Proximity proteomics in a marine diatom reveals a putative cell surface-to-chloroplast iron trafficking pathway

Jernej Turnšek[1,2,3,4,5,6,#], John K. Brunson[6,7], Maria del Pilar Martinez Viedma[8,§], Thomas J. Deerinck[9], Aleš Horák[10,11], Miroslav Oborník[10,11], Vincent A. Bielinski[12], 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]Informatics, J. Craig Venter Institute, La Jolla, CA 92037, USA; [9]National Center for Microscopy and Imaging Research, University of California San Diego, La Jolla, CA 92093, USA; [10]Biology Centre CAS, Institute of Parasitology, Branišovská 31, 370 05 České Budějovice, Czech Republic; [11]University of South Bohemia, Faculty of Science, Branišovská 31, 370 05 České Budějovice, Czech Republic; [12]Synthetic Biology and Bioenergy, J. Craig Venter Institute, La Jolla, CA 92037, USA; #Present affiliations: Department of Plant & Microbial Biology, University of California, Berkeley, Berkeley, CA 94720, USA; Howard Hughes Medical Institue, 4000 Jones Bridge Rd, Chevy Chase, MD 20815; §Present affiliation: PHARMA MAR, S.A., Avda. de los Reyes, 1, Pol. Ind. La Mina, 28770 Colmenar Viejo, Madrid, Spain

***For correspondence:** aallen@ucsd.edu & aallen@jcvi.org (A.E.A.)

Abstract

Iron is a biochemically critical metal cofactor in enzymes involved in photosynthesis, cellular 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 endocytosed, 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 marine 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

ocean, iron cycle, diatoms, *Phaeodactylum tricornutum*, phytotransferrin, endocytosis, metal trafficking, chloroplast, proximity-dependent proteomic mapping, APEX2

Glossary

AA/AAs	amino acid residue/amino acid residues
APEX	engineered ascorbate peroxidase (APX)
CREG	cellular repressor of E1A-stimulated genes
DAB	3,3'-diaminobenzidine
EYFP	enhanced yellow fluorescent protein
Fe'	dissolved labile iron (all unchelated iron species)
FMN	flavin mononucleotide
HNLC	high-nutrient, low-chlorophyll
ISIP	iron starvation induced protein
MS	mass spectrometry
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**) and is fundamental to almost all extant metabolisms (**Aguirre et al., 2012**) serving as a cofactor in enzymes involved in DNA replication, photosynthesis, cellular respiration, nitrate assimilation, nitrogen fixation, and reactive oxygen species defense (**Crichton, 2016**). Early anoxic oceans were rich in readily bioavailable ferrous, Fe^{2+}, 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 the ferrous iron oxidized into insoluble ferric, Fe^{3+}, 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 iron uptake mechanisms.

Iron availability limits primary productivity in high nutrient, low chlorophyll (HNLC) regions which cover ~25% of the modern ocean habitat as demonstrated by numerous large-scale Fe fertilization experiments and natural Fe upwelling events invariably resulting in diatom-dominated phytoplankton blooms (**Ardyna et al., 2019**; **de Baar et al., 2005**; **Martin et al., 1994**). Although dissolved Fe in marine environments is primarily found complexed to organic ligands (**Hutchins and Boyd, 2016**; **Tagliabue et al., 2018**) such as bacterially produced siderophores and hemes (**Boiteau et al., 2016**; **Hogle et al., 2014**), low (pM) amounts of unchelated labile Fe, Fe', are crucial for eukaryotic phytoplankton, particularly diatoms (**Morel et al., 2008**), widespread single-celled microalgae critical to oceanic primary production (**Benoiston et al., 2017**). Diatoms have convergently evolved phytotransferrin (pTF), which serves as the basis of a high-affinity ferric, Fe^{3+}, 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 transcriptomic 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**). They are phylogenetically related to Fe-assimilation (FEA) domain-containing proteins found in freshwater and marine single-celled algae, and represent functional analogues of iron delivery proteins transferrins (**Behnke and LaRoche, 2020**; **Cheng et al., 2004**; **McQuaid et al., 2018**; **Scheiber et al., 2019**). *pTF* (Ensembl ID: Phatr3_J54465, UniProt ID: B7FYL2) from the model marine diatom *Phaeodactylum tricornutum* (**Bowler et al., 2008**) was first identified as an iron

deficiency-sensitive gene and named *ISIP2a* (iron starvation induced protein 2a) (***Allen et al., 2008***), although its expression is 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 labile ferric iron is coordinated synergistically with carbonate anion (CO_3^{2-}), a common feature of all transferrins (***Cheng et al., 2004***; ***McQuaid et al., 2018***; ***Zak et al.***, **2002**). 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***).

While accessing and binding dilute dissolved labile ferric iron on the cell surface is important, its subsequent internalization and delivery to target sites within complex cellular milieu are also critical. Fe^{3+} is highly insoluble and conversion between Fe^{3+} and Fe^{2+} 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 ferric 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^{3+} is then reduced by six-transmembrane epithelial antigen of prostate 3 (STEAP3), exported to the cytoplasm through divalent metal transporter 1 (DMT1; also NRAMP2) or Zrt- and Irt-like protein (ZIP) 14, and offloaded to iron chaperones for distribution to cellular iron sinks (***Bogdan et al., 2016***; ***Eckenroth et al., 2011***; ***Ohgami et al., 2005***; ***Philpott and Jadhav, 2019***; ***Wang and Pantopoulos, 2011***). In contrast, proteins that mediate intracellular allocation of internalized Fe^{3+} and conduct specific biochemical transformations downstream of diatom pTF are unknown. Elucidation of these pathways will advance our understanding of diatom adaptation to survival in iron-deficient ocean waters and provide insight into how climate change will impact diatom physiology and fitness.

To investigate the proximal proteomic neighborhood of *P. tricornutum* pTF, we employed APEX2, an engineered soybean ascorbate peroxidase that functions as a dual probe for electron microscopy and proximity proteomics (***Hung et al., 2016***; ***Lam et al., 2015***; ***Martell et al., 2017***). When fused with a protein of interest (bait protein), 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 acid residues—primarily tyrosine—on the surface of nearby endogenous proteins. Tagged proteins can then be isolated by purification with streptavidin beads and analyzed using mass spectrometry (MS). While the "biotinylation radius" in APEX2 experiments is estimated to be in the 10–20 nm range, it should rather be seen as a "probability gradient" where the likelihood of tagging decreases with distance away from the APEX2-tagged protein of interest (***Gingras et al., 2019***; ***Hung et al., 2016***; ***Lam et al., 2015***). Proteomic hits can therefore include vicinal proteins which interact with the APEX2-tagged bait protein indirectly in addition to direct interactors of varying strength (***Lundberg and Borner, 2019***). Tha ability to capture weakly and transiently interacting proteins in a biological process of interest is a major advantage APEX2 offers over traditional pull-down approaches (***Gingras et al., 2019***). Swaping biotin-phenol for diaminobendizine (DAB)—another APEX2 substrate—enables one to determine fusion protein localization via high-resolution electron microscopy making APEX2 a powerful bifunctional probe (***Martell et al., 2017***; ***Martell et al., 2012***).

APEX2 and related enzyme-based chemical biology tools have over the past decade been applied to investigate cellular compartments and processes in diverse model systems including mammalian cell culture, yeast, protists, and plant protoplasts (***Gingras et al., 2019***; ***Trinkle, 2019***). APEX2 was recently used to map proteomes associated with post-Golgi vesicle trafficking and endocytosis (***Del Olmo et al., 2019***; ***Otsuka et al., 2019***) as well as receptor-mediated signaling (***Paek et al., 2017***), successes that are all directly pertinent to the goals of our study.

Here, we provide further evidence for *P. tricornutum* phytotransferrin pTF endocytosis, demonstrate correct subcellular localization and activity of its APEX2 fusion, and identify nearly 40 likely proximal proteins through a proximity-dependent proteomic mapping experiment conducted in quintuplicate. To the best of our knowledge, this study represents the first application of APEX2 in a marine microbial model system (***Figure 1***). We then provide initial characterization, using dual fluorophore protein tagging fluorescence microscopy, phylogenetic, domain, and biochemical analyses, for three proteins expressed from a chromosomal gene cluster (pTF.CREG1, pTF.CatCh1, pTF.ap1) believed to act in association with endocytosed pTF downstream of ferric iron binding at the cell surface. Bacterially expressed recombinant pTF.CREG1 was found to be catalytically dead indicating it may have a regulatory, rather than enzymatic, role in pTF endocytosis, similar to functions proposed for human CREG1. Meanwhile, bioinformatic interrogation of pTF.CatCh1 and its localization pattern suggest it might be a

166 chloroplast-associated iron-binding protein. Finally, we introduce a model for an intracellular ferric
167 iron allocation pathway in *P. tricornutum* based on our proximity proteomics hits.

Results

P. tricornutum phytotransferrin (pTF) is localized to intracellular vesicles.

To confirm pTF association with intracellular vesicles as suggested by ***McQuaid et al., 2018***, a pTF-mCherry encoding episome was conjugated into *ΔpTF P. tricornutum* cells (***Diner et al., 2016***; ***Karas et al., 2015***; ***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, colocalized mCherry and MDY-64 signals were observed on the cell surface and within intracellular vesicles (***Figure 2A***).

APEX2 fused with *P. tricornutum* phytotransferrin (pTF) is enzymatically active *in vivo*.

To generate pTF-APEX2 expressing diatom cell lines, an episome encoding pTF with C-terminal APEX2 was assembled and conjugated into WT *P. tricornutum* cells (***Figure 2B***). Considering the predicted pTF domains (***Figure 2—figure supplement 1A***), APEX2 was likely facing the cytosol 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 four out of five tested transconjugant cell lines. The protein was present in the insoluble cell lysate fraction further indicative of its membrane localization (***Figure 2C***). Amplex UltraRed, a highly sensitive APEX2 substrate (***Hung et al., 2016***), was used to assay live pTF-APEX2 expressing cells. A resorufin (reaction product of Amplex UltraRed and APEX2) 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***; ***Figure 2—source data 1***), respectively, indicating active APEX2 with incorporated heme. Resorufin was also directly visualized by confocal microscopy and a strong cytosolic signal not tightly localized to the expected site of origin, similar to previous reports (***Martell et al., 2012***), was observed (***Figure 2—figure supplement 1B***), perhaps indirectly supporting our predicted APEX2 orientation. 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.

198 **pTF-APEX2 is localized to the cell membrane and intracellular vesicles.**
199
200 To confirm pTF-APEX2 is localized to the cell surface and intracellular vesicles similar to the pTF-
201 mCherry fusion protein, pTF-APEX2 expressing cells (cell line s2, ***Figure 2C–D***) were treated
202 with 25 mM 3,3'-diaminobenzidine (DAB) in the presence of 3 mM H_2O_2 for 15 min on ice (***Figure***
203 ***3A***). This reaction leads to DAB polymerization and local precipitation around APEX2 that can be
204 stained with osmium tetroxide and visualized with an electron microscope. Cells were embedded
205 in a 3% agar matrix to prevent losses during numerous washing steps (***Figure 3—figure***
206 ***supplement 1A***). Tightly localized signal was observed on the cell membrane and in intracellular
207 vesicles indicating that pTF-APEX2 fusion trafficked to the correct subcellular sites (***Figure 3B***).
208 Additionally, we observed mitochondrial signal in WT and pTF-APEX2 cell lines subjected to the
209 DAB reaction (***Figure 3B***; ***Figure 3—figure supplement 1B***). Analysis of the *P. tricornutum*
210 proteome revealed eight peroxidases with Arg38, His42, His163, and Asp208; catalytic amino
211 acid residues that are conserved across all identified ascorbate peroxidases (APX) including
212 soybean APX (***Raven et al., 2003***) and its derivative APEX2 (***Figure 3—figure supplement 1C***).
213 Three of them contain proline in place of alanine at position 134 relative to APEX2, a substrate-
214 binding loop mutation rendering APEX2 highly active (***Lam et al., 2015***), and another one is
215 predicted to localize to mitochondria (***Figure 3—figure supplement 1D–E***). It is plausible that
216 one or more of these endogenous APEX2-like peroxidases are responsible for the observed
217 mitochondrial signal. APEX2 was inactivated in iron-deplete conditions (40 nM total Fe) and could
218 be reactivated by supplementing fixed cells with 10 μM hemin chloride for 3 hrs (data not shown),
219 but this substantially increased background in the Amplex UltraRed assay. Therefore, subsequent
220 proximity labeling experiments were carried out on cells growing in iron-replete conditions.

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 (cell line s2, ***Figure 2C–D***) were grown to mid- to late-exponential phase and supplemented with 2.5 mM biotin-phenol and 1 mM hydrogen peroxide, following the APEX2 proximity labeling protocol developed for yeast (***Hwang and Espenshade, 2016***; ***Hwang et al., 2016***). These labeling reaction steps were performed at 4°C and the incubation with hydrogen peroxide was extended to 20 min to mirror Amplex UltraRed assay and DAB reaction conditions (***Figure 4A***). Increasing 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 of which are critical for efficient labeling in yeast (***Hwang and Espenshade, 2016***)—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 that of 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 sample, formed two distinct clusters (***Figure 4—figure supplement 1***; ***Figure 4—figure supplement 1—source data 1***) indicating minimal technical variability. Thirty-eight statistically significant proteins (P value ≤0.05) with APEX2/WT ratios of at least 1.5 were identified (***Figure 4B***; ***Figure 4—source data 1***). These ratios were obtained from average total peptide counts across quintuplicates (***Figure 4—source data 1***). Endogenous biotinylated proteins were also detected and had APEX2/WT ratios close to one, thus acting as an intrinsic pull-down control (***Supplemental file 1—Table S1***; ***Figure 4—source data 1***). Some background enrichment in WT cells was likely due to endogenous APEX2-like peroxidases and one would therefore expect mitochondrial proteins with APEX2/WT ratios close to one 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) (***Figure 4—source data 1***). Fourteen proteins with an APEX2/WT ratio of at least two were detected. Of these 14 proteins, nine are known to be transcriptionally sensitive to iron availability; this is also the case for an additional five proteins with APEX2/WT ratio of at least 1.5 (***Figure 5—figure supplement 1—source data 1***; ***Smith et al., 2016***). We note that pTF was present, but not enriched, in our pTF-APEX2 proteomics samples (***Figure 4—source data 1***).

One possible explanation for this result is that pTF may not have surface exposed amino acid residues that are permissive to biotinylation by APEX2-generated phenoxyl radicals 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 critically involved in all of the major photosynthetic complexes—photosystem II, photosystem I, and cytochrome *b6f* (***Rochaix, 2011***)—we asked if some of the proteins are perhaps predicted to be targeted to the chloroplast. Indeed, five are, 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***; ***Kazamia et al., 2018***; ***Figure 4C***). We elaborate on the potential role for these proteins in intracellular iron trafficking in the Discussion.

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

We focused on three proteins with APEX2/WT ratio of at least two—Phatr3 IDs: J51183 (hereafter pTF.CREG1), J52498 (hereafter pTF.CatCh1), and J54986 (hereafter pTF.ap1)—that are expressed from a previously identified iron- and silicon-sensitive gene cluster on chromosome 20 (***Allen et al., 2008***; ***Sapriel et al., 2009***; ***Figure 5A***). One additional protein, ISIP2b (Phatr3_J54987, UniProt ID: B7G9B1), 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.CREG1 and pTF.ap1) are predicted to go 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***; ***Supplemental file 1—Table S2***). To test whether they 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 (***Figure 5—figure supplement 2A–B***). Imaging conditions were optimized with mCherry and Venus (yellow fluorescent protein with spectral properties very similar to those of EYFP) expressing *P. tricornutum* cell lines for minimal cross-channel bleed-through (***Figure 5—figure supplement 2C***). pTF.CREG1 and pTF.CatCh1 colocalized with pTF in the chloroplast vicinity and on the chloroplast margin, respectively (***Figure 5B***). Both proteins consistently exhibited these distinct colocalization patterns as demonstrated by additional three mCherry- and EYFP-positive cells from corresponding co-expression transconjugant cell lines (***Figure 5—figure supplement 3A–B***). Colocalization of pTF.ap1 with pTF close to the chloroplast was also evident, although somewhat less precise (***Figure 5—figure supplement 4***).

295 **pTF.CREG1 has homologs across the tree of life and may possess a non-enzymatic**
296 **role similar to human CREG1.**
297
298 To shed light on possible functions of the uncharacterized proteins colocalizing with pTF and to
299 examine their occurrence in other diatoms and marine phytoplankton beyond *Phaeodactylum*
300 *tricornutum*, phylogenetic analysis was performed. pTF.CREG1 homologs were identified across
301 the tree of life (***Figure 6***; ***Figure 6—source data 1***). Further investigation of the protein sequence
302 alignment underlying the phylogenetic tree revealed 12 amino acid residues that are at least 90%
303 conserved across all homologs (***Figure 6—figure supplement 1***). Diatom homologs can be seen
304 in a crown group with proteins from other complex plastid-containing algae such as cryptophytes,
305 haptophytes, pelagophytes, chlorarachniophytes, and dinoflagellates (***Figure 6***; ***Archibald, 2009***;
306 ***Oborník and Füssy, 2018***). pTF.CREG1 and its iron-insensitive paralog Phatr3_J10972—which
307 phylogenetically cluster together—have ~50% amino acid sequence similarity and belong to the
308 flavin mononucleotide (FMN)-binding split barrel fold protein superfamily (SSF50475) according
309 to their Ensembl Protists profiles.
310
311 We analysed pTF.CREG1 amino acid sequence (UniProt ID: B7G9B3, 234 AAs, ~26.4 kDa) and
312 compared it to the human CREG1 ortholog (UniProt ID: O75629, 220 AAs, ~24 kDa). While
313 human CREG1 is evolutionarily and structurally closely related to flavin mononucleotide (FMN)-
314 coordinating dimeric oxidases and oxidoreductases such as pyridoxine 5'-phosphate (PNP)
315 oxidases, it is unable to bind FMN and catalyze NADPH oxidation (***Ghobrial et al., 2018***; ***Sacher***
316 ***et al., 2005***). pTF.CREG1 has a predicted N-terminal signal peptide ending in SGA that may be
317 recognized by an endoplasmic reticulum-localized signal peptidase (***Benham, 2012***; ***Tuteja,***
318 ***2005***), and is likely modified by both N- and O-glycosylation, in line with human CREG1 (***Figure***
319 ***6—figure supplement 2A***; ***Sacher et al., 2005***). These features suggest that pTF.CREG1 enters
320 and travels along the secretory pathway (***Bard and Chia, 2016***; ***Benham, 2012***). Furthermore,
321 the amino acid sequence alignment between pTF.CREG1 and human CREG1 revealed a similar
322 tetrapeptide motif DP(Q/E)S which is located on the loop occluding the FMN-binding pocket at
323 the human CREG1 dimer interface (***Figure 6—figure supplement 1***; ***Figure 6—figure***
324 ***supplement 2A***; ***Sacher et al, 2005***). This short amino acid sequence—while thought to
325 contribute to the lack of catalytic activity in human CREG1—was found to be critical for mediating
326 cell growth inhibition in human teratocarcinoma cell line NTERA-2 (***Sacher et al., 2005***).
327 Structural comparison between a pTF.CREG1 homology model created using Phyre2 web portal
328 (***Kelley et al., 2015***) and the human CREG1 monomer (***Sacher et al., 2005***; PDB ID: 1XHN)
329 supports our hypothesis that pTF.CREG1 also contains a loop preventing FMN binding (***Figure***

330 ***6—figure supplement 2B***). Two arginine amino acid residues involved in binding the terminal
331 FMN phosphate in *S. cerevisiae* pyridoxine 5′-phosphate oxidase (UniProt ID: P38075, PDB ID:
332 1CI0) are mutated to aspartic acid and methionine amino acid residues in human CREG1 (***Sacher***
333 ***et al., 2005***), and one of these two key arginine residues is absent from pTF.CREG1 (***Figure 6—***
334 ***figure supplement 3A***). Finally, we note that a putative homodimerization region in pTF.CREG1
335 homology model contains many hydrophobic amino acid residues and overall follows the
336 compositional character of the human CREG1 dimerization interface suggesting pTF.CREG1 is
337 acting as a dimer *in vivo* (***Figure 6—figure supplement 3B***).
338
339 Given the apparent structural similarity of the *P. tricornutum* pTF.CREG1 to the catalytically
340 inactive human ortholog, we opted to test its activity, or lack thereof, to delineate it from the
341 evolutionarily related FMN-binding enzymes. Various truncations of His6-tagged pTF.CREG1
342 were expressed in *Escherichia coli* and purified using cobalt, Co^{2+}, immobilized metal affinity
343 chromatography (IMAC) to test solubility. A construct encoding pTF.CREG1 without the 31 N-
344 terminal amino acid residues and leading to a soluble protein was selected for larger-scale
345 purification (***Figure 6—figure supplement 4A***). The Co^{2+}-bound protein fraction (***Figure 6—***
346 ***figure supplement 4B***) exhibited no flavin reductase activity as measured by the oxidation of
347 NADPH in the presence of potential flavin substrates flavin mononucleotide (FMN) and riboflavin
348 (***Figure 6—figure supplement 2C***; ***Figure 6—figure supplement 2—source data 1***).
349
350 In conclusion, our homology modeling analyses and enzymatic assays indicate that pTF.CREG1
351 does not bind FMN and does not reduce flavins in an NADPH-dependent manner, respectively,
352 both in agreement with our current knowledge about human CREG1. As such, pTF.CREG1 is an
353 enticing candidate for further study as a key participant in the endocytic pTF receptor-mediated
354 iron uptake pathway in diatoms. We put forth a hypothesis for its function in the Discussion.

pTF.CatCh1 is a heterokont-specific protein with conserved CX(X)C motifs.

In contrast to pTF.CREG1, pTF.CatCh1 homologs were identified only in diatoms and other single-celled heterokonts (***Figure 7A***; ***Figure 7—source data 1***). pTF.CatCh1 is a ~37.4 kDa (349 AAs) large protein with a predicted signal peptide (AAs 1–26) ending in ASA, two N-glycosylation NX(S/T) sequons (i.e., asparagines in the identified NST and NLS motifs may be glycosylated), and 13 predicted O-glycosylation sites (***Rao and Wollenweber, 2010***; data not shown). This protein is paralogous to the iron starvation induced protein 2b (ISIP2b/Phatr3_J54987) whose gene lies adjacently to *pTF.CatCh1* in the same chromosome 20 gene cluster (***Figure 5A***). They share three CXC and one CXXC motifs typical of metal-binding, redox-active, and iron–sulfur cluster-containing proteins (***Figure 7B***; ***Blaby-Haas et al., 2014***; ***Fomenko and Gladyshev, 2003***; ***Poole, 2014***; ***Przybyla-Toscano et al., 2018***). One of the CXC motifs and the CXXC motif are conserved across all but two homologs in our phylogenetic analysis (***Figure 7B***; ***Figure 7—figure supplement 1A***), and the latter is predicted to be located in one of the two disordered C-terminal regions in pTF.CatCh1 (AAs 196–292, AAs 321–349; ***Figure 7—figure supplement 1B***). With additional 12 cysteines conserved between pTF.CatCh1 and ISIP2b, pTF.CatCh1 contains a total of 20 cysteine amino acid residues (all between AAs 34 and 226). Six of these 20 are 100% conserved across all homologs (***Figure 7B***). The predicted transmembrane domain (AAs 296–315) is flanked by a polyserine ($[Ser]_6$) and a short arginine-rich stretch ($RKL[R]_3$). Flexible polyserine linkers are found in modular, multidomain proteins (***Uversky, 2015a***), and positively charged amino acid residue tracts in the vicinity of transmembrane regions were shown to be orientation determinants in outer chloroplast membrane proteins (***May and Soll, 1998***). These observations are synthesized in a domain organization schematic shown in ***Figure 7—figure supplement 1C***. We elaborate on possible roles of the conserved CX(X)C motifs, as well as other identified motifs and domains in pTF.CatCh1 in the Discussion.

pTF.ap1 homologs are present in marine microeukaryotes including diatoms.

Phylogenetic characterization was performed with pTF.ap1 as well and numerous homologs with at least five highly conserved motifs were identified in diatoms and other marine microeukaryotes, including chlorophytes, cryptophytes, and haptophytes (***Figure 7—figure supplement 2A–B***; ***Figure 7—figure supplement 2—source data 1***). These pTF.ap1 homologs contain five 100% conserved cysteine amino acid residues, two of which are shown in ***Figure 7—figure supplement 2B***, as well as 100% conserved tyrosine, histidine, and aspartic acid residues,

389 indicating this protein may also be redox-active and/or have a metal ion processing role
390 (***Dokmanić et al., 2007***).

Discussion

In this study, we used the iron receptor protein phytotransferrin (pTF) to implement APEX2-mediated proximity-dependent proteomic mapping in the model marine diatom *Phaeodactylum tricornutum*. The resulting data provide several advances regarding the function of pTF and its putative protein interaction partners downstream of iron binding at the cell surface.

pTF (ISIP2a; iron starvation induced protein 2a) was identified in *P. tricornutum* as a marker for iron (Fe) limitation (***Allen et al., 2008***). It was subsequently shown to be involved in Fe acquisition (***Morrissey et al., 2015***) and to be widespread in marine and freshwater microeukaryotic phytoplankton communities (***Bertrand et al., 2015***; ***Carradec et al., 2018***; ***McQuaid 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 labile ferric, Fe^{3+}, iron, a critical micronutrient for cellular biochemistry. ***McQuaid et al., 2018***, showed that pTF is essential for high-affinity iron uptake in *P. tricornutum*, and that carbonate, CO_3^{2-}, and Fe^{3+} interact synergistically to control the iron uptake rate. The finding that carbonate anions are required for the activity of a key diatom iron transport system is noteworthy and suggests that ocean acidification might inhibit microbial iron uptake. The occurrence of transferrin in marine diatoms and other single-celled marine and freshwater phytoplankton raises significant questions concerning its evolutionary origin and cellular role.

McQuaid et al., 2018, observed pTF in intracellular puncta, showed that the clathrin-mediated endocytosis inhibitor Pitstop 2 reduces iron (Fe) uptake rates, visualized vesicles after adding iron to an iron-limited WT *P. tricornutum* culture stained with the membrane dye FM 1-43, but our MDY-64 labeling data provide the first direct evidence that diatom pTF is indeed associated with intracellular membranous compartments. Further work is needed to elucidate pTF dynamic upon cellular internalization—especially after iron addition to iron-deplete cells which mimicks oceanic iron repletion events—and its trafficking in and potential exit from the endomembrane system in relation to cell cycle phases and light-dark cycles (***Mayle et al., 2012***; ***Naslavsky and Kaplan, 2018***; ***Smith et al., 2016***).

While APEX2 was expected to be active in *P. tricornutum* given previous heterologous expression of functional horseradish peroxidase (HRP) in the model marine diatom *Thalassiosira pseudonana* (***Sheppard et al., 2012***), it was unclear whether endogenous peroxidases would

425 contribute to the background signal in various APEX2 assays. In our Amplex UltraRed assay
426 experiments, an order of magnitude higher signal-to-noise ratio (i.e., resurofin signal in pTF-
427 APEX2 expressing versus WT *P. tricornutum* cells) was detected when live cells were reacted on
428 ice as opposed to on room temperature, suggesting endogenous *P. tricornutum* peroxidases, but
429 not exogenous APEX2, are largely inhibited on ice. This is notable as it indicates that performing
430 APEX2 assays on ice may represent a generalizable experimental strategy for mesophilic
431 microeukaryotic phytoplankton model systems.
432
433 Identifying Amplex UltraRed assay conditions which minimized WT background was promising,
434 but not a guarantee that some endogenous enzymes would not be able to process other APEX2
435 substrates, such as those crucial for electron microscopy (3,3'-diaminobenzidine/DAB) and
436 proximity proteomics (biotin-phenol). Indeed, mitochondrial background signal was detected in
437 our electron microscopy experiments both in WT and pTF-APEX2 expressing cells, and our *P.*
438 *tricornutum* proteome analysis revealed candidate endogenous APEX2-like peroxidases that
439 could explain this result. Nevertheless, C-terminal tagging of pTF with APEX2 resulted in the
440 expected cell surface and what appeared to be vesicular fusion protein localization away from
441 mitochondria indicating no spatial overlap between endogenous APEX2-like enzymes and pTF-
442 APEX2. Inconsistent size and shape of vesicle-like structures observed in our electron
443 micrographs were conceivably due to a combination of arbitrary cell orientation inside agar blocks
444 (where diatom cells were embedded) and the exact z position of ultramicrotome cut sites. On the
445 whole, APEX2-based imaging is complementary to super-resolution microscopy in diatoms and
446 represents a basis for electron tomography applications (***Gröger et al., 2016***; ***Sengupta et al.,***
447 ***2019***). Some off-target biotinylation in our proximity labeling experiments was likely due to
448 background mitochondrial peroxidase activity as indicated by the presence of mitochondrial
449 proteins in our mass spectrometry data. Their APEX2/WT ratios close to one give us more
450 confidence that those with elevated ratios above 1.5 were enriched due to APEX2 activity. Many
451 of the latter have functional annotations for proteins we might expect to play a role in endosomal
452 iron assimilation, and many have previously been shown to have iron-sensitive transcripts, both
453 of which further support this notion. Future efforts to express pTF tagged with N-terminal APEX2,
454 which should lead to intravesicular proximity labeling according to our pTF orientation prediction,
455 optimizing conditions for APEX2 reactivation in iron-deplete conditions with hemin chloride, and
456 more cell wall-permeant APEX2 probes (***Tian et al., 2020***) will allow for identification of additional
457 candidate proteins. Considering iron metabolism genes in *P. tricornutum*, including *pTF*, are
458 highly expressed at night, proteomic maps with improved spatial and temporal resolution may

459 emerge from proximity-dependent proteomic mapping experiments conducted with synchronized
460 cell lines sampled at different timepoints throughout diel cycle (**Huysman et al., 2014**; **Lundberg**
461 **and Borner, 2019**; **Smith et al., 2016**). The successful implementation of TurboID—an
462 engineered biotin ligase for proximity proteomics (**Branon et al., 2018**)—in model plants
463 *Arabidopsis thaliana* (**Kim et al., 2019**; **Mair et al., 2019**) and *Nicotiana benthamiana* (**Zhang et**
464 **al., 2019**) suggests its adoption in existing and emerging single-celled marine phytoplankton
465 model systems, including diatoms, is likely (**Falciatore et al., 2020**; **Faktorová et al., 2020**).
466 TurboID is an iron-independent protein and could thus be particuarly valuable to further advance
467 iron metabolism studies in *P. tricornutum*. These proposed improvements and alternative
468 experimental approaches should cross-validate the proteins identified in this study as well as
469 illuminate new ones.
470
471 Functionally uncharacterized gene cluster on chromosome 20 we zoomed in on is transcriptionally
472 upregulated in iron-deplete (**Allen et al., 2008**; **Smith et al., 2016**) and silicic acid-replete (**Sapriel**
473 **et al., 2009**) conditions corroborating a known link between iron and silicon metabolism in diatoms
474 (**Brzezinski et al., 2015**; **Durkin et al., 2012**; **Hutchins et al., 1998**; **Leynaert et al., 2004**).
475 Physically clustered functionally related genes are a hallmark of prokaryotic biology and operon-
476 like genomic elements have been described in fungi and plants (**Osbourn and Field, 2009**;
477 **Nützmann et al., 2018**), however identifying and characterizing them in marine diatom genomes
478 is in its relative infancy. For example, genes encoding light harvesting fucoxanthin-chlorophyll-
479 binding proteins (FCPs) are co-localized on chromosome 2 in *P. tricornutum* (**Bhaya and**
480 **Grossman, 1993**), and a compact ~7 kbp gene cluster encoding enzymes for domoic acid
481 neurotoxin production was characterized in a cosmopolitan diatom *Pseudo-nitzschia multiseries*
482 (**Brunson et al., 2018**). Given that diatom genomes have likely been extensively shaped by
483 horizontal gene transfer (**Bowler et al., 2008**; **Diner et al., 2017**), it is interesting to note that
484 some yeast species acquired iron acquisition-enabling operon-like genomic elements via
485 horizontal operon transfer (HOT) from an ancient bacterium (**Kominek et al., 2019**). Our
486 phylogenetic analyses suggest all three proteins—pTF.CREG1, pTF.CatCh1, pTF.ap1—encoded
487 by this locus are present in diatoms and other (marine) microeukaryotes, while pTF.CREG1
488 homologs are found across the tree of life. They, alongside the rest of our proteomic data, paint
489 an emerging view of intracellular ferric, Fe^{3+}, iron processing events downstream of its
490 phytotransferrin-mediated binding at the cell surface.
491

492 In humans, endosome acidification causes structural rearrangements in the transferrin receptor
493 protein and protonation of synergistically coordinated carbonate anion leading to ferric iron
494 release (**Cheng et al., 2004**). This event is concurrent with or followed by endosomal reductase
495 STEAP3-catalyzed ferric iron reduction and ferrous, Fe^{2+}, iron export through divalent metal
496 transporter DMT1/NRAMP2 or Zrt- and Irt-like protein (ZIP) 14 (**Bogdan et al., 2016**). Two
497 putative proton, H^+, pumps were present in our mass spectrometry data: Phatr3_J23497—
498 predicted as a P-type ATPase—and Phatr3_J27923—a vacuolar-type H^+-ATPase subunit A.
499 Multisubunit V-ATPases regulate pH homeostasis in virtually all eukaryotes and are known to
500 function as endosome acidifying protein complexes (**Finbow and Harrison, 1997**; **Maxson and**
501 **Grinstein, 2014**), which makes Phatr3_J27923 a candidate protein involved in acidification of
502 phytotransferrin-rich endosomes in *P. tricornutum*. Notably, V-ATPase transcripts were
503 overrepresented in diatoms following iron enrichment, possibly to account for an increase in the
504 number of endosomes associated with pTF-driven iron acquisition (**Marchetti et al., 2012**), and
505 V-ATPase was shown to be involved in diatom silica deposition vesicle (SDV) acidification
506 (**Hildebrand et al., 2018**; **Yee et al., 2019**). Endosome acidification, perhaps via V-ATPase, may
507 facilitate phytotransferrin-bound ferric, Fe^{3+}, iron release, possibly for processing by an iron
508 reducing enzyme similar to the human metalloreductase STEAP3.
509
510 Human cellular repressor of E1A-stimulated genes 1 (CREG1) is a glycoprotein involved in
511 embryonic development, growth, differentiation, and senescence as part of the endosomal-
512 lysosomal system (**Ghobrial et al., 2018**; **Schähs et al., 2008**). Contrary to evolutionarily related
513 flavin mononucleotide (FMN)-binding split barrel fold oxidases and oxidoreductases such as
514 pyridoxine 5′-phosphate (PNP) oxidase, an enzyme involved in vitamin B_6 metabolism (**Musayev**
515 **et al., 2003**), CREG1 forms a homodimer unable to bind FMN *in vitro* suggesting it does not act
516 as an enzyme *in vivo* (**Sacher et al., 2005**). The lack of FMN in human CREG1 is due to a short
517 loop protruding into the FMN-binding pocket at the dimer interface, and aspartic acid and
518 methionine amino acid residues replacing the corresponding two yeast PNP oxidase (PDB ID:
519 1CI0) arginines responsible for binding the terminal FMN phosphate (**Sacher et al., 2005**).
520
521 CREG1-like protein pTF.CREG1 was proposed to be central to sustained growth of iron-limited
522 *P. tricornutum* cells (**Allen et al., 2008**) or to serve as a cell surface iron receptor (**Lommer et al.,**
523 **2012**). CREG transcripts were more common in diatom transcriptomes from the Southern Ocean
524 as opposed to non-Southern Ocean regions indicating these proteins are important in coping with
525 iron limitation in this major high nutrient, low chlorophyll (HNLC) zone (**Moreno et al., 2018**). This

