Labeling

25 mL WT and pTF-APEX2 expressing cells in mid- to late-exponential phase—cell density of all cultures at the time of harvest was ~2 x 10^7 cells/mL which corresponds to OD_{750} ~0.4—were cooled on ice for 10 min and pelleted (4000 rpm, 4°C, 10 min). Supernatants were discarded, pellets resuspended in 0.5 mL ice-cold PBS (pH 7.4), transferred to microcentrifuge tubes, and spun down (4000 rpm, 4°C, 10 min). Cells were then resuspended in 0.5 mL ice-cold 1.2 M D-sorbitol in PBS (pH 7.4), supplemented with 2.5 mM biotin-phenol (Berry & Associates, Dexter, MI, Catalog #BT 1015), incubated on a tube rotator at 4°C for 90 min, supplemented with 1 mM H_2O_2, and incubated on a tube rotator at 4°C for another 20 min. Labeling reaction was quenched by washing cells twice (4000 rpm, 4°C, 5 min) with 0.5 mL ice-cold quenching solution (10 mM sodium ascorbate (VWR International, Radnor, PA, Catalog #95035-692), 5 mM Trolox (Sigma-Aldrich, Catalog #238813-5G), 10 mM sodium azide (VWR International, Catalog #AA14314-22) in PBS (pH 7.4)). 50 µL quenched cells was saved for streptavidin blot. The remaining 450 µL was spun down (4000 rpm, 4°C, 10 min) and lysed in 250 µL cell lysis buffer (50 mM Tris-HCl, 200 mM NaCl, 1 mM DTT, 1 mM PMSF, pH 8.5) by sonication for 15 min (30 sec on, 1 min off). Cell lysates were spun down at 4000 rpm and 4°C for 45 min and protein concentration in supernatants measured using Bradford Assay Kit. For streptavidin blot, 25 µL saved quenched cells were first sonicated for 15 min (30 sec on, 1 min off). 2.5 µL whole cell lysate and 1 ng biotinylated BSA positive control (Thermo Fischer Scientific, Catalog #29130) were then resolved on NuPage 4-12% Bis-Tris 1.5 mm gel and wet transferred to a PVDF membrane. Membrane was washed twice with 15 mL PBST (PBS (pH 7.4) with 0.1% Tween-20) for 5 min, left blocking overnight at RT and gentle shaking in PBST supplemented with 5% BSA (Sigma-Aldrich, Catalog #A9647-100G), and washed once with 15 mL PBST for 5 min. It was then incubated for 1 h at RT and gentle shaking in 15 mL PBST supplemented with 5% BSA and 1:15,000 HRP-conjugated streptavidin (Thermo Fischer Scientific, Catalog #S911). Finally, membrane was washed three times with 15 mL PBST supplemented with 5% BSA for 5 min and once with 15 mL PBST for

10 min after which it was visualized with SuperSignal™ West Dura Extended Duration reagent (Thermo Fischer Scientific, Catalog #34075) using C-DiGit® Blot Scanner (Li-COR, Lincoln, NE).

Protein ladder: WesternSure® Pre-stained Chemiluminescent Protein Ladder (HRP-conjugated protein ladder) (Li-COR, Catalog #926-98000).

Part 2: Pull-down and mass spectrometry

Pull-down: 50 µL streptavidin beads (Thermo Fischer Scientific, Catalog #88816) were pelleted with a magnetic rack (~5 min to pellet fully). Supernatant was discarded and beads washed twice with 1 mL ice-cold cell lysis buffer (50 mM Tris-HCl, 200 mM NaCl, 1 mM DTT, 1 mM PMSF, pH 8.5). 360 µg protein in 500 µL total volume (x µL supernatant from cell lysis with 360 µg protein and 500-x µL cell lysis buffer) was incubated on a tube rotator overnight at 4°C. Streptavidin beads were washed to remove nonspecific binders the next morning as follows: 2x 1 mL cell lysis buffer, 1x 1 mL 1 M KCl, 1x 1 mL 0.1 M Na_2CO_3 (pH 11.5), 1x 1 mL 2 M urea (pH 8) in 10 mM Tris-HCl, 2x 1 mL cell lysis buffer, 2x 1 mL PBS (pH 7.4). PBS after the final washing step was removed before storing beads at -80°C. Notes: All the collection steps were 5 min. All the washing solutions were ice-cold. Microcentrifuge tubes were either very briefly vortexed (~2 sec) or tapped by hand between each wash step to promote bead resuspension.

On-bead digestion and TMT labeling: Samples were prepared as previously described (***Kalocsay, 2019***). Liquid reagents used were HPLC quality grade. Washed beads were resuspended in 50 µL of 200 mM EPPS (4-(2-Hydroxyethyl)-1-piperazinepropanesulfonic acid) buffer (pH 8.5) and 2% (v/v) acetonitrile with 1 µL of 2 mg/mL lysil endoproteinase Lys-C stock solution (FUJIFILM Wako Pure Chemical Corporation, Richmond, VA, Catalog #125-05061), vortexed briefly, and incubated at 37°C for 3 hours. 50 µL of trypsin stock (Promega, Madison, WI, Catalog #V5111) diluted 1:100 (v/v) in 200 mM EPPS (pH 8.5) was then added. After mixing, digests

were incubated at 37°C overnight and beads were magnetically removed. Peptides were then directly labeled as follows: acetonitrile was added to 30% (v/v) concentration and peptides were labeled with TMT 10-plex reagent (Thermo Fisher Scientific, Catalog #90406) for 1 hour. Labeling reactions were quenched with hydroxylamine at a final concentration of 0.3% (v/v) for 15 min and 1% of labeled peptides was analyzed for efficiency of label incorporation by mass spectrometry. After quenching, peptide solutions were first acidified with formic acid, trifluoroacetic acid (TFA) was then added to a concentration of 0.1%, and peptides were desalted by acidic C_{18} solid phase extraction (StageTip). Labeled peptides were finally resuspended in 1% (v/v) formic acid and 3% (v/v) acetonitrile.

Mass spectrometry: Data were collected with a MultiNotch MS3 TMT method (***McAlister et al., 2014***) using an Orbitrap Lumos mass spectrometer coupled to a Proxeon EASY-nLC 1200 Liquid Chromatography (LC) system (both Thermo Fisher Scientific). The used capillary column was packed with C_{18} resin (35 cm length, 100 µm inner diameter, 2.6 µm Accucore matrix (Thermo Fisher Scientific)). Peptides were separated for 3 or 4 hours over acidic acetonitrile gradients by LC prior to mass spectrometry (MS) analysis. Data from two 4 hour runs and one 3 hour run were recorded and combined. After an initial MS^1 scan (Orbitrap analysis; resolution 120,000; mass range 400−1400 Th), MS^2 analysis used collision-induced dissociation (CID, CE = 35) with a maximum ion injection time of 150–300 ms and an isolation window of 0.5 m/z. In order to obtain quantitative information, MS^3 precursors were fragmented by high-energy collision-induced dissociation (HCD) and analyzed in the Orbitrap at a resolution of 50,000 at 200 Th. Further details on LC and MS parameters and settings were described recently (***Paulo et al., 2016***).

pTF-mCherry and MS hit-EYFP co-expression and imaging

Episomes with pTF-mCherry and MS hit-EYFP expression cassettes were conjugated into WT *P. tricornutum* cells, resulting transconjugants genotyped and screened for fluorescence. 5 µL fluorescent cells in mid- to late-exponential phase were imaged with

settings that minimized cross-channel bleed-through: argon laser strength at 30%, 514 nm laser line at 50%, mCherry emission window: 620–640 nm, EYFP emission window: 520–540 nm, autofluorescence emission window: 700–750 nm.

pTF.CREGr protein expression conditions

Small scale expression testing for the Δ31_pTF.CREGr-His6 (and all other pTF.CREGr truncations not described here) was performed as follows: an overnight culture of BL21 *E. coli* cells carrying the PtpBAD-Δ31_pTF.CREGr-CTHF (=pJT_Δ31_pTF.CREGr-His6) expression vector was used to inoculate 50 mL of Terrific Broth (TB) supplemented with tetracycline (10 ug/mL). The cultures were grown in a shaking refrigerated incubator (37°C, 200 rpm) until OD_{600} of 0.4–0.6 was reached. The temperature in the shaker was then lowered to 18°C and flasks were allowed to adjust to this temperature for about 30 min before arabinose was added to a final concentration of 0.5% (w/v). Growth at 18°C was continued overnight (12–18 hrs) after which 10 mL of cultures were harvested by centrifugation at 6000 g for 10 min. Large scale Δ31_pTF.CREGr-His6 expression was performed as above, but at 1 L total volume. Following arabinose induction and overnight growth of the 1 L culture, 500 mL was harvested for further processing.

pTF.CREGr purification conditions

All purifications were performed using cobalt, Co^{2+}, TALON Metal Affinity Resin (Takara Bio, Catalog #635502) which binds His6-tagged proteins with high affinity. Pellets from small scale expression testing (10 mL culture) were resuspended in 800 uL of lysis buffer (50 mM NaH_2PO_4 (pH 7.5), 500 mM NaCl, 0.1% Triton X-100, 10 mM imidazole, 1 mg/mL lysozyme, and 10 μM β-mercaptoethanol) and subjected to microtip sonication on ice until lysis was complete. Lysates were then clarified by centrifugation (10 min, 15000 g, 4°C) and the supernatant (700 uL) was set aside. For each purification, 25 uL of TALON resin was equilibrated by washing three times with 10 volumes of lysis buffer and pelleting by centrifugation after each wash (30 sec, 3000 g, room temperature). After the third wash, 25 uL of lysis buffer was added to the resin to make a 50 uL slurry.

The supernatants from the previously clarified lysates were then added to the TALON resin slurry and incubated for 1 hr at room temperature with end-over-end mixing. Following the 1 hr incubation, the resin was pelleted by centrifugation (30 sec, 3000 g, room temperature) and washed three times with 10 volumes of wash buffer (50 mM NaH_2PO_4 (pH 7.5), 500 mM NaCl, and 30 mM imidazole). Protein was then eluted with 50 uL of elution buffer (50 mM NaH_2PO_4 (pH 7.5), 500 mM NaCl, 250 mM imidazole, and 10% glycerol) and was either subjected to immediate SDS-PAGE electrophoresis or stored at -80 °C.

Pellets from large scale expression (500 mL culture) were processed with a few modifications. For lysis, cell pellets were resuspended in 20 mL of lysis buffer. For purification, 1 mL of TALON resin (i.e. 2 mL of equilibrated slurry) was used in combination with approx. 19 mL of clarified lysate and allowed to incubate overnight at 8 °C with end-over-end mixing. Final protein elution was performed using 5 mL of elution buffer. Purified protein was subjected to SDS-PAGE electrophoresis and concentrated using an Amicon Ultra-15 10 kDa cutoff concentrator (Millipore Sigma, Burlington, MA, Catalog #UFC901024). Following concentration, buffer exchange into 20 mM HEPES (pH 8.0), 10% glycerol, 300 mM KCl was performed using a PD-10 desalting column (GE Healthcare, Chicago, IL, Catalog #17-0851-01). Protein was then concentrated again and total protein content was determined by the Bradford method using the Protein Assay Dye Reagent (Bio-Rad, Hercules, CA, Catalog #5000006).

pTF.CREGr enzymatic assay

To test for flavin reductase activity, we set up the following enzyme assay conditions in 100 uL final volume: 50 mM Tris-HCl (pH 7.5), 100 mM KCl, 10% glycerol, 30 µM flavin mononucleotide (FMN), 30 µM riboflavin, and 250 µM NADPH (***Coves and Fontecave, 1993***). Assays were started by the addition of 50 µg of purified pTF.CREGr or Milli-Q water in the case of the no-enzyme controls. Oxidation of NADPH was measured by the decrease in absorbance at 320 nm on a Flexstation 3 microtiter plate reader. To demonstrate the role of flavins as a substrate, we also omitted both FMN and flavins in

additional assays, both in the presence and absence of enzyme. All assays were conducted in triplicate.

Bioinformatic and data analyses

Mass spectrometry data analysis: Peptide-spectrum matches used a SEQUEST (v.28, rev. 12) algorithm (***Eng et al., 1994***). Data were searched against a size-sorted forward and reverse database of the *Phaeodactylum tricornutum* proteome (strain CCAP 1055/1, UniProt reference proteome UP000000759) with added common contaminant proteins and the pTF-APEX2 fusion protein sequence. Spectra were first converted to mzXML and searches were then performed using a mass tolerance of 50 ppm for precursors and a fragment ion tolerance of 0.9 Da. For the searches, maximally 2 missed cleavages per peptide were allowed. We searched dynamically for oxidized methionine residues (+15.9949 Da) and applied a target decoy database strategy. A false discovery rate (FDR) of 1% was set for peptide-spectrum matches following filtering by linear discriminant analysis (LDA) (***Beausoleil et al., 2006***; ***Huttlin et al., 2010***). The FDR for final collapsed proteins was 1%. Quantitative information on peptides was derived from MS^3 scans. Quant tables were generated with the following filter criteria: MS^2 isolation specificity of >70% for each peptide and a sum of TMT signal to noise (s/n) of >200 over all channels per peptide. Quant tables were exported to Excel and further processed therein. Details of the TMT intensity quantification method and additional search parameters applied were described previously (***Paulo et al., 2016***). Scaled proteomic data were subjected to two-way hierarchical clustering (Ward's method) using JMP software package. Vulcano plot with log_2-transformed average APEX2/WT ratios and associated P-values was made in R using ggplot2 data visualization package (***R Core Team, 2013***). Gene IDs corresponding to protein hits were inferred using Ensembl Phatr3 *P. tricornutum* genomic database.

Protein feature identification: Length, size and isoelectric point (pI) of proteins was determined with ProtParam (***Gasteiger et al., 2005***). Signal peptides and transmembrane regions were determined with SignalP 4.1 (***Petersen et al., 2011***) and

TMHMM Server v. 2.0 (***Krogh et al., 2001***), respectively. Protein localizations were predicted with a combination of tools: TargetP 1.1 (***Emanuelsson et al., 2000***), SignalP 4.1, and ASAFind (***Gruber et al., 2015***) (version 1.1.7). All putative chloroplastic localizations in the study mean a protein was predicted to be chloroplastic by ASAFind (state-of-the-art plastidial localization prediction tool for diatoms); chloroplast localization prediction confidences are noted. Low confidence prediction by ASAFind means that the protein satisfies the following filtering criteria: (1) it contains a signal peptide as detected by SignalP 4.1, (2) +1 position of ASAFind predicted cleavage site is phenylalanine (F), tryptophan (W), tyrosine (Y), or leucine (L) (making the protein "potentially plastid targeted"), and (3) one or both of the following is false: the ASAFind predicted cleavage site coincides with the SignalP 4.1 prediction and the transit peptide score is higher than 2. Peroxidase class prediction was done in RedoxiBase (***Savelli et al., 2019***).