526 may also explain why more diatom homologs are not present in our phylogenetic tree as
527 transcriptomes from the Marine Microbial Eukaryote Transcriptome Sequencing Project
528 (MMETSP) database used in our phylogenetic analysis were not obtained from iron-limited
529 cultures (***Caron et al., 2016***; ***Keeling et al., 2014***). Our protein feature analysis revealed a signal
530 peptide and multiple putative glycosylation sites which suggests that pTF.CREG1 enters the
531 secretory pathway (***Bard and Chia, 2016***). While the nature of pTF.CREG1 recruitment to and
532 colocalization with pTF-rich endosomes in the vicinity of chloroplast periphery—as suggested by
533 our fluorescence microscopy results—is unclear, it is possible that pTF.CREG1-containing
534 vesicles intersect with pTF-rich endosomes *en route* through the secretory pathway. Vesicular
535 trafficking is highly dynamic and bidirectional (***Naslavsky and Kaplan, 2018***; ***Progida and***
536 ***Bakke, 2016***), and early endosomal traffic consists of sequential vesicle fusion events (***Brenner,***
537 ***2012***), so this is at least in principle plausible. Punctate pTF.CREG1 colocalization with pTF
538 suggests that this protein acts in an endosomal-lysosomal system, and this localization pattern is
539 consistent with the observation that proteins with low confidence ASAFind chloroplast prediction
540 (such as pTF.CREG1) can be associated with “blob-like structures” (BLBs) adjacent to, but not
541 inside, the chloroplast (***Gruber et al., 2015***; ***Kilian and Kroth, 2005***). Given that our transgenic
542 colocalization *P. tricornutum* cell lines were neither synchronized nor grown in iron-deplete
543 conditions, it is plausible the protein also acts closer to the cell surface. Likely, but not confirmed,
544 APEX2 orientation away from the cell surface and endosome interior in our proximity labeling
545 experiments, and the reported inability of APEX2-generated phenoxyl radicals to penetrate
546 endomembranes (***Rhee et al., 2013***), would suggest that pTF.CREG1—at some point during its
547 recruitment to pTF-rich endosomes—has to be exposed to the cytosol to promote efficient
548 biotinylation. Alternatively, it is impossible to rule out the possibility that lipid composition of *P.*
549 *tricornutum* endomembranes is permissive to phenoxyl radical diffusion (***Zulu et al., 2018***).
550
551 Domain annotation in pTF.CREG1 (UniProt ID: B7G9B3) encouraged us to perform amino acid
552 sequence and structural alignments with human CREG1 (UniProt ID: O75629, PDB ID: 1XHN).
553 These analyses revealed that pTF.CREG1 lacks at least one arginine amino acid residue required
554 for terminal FMN phosphate binding and likely contains an FMN binding pocket-occluding loop,
555 two key features delineating human CREG1 from the yeast PNP oxidase (***Sacher et al., 2005***).
556 Our hypothesis that pTF.CREG1 is unable to coordinate FMN was initially supported by our
557 protein purifications which were not yellow as expected from a typical flavoprotein (***Theorell,***
558 ***1935***), and was subsequently strengthened by their catalytic inactivity upon supplementation with
559 biologically relevant flavin cofactors. The demonstrated lack of enzymatic activity in pTF.CREG1

560 and localization data from *in vivo* fluorescent reporter experiments represent an important
561 extension of the human CREG1 literature to the single-celled eukaryotic phytoplankton field.
562
563 Secreted glycosylated human CREG1 was shown to delay G1/S cell cycle phase transition via
564 direct interaction with mannose 6-phosphate/insulin-like growth factor II receptor (M6P/IGF2R), a
565 receptor internalized via clathrin-mediated endocytosis (***Di Bacco and Gill, 2003***; ***Ghosh et al.,***
566 ***2003***; ***Han et al., 2009***). This cellular growth inhibition was demonstrated to, at least partially,
567 depend on a tetrapeptide motif within an FMN binding pocket-occluding loop (***Sacher et al.,***
568 ***2005***). In conjunction with our amino acid sequence alignments and structural prediction which
569 point to the conservation of these two features in pTF.CREG1, these studies lend additional
570 support to our data suggesting colocalization and potential direct physical association between
571 phytotransferrin and pTF.CREG1 is non-enzymatic in nature. Internalization of the human
572 transferrin-transferrin receptor complex is not a constitutive, but rather a regulated kinase activity-
573 dependent process (***Cao et al., 2016***). We currently hypothesize that pTF.CREG1 may serve as
574 a "checkpoint" to keep phytotransferrin endocytosis at a low level in iron-replete conditions, but
575 when iron is scarce instead positively regulate this pathway. Indeed, transcriptomics experiments
576 show that a second transcriptional peak for *pTF.CREG1* occurs in iron-deplete medium approx.
577 8 hrs prior to the *pTF* peak, a transcriptional dynamic which would be consistent with our
578 hypothesis (***Figure 5—figure supplement 1***; ***Smith et al., 2016***). Given that glycosylation in
579 human CREG1 appears to be important for interaction with its cognate receptor (***Han et al., 2011***;
580 ***Sacher et al., 2005***; ***Schähs et al., 2008***), posttranslational modifications, perhaps
581 glycosylation(s), of pTF.CREG1 may further drive this regulatory activity. The presence of
582 Phatr3_J10972, an iron-insensitive pTF.CREG1 paralog, in the *P. tricornutum* genome suggests
583 that CREG-like proteins are implicated in diverse endocytosis pathways in single-celled
584 eukaryotic phytoplankton. However, the association between iron-sensitive pTF.CREG1 and
585 phytotransferrin in *P. tricornutum* implies a specific pTF.CREG1 function in non-reductive iron
586 uptake. It will be very interesting to define the exact pTF.CREG1 role in further cellular and
587 mechanistic studies and compare it to that of its homologs in multicellular organisms.
588 Conceivably, an ancient protein like pTF.CREG1 could have remained at the core of linking cell
589 cycle progression, cellular growth, and nutrient homeostasis across the tree of life.
590
591 In contrast to pTF.CREG1, no functional annotation could be assigned to and no structural
592 homology models were generated for pTF.CatCh1 in the Pfam database (***El-Gebali et al., 2019***)
593 and by Phyre2, respectively. This protein has a predicted N-terminal signal peptide and numerous

N- and O-glycosylation sites suggesting it enters and trafficks along the secretory pathway. The outermost complex plastid membrane in diatoms corresponds to the chloroplast endoplasmic reticulum (cER) and is continuous with the nuclear envelope (***Gibbs et al., 1981***; ***Prihoda et al., 2012***). Nuclear-encoded *P. tricornutum* proteins destined for the periplastidial compartment (vastly reduced endosymbiont cytoplasm between the inner two and the outer two complex plastid membranes) are first cotranslationally imported into the cER lumen (***Grosche et al., 2014***; ***Peschke et al., 2013***). Our microscopy results show that pTF.CatCh1 is localized on the chloroplast margin, perhaps anchored to the outer chloroplast (cER) membrane with its C-terminal transmembrane domain, thus raising a possibility that pTF.CatCh1 gets to its final destination via cotranslational membrane insertion.

The cysteine-rich domain contains one CXC and one CXXC motifs which are nearly invariably present in pTF.CatCh1 homologs. Two cysteines spaced by two non-cysteine amino acids (the CXXC motif) are commonly found in metal-binding proteins, including those that bind iron (***Fomenko et al., 2008***; ***Gladyshev et al., 2004***). This motif in pTF.CatCh1 is preceded by an aspartic acid residue resulting in DCXXC which is known to be particularly permissive to binding divalent metal ions (***Figure 7—figure supplement 1A***; ***Banci et al., 2002***; ***Banci et al., 2006***). Histidine, tyrosine, methionine, aspartic acid, and glutamic acid residues can also be involved in iron cation coordination, but none of these are conserved across pTF.CatCh1 homologs (***Figure 7—source data 1***; ***Dokmanić et al., 2007***). Given the nearly completely conserved C(X)XC motifs and additional six 100% conserved cysteine residues, it is conceivable they are critical to pTF.CatCh1 activity.

Based on the fluorescence microscopy data, the conserved cysteine amino acid residues and cysteine-containing motifs, and given that disordered protein domains—two are predicted to flank the pTF.CatCh1 transmembrane domain—are mediators of protein-protein interactions and play an important role in the formation of protein complexes (***Uversky, 2015b***; ***Uversky, 2016***), we are considering two hypotheses for *in vivo* pTF.CatCh1 function.

First, pTF.CatCh1 may serve as a metallochaperone that binds ferrous, Fe^{2+}, iron downstream of the endosomal ferric, Fe^{3+}, iron reduction step, and directs it to the chloroplast interior. Metallochaperones are intracellular proteins responsible for protecting and guiding metal ions on their way to correct metalloenzyme sinks (***O'Halloran and Culotta, 2000***). They engage in protein-protein interactions and help prevent metal-induced toxicity caused by free radical-

generating Fenton reactions which are especially common with iron and copper (***Philpott and Jadhav, 2019***; ***Rosenzweig, 2002***; ***Valko et al., 2005***).

Second, pTF.CatCh1 may recruit pTF to the chloroplast periphery via one of its two C-terminal disordered regions flanking the predicted transmembrane domain, and thus serve as a docking site for iron-laden pTF. It could achieve this via the “fly-casting mechanism” which involves binding-induced protein folding (***Shoemaker et al., 2000***). The observed colocalization of pTF.CatCh1 with pTF may indicate that Fe^{3+} can be offloaded directly from pTF to pTF.CatCh1 which would imply an additional, non-endosomal, intracellular Fe^{3+} reduction step. Notably, a chloroplast-associated Fe^{3+} reduction pathway mediated by Fe^{3+} chelate reductase FRO7 is present in *Arabidopsis thaliana* (***Jeong et al., 2008***). Abberant cellular localization of heterologously expressed fluorescent protein fusions is a widely observed and acknowledged phenomenon (***Jensen, 2012***), but we note that pTF localization on the chloroplast margin insdistinguishable from that in this study was reported by ***McQuaid et al., 2018*** suggesting it is biologically relevant. Similarly to *pTF.CREG1*, *pTF.CatCh1* exhibits a second transcriptional peak in iron-deplete medium approx. 16 hrs prior to the *pTF* peak (***Figure 5—figure supplement 1***). It would make sense to synthesize a protein critical for iron-laden pTF recruitment to the chloroplast margin prior to significant increase in basal pTF endocytosis.

All in all, our data suggest pTF.CatCh1 is a chloroplast margin-localized, pTF-associated, *P. tricornutum* protein with a possible role in chloroplast iron homeostasis. The two hypotheses are put forth here to help frame its further genetic, cellular, biochemical, and structural dissection.

Chloroplasts are the most iron-rich system in plant cells (***López-Millán et al., 2016***). Iron—in particular in the form of iron–sulfur (Fe–S) clusters—serves as an essential cofactor for the photosynthetic electron transfer chain, chlorophyll biosynthesis, and chloroplast protein import (***Balk and Schaedler, 2014***; ***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 (***Hertle et al., 2020***; ***Lindquist and Aronsson, 2018***), an endoplasmic reticulum to Golgi to chloroplast protein trafficking pathway exists (***Radhamony and Theg, 2006***; ***Villarejo et al., 2005***), and additional hypotheses for such specialized vesicle-mediated plastid targeting of proteins are established (***Baslam et al., 2016***). Therefore, a direct link between endocytic internalization of phytotransferrin-bound iron and its offloading to or adjacent to chloroplasts as suggested by our results seems plausible. “Kiss and run” mechanism that allows erythroid cell

662 mitochondria to access iron via direct interaction with transferrin-rich endosomes further
663 underscores this idea (**Hamdi et al., 2016**). Analysis of our mass spectrometry (MS) hits that were
664 not co-expressed with pTF-mCherry reveals three additional proteins—ISIP1 (Phatr3_J55031),
665 FBP1 (Phatr3_J46929), and FBAC5 (Phatr3_J41423)—that support the idea of a direct cell
666 surface-to-chloroplast iron trafficking pathway in *Phaeodactylum tricornutum*.
667
668 *ISIP1* (iron starvation induced protein 1; Phatr3_J55031) gene expression in *P. tricornutum* and
669 other marine microeukaryotes is induced by iron limitation (**Allen et al., 2008**; **Kazamia et al.,**
670 **2018**; **Marchetti et al., 2012**). *P. tricornutum* ISIP1—the second most enriched protein in our MS
671 dataset—has a role in siderophore-bound iron uptake and colocalizes with phytotransferrin (pTF)
672 adjacently to the chloroplast in an "iron-processing compartment" (**Kazamia et al., 2019**; **Kazamia**
673 **et al., 2018**). FBP1 (Phatr3_J46929)—iron-sensitive siderophore ferrichrome-binding protein with
674 punctate intracellular localization and central to the carbonate-independent iron acquisition
675 pathway in *P. tricornutum* (**Allen et al., 2008**; **Coale et al., 2019**)—was another highly enriched
676 protein in our dataset indicating the internalization pathways for inorganic (i.e., phytotransferrin-
677 bound) and organic (i.e., siderophore-bound) iron in *P. tricornutum* may be physically coupled
678 intracellularly and directed to iron-accumulating vesicles close to the chloroplast. Metal
679 homeostasis by such discrete membranous compartments is widespread in nature (**Blaby-Haas**
680 **and Merchant, 2014**). Copper- and zinc-storing compartments—cuprosomes and zincosomes,
681 respectively—are present in the model chlorophyte microalga *Chlamydomonas reinhardtii* (**Aron**
682 **et al., 2015**; **Hong-Hermesdorf et al., 2014**; **Merchant, 2019**), and lipid-bound iron-accumulating
683 ferrosome compartments exist in diverse bacteria and archaea (**Grant and Komeili, 2020**).
684
685 Finally, the most enriched protein in our MS dataset was FBAC5 (Phatr3_J41423)—iron-sensitive
686 iron-independent class I fructose-biophosphate aldolase—which was shown to be localized to the
687 *P. tricornutum* pyrenoid—proteinaceous RuBisCO-containing subchloroplastic organelle—and
688 proposed to link Calvin-Benson-Bassham cycle activity with the CO_2 concentrating mechanism
689 (**Allen et al., 2012**; **Matsuda et al., 2017**). FBAC5 is one of the most reliable markers for diatom
690 iron stress as demonstrated by laboratory culture (**Cohen et al., 2018**; **Lommer et al., 2012**),
691 microcosm incubation (**Cohen et al., 2017**; **Cohen et al., 2017**), and field studies (**Bertrand et**
692 **al., 2015**), however its exact role has remained elusive. Fructose-biphosphate aldolases (FBAs)
693 are glycolytic enzymes known for their ability to moonlight (**Boukouris et al., 2016**), a term for
694 proteins with multiple physiologically relevant functions encoded by one polypeptide chain
695 (**Jeffery, 2018**). Moonlighting FBAs have roles in actin cytoskeleton dynamic, cellular growth

696 arrest, and apoptosis (in human cancer cell lines; ***Gizak et al., 2019***), transcriptional regulation
697 (in yeast; ***Cieśla et al., 2014***), and endosome acidification via direct interaction with vacuolar-type
698 ATPase (in a mouse cell line; ***Merkulova et al., 2011***). These reported non-catalytic FBA activities
699 suggest that FBAC5 acting outside of the *P. tricornutum* chloroplast, perhaps in the endosomal-
700 lysosomal system as our data would imply, is possible.
701
702 Our initial characterization of pTF.CREG1 and pTF.CatCh1, together with the discussion on the
703 four other mass spectrometry hits, enable proposition of a model connecting cell surface ferric
704 iron binding, internalization, and intracellular trafficking in *Phaeodactylum tricornutum* with many
705 outstanding questions ripe for future investigation (***Figure 8***). Importantly, it remains to be
706 determined which protein is responsible for intracellular ferric, Fe^{3+}, iron reduction in *P.*
707 *tricornutum*, activity that would be consistent with the non-reductive cell surface ferric, Fe^{3+}, iron
708 uptake proposed for phytotransferrin (pTF) (***McQuaid et al., 2018***; ***Morrissey et al., 2015***).
709 Perhaps notably, neither of the five proteins with clear ferric reductase annotation (FRE1–5,
710 Phatr3 IDs: J54486, J46928, J54940, J54409, J54982) nor additional five hypothetical ferric
711 reductases (***Coale et al., 2019***; ***Smith et al., 2016***) were present in our mass spectrometry data.
712 Further proximity-dependent proteomic mapping experiments will be essential for determining the
713 full set of proteins in the non-reductive iron acquisition pathway in *P. tricornutum* as enzymatic
714 Fe^{3+} reduction may happen later in the pathway and/or in a separate subcellular compartment.
715
716 In conclusion, the identified proteins and the associated model we introduce in this work represent
717 an important advance for the marine phytoplankton field, given that diatoms are central to the
718 study of cellular iron homeostasis and its link to fluctuating iron levels in the ocean. The presented
719 insights are vital as our understanding of intracellular iron trafficking is relatively incomplete even
720 in significantly more established model systems such as mammalian cells, and our knowledge of
721 chloroplast iron uptake is in its infancy (***Blaby-Haas and Merchant, 2012***; ***Philpott and Jadhav,***
722 ***2019***). We anticipate further molecular dissection of our proteomics dataset will not only provide
723 details on the transformation dynamic of acquired ferric iron and its cellular sinks but will thereby
724 also illuminate the extent to which phytotransferrin pathway mirrors the metazoan transferrin
725 cycle.
726
727 Implementing the use of APEX2 in a model marine diatom is itself a significant and timely advance
728 as the field is expanding and moving from broad -omics-based surveys to detailed molecular
729 studies (***Falciatore et al., 2020***; ***Faktorová et al., 2020***). APEX2 and related molecular tools hold

great promise to dissect additional key subcellular compartments in diatoms such as pyrenoids and silica deposition vesicles (***Figure 8—figure supplement 1***; ***Barrett et al., 2021***; ***Hildebrand et al., 2018***; ***Matsuda et al. 2017***; ***Wang and Jonikas, 2020***), two exciting future avenues for these single-celled heterokonts uniquely positioned at the nexus of evolutionary cell biology and blue biotechnology.

735 ## Materials and Methods

736

737 **Vector cloning**

738

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

747

748 **pJT_native_pTF-mCherry:** Gibson Assembly was performed with three DNA amplicons: (1)
749 pPtPBR1 episome backbone split into two fragments and (2) expression cassette including *pTF*,
750 *mCherry*, and *pTF* promoter and terminator.

751

752 **pJT_pTF-mCherry_MS hit-EYFP:** pJT_native_pTF-mCherry was split into two fragments via
753 PCR stitching (keeping ampicillin resistance gene split). Genes corresponding to MS hits were
754 amplified using cDNA from iron-starved WT *P. tricornutum* cells and assembled with *EYFP* and
755 *Phatr3_J23658* (flavodoxin-encoding gene) promoter and terminator through two PCR stitching
756 rounds into a single fragment. Gibson Assembly was then performed to combine all three final
757 amplicons. Detailed assembly scheme is presented in ***Figure 5—figure supplement 2A***.

758

759 **pJT_Δ31_pTF.CREG1-His6:** Cloning of gene fragments into the *E. coli* protein expression vector
760 PtpBAD-CTHF was performed as described previously (***Brunson et al., 2018***). Briefly, PtpBAD-
761 CTHF was linearized by digestion with XhoI (New England Biolabs (NEB), Ipswich, MA, Catalog
762 #R0146S) and the resulting DNA was column purified. *Δ18_pTF.CREG1* gene was obtained by
763 PCR from *P. tricornutum* gDNA with PrimeSTAR GXL DNA Polymerase (Takara Bio, Kusatsu,
764 Japan, Catalog #R050B) and primer set JT31/JT32 to incorporate the appropriate Gibson
765 Assembly overhangs and remove the predicted N-terminal signal peptide (amino acid residues
766 1–18). Insertion of truncated *pTF.CREG1* into linearized PtpBAD-CTHF was performed using
767 Gibson Assembly Master Mix (NEB, Catalog #E2611S), 1 uL of the Gibson Assembly reaction
768 mixture was transformed via heat shock into chemically competent NEB 5-alpha cells (NEB,

769 Catalog #C2988J), and cells were incubated on lysogeny broth with 10 µg/mL tetracycline (LB-
770 Tet10) 1% agar plates overnight at 37°C. Transformants were screened by colony PCR using the
771 primer set JT37/JT38 and Sapphire polymerase (Takara Bio), and positive clones were selected
772 for outgrowth. Isolated plasmids were sequence-validated by Sanger sequencing (Eurofins,
773 Luxembourg, Luxembourg). A sequence-validated clone was designated as PtpBAD-
774 Δ18_pTF.CREG1-CTHF and transformed into chemically competent BL21 *E. coli* cells (NEB,
775 Catalog #C2530H) which were spread on LB-Tet10 1% agar plates. The resulting transformants
776 were used for subsequent Δ18_pTF.CREG1 expression experiments. Following unsuccessful
777 expression testing of the Δ18 N-terminal truncation construct, an additional set of genes encoding
778 pTF.CREG1 with N-terminal truncations (Δ26, Δ31, Δ39, Δ43) was generated by PCR using
779 PrimeStar GXL DNA Polymerase (Takara Bio), primer sets JT33/JT32, JT34/JT32, JT35/JT32,
780 JT36/JT32, and PtpBAD-Δ18_pTF.CREG1-CTHF as a template. Assembly, colony PCR
781 screening, plasmid isolation, sequencing, and transformation into chemically competent BL21 *E.*
782 *coli* cells of all additional expression vectors, including PtpBAD-Δ31_pTF.CREG1-CTHF
783 (=pJT_Δ31_pTF.CREG1-His6), was performed as above. An additional construct encoding
784 pTF.CREG1 with an N-terminal His6 and Δ39 N-terminal truncation was built (PtpBAD-NTH-
785 Δ39_pTF.CREG1) using the vector NTH-PtpBAD (constructed similarly to PtpBAD-CTHF; see
786 ***Brunson et al., 2018*** and ***Savitsky et al., 2010***), but was not pursued beyond initial expression
787 testing.
788
789 *pTF* (*Phatr3_J54465*) in all episomes was in its native form (3 exons, 2 introns). Molecular cloning
790 primers are listed in ***Supplemental file 1—Table S3.*** Vector details, further amplicon information,
791 and fusion protein sequences are catalogued in ***Supplemental file 1—Table S4***.
792
793 Related resource: Turnšek J. and Gholami P. 2017. Guidelines for highly efficient construction
794 of diatom episomes using Gibson Assembly. *protocols.io*. DOI:
795 dx.doi.org/10.17504/protocols.io.jy7cpzn
796
797 **Diatom culturing, conjugation, and genotyping**
798
799 **Culturing:** Sequenced *Phaeodactylum tricornutum* strain CCMP632 (synonymous to CCMP2561
800 and CCAP 1055/1; National Center for Marine Algae and Microbiota (NCMA), East Boothbay,
801 ME) was used throughout the study and grown at 18°C, 300 µmol quanta m^{-2} s^{-1}, and a 10 h:14
802 h dark:light cycle in biotin-free L1 medium prepared by mixing 1 L Aquil salts, 2 mL nitrate and

803 phosphate (NP) nutrient stock, 1 mL trace metal stock, and 1 mL thiamine hydrochloride and
804 cyanocobalamin (TC) stock unless otherwise noted. Preparation of Aquil salts: 0.5 L anhydrous
805 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
806 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
807 $MgCl_2$ x $6H_2O$, 1.54 g $CaCl_2$ x $2H_2O$, 100 µL 170 mg/mL $SrCl_2$ x $6H_2O$ stock) were prepared
808 separately, combined, filter sterilized (0.2 µm), and stored at room temperature. Preparation of
809 NP nutrient stock: 37.5 g $NaNO_3$ and 2.5 g NaH_2PO_4 were dissolved in 100 mL Milli-Q, filter
810 sterilized (0.2 µm), and stored at 4°C. Preparation of trace metal stock (for 1 L 1,000x stock): 3.15
811 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
812 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 $4H_2O$,
813 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
814 K_2CrO_4, and Milli-Q up to 1 L were combined, filter sterilized (0.2 µm), and kept at 4°C. Preparation
815 of TC stock: 20 mg thiamine hydrochloride and 0.1 mL 1 g/L cyanocobalamin stock were mixed
816 in 100 mL Milli-Q. The resulting solution was stored at 4°C. *ΔpTF P. tricornutum* cell line
817 (maintained in the Allen Lab; ***McQuaid et al., 2018***) was additionally supplemented with 200
818 µg/mL nourseothricin (GoldBio, Saint Louis, MO, Catalog #N-500-1). All transconjugant *P.*
819 *tricornutum* cell lines were supplemented with 50 or 100 µg/mL phleomycin (InvivoGen, San
820 Diego, CA, Catalog #ant-ph-10p).
821
822 **Conjugation:** (1) Bacterial donor preparation: Chemically competent pTA-Mob-containing
823 TransforMax™ EPI300™ *E. coli* cells (***Strand et al., 2014***; Lucigen, Middleton, WI, Catalog
824 #EC300110) were transformed via heat shock with sequence-verified pPtPBR1 episomes.
825 Transformants were selected on gentamycin-, carbenicillin-, and tetracycline-containing LB 1%
826 agar plates. 3 mL overnight LB cultures supplemented with antibiotics were inoculated from
827 glycerol stocks. (2) *P. tricornutum* preparation: ~2 x 10^8 *P. tricornutum* cells (in 200 µL) in mid- to
828 late-exponential phase were spread on pre-dried (i.e., plates with lid half open and kept in the
829 laminar flow hood for at least 90 min) ½ L1 1% agar plates with 5% LB and left growing for 1 or 2
830 days. Plates were additionally supplemented with 200 µg/mL nourseothricin for conjugation of
831 *ΔpTF P. tricornutum* cells. (3) Conjugation: Overnight donor bacterial cultures were diluted 1:50
832 in 25 mL LB supplemented with antibiotics, grown at 37°C until OD_{600} 0.8–1, spun down,
833 resuspended in 150 uL Super Optimal broth with Catabolite repression (SOC) medium, and
834 spread as evenly as possible on top of a *P. tricornutum* lawn. Plates with donor-*P. tricornutum*
835 co-culture were first left in dark and 30°C for 90 min, then for 1 or 2 days at standard growth
836 conditions. (4) Selection: Co-culture lawn was scraped off of plates 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 liquid 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 cell line was passaged in 30 mL L1 medium supplemented with 50 or 100 µg/mL phleomycin (and 200 µg/mL nourseothricin in case of *ΔpTF P. tricornutum* cells).

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 WT *P. tricornutum* culture was grown in iron-deplete conditions (L1 medium with 7.5 nM total iron) for two weeks. Cells were then centrifuged, supernatants discarded, pellets flash frozen in liquid nitrogen, and stored at -80°C.

RNA extraction: Direct-zol™ RNA Miniprep Plus RNA extraction kit was used (Zymo Research, Irvine, CA, Catalog #R2070). Briefly, cell 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 estimated with Qubit 2.0 Fluorometer (Thermo Fischer Scientific; measured concentration was 27.6 ng/µL), and samples stored at -80°C.

871 **cDNA synthesis:** cDNA synthesis kit from Thermo Fischer Scientific was used (SuperScript™ III
872 First-Strand Synthesis System, Catalog #18080-051). Briefly, 1 µL total RNA was combined with
873 0.5 µL Oligo$(dT)_{20}$ Primer, 0.5 µL 10 mM dNTP Mix, and 3 µL nuclease-free water followed by 5
874 min and 1 min incubation at 65°C and on ice, respectively. After 5 µL cDNA Synthesis Mix (1 µL
875 10x Reverse Transcription buffer, 2 µL 25 mM $MgCl_2$, 1 µL 0.1 M DTT, 0.5 µL RNase OUT, 0.5
876 µL SuperScript III Reverse Transcriptase) was added, samples were incubated for 50 min and 5
877 min at 50°C and 85°C, respectively, then chilled on ice. Finally, 1 µL RNase H was added and
878 sample kept for 20 min at 37°C. cDNA was stored at -20°C until use.
879
880 **MDY-64 labeling and imaging**
881
882 10 mL of a pTF-mCherry expressing *ΔpTF P. tricornutum* cell line grown in iron-deplete conditions
883 (L1 medium with 7.5 nM total iron) were spun down, supernatant discarded, and pellet
884 resuspended in 50 µL phosphate-buffered saline (PBS) (pH 7.4). 0.5 µL 10 mM MDY-64 stock (in
885 DMSO; Thermo Fischer Scientific, Catalog #Y7536) was added, cells incubated for 10 min at RT,
886 pelleted, and resuspended in 50 µL fresh PBS (pH 7.4). Cell suspension was prepared for imaging
887 as follows: 5 µL was placed between a 1.5 mm microscope slide and a cover slip (this setup
888 applies to all imaging experiments in the study). Imaging conditions: Leica TCS SP5 confocal
889 microscope (Leica Microsystems, Wetzlar, Germany), argon laser strength set to 30%, 458 nm
890 laser line at 50% maximum strength, emission window set to 477–517 nm (for visualizing MDY-
891 64; MDY-64 excitation and emission maxima are 451 nm and 497 nm, respectively), 514 nm laser
892 line at 50% maximum strength, emission window set to 620–640 nm (for visualizing mCherry;
893 mCherry excitation and emission maxima are 587 nm and 610 nm, respectively).
894
895 **Protein expression analyses**
896
897 **pTF-APEX2 detection:** Cell pellets from 8 mL mid- to late-exponential phase WT or
898 transconjugant *P. tricornutum* cultures were resuspended in 150 µL cell lysis buffer (50 mM Tris-
899 HCl, 200 mM NaCl, 1 mM DTT, 1 mM PMSF, pH 8.5) and sonicated for 5 min (30 sec on, 1 min
900 off) with Bioruptor UCD-200TM (Diagenode, Liège, Belgium). The resulting cell lysates were
901 centrifuged, total protein content in supernatants measured with Bradford Assay Kit (Thermo
902 Fischer Scientific, Catalog #23236), and insoluble fractions resuspended in 150 µL cell lysis
903 buffer. 1 µg soluble protein samples and 0.5 µL resuspended insoluble protein fractions were
904 resolved on a NuPage 4-12% Bis-Tris 1.5 mm gel (Thermo Fischer Scientific, Catalog

905 #NP0335BOX), wet transferred to polyvinylidene difluoride (PVDF) membranes (Thermo Fischer
906 Scientific, Catalog #LC2005), and visualized with WesternBreeze™ Chemiluminescent Kit, anti-
907 rabbit (Thermo Fischer Scientific, Catalog #WB7106).
908
909 **pTF-mCherry detection:** Cell pellets from 400 µL mid- to late-exponential phase WT or
910 transconjugant *P. tricornutum* cultures were resuspended in 50 µL cell lysis buffer (50 mM Tris-
911 HCl, 200 mM NaCl, 1 mM DTT, 1 mM PMSF, pH 8.5) and sonicated 15 min (30 sec on, 1 min off)
912 with Bioruptor UCD-200TM. 1 µL of each whole cell lysate was resolved on a NuPage 4-12% Bis-
913 Tris 1.5 mm gel, wet transferred to PVDF membranes, and visualized with WesternBreeze™
914 Chemiluminescent Kit, anti-rabbit.
915
916 pTF antibody details: Amino acid residues 32–223 served as the immunogen (N-terminal pTF
917 region just downstream of the signal peptide). The antibody was produced in a rabbit. 1:10,000
918 dilution of 1.14 mg/mL antibody stock was used in this study for background-free results.
919
920 Protein ladder: MagicMark™ XP Western Protein Standard (Thermo Fischer Scientific, Catalog
921 #LC5602).
922
923 Related resource: Turnšek J. 2017. HA tag enables highly efficient detection of heterologous
924 proteins in *Phaeodactylum tricornutum* (*Pt*) exconjugants. *protocols.io*. DOI:
925 dx.doi.org/10.17504/protocols.io.j7ncrme
926
927 **Amplex UltraRed assay and resorufin imaging**
928
929 **Amplex UltraRed assay:** 5 mL of triplicate WT or pTF-APEX2 expressing *P. tricornutum* cultures
930 in mid- to late-exponential phase were incubated on ice for 5 min, spun down, supernatant
931 discarded, pellet resuspended in 500 µL ice-cold PBS (pH 7.4), and transferred to microcentrifuge
932 tubes. Cells were spun down again, resuspended in 200 µL ice-cold reaction buffer (50 µM
933 Amplex UltraRed (AUR; Thermo Fischer Scientific, Catalog #A36006), 2 mM H_2O_2, in PBS (pH
934 7.4)), and incubated on ice for 15 min unless otherwise noted. 50 µL supernatant was mixed with
935 50 µL PBS (pH 7.4) and resorufin fluorescence measured in a black microtiter plate with black
936 bottom using Flexstation 3 microtiter plate reader (Molecular Devices, San Jose, CA; excitation:
937 544 nm, emission: 590 nm; resorufin excitation and emission maxima are 568 nm and 581 nm,
938 respectively). Horseradish peroxidase (HRP) was always included as a positive assay control.

939 Fluorescence was normalized to OD_{750} of experimental *P. tricornutum* cultures. Amplex UltraRed
940 was prepared as a 10 mM stock in DMSO and stored in 20 µL aliquots at -20°C. 3% (w/w) H_2O_2
941 stock (Sigma-Aldrich, Saint Louis, MO, Catalog #323381-25ML) was stored in 100 µL aliquots at
942 -20°C.
943
944 **Resorufin imaging:** WT and pTF-APEX2 expressing *P. tricornutum* cells after performing the
945 Amplex UltraRed assay were imaged with Leica TCS SP5 confocal microscope using the
946 following parameters: argon laser strength at 30%, 514 nm laser line at 50% maximum strength,
947 resorufin emission window: 575–605 nm, autofluorescence emission window: 700–750 nm.
948
949 **Transmission electron microscopy (TEM)**
950
951 **Part 1: Labeling**
952 5 mL of triplicate WT or pTF-APEX2 expressing *P. tricornutum* cultures in mid- to late-exponential
953 phase were spun down (4,000 rpm, 4°C, 10 min) and fixed in 5 mL ice-cold 2% (w/v)
954 paraformaldehyde (PFA) and 2% (v/v) glutaraldehyde in 0.15 M sodium cacodylate buffer (pH
955 7.4) for 30 min on ice. Cells were rinsed in 5 mL 0.15 M sodium cacodylate buffer (pH 7.4) five
956 times for 3 min on ice, then once again in 5 mL 0.15 M sodium cacodylate buffer (pH 7.4) with 10
957 mM glycine for 3 min on ice. Cells were then treated with 25 mM 3,3'-diaminobenzidine (DAB;
958 Sigma-Aldrich, Catalog #D8001) as follows: 5.36 mg DAB was dissolved in 1 mL 0.1 N HCl and
959 sonicated for 45 min. 5 mL of 0.3 M sodium cacodylate buffer (pH 7.4) was added to dissolved
960 DAB, final volume adjusted to 10 ml with ddH_2O, solution filtered through a 0.22 µm syringe filter,
961 and 3 µL 30% (w/w) H_2O_2 added for 3 mM final concentration. Cells were incubated in 5 mL of
962 this reaction buffer for 15 min on ice, rinsed five times for 3 min in 5 mL 0.15 M sodium cacodylate
963 buffer (pH 7.4), post-fixed in 2 mL 1% osmium tetroxide (OsO_4) (Electron Microscopy Sciences,
964 Hatfield, PA, Catalog #19150) in 0.15 M sodium cacodylate buffer (pH 7.4) for 30 min on ice, and
965 rinsed in 5 mL ice-cold ddH_2O five times for 3 min. Cells were then resuspended in 300 µL melted
966 3% agar, poured onto a glass slide sitting on ice, left to solidify, and cut into small ~3 x 3 x 1 mm
967 pieces which were transferred into scintillation vials with 10 mL ice-cold ddH_2O. All shorter spin
968 down steps were done at 4,000 rpm and 4°C for 1.5 min. All buffers were used ice-cold.
969
970 **Part 2: TEM sample preparation**
971 Agar blocks with embedded *P. tricornutum* cells were fixed in 10 mL 2% (v/v) glutaraldehyde in
972 ddH_2O to crosslink agar, rinsed five times for 2 min in 5 mL ice-cold ddH_2O, and left incubating in

973 5 mL 2% uranyl acetate (Electron Microscopy Sciences, Catalog #22400) overnight at 4°C. Next
974 morning, agar blocks were first dehydrated in the following ethanol series: 20%, 50%, 70%, 90%,
975 100% (on ice, 10 mL), 100%, 100% (at RT, 10 mL), then infiltrated with 10 mL 50% epoxy resin
976 for ~1 h. To prepare 20 mL 100% resin, Durcupan ACM mixture (Electron Microscopy Sciences,
977 Catalog #14040) components were combined as follows: 11.4 g A (epoxy resin), 10 g B (964
978 hardener), 0.3 g C (964 accelerator), and 0.1 g D (dibutyl phthalate) in this exact order (for 50%
979 resin 1 part 100% resin was combined with 1 part 100% ethanol). Agar blocks were transferred
980 into 10 mL 100% resin for ~4 h, then into fresh 10 mL 100% resin overnight, again into fresh 10
981 mL 100% resin for 4 h the following morning before finally being poured into aluminum boats and
982 left to polymerize in 60°C oven for at least 48 h (over the weekend). Polymerized resins were
983 detached from aluminum boats, agar blocks dense with cells cut out and glued to "dummy" blocks
984 by incubation in a 60°C oven for at least 15 min. ~500 nm thick sections were cut with an
985 ultramicrotome (Leica Ultracut), stained with 1% toluidine blue, and observed under light
986 microscope to make sure embedded cells were exposed. ~100 nm thick sections were then cut,
987 placed on TEM grids, labeled, and saved until imaging. Imaging was performed with JEOL JEM-
988 1200 (Japan Electron Optics Laboratory, Akishima, Tokyo, Japan) transmission electron
989 microscope at 80 keV. Backscatter scanning electron microscopy was performed with Zeiss
990 Merlin (Oberkochen, Germany) scanning electron microscope (SEM) at 2 keV by placing ~80 nm
991 thick sections on a silicon wafer and imaged with inverted contrast which gives a TEM-like image.
992

993 **<u>Proximity-dependent proteomic mapping</u>**

994

995 **Part 1: Labeling**

996 25 mL of quintuplicate WT and pTF-APEX2 expressing *P. tricornutum* cultures in mid- to late-
997 exponential phase—cell density of all cultures just prior to harvest was ~2 x 10^7 cells/mL which
998 corresponds to OD_{750} ~0.4—were cooled on ice for 10 min and pelleted (4,000 rpm, 4°C, 10 min).
999 Supernatants were discarded, pellets resuspended in 0.5 mL ice-cold PBS (pH 7.4), transferred
1000 to microcentrifuge tubes, and spun down (4,000 rpm, 4°C, 10 min). Cells were then resuspended
1001 in 0.5 mL ice-cold 1.2 M D-sorbitol in PBS (pH 7.4), supplemented with 2.5 mM biotin-phenol
1002 (Berry & Associates, Dexter, MI, Catalog #BT 1015), incubated on a tube rotator at 4°C for 90
1003 min, supplemented with 1 mM H_2O_2, and incubated on a tube rotator at 4°C for another 20 min.
1004 Labeling reaction was quenched by washing cells twice (4,000 rpm, 4°C, 5 min) with 0.5 mL ice-
1005 cold quenching solution (10 mM sodium ascorbate (VWR International, Radnor, PA, Catalog
1006 #95035-692), 5 mM Trolox (Sigma-Aldrich, Catalog #238813-5G), 10 mM sodium azide (VWR

1007 International, Catalog #AA14314-22) in PBS (pH 7.4)). 50 µL of quenched cell suspension was
1008 saved for a streptavidin blot. The remaining 450 µL was spun down (4,000 rpm, 4°C, 10 min) and
1009 lysed in 250 µL cell lysis buffer (50 mM Tris-HCl, 200 mM NaCl, 1 mM DTT, 1 mM PMSF, pH 8.5)
1010 by sonication for 15 min (30 sec on, 1 min off). Cell lysates were spun down at 4,000 rpm and
1011 4°C for 45 min and protein concentration in supernatants measured using Bradford Assay Kit. For
1012 streptavidin blot, 25 µL of saved quenched cells were first sonicated for 15 min (30 sec on, 1 min
1013 off). 2.5 µL of whole cell lysates and 1 ng biotinylated BSA positive control (Thermo Fischer
1014 Scientific, Catalog #29130) were then resolved on a NuPage 4-12% Bis-Tris 1.5 mm gel and wet
1015 transferred to a PVDF membrane. Membrane was washed twice with 15 mL PBST (PBS (pH 7.4)
1016 with 0.1% (v/v) Tween-20) for 5 min, left blocking overnight at RT and gentle shaking in PBST
1017 supplemented with 5% BSA (Sigma-Aldrich, Catalog #A9647-100G), and washed once the next
1018 morning with 15 mL PBST for 5 min. It was then incubated for 1 h at RT and gentle shaking in 15
1019 mL PBST supplemented with 5% BSA and 1:15,000 HRP-conjugated streptavidin (Thermo
1020 Fischer Scientific, Catalog #S911). Finally, membrane was washed three times with 15 mL PBST
1021 supplemented with 5% BSA for 5 min and once with 15 mL PBST for 10 min after which it was
1022 visualized with SuperSignal™ West Dura Extended Duration reagent (Thermo Fischer Scientific,
1023 Catalog #34075) using C-DiGit® Blot Scanner (Li-COR, Lincoln, NE).
1024
1025 Protein ladder: WesternSure® Pre-stained Chemiluminescent Protein Ladder (HRP-conjugated
1026 protein ladder) (Li-COR, Catalog #926-98000).
1027
1028 **Part 2: Pull-down and mass spectrometry**
1029
1030 Pull-down: 50 µL streptavidin beads (Thermo Fischer Scientific, Catalog #88816) transferred to
1031 microcentrifuge tubes were pelleted in a magnetic rack (tubes were left standing for ~5 min to
1032 pellet fully). Supernatants were discarded and beads were washed twice with 1 mL ice-cold cell
1033 lysis buffer (50 mM Tris-HCl, 200 mM NaCl, 1 mM DTT, 1 mM PMSF, pH 8.5). 360 µg proteins in
1034 500 µL total volume (x µL supernatant from cell lysis with 360 µg proteins and 500-x µL cell lysis
1035 buffer) were incubated on a tube rotator overnight at 4°C. Streptavidin beads were washed to
1036 remove nonspecific binders the next morning as follows: 2x 1 mL cell lysis buffer, 1x 1 mL 1 M
1037 KCl, 1x 1 mL 0.1 M Na_2CO_3 (pH 11.5), 1x 1 mL 2 M urea (pH 8.0) in 10 mM Tris-HCl, 2x 1 mL
1038 cell lysis buffer, 2x 1 mL PBS (pH 7.4). PBS after the final washing step was removed before
1039 storing beads at -80°C. Notes: All streptavidin beads collection steps were 5 min long. All the

1040 washing solutions were ice-cold. Microcentrifuge tubes were either very briefly vortexed (~2 sec)
1041 or tapped by hand between each washing step to promote bead resuspension.
1042
1043 On-bead digestion and TMT labeling: Samples were prepared as previously described
1044 (***Kalocsay, 2019***). Liquid reagents used were HPLC quality grade. Washed beads were
1045 resuspended in 50 µL of 200 mM EPPS (4-(2-Hydroxyethyl)-1-piperazinepropanesulfonic acid)
1046 buffer (pH 8.5) and 2% (v/v) acetonitrile with 1 µL of 2 mg/mL lysil endoproteinase Lys-C stock
1047 solution (FUJIFILM Wako Pure Chemical Corporation, Richmond, VA, Catalog #125-05061),
1048 briefly vortexed, and incubated at 37°C for 3 hours. 50 µL of trypsin stock (Promega, Madison,
1049 WI, Catalog #V5111) diluted 1:100 (v/v) in 200 mM EPPS (pH 8.5) was then added. After mixing,
1050 digests were incubated at 37°C overnight and beads were magnetically removed. Peptides were
1051 then directly labeled as follows: acetonitrile was added to 30% (v/v) concentration and peptides
1052 were labeled with TMT 10-plex reagent (Thermo Fisher Scientific, Catalog #90406) for 1 hour.
1053 Labeling reactions were quenched with hydroxylamine at a final concentration of 0.3% (v/v) for
1054 15 min and 1% of labeled peptides was analyzed for label incorporation efficiency by mass
1055 spectrometry. After quenching, peptide solutions were first acidified with formic acid,
1056 trifluoroacetic acid (TFA) was then added to a concentration of 0.1% (v/v), and peptides were
1057 desalted by acidic C_{18} solid phase extraction (StageTip). Labeled peptides were finally
1058 resuspended in 1% (v/v) formic acid and 3% (v/v) acetonitrile.
1059
1060 Mass spectrometry: Data were collected with a MultiNotch MS3 TMT method (***McAlister et al.,***
1061 ***2014***) using an Orbitrap Lumos mass spectrometer coupled to a Proxeon EASY-nLC 1200 Liquid
1062 Chromatography (LC) system (both Thermo Fisher Scientific). The used capillary column was
1063 packed with C_{18} resin (35 cm length, 100 µm inner diameter, 2.6 µm Accucore matrix (Thermo
1064 Fisher Scientific)). Peptides were separated for 3 or 4 hours over acidic acetonitrile gradients by
1065 LC prior to mass spectrometry (MS) analysis. Data from two 4 hour runs and one 3 hour run were
1066 recorded and combined. After an initial MS^1 scan (Orbitrap analysis; resolution 120,000; mass
1067 range 400−1,400 Th), MS^2 analysis used collision-induced dissociation (CID, CE = 35) with a
1068 maximum ion injection time of 150–300 ms and an isolation window of 0.5 m/z. In order to obtain
1069 quantitative information, MS^3 precursors were fragmented by high-energy collision-induced
1070 dissociation (HCD) and analyzed in the Orbitrap at a resolution of 50,000 at 200 Th. Further details
1071 on LC and MS parameters and settings were described recently (***Paulo et al., 2016***). The mass
1072 spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the
1073 PRIDE partner repository (***Perez-Riverol et al., 2019***) with the dataset identifier PXD018022.

1074 **pTF-mCherry and MS hit-EYFP co-expression and imaging**

1075

1076 Sequence-validated episomes with pTF-mCherry and MS hit-EYFP expression cassettes were
1077 conjugated into WT *P. tricornutum* cells, resulting transconjugants genotyped, and screened for
1078 fluorescence. 5 µL of fluorescent transconjugant cell lines in mid- to late-exponential phase were
1079 imaged with settings that minimized cross-channel bleed-through: argon laser strength at 30%,
1080 514 nm laser line at 50% maximum strength, mCherry emission window: 620–640 nm, EYFP
1081 emission window: 520–540 nm, autofluorescence emission window: 700–750 nm.

1082

1083 **pTF.CREG1 protein expression conditions**

1084

1085 Small scale expression testing for the Δ31_pTF.CREG1-His6 (and all other pTF.CREG1
1086 truncations not described here) was performed as follows: an overnight culture of BL21 *E. coli*
1087 cells carrying the PtpBAD-Δ31_pTF.CREG1-CTHF (=pJT_Δ31_pTF.CREG1-His6) expression
1088 vector was used to inoculate 50 mL of Terrific Broth (TB) supplemented with tetracycline (10
1089 ug/mL). Cultures were grown in a shaking incubator (37°C, 200 rpm) until OD_{600} of 0.4–0.6 was
1090 reached. Incubator temperature was then lowered to 18°C and flasks were allowed to adjust to
1091 this temperature for about 30 min before arabinose was added to a final concentration of 0.5%
1092 (w/v). Growth at 18°C was continued overnight (12–18 hrs) after which 10 mL of cultures were
1093 harvested by centrifugation at 6,000 g for 10 min. Large-scale Δ31_pTF.CREG1-His6 expression
1094 was performed as above, but at 2 L total volume. Following arabinose induction and overnight
1095 growth of the 2 L culture, 500 mL was harvested for further processing.

1096

1097 **pTF.CREG1 purification conditions**

1098

1099 All purifications were performed using cobalt, Co^{2+}, TALON Metal Affinity Resin (Takara Bio,
1100 Catalog #635502) which binds His6-tagged proteins with high affinity. Pellets from small scale
1101 expression testing (10 mL culture) were resuspended in 800 uL of lysis buffer (50 mM NaH_2PO_4
1102 (pH 7.5), 500 mM NaCl, 0.1% (v/v) Triton X-100, 10 mM imidazole, 1 mg/mL lysozyme, and 10
1103 µM β-mercaptoethanol) and subjected to microtip sonication on ice until lysis was complete.
1104 Lysates were then clarified by centrifugation (10 min, 15,000 g, 4°C) and supernatants (700 uL)
1105 were set aside. For each purification, 25 uL of TALON resin was equilibrated by washing three
1106 times with 10 volumes of lysis buffer and pelleting by centrifugation after each wash (30 sec,
1107 3,000 g, RT). After third wash, 25 uL of lysis buffer was added to the resin to make a 50 uL slurry.

1108 Supernatants from previously clarified cell lysates were then added to the TALON resin slurry and
1109 incubated for 1 hr at RT with end-over-end mixing. Following the 1 hr incubation, the resin was
1110 pelleted by centrifugation (30 sec, 3,000 g, RT) and washed three times with 10 volumes of wash
1111 buffer (50 mM NaH_2PO_4 (pH 7.5), 500 mM NaCl, 30 mM imidazole). Proteins were eluted with 50
1112 uL of elution buffer (50 mM NaH_2PO_4 (pH 7.5), 500 mM NaCl, 250 mM imidazole, 10% glycerol)
1113 and were either subjected to immediate SDS-PAGE electrophoresis or stored at -80°C.
1114
1115 Pellets from large-scale expression (500 mL harvested bacterial culture) were processed with a
1116 few modifications. For lysis, cell pellets were resuspended in 20 mL of lysis buffer. For purification,
1117 1 mL of TALON resin (i.e., 2 mL of equilibrated slurry) was used in combination with approx. 19
1118 mL of clarified cell lysate which was allowed to incubate overnight at 8°C with end-over-end
1119 mixing. Final protein elution was performed using 5 mL of elution buffer. Purified protein was
1120 subjected to SDS-PAGE electrophoresis and concentrated using an Amicon Ultra-15 10 kDa
1121 cutoff concentrator (Millipore Sigma, Burlington, MA, Catalog #UFC901024). Following the
1122 concentration step, buffer exchange into 20 mM HEPES (pH 8.0), 10% glycerol, 300 mM KCl was
1123 performed using a PD-10 desalting column (GE Healthcare, Chicago, IL, Catalog #17-0851-01).
1124 Protein was then concentrated again and total protein content was determined by the Bradford
1125 method using the Protein Assay Dye Reagent (Bio-Rad, Hercules, CA, Catalog #5000006).
1126
1127 **Flavin reduction assays**
1128
1129 To test for flavin reductase activity, we set up the following enzyme assay conditions in 100 uL
1130 final volume: 50 mM Tris-HCl (pH 7.5), 100 mM KCl, 10% glycerol, 30 μM flavin mononucleotide
1131 (FMN), 30 μM riboflavin, and 250 μM NADPH (***Coves and Fontecave, 1993***). Assays were
1132 started by the addition of 50 μg of pTF.CREG1, BSA, or Milli-Q water for no-enzyme controls.
1133 Oxidation of NADPH was measured by the decrease in absorbance at 320 nm on a Flexstation 3
1134 microtiter plate reader and the signal for each reaction was tracked over a period of 60 min. All
1135 assays were conducted in triplicate.
1136
1137 **Bioinformatic and data analyses**
1138
1139 **Mass spectrometry data analysis:** Peptide-spectrum matches used a SEQUEST (v. 28, rev.
1140 12) algorithm (***Eng et al., 1994***). Data were searched against a size-sorted forward and reverse
1141 database of the *Phaeodactylum tricornutum* proteome (strain CCAP 1055/1, UniProt reference

1142 proteome UP000000759) with added common contaminant proteins and the pTF-APEX2 fusion
1143 protein sequence. Spectra were first converted to mzXML and searches were then performed
1144 using a mass tolerance of 50 ppm for precursors and a fragment ion tolerance of 0.9 Da. For the
1145 searches, maximally two missed cleavages per peptide were allowed. We searched dynamically
1146 for oxidized methionine residues (+15.9949 Da) and applied a target decoy database strategy. A
1147 false discovery rate (FDR) of 1% was set for peptide-spectrum matches following filtering by linear
1148 discriminant analysis (LDA) (***Beausoleil et al., 2006***; ***Huttlin et al., 2010***). The FDR for final
1149 collapsed proteins was 1%. Quantitative peptide information was derived from MS^3 scans. Quant
1150 tables were generated with the following filter criteria: MS^2 isolation specificity of >70% for each
1151 peptide and a sum of TMT signal-to-noise (s/n) of >200 over all channels per peptide. Quant
1152 tables were exported to Excel and further processed therein. Details of the TMT intensity
1153 quantification method and additional applied search parameters were described previously
1154 (***Paulo et al., 2016***). Scaled proteomic data were subjected to two-way hierarchical clustering
1155 (Ward's method) using JMP software package (SAS Institute). Volcano plot with log_2-transformed
1156 average APEX2/WT ratios and associated P values was made in R using ggplot2 data
1157 visualization package (***R Core Team, 2013***). Gene IDs corresponding to protein hits were inferred
1158 using Ensembl Protists Phatr3 *P. tricornutum* genomic database.
1159
1160 **Protein feature identification:** Protein lengths, molecular weights, and isoelectric points (pI)
1161 were determined with ProtParam (***Gasteiger et al., 2005***). Signal peptides and transmembrane
1162 regions were identified with SignalP 4.1 (***Petersen et al., 2011***) and TMHMM Server v. 2.0 (***Krogh***
1163 ***et al., 2001***), respectively. Protein localizations were predicted with a combination of tools:
1164 TargetP 1.1 (***Emanuelsson et al., 2000***), SignalP 4.1, and ASAFind (***Gruber et al., 2015***) version
1165 1.1.7. All putative chloroplastic localizations in the study mean that a protein was predicted to be
1166 chloroplastic by ASAFind (state-of-the-art plastidial protein localization prediction tool for
1167 diatoms); chloroplast localization prediction confidences are noted. Low confidence prediction by
1168 ASAFind means that a protein satisfies the following filtering criteria: (1) it contains a signal
1169 peptide as detected by SignalP 4.1, (2) +1 position of ASAFind predicted cleavage site is
1170 phenylalanine (F), tryptophan (W), tyrosine (Y), or leucine (L) (making the protein "potentially
1171 plastid targeted"; it was tyrosine in pTF.CREG1), and (3) one or both of the following is false: the
1172 ASAFind predicted cleavage site coincides with the SignalP 4.1 prediction and the transit peptide
1173 score is higher than 2 (the latter is false for pTF.CREG1: transit peptide score was ~0.81).
1174 Peroxidase class prediction was done in RedoxiBase (***Savelli et al., 2019***). N-glycosylation and

1175 GalNAc-type O-glycosylation predictions were performed with NetNGlyc 1.0 (***Gupta et al., 2004***)
1176 and NetOGlyc 4.0 (***Steentoft et al., 2013***), respectively.
1177
1178 **Amino acid sequence alignments:** APEX2-like *P. tricornutum* peroxidases were aligned with
1179 Clustal Omega (***Sievers et al., 2011***) using the default settings and CLC Sequence Viewer 7.7
1180 (QUIAGEN) using the following parameters: Gap open cost: 10.0, Gap extension cost: 1.0, End
1181 gap cost: As any other, Alignment: Very accurate (slow). Putative substrate-binding loops in these
1182 peroxidases were evaluated using the Clustal Omega alignment. Conserved motifs in amino acid
1183 sequence alignments underlying the phylogenetic trees were displayed in CLC Sequence Viewer
1184 7.7.
1185
1186 **Phylogenetic analyses:** Homologs of respective *P. tricornutum* proteins were retrieved from the
1187 non-redundant National Center for Biotechnology Information (NCBI) and the Marine Microbial
1188 Eukaryote Transcriptome Sequencing Project (MMETSP) databases using the BLASTP algorithm
1189 (e-value threshold set to e^{-15}) (***Altschul et al., 1990***). The BLAST search retrieved only a handful
1190 of homologs for pTF.ap1 and no homologs for pTF.CatCh1 even among closely related diatoms.
1191 Therefore, more sensitive HMMER (***Eddy, 1998***) searches were employed to extend the datasets,
1192 which were afterward aligned using the Localpair algorithm as implemented in MAFFT (***Katoh et***
1193 ***al., 2002***). Ambiguously aligned regions, regions composed mostly of gaps, and short fragments
1194 were manually removed in SeaView 4 (***Gouy et al., 2010***). For each alignment, the maximum
1195 likelihood analysis was carried out in IQ-TREE (***Nguyen et al., 2015***) under the best-fitting
1196 substitution matrix as inferred by the built-in model finder. Branching support was estimated using
1197 "thorough" non-parametric bootstrap analysis from 500 replicates in IQ-TREE.
1198
1199 **Identification of disordered protein regions:** pTF.CatCh1 amino acid sequence was analyzed
1200 for the presence of disordered protein regions with PONDR (Predictor of Naturally Disordered
1201 Regions) VSL2 predictor (***Peng et al., 2006***) and IUPred2a long disorder prediction type
1202 (***Mészáros et al., 2018***).
1203
1204 **Homology modeling of pTF.CREG1:** Full-length pTF.CREG1 amino acid sequence was used
1205 as an input for Phyre2, an online protein structure prediction server (***Kelley et al., 2015***). Normal
1206 modeling mode was selected. Human CREG1 (PDB ID: 1XHN) was identified as the best
1207 modeling template (human CREG1 is ~37% identical to pTF.CREG1 as inferred from their Clustal
1208 Omega alignment). pTF.CREG1 was modeled with >90% confidence across 182 modeled amino

1209 acid residues (Pro43–Lys224) indicating the predicted structure is representative of the actual
1210 structure. pTF.CREG1 homology model was aligned with human CREG1 monomer in UCSF
1211 Chimera version 1.11.1 using MatchMaker tool with default settings (***Pettersen et al., 2004***)
1212
1213 **Other data analyses:** Amplex UltraRed assay and transcriptomic data were plotted in R using
1214 ggplot2 data visualization package (***R Core Team, 2013***). Flavin reduction assay data were
1215 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 H. Coale, and Sarah R. 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), the United States Department of Energy Genomics Science program grant DE-SC0018344 (AEA), the Czech Science Foundation grant 18-13458S (MO), and ERDF/ESF Centre for Research of Pathogenicity and Virulence of Parasites grant No.CZ.02.1.01/0.0/0.0/16_019/0000759 (MO). 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).

1232 ## Additional information

1233

1234 ### Funding

1235

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
The Czech Science Foundation	18-13458S	Miroslav Oborník
ERDF/ESF Centre for Research of Pathogenicity and Virulence of Parasites	No.CZ.02.1.01/0.0/0.0/ 16_019/0000759	Miroslav Oborník
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.

1236

1237 *Jernej Turnšek originated and proposed the idea to implement the use of APEX2 in model marine diatoms

1238 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
Maria del Pilar Martinez Viedma	https://orcid.org/0000-0002-8885-5008
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

1246 ## Additional files

1247

1248 ### Supporting files

1249

1250 **Supplemental file 1.**

1251 **Table S1:** Predicted endogenous biotinylated proteins are present at similar levels in WT and
1252 pTF-APEX2 proteomic samples. **Table S2:** Features of the three proteins co-expressed from a
1253 gene cluster on chromosome 20. **Table S3:** Molecular cloning primers. **Table S4:** Constructed
1254 *Phaeodactylum tricornutum* episomes and *Escherichia coli* vectors with expression cassette
1255 details and fusion protein amino acid sequences.

1256

1257 **Key Resources Table.**

1258

1259 The quantitative mass spectrometry proteomics data used to generate ***Figure 4—source data 1***
1260 and ***Figure 4—figure supplement 1—source data 1*** have been deposited to the
1261 ProteomeXchange Consortium via the PRIDE partner repository (***Perez-Riverol et al., 2019***) with
1262 the dataset identifier PXD018022. Data can be accessed via a reviewer account using Username:
1263 reviewer15941@ebi.ac.uk and Password: ZCd0fJ8p.

1264

1265 The amino acid sequence alignment LINSI files ***Figure 6—source data 1***, ***Figure 7—source***
1266 ***data 1***, and ***Figure 7—figure supplement 2—source data 1*** can be read and interrogated with
1267 AlignmentViewer available at https://alignmentviewer.org/.

1268

1269 ### Data availability

1270

1271 All data generated or analyzed during this study are included in the manuscript and supporting
1272 files.

1273

1274 ## Competing interests

1275

1276 The authors declare that no competing interests exist.

References

Aguirre JD, Clark HM, McIlvin M, Vazquez C, Palmere SL, Grab DJ, Seshu J, Hart PJ, Saito M, Culotta VC. 2013. A Manganese-rich Environment Supports Superoxide Dismutase Activity in a Lyme Disease Pathogen, *Borrelia burgdorferi*. *J Biol Chem* **288**:8468–8478. DOI: https://doi.org/10.1074/jbc.M112.433540, PMID: 23376276

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

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

Banci L, Bertini I, Ciofi-Baffoni S, Finney LA, Outten CE, O'Halloran TV. 2002. A New Zinc-protein Coordination Site in Intracellular Metal Trafficking: Solution Structure of the Apo and Zn(II) forms of ZntA(46-118). *J Mol Biol* **323**:883–897. DOI: https://doi.org/10.1016/s0022-2836(02)01007-0, PMID: 12417201

Banci L, Bertini I, Ciofi-Baffoni S, Su X-C, Miras R, Bal N, Mintz E, Catty P, Shokes JE, Scott RA. 2006. Structural basis for metal binding specificity: the N-terminal cadmium binding domain of the P1-type ATPase CadA. *J Mol Biol* **356**:638–650. DOI: https://doi.org/10.1016/j.jmb.2005.11.055, PMID: 16388822

Bard F, Chia J. 2016. Cracking the Glycome Encoder: Signaling, Trafficking, and Glycosylation. *Trends Cell Biol* **26**:379–388. DOI: https://doi.org/10.1016/j.tcb.2015.12.004, PMID: 26832820

Barrett J, Girr P, Mackinder LCM. 2021. Pyrenoids: CO_2-fixing phase separated liquid organelles. *Biochim Biophys Acta Mol Cell Res* **1868**:118949. DOI: https://doi.org/10.1016/j.bbamcr.2021.118949, PMID: 33421532

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

Beausoleil SA, Villén J, Gerber SA, Rush J, Gygi SP. 2006. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* **24**:1285–1292. DOI: https://doi.org/10.1038/nbt1240, PMID: 16964243

Behnke J, LaRoche J. 2020. Iron uptake proteins in algae and the role of Iron Starvation-Induced Proteins (ISIPs). *Eur J Phycol* **55**:339–360. DOI: https://doi.org/10.1080/09670262.2020.1744039

Benham AM. 2012. Protein Secretion and the Endoplasmic Reticulum. *Cold Spring Harb Perspect Biol* **4**:a012872. DOI: https://doi.org/10.1101/cshperspect.a012872, PMID: 22700933

Benoiston A-S, Ibarbalz FM, Bittner L, Guidi L, Jahn O, Dutkiewicz S, Bowler C. 2017. The evolution of diatoms and their biogeochemical functions. *Philos Trans R Soc Lond B Biol Sci* **372**:20160397. DOI: https://doi.org/10.1098/rstb.2016.0397, PMID: 28717023

Bertrand EM, McCrow JP, Moustafa A, Zheng H, McQuaid JB, Delmont TO, Post AF, Sipler RE, Spackeen JL, Xu K, Bronk DA, Hutchins DA, Allen AE. 2015. Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc Natl Acad Sci U S A* **112**:9938–9943. DOI: https://doi.org/10.1073/pnas.1501615112, PMID: 26221022

Bhaya D, Grossman AR. 1993. Characterization of gene clusters encoding the fucoxanthin chlorophyll proteins of the diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res* **21**:4458–4466. DOI: https://doi.org/10.1093/nar/21.19.4458, PMID: 8233779

Blaby-Haas CE, Merchant SS. 2012. The ins and outs of algal metal transport. *Biochim Biophys Acta* **1823**:1531–1552. DOI: https://doi.org/10.1016/j.bbamcr.2012.04.010, PMID: 22569643

Blaby-Haas CE, Merchant SS. 2014. Lysosome-related Organelles as Mediators of Metal Homeostasis. *J Biol Chem* **289**:28129–28136. DOI: https://doi.org/10.1074/jbc.R114.592618, PMID: 25160625

Blaby-Haas CE, Padilla-Benavides T, Stübe R, Argüello JM, Merchant SS. 2014. Evolution of a plant-specific copper chaperone family for chloroplast copper homeostasis. *Proc Natl Acad Sci U S A* **111**:E5480–E5487. DOI: https://doi.org/10.1073/pnas.1421545111, PMID: 25468978

Bogdan AR, Miyazawa M, Hashimoto K, Tsuji Y. 2016. Regulators of Iron Homeostasis: New Players in Metabolism, Cell Death, and Disease. *Trends Biochem Sci* **41**:274–286. DOI: https://doi.org/10.1016/j.tibs.2015.11.012, PMID: 26725301

Boiteau RM, Mende DR, Hawco NJ, McIlvin MR, Fitzsimmons JN, Saito MA, Sedwick PN, DeLong EF, Repeta DJ. 2016. Siderophore-based microbial adaptations to iron scarcity across the eastern Pacific Ocean. *Proc Natl Acad Sci U S A* **113**:14237–14242. DOI: https://doi.org/10.1073/pnas.1608594113, PMID: 27911777

Bonfio C, Valer L, Scintilla S, Shah S, Evans DJ, Jin L, Szostak JW, Sasselov DD, Sutherland JD, Mansy SS. 2017. UV-light-driven prebiotic synthesis of iron-sulfur clusters. *Nat Chem* **9**:1229–1234. DOI: https://doi.org/10.1038/nchem.2817, PMID: 29168482

Boukouris AE, Zervopoulos SD, Michelakis ED. 2016. Metabolic Enzymes Moonlighting in the Nucleus: Metabolic Regulation of Gene Transcription. *Trends Biochem Sci* **41**:712–730. DOI: https://doi.org/10.1016/j.tibs.2016.05.013, PMID: 27345518

Bowler C, (75 authors), Grigoriev IV. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**:239–244. DOI: https://doi.org/10.1038/nature07410, PMID: 18923393

Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, Feldman JL, Perrimon N, Ting AY. 2018. Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* **36**:880–887. DOI: https://doi.org/10.1038/nbt.4201, PMID: 30125270

Brenner MP. 2012. Endocytic Traffic: Vesicle Fusion Cascade in the Early Endosomes. *Curr Biol*. **22**:R597–598. DOI: https://doi.org/10.1016/j.cub.2012.06.009, PMID: 22877778

Brunson JK, McKinnie SMK, Chekan JR, McCrow JP, Miles ZD, Bertrand EM, Bielinski VA, Luhavaya H, Oborník M, Smith GJ, Hutchins DA, Allen AE, Moore BS. 2018. Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. *Science* **361**:1356–1358. DOI: https://doi.org/10.1126/science.aau0382, PMID: 30262498

Brzezinski MA, Krause JW, Bundy RM, Barbeau KA, Franks P, Goericke R, Landry MR, Stukel MR. 2015. Enhanced silica ballasting from iron stress sustains carbon export in a frontal zone within the California Current. *J Geophys Res C: Oceans* **120**:4654–4669. DOI: https://doi.org/10.1002/2015JC010829

Camacho A, Walter XA, Picazo A, Zopfi J. 2017. Photoferrotrophy: Remains of an Ancient Photosynthesis in Modern Environments. *Front Microbiol* **8**:323. DOI: https://doi.org/10.3389/fmicb.2017.00323, PMID: 28377745

Cao H, Schroeder B, Chen J, Schott MB, McNiven MA. 2016. The Endocytic Fate of the Transferrin Receptor Is Regulated by c-Abl Kinase. *J Biol Chem* **291**:16424–16437. DOI: https://doi.org/10.1074/jbc.M116.724997, PMID: 27226592

Caron DA, Alexander H, Allen AE, Archibald JM, Armbrust EV, Bachy C, Bell CJ, Bharti A, Dyhrman ST, Guida SM, Heidelberg KB, Kaye JZ, Metzner J, Smith SR, Worden AZ. 2017. Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nat Rev Microbiol* **15**:6–20. DOI: https://doi.org/10.1038/nrmicro.2016.160, PMID: 27867198

Carradec Q, (37 authors), Wincker P. 2018. A global ocean atlas of eukaryotic genes. *Nat Commun* **9**:373. DOI: https://doi.org/10.1038/s41467-017-02342-1, PMID: 29371626

Cheng Y, Zak O, Aisen P, Harrison SC, Walz T. 2004. Structure of the Human Transferrin Receptor-Transferrin Complex. *Cell* **116**:565–576. DOI: https://doi.org/10.1016/s0092-8674(04)00130-8, PMID: 14980223

Cieśla M, Mierzejewska J, Adamczyk M, Farrants A-KÖ, Boguta M. 2014. Fructose bisphosphate aldolase is involved in the control of RNA polymerase III-directed transcription. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1843**:1103–1110. DOI: https://10.1016/j.bbamcr.2014.02.007, PMID: 24576411

Coale TH, Mooseburner M, Horák A., Oborník M., Barbeau KA, Allen AE. 2019. Reduction-dependent siderophore assimilation in a model pennate diatom. *Proc Natl Acad Sci U S A*. **116**:23609–23617. DOI: https://doi.org/10.1073/pnas.1907234116, PMID: 31685631

Cohen NR, A. Ellis K, Burns WG, Lampe RH, Schuback N, Johnson Z, Sañudo-Wilhelmy S, Marchetti A. 2017. Iron and vitamin interactions in marine diatom isolates and natural assemblages of the Northeast Pacific Ocean: Iron and vitamin interactions in diatoms. *Limnol Oceanogr* **62**:2076–2096. DOI: https://doi.org/10.1002/lno.10552

Cohen NR, Ellis KA, Lampe RH, McNair H, Twining BS, Maldonado MT, Brzezinski MA, Kuzminov FI, Thamatrakoln K, Till CP, Bruland KW, Sunda WG, Bargu S, Marchetti A. 2017. Diatom Transcriptional and Physiological Responses to Changes in Iron Bioavailability across Ocean Provinces. *Frontiers in Marine Science* **4**:360. DOI: https://doi.org/10.3389/fmars.2017.00360