Protein alignments: Protein sequences of APEX2-like *P. tricornutum* peroxidases were aligned with Clustal Omega (***Sievers et al., 2011***) using the default settings and CLC Sequence Viewer 7.7 (QUIAGEN) using the following parameters: Gap open cost: 10.0, Gap extension cost: 1.0, End gap cost: As any other, Alignment: Very accurate (slow). Putative substrate-binding loops were evaluated using the Clustal Omega alignment. Conserved motifs in protein alignments behind phylogenetic trees were displayed in CLC Sequence Viewer 7.7.

Phylogenetic analyses: Homologs of respective *P. tricornutum* proteins were retrieved from the National Center for Biotechnology Information (NCBI) non-redundant and the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) databases using the blastp algorithm (e-value threshold set to 1e-15) (***Altschul et al., 1990***). The blast search retrieved only a handful of homologs for pTF.ap1 and no homologs were found for pTF.CatCh1 even among closely related diatoms. Therefore, more sensitive HMMER (***Eddy, 1998***) searches were employed to extend the datasets, which were afterward aligned using the Localpair algorithm as implemented in MAFFT (***Katoh et al., 2002***). Ambiguously aligned regions, regions composed mostly of gaps as well as

short fragments were manually removed in SeaView 4 (***Guoy et al., 2009***). For each alignment, the maximum likelihood analysis was carried out in IQ-TREE (***Nguyen et al., 2015***) under the best-fitting substitution matrix as inferred by the built-in model finder. Branching support was estimated using “thorough” non-parametric bootstrap analysis from 500 replicates in IQ-TREE.

Identification of disordered protein regions: pTF.CatCh1 amino acid sequence was analyzed for the presence of disordered protein regions with PONDR (Predictor of Naturally Disordered Regions) VSL2 predictor (***Peng et al., 2006***) and IUPred2a long disorder prediction type (***Mészáros et al., 2018***).

Other data analyses: Amplex UltraRed assay and transcriptomic data were plotted with R using ggplot2 data visualization package (***R Core Team, 2013***). Biochemical assay data were plotted in Excel.

Acknowledgements

We would like to thank Andrea Thor and Mason Mackey for help with electron microscopy sample preparation and imaging; Marian Kalocsay for assistance with mass spectrometry experiments and data analyses; Pardis Gholami and Hong Zheng for laboratory assistance; Jeffrey D. Martell, Jeffrey B. McQuaid, Tyler Coale, and Sarah Smith for fruitful discussions. This project was supported by the Gordon and Betty Moore Foundation grants GBMF4958 (JT), GBMF3828 (AEA), and GBMF5006 (AEA), the National Science Foundation grants NSF-OCE-1756884 (AEA) and NSF-MCB-1818390 (AEA), and the United States Department of Energy Genomics Science program grant DE-SC0018344 (AEA). JKB was supported by the National Institutes of Health (NIH) Ruth L. Kirschstein Predoctoral Individual National Research Service Award F31 1F31ES030613-01. National Center for Microscopy and Imaging Research (NCMIR) is supported by the National Institute for General Medical Sciences (NIGMS) of the National Institutes of Health (NIH).

Additional information

Funding

Funder	Grant reference number	Author
Gordon and Betty Moore Foundation (GBMF), Marine Microbial Initiative (MMI)	GBMF4958	Jernej Turnšek*
National Institutes of Health (NIH) Ruth L. Kirschstein Predoctoral Individual National Research Service Award F31	1F31ES030613-01	John K. Brunson
Gordon and Betty Moore Foundation (GBMF), Marine Microbial Initiative (MMI)	GBMF3828	Andrew E. Allen
Gordon and Betty Moore Foundation (GBMF), Marine Microbial Initiative (MMI)	GBMF5006	Andrew E. Allen
National Science Foundation, Biological Oceanography	NSF-OCE-1756884	Andrew E. Allen
National Science Foundation, Molecular and Cellular Biology	NSF-MCB-1818390	Andrew E. Allen
Department of Energy (DOE), Office of Biological and Environmental Research (BER)	DE-SC0018344	Andrew E. Allen

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

*Jernej proposed the idea to use APEX2 in diatoms and wrote the proposal jointly with James J. Russell, Pamela A. Silver, and Julie A. Theriot.

Author ORCIDs

Jernej Turnšek https://orcid.org/0000-0002-9056-3565
Aleš Horák https://orcid.org/0000-0002-3364-7383
Vincent A. Bielinski https://orcid.org/0000-0002-8107-8752
Andrew E. Allen https://orcid.org/0000-0001-5911-6081

Additional files

Supplementary files

• Supplementary file 1. **Table S1:** Predicted endogenous biotinylated proteins are present at similar levels in WT and pTF-APEX2 proteomic samples. **Table S2:** Features of the three proteins co-expressed with pTF. **Table S3:** Molecular cloning primers. **Table S4:** Constructed *Phaeodactylum tricornutum* episomes and *Escherichia coli* vectors with expression cassette details and fusion protein amino acid sequences. **Table S5:** Key chemicals and reagents used in the study.

• Supplementary file 2. Raw Amplex UltraRed assay data.

• Supplementary file 3. Scaled quantitative mass spectrometry data.

• Supplementary file 4. Raw quantitative mass spectrometry data, transcriptomic data, and functional annotations for proteins with APEX2/WT ratios ≥ 1.5 and P-value ≤ 0.05.

• Supplementary file 5. Raw pTF.CREGr biochemical assay data.

Data availability

All data generated or analyzed during this study are included in the manuscript and supporting files.

Competing interests

The authors declare that no competing interests exist.

References

Allen AE, Laroche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, Finazzi G, Fernie AR, Bowler C. 2008. Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc Natl Acad Sci U S A* **105**:10438–10443. DOI: https://doi.org/10.1073/pnas.0711370105, PMID: 18653757

Allen AE, Moustafa A, Montsant A, Eckert A, Kroth PG, Bowler C. 2012. Evolution and Functional Diversification of Fructose Bisphosphate Aldolase Genes in Photosynthetic Marine Diatoms. *Mol Biol Evol* **29**:367–379. DOI: https://doi.org/10.1093/molbev/msr223, PMID: 21903677

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**:403–410. DOI: https://doi.org/10.1016/S0022-2836(05)80360-2, PMID: 2231712

Anonymous. 1703. Two letters from a Gentleman in the Country, relating to Mr. Leeuwenhoek's Letter in Transaction, no. 283. Communicated by Mr. C. *Philos Trans R Soc Lond B* **23**:1494–1501. DOI: https://doi.org/10.1098/rstl.1702.0065

Archibald JM. 2009. The puzzle of plastid evolution. *Curr Biol* **19**:R81–8. DOI: https://doi.org/10.1016/j.cub.2008.11.067, PMID: 19174147

Aron AT, Ramos-Torres KM, Cotruvo JA Jr, Chang CJ. 2015. Recognition- and Reactivity-Based Fluorescent Probes for Studying Transition Metal Signaling in Living Systems. *Acc Chem Res* **48**:2434–2442. DOI: https://doi.org/10.1021/acs.accounts.5b00221, PMID: 26215055

Ardyna M, Lacour L, Sergi S, d'Ovidio F, Sallée J-B, Rembauville M, Blain S, Tagliabue A, Schlitzer R, Jeandel C, Arrigo KR, Claustre H. 2019. Hydrothermal vents trigger massive phytoplankton blooms in the Southern Ocean. *Nat Commun* **10**:2451. DOI: https://doi.org/10.1038/s41467-019-09973-6, PMID: 31165724

Balk J, Schaedler TA. 2014. Iron Cofactor Assembly in Plants. *Annu Rev Plant Biol* **65**:125–153. DOI: https://doi.org/10.1146/annurev-arplant-050213-035759, PMID: 24498975

Baslam M, Oikawa K, Kitajima-Koga A, Kaneko K, Mitsui T. 2016. Golgi-to-plastid trafficking of proteins through secretory pathway: Insights into vesicle-mediated import toward the plastids. *Plant Signal Behav* **11**:e1221558. DOI: https://doi.org/10.1080/15592324.2016.1221558, PMID: 27700755

1400 **Beausoleil SA**, Villén J, Gerber SA, Rush J, Gygi SP. 2006. A probability-based approach for high-
1401 throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* **24**:1285–1292. DOI:
1402 https://doi.org/10.1038/nbt1240, PMID: 16964243
1403
1404 **Benoiston A-S**, Ibarbalz FM, Bittner L, Guidi L, Jahn O, Dutkiewicz S, Bowler C. 2017. The evolution of
1405 diatoms and their biogeochemical functions. *Philos Trans R Soc Lond B Biol Sci* **372**:20160397. DOI:
1406 https://doi.org/10.1098/rstb.2016.0397, PMID: 28717023
1407
1408 **Bersuker K**, Peterson CWH, To M, Sahl SJ, Savikhin V, Grossman EA, Nomura DK, Olzmann JA. 2018.
1409 A Proximity Labeling Strategy Provides Insights into the Composition and Dynamics of Lipid Droplet
1410 Proteomes. *Dev Cell* **44**:97–112. DOI: https://doi.org/10.1016/j.devcel.2017.11.020, PMID: 29275994
1411
1412 **Bertrand EM**, McCrow JP, Moustafa A, Zheng H, McQuaid JB, Delmont TO, Post AF, Sipler RE,
1413 Spackeen JL, Xu K, Bronk DA, Hutchins DA, Allen AE. 2015. Phytoplankton–bacterial interactions
1414 mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc Natl Acad Sci U S A*
1415 **112**:9938–9943. DOI: https://doi.org/10.1073/pnas.1501615112, PMID: 26221022
1416
1417 **Blaby-Haas CE**, Merchant SS. 2012. The ins and outs of algal metal transport. *Biochim Biophys Acta*
1418 **1823**:1531–1552. DOI: https://doi.org/10.1016/j.bbamcr.2012.04.010, PMID: 22569643
1419
1420 **Blaby-Haas CE**, Merchant SS. 2014. Lysosome-related Organelles as Mediators of Metal Homeostasis. *J*
1421 *Biol Chem* **289**:28129–28136. DOI: https://doi.org/10.1074/jbc.R114.592618, PMID: 25160625
1422
1423 **Blaby-Haas CE**, Padilla-Benavides T, Stübe R, Argüello JM, Merchant SS. 2014. Evolution of a plant-
1424 specific copper chaperone family for chloroplast copper homeostasis. *Proc Natl Acad Sci U S A*
1425 **111**:E5480–E5487. DOI: https://doi.org/10.1073/pnas.1421545111, PMID: 25468978
1426
1427 **Bogdan AR**, Miyazawa M, Hashimoto K, Tsuji Y. 2016. Regulators of Iron Homeostasis: New Players in
1428 Metabolism, Cell Death, and Disease. *Trends Biochem Sci* **41**:274–286. DOI:
1429 https://doi.org/10.1016/j.tibs.2015.11.012, PMID: 26725301
1430
1431 **Boiteau RM**, Mende DR, Hawco NJ, McIlvin MR, Fitzsimmons JN, Saito MA, Sedwick PN, DeLong EF,
1432 Repeta DJ. 2016. Siderophore-based microbial adaptations to iron scarcity across the eastern Pacific
1433 Ocean. *Proc Natl Acad Sci U S A* **113**:14237–14242. DOI: https://doi.org/10.1073/pnas.1608594113,
1434 PMID: 27911777

1435 **Bölter B**. 2018. En route into chloroplasts: preproteins' way home. *Photosynth Res* **138**:263–275. DOI:
1436 https://doi.org/10.1007/s11120-018-0542-8, PMID: 29943212
1437
1438 **Bonfio C**, Valer L, Scintilla S, Shah S, Evans DJ, Jin L, Szostak JW, Sasselov DD, Sutherland JD, Mansy
1439 SS. 2017. UV-light-driven prebiotic synthesis of iron-sulfur clusters. *Nat Chem* **9**:1229–1234. DOI:
1440 https://doi.org/10.1038/nchem.2817, PMID: 29168482
1441
1442 **Boycheva S**, Daviet L, Wolfender J-L, Fitzpatrick TB. 2014. The rise of operon-like gene clusters in
1443 plants. *Trends Plant Sci* **19**:447–459. DOI: https://doi.org/10.1016/j.tplants.2014.01.013, PMID: 24582794
1444
1445 **Boyd PW**, Ellwood MJ. 2010. The biogeochemical cycle of iron in the ocean. *Nat Geosci* **3**:675–682.
1446 DOI: https://doi.org/10.1038/ngeo964
1447
1448 **Bowler C**, (75 authors), Grigoriev IV. 2008. The *Phaeodactylum* genome reveals the evolutionary history
1449 of diatom genomes. *Nature* **456**:239–244. DOI: https://doi.org/10.1038/nature07410, PMID: 18923393
1450
1451 **Branon TC**, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, Feldman JL, Perrimon N, Ting AY.
1452 2018. Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* **36**:880–887.
1453 DOI: https://doi.org/10.1038/nbt.4201, PMID: 30125270
1454
1455 **Brett CL**, Donowitz M, Rao R. 2006. Does the proteome encode organellar pH? *FEBS Lett* **580**:717–719.
1456 DOI: https://doi.org/10.1016/j.febslet.2005.12.103, PMID: 16413548
1457
1458 **Brunson JK**, McKinnie SMK, Chekan JR, McCrow JP, Miles ZD, Bertrand EM, Bielinski VA, Luhavaya H,
1459 Oborník M, Smith GJ, Hutchins DA, Allen AE, Moore BS. 2018. Biosynthesis of the neurotoxin domoic
1460 acid in a bloom-forming diatom. *Science* **361**:1356–1358. DOI: https://doi.org/10.1126/science.aau0382,
1461 PMID: 30262498
1462
1463 **Brzezinski MA**, Krause JW, Bundy RM, Barbeau KA, Franks P, Goericke R, Landry MR, Stukel MR.
1464 2015. Enhanced silica ballasting from iron stress sustains carbon export in a frontal zone within the
1465 California Current. *J Geophys Res C: Oceans* **120**:4654–4669. DOI: https://doi.org/10.1002/
1466 2015JC010829.
1467
1468 **Camacho A**, Walter XA, Picazo A, Zopfi J. 2017. Photoferrotrophy: Remains of an Ancient
1469 Photosynthesis in Modern Environments. *Front Microbiol* **8**:323. DOI:
1470 https://doi.org/10.3389/fmicb.2017.00323, PMID: 28377745