Cohen NR, Gong W, Moran DM, Mcilvin MR, Saito MA, Marchetti A. 2018. Transcriptomic and proteomic responses of the oceanic diatom *Pseudo-nitzschia granii* to iron limitation. *Environ Microbiol* **20**:3109–3126. DOI: https://doi.org/10.1111/1462-2920.14386, PMID: 30117243

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, Inc., 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

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**:e47192. DOI: https://doi.org/10.15252/embr.201847192, PMID: 30610016

Di Bacco A, Gill G. 2003. The secreted glycoprotein CREG inhibits cell growth dependent on the mannose-6-phosphate/insulin-like growth factor II receptor. *Oncogene* **22**:5436–5445. DOI: https:// doi.org/10.1038/sj.onc.1206670, PMID: 12934103

1475 **Diner RE**, Bielinski VA, Dupont CL, Allen AE, Weyman PD. 2016. Refinement of the Diatom
1476 Episome Maintenance Sequence and Improvement of Conjugation-Based DNA Delivery
1477 Methods. *Front Bioeng Biotechnol* **4**:65. DOI: https:// doi.org/10.3389/fbioe.2016.00065, PMID:
1478 27551676
1479
1480 **Diner RE**, Noddings CM, Lian NC, Kang AK, McQuaid JB, Jablanovic J, Espinoza JL, Nguyen
1481 NA, Anzelmatti MA Jr, Jansson J, Bielinski VA, Karas BJ, Dupont CL, Allen AE, Weyman PD.
1482 2017. Diatom centromeres suggest a mechanism for nuclear DNA acquisition. *Proc Natl Acad*
1483 *Sci U S A* **114**:E6015–E6024. DOI: https:// doi.org/10.1073/pnas.1700764114, PMID: 28673987
1484
1485 **Dokmanić I**, Sikić M, Tomić S. 2008. Metals in proteins: correlation between the metal-ion type,
1486 coordination number and the amino-acid residues involved in the coordination. *Acta Crystallogr*
1487 *D Biol Crystallogr* **64**:257–263. DOI: https:// doi.org/10.1107/S090744490706595X, PMID:
1488 18323620
1489
1490 **Durkin CA**, Marchetti A, Bender SJ, Truong T, Morales R, Mock T, Armbrust EV. 2012.
1491 Frustule-related gene transcription and the influence of diatom community composition on silica
1492 precipitation in an iron-limited environment. *Limnol Oceanogr* **57**:1619–1633. DOI:
1493 https://doi.org/10.4319/lo.2012.57.6.1619
1494
1495 **Eckenroth BE**, Steere AN, Chasteen ND, Everse SJ, Mason AB. 2011. How the binding of
1496 human transferrin primes the transferrin receptor potentiating iron release at endosomal pH.
1497 *Proc Natl Acad Sci U S A* **108**:13089–13094. DOI: https://doi.org/10.1073/pnas.1105786108,
1498 PMID: 21788477
1499
1500 **Eddy SR**. 1998. Profile hidden Markov models. *Bioinformatics* **14**:755–763. DOI:
1501 https://doi.org/10.1093/bioinformatics/14.9.755, PMID: 9918945
1502
1503 **El-Gebali S**, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ,
1504 Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn
1505 RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res* **47**:D427–D432. DOI:
1506 https://doi.org/10.1093/nar/gky995, PMID: 30357350
1507
1508 **Emanuelsson O**, Nielsen H, Brunak S, von Heijne G. 2000. Predicting Subcellular Localization
1509 of Proteins Based on their N-terminal Amino Acid Sequence. *J Mol Biol* **300**:1005–1016. DOI:
1510 https://doi.org/10.1006/jmbi.2000.3903, PMID: 10891285
1511
1512 **Eng JK**, McCormack AL, Yates JR. 1994. An approach to correlate tandem mass spectral data
1513 of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**:976–
1514 989. DOI: https://doi.org/10.1016/1044-0305(94)80016-2, PMID: 24226387
1515
1516 **Faktorová D**, (104 authors), Lukeš, J. 2020. Genetic tool development in marine protists:
1517 Emerging model organisms for experimental cell biology. *Nat Methods* **17**:481–494. DOI:
1518 https://doi.org/10.1038/s41592-020-0796-x, PMID: 32251396
1519
1520 **Falciatore A**, Jaubert M, Bouly J-P, Bailleul B, Mock T. 2020. Diatom Molecular Research
1521 Comes of Age: Model Species for Studying Phytoplankton Biology and Diversity. *Plant Cell*
1522 **32**:547–572. DOI: https://doi.org/10.1105/tpc.19.00158, PMID: 31852772
1523
1524 **Finbow ME**, Harrison MA. 1997. The vacuolar H^+-ATPase: a universal proton pump of
1525 eukaryotes. *Biochem J* **324**:697–712. DOI: https://doi.org/10.1042/bj3240697, PMID: 9210392

Fomenko DE, Gladyshev VN. 2003. Identity and Functions of CxxC-Derived Motifs. *Biochemistry* **42**:11214–11225. DOI: https://doi.org/10.1021/bi034459s, PMID: 14503871

Fomenko DE, Marino SM, Gladyshev VN. 2008. Functional Diversity of Cysteine Residues in Proteins and Unique Features of Catalytic Redox-Active Cysteines in Thiol Oxidoreductases. *Mol Cells* **26**:228–235. PMID: 18648218

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 MR, Appel RD, Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. (In:) The Proteomics Protocols Handbook, (Editor:) John M. Walker. Humana Press, Inc., 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

Ghosh P, Dahms NM, Kornfeld S. 2003. Mannose 6-phosphate receptors: new twists in the tale. *Nat Rev Mol Cell Biol* **4**:202–212. DOI: https://doi.org/10.1038/nrm1050, PMID: 12612639

Gibbs SP. 1981. The Chloroplast Endoplasmic Reticulum: Structure, Function, and Evolutionary Significance (In:) International Review of Cytology, (Editors:) Bourne GH, Danielli JF, Jeon KW. Academic Press, Inc., New York, NY, USA. 49–99. DOI: https://doi.org/10.1016/S0074-7696(08)61194-8

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

Gingras A-C, Abe KT, Raught B. 2019. Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles. *Curr Opin Chem Biol* **48**:44–54. DOI: https://10.1016/j.cbpa.2018.10.017, PMID: 30458335

Gizak A, Wiśniewski J, Heron P, Mamczur P, Sygusch J, Rakus D. 2019. Targeting a moonlighting function of aldolase induces apoptosis in cancer cells. *Cell Death Dis* **10**:712. DOI: https://doi.org/10.1038/s41419-019-1968-4, PMID: 31558701

Gladyshev VN, Kryukov GV, Fomenko DE, Hatfield DL. 2004. Identification of Trace Element-Containing Proteins in Genomic Databases. *Annu Rev Nutr* **24**:579–596. DOI: https://doi.org/10.1146/annurev.nutr.24.012003.132241, PMID: 15189132

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

Grant CR, Komeili A. 2020. Ferrosomes are iron storage organelles formed by broadly conserved gene clusters in bacteria and archaea. *bioRxiv*. DOI: https://doi.org/10.1101/2020.01.10.902569

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

Grosche C, Hempel F, Bolte K, Zauner S, Maier UG. 2014. The periplastidal compartment: a naturally minimized eukaryotic cytoplasm. *Curr Opin Microbiol* **22**:88–93. DOI: https://doi.org/10.1016/j.mib.2014.09.017, PMID: 25460801

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

Gupta R, Jung E., Brunak S. 2004. Prediction of N-glycosylation sites in human proteins. In preparation. URL: https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0

Hamdi A, Roshan TM, Kahawita TM, Mason AB, Sheftel AD, Ponka P. 2016. Erythroid cell mitochondria receive endosomal iron by a “kiss-and-run” mechanism. *Biochim Biophys Acta* **1863**:2859–2867. DOI: https://doi.org/10.1016/j.bbamcr.2016.09.008, PMID: 27627839

Han Y, Cui J, Tao J, Guo L, Guo P, Sun M, Kang J, Zhang X, Yan C, Li S. 2009. CREG inhibits migration of human vascular smooth muscle cells by mediating IGF-II endocytosis. *Exp Cell Res* **315**:3301–3311. DOI: https://doi.org/10.1016/j.yexcr.2009.09.013, PMID: 19769965

Han Y, Luan B, Sun M, Guo L, Guo P, Tao J, Deng J, Wu G, Liu S, Yan C, Li S. 2011. Glycosylation-independent binding to extracellular domains 11-13 of mannose-6-phosphate/insulin-like growth factor-2 receptor mediates the effects of soluble CREG on the phenotypic modulation of vascular smooth muscle cells. *J Mol Cell Cardiol* **50**:723–730. DOI: https://doi.org/10.1016/j.yjmcc.2010.12.013, PMID: 21195083

Hertle AP, García-Cerdán JG, Armbruster U, Shih R, Lee JJ, Wong W, Niyogi KK. 2020. A Sec14 domain protein is required for photoautotrophic growth and chloroplast vesicle formation in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **117**:9101–9111. DOI: https://doi.org/10.1073/pnas.1916946117, PMID: 32245810

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

Jensen EC. 2012. Use of fluorescent probes: their effect on cell biology and limitations. *Anat Rec* **295**:2031–2036. DOI: https://doi.org/10.1002/ar.22602, PMID: 23060362

Jeong J, Cohu C, Kerkeb L, Pilon M, Connolly EL, Guerinot ML. 2008. Chloroplast Fe(III) chelate reductase activity is essential for seedling viability under iron limiting conditions. *Proc Natl Acad Sci U S A* **105**:10619–10624. DOI: https://doi.org/10.1073/pnas.0708367105, PMID: 18647837

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, Inc., 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

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**:845–858. DOI: https://doi.org/10.1038/nprot.2015.053, PMID: 25950237

Kilian O, Kroth PG. 2005. Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. *Plant J* **41**:175–183. DOI: https://doi.org/10.1111/j.1365-313X.2004.02294.x, PMID: 15634195

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 of TurboID-mediated proximity labeling for mapping a GSK3 kinase signaling network in *Arabidopsis*. *bioRxiv*. DOI: https://doi.org/10.1101/636324

Knoll AH, Bergmann KD, Strauss JV. 2016. Life: the first two billion years. *Philos Trans R Soc Lond B Biol Sci* **371**:20150493. DOI: https://doi.org/10.1098/rstb.2015.0493, PMID: 27672146

Kominek J, Doering DT, Opulente DA, Shen X-X, Zhou X, DeVirgilio J, Hulfachor AB, Groenewald M, Mcgee MA, Karlen SD, Kurtzman CP, Rokas A, Hittinger CT. 2019. Eukaryotic Acquisition of a Bacterial Operon. *Cell* **176**:1356–1366.e10. DOI: https://doi.org/10.1016/j.cell.2019.01.034, PMID: 30799038

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**:567–580. DOI: https://doi.org/10.1006/jmbi.2000.4315, PMID: 11152613

Kustka AB, Allen AE, Morel FMM. 2007. SEQUENCE ANALYSIS AND TRANSCRIPTIONAL REGULATION OF IRON ACQUISITION GENES IN TWO MARINE DIATOMS. *J Phycol* **43**:715–729. DOI: https://doi.org/10.1111/j.1529-8817.2007.00359.x

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

Lampe RH, Mann EL, Cohen NR, Till CP, Thamatrakoln K, Brzezinski MA, Bruland KW, Twining BS, Marchetti A. 2018. Different iron storage strategies among bloom-forming diatoms. *Proc Natl Acad Sci U S A* **115**:E12275–E12284. DOI: https://doi.org/10.1073/pnas.1805243115, PMID: 30538208

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

Lindquist E, Aronsson H. 2018. Chloroplast vesicle transport. *Photosynth Res* **138**:361–371. DOI: https://doi.org/10.1007/s11120-018-0566-0, PMID: 30117121

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* **8**:e47864. 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

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

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

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, 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

Musayev FN, Di Salvo ML, Ko T-P, Schirch V, Safo MK. 2003. Structure and properties of recombinant human pyridoxine 5'-phosphate oxidase. *Protein Sci* **12**:1455–1463. DOI: https://doi.org/10.1110/ps.0356203, PMID: 12824491

Naslavsky N, Caplan S. 2018. The enigmatic endosome – sorting the ins and outs of endocytic trafficking. *J Cell Sci* **131**:jcs216499. DOI: https://doi.org/10.1242/jcs.216499, PMID: 29980602

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**:268–274. DOI: https://doi.org/10.1093/molbev/msu300, PMID: 25371430

Nützmann H-W, Scazzocchio C, Osbourn A. 2018. Metabolic Gene Clusters in Eukaryotes. *Annu Rev Genet* **52**:159–183. DOI: https://doi.org/10.1146/annurev-genet-120417-031237, PMID: 30183405

Oborník M, Füssy Z. 2018. Complex Endosymbioses I: From Primary to Complex Plastids, Multiple Independent Events. *(*In:) Plastids. Methods in Molecular Biology, (Editor:) Eric Maréchal. Humana Press, Inc., Totowa, NJ, USA. **1829**:17–35. DOI: https://doi.org/10.1007/978-1-4939-8654-5_2

Och LM, Shields-Zhou GA. 2012. The Neoproterozoic oxygenation event: Environmental perturbations and biogeochemical cycling. *Earth-Sci Rev* **110**:26–57. DOI: https://doi.org/10.1016/j.earscirev.2011.09.004

Ohgami RS, Campagna DR, Greer EL, Antiochos B, McDonald A, Chen J, Sharp JJ, Fujiwara Y, Barker JE, Fleming MD. 2005. Identification of a ferrireductase required for efficient transferrin-dependent iron uptake in erythroid cells. *Nat Genet* **37**:1264–1269. DOI: https://doi.org/10.1038/ng1658, PMID: 16227996

O'Halloran TV, Culotta VC. 2000. Metallochaperones, an Intracellular Shuttle Service for Metal Ions. *J Biol Chem* **275**:25057–25060. DOI: https://doi.org/10.1074/jbc.R000006200, PMID: 10816601

Osbourn AE, Field B. 2009. Operons. *Cell Mol Life Sci* **66**:3755–3775. DOI: https://doi.org/10.1007/s00018-009-0114-3, PMID: 19662496

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

Peschke M, Moog D, Klingl A, Maier UG, Hempel F. 2013. Evidence for glycoprotein transport into complex plastids. *Proc Natl Acad Sci U S A* **110**:10860–10865. DOI: https://doi.org/10.1073/pnas.1301945110, PMID: 23754425

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

Perez-Riverol Y, (21 authors), Vizcaíno JA. 2019. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**:D442–D450. DOI: https://doi.org/10.1093/nar/gky1106, PMID: 30395289

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

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**:1605–1612. DOI: https://doi.org/10.1002/jcc.20084, PMID: 15264254

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

Prihoda J, Tanaka A, de Paula WBM, Allen JF, Tirichine L, Bowler C. 2012. Chloroplast-mitochondria cross-talk in diatoms. *J Exp Bot* **63**:1543–1557. DOI: https://doi.org/10.1093/jxb/err441, PMID: 22268145

Progida C, Bakke O. 2016. Bidirectional traffic between the Golgi and the endosomes - machineries and regulation. *J Cell Sci* **129**:3971–3982. DOI: https://doi.org/10.1242/jcs.185702, PMID: 27802132

Przybyla-Toscano J, Roland M, Gaymard F, Couturier J, Rouhier N. 2018. Roles and maturation of iron-sulfur proteins in plastids. *J Biol Inorg Chem* **23**:545–566. DOI: https://doi.org/10.1007/s00775-018-1532-1, PMID: 29349662

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/

Radhamony RN, Theg SM. 2006. Evidence for an ER to Golgi to chloroplast protein transport pathway. *Trends Cell Biol* **16**:385–387. DOI: https://doi.org/10.1016/j.tcb.2006.06.003, PMID: 16815014

Rao RSP, Wollenweber B. 2010. Do N-glycoproteins have preference for specific sequons? *Bioinformation* **5**:208–212. DOI: https://doi.org/10.6026/97320630005208, PMID: 21364799

Raven EL. 2003. Understanding functional diversity and substrate specificity in haem peroxidases: what can we learn from ascorbate peroxidase? *Nat Prod Rep* **20**:367–381. DOI: https://doi.org/10.1039/b210426c, PMID: 12964833

Ree R, Varland S, Arnesen T. 2018. Spotlight on protein N-terminal acetylation. *Exp Mol Med* **50**:90. DOI: https://doi.org/10.1038/s12276-018-0116-z, PMID: 30054468

Rhee H-W, Zou P, Udeshi ND, Martell JD, Mootha VK, Carr SA, Ting AY. 2013. Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science* **339**:1328–1331. DOI: https://doi.org/10.1126/science.1230593, PMID: 23371551

Rochaix J-D. 2011. Regulation of photosynthetic electron transport. *Biochim Biophys Acta* **1807**:375–383. DOI: https://doi.org/10.1016/j.bbabio.2010.11.010, PMID: 21118674

Rosenzweig AC. 2002. Metallochaperones: Bind and Deliver Minireview. *Chem Biol* **9**:673–677. DOI: https://doi.org/10.1016/S1074-5521(02)00156-4, PMID: 12079778

Sacher M, Di Bacco A, Lunin VV, Ye Z, Wagner J, Gill G, Cygler M. 2005. The crystal structure of CREG, a secreted glycoprotein involved in cellular growth and differentiation. *Proc Natl Acad Sci U S A* **102**:18326–18331. DOI: https://doi.org/10.1073/pnas.0505071102, PMID: 16344469

Schähs P, Weidinger P, Probst OC, Svoboda B, Stadlmann J, Beug H, Waerner T, Mach L. 2008. Cellular repressor of E1A-stimulated genes is a *bona fide* lysosomal protein which undergoes proteolytic maturation during its biosynthesis. *Exp Cell Res* **314**:3036–3047. DOI: https://doi.org/10.1016/j.yexcr.2008.06.015, PMID: 18621046

Sapriel G, Quinet M, Heijde M, Jourdren L, Tanty V, Luo G, Le Crom S, Lopez PJ. 2009. Genome-Wide Transcriptome Analyses of Silicon Metabolism in *Phaeodactylum tricornutum* Reveal the Multilevel Regulation of Silicic Acid Transporters. *PLoS ONE* **4**:e7458. DOI: https://doi.org/10.1371/journal.pone.0007458, PMID: 19829693

Savelli B, Li Q, Webber M, Jemmat AM, Robitaille A, Zamocky M, Mathé C, Dunand C. 2019. RedoxiBase: A database for ROS homeostasis regulated proteins. *Redox Biol* **26**:101247. DOI: https://doi.org/10.1016/j.redox.2019.101247, PMID: 31228650

Savitsky P, Bray J, Cooper CDO, Marsden BD, Mahajan P, Burgess-Brown NA, Gileadi O. 2010. High-throughput production of human proteins for crystallization: the SGC experience. *J Struct Biol* **172**:3–13. DOI: https://doi.org/10.1016/j.jsb.2010.06.008, PMID: 20541610

Scheiber IF, Pilátová J, Malych R, Kotabova E, Krijt M, Vyoral D, Mach J, Léger T, Camadro J-M, Prášil O, Lesuisse E, Sutak R. 2019. Copper and iron metabolism in *Ostreococcus tauri* - the role of phytotransferrin, plastocyanin and a chloroplast copper-transporting ATPase. *Metallomics* **11**:1657–1666. DOI: https://doi.org/10.1039/c9mt00078j, PMID: 31380866

Shoemaker BA, Portman JJ, Wolynes PG. 2000. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A* **97**:8868–8873. DOI: https://doi.org/10.1073/pnas.160259697, PMID: 10908673

Sengupta R, Poderycki MJ, Mattoo S. 2019. CryoAPEX – an electron tomography tool for subcellular localization of membrane proteins. *J Cell Sci* **132**:jcs222315. DOI: https://doi.org/10.1242/jcs.222315, PMID: 30886003

Sheppard VC, Scheffel A, Poulsen N, Kröger N. 2012. Live Diatom Silica Immobilization of Multimeric and Redox-Active Enzymes. *Appl Environ Microbiol* **78**:211–218. DOI: https://doi.org/10.1128/AEM.06698-11, PMID: 22057862

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, Mcwilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**:539. DOI: https://doi.org/10.1038/msb.2011.75, PMID: 21988835

Smith SR, Gillard JTF, Kustka AB, McCrow JP, Badger JH, Zheng H, New AM, Dupont CL, Obata T, Fernie AR, Allen AE. 2016. Transcriptional Orchestration of the Global Cellular Response of a Model Pennate Diatom to Diel Light Cycling under Iron Limitation. *PLoS Genet* **12**:e1006490. DOI: https://doi.org/10.1371/journal.pgen.1006490, PMID: 27973599

Soll J, Schleiff E. 2004. Protein import into chloroplasts. *Nat Rev Mol Cell Biol* **5**:198–208. DOI: https://doi.org/10.1038/nrm1333, PMID: 14991000

Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT-BG, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, Gupta R, Bennett EP, Mandel U, Brunak S, Wandall HH, Levery SB, Clausen H. 2013. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J* **32**:1478–1488. DOI: https://doi.org/10.1038/emboj.2013.79, PMID: 23584533

Strand TA, Lale R, Degnes KF, Lando M, Valla S. 2014. A New and Improved Host-Independent Plasmid System for RK2-based Conjugal Transfer. *PLoS One* **9**:e90372. DOI: https://doi.org/10.1371/journal.pone.0090372, PMID: 24595202

Tagliabue A, Bowie AR, Boyd PW, Buck KN, Johnson KS, Saito MA. 2017. The integral role of iron in ocean biogeochemistry. *Nature* **543**:51–59. DOI: https://doi.org/10.1038/nature21058, PMID: 28252066

Tesson B, Genet MJ, Fernandez V, Degand S, Rouxhet PG, Martin-Jézéquel V. 2009. Surface Chemical Composition of Diatoms. *Chembiochem* **10**:2011–2024. DOI: https://doi.org/10.1002/cbic.200800811, PMID: 19623594

Theorell H. 1935. Preparation in pure state of the effect group of yellow enzymes. *Biochemische Zeitschrift* **275**:344–346.

Li Y, Tian C, Liu K, Zhou Y, Yang J, Zou P. 2020. A Clickable APEX Probe for Proximity-Dependent Proteomic Profiling in Yeast. *Cell Chem Biol* **27**:858-865.e8. DOI: https://doi.org/10.1016/j.chembiol.2020.05.006; PMID: 32470320

Trinkle-Mulcahy L. 2019. Recent advances in proximity-based labeling methods for interactome mapping. *F1000Res* **8**:135. DOI: https://doi.org/10.12688/f1000research.16903.1, PMID: 30774936

Tuteja R. 2005. Type I signal peptidase: An overview. *Arch Biochem Biophys* **441**:107–111. DOI: https://doi.org/10.1016/j.abb.2005.07.013, PMID: 16126156

Uversky VN. 2015a. The intrinsic disorder alphabet. III. Dual personality of serine. *Intrinsically Disord Proteins* **3**:e1027032. DOI: https://doi.org/10.1080/21690707.2015.1027032, PMID: 28232888

Uversky VN. 2015b. The multifaceted roles of intrinsic disorder in protein complexes. *FEBS Lett* **589**:2498–2506. DOI: https://doi.org/10.1016/j.febslet.2015.06.004, PMID: 26073257

Uversky VN. 2016. Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins. *J Biol Chem* **291**:6681–6688. DOI: https://doi.org/10.1074/jbc.R115.685859, PMID: 26851286

Valko M, Morris H, Cronin MTD. 2005. Metals, Toxicity and Oxidative Stress. *Curr Med Chem* **12**:1161–1208. DOI: https://doi.org/10.2174/0929867053764635, PMID: 15892631

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

Wang L, Jonikas MC. 2020. The pyrenoid. *Curr Biol* **30**:R456–R458. DOI: https://doi.org/10.1016/j.cub.2020.02.051, PMID: 32428480

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

Yee DP, Hildebrand M, Tresguerres M. 2019. Dynamic subcellular translocation of V-type H^+-ATPase is essential for biomineralization of the diatom silica cell wall. *New Phytol* **225**:2411–2422. DOI: https://doi.org/10.1111/nph.16329, PMID: 31746463.

Zak O, Ikuta K, Aisen P. 2002. The Synergistic Anion-Binding Sites of Human Transferrin: Chemical and Physiological Effects of Site-Directed Mutagenesis. *Biochemistry* **41**:7416–7423. DOI: https://doi.org/10.1021/bi0160258, PMID: 12044175

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

Zulu NN, Zienkiewicz K, Vollheyde K, Feussner I. 2018. Current trends to comprehend lipid metabolism in diatoms. *Prog Lipid Res* **70**:1–16. DOI: https://doi.org/10.1016/j.plipres.2018.03.001, PMID: 29524459

2069 **Figure 1.** Functional APEX2-enabled proteogenomics with pTF (phytotransferrin/ISIP2a/Phatr3_J54465)
2070 in the model marine diatom *Phaeodactylum tricornutum*.
2071 **(A)** pTF-APEX2 encoding episome is introduced into *P. tricornutum* cells using bacterial conjugation.
2072 **(B)** Resulting transconjugants are genotyped and evaluated for fusion protein expression. APEX2 activity
2073 and fusion protein localization are then confirmed with an enzymatic assay and electron microscopy,
2074 respectively.
2075 **(C)** In a proximity-dependent proteomic mapping (proximity proteomics) experiment, pTF-APEX2
2076 expressing cell line is supplemented with biotin-phenol and hydrogen peroxide, reaction quenched, and
2077 cells lysed. Cell lysate is then subjected to streptavidin pull-down, proteins analyzed with mass
2078 spectrometry (MS), peptides mapped to a *P. tricornutum* proteome database, and corresponding genes
2079 identified.
2080 **(D)** Interesting MS hits are further evaluated experimentally (e.g., for colocalization with the bait protein
2081 (i.e., pTF) and/or for predicted biochemical activity) as well as bioinformatically. Created with
2082 BioRender.com.

Figure 2. pTF is localized to intracellular vesicles and APEX2 is enzymatically active in live *Phaeodactylum tricornutum* cells.

(A) pTF-mCherry colocalizes with the membrane dye MDY-64 on the cell surface and within intracellular vesicles. The fluorescent fusion protein was expressed under the *pTF* promoter and terminator in a *ΔpTF P. tricornutum* genetic background. Cells were stained with 100 µM MDY-64 for 10 min at room temperature. Scale bar is 10 µm. Fusion protein schematic created with BioRender.com.

(B) Top: *pTF-APEX2* flanked by nitrate-inducible *NR* (nitrate reductase) promoter, terminator, and untranslated regions (UTRs). Linker sequence ("*link*") between *pTF* and *APEX2* encoded 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 pTF-APEX2 fusion protein (84.73 kDa) expression in four transconjugant *P. tricornutum* cell lines (lanes s1–s4; left: soluble cell lysate fraction, right: insoluble cell lysate fraction). Bacterial conjugation was performed with WT *P. tricornutum* cells which is why native pTF (57.05 kDa) bands were present in our samples. Teal circle: APEX2, white circle: linker, gray 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 (cell line s2) were used in this experiment. Standard deviation is shown. *P. tricornutum* cartoons created with BioRender.com.

Figure 2—source data 1. Raw Amplex UltraRed assay data underpinning the chart in ***Figure 2D***.

Figure 2—source data 2. R script to reproduce the chart in ***Figure 2D***.

Figure 2—source data 3. CSV file to be used in conjuction with **Figure 2—source data 2**.

2110 **Figure 3.** pTF-APEX2 is localized to the cell membrane and intracellular vesicles.
2111 **(A)** Electron microscopy protocol summary. Briefly: WT and pTF-APEX2 expressing cells (grown in
2112 triplicate) were fixed, treated with diaminobenzidine (DAB) and hydrogen peroxide, post-fixed with osmium
2113 tetroxide, embedded in agar, negatively stained with uranyl acetate, dehydrated, embedded in resin, and
2114 visualized. Created with BioRender.com.
2115 **(B)** Expected cell surface and intracellular pTF-APEX2 localization. Teal and brown arrows point to APEX2-
2116 induced signal and cell wall, respectively. Zoomed in (top left image) is cell periphery where it can clearly
2117 be seen that cell membrane, not cell wall, is occupied by the fusion protein. pTF-APEX2 containing vesicles
2118 were observed (bottom two left images; acquired with backscatter scanning electron microscopy). Zoomed
2119 in (middle left image) are cell membrane and vesicles with pTF-APEX2. Mitochondrial signal in both WT
2120 and pTF-APEX2 expressing cells is likely due to endogenous (mitochondrial) APEX2-like peroxidases.
2121 Scale bar is 1 µm. *P. tricornutum* cartoons created with BioRender.com.

Figure 4. Proximity-dependent proteomic mapping with APEX2 reveals candidate proteins involved in phytotransferrin (pTF) endocytosis in *Phaeodactylum tricornutum.*

(A) Summary of the proximity-dependent proteomic mapping experiment. Briefly: WT and pTF-APEX2 expressing cells (grown in quintuplicate) 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. Cleared cell lysates 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: Volcano 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 encoded by a known iron-sensitive gene cluster on chromosome 20 are highlighted (arrowed Ensembl IDs). *P. tricornutum* cartoons created with BioRender.com.

(C) 14/38 (10 shown here) MS hits are proteins with iron-sensitive transcripts. Proteins chosen for further characterization 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 proteinaceus organelle—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***; ***Kazamia et al., 2018***).

Figure 4—source data 1. Pre-processed quantitative mass spectrometry proteomics data underpinning the Volcano plot and the table in ***Figure 4B*** and ***Figure 4C***, respectively. The table in ***Figure 4C*** was additionally informed by **Figure 5—figure supplement 1—source data 1**.

Figure 4—source data 2. R script to reproduce the chart in ***Figure 4B***.

Figure 4—source data 3. CSV file to be used in conjuction with **Figure 4—source data 2**.

2150 **Figure 5.** Proteins encoded by a known iron- and silicon-sensitive locus in *Phaeodactylum tricornutum*
2151 colocalize with pTF in the chloroplast vicinity.
2152 **(A)** Genes corresponding to three statistically significant proteomic hits are clustered on chromosome 20.
2153 All three proteins were co-expressed with mCherry-tagged pTF as fusions with a yellow fluorescent protein.
2154 Corresponding Ensembl and UniProt IDs are noted. Numbers indicate base pairs. Created with
2155 BioRender.com.
2156 **(B)** Two proteins—pTF.CREG1 and pTF.CatCh1—show clear, yet distinct, colocalization with pTF and the
2157 chloroplast periphery. pTF.CREG1-EYFP was consistently punctate whereas pTF.CatCh1-EYFP lined the
2158 chloroplast margin. pTF-mCherry punctum was almost exclusively positioned next to this “chloroplast lining”
2159 pattern. *P. tricornutum* is pleiomorphic which explains why different cell morphologies were observed in
2160 pTF.CREG1-EYFP (fusiform morphotype) and pTF.CatCh1-EYFP (triradiate morphotype) cell lines. Scale
2161 bar is 10 µm. Fusion protein schematics created with BioRender.com.

2162 **Figure 6.** pTF.CREG1 homologs are present across the tree of life.
2163 pTF.CREG1 homolog search was performed with the National Center for Biotechnology Information (NCBI)
2164 and the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) databases (***Caron et***
2165 ***al., 2016***; ***Keeling et al., 2014***). Homologs from algae with complex plastids (including diatoms) form a
2166 crown group away from animal proteins. pTF.CREG1 clusters with its paralog Phatr3_J10972 (black
2167 rectangle). Red dots indicate predicted gene duplication events. Scale bar: 0.5 substitutions per position.
2168
2169 **Figure 6—source data 1.** Amino acid sequence alignment underpinning the phylogenetic tree in ***Figure***
2170 ***6***.

2171 **Figure 7.** pTF.CatCh1 is a cysteine-rich heterokont-specific protein.
2172 **(A)** pTF.CatCh1 homolog search was performed with NCBI and marine microbial eukaryote (MMETSP)
2173 databases. Numerous, almost exclusively diatom, proteins were identified. Interestingly, pTF.CatCh1 is
2174 paralogous to ISIP2b (black arrows). Scale bar: 0.1 substitutions per position.
2175 **(B)** pTF.CatCh1 and ISIP2b share three CXC and one CXXC motifs (boxed) commonly found in redox-
2176 active proteins and proteins involved in cellular metal homeostasis. Black boxes indicate the two CX(X)C
2177 motifs conserved across the majority of pTF.CatCh1 homologs. Blue triangles: additional conserved
2178 cysteine amino acid residues between pTF.CatCh1 and ISIP2b for a total of 20. Asterisks: cysteine amino
2179 acid residues that are 100% conserved across all homologs.
2180
2181 **Figure 7—source data 1.** Amino acid sequence alignment underpinning the phylogenetic tree in ***Figure***
2182 ***7A***.

2183 **Figure 8.** Emerging view of intracellular inorganic iron allocation pathway in *Phaeodactylum tricornutum.*
2184 We propose a model for clathrin-mediated internalization of phytotransferrin (pTF)-bound ferric, Fe^{3+}, iron,
2185 and its trafficking to the chloroplast periphery. In this model, pTF releases Fe^{3+} after undergoing a structural
2186 change upon endosome acidification driven by a vacuolar-type H^{+}-ATPase. pTF.CREG1 may regulate this
2187 process at the endosome periphery and perhaps also in earlier endocytosis stages closer to the cell surface.
2188 It is unclear how intravesicular reduction of ferric iron is achieved and how the resulting ferrous iron gets
2189 exported. Given that no DMT1/NRAMP2 homologs are present in *P. tricornutum*, Zrt- and Irt-like proteins
2190 (ZIPs)—known to export Fe^{2+} from intracellular compartments—may mediate the latter (***Blaby-Haas and***
2191 ***Merchant, 2012***; ***Kustka et al., 2007***; ***Lampe et al., 2018***). Ferrous iron could be offloaded to pTF.CatCh1
2192 on the outer chloroplast membrane and ultimately transported further into the chloroplast interior for cellular
2193 assimilation and use (e.g., iron–sulfur cluster biosynthesis for incorporation into photosystems I and II).
2194 pTF.CatCh1 may alternatively (or additionally) be responsible for pTF recruitment to the chloroplast margin.
2195 The enrichment of FBP1 and ISIP1—both previously shown to localize to intracellular vesicles in *P.*
2196 *tricornutum* (***Coale et al., 2019***; ***Kazamia et al., 2019***)—in our proximity proteomics experiment suggests
2197 the whole pathway intersects with that for internalizing organic, siderophore-bound, iron. Gray panel:
2198 pytotransferrin (pTF). Blue panels: proteins colocalized with pTF. Green panels: additional proteins
2199 enriched in our APEX2 experiment. One phytotransferrin-Fe^{3+} complex is shown for clarity. Orange vesicle
2200 interior indicates pH drop due to ATPase activity. Detailed chloroplast membrane layers are omitted (diatom
2201 chloroplasts contain four membranes). Created with BioRender.com.