1471 **Caron DA**, Alexander H, Allen AE, Archibald JM, Armbrust EV, Bachy C, Bell CJ, Bharti A, Dyhrman ST,
1472 Guida SM, Heidelberg KB, Kaye JZ, Metzner J, Smith SR, Worden AZ. 2017. Probing the evolution,
1473 ecology and physiology of marine protists using transcriptomics. *Nat Rev Microbiol* **15**:6–20. DOI:
1474 https://doi.org/10.1038/nrmicro.2016.160, PMID: 27867198
1475
1476 **Carra S**, Alberti S, Benesch JLP, Boelens W, Buchner J, Carver JA, Cecconi C, Ecroyd H, Gusev N,
1477 Hightower LE, Klevit RE, Lee HO, Liberek K, Lockwood B, Poletti A, Timmerman V, Toth ME, Vierling E,
1478 Wu T, Tanguay RM. 2019. Small heat shock proteins: multifaceted proteins with important implications for
1479 life. *Cell Stress Chaperones* **24**:295–308. DOI: https://doi.org/ 10.1007/s12192-019-00979-z, PMID:
1480 30758704
1481
1482 **Carradec Q**, (37 authors), Wincker P. 2018. A global ocean atlas of eukaryotic genes. *Nat Commun*
1483 **9**:373. DOI: https://doi.org/10.1038/s41467-017-02342-1, PMID: 29371626
1484
1485 **Cheng Y**, Zak O, Aisen P, Harrison SC, Walz T. 2004. Structure of the Human Transferrin Receptor-
1486 Transferrin Complex. *Cell* **116**:565–576. DOI: https://doi.org/10.1016/s0092-8674(04)00130-8, PMID:
1487 14980223
1488
1489 **Coale TH**, Horák A., Oborník M., Allen AE. (In Press) Reduction-dependent siderophore assimilation in a
1490 model pennate diatom. (In Press, *Proc Natl Acad Sci U S A*).
1491
1492 **Cohen NR**, A. Ellis K, Burns WG, Lampe RH, Schuback N, Johnson Z, Sañudo-Wilhelmy S, Marchetti A.
1493 2017. Iron and vitamin interactions in marine diatom isolates and natural assemblages of the Northeast
1494 Pacific Ocean: Iron and vitamin interactions in diatoms. *Limnol Oceanogr* **62**:2076–2096. DOI:
1495 https://doi.org/10.1002/lno.10552
1496
1497 **Cohen NR**, Ellis KA, Lampe RH, McNair H, Twining BS, Maldonado MT, Brzezinski MA, Kuzminov FI,
1498 Thamatrakoln K, Till CP, Bruland KW, Sunda WG, Bargu S, Marchetti A. 2017. Diatom Transcriptional
1499 and Physiological Responses to Changes in Iron Bioavailability across Ocean Provinces. *Frontiers in*
1500 *Marine Science* **4**:360. DOI: https://doi.org/10.3389/fmars.2017.00360
1501
1502 **Cohen NR**, Gong W, Moran DM, Mcilvin MR, Saito MA, Marchetti A. 2018. Transcriptomic and proteomic
1503 responses of the oceanic diatom *Pseudo-nitzschia granii* to iron limitation. *Environ Microbiol* **20**:3109–
1504 3126. DOI: https://doi.org/10.1111/1462-2920.14386, PMID: 30117243
1505
1506 **Cong Q**, Anishchenko I, Ovchinnikov S, Baker D. 2019. Protein interaction networks revealed by
1507 proteome coevolution. *Science* **365**:185–189. DOI: https://doi.org/10.1126/science.aaw6718, PMID:
1508 31296772

Conrad ME, Umbreit JN, Moore EG. 1993. Rat Duodenal Iron-Binding Protein Mobilferrin is a Homologue of Calreticulin. *Gastroenterology* **104**:1700–1704. DOI: https://doi.org/10.1016/0016-5085(93)90648-v, PMID: 8500729

Conrad ME, Umbreit JN, Moore EG, Heiman D. 1996. Mobilferrin is an intermediate in iron transport between transferrin and hemoglobin in K562 cells. *J Clin Invest* **98**:1449–1454. DOI: https://doi.org/10.1172/JCI118933, PMID: 8823311

Coves J, Fontecave M. 1993. Reduction and mobilization of iron by a NAD(P)H:flavin oxidoreductase from *Escherichia coli*. *Eur J Biochem* **211**:635–641. DOI: https://doi.org/10.1111/j.1432-1033.1993.tb17591.x, PMID: 8436123

Crichton R. 2016. Iron Metabolism: From Molecular Mechanisms to Clinical Consequences, Fourth Edition. John Wiley & Sons, Ltd., Hoboken, NJ, USA.1–556. DOI: https://doi.org/10.1002/9781118925645

de Baar HJW, (32 authors), Wong CS. 2005. Synthesis of iron fertilization experiments: From the Iron Age in the Age of Enlightenment. *J Geophys Res* **110**:C09S16. DOI: https://doi.org/10.1029/2004JC002601

de Vargas C, (52 authors), Karsenti E. 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**:1261605 DOI: https://doi.org/10.1126/science.1261605, PMID: 25999516

Del Olmo T, Lauzier A, Normandin C, Larcher R, Lecours M, Jean D, Lessard L, Steinberg F, Boisvert F-M, Jean S. 2019. APEX2-mediated RAB proximity labeling identifies a role for RAB21 in clathrin-independent cargo sorting. *EMBO Rep* **20**. DOI: https:// doi.org/10.15252/embr.201847192, PMID: 30610016

Dickinson MS, Anderson LN, Webb-Robertson B-JM, Hansen JR, Smith RD, Wright AT, Hybiske K. 2019. Proximity-dependent proteomics of the *Chlamydia trachomatis* inclusion membrane reveals functional interactions with endoplasmic reticulum exit sites. *PLoS Pathog* **15**:e1007698. DOI: https://doi.org/10.1371/journal.ppat.1007698, PMID: 30943267

Diner RE, Bielinski VA, Dupont CL, Allen AE, Weyman PD. 2016. Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-Based DNA Delivery Methods. *Front Bioeng Biotechnol* **4**:65. DOI: https:// doi.org/10.3389/fbioe.2016.00065, PMID: 27551676

Durkin CA, Marchetti A, Bender SJ, Truong T, Morales R, Mock T, Armbrust EV. 2012. Frustule-related gene transcription and the influence of diatom community composition on silica precipitation in an iron-limited environment. *Limnol Oceanogr* **57**:1619–1633. DOI: https://doi.org/10.4319/lo.2012.57.6.1619

Eckenroth BE, Steere AN, Chasteen ND, Everse SJ, Mason AB. 2011. How the binding of human transferrin primes the transferrin receptor potentiating iron release at endosomal pH. *Proc Natl Acad Sci U S A* **108**:13089–13094. DOI: https://doi.org/10.1073/pnas.1105786108, PMID: 21788477

Eddy SR. 1998. Profile hidden Markov models. *Bioinformatics* **14**:755–763. DOI: https://doi.org/10.1093/bioinformatics/14.9.755, PMID: 9918945

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res* **47**:D427–D432. DOI: https://doi.org/10.1093/nar/gky995, PMID: 30357350

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. *J Mol Biol* **300**:1005–1016. DOI: https://doi.org/10.1006/jmbi.2000.3903, PMID: 10891285

Eng JK, McCormack AL, Yates JR. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**:976–989. DOI: https://doi.org/10.1016/1044-0305(94)80016-2, PMID: 24226387

Erdős G, Mészáros B, Reichmann D, Dosztányi Z. 2019. Large-Scale Analysis of Redox-Sensitive Conditionally Disordered Protein Regions Reveals Their Widespread Nature and Key Roles in High-Level Eukaryotic Processes. *Proteomics* **19**:1800070. DOI: https://doi.org/10.1002/pmic.201800070, PMID: 30628183

Faktorová D., (104 authors), Lukeš, J. 2019. Genetic tool development in marine protists: Emerging model organisms for experimental cell biology. *bioRxiv*. DOI: https://doi.org/10.1101/718239

Fazal FM, Han S, Parker KR, Kaewsapsak P, Xu J, Boettiger AN, Chang HY, Ting AY. 2019. Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* **178**:473–490. DOI: https://doi.org/10.1016/j.cell.2019.05.027, PMID: 31230715

Finbow ME, Harrison MA. 1997. The vacuolar H^+-ATPase: a universal proton pump of eukaryotes. *Biochem J* **324**:697–712. DOI: https://doi.org/10.1042/bj3240697, PMID: 9210392

Fontecave M, Covès J, Pierre JL. 1994. Ferric reductases or flavin reductases? *Biometals* **7**:3–8. DOI: https://doi.org/10.1007/BF00205187, PMID: 8118169

Francius G, Tesson B, Dague E, Martin-Jézéquel V, Dufrêne YF. 2008. Nanostructure and nanomechanics of live *Phaeodactylum tricornutum* morphotypes. *Environ Microbiol* **10**:1344–1356. DOI: https://doi.org/10.1111/j.1462-2920.2007.01551.x, PMID: 18248452

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. (In:) The Proteomics Protocols Handbook, (Editor:) John M. Walker. Humana Press, Totowa, NJ, USA. 571–607. DOI: https://doi.org/10.1385/1-59259-890-0:571

Ghobrial G, Araujo L, Jinwala F, Li S, Lee LY. 2018. The Structure and Biological Function of CREG. *Front Cell Dev Biol* **6**:136. DOI: https://doi.org/10.3389/fcell.2018.00136, PMID: 30416997

Gibson DG, Young L, Chuang R-Y, Craig Venter J, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**:343–345. DOI: https://doi.org/10.1038/nmeth.1318, PMID: 19363495

Go CD, Knight JDR, Rajasekharan A, Rathod B, Hesketh GG, Abe KT, Youn J-Y, Samavarchi-Tehrani P, Zhang H, Zhu LY, Popiel E, Lambert J-P, Coyaud É, Cheung SWT, Rajendran D, Wong CJ, Antonicka H, Pelletier L, Raught B, Palazzo AF, Shoubridge EA, Gingras A-C. A proximity biotinylation map of a human cell. *bioRxiv.* DOI: https://doi.org/10.1101/796391

Gröger P, Poulsen N, Klemm J, Kröger N, Schlierf M. 2016. Establishing super-resolution imaging for proteins in diatom biosilica. *Sci Rep* **6**:36824. DOI: https://doi.org/10.1038/srep36824, PMID: 27827427

Gruber A, Rocap G, Kroth PG, Armbrust EV, Mock T. 2015. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J* **81**:519–528. DOI: https://doi.org/10.1111/tpj.12734, PMID: 25438865

Gouy M, Guindon S, Gascuel O. 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol Biol Evol* **27**:221–224. DOI: https://doi.org/10.1093/molbev/msp259, PMID: 19854763

Han S, Udeshi ND, Deerinck TJ, Svinkina T, Ellisman MH, Carr SA, Ting AY. 2017. Proximity Biotinylation as a Method for Mapping Proteins Associated with mtDNA in Living Cells. *Cell Chem Biol* **24**:404–414. DOI: https://doi.org/10.1016/j.chembiol.2017.02.002, PMID: 28238724

Hildebrand M, Lerch SJL, Shrestha RP. 2018. Understanding Diatom Cell Wall Silicification—Moving Forward. *Frontiers in Marine Science* **5**:125. DOI: https://doi.org/10.3389/fmars.2018.00125

Hogle SL, Barbeau KA, Gledhill M. 2014. Heme in the marine environment: from cells to the iron cycle. *Metallomics* **6**:1107–1120. DOI: https://doi.org/10.1039/c4mt00031e, PMID: 24811388

Hong-Hermesdorf A, Miethke M, Gallaher SD, Kropat J, Dodani SC, Chan J, Barupala D, Domaille DW, Shirasaki DI, Loo JA, Weber PK, Pett-Ridge J, Stemmler TL, Chang CJ, Merchant SS. 2014. Subcellular metal imaging identifies dynamic sites of Cu accumulation in *Chlamydomonas*. *Nat Chem Biol* **10**:1034–1042. DOI: https://doi.org/10.1038/nchembio.1662, PMID: 25344811

Hung V, Udeshi ND, Lam SS, Loh KH, Cox KJ, Pedram K, Carr SA, Ting AY. 2016. Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nat Protoc* **11**:456–475. DOI: https://doi.org/10.1038/nprot.2016.018, PMID: 26866790

Hutchins DA, Boyd PW. 2016. Marine phytoplankton and the changing ocean iron cycle. *Nat Clim Chang* **6**:1072–1079. DOI: https://doi.org/10.1038/nclimate3147

Hutchins DA, Bruland KW. 1998. Iron-limited diatom growth and Si:N uptake ratios in a coastal upwelling regime. *Nature* **393**:561–564. DOI: https://doi.org/10.1038/31203

Huttlin EL, Jedrychowski MP, Elias JE, Goswami T, Rad R, Beausoleil SA, Villén J, Haas W, Sowa ME, Gygi SP. 2010. A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression. *Cell* **143**:1174–1189. DOI: https://doi.org/10.1016/j.cell.2010.12.001, PMID: 21183079

Huysman MJJ, Vyverman W, De Veylder L. 2014. Molecular regulation of the diatom cell cycle. *J Exp Bot* **65**:2573–2584. DOI: https://doi.org/10.1093/jxb/ert387, PMID: 24277280

Hwang J, Espenshade PJ. 2016. Proximity-dependent biotin labelling in yeast using the engineered ascorbate peroxidase APEX2. *Biochem J* **473**:2463–2469. DOI: https://doi.org/10.1042/BCJ20160106, PMID: 27274088

Hwang J, Ribbens D, Raychaudhuri S, Cairns L, Gu H, Frost A, Urban S, Espenshade PJ. 2016. A Golgi rhomboid protease Rbd2 recruits Cdc48 to cleave yeast SREBP. *EMBO J* **35**:2332–2349. DOI: https://doi.org/10.15252/embj.201693923, PMID: 27655872

Jeffery CJ. 2018. Protein moonlighting: what is it, and why is it important? *Philos Trans R Soc Lond B Biol Sci* **373**:20160523. DOI: https://doi.org/10.1098/rstb.2016.0523, PMID: 29203708

Jin L, Engelhart AE, Zhang W, Adamala K, Szostak JW. 2018. Catalysis of Template-Directed Nonenzymatic RNA Copying by Iron(II). *J Am Chem Soc* **140**:15016–15021. DOI: https://doi.org/10.1021/jacs.8b09617, PMID: 30335371

Kalocsay M. 2019. APEX Peroxidase-Catalyzed Proximity Labeling and Multiplexed Quantitative Proteomics. (In:) Proximity Labeling, (Editors:) Murat Sunbul and Andreas Jäschke. Human Press, Totawa, NJ, USA. 41–55. DOI: https://doi.org/10.1007/978-1-4939-9537-0_4, PMID: 31124087

Karas BJ, Diner RE, Lefebvre SC, McQuaid J, Phillips APR, Noddings CM, Brunson JK, Valas RE, Deerinck TJ, Jablanovic J, Gillard JTF, Beeri K, Ellisman MH, Glass JI, Hutchison CA 3rd, Smith HO, Venter JC, Allen AE, Dupont CL, Weyman PD. 2015. Designer diatom episomes delivered by bacterial conjugation. *Nat Commun* **6**:6925. DOI: https://doi.org/10.1038/ncomms7925, PMID: 25897682

Katoh K, Misawa K, Kuma K-I, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**:3059–3066. DOI: https://doi.org10.1093/nar/gkf436/, PMID: 12136088

Kazamia E, Sutak R, Paz-Yepes J, Dorrell RG, Vieira FRJ, Mach J, Morrissey J, Leon S, Lam F, Pelletier E, Camadro J-M, Bowler C, Lesuisse E. 2018. Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. *Sci Adv* **4**:eaar4536. DOI: https://doi.org/10.1126/sciadv.aar4536, PMID: 29774236

Kazamia E, Mach J, McQuaid J, Olsinova M, Machan R, Coale T, Lesuisse E, Allen, A, Sutak R, Bowler, C. 2019. Monitoring the localization of three iron starvation induced proteins (ISIPs) in a single transformant of *Phaeodactylum tricornutum* under a range of iron supplementation regimes. The molecular life of diatoms Programme and Abstract Book: 95–96. URL: http://meetings.embo.org/files/abstractbooks/Program_%20Abstract%20book_MLD_5_for_print.pdf

Keeling PJ, (79 authors), Worden AZ. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans Through Transcriptome Sequencing. *PLoS Biol* **12**:e1001889. DOI: https://doi.org/10.1371/journal.pbio.1001889, PMID: 24959919

1690 **Kilian O**, Kroth PG. 2005. Identification and characterization of a new conserved motif within the
1691 presequence of proteins targeted into complex diatom plastids. *Plant J* **41**:175–183. DOI:
1692 https://doi.org/10.1111/j.1365-313X.2004.02294.x, PMID: 15634195
1693
1694 **Kim T-W**, Park CH, Hsu C-C, Zhu J-Y, Hsiao Y, Branon T, Xu S-L, Ting AY, Wang Z-Y. 2019. Application
1695 of TurboID-mediated proximity labeling for mapping a GSK3 kinase signaling network in *Arabidopsis*.
1696 *bioRxiv*. DOI: https://doi.org/10.1101/636324
1697
1698 **Kitadai N**, Nakamura R, Yamamoto M, Takai K. 2019. Metals likely promoted protometabolism in early
1699 ocean alkaline hydrothermal systems. *Sci Adv* **5**:eaav7848. DOI: https://doi.org/10.1126/sciadv.aav7848,
1700 PMID: 31223650
1701
1702 **Knoll AH**, Bergmann KD, Strauss JV. 2016. Life: the first two billion years. *Philos Trans R Soc Lond B*
1703 *Biol Sci* **371**:20150493. DOI: https://doi.org/10.1098/rstb.2015.0493, PMID: 27672146
1704
1705 **Kohli P**, Höhne M, Jüngst C, Bertsch S, Ebert LK, Schauss AC, Benzing T, Rinschen MM, Schermer B.
1706 2017. The ciliary membrane-associated proteome reveals actin-binding proteins as key components of
1707 cilia. *EMBO Rep* **18**:1521–1535. DOI: https://doi.org/10.15252/embr.201643846, PMID: 28710093
1708
1709 **Komeili A.** 2018. Molecular Mechanisms of Organelle Formation in Bacteria. Grantome research
1710 proposal Abstract. URL: http://grantome.com/grant/NIH/R37-GM042143-27
1711
1712 **Kröger N**, Brunner E. 2014. Complex-shaped microbial biominerals for nanotechnology. *Wiley Interdiscip*
1713 *Rev Nanomed Nanobiotechnol* **6**:615–627. DOI: https://doi.org/10.1002/wnan.1284, PMID: 25154474
1714
1715 **Krogh A**, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology
1716 with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**:567–580. DOI:
1717 https://doi.org/10.1006/jmbi.2000.4315, PMID: 11152613
1718
1719 **Kroth PG**, Bones AM, Daboussi F, Ferrante MI, Jaubert M, Kolot M, Nymark M, Río Bártulos C, Ritter A,
1720 Russo MT, Serif M, Winge P, Falciatore A. 2018. Genome editing in diatoms: achievements and goals.
1721 *Plant Cell Rep* **37**:1401–1408. DOI: https://doi.org/10.1007/s00299-018-2334-1, PMID: 30167805
1722
1723 **Kudla J**, Bock R. 2016. Lighting the Way to Protein-Protein Interactions: Recommendations on Best
1724 Practices for Bimolecular Fluorescence Complementation Analyses. *Plant Cell* **28**:1002–1008. DOI:
1725 https://doi.org/10.1105/tpc.16.00043, PMID: 27099259

Lam SS, Martell JD, Kamer KJ, Deerinck TJ, Ellisman MH, Mootha VK, Ting AY. 2015. Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods* **12**:51–54. DOI: https://doi.org/10.1038/nmeth.3179, PMID: 25419960

Leynaert A, Bucciarelli E, Claquin P, Dugdale RC, Martin-Jézéquel V, Pondaven P, Ragueneau O. 2004. Effect of iron deficiency on diatom cell size and silicic acid uptake kinetics. *Limnol Oceanogr* **49**:1134–1143. DOI: https://doi.org/10.4319/lo.2004.49.4.1134

López-Millán AF, Duy D, Philippar K. 2016. Chloroplast Iron Transport Proteins - Function and Impact on Plant Physiology. *Front Plant Sci* **7**:178. DOI: https://doi.org/10.3389/fpls.2016.00178, PMID: 27014281

Lommer M, Specht M, Roy A-S, Kraemer L, Andreson R, Gutowska MA, Wolf J, Bergner SV, Schilhabel MB, Klostermeier UC, Beiko RG, Rosenstiel P, Hippler M, LaRoche J. 2012. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol* **13**:R66. DOI: https://doi.org/10.1186/gb-2012-13-7-r66, PMID: 22835381

Lundberg E, Borner GHH. 2019. Spatial proteomics: a powerful discovery tool for cell biology. *Nat Rev Mol Cell Biol* **20**:285–302. DOI: https://doi.org/10.1038/s41580-018-0094-y, PMID: 30659282

Mair A, Xu S-L, Branon TC, Ting AY, Bergmann DC. 2019. Proximity labeling of protein complexes and cell type-specific organellar proteomes in Arabidopsis enabled by TurboID. *eLife* [Epub ahead of print]. DOI: https://doi.org/10.7554/eLife.47864, PMID: 31535972

Marchand J, Heydarizadeh P, Schoefs B, Spetea C. 2018. Ion and metabolite transport in the chloroplast of algae: lessons from land plants. *Cell Mol Life Sci* **75**:2153–2176. DOI: https://doi.org/10.1007/s00018-018-2793-0, PMID: 29541792

Marchetti A, Schruth DM, Durkin CA, Parker MS, Kodner RB, Berthiaume CT, Morales R, Allen AE, Armbrust EV. 2012. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc Natl Acad Sci U S A* **109**:E317–E325. DOI: https://doi.org/10.1073/pnas.1118408109, PMID: 22308424

Markmiller S, Soltanieh S, Server KL, Mak R, Jin W, Fang MY, Luo E-C, Krach F, Yang D, Sen A, Fulzele A, Wozniak JM, Gonzalez DJ, Kankel MW, Gao F-B, Bennett EJ, Lécuyer E, Yeo GW. 2018. Context-Dependent and Disease-Specific Diversity in Protein Interactions within Stress Granules. *Cell* **172**:590–604. DOI: https://doi.org/10.1016/j.cell.2017.12.032, PMID: 29373831

Marshall KE, Robinson EW, Hengel SM, Paša-Tolić L, Roesijadi G. 2012. FRET Imaging of Diatoms Expressing a Biosilica-Localized Ribose Sensor. *PLoS One* **7**:e33771. DOI: https://doi.org/10.1371/journal.pone.0033771, PMID: 22470473

Martell JD, Deerinck TJ, Lam SS, Ellisman MH, Ting AY. 2017. Electron microscopy using the genetically encoded APEX2 tag in cultured mammalian cells. *Nat Protoc* **12**:1792–1816. DOI: https://doi.org/10.1038/nprot.2017.065, PMID: 28796234

Martell JD, Deerinck TJ, Sancak Y, Poulos TL, Mootha VK, Sosinsky GE, Ellisman MH, Ting AY. 2012. Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat Biotechnol* **30**:1143–1148. DOI: https://doi.org/10.1038/nbt.2375, PMID: 23086203

Martin JH, (42 authors), Tindale NW. 1994. Testing the iron hypothesis in ecosystems of the equatorial Pacific Ocean. *Nature* **371**:123–129. DOI: https://doi.org/10.1038/371123a0

Matsuda Y, Hopkinson BM, Nakajima K, Dupont CL, Tsuji Y. 2017. Mechanisms of carbon dioxide acquisition and CO_2 sensing in marine diatoms: a gateway to carbon metabolism. *Philos Trans R Soc Lond B Biol Sci*. B **372**:20160403 DOI: https://doi.org/10.1098/rstb.2016.0403, PMID: 28717013

May T, Soll J. 1998. Positive Charges Determine the Topology and Functionality of the Transmembrane Domain in the Chloroplastic Outer Envelope Protein Toc34. *J Cell Biol* **141**:895–904. DOI: https://doi.org/10.1083/jcb.141.4.895, PMID: 9585409

May T, Soll J. 2000. 14-3-3 Proteins Form a Guidance Complex with Chloroplast Precursor Proteins in Plants. *Plant Cell* **12**:53–64. DOI: https://doi.org/10.1105/tpc.12.1.53, PMID: 10634907

Maxson ME, Grinstein S. 2014. The vacuolar-type H^+-ATPase at a glance – more than a proton pump. *J Cell Sci* **127**:4987–4993. DOI: https://doi.org/10.1242/jcs.158550, PMID: 25453113

Mayle KM, Le AM, Kamei DT. 2012. The Intracellular Trafficking Pathway of Transferrin. *Biochim Biophys Acta* **1820**:264–281. DOI: https://doi.org/10.1016/j.bbagen.2011.09.009, PMID: 21968002

McAlister GC, Nusinow DP, Jedrychowski MP, Wühr M, Huttlin EL, Erickson BK, Rad R, Haas W, Gygi SP. 2014. MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression Across Cancer Cell Line Proteomes. *Anal Chem* **86**:7150–7158. DOI: https://doi.org/10.1021/ac502040v, PMID: 24927332

McQuaid JB, Kustka AB, Oborník M, Horák A, McCrow JP, Karas BJ, Zheng H, Kindeberg T, Andersson AJ, Barbeau KA, Allen AE. 2018. Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. *Nature* **555**:534–537. DOI: https://doi.org/10.1038/nature25982, PMID: 29539640

Merchant S. 2019. Nutritional Copper Signaling and Homeostasis. Grantome research proposal Abstract. URL: http://grantome.com/grant/NIH/R35-GM127114-01

Merkulova M, Hurtado-Lorenzo A, Hosokawa H, Zhuang Z, Brown D, Ausiello DA, Marshansky V. 2011. Aldolase directly interacts with ARNO and modulates cell morphology and acidic vesicle distribution. *Am J Physiol Cell Physiol* **300**:C1442–C1455. DOI: https://doi.org/10.1152/ajpcell.00076.2010, PMID: 21307348

Merkulova M, Păunescu TG, Azroyan A, Marshansky V, Breton S, Brown D. 2015. Mapping the H^+ (V)-ATPase interactome: identification of proteins involved in trafficking, folding, assembly and phosphorylation. *Sci Rep* **5**:14827. DOI: https://doi.org/10.1038/srep14827, PMID: 26442671

Mészáros B, Erdos G, Dosztányi Z. 2018. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res* **46**:W329–W337. DOI: https://doi.org/10.1093/nar/gky384, PMID: 29860432

Morel FMM, Kustka AB, Shaked Y. 2008. The role of unchelated Fe in the iron nutrition of phytoplankton. *Limnol Oceanogr* **53:**400–404. DOI: doi.org/10.4319/lo.2008.53.1.0400

Moreno CM, Lin Y, Davies S, Monbureau E, Cassar N, Marchetti A. 2018. Examination of gene repertoires and physiological responses to iron and light limitation in Southern Ocean diatoms. *Polar Biol* **41**:679–696. DOI: https://doi.org/10.1007/s00300-017-2228-7

Morrissey J, Baxter IR, Lee J, Li L, Lahner B, Grotz N, Kaplan J, Salt DE, Guerinot ML. 2009. The Ferroportin Metal Efflux Proteins Function in Iron and Cobalt Homeostasis in Arabidopsis. *Plant Cell* **21**:3326–3338. DOI: https://doi.org/10.1105/tpc.109.069401, PMID: 19861554