2202 **Figure 2—figure supplement 1.** *Phaeodactylum tricornutum* phytotransferrin (pTF) features, resorufin
2203 imaging in pTF-APEX2 expressing cells, and resorufin signal co-dependence on Amplex UltraRed and
2204 hydrogen peroxide.
2205 **(A)** pTF is a 537 amino acid residues (AAs) long cell membrane-associated iron receptor protein consisting
2206 of a signal peptide (AAs 1–18; predicted with SignalP 4.1), an iron-binding domain (AAs 19–493), a
2207 transmembrane domain (AAs 494–516; TMHMM Server v. 2.0 prediction output is shown to the right), and
2208 a tail domain (AAs 517–537). The latter is predicted to contain an endocytosis motif (***Lommer et al., 2012***).
2209 Bolded and light blue are conserved putative iron- (Asp48, Tyr71, Tyr214, His294) and carbonate-
2210 coordinating (Arg183) amino acid residues, respectively.
2211 **(B)** Direct visualization of resorufin (yellow) in pTF-APEX2 expressing cells. Both WT and pTF-APEX2
2212 expressing cells were incubated in reaction buffer (50 µM Amplex UltraRed, 2 mM H_2O_2, PBS (pH 7.4)) for
2213 6 min at room temperature prior to imaging. Red is chloroplast autofluorescence.
2214 **(C)** Supernatants after incubating cells from four pTF-APEX2 expressing cell lines (s1–s4) in reaction buffer
2215 (50 µM Amplex UltraRed, 2 mM H_2O_2, PBS (pH 7.4)) for 6 min at room temperature. Resorufin signal was
2216 observed only when both APEX2 substrates (Amplex UltraRed and hydrogen peroxide) were present.

2217 **Figure 3—figure supplement 1.** Analysis of transcriptionally active *Phaeodactylum tricornutum*
2218 peroxidases reveals putative APEX2-like and mitochondrial enzymes.
2219 **(A)** Detail from the electron microscopy sample preparation protocol: agar blocks with embedded
2220 diaminobenzidine (DAB)-treated pTF-APEX2 expressing diatom cells. *P. tricornutum* cartoon and "zoom
2221 in" objects created with BioRender.com.
2222 **(B)** Mitochondrial signal as shown here was observed in electron microscopy experiments with both WT
2223 and pTF-APEX2 expressing cells subjected to DAB and hydrogen peroxide. "Zoom in" objects created with
2224 BioRender.com.
2225 **(C)** Amino acid sequence alignment of putative APEX2-like *P. tricornutum* peroxidases. Amino acid
2226 residues crucial for ascorbate peroxidase (APX) and likely crucial for APEX2 activity (Arg38, His42, His163,
2227 and Asp208; numbering is relative to APEX2; depicted with black triangles and shaded) are conserved in
2228 eight *P. tricornutum* peroxidases. Numbers above amino acid sequences indicate alignment position.
2229 **(D)** Amino acid sequence alignment of putative substrate binding loops in the identified APEX2-like
2230 peroxidases. Black and blue arrows point to enzymes with proline at position 134 (shaded; numbering is
2231 relative to APEX2)—an amino acid residue change rendering APEX2 much more active than its APEX
2232 predecessor (***Lam et al., 2015***)—and predicted mitochondrial localization, respectively.
2233 **(E)** Some additional properties of the identified APEX2-like peroxidases. They all belong to class I
2234 (intracellular) haem peroxidases; one may be membrane-associated. [1]RedoxiBase database. [2]APEX2 and
2235 horseradish peroxidase (HRP) have proline at this position. [3]TargetP 1.1. [4]TMHMM Server v. 2.0.

2236 **Figure 4—figure supplement 1.** WT and pTF-APEX2 proteomic replicates form distinct clusters.
2237 Hierarchical clustering (Ward's method as implemented in the JMP software package) of scaled quantitative
2238 mass spectrometry proteomics data reveals WT- and pTF-APEX2-specific clusters. The heat map as a
2239 whole suggests minimal technical variability in our proteomic experiment even though samples WT5 and
2240 pTF-APEX2.4 deviate slightly from other corresponding samples.
2241
2242 **Figure 4—figure supplement 1—source data 1.** Scaled quantitative mass spectrometry proteomics data.

Figure 5—figure supplement 1. *pTF.CREG1*, *pTF.CatCh1*, and *pTF.ap1* closely resemble *pTF* transcriptional profile.
Transcriptional profiles of *pTF* and the three genes from the identified iron- and silicon-sensitive gene cluster on chromosome 20 across different growth conditions (L (low): 20 pM Fe', M (medium): 40 pM Fe', H (high): 400 pM Fe') and sampling timepoints (10AM, 2PM, 6PM, 10PM, 2AM, 6AM) are shown. Fe': sum of all iron species not complexed to EDTA. Red dashed lines correspond to genes' transcriptional peaks (late night). Average RPKM: Reads Per Kilobase of transcript, per Million mapped reads averaged across across two (L and M) or three (H) replicates.

Figure 5—figure supplement 1—source data 1. Transcriptomic data and functional annotations for proteins with APEX2/WT ratios ≥1.5 and P values ≤0.05 underpinning ***Figure 5—figure supplement 1*** and ***Figure 4C***.

Figure 5—figure supplement 1—source data 2. R script to reproduce the chart in ***Figure 5—figure supplement 1***.

Figure 5—figure supplement 2—source data 3. CSV file to be used in conjuction with **Figure 5—figure supplement 1—source data 2**.

2261 **Figure 5—figure supplement 2.** Co-expression episome assembly overview, assessment of co-
2262 expression cell lines, and confocal microscopy bleed-through controls.
2263 **(A)** Summary of the pTF-mCherry MS hit-EYFP co-expression episome assembly strategy. Created with
2264 BioRender.com.
2265 **(B)** Colony PCR and western blot analyses indicate the presence of full-length fusion proteins in our co-
2266 expression cell lines. Left: Primers binding to *Phatr3_J23658* (flavodoxin-encoding gene) promoter and
2267 terminator flanking MS hit-EYFP coding regions were used in PCR reactions which resulted in the expected
2268 ~3 kbp amplicons. Positive controls: pure episomes used for conjugation. Negative controls: WT *P.*
2269 *tricornutum* cells and water which yielded no amplicons (not shown). Right: Anti-pTF western blot revealed
2270 the expected pTF-mCherry (83.75 kDa) bands alongside the bands corresponding to native pTF (57.05
2271 kDa) and what appears to be a pTF-mCherry truncation (the additional band at ~70 kDa). Protein
2272 schematics created with BioRender.com.
2273 **(C)** The observed lack of significant cross-channel bleed-through supports colocalization imaging data and
2274 conclusions. Imaging conditions with minimal bleed-through were determined with WT *P. tricornutum* cells
2275 (top row), *P. tricornutum* cells expressing mCherry alone (middle row), and *P. tricornutum* cells expressing
2276 a Venus-tagged protein (bottom row). Scale bar is 10 μm.

2277 **Figure 5—figure supplement 3.** Additional representative examples of pTF.CREG1 and pTF.CatCh1
2278 colocalization with pTF.
2279 **(A)** Three additional cells from pTF-mCherry and pTF.CREG1-EYFP expressing cell line described in the
2280 main text exhibit similar punctate colocalization of both fusion proteins at the chloroplast periphery.
2281 **(B)** Three additional cells from pTF-mCherry and pTF.CatCh1-EYFP expressing cell line described in the
2282 main text exhibit similar “chloroplast lining” colocalization pattern of both fusion proteins at the chloroplast
2283 periphery. pTF-mCherry punctum was typically positioned adjacently to this signal. Scale bar is 10 μm.

2284 **Figure 5—figure supplement 4.** pTF.ap1 colocalizes with pTF.
2285 pTF.ap1-EYFP was co-expressed with pTF-mCherry and colocalization of the two fusion proteins—
2286 although not as obvious as in the case of pTF.CREG1 and pTF.CatCh1—was observed. These cells are
2287 representative of a pooled non-clonal *P. tricornutum* culture. Scale bar is 10 µm. Fusion protein schematics
2288 created with BioRender.com.

2289 **Figure 6—figure supplement 1.** Conserved amino acid residues across pTF.CREG1 homologs.
2290 Protein sequence alignment (***Figure 6—source data 1***) leading to the phylogenetic tree shown in ***Figure***
2291 ***6*** was examined for amino acid residues that are at least 90% conserved across all homologs. 12 such
2292 amino acid residues (blue triangles) in nine examined regions were identified. Additional few amino acid
2293 residues surrounding each conserved site are shown. Putative loop tetrapeptide motifs (see the main text
2294 for details) shown to the far right directly precede the conserved proline and cysteine amino acid residues
2295 in region 4. The DPES tetrapeptide motif in pTF.CREG1 is also found in *Pteridomonas danica* (single-celled
2296 heterotrophic heterokont), a *Pterosperma* species (single-celled green alga), *Gossypium barbadense*,
2297 *Durio zibethinus*, and *Theobroma cacao* (all three are plants). Black arrows point to pTF.CREG1 and its
2298 paralog Phatr3_J10972. Human CREG1 and pTF.CREG1 sequences are shaded. Numbers above amino
2299 acid sequences indicate alignment position. *17 AAs between glycine (G) and arginine (R) in the conserved
2300 region 1 were omitted.

Figure 6—figure supplement 2. Bioinformatic, structural, and enzymatic insights suggest pTF.CREG1 has a non-catalytic function similar to human CREG1.

(A) pTF.CREG1 (UniProt ID: B7G9B3) features and its amino acid sequence alignment with human CREG1 (UniProt ID: O75629). pTF.CREG1 has a signal peptide that ends in SGA (blue underline; predicted with SignalP 4.1). This motif is closely related to Type I signal peptidase recognition site AXA indicating that pTF.CREG1 may enter the secretory pathway (***Tuteja, 2005***). Blue triangle points to a putative peptidase cut site. The core signal peptide sequence is enriched in hydrophobic amino acid residues, a common signal peptide feature (black underline). N-terminal methionine in Met-Phe (gray underline) may be acetylated by an N-terminal acetyltransferase (NAT), perhaps NatC or NatE which use Met-Phe sequence as their substrate (***Ree et al., 2018***). Tyrosine immediately succeeding SGA (green triangle) was used by ASAFind version 1.1.7 to classify pTF.CREG1 as a low confidence chloroplast-directed protein. Ser32 and Ser133 (red triangles) are predicted to undergo mucin-type O-glycosylation. Yellow triangle indicates an asparagine amino acid residue (Asn132) that is predicted to be N-glycosylated in human CREG1 (***Sacher et al*., 2005**). These putative glycosylation sites further support the notion that the protein trafficks along the secretory pathway. Amino acid residues 141–144 in human CREG1 (DPQS)—which are located on the FMN binding pocket-occluding loop (***Sacher et al*., 2005**)—are largely conserved (DPES) in pTF.CREG1 (shaded black rectangle). Finally, magenta triangle points to the arginine mutation site, Asn79, which may—like the corresponding Asp87 in human CREG1—prevent coordination of the terminal FMN phosphate (***Figure 6—figure supplement 3A***; ***Sacher et al*., 2005**).

(B) Structural comparison between human CREG1 (PDB ID: 1XHN) and pTF.CREG1 homology model (created with Phyre2 web portal) suggests pTF.CREG1 is unable to coordinate FMN at the possible dimer interface (***Figure 6—figure supplement 3B***). Human CREG1 does not bind FMN *in vitro* (***Sacher et al*., 2005**) due to the FMN binding pocket-occluding loop (gray rectangle). Pro146 and Glu147 in pTF.CREG1 are predicted to protrude into this pocket similarly to their human CREG1 counterparts Pro142 and Gln143. Gray: human CREG1 chain A (i.e., monomer 1), colored by secondary structure: human CREG1 chain B (i.e., monomer 2), blue: pTF.CREG1 homology model. N- and C-termini in protein structures are depicted. Created with UCSF Chimera version 1.11.1 and BioRender.com.

(C) pTF.CREG1 does not reduce physiologically relevant flavins. Left: Comparison of NADPH oxidation (measured by decrease in absorbance at 320 nm) by recombinant pTF.CREG1 (long dashed black line) supplemented with flavins (flavin mononucleotide (FMN) and riboflavin) versus a no enzyme, water control (solid gray line). Right: The same assay, but here comparing a negative control protein without known NADPH oxidation activity (bovine serum albumin, BSA; short dashed black line) versus a no enzyme, water control (solid gray line). Experiments were started by the addition of protein or water blank and carried out in triplicate for each treatment. Background 320 nm absorbance of water blanks was subtracted from each replicate. Standard deviation is shown.

Figure 6—figure supplement 2—source data 1. Flavin reduction enzymatic assay data underpinning charts in ***Figure 6—figure supplement 2C***.

Figure 6—figure supplement 3. pTF.CREG1 is unlikely able to coordinate the terminal FMN phosphate and may act as a dimer *in vivo*.

(A) Two mutated arginine amino acid residues in human CREG1—one (Asp87) can be inferred from ***Sacher et al., 2005***—are believed to prevent the protein from coordinating the terminal FMN phosphate. Asn79 in pTF.CREG1 is a corresponding amino acid residue to the human Asp87 (inset). The steric similarity between their side chains sugests that pTF.CREG1 would also be unable to coordindate the terminal FMN phosphate even in the absence of the FMN binding pocket-occluding loop. Gray: human CREG1 chain A (i.e., monomer 1 from PDB ID: 1XHN), colored by secondary structure: human CREG1 chain B (i.e., monomer 2 from PDB ID: 1XHN), blue: pTF.CREG1 homology model. Created with UCSF Chimera version 1.11.1 and BioRender.com.

(B) Putative homodimerization region in the pTF.CREG1 homology model exhibits features of the human CREG1 dimer one. Human CREG1 chain A (i.e., monomer 1 from PDB ID: 1XHN) from the visualization presented in ***Figure 6—figure supplement 2B*** was hidden which left us with a chimeric human-diatom dimer. Colored by secondary structure: human CREG1 chain B (i.e., monomer 2 from PDB ID: 1XHN), blue: pTF.CREG1 homology model. Gray rectangles indicate the FMN binding pocket-occluding loop (see the main text). Black rectangle depicts a modeled β-sheet region in pTF.CREG1 that is absent from human CREG1. Colored in magenta, red, orange, and green is a possible homodimerization interface in pTF.CREG1. Amino acid residue composition (written in N-terminal→C-terminal direction) corresponding to each color is as follows: magenta: **P**E**S**<u>**PV**</u>C<u>**A**</u>R<u>**L**</u>T<u>**L**</u>S<u>**G**</u>, red: K<u>**V**</u>S<u>**F**</u>T<u>**L**</u>T<u>E</u><u>**A**</u>S<u>**L**</u>, orange: <u>**WGVL**</u>TT<u>**I**</u>, green: <u>**P**</u>D<u>**G**</u>Q<u>**PFG**</u>N<u>**VY**</u>S<u>**F**</u>. Bold: hydrophobic, underlined: hydrophobic and/or following ***Sacher et al., 2005***. Created with UCSF Chimera version 1.11.1 and BioRender.com

2360 **Figure 6—figure supplement 4.** pTF.CREG1 expression and purification.
2361 Briefly, genes encoding various pTF.CREG1 truncations were placed under arabinose-inducible promoter,
2362 expressed in BL21 *E. coli* cells, and purified using cobalt, Co^{2+}, resin-based gravity flow columns. CTHF:
2363 C-terminal His6 and FLAG tag. NTH: N-terminal His6 tag. L: PageRuler Plus Prestained Protein Ladder.
2364 **(A)** Final elutions of various pTF.CREG1 truncations from small scale (50 mL) protein expression
2365 experiments. Δ31_pTF.CREG1-His6 (bolded and black rectangle) was selected for scale-up.
2366 **(B)** The final Δ31_pTF.CREG1-His6 elution from a large-scale (2 L) protein expression experiment. NuPage
2367 4-12% Bis-Tris gels were used to resolve protein samples.

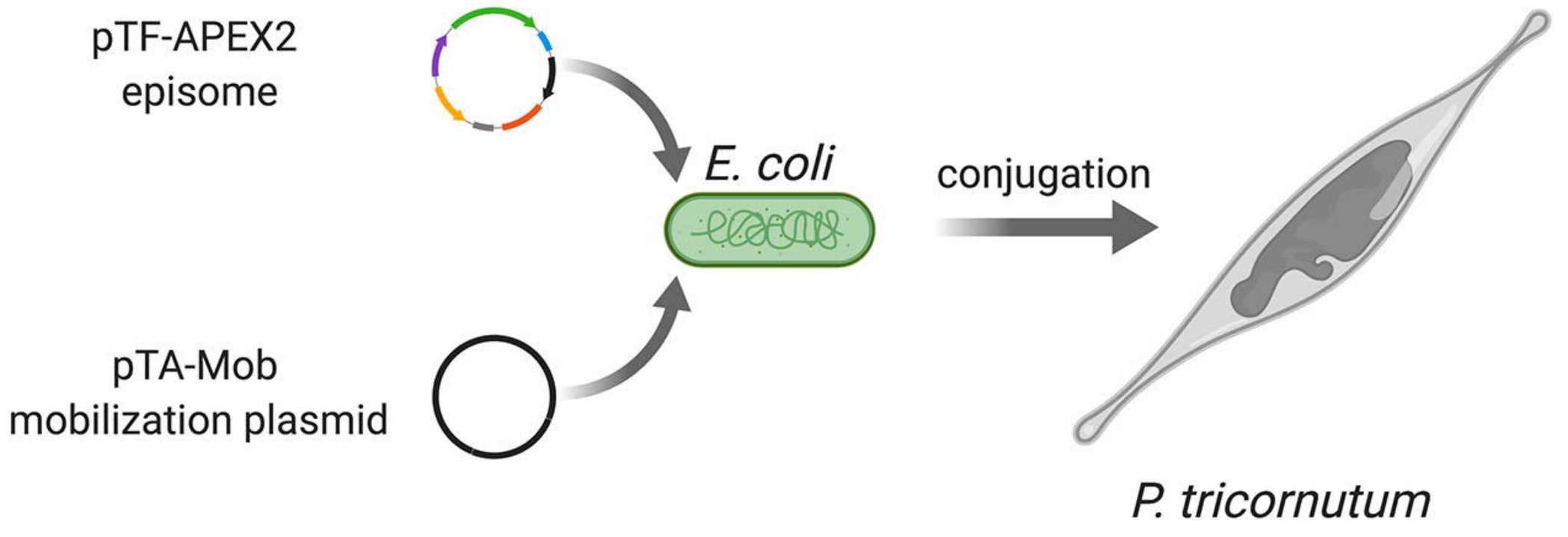
2368 **Figure 7—figure supplement 1.** Conservation of CX(X)C motifs in pTF.CatCh1 homologs and
2369 pTF.CatCh1 features.
2370 **(A)** CXXC and the adjacent CXC motifs are present in all but two identified pTF.CatCh1 homologs. Six
2371 additional cysteines are 100% conserved (***Figure 7—source data 1***). pTF.CatCh1 and ISIP2b (paralogs)
2372 are marked with black arrows. Amino acid residues immediately preceding the CXXC motif are also shown
2373 as aspartic acid residue (D) at this position is known to promote divalent metal ion binding (***Banci et al.,***
2374 ***2002***; ***Banci et al., 2006***). Numbers above amino acid sequences indicate alignment position.
2375 **(B)** pTF.CatCh1 contains two disordered C-terminal regions flanking the transmembrane domain. CXXC is
2376 predicted to be in one of them by PONDR VSL2 (top). No cysteines are predicted to be in this disordered
2377 region by IUPred2a (bottom). Approximate locations of the conserved CXC and CXXC motifs are depicted
2378 with blue and orange triangles, respectively.
2379 **(C)** pTF.CatCh1 features assuming disorder prediction with PONDR VSL2. Signal peptide (AAs 1–26; gray)
2380 is followed by a cysteine-rich domain (20 cysteines between amino acid residue 34 and 226; pink), and two
2381 disordered regions (green). The transmembrane (TM) domain (AAs 296–315; blue) nested between the
2382 predicted disordered regions is flanked by a polyserine ($[S]_6$) and an arginine-rich ($RKL[R]_3$) amino acid
2383 residue stretch. N- and C-terminus are shown. Created with BioRender.com.

Figure 7—figure supplement 2. pTF.ap1 phylogeny and conserved motifs.
(A) pTF.ap1 (black arrow) 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 the genus *Pseudo-nitzschia*, haptophyte *Prymnesium parvum*, and pelagophyte *Aureococcus anophagefferens*), cosmopolitan diatom *Ditylum brightwelli*, haptophyte *Emiliana huxleyi*, ultraplanktonic species *Pelagomonas calceolate*, colonial species *Prasinoderma coloniale* and *Phaeocystis antarctica*. Scale bar: 0.1 substitutions per position.
(B) pTF.ap1 (black arrow) contains highly conserved motifs (five shown here). There are three 100% conserved cysteine amino acid residues across pTF.ap1 homologs (***Figure 7—figure supplement 2—source data 1***) in addition to the two here (blue triangles and gray shade). Numbers above amino acid sequences indicate alignment position.

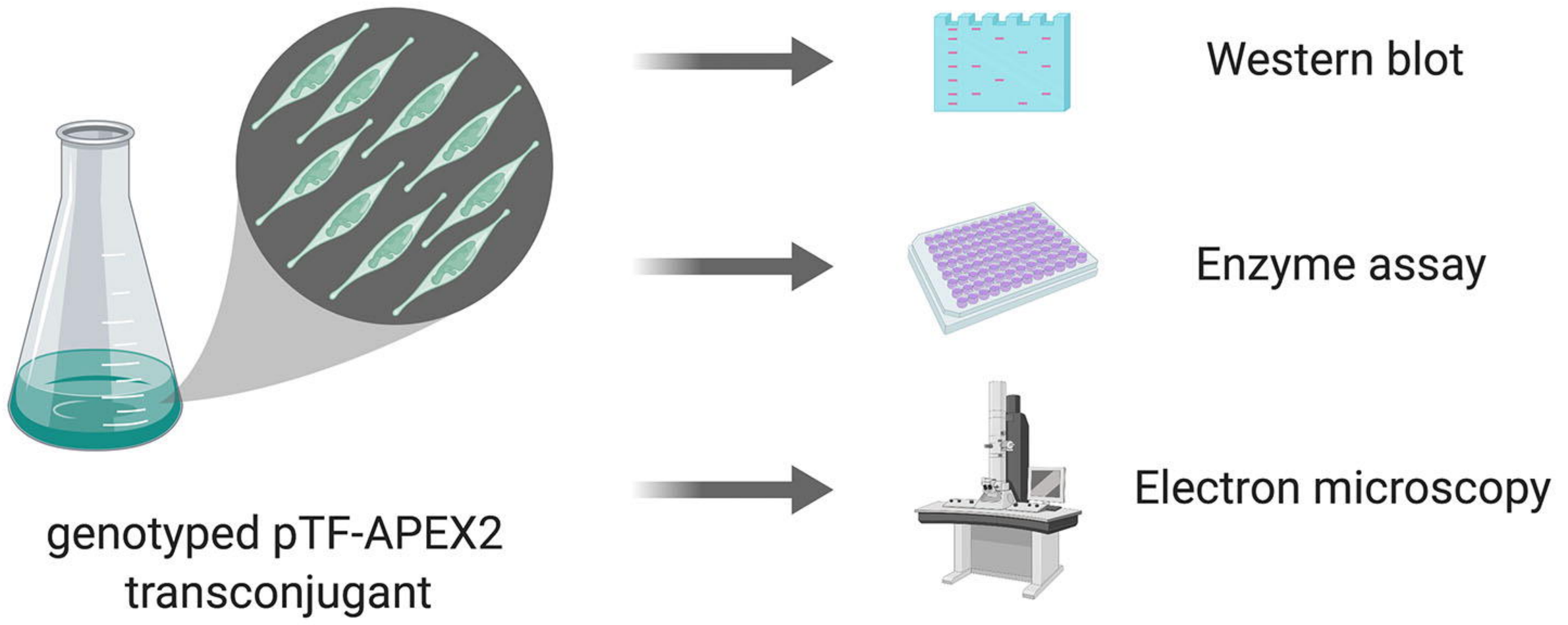
Figure 7—figure supplement 2—source data 1. Amino acid sequence alignment underpinning the phylogenetic tree in ***Figure 7—figure supplement 2A***.

2399 **Figure 8—figure supplement 1.** Outstanding biological problems in established model marine diatoms
2400 poised to be advanced with proximity-dependent proteomic mapping approaches.
2401 Two diatom compartments of immediate interest are pyrenoid—proteinaceous RuBisCO-containing
2402 subchloroplastic organelle—and silica deposition vesicle (SDV)—diatom cell wall biosynthesis site. *P.*
2403 *tricornutum* and *T. pseudonana*—genetically tractable model marine diatoms—are appropriate for
2404 immediate dissection of pyrenoid and SDV proteomic composition, respectively. Created with
2405 BioRender.com.

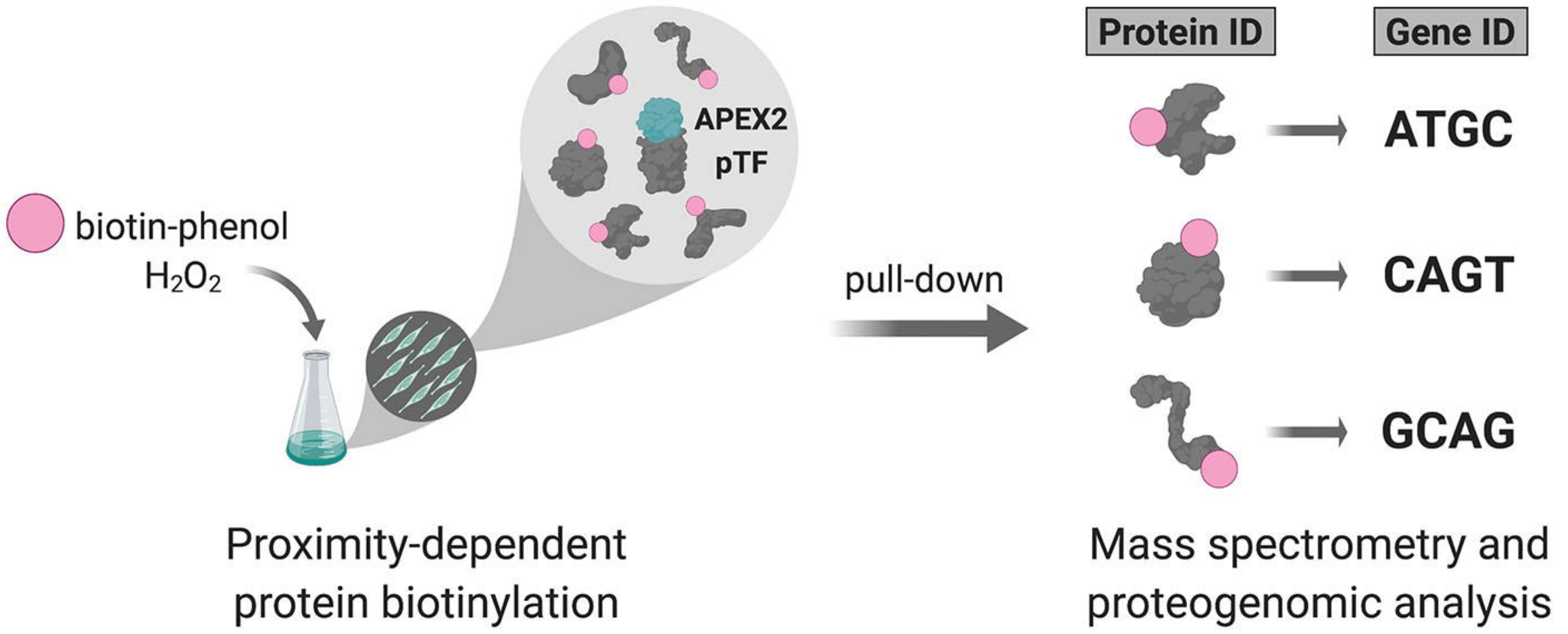
(A) Generate pTF-APEX2 *P. tricornutum* cell line



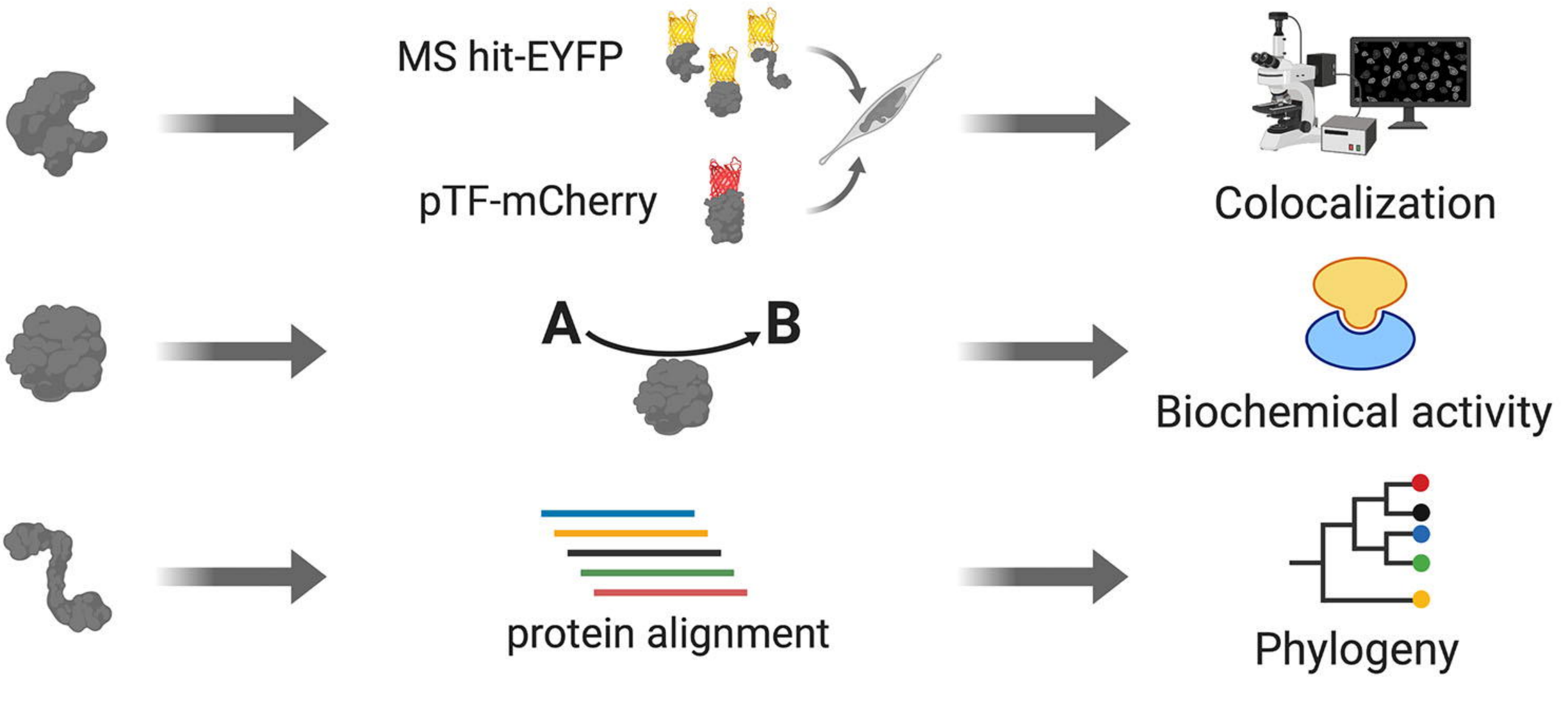
(B) Evaluate pTF-APEX2 expression, activity, and localization