Morrissey J, Sutak R, Paz-Yepes J, Tanaka A, Moustafa A, Veluchamy A, Thomas Y, Botebol H, Bouget F-Y, McQuaid JB, Tirichine L, Allen AE, Lesuisse E, Bowler C. 2015. A Novel Protein, Ubiquitous in Marine Phytoplankton, Concentrates Iron at the Cell Surface and Facilitates Uptake. *Curr Biol* **25**:364–371. DOI: https://doi.org/10.1016/j.cub.2014.12.004, PMID: 25557662

1833 **Myers SA**, Wright J, Peckner R, Kalish BT, Zhang F, Carr SA. 2018. Discovery of proteins associated
1834 with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. *Nat Methods* **15**:437–439.
1835 DOI: https://doi.org/10.1038/s41592-018-0007-1, PMID: 29735997
1836
1837 **Naslavsky N**, Caplan S. 2018. The enigmatic endosome – sorting the ins and outs of endocytic
1838 trafficking. *J Cell Sci* **131**:jcs216499. DOI: https://doi.org/10.1242/jcs.216499, PMID: 29980602
1839
1840 **Nguyen L-T**, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic
1841 algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**:268–274. DOI:
1842 https://doi.org/10.1093/molbev/msu300, PMID: 25371430
1843
1844 **Nitschke W**, McGlynn SE, Milner-White EJ, Russell MJ. 2013. On the antiquity of metalloenzymes and
1845 their substrates in bioenergetics. *Biochim Biophys Acta* **1827**:871–881. DOI:
1846 https://doi.org/10.1016/j.bbabio.2013.02.008, PMID: 23454059
1847
1848 **Nützmann H-W**, Huang A, Osbourn A. 2016. Plant metabolic clusters – from genetics to genomics. *New*
1849 *Phytol* **211**:771–789. DOI: https://doi.org/10.1111/nph.13981, PMID: 27112429
1850
1851 **Oborník M**, Füssy Z. 2018. Complex Endosymbioses I: From Primary to Complex Plastids, Multiple
1852 Independent Events. (In:) Plastids. Methods in Molecular Biology, (Editor:) Eric Maréchal. Humana Press,
1853 New York, NY. **1829**:17–35. DOI: https://doi.org/10.1007/978-1-4939-8654-5_2
1854
1855 **Och LM**, Shields-Zhou GA. 2012. The Neoproterozoic oxygenation event: Environmental perturbations
1856 and biogeochemical cycling. *Earth-Sci Rev* **110**:26–57. DOI:
1857 https://doi.org/10.1016/j.earscirev.2011.09.004
1858
1859 **Ohgami RS**, Campagna DR, Greer EL, Antiochos B, McDonald A, Chen J, Sharp JJ, Fujiwara Y, Barker
1860 JE, Fleming MD. 2005. Identification of a ferrireductase required for efficient transferrin-dependent iron
1861 uptake in erythroid cells. *Nat Genet* **37**:1264–1269. DOI: https://doi.org/10.1038/ng1658, PMID:
1862 16227996
1863
1864 **O'Halloran TV**. 1993. Transition metals in control of gene expression. *Science* **261**:715 –725. DOI:
1865 https://doi.org/10.1126/science.8342038, PMID: 8342038
1866
1867 **O'Halloran TV**, Culotta VC. 2000. Metallochaperones, an Intracellular Shuttle Service for Metal Ions. *J*
1868 *Biol Chem* **275**:25057–25060. DOI: https://doi.org/10.1074/jbc.R000006200, PMID: 10816601

Orr DJ, Pereira AM, da Fonseca Pereira P, Pereira-Lima ÍA, Zsögön A, Araújo WL. 2017. Engineering photosynthesis: progress and perspectives. *F1000Res* **6**:1891. DOI: https://doi.org/10.12688/f1000research.12181.1, PMID: 29263782

Otsuka Y, Satoh T, Nakayama N, Inaba R, Yamashita H, Satoh AK. 2019. Parcas is the predominant Rab11GEF for rhodopsin transport in *Drosophila* photoreceptors. *J Cell Sci*. **132**:1–13. DOI: https://doi.org/10.1242/jcs.231431, PMID: 31296556

Paek J, Kalocsay M, Staus DP, Wingler L, Pascolutti R, Paulo JA, Gygi SP, Kruse AC. 2017. Multidimensional Tracking of GPCR Signaling via Peroxidase-Catalyzed Proximity Labeling. *Cell* **169**:338–349. DOI: https://doi.org/10.1016/j.cell.2017.03.028, PMID: 28388415

Paulo JA, O'Connell JD, Everley RA, O'Brien J, Gygi MA, Gygi SP. 2016. Quantitative mass spectrometry-based multiplexing compares the abundance of 5000 *S. cerevisiae* proteins across 10 carbon sources. *J Proteomics* **148**:85–93. DOI: https://doi.org/10.1016/j.jprot.2016.07.005, PMID: 27432472

Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z. 2006. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* **7**:208. DOI: https://doi.org/10.1186/1471-2105-7-208, PMID: 16618368

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**:785–786. DOI: https://doi.org/10.1038/nmeth.1701, PMID: 21959131

Philpott CC, Jadhav S. 2019. The ins and outs of iron: Escorting iron through the mammalian cytosol. *Free Radic Biol Med* **133**:112–117. DOI: https://doi.org/10.1016/j.freeradbiomed.2018.10.411, PMID: 30321701

Poole LB. 2015. The basics of thiols and cysteines in redox biology and chemistry. *Free Radic Biol Med* **80**:148–157. DOI: https://doi.org/10.1016/j.freeradbiomed.2014.11.013, PMID: 25433365

Qiu W, Xu Z, Zhang M, Zhang D, Fan H, Li T, Wang Q, Liu P, Zhu Z, Du D, Tan M, Wen B, Liu Y. 2019. Determination of local chromatin interactions using a combined CRISPR and peroxidase APEX2 system. *Nucleic Acids Res* **47**:e52. DOI: https://doi.org/10.1093/nar/gkz134, PMID: 30805613

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/

1907 **Radhamony RN**, Theg SM. 2006. Evidence for an ER to Golgi to chloroplast protein transport pathway.
1908 *Trends Cell Biol* **16**:385–387. DOI: https://doi.org/10.1016/j.tcb.2006.06.003, PMID: 16815014
1909
1910 **Raven EL**. 2003. Understanding functional diversity and substrate specificity in haem peroxidases: what
1911 can we learn from ascorbate peroxidase? *Nat Prod Rep* **20**:367–381. DOI:
1912 https://doi.org/10.1039/b210426c, PMID: 12964833
1913
1914 **Rhee H-W**, Zou P, Udeshi ND, Martell JD, Mootha VK, Carr SA, Ting AY. 2013. Proteomic Mapping of
1915 Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science* **339**:1328–1331. DOI:
1916 https://doi.org/10.1126/science.1230593, PMID: 23371551
1917
1918 **Robinson NJ**, Winge DR. 2010. Copper Metallochaperones. *Annu Rev Biochem* **79**:537–562. DOI:
1919 https://doi.org/10.1146/annurev-biochem-030409-143539, PMID: 20205585
1920
1921 **Rochaix J-D**. 2011. Regulation of photosynthetic electron transport. *Biochim Biophys Acta* **1807**:375–
1922 383. DOI: https://doi.org/10.1016/j.bbabio.2010.11.010, PMID: 21118674
1923
1924 **Romero-Isart N**, Vasák M. 2002. Advances in the structure and chemistry of metallothioneins. *J Inorg*
1925 *Biochem* **88**:388–396. https://doi.org/10.1016/S0162-0134(01)00347-6, PMID: 11897355
1926
1927 **Rosenzweig AC** 2002. Metallochaperones: Bind and Deliver Minireview. *Chem Biol* **9**:673–677. DOI:
1928 https://doi.org/10.1016/S1074-5521(02)00156-4, PMID: 12079778
1929
1930 **Sacher M**, Di Bacco A, Lunin VV, Ye Z, Wagner J, Gill G, Cygler M. 2005. The crystal structure of CREG,
1931 a secreted glycoprotein involved in cellular growth and differentiation. *Proc Natl Acad Sci U S A*
1932 **102**:18326–18331. DOI: https://doi.org/10.1073/pnas.0505071102, PMID: 16344469
1933
1934 **Sapriel G**, Quinet M, Heijde M, Jourdren L, Tanty V, Luo G, Le Crom S, Lopez PJ. 2009. Genome-Wide
1935 Transcriptome Analyses of Silicon Metabolism in *Phaeodactylum tricornutum* Reveal the Multilevel
1936 Regulation of Silicic Acid Transporters. *PLoS ONE* **4**:e7458. DOI:
1937 https://doi.org/10.1371/journal.pone.0007458, PMID: 19829693
1938
1939 **Savelli B**, Li Q, Webber M, Jemmat AM, Robitaille A, Zamocky M, Mathé C, Dunand C. 2019.
1940 RedoxiBase: A database for ROS homeostasis regulated proteins. *Redox Biol* **26**:101247. DOI:
1941 https://doi.org/10.1016/j.redox.2019.101247, PMID: 31228650

1942 **Savitsky P**, Bray J, Cooper CDO, Marsden BD, Mahajan P, Burgess-Brown NA, Gileadi O. 2010. High-
1943 throughput production of human proteins for crystallization: the SGC experience. *J Struct Biol* **172**:3–13.
1944 DOI: https://doi.org/10.1016/j.jsb.2010.06.008, PMID: 20541610
1945
1946 **Scheiber IF**, Pilátová J, Malych R, Kotabova E, Krijt M, Vyoral D, Mach J, Léger T, Camadro J-M, Prášil
1947 O, Lesuisse E, Sutak R. 2019. Copper and iron metabolism in *Ostreococcus tauri* - the role of
1948 phytotransferrin, plastocyanin and a chloroplast copper-transporting ATPase. *Metallomics*. [Epub ahead
1949 of print]. DOI: https://doi.org/10.1039/c9mt00078j, PMID: 31380866
1950
1951 **Schober AF**, Río Bártulos C, Bischoff A, Lepetit B, Gruber A, Kroth PG. 2019. Organelle Studies and
1952 Proteome Analyses on Mitochondria and Plastids Fractions from the Diatom *Thalassiosira pseudonana*.
1953 *Plant Cell Physiol* **0**:1–18. DOI: https://doi.org/10.1093/pcp/pcz097, PMID: 31179502
1954
1955 **Schröder I**, Johnson E, de Vries S. 2003. Microbial ferric iron reductases. *FEMS Microbiol Rev* **27**:427–
1956 447. DOI: https://doi.org/10.1016/S0168-6445(03)00043-3, PMID: 12829278
1957
1958 **Sengupta R**, Poderycki MJ, Mattoo S. 2019. CryoAPEX – an electron tomography tool for subcellular
1959 localization of membrane proteins. *J Cell Sci* **132**. DOI: https://doi.org/10.1242/jcs.222315, PMID:
1960 30886003
1961
1962 **Sheppard VC**, Scheffel A, Poulsen N, Kröger N. 2012. Live Diatom Silica Immobilization of Multimeric
1963 and Redox-Active Enzymes. *Appl Environ Microbiol* **78**:211–218. DOI:
1964 https://doi.org/10.1128/AEM.06698-11, PMID: 22057862
1965
1966 **Sievers F**, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, Mcwilliam H, Remmert M, Söding J,
1967 Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence
1968 alignments using Clustal Omega. *Mol Syst Biol* **7**:539. DOI: https://doi.org/10.1038/msb.2011.75, PMID:
1969 21988835
1970
1971 **Smith SR**, Gillard JTF, Kustka AB, McCrow JP, Badger JH, Zheng H, New AM, Dupont CL, Obata T,
1972 Fernie AR, Allen AE. 2016. Transcriptional Orchestration of the Global Cellular Response of a Model
1973 Pennate Diatom to Diel Light Cycling under Iron Limitation. *PLoS Genet* **12**:e1006490. DOI:
1974 https://doi.org/10.1371/journal.pgen.1006490, PMID: 27973599
1975
1976 **Soll J**, Schleiff E. 2004. Protein import into chloroplasts. *Nat Rev Mol Cell Biol* **5**:198–208. DOI:
1977 https://doi.org/10.1038/nrm1333, PMID: 14991000

1978 **Tagliabue A**, Bowie AR, Boyd PW, Buck KN, Johnson KS, Saito MA. 2017. The integral role of iron in
1979 ocean biogeochemistry. *Nature* **543**:51–59. DOI: https://doi.org/10.1038/nature21058, PMID: 28252066
1980
1981 **Tanaka A**, De Martino A, Amato A, Montsant A, Mathieu B, Rostaing P, Tirichine L, Bowler C. 2015.
1982 Ultrastructure and Membrane Traffic During Cell Division in the Marine Pennate Diatom *Phaeodactylum*
1983 *tricornutum*. *Protist* **166**:506–521. DOI: https://doi.org/10.1016/j.protis.2015.07.005, PMID: 26386358
1984
1985 **Tesson B**, Genet MJ, Fernandez V, Degand S, Rouxhet PG, Martin-Jézéquel V. 2009. Surface Chemical
1986 Composition of Diatoms. *Chembiochem* **10**:2011–2024. DOI: https://doi.org/10.1002/cbic.200800811,
1987 PMID: 19623594
1988
1989 **Tsuji Y**, Nakajima K, Matsuda Y. 2017. Molecular aspects of the biophysical CO_2-concentrating
1990 mechanism and its regulation in marine diatoms. *J Exp Bot* **68**:3763–3772. DOI:
1991 https://doi.org/10.1093/jxb/erx173, PMID: 28633304
1992
1993 **Umbreit JN**, Conrad ME, Hainsworth LN, Simovich M. 2002. The ferrireductase paraferritin contains
1994 divalent metal transporter as well as mobilferrin. *Am J Physiol Gastrointest Liver Physiol* **282**:G534–9.
1995 DOI: https://doi.org/10.1152/ajpgi.00199.2001, PMID: 11842004
1996
1997 **Umbreit JN**, Conrad ME, Moore EG, Desai MP, Turrens J. 1996. Paraferritin: A Protein Complex with
1998 Ferrireductase Activity is Associated with Iron Absorption in Rats. *Biochemistry* **35**:6460–6469. DOI:
1999 https://doi.org/10.1021/bi951927s, PMID: 8639593
2000
2001 **Uversky VN**. 2015. The multifaceted roles of intrinsic disorder in protein complexes. *FEBS Lett*
2002 **589**:2498–2506. DOI: https://doi.org/10.1016/j.febslet.2015.06.004, PMID: 26073257
2003
2004 **Uversky VN**. 2016. Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically
2005 Disordered Proteins. *J Biol Chem* **291**:6681–6688. DOI: https://doi.org/10.1074/jbc.R115.685859, PMID:
2006 26851286
2007
2008 **Valko M**, Morris H, Cronin MTD. 2005. Metals, Toxicity and Oxidative Stress. *Curr Med Chem* **12**:1161–
2009 1208. DOI: https://doi.org/10.2174/0929867053764635, PMID: 15892631
2010
2011 **Vartanian M**, Desclés J, Quinet M, Douady S, Lopez PJ. 2009. Plasticity and robustness of pattern
2012 formation in the model diatom *Phaeodactylum tricornutum*. *New Phytol* **182**:429–442. DOI:
2013 https://doi.org/10.1111/j.1469-8137.2009.02769.x, PMID: 19210721

Villarejo A, Burén S, Larsson S, Déjardin A, Monné M, Rudhe C, Karlsson J, Jansson S, Lerouge P, Rolland N, von Heijne G, Grebe M, Bako L, Samuelsson G. 2005. Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. *Nat Cell Biol* **7**:1224–1231. DOI: https://doi.org/10.1038/ncb1330, PMID: 16284624

Waller RF, Cleves PA, Rubio-Brotons M, Woods A, Bender SJ, Edgcomb V, Gann ER, Jones AC, Teytelman L, von Dassow P, Wilhelm SW, Collier JL. 2018. Strength in numbers: Collaborative science for new experimental model systems. *PLoS Biol* **16**:e2006333. DOI: https://doi.org/10.1371/journal.pbio.2006333, PMID: 29965960

Wang J, Pantopoulos K. 2011. Regulation of cellular iron metabolism. *Biochem J* **434**:365–381. DOI: https://doi.org/10.1042/BJ20101825, PMID: 21348856

Ward DM, Kaplan J. 2012. Ferroportin-mediated iron transport: expression and regulation. *Biochim Biophys Acta* **1823**:1426–1433. DOI: https://doi.org/10.1016/j.bbamcr.2012.03.004, PMID: 22440327

Zhang Y, Song G, Lal NK, Nagalakshmi U, Li Y, Zheng W, Huang P, Branon TC, Ting AY, Walley JW, Dinesh-Kumar SP. 2019. TurboID-based proximity labeling reveals that UBR7 is a regulator of N NLR immune receptor-mediated immunity. *Nat Commun* **10**:3252. DOI: https://doi.org/10.1038/s41467-019-11202-z, PMID: 31324801

Supplementary Figures

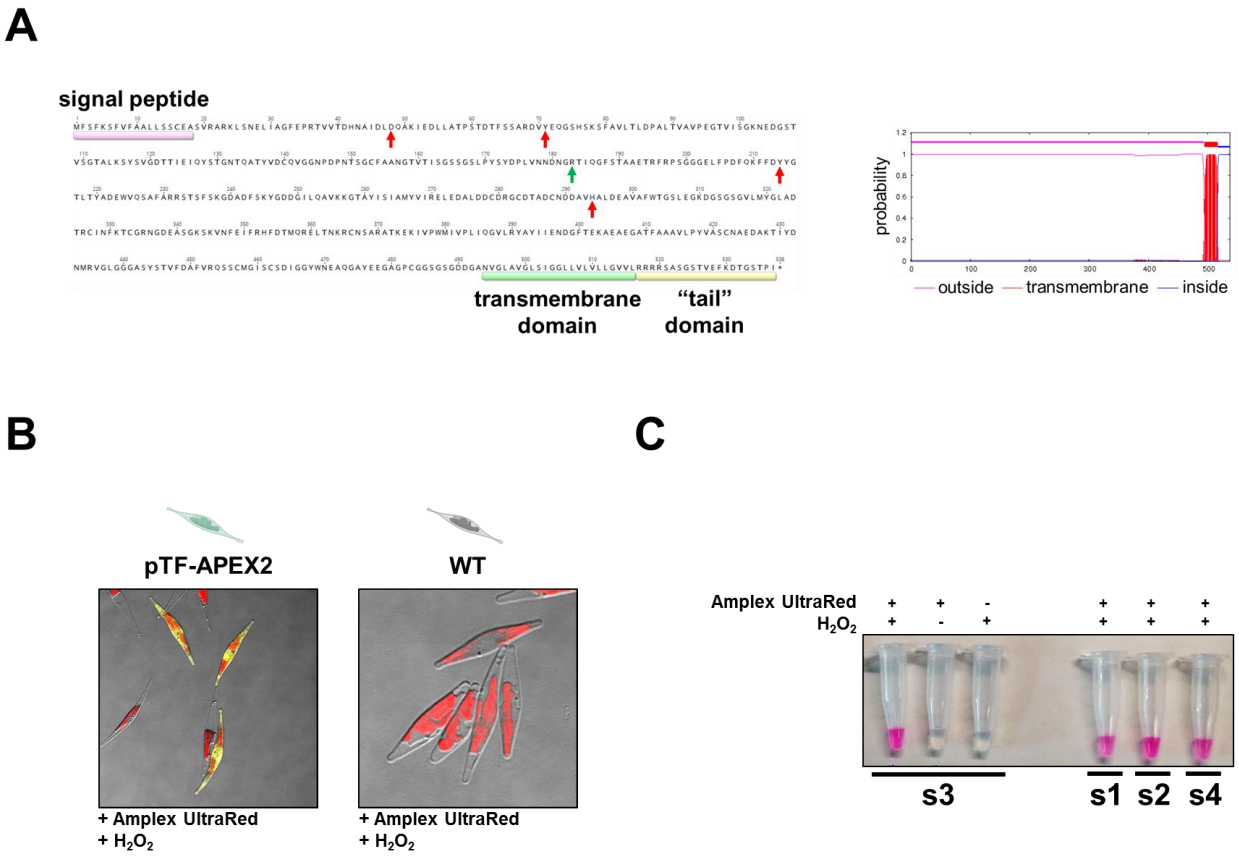
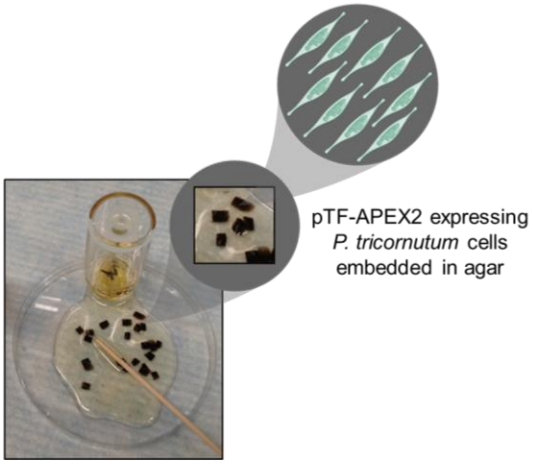
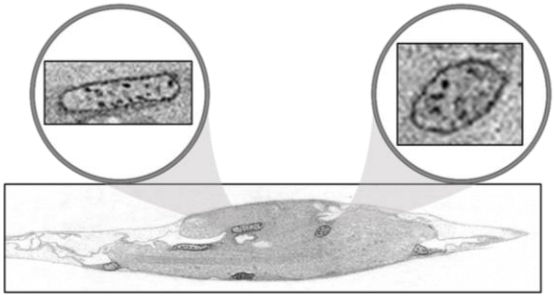


Figure 2—figure supplement 1. pTF features, resorufin imaging in pTF-APEX2 expressing cells, and resorufin signal co-dependence on Amplex UltraRed and hydrogen peroxide. (A) pTF is a 537 amino acid residues (aa) long membrane-associated iron receptor protein consisting of a signal peptide (aa 1–18; predicted with SignalP 4.1), an iron-binding domain (aa 19–493), a transmembrane domain (aa 494–516; predicted with TMHMM Server v. 2.0), and a "tail" domain (aa 517–537). Red and green arrows are conserved putative iron- (Asp48, Tyr71, Tyr214, His294) and carbonate-coordinating (Arg183) residues, respectively. (B) Direct visualization of resorufin (yellow) in pTF-APEX2 expressing cells. Red is chloroplast autofluorescence. (C) Supernatants after incubating cells from 4 pTF-APEX2 expressing strains (s1–s4) in a reaction buffer (50 μM Amplex UltraRed, 2 mM H_2O_2, PBS (pH 7.4)) for 6 min at room temperature. Clear resorufin signal was observed only when both APEX2 substrates were present.

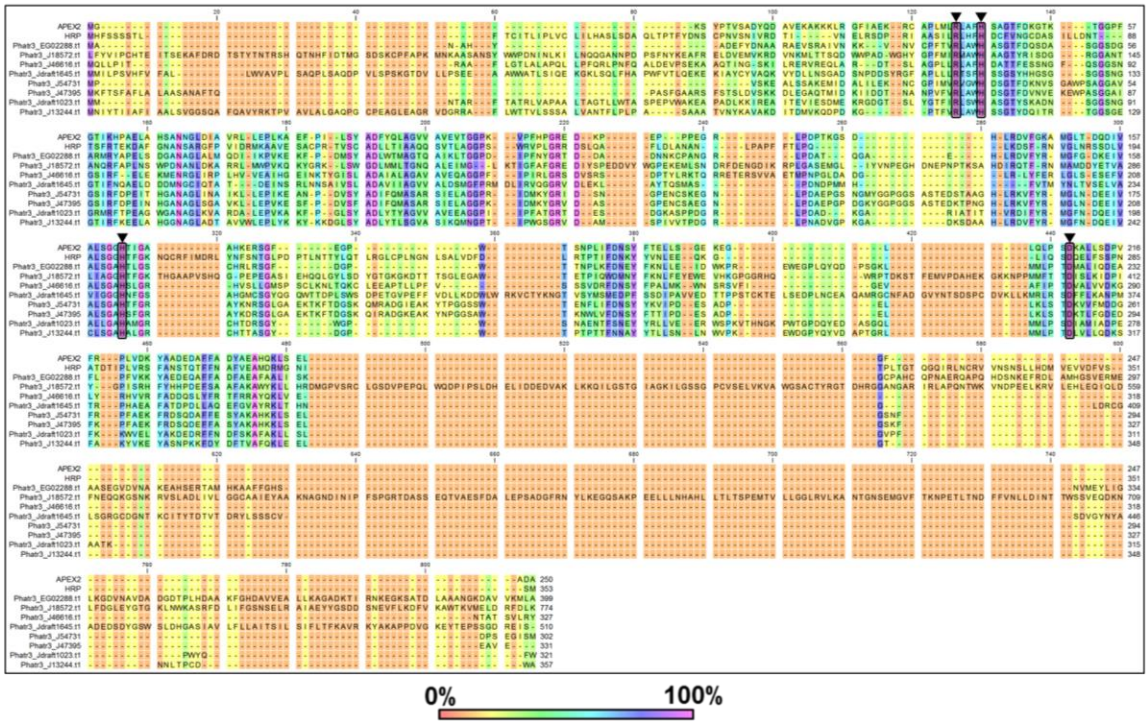
A



B



C



D

Gene ID	Peroxidase class[1]	Position 134 relative to APEX2[2]	Predicted localization[3]	Nr. of TM domains[4]
Phatr3_EG02288	I	Ala	cytosol	/
Phatr3_J18572	I	**Pro**	cytosol	/
Phatr3_J46616	I	**Pro**	secretory pathway	/
Phatr3_Jdraft1645	I	**Pro**	secretory pathway	1
Phatr3_J54731	I	Ala	cytosol	/
Phatr3_J47395	I	Ala	secretory pathway	/
Phatr3_Jdraft1023	I	Ala	**mitochondria**	/
Phatr3_J13244	I	Ala	secretory pathway	/

E

2049

Figure 3—figure supplement 1. Analysis of transcriptionally active *Phaeodactylum tricornutum* peroxidases reveals putative APEX2-like and mitochondrial enzymes. (A) Excerpt from the TEM labeling protocol: agar blocks with embedded DAB-treated pTF-APEX2 expressing diatom cells. *P. tricornutum* cartoon and "zoom in" objects created with BioRender.com. (B) Mitochondrial signal as shown here was observed in both WT and pTF-APEX2 expressing cells subjected to DAB and hydrogen peroxide. "Zoom in" objects created with BioRender.com. (C) Alignment of putative APEX2-like *P. tricornutum* peroxidases. Amino acid residues crucial for APX and likely crucial for APEX2 activity (Arg38, His42, His163 and Asp208) are conserved in eight *P. tricornutum* peroxidases (depicted with black triangles and shaded). (D) Three of the identified APEX2-like peroxidases contain proline instead of alanine at the critical position 134, an amino acid change rendering APEX2 much more active than its APEX predecessor. One of them may be mitochondrial. [1]RedoxiBase database. [2]APEX2 and horseradish peroxidase (HRP) have proline at this position. [3]TargetP 1.1. [4]TMHMM Server v. 2.0. (E) Alignment of putative substrate binding loops in the identified APEX2-like *P. tricornutum* peroxidases. Black and blue arrows point to enzymes with proline at position 134 (shaded gray) and predicted mitochondrial localization, respectively.

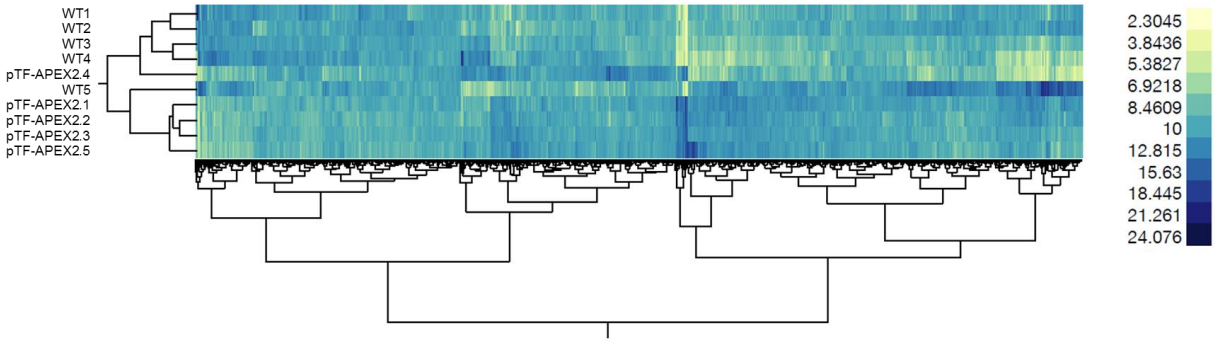


Figure 4—figure supplement 1. WT and pTF-APEX2 proteomic replicates form distinct clusters. Hierarchical clustering (Ward's method as implemented in the JMP software package) of scaled quantitative proteomics data reveal WT- and pTF-APEX2-specific clusters. Samples "WT5" and "pTF-APEX2.4" differ slightly from other corresponding samples.