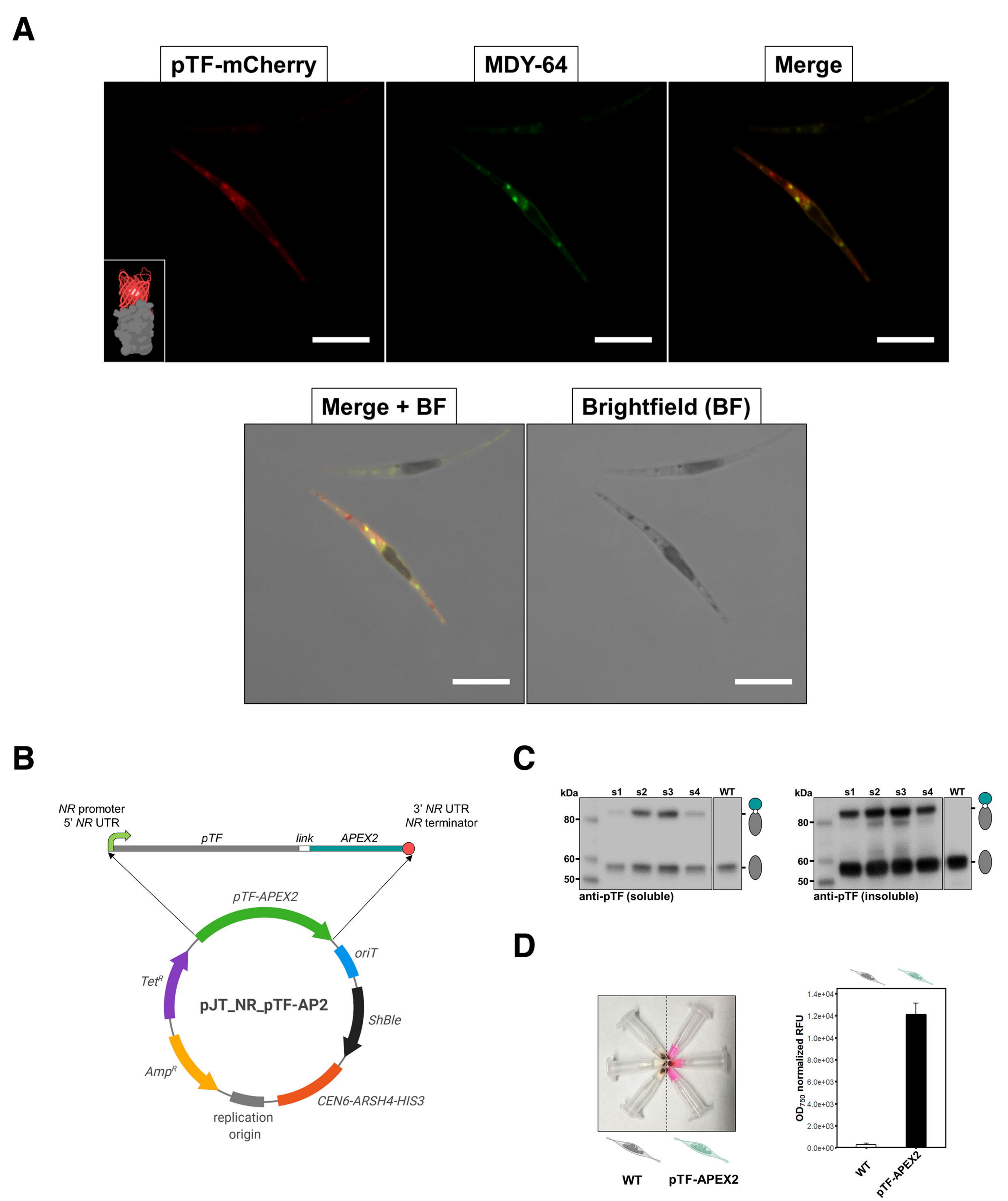


(C) Proximity proteomics with pTF-APEX2 cell line

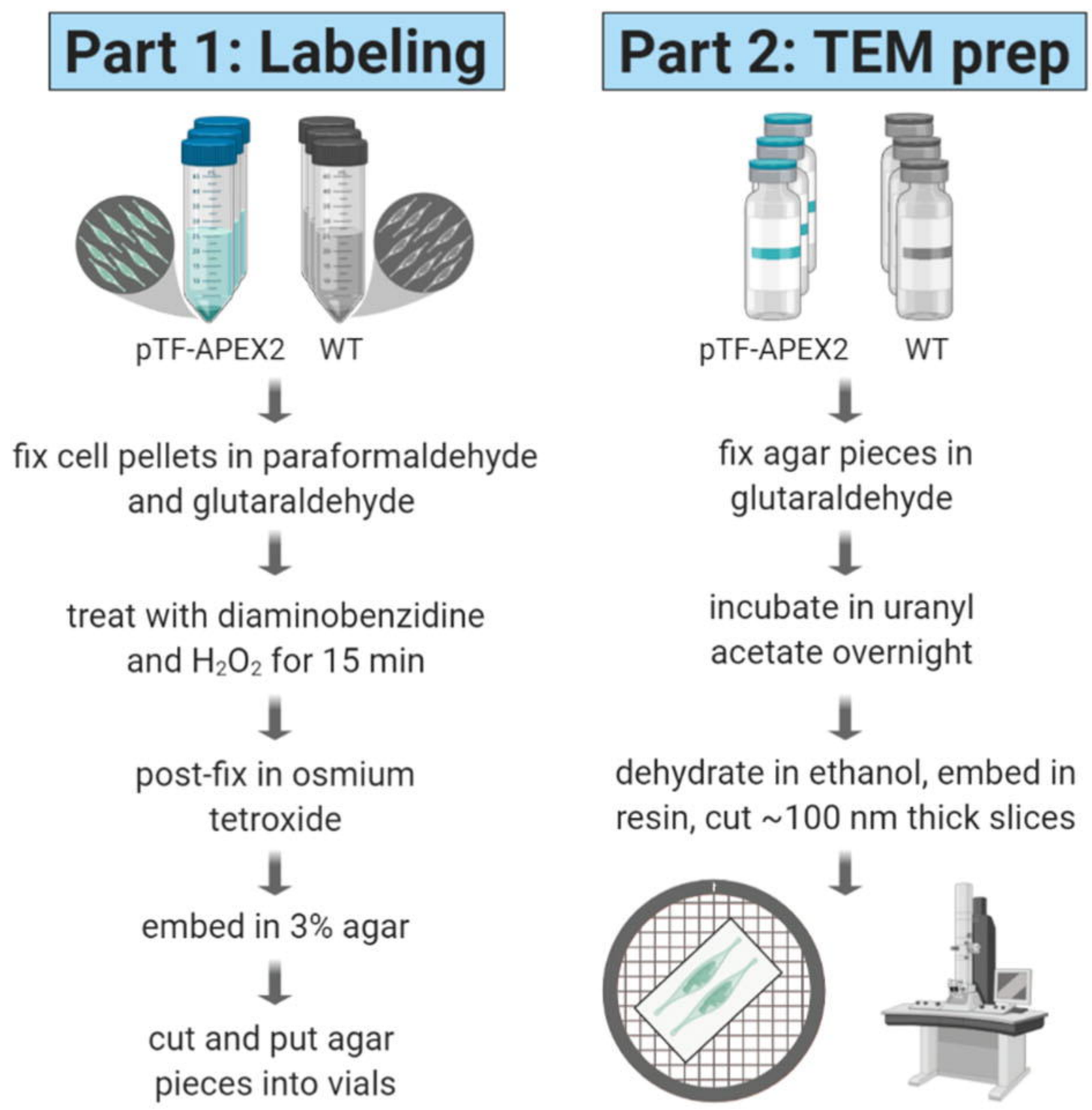


(D) Characterization of mass spectrometry hits

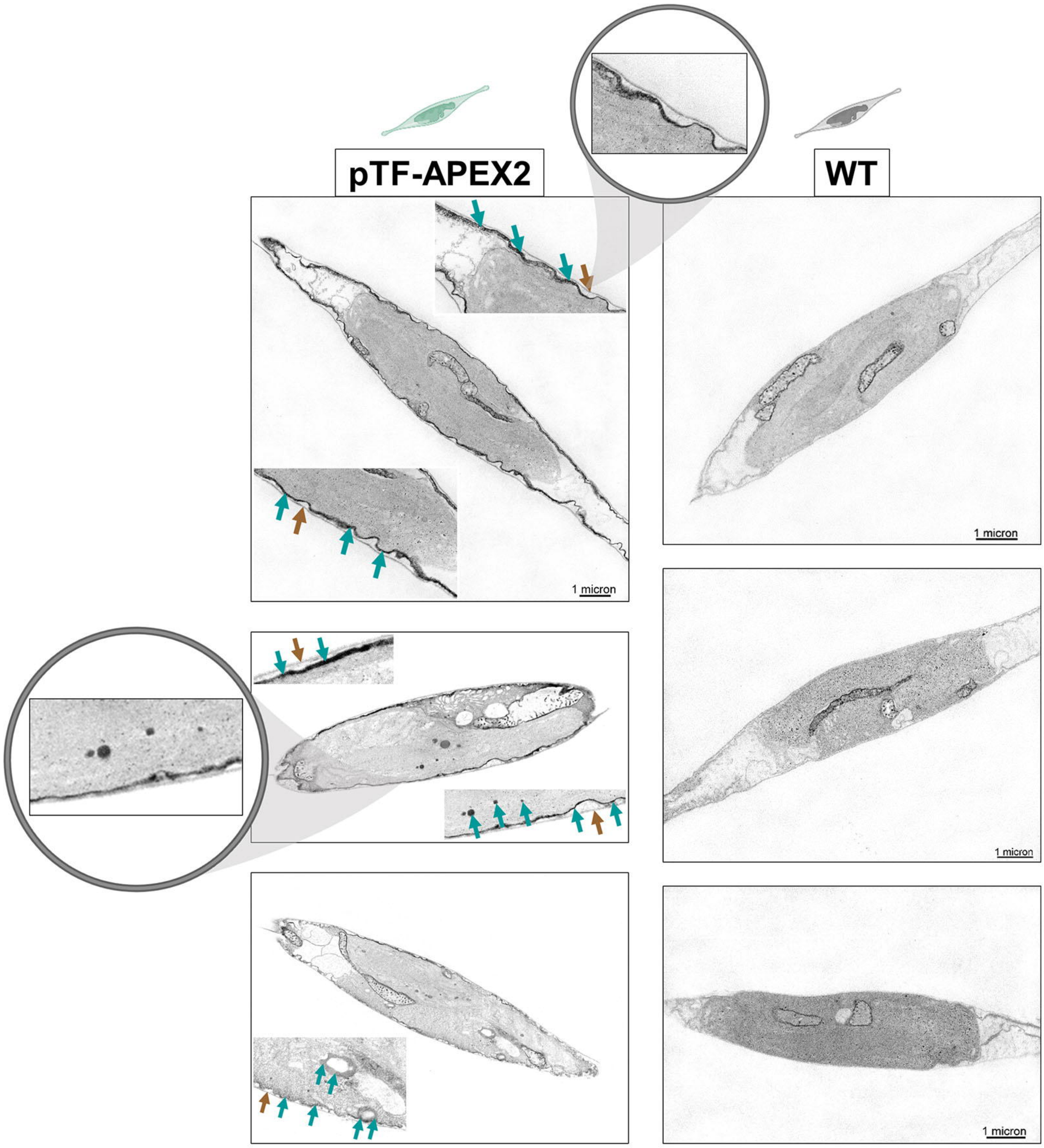



A
pTF-mCherry
MDY-64
Merge
Merge + BF
Brightfield (BF)
B
NR promoter
5' NR UTR
pTF
link
APEX2
3' NR UTR
NR terminator
pTF-APEX2
oriT
ShBle
CEN6-ARSH4-HIS3
replication origin
AmpR
TetR
pJT_NR_pTF-AP2
C
kDa
s1
s2
s3
s4
WT
80
60
50
anti-pTF (soluble)
anti-pTF (insoluble)
D
WT
pTF-APEX2
OD750 normalized RFU
1.4e+04
1.2e+04
1.0e+04
8.0e+03
6.0e+03
4.0e+03
2.0e+03
0.0e+00

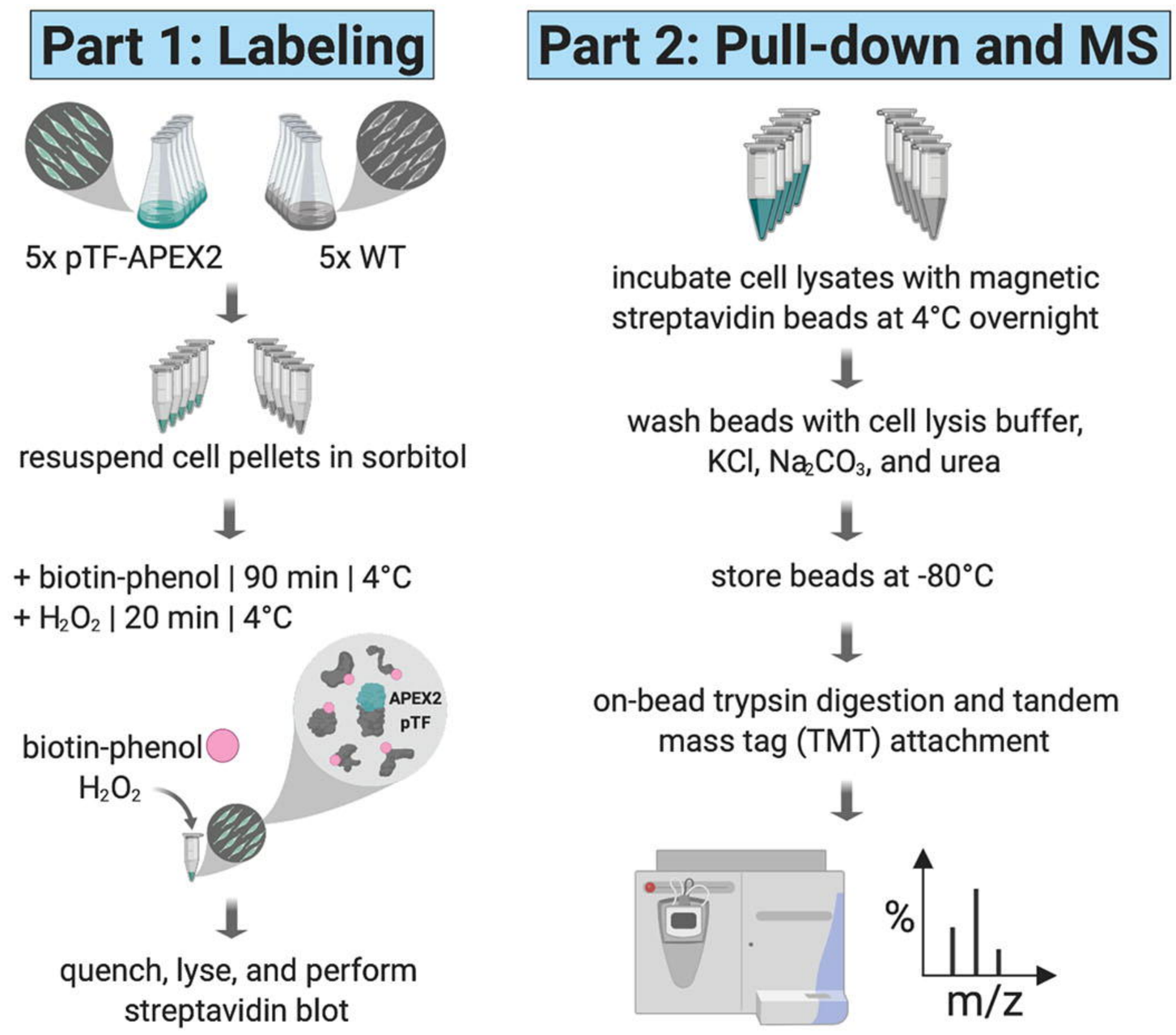
A


Part 1: Labeling
pTF-APEX2
WT
fix cell pellets in paraformaldehyde and glutaraldehyde
treat with diaminobenzidine and H_2O_2 for 15 min
post-fix in osmium tetroxide
embed in 3% agar
cut and put agar pieces into vials
Part 2: TEM prep
pTF-APEX2
WT
fix agar pieces in glutaraldehyde
incubate in uranyl acetate overnight
dehydrate in ethanol, embed in resin, cut ~100 nm thick slices

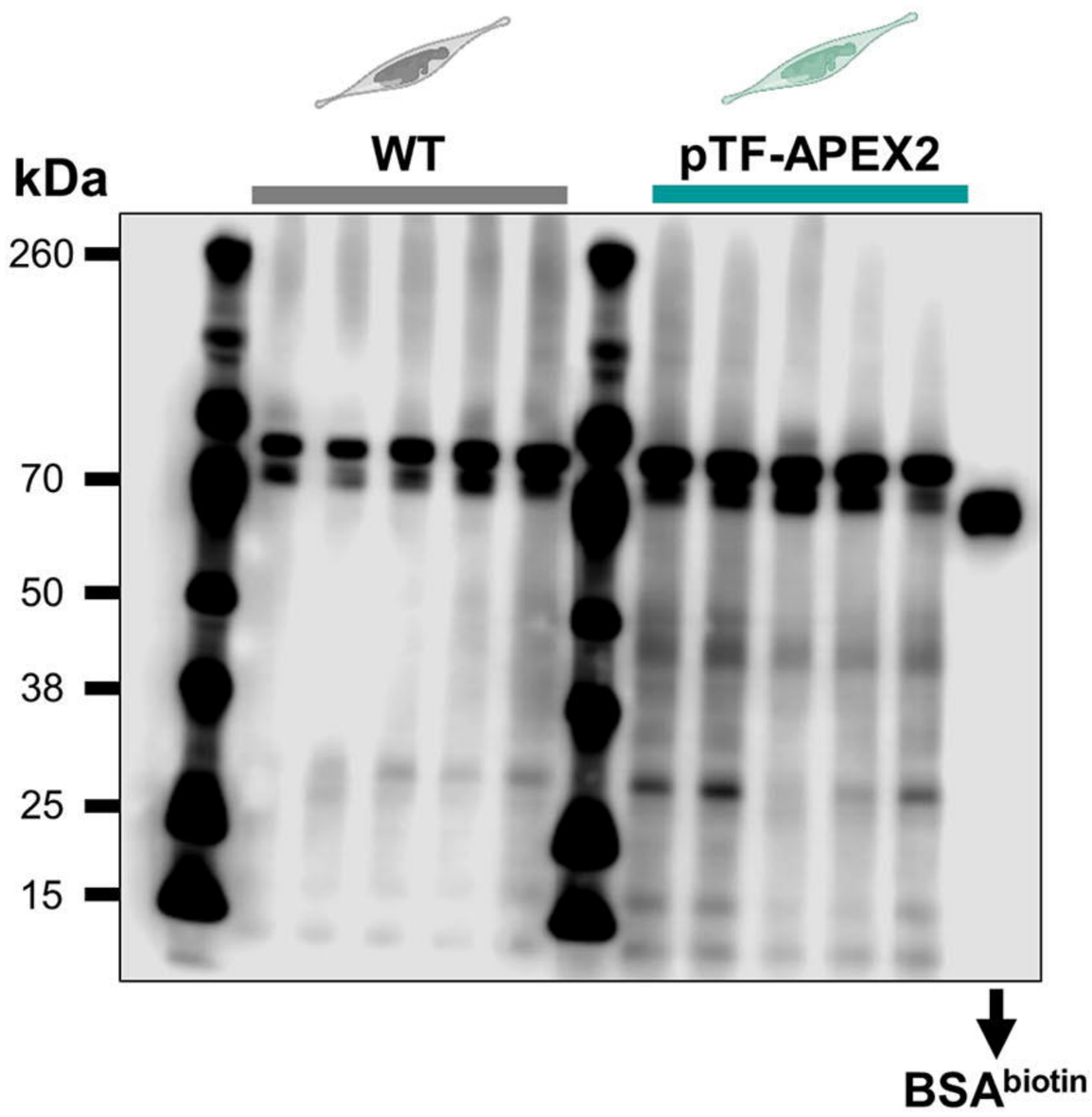
B

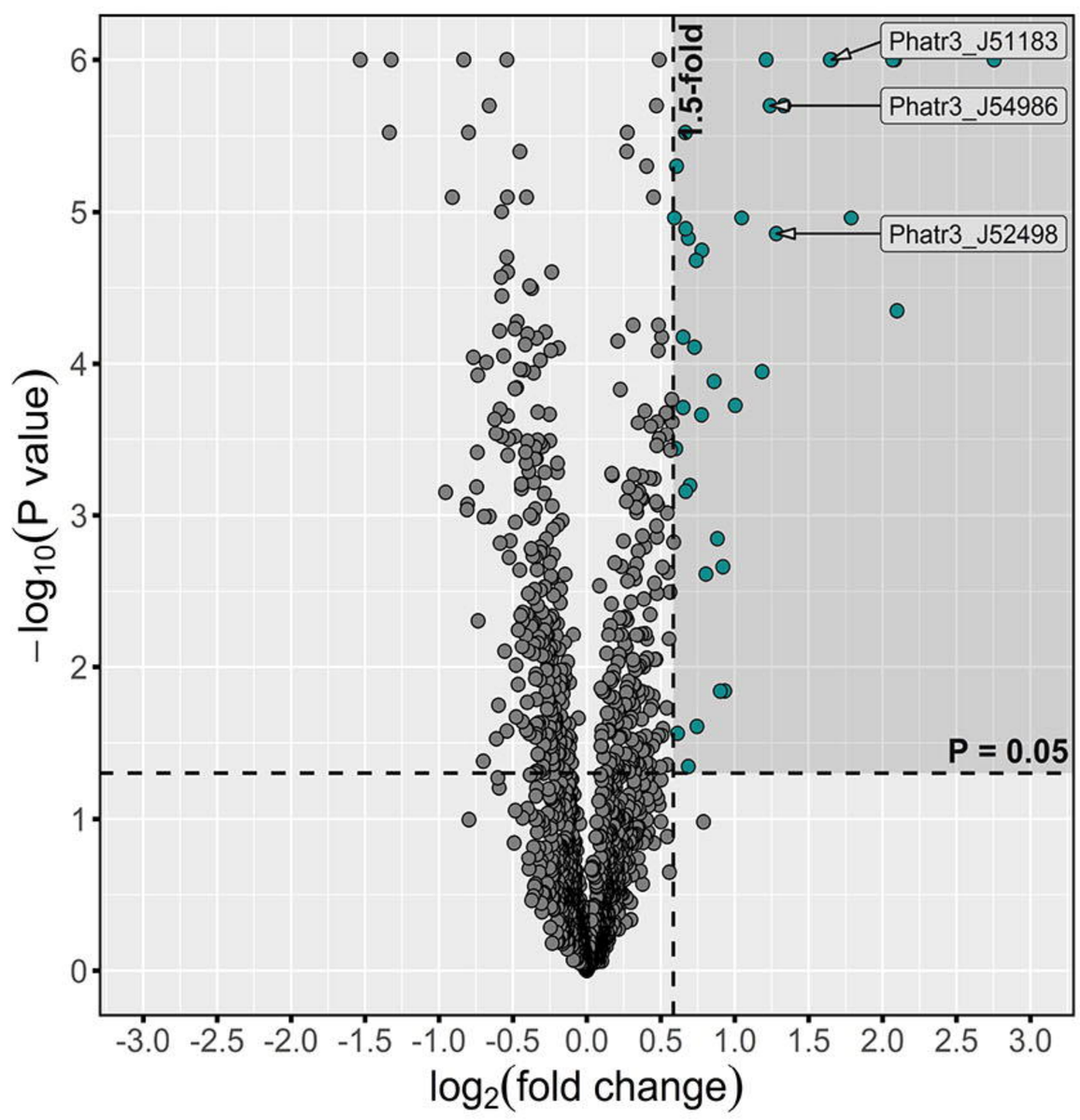

pTF-APEX2
WT
1 micron
1 micron
1 micron
1 micron

A



B

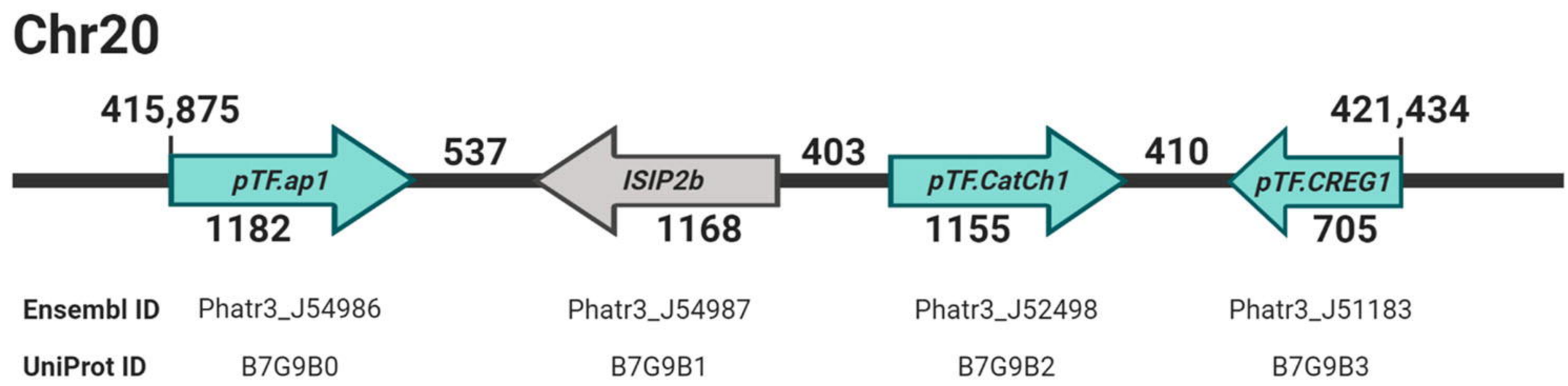




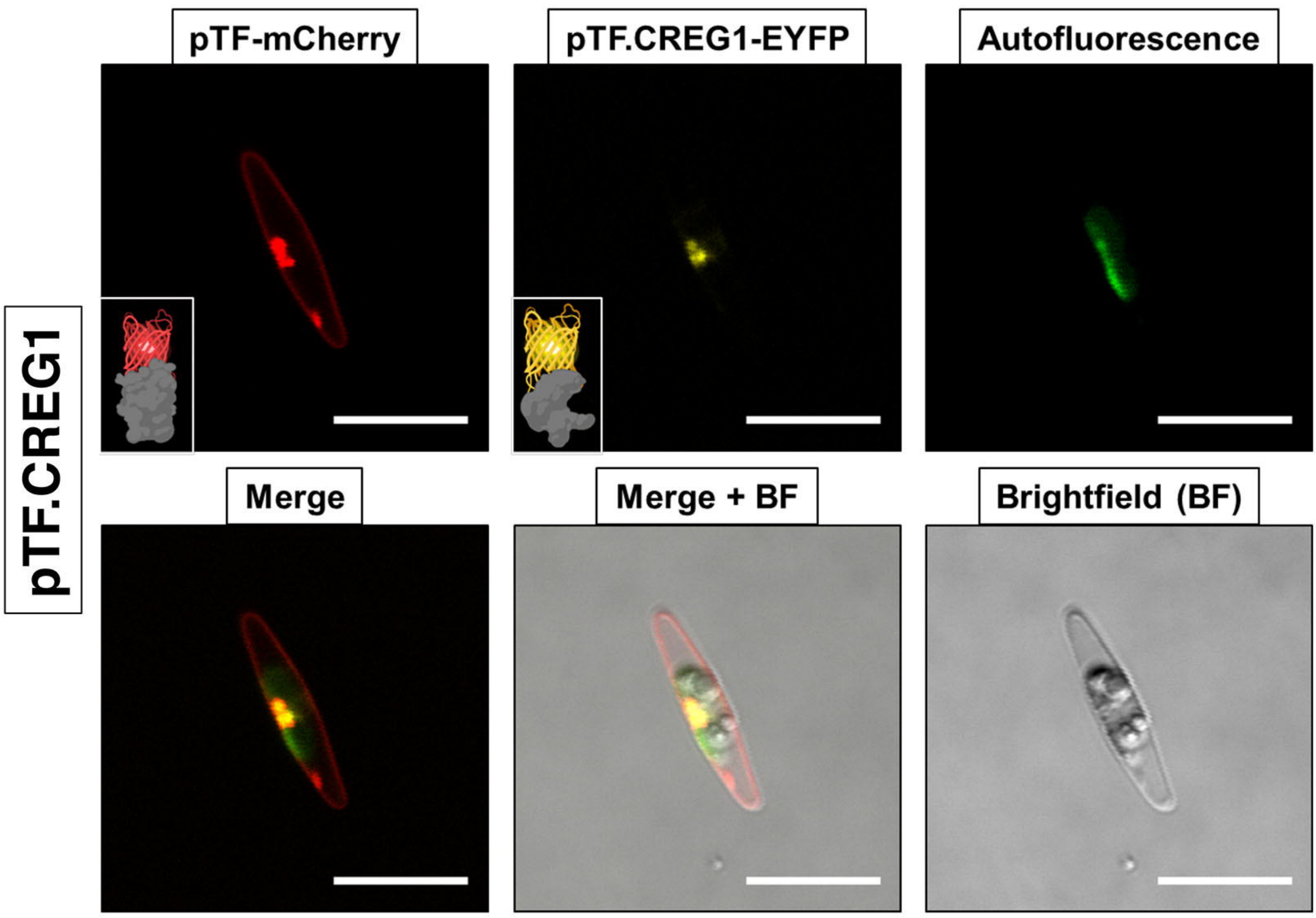
C

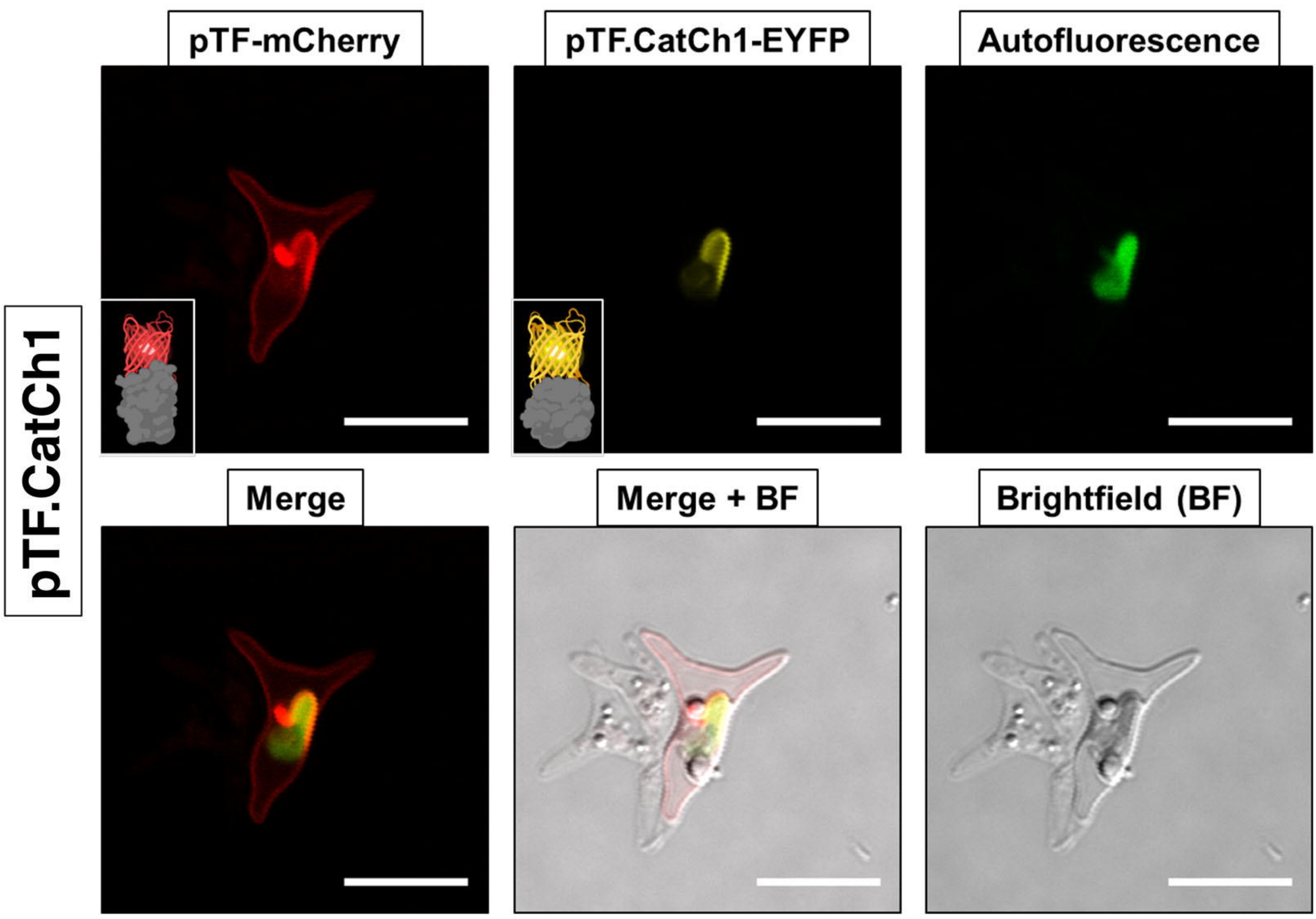
UniProt ID	APEX2/WT ratio	Ensembl ID	Iron-sensitive transcript	Chloroplast localization/ confidence	Annotation
B7GE67	6.75	Phatr3_J41423	yes	yes/low*	fructose-biphosphate aldolase FBAC5
B7GA90	4.20	Phatr3_J55031	yes	yes/low**	iron starvation induced protein (ISIP) 1
B5Y488	3.46	Phatr3_J46929	yes	no	siderophore-bound 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

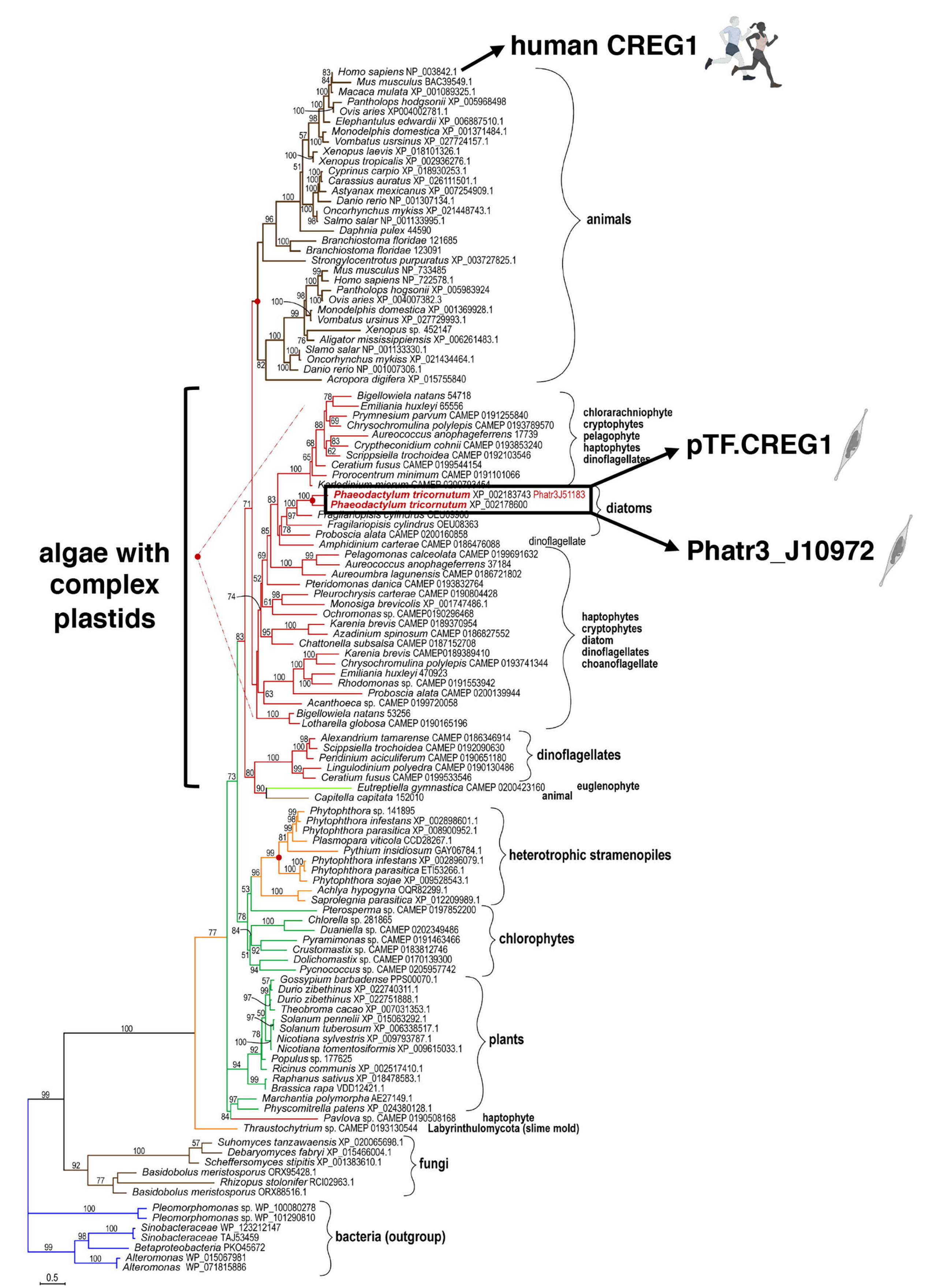
A

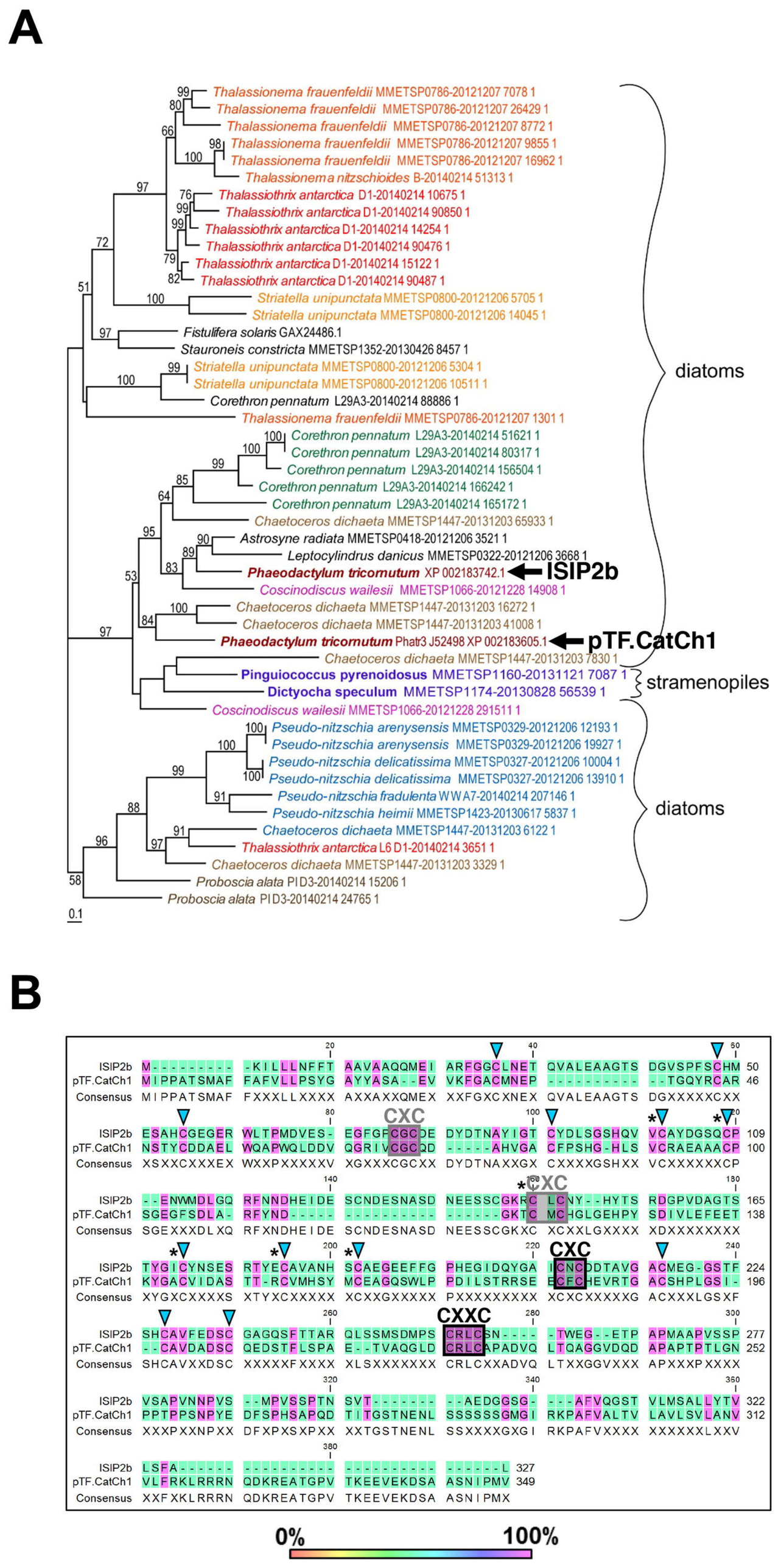

Chr20
415,875
421,434
537
403
410
pTF.ap1
ISIP2b
pTF.CatCh1
pTF.CREG1
1182
1168
1155
705
Ensembl ID
Phatr3_J54986
Phatr3_J54987
Phatr3_J52498
Phatr3_J51183
UniProt ID
B7G9B0
B7G9B1
B7G9B2
B7G9B3

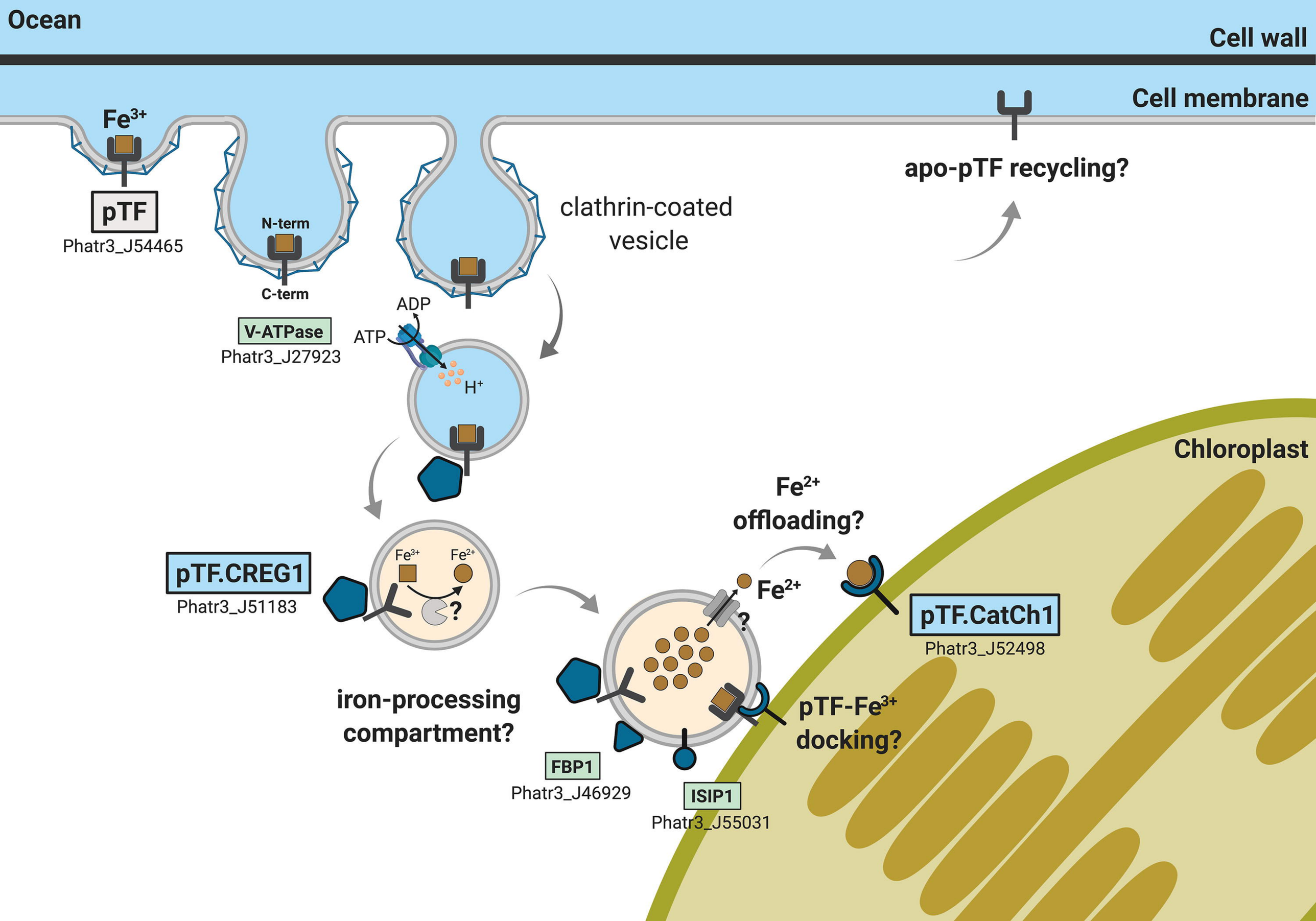
B


pTF.CREG1
pTF-mCherry
pTF.CREG1-EYFP
Autofluorescence
Merge
Merge + BF
Brightfield (BF)


pTF.CatCh1
pTF-mCherry
pTF.CatCh1-EYFP
Autofluorescence
Merge
Merge + BF
Brightfield (BF)


human CREG1
Homo sapiens NP_003842.1
Mus musculus BAC39549.1
Macaca mulata XP_001089325.1
Pantholops hodgsonii XP_005968498
Ovis aries XP004002781.1
Elephantulus edwardii XP_006887510.1
Monodelphis domestica XP_001371484.1
Vombatus usrsinus XP_027724157.1
Xenopus laevis XP_018101326.1
Xenopus tropicalis XP_002936276.1
Cyprinus carpio XP_018930253.1
Carassius auratus XP_026111501.1
Astyanax mexicanus XP_007254909.1
Danio rerio NP_001307134.1
Oncorhynchus mykiss XP_021448743.1
Salmo salar NP_001133995.1
Daphnia pulex 44590
Branchiostoma floridae 121685
Branchiostoma floridae 123091
Strongylocentrotus purpuratus XP_003727825.1
Mus musculus NP_733485
Homo sapiens NP_722578.1
Pantholops hogsonii XP_005983924
Ovis aries XP_004007382.3
Monodelphis domestica XP_001369928.1
Vombatus ursinus XP_027729993.1
Xenopus sp. 452147
Aligator mississippiensis XP_006261483.1
Slamo salar NP_001133330.1
Oncorhynchus mykiss XP_021434464.1
Danio rerio NP_001007306.1
Acropora digifera XP_015755840
animals
algae with complex plastids
Bigellowiela natans 54718
Emiliania huxleyi 65556
Prymnesium parvum CAMEP 0191255840
Chrysochromulina polylepis CAMEP 0193789570
Aureococcus anophagefferens 17739
Cryptheconidium cohnii CAMEP 0193853240
Scrippsiella trochoidea CAMEP 0192103546
Ceratium fusus CAMEP 0199544154
Prorocentrum minimum CAMEP 0191101066
chlorarachniophyte
cryptophytes
pelagophyte
haptophytes
dinoflagellates
pTF.CREG1
Phaeodactylum tricornutum XP_002183743 Phatr3J51183
Phaeodactylum tricornutum XP_002178600
Fragilariopisis cylindrus OEU08363
Proboscia alata CAMEP 0200160858
diatoms
Phatr3_J10972
Amphidinium carterae CAMEP 0186476088
dinoflagellate
Pelagomonas calceolata CAMEP 0199691632
Aureococcus anophagefferens 37184
Aureoumbra lagunensis CAMEP 0186721802
Pteridomonas danica CAMEP 0193832764
Pleurochrysis carterae CAMEP 0190804428
Monosiga brevicolis XP_001747486.1
Ochromonas sp. CAMEP0190296468
Karenia brevis CAMEP 0189370954
Azadinium spinosum CAMEP 0186827552
Chattonella subsalsa CAMEP 0187152708
Karenia brevis CAMEP0189389410
Chrysochromulina polylepis CAMEP 0193741344
Emiliania huxleyi 470923
Rhodomonas sp. CAMEP 0191553942
Proboscia alata CAMEP 0200139944
Acanthoeca sp. CAMEP 0199720058
Bigellowiela natans 53256
Lotharella globosa CAMEP 0190165196
haptophytes
cryptophytes
diatom
dinoflagellates
choanoflagellate
Alexandrium tamarense CAMEP 0186346914
Scippsiella trochoidea CAMEP 0192090630
Peridinium aciculiferum CAMEP 0190651180
Lingulodinium polyedra CAMEP 0190130486
Ceratium fusus CAMEP 0199533546
dinoflagellates
Eutreptiella gymnastica CAMEP 0200423160
euglenophyte
Capitella capitata 152010
animal
Phytophthora sp. 141895
Phytophthora infestans XP_002898601.1
Phytophthora parasitica XP_008900952.1
Plasmopara viticola CCD28267.1
Pythium insidiosum GAY06784.1
Phytophthora infestans XP_002896079.1
Phytophthora parasitica ETI53266.1
Phytophthora sojae XP_009528543.1
Achlya hypogyna OQR82299.1
Saprolegnia parasitica XP_012209989.1
heterotrophic stramenopiles
Pterosperma sp. CAMEP 0197852200
Chlorella sp. 281865
Duaniella sp. CAMEP 0202349486
Pyramimonas sp. CAMEP 0191463466
Crustomastix sp. CAMEP 0183812746
Dolichomastix sp. CAMEP 0170139300
Pycnococcus sp. CAMEP 0205957742
chlorophytes
Gossypium barbadense PPS00070.1
Durio zibethinus XP_022740311.1
Durio zibethinus XP_022751888.1
Theobroma cacao XP_007031353.1
Solanum pennelii XP_015063292.1
Solanum tuberosum XP_006338517.1
Nicotiana sylvestris XP_009793787.1
Nicotiana tomentosiformis XP_009615033.1
Populus sp. 177625
Ricinus communis XP_002517410.1
Raphanus sativus XP_018478583.1
Brassica rapa VDD12421.1
Marchantia polymorpha AE27149.1
Physcomitrella patens XP_024380128.1
plants
Pavlova sp. CAMEP 0190508168
haptophyte
Thraustochytrium sp. CAMEP 0193130544
Labyrinthulomycota (slime mold)
Suhomyces tanzawaensis XP_020065698.1
Debaryomyces fabryi XP_015466004.1
Scheffersomyces stipitis XP_001383610.1
Basidobolus meristosporus ORX95428.1
Rhizopus stolonifer RCI02963.1
Basidobolus meristosporus ORX88516.1
fungi
Pleomorphomonas sp. WP_100080278
Pleomorphomonas sp. WP_101290810
Sinobacteraceae WP_123212147
Sinobacteraceae TAJ53459
Betaproteobacteria PKO45672
Alteromonas WP_015067981
Alteromonas WP_071815886
bacteria (outgroup)
0.5


A
Thalassionema frauenfeldii MMETSP0786-20121207 7078 1
Thalassionema frauenfeldii MMETSP0786-20121207 26429 1
Thalassionema frauenfeldii MMETSP0786-20121207 8772 1
Thalassionema frauenfeldii MMETSP0786-20121207 9855 1
Thalassionema frauenfeldii MMETSP0786-20121207 16962 1
Thalassionema nitzschioides B-20140214 51313 1
Thalassiothrix antarctica D1-20140214 10675 1
Thalassiothrix antarctica D1-20140214 90850 1
Thalassiothrix antarctica D1-20140214 14254 1
Thalassiothrix antarctica D1-20140214 90476 1
Thalassiothrix antarctica D1-20140214 15122 1
Thalassiothrix antarctica D1-20140214 90487 1
Striatella unipunctata MMETSP0800-20121206 5705 1
Striatella unipunctata MMETSP0800-20121206 14045 1
Fistulifera solaris GAX24486.1
Stauroneis constricta MMETSP1352-20130426 8457 1
Striatella unipunctata MMETSP0800-20121206 5304 1
Striatella unipunctata MMETSP0800-20121206 10511 1
Corethron pennatum L29A3-20140214 88886 1
Thalassionema frauenfeldii MMETSP0786-20121207 1301 1
Corethron pennatum L29A3-20140214 51621 1
Corethron pennatum L29A3-20140214 80317 1
Corethron pennatum L29A3-20140214 156504 1
Corethron pennatum L29A3-20140214 166242 1
Corethron pennatum L29A3-20140214 165172 1
Chaetoceros dichaeta MMETSP1447-20131203 65933 1
Astrosyne radiata MMETSP0418-20121206 3521 1
Leptocylindrus danicus MMETSP0322-20121206 3668 1
Phaeodactylum tricornutum XP 002183742.1
ISIP2b
Coscinodiscus wailesii MMETSP1066-20121228 14908 1
Chaetoceros dichaeta MMETSP1447-20131203 16272 1
Chaetoceros dichaeta MMETSP1447-20131203 41008 1
Phaeodactylum tricornutum Phatr3 J52498 XP 002183605.1
pTF.CatCh1
Chaetoceros dichaeta MMETSP1447-20131203 7830 1
Pinguiococcus pyrenoidosus MMETSP1160-20131121 7087 1
Dictyocha speculum MMETSP1174-20130828 56539 1
Coscinodiscus wailesii MMETSP1066-20121228 291511 1
Pseudo-nitzschia arenysensis MMETSP0329-20121206 12193 1
Pseudo-nitzschia arenysensis MMETSP0329-20121206 19927 1
Pseudo-nitzschia delicatissima MMETSP0327-20121206 10004 1
Pseudo-nitzschia delicatissima MMETSP0327-20121206 13910 1
Pseudo-nitzschia fradulenta WWA7-20140214 207146 1
Pseudo-nitzschia heimii MMETSP1423-20130617 5837 1
Chaetoceros dichaeta MMETSP1447-20131203 6122 1
Thalassiothrix antarctica L6 D1-20140214 3651 1
Chaetoceros dichaeta MMETSP1447-20131203 3329 1
Proboscia alata PID3-20140214 15206 1
Proboscia alata PID3-20140214 24765 1
0.1
diatoms
stramenopiles
diatoms
B
ISIP2b
pTF.CatCh1
Consensus
CXC
CXXC
0%
100%

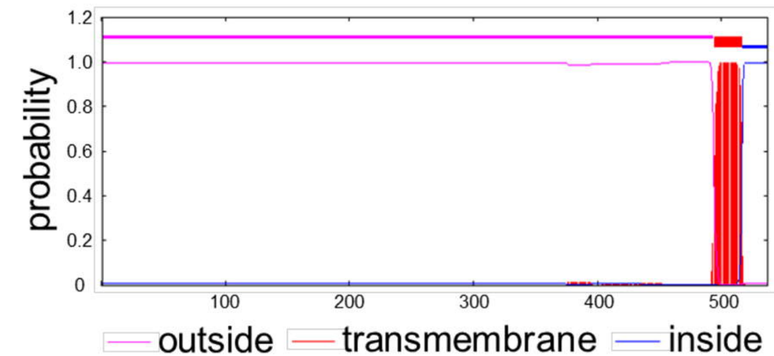

Ocean
Cell wall
Cell membrane
Fe3+
pTF
Phatr3_J54465
N-term
C-term
V-ATPase
Phatr3_J27923
ADP
ATP
H+
clathrin-coated vesicle
apo-pTF recycling?
Chloroplast
Fe2+ offloading?
Fe2+
?
pTF.CatCh1
Phatr3_J52498
pTF.CREG1
Phatr3_J51183
Fe3+
Fe2+
?
iron-processing compartment?
pTF-Fe3+ docking?
FBP1
Phatr3_J46929
ISIP1
Phatr3_J55031

A

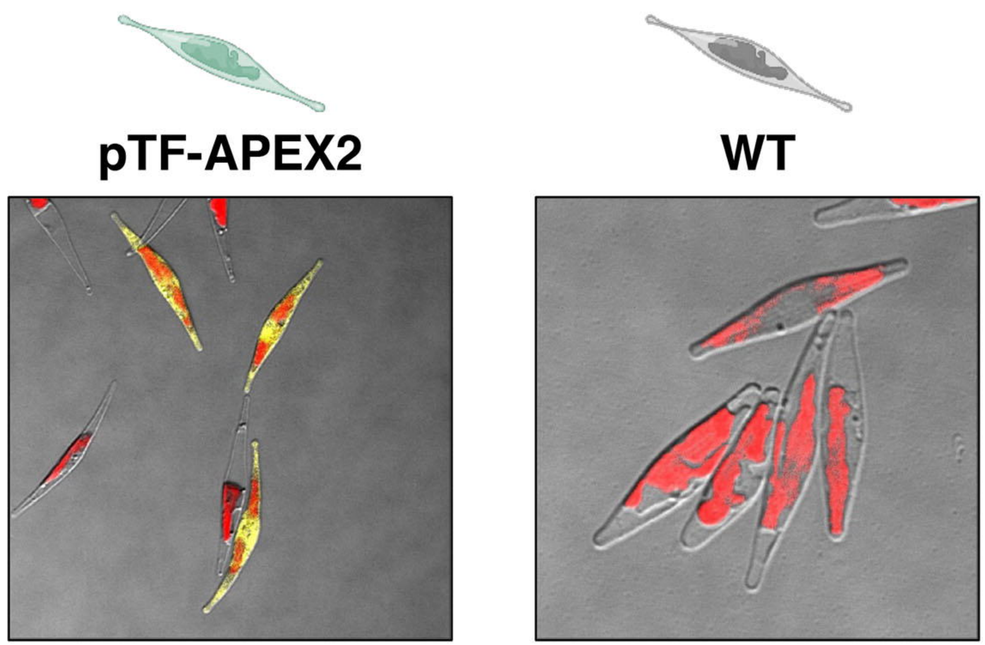
Signal peptide

```
MFSFKSFVFAALLSSCEASVRARKLSNELIAGFEPRTVVTDHNAIDLDQAKIEDLLATPSTDTFSSARDVYEQGSHS
KSFAVLTLDPALTVAVPEGTVISGKNEDGSTVSGTALKSYSVGDTTIEIQYSTGNTQATYVDCQVGGNPDPNTSGCF
AANGTVTISGSSGSLPYSYDPLVNNDNGRTIQGFSTAAETRFRPSGGGELFPDFQKFFDYYGTLTYADEWVQSAFAR
RSTSFSKGDADFSKYGDDGILQAVKKGTAYISIAMYVIRELEDALDDCDRGCDTADCNDDAVHALDEAVAFWTGSLE
GKDGSGSGVLMYGLADTRCINFKTCGRNGDEASGKSKVNFEIFRHFDTMQRELTNKRCNSARATKEKIVPWMIVPLI
QGVLRYAYIIENDGFTEKAEAEGATFAAAVLPYVASCNAEDAKTIYDNMRVGLGGGASYSTVFDAFVRQSSCMGISC
SDIGGYWNEAQGAYEEGAGPCGGSGSGDDGANVGLAVGLSIGGLLVLVLLGVVLRRRRSASGSTVEFKDTGSTPI*
```