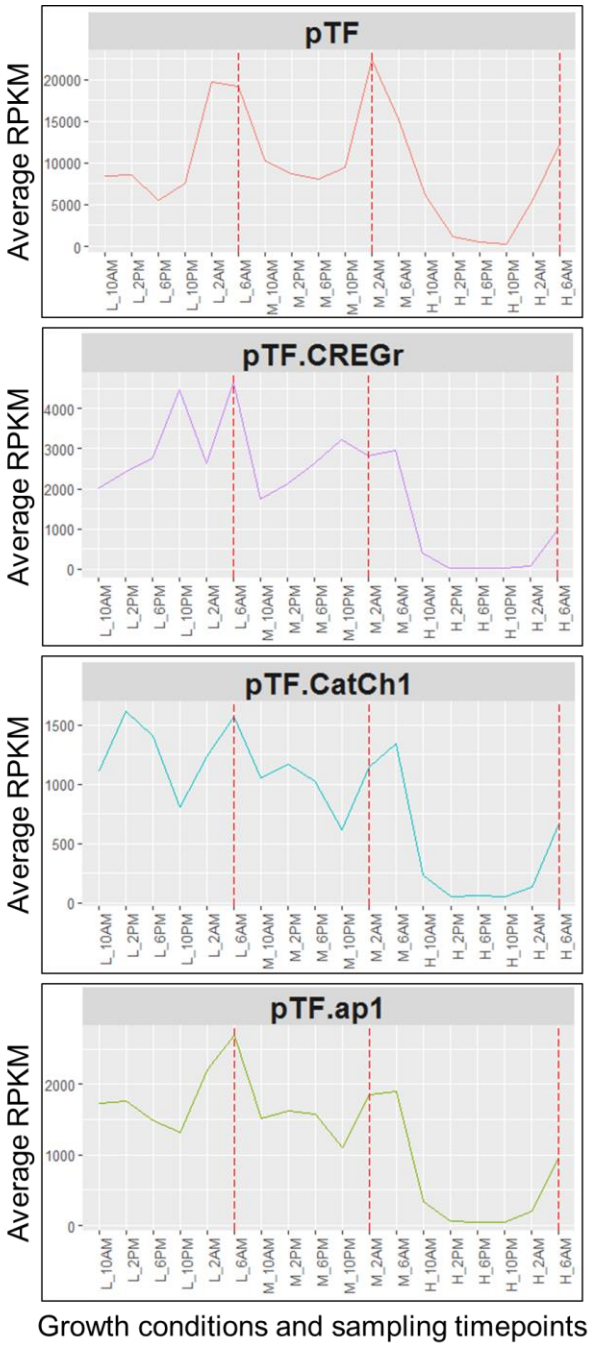


Figure 5—figure supplement 1. Transcriptional profiles of *pTF* and the three genes from the identified iron-sensitive gene cluster on chromosome 20 across different growth conditions (L (low): 20 pM Fe', M (medium): 40 pM Fe', H (high/replete): 400 pM Fe') and sampling timepoints (10AM, 2PM, 6PM, 10PM, 2AM, 6AM). Fe': sum of all Fe species not complexed to EDTA. Red dashed lines correspond to genes' transcriptional peaks (late night). See ***Supplementary file 4*** for detailed description of the source of raw data.

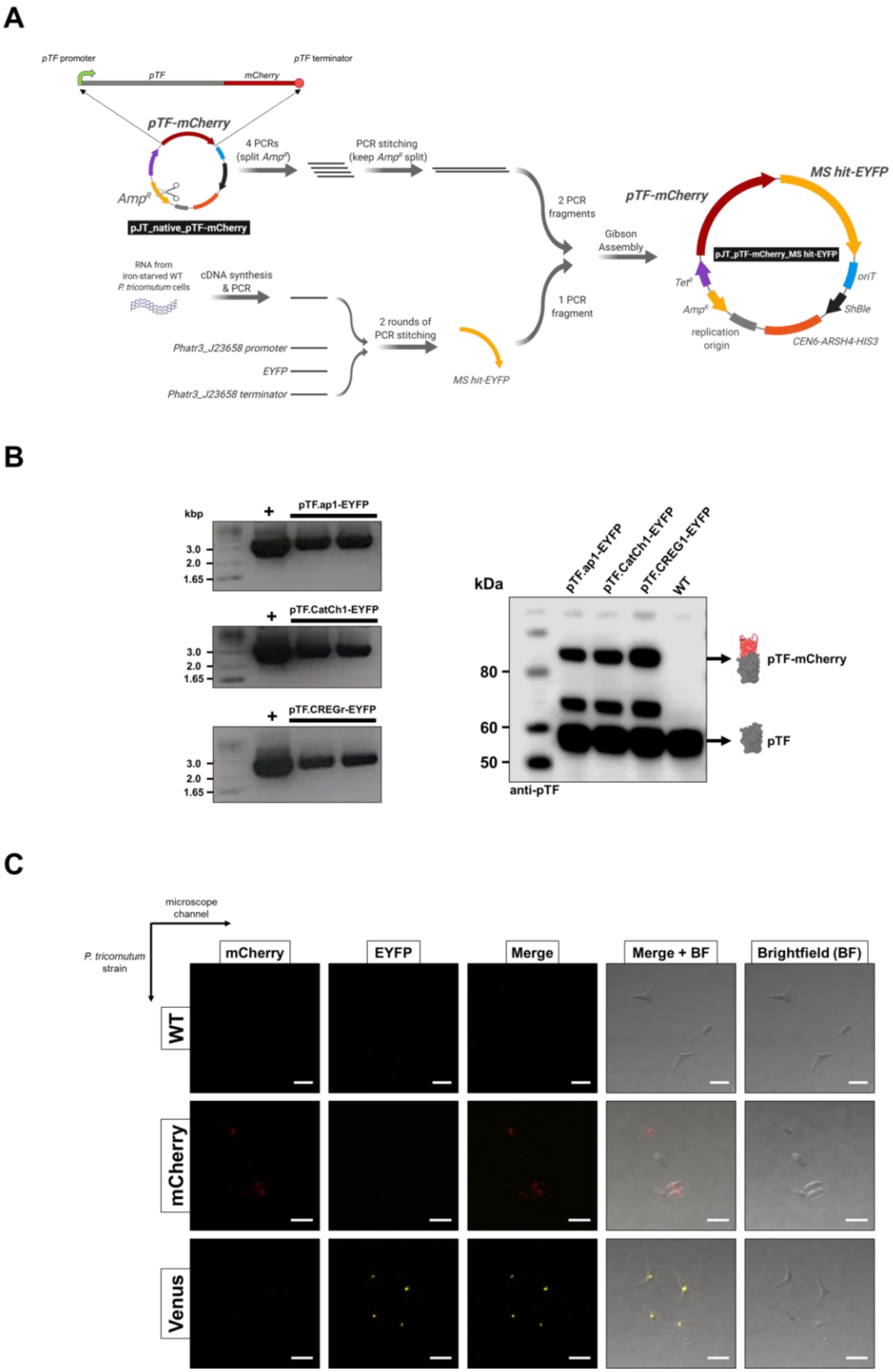


Figure 5—figure supplement 2. Co-expression episome assembly overview, assessment of co-expression cell lines, and confocal microscopy bleed-through controls. (A) Summary of our co-expression episome assembly strategy. Created with BioRender.com. (B) Colony PCRs and Western blots suggesting full length proteins are present in co-expression cell lines. Left: Primers binding to flavodoxin promoter and terminator flanking MS hit-EYFP coding regions were used for PCR which resulted in the expected ~3 kbp amplicons. Positive controls: purified episomes. Negative controls: WT *P. tricornutum* and water which yielded no amplicons (not shown). Right: Anti-pTF Western blot revealed the expected ~83.8 kDa pTF-mCherry bands alongside native pTF. Protein schemes created with BioRender.com. (C) Lack of significant cross-channel bleed-through supports colocalization imaging data and conclusions. Imaging conditions with minimal bleed-through were determined with WT cells (top row), cells expressing only mCherry (middle row), and cells expressing a Venus-tagged protein (bottom row). Scale bar is 10 µm.

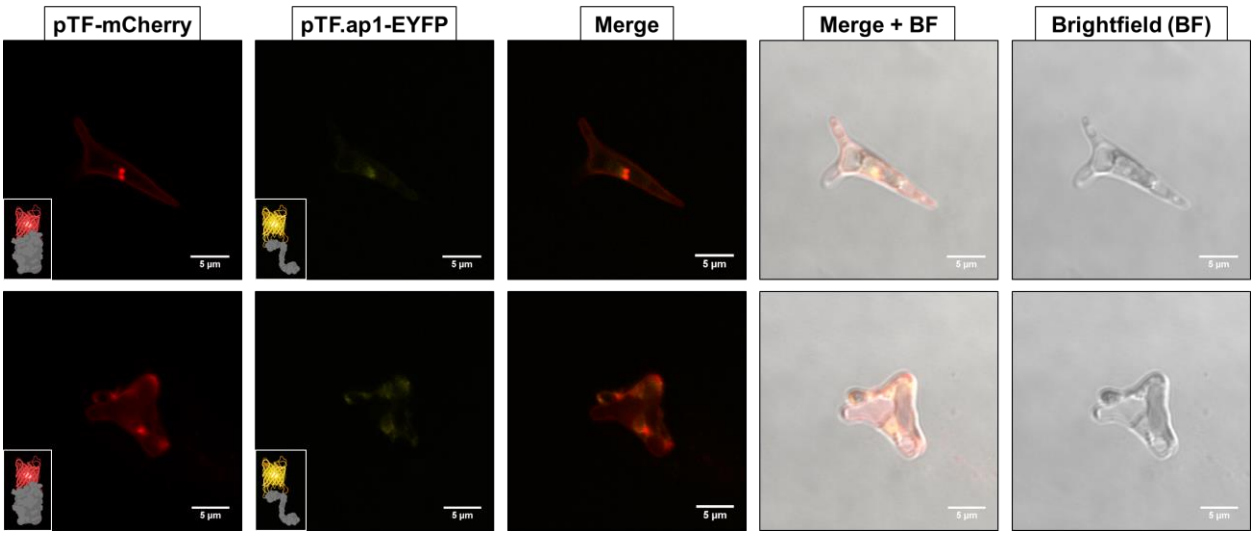


Figure 5—figure supplement 3. pTF.ap1-EYFP was co-expressed with pTF-mCherry and colocalization of the two proteins—although not as precise—was observed. Scale bar is 5 µm. Protein fusion schemes created with BioRender.com.

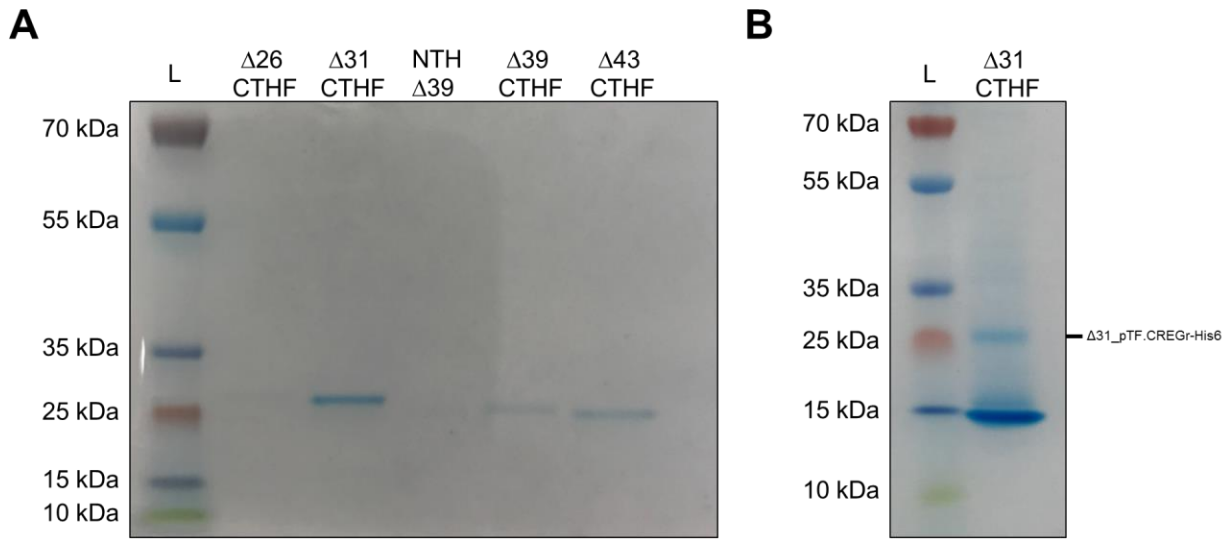


Figure 7—figure supplement 1. Summary of pTF.CREGr expression and purification screen. NuPage 4-12% Bis-Tris gels loaded with (A) PageRuler Plus Prestained Protein Ladder and final elutions of various purified pTF.CREGr truncations from small scale protein expression experiments, and (B) PageRuler Plus Prestained Protein Ladder and a final elution of purified Δ31_pTF.CREGr-His6 from a large scale protein expression experiment. CTHF: C-terminal His6 and FLAG tag. NTH: N-terminal His6 tag. L: protein ladder.

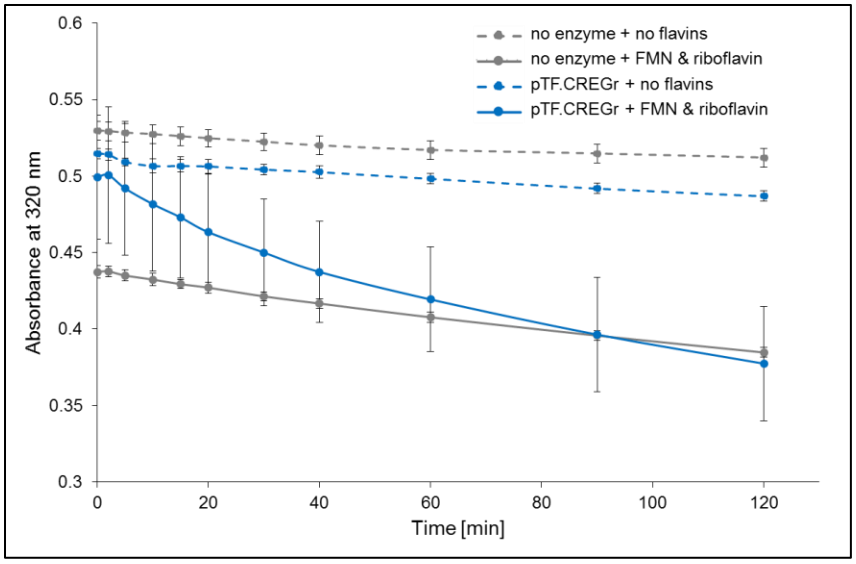
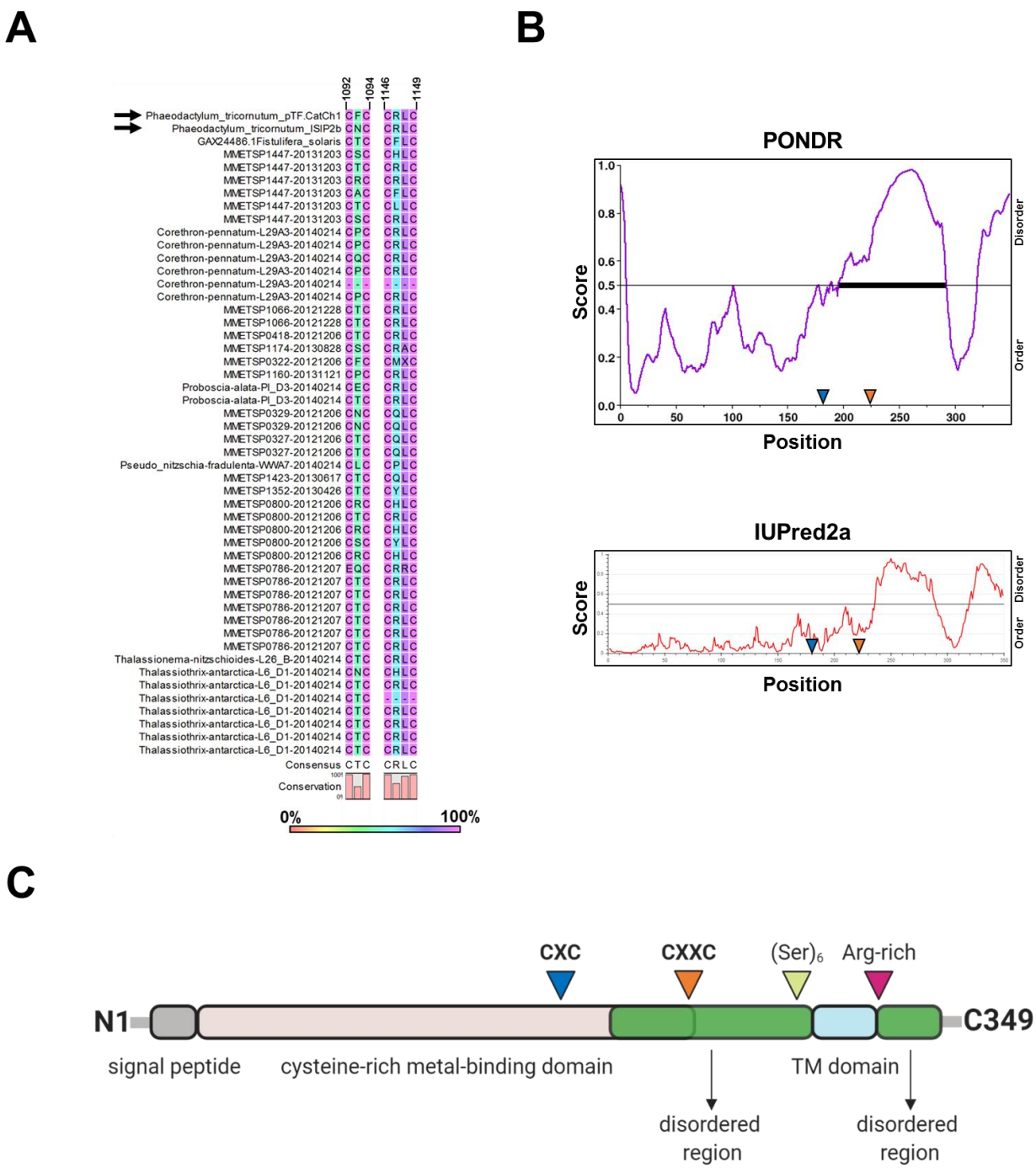


Figure 7—figure supplement 2. Comparison of NADPH oxidation (measured by decrease in absorbance at 320 nm) between pTF.CREGr supplemented with flavins (flavin mononucleotide (FMN) and riboflavin), a flavin-only, no enzyme treatment, pTF.CREGr with no additional flavins, and an NADPH-only treatment with no enzyme or flavins added. Assays were started by adding enzyme or water blank and were carried out in triplicate for each treatment. Standard deviation is shown.



2111

Figure 8—figure supplement 1. Conservation of putative metal-binding sites in pTF.CatCh1 in marine microeukaryotes and proposed pTF.CatCh1 domain organization. (A) CXXC and adjacent CXC motifs are present in all, but two, identified pTF.CatCh1 homologs. 6 additional cysteines are 100% conserved (data not shown). pTF.CatCh1 and ISIP2b are marked with arrows. (B) pTF.CatCh1 contains a disordered C-terminal domain which may be involved in protein-protein interactions. CXXC is predicted to be in this region by PONDR VSL2 (top) possibly suggesting metal binding and disorder-order transitions are dependent on local redox environment. No cysteines are predicted to be in the disordered region by IUPred2a (bottom). The locations of conserved CXC and CXXC motifs are depicted with blue and orange triangles, respectively. (C) Features and proposed domain architecture in pTF.CatCh1 assuming disorder prediction with PONDR. Signal peptide (aa 1–26; grey) is followed by a cysteine-rich domain (20 cysteines between amino acid residues 27 and 226; pale pink) possibly responsible for binding iron, and a disordered region (green) that may be involved in protein-protein interactions. The TM domain (aa 296–315; light blue) splitting the disordered region in two is flanked by a polyserine and an arginine-rich stretch. N- and C-terminus with numbered first and last residue are shown. Created with BioRender.com.

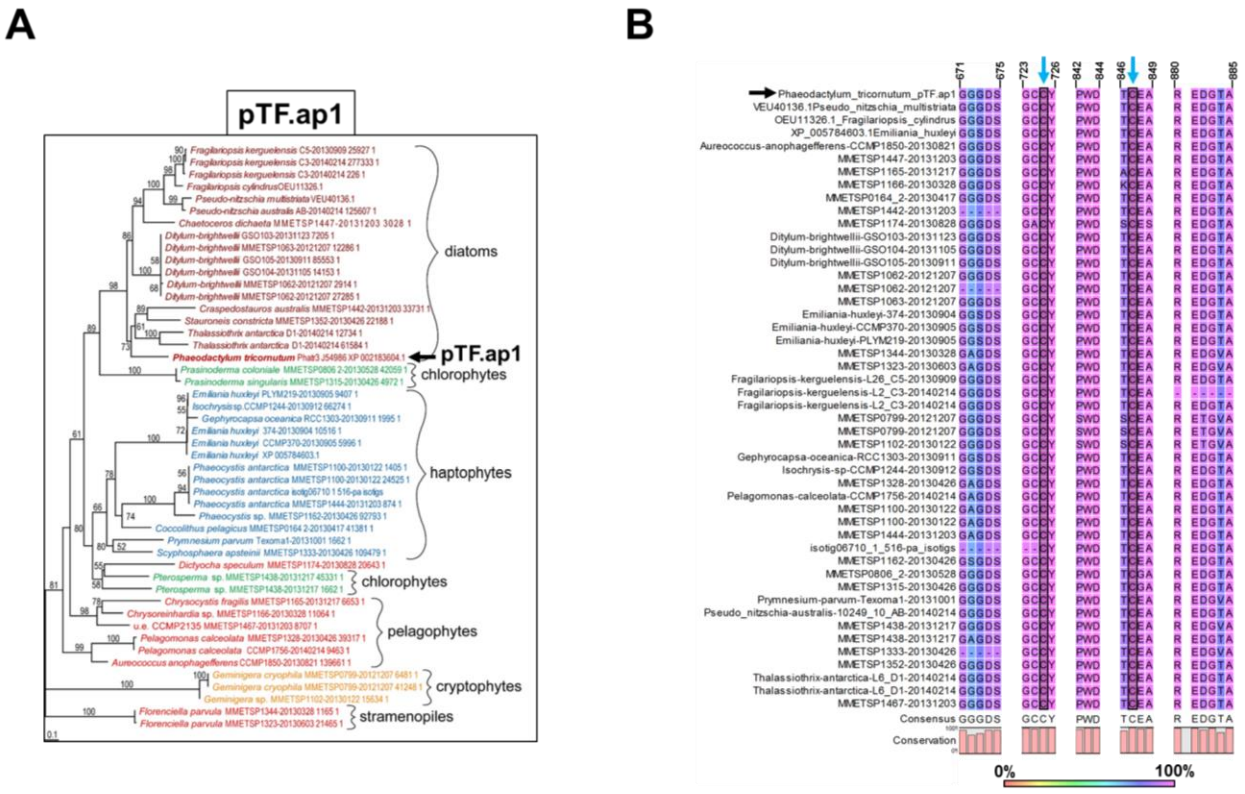


Figure 8—figure supplement 2. (A) pTF.ap1 homologs are present in diatoms and other marine microeukaryotes. Homolog search was performed with NCBI and marine microbial eukaryote (MMETSP) databases. Some notable species include: common polar diatoms *Fragilariopsis kergulensis* and *Fragilariopsis cylindrus*; toxin-producing microeukaryotes (diatoms from genus *Pseudo-nitzschia*, haptophyte *Prymnesium parvum*, and pelagophyte *Aureococcus anophagefferens*), cosmopolitan diatom *Ditylum brightwelli*, haptophyte *Emiliana huxleyi*, and an ultraplanktonic species *Pelagomonas calceolate;* colonial species *Prasinoderma coloniale* and *Phaeocystis antarctica*. Scale bar: 0.1 substitutions per position. (B) pTF.ap1 (arrow) contains highly conserved motifs. There are additional three 100% conserved cysteines across pTF.ap1 homologs in addition to the two presented here (blue arrows and grey shade) indicating this protein may be involved in metal binding similar to pTF.CatCh1 (data not shown).

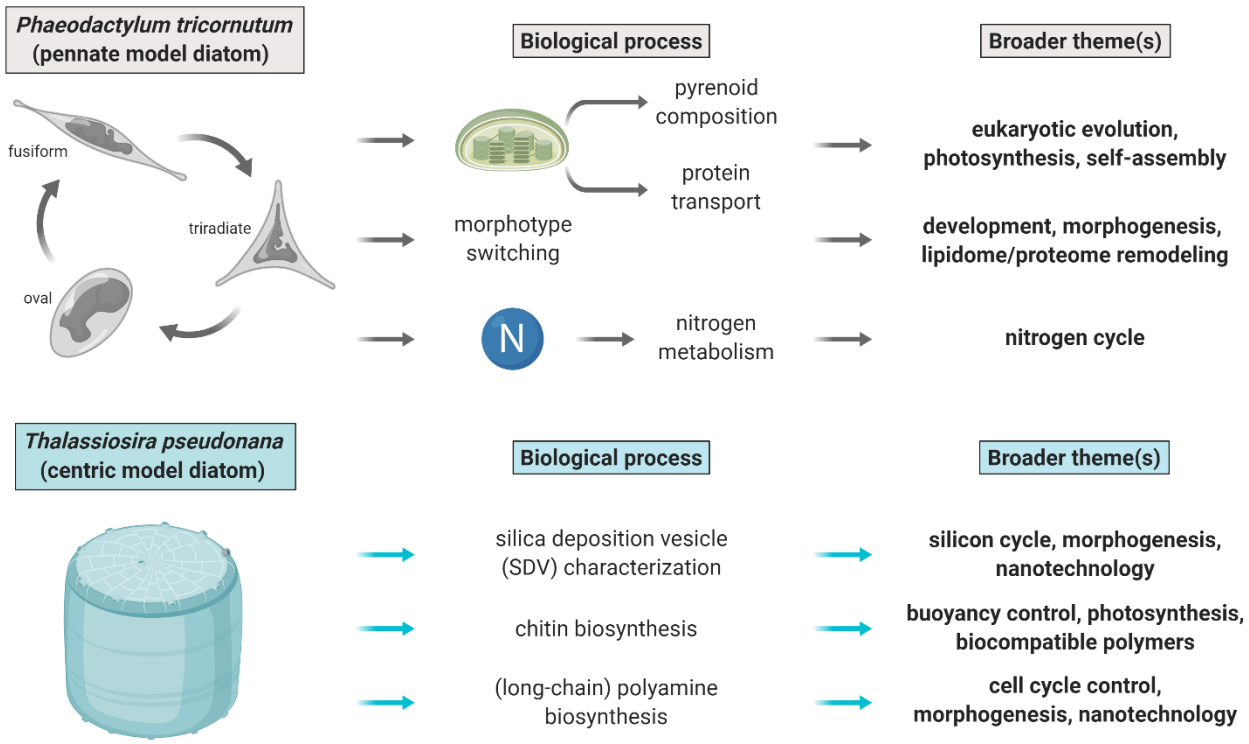


Figure 9—figure supplement 1. Vision board summarizing biological problems in established model marine diatoms that could benefit from proximity proteomics approaches. Two diatom compartments of immediate interest are pyrenoid—proteinaceous RuBisCO-containing suborganelle—and silica deposition vesicle (SDV)—site of diatom cell wall biomineralization. We intentionally constrained ourselves to these two already genetically tractable diatoms appropriate for the immediate study of pyrenoid and SDV composition, but we expect APEX2 and related proximity proteomics approaches will be adopted in other emerging diatom models as well as other marine microeukaryotes. Created with BioRender.com.