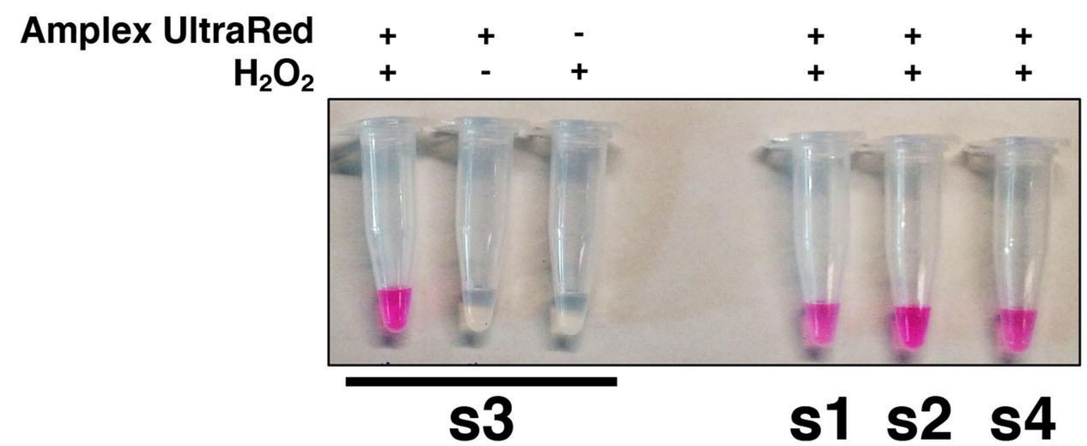
Transmembrane domain **Tail domain**



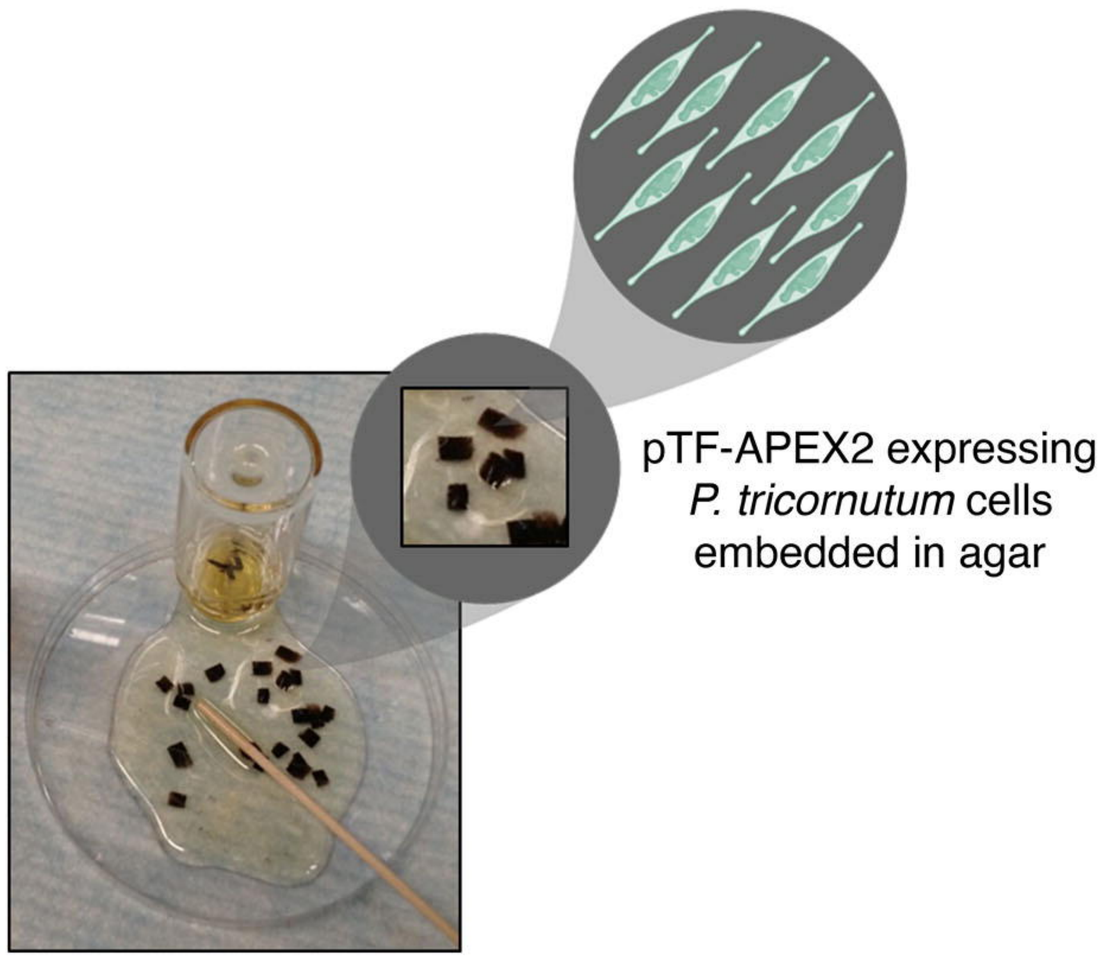
B



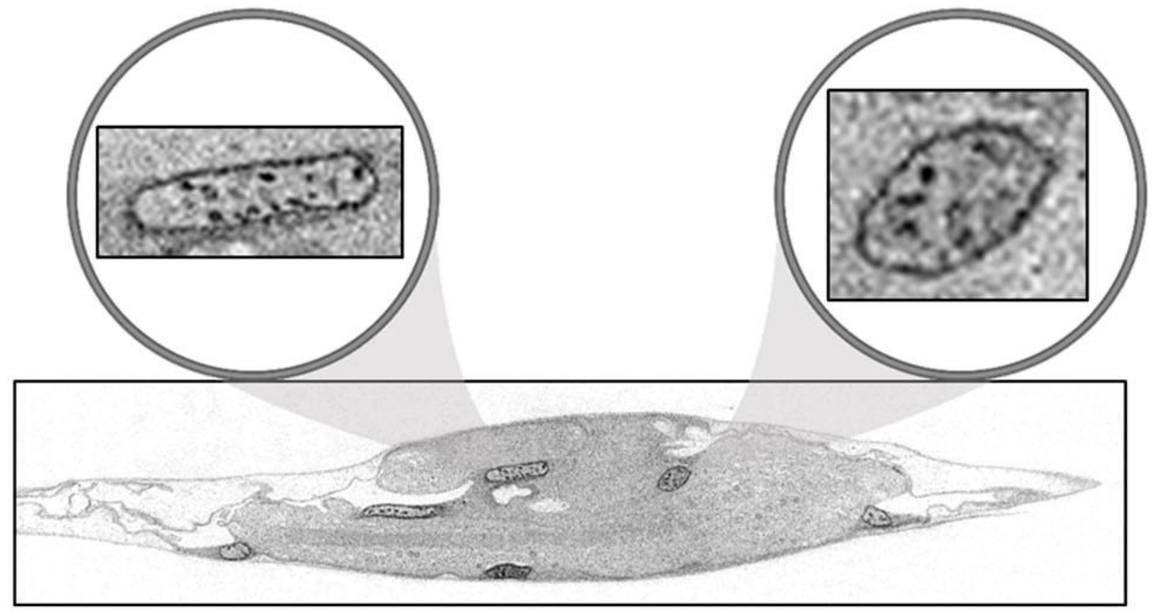
C



A

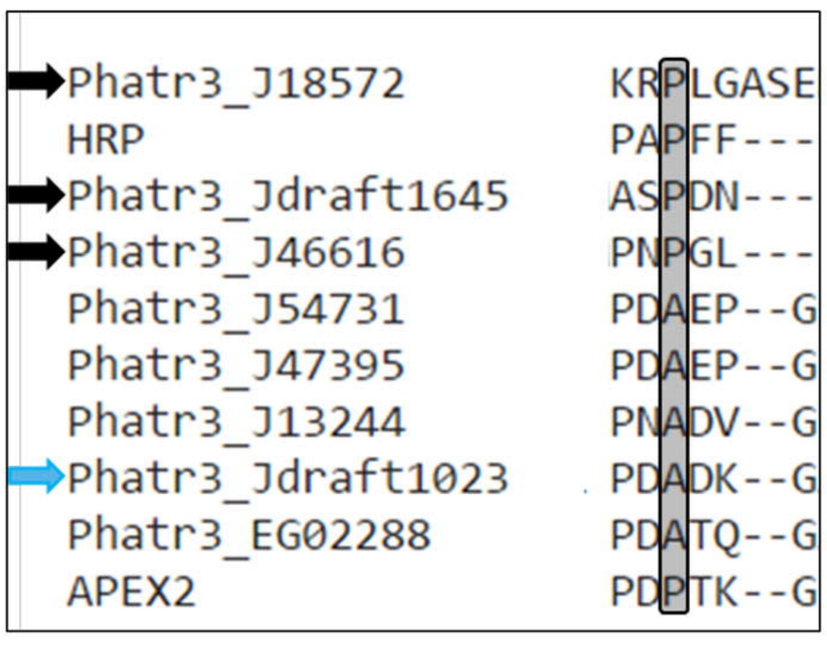


B



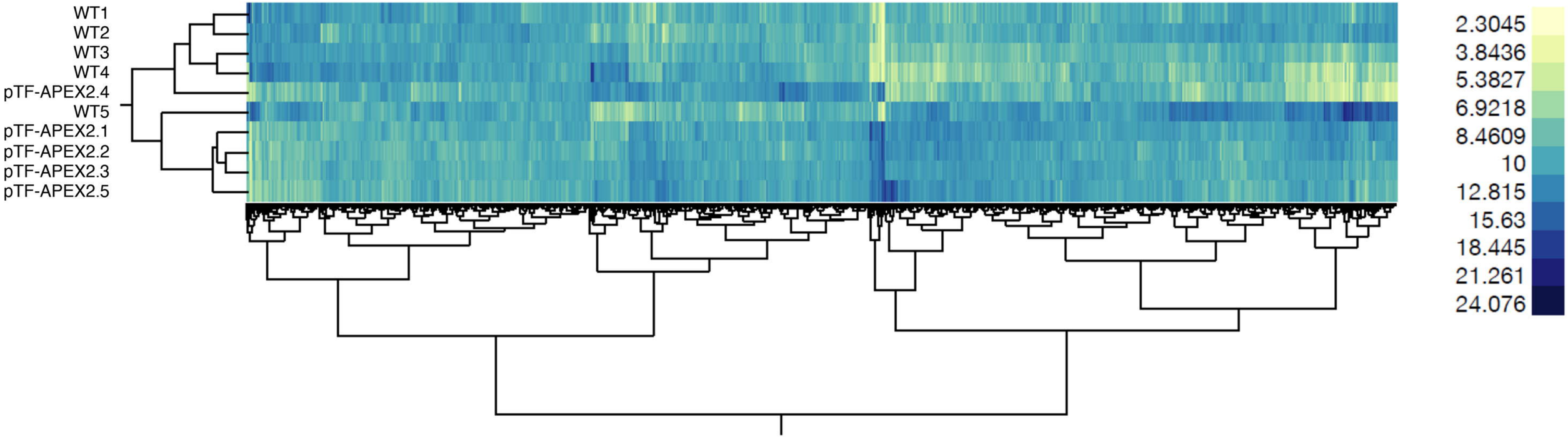
C

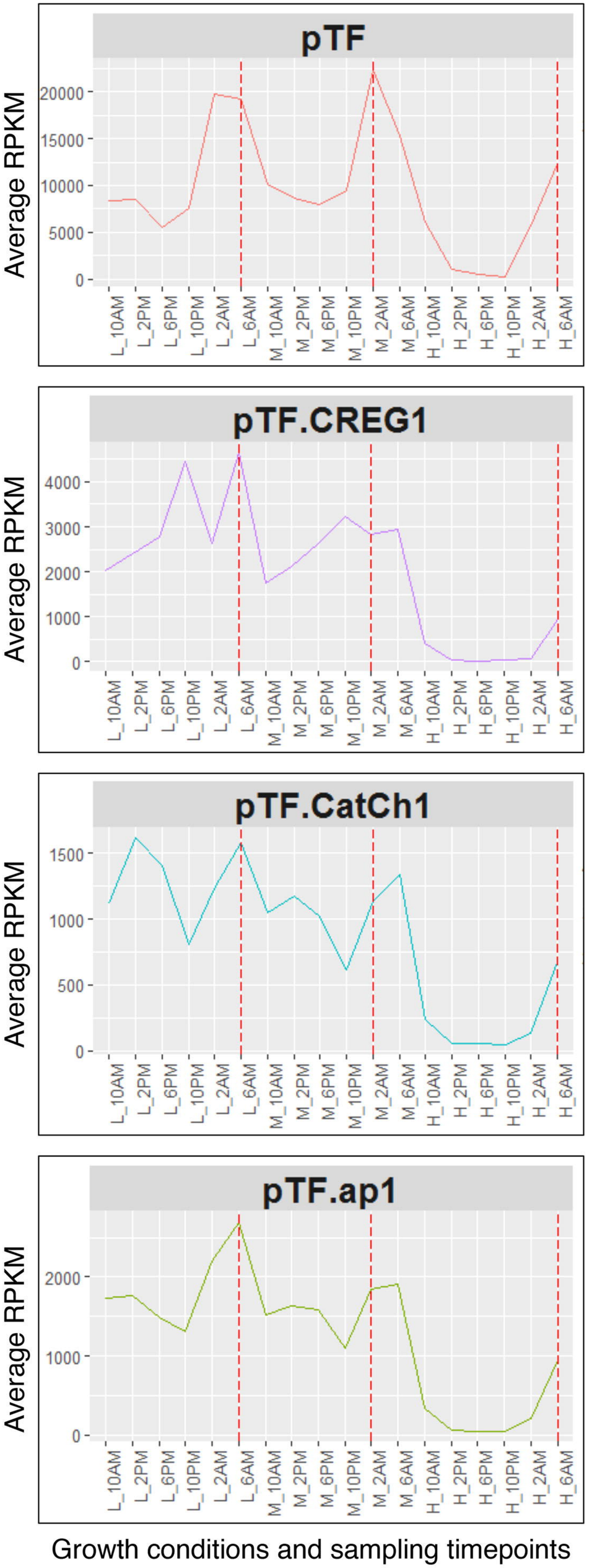
D

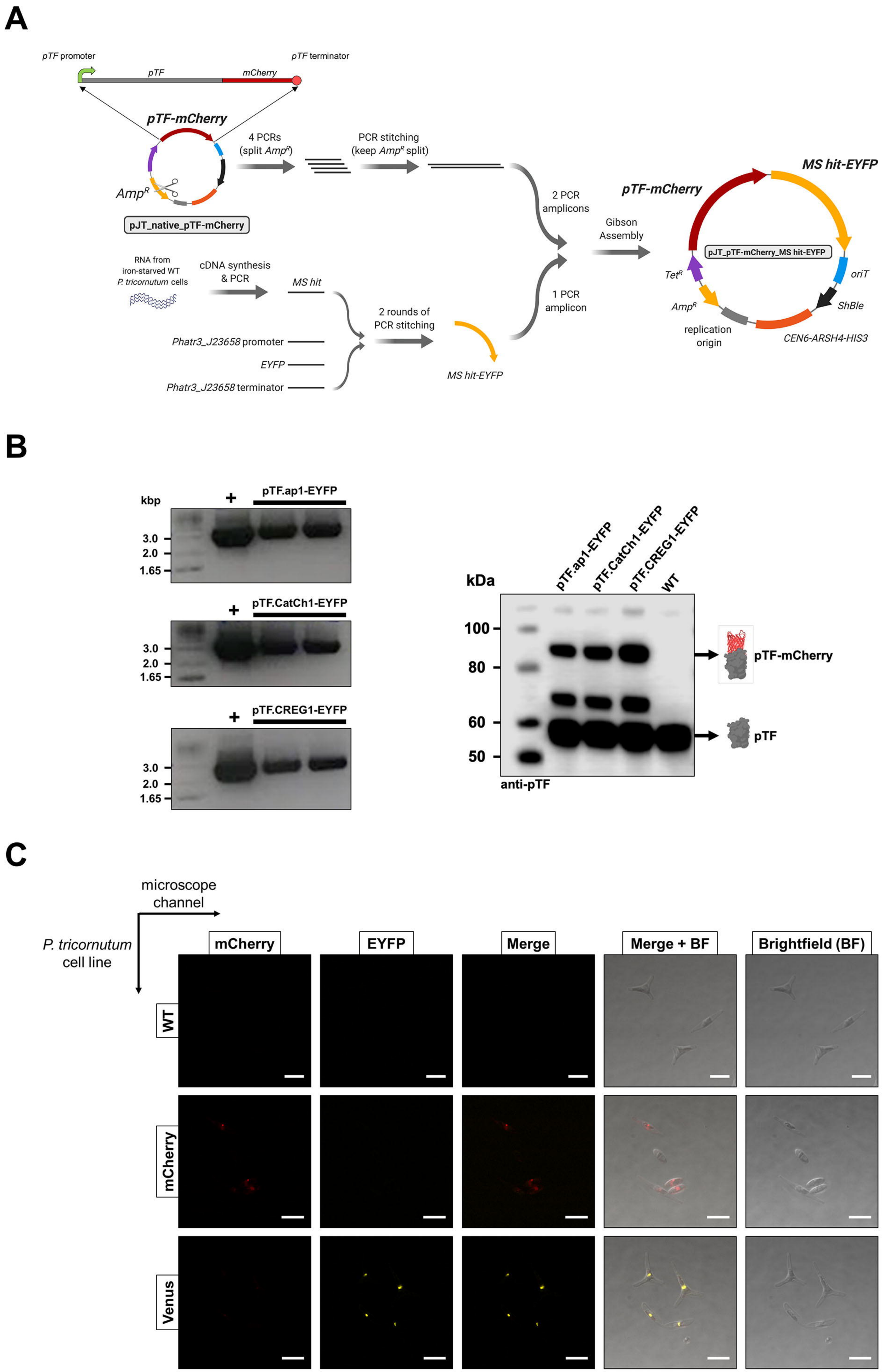


E

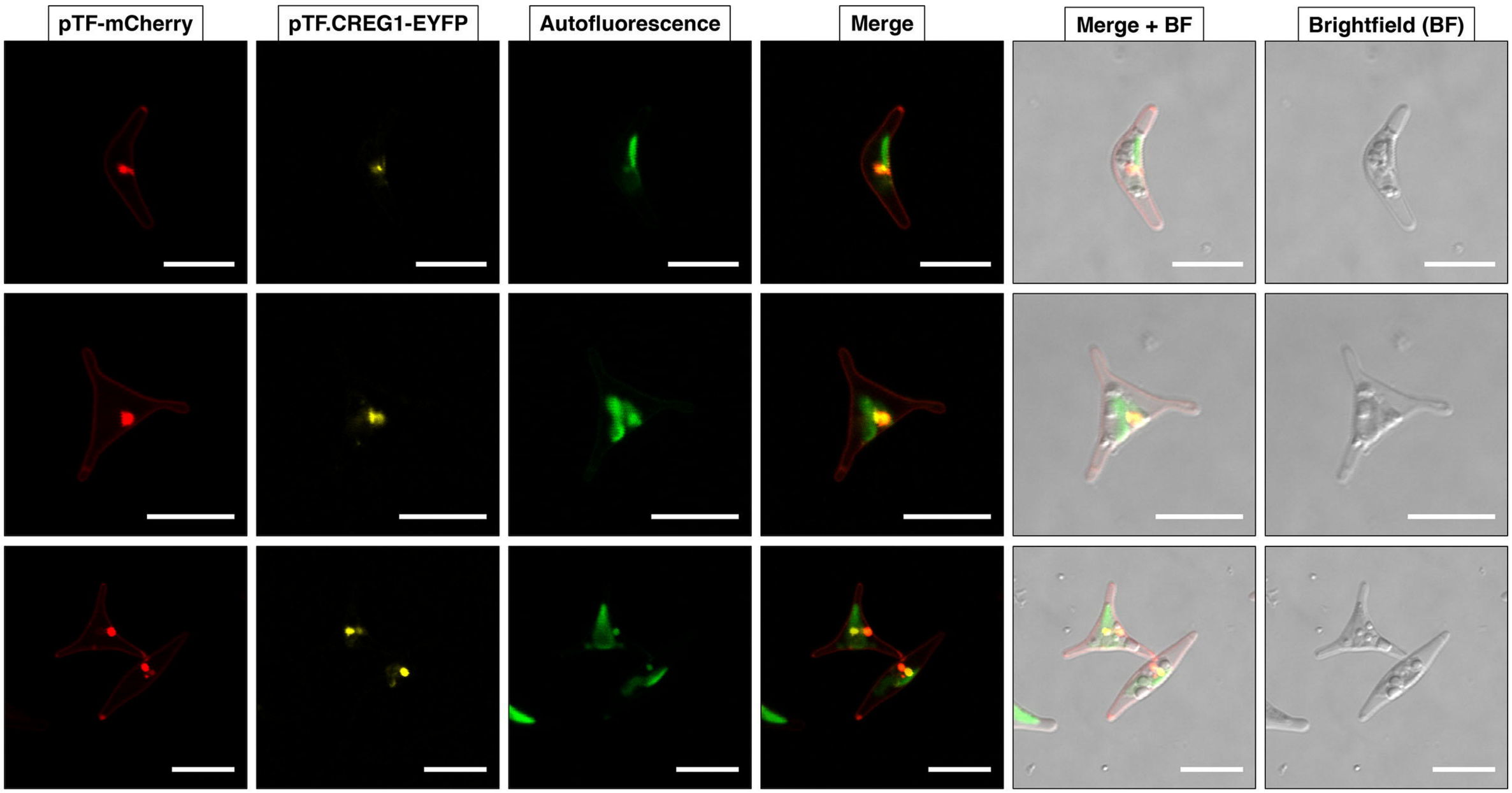
Ensembl ID	Peroxidase class[1]	Position 134 relative to APEX2[2]	Predicted localization[3]	Number of transmembrane domains[4]
Phatr3_J18572	I	**Pro**	cytosol	/
Phatr3_Jdraft1645	I	**Pro**	secretory pathway	1
Phatr3_J46616	I	**Pro**	secretory pathway	/
Phatr3_J54731	I	Ala	cytosol	/
Phatr3_J47395	I	Ala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


WT1
WT2
WT3
WT4
pTF-APEX2.4
WT5
pTF-APEX2.1
pTF-APEX2.2
pTF-APEX2.3
pTF-APEX2.5
2.3045
3.8436
5.3827
6.9218
8.4609
10
12.815
15.63
18.445
21.261
24.076

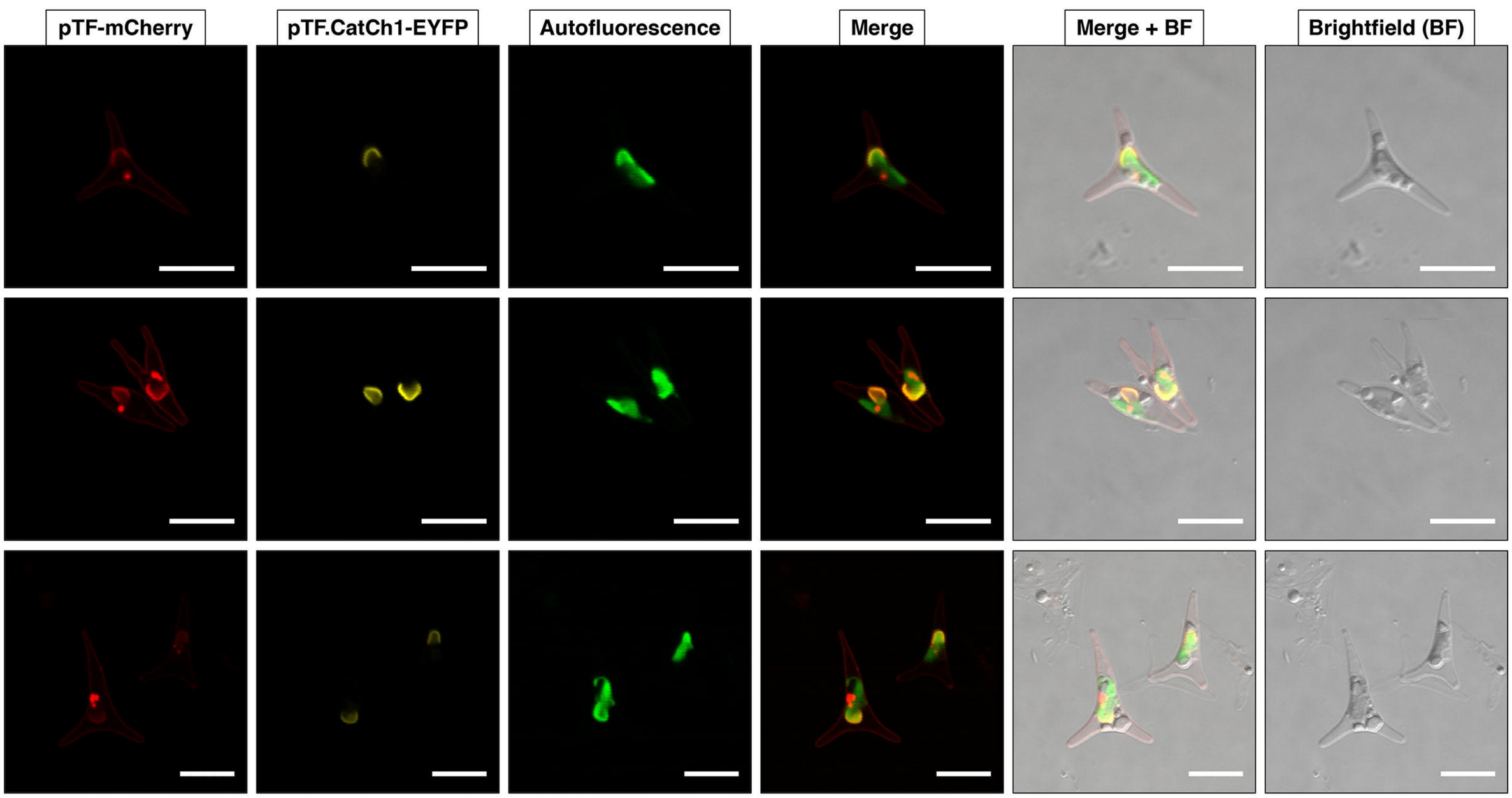

pTF
pTF.CREG1
pTF.CatCh1
pTF.ap1
Average RPKM
L_10AM
L_2PM
L_6PM
L_10PM
L_2AM
L_6AM
M_10AM
M_2PM
M_6PM
M_10PM
M_2AM
M_6AM
H_10AM
H_2PM
H_6PM
H_10PM
H_2AM
H_6AM
Growth conditions and sampling timepoints

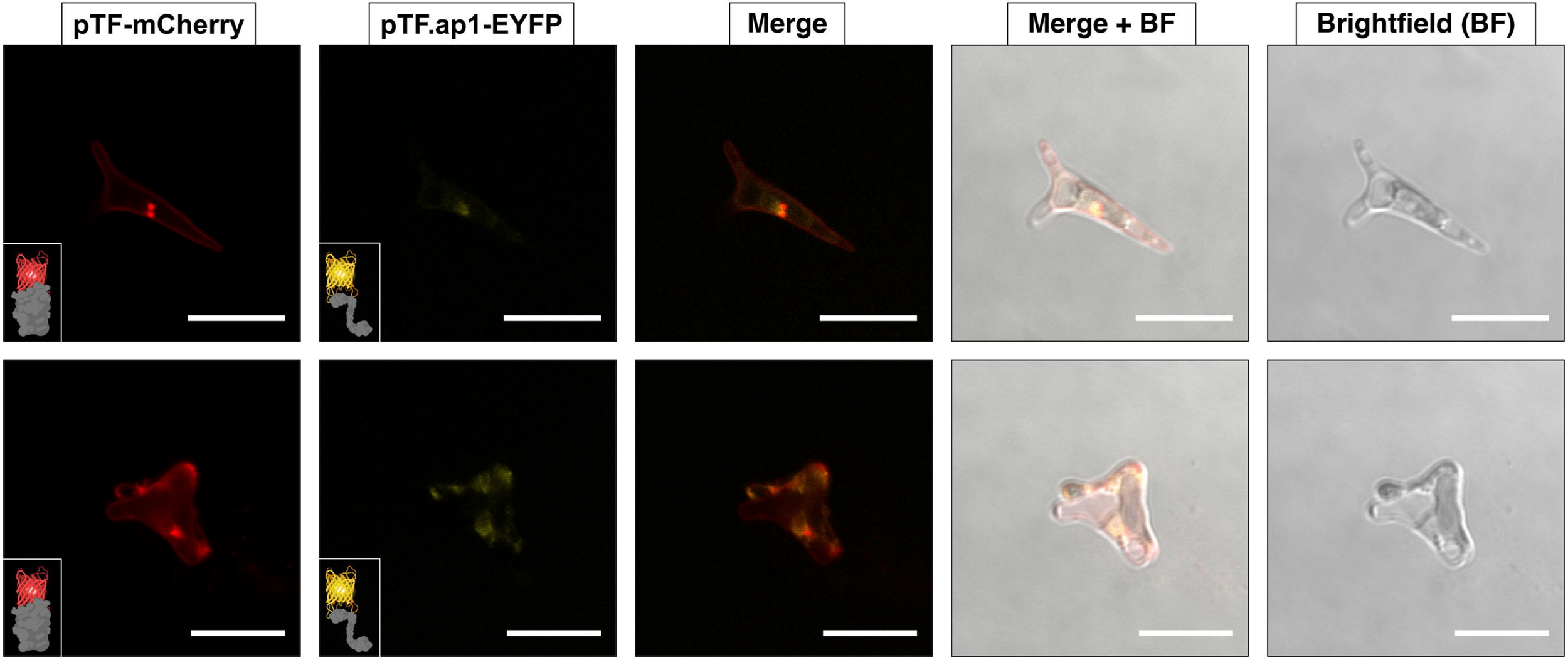

A
pTF promoter
pTF terminator
pTF
mCherry
pTF-mCherry
Amp^R
pJT_native_pTF-mCherry
4 PCRs
(split Amp^R)
PCR stitching
(keep Amp^R split)
2 PCR
amplicons
Gibson
Assembly
MS hit-EYFP
pTF-mCherry
pJT_pTF-mCherry_MS hit-EYFP
Tet^R
oriT
Amp^R
ShBle
replication
origin
CEN6-ARSH4-HIS3
RNA from
iron-starved WT
P. tricornutum cells
cDNA synthesis
& PCR
MS hit
Phatr3_J23658 promoter
EYFP
Phatr3_J23658 terminator
2 rounds of
PCR stitching
MS hit-EYFP
1 PCR
amplicon
B
kbp
+
pTF.ap1-EYFP
3.0
2.0
1.65
pTF.CatCh1-EYFP
pTF.CREG1-EYFP
kDa
pTF.ap1-EYFP
pTF.CatCh1-EYFP
pTF.CREG1-EYFP
WT
100
80
60
50
anti-pTF
pTF-mCherry
pTF
C
microscope
channel
P. tricornutum
cell line
mCherry
EYFP
Merge
Merge + BF
Brightfield (BF)
WT
mCherry
Venus

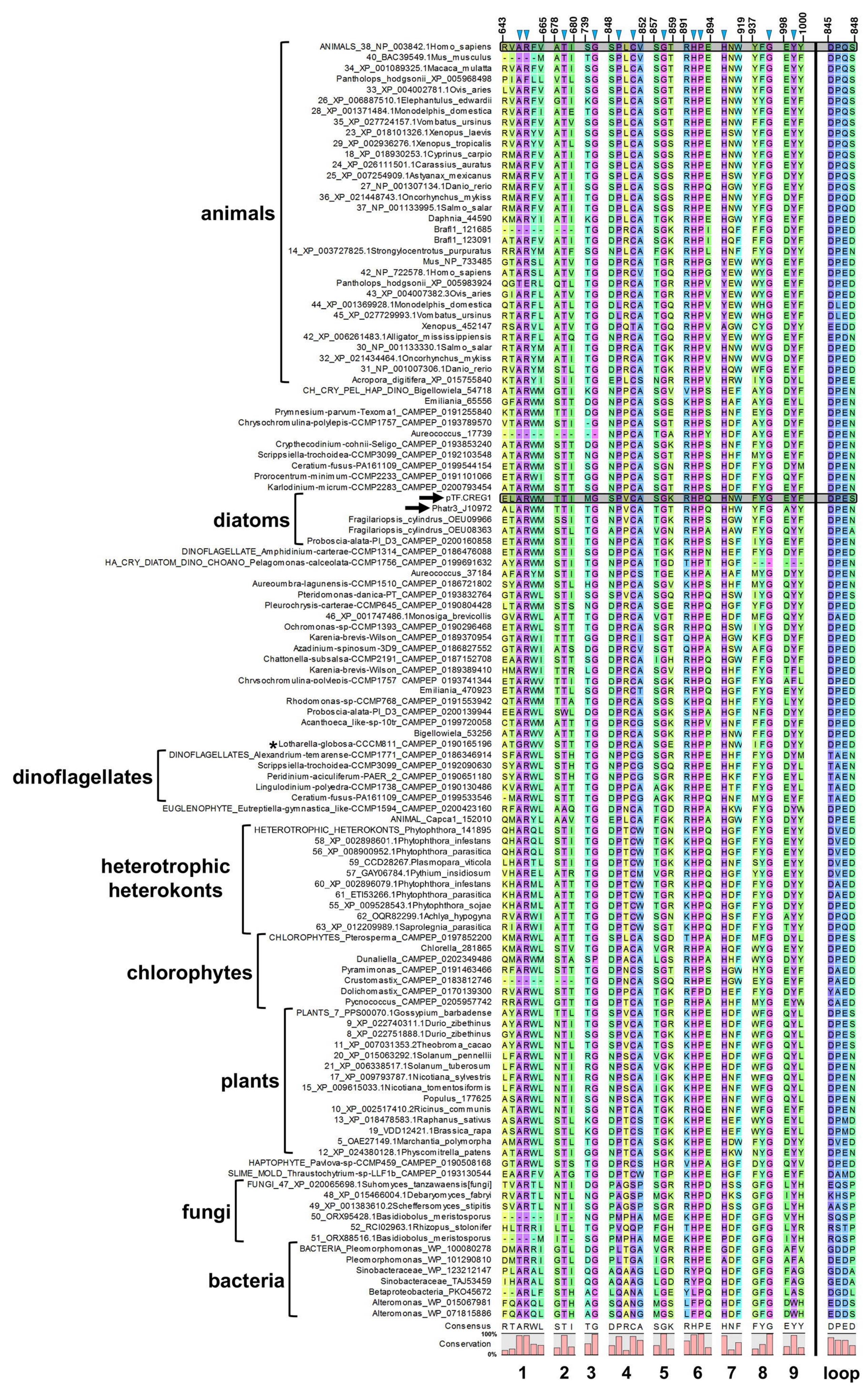
A


pTF-mCherry
pTF.CREG1-EYFP
Autofluorescence
Merge
Merge + BF
Brightfield (BF)

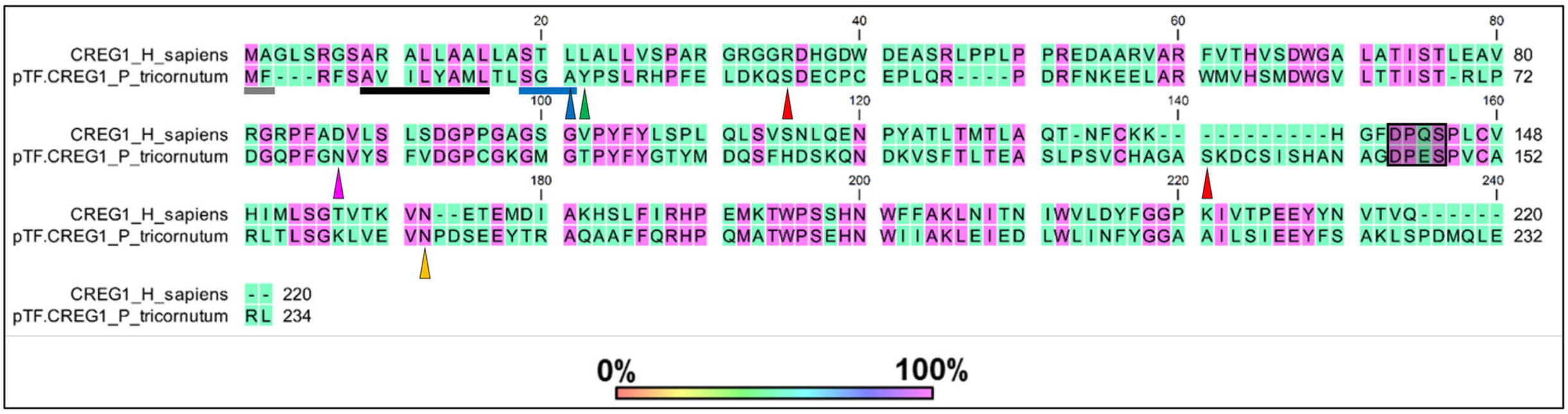
B


pTF-mCherry
pTF.CatCh1-EYFP
Autofluorescence
Merge
Merge + BF
Brightfield (BF)

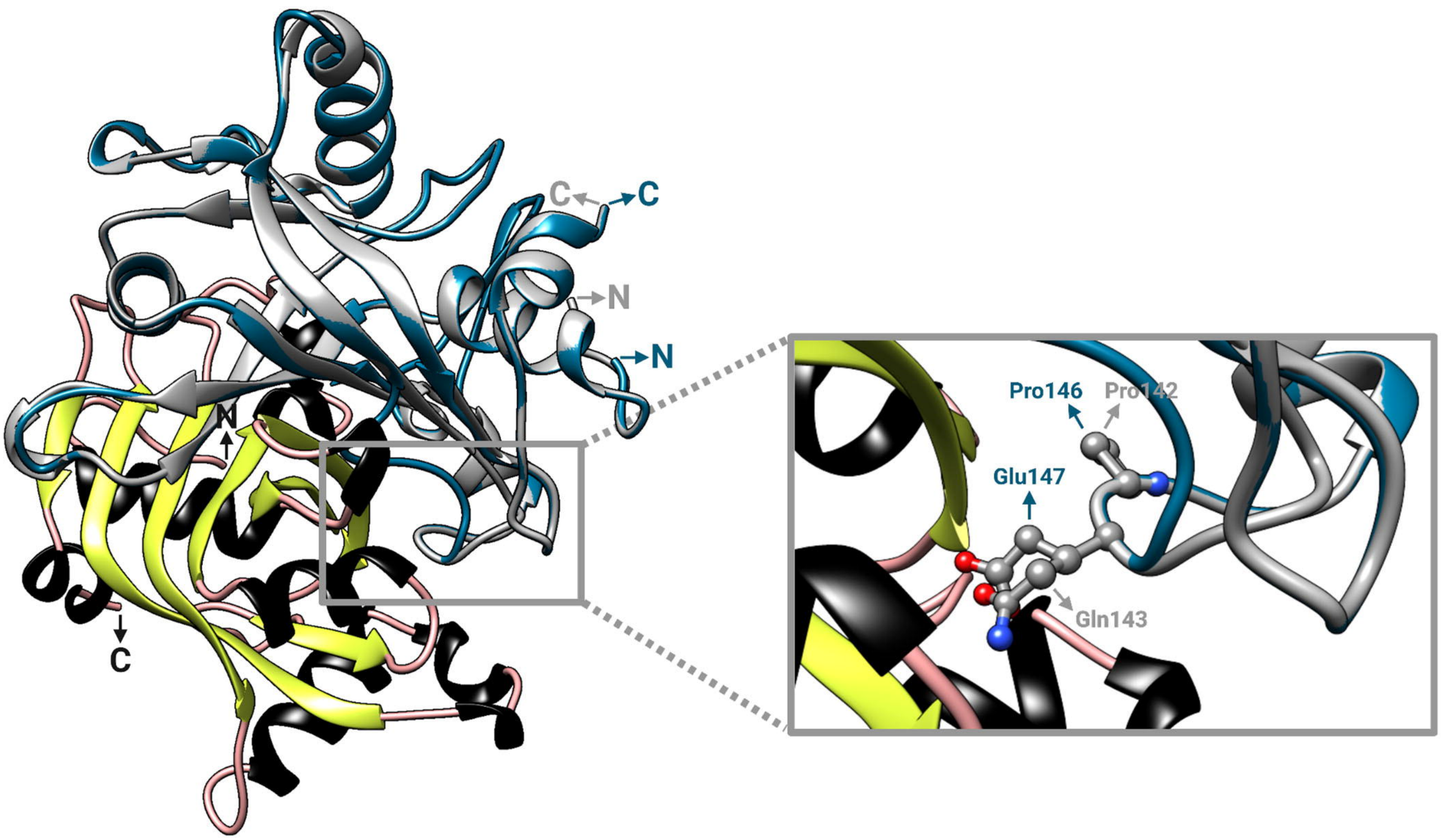

pTF-mCherry
pTF.ap1-EYFP
Merge
Merge + BF
Brightfield (BF)


643 665 678 680 739 848 852 857 859 891 894 919 937 998 1000 845 848
animals
ANIMALS_38_NP_003842.1Homo_sapiens RVARFV ATI SG SPLCV SGT RHPE HNW YFG EYY DPQS
40_BAC39549.1Mus_musculus -----M ATI TG SPLCV SGT RHPE HNW YFG EYF DPQS
34_XP_001089325.1Macaca_mulatta RVARFV ATI SG SPLCA SGT RHPE HNW YFG EYY DPQS
Pantholops_hodgsonii_XP_005968498 PIAFLL VTL SG SPLCA SGT RHPE HNW YFG EYY DPQS
33_XP_004002781.1Ovis_aries LVARFV ATI SG SPLCA SGT RHPE HNW YFG EYY DPQS
26_XP_006887510.1Elephantulus_edwardii RVARFV GTI KG SPLCA SGT RHPE HNW YYG EYF DPQS
28_XP_001371484.1Monodelphis_domestica RVARFI ATE TG SPLCA SGT RHPE HNW YFG EYY DPQS
35_XP_027724157.1Vombatus_ursinus RVARFV ATV SG SPLCA SGT RHPE HNW FFG EYF DPQS
23_XP_018101326.1Xenopus_laevis RVARYV ATI SG SPLCA SGS RHPE HSW YFG EYY DPQS
29_XP_002936276.1Xenopus_tropicalis RVARYV ATL SG SPLCA SGS RHPE HNW YFG EYY DPQS
18_XP_018930253.1Cyprinus_carpio RMARFV ATI TG SPLCA SGS RHPE HNW YFG EYY DPQS
24_XP_026111501.1Carassius_auratus RMARFV ATI TG SPLCA SGS RHPE HNW YFG EYY DPQS
25_XP_007254909.1Astyanax_mexicanus RMARFV ATI TG SPLCA SGS RHPE HSW YFG DYY DPQS
27_NP_001307134.1Danio_rerio RMARFV ATI SG SPLCA SGS RHPQ HGW YFG EYY DPQS
36_XP_021448743.1Oncorhynchus_mykiss RMARFV ATI TG DPLCA SGS RHPE HNW YFG DYF DPQD
37_NP_001133995.1Salmo_salar RMARFV ATI TG DPLCA SGS RHPE HNW YFG DYF DPQD
Daphnia_44590 KMARYI ATI KG DPLCA TGK RHPE HGW YFG EYF DPED
Brafl1_121685 ------ --- TG DPRCA SGK RHPI HQF FFG DYF DPED
Brafl1_123091 ATARFV ATI TG DPRCA SGK RHPI HQF FFG DYF DPED
14_XP_003727825.1Strongylocentrotus_purpuratus RRARYM ATF SG NPLCA FGK RHPL HNF FYG DYY DPEN
Mus_NP_733485 GTARSL ATV TG DPRCA TGR RHPG YEW WYG EYF DPED
42_NP_722578.1Homo_sapiens ATARSL ATV TG DPRCV TGQ RHPG YEW WYG EYF DPED
Pantholops_hodgsonii_XP_005983924 QGTERL QTL TG DPRCA TGR RHPV YEW WYG EYF DPED
43_XP_004007382.3Ovis_aries GIARFL ATV TG DPRCA TGR RHPV YEW WYG EYF DPED
44_XP_001369928.1Monodelphis_domestica QTARFL ATL TG DLRCA TGQ RHPV YEW WHG EYF DLED
45_XP_027729993.1Vombatus_ursinus RTARFL ATV TG DLRCA TGQ RHPV YEW WHG EYF DLED
Xenopus_452147 RSARVL ATV TG DPQTA TGQ RHPV AGW CYG DYY EEDD
42_XP_006261483.1Alligator_mississippiensis RTARFL ATQ TG NPRCA TGQ RHPV YEW WYG EYF EPDN
30_NP_001133330.1Salmo_salar RTARYM ATI TG DPRCA TGK RHPV HNW WVG DYF DPED
32_XP_021434464.1Oncorhynchus_mykiss RTARYM ATI TG DPRCA TGK RHPV HNW WVG DYF DPED
31_NP_001007306.1Danio_rerio RVARYM STL TG DPRCA TGK RHPV HQW WFG EYF DPED
Acropora_digitifera_XP_015755840 KTARYI SII TG EPLCS NGR RHPV HRW IYG DYL DPEE
CH_CRY_PEL_HAP_DINO_Bigellowiela_54718 ATARWM GTI KG NPPCA SGV VHPS HEF AYG DYF DPEN
Emiliania_65556 GFARWM STT DG NPPCA SGS KHPS HAF AYG EYL DPEN
Prymnesium-parvum-Texoma1_CAMPEP_0191255840 KTARWM TTI DG NPPCA SGE RHPS HNF EYG DYF DPEN
Chrysochromulina-polylepis-CCMP1757_CAMPEP_0193789570 VTARWM STI -G NPPCA SGT RHPS HDF AYG DYF DPEN
Aureococcus_17739 ------ --- -- NPPCA TGA RHPY HDF AYG EYF DPEN
Crypthecodinium-cohnii-Seligo_CAMPEP_0193853240 ATARWM STT DG NPPCA SGK RHPS HNF FYG DYF DPEN
Scrippsiella-trochoidea-CCMP3099_CAMPEP_0192103548 ATARWM STT NG NPPCA SGT RHPS HHF MYG EYF DPEN
Ceratium-fusus-PA161109_CAMPEP_0199544154 ETARWM STI SG NPPCA SGN RHPS HDF FYG DYM DPEN
Prorocentrum-minimum-CCMP2233_CAMPEP_0191101066 ETARWI STT GG NPPCA SGT RHPS HDF FYG QYF DPEN
Karlodinium-micrum-CCMP2283_CAMPEP_0200793454 ATARWM STT SG NPPCA SGT RHPS HDF FYG EYF DPEN
diatoms
pTF.CREG1 ELARWM TTI MG SPVCA SGK RHPQ HNW FYG EYF DPES
Phatr3_J10972 ALARWM TTI TG NPVCA TGT RHPQ HHW YFG AYY DPEN
Fragilariopsis_cylindrus_OEU09966 ETARWM SSI TG NPVCA TGK RHPS HGW YFG EYF DPEN
Fragilariopsis_cylindrus_OEU08363 ATARWM STL TG APPCA VGN RHPA HAW FYG QYY DPEA
Proboscia-alata-PI_D3_CAMPEP_0200160858 ETARWM STI TG NPPCA TGK RHPS HSF IYG EYF DPEN
DINOFLAGELLATE_Amphidinium-carterae-CCMP1314_CAMPEP_0186476088 ETARWM STI SG DPPCA TGK RHPN HEF FYG DYF DPED
HA_CRY_DIATOM_DINO_CHOANO_Pelagomonas-calceolata-CCMP1756_CAMPEP_0199691632 AYARWM STI TG NPPCA TGD THPT HGF --- --- DPEN
Aureococcus_37184 AFARYM STI TG NPPCS TGE KHPA HAF MYG DYY DPEN
Aureoumbra-lagunensis-CCMP1510_CAMPEP_0186721802 SYARWM STL HG NPPCA SGV KHPS HHF MYG QYY DPEN
Pteridomonas-danica-PT_CAMPEP_0193832764 GTARWL STI SG SPVCA SGQ RHPQ HSW IYG QYY DPES
Pleurochrysis-carterae-CCMP645_CAMPEP_0190804428 LTARWM STS NG DPRCA SGE RHPQ HGF YFG DYY DPED
46_XP_001747486.1Monosiga_brevicollis GVARWL ATT TG DPRCA VGT RHPE HDF MFG QYY DAED
Ochromonas-sp-CCMP1393_CAMPEP_0190296468 ETARWL STT TG DPRCA SGR RHPQ HSW IYG DYY DPED
Karenia-brevis-Wilson_CAMPEP_0189370954 GTARWI TTT GG DPRCI SGT QHPA HGW KFG DYF DPED
Azadinium-spinosum-3D9_CAMPEP_0186827552 GTARWI ATS DG DPRCV SGT QHPA HSW AFG DYF DPED
Chattonella-subsalsa-CCMP2191_CAMPEP_0187152708 EAARWI STT SG DPRCA IGH RHPQ HGW FFG DYF DPED
Karenia-brevis-Wilson_CAMPEP_0189389410 HMARWI TTR LG DPRCA SGV RHPQ HHF FYG TFL DPED
Chrysochromulina-polylepis-CCMP1757_CAMPEP_0193741344 ETARWV TTI TG DPRCA SGK RHPQ HGF FYG AFL DPED
Emiliania_470923 ETARWM TTL SG DPRCT SGR RHPQ HGF FYG EYY DPED
Rhodomonas-sp-CCMP768_CAMPEP_0191553942 QTARWM TTA TG DPRCA SGS KHPQ HHF FYG LYY DPED
Proboscia-alata-PI_D3_CAMPEP_0200139944 EEARWL SWL DG DPRCA SGK SHPA HGF NFG DYY DPED
Acanthoeca_like-sp-10tr_CAMPEP_0199720058 CTARWM ATT TG DPRCG SGK RHPV HNF FFG DYF DPED
Bigellowiela_53256 ATARWV ATT TG DPRCA SGK RHPP HNW FFG DYY DPED
*Lotharella-globosa-CCCM811_CAMPEP_0190165196 ATGRWV STT TG DPRCA SGE RHPP HNW IFG DYY DPED
dinoflagellates
DINOFLAGELLATES_Alexandrium-tamarense-CCMP1771_CAMPEP_0186346914 SFARWL STH TG NPPCG SGR RHPE HHF FYG DYM TAEN
Scrippsiella-trochoidea-CCMP3099_CAMPEP_0192090630 SYARWL STH TG NPPCG SGQ RHPE HHF FYG DYL TAEN
Peridinium-aciculiferum-PAER_2_CAMPEP_0190651180 SYARWL STH TG NPPCG SGR RHPE HKF FYG EYL TAEN
Lingulodinium-polyedra-CCMP1738_CAMPEP_0190130486 KVARWL ATT TG DPPCA AGK RHPE HTF FYG EYL TAED
Ceratium-fusus-PA161109_CAMPEP_0199533546 -MARWL STT TG DPPCG AGK RHPQ HNF MYG EYF TAED
EUGLENOPHYTE_Eutreptiella-gymnastica_like-CCMP1594_CAMPEP_0200423160 RFARWL AAQ TG DPNCA TGD RHPA HKW FYG DYW DPED
ANIMAL_Capca1_152010 QMARYL AAV TG EPLCA FGK RHPA HGW FYG DYY DPEE
heterotrophic heterokonts
HETEROTROPHIC_HETEROKONTS_Phytophthora_141895 QHARQL STI TG DPTCW TGN KHPQ HGF FYG EYY DVED
58_XP_002898601.1Phytophthora_infestans QHARQL STI TG DPTCW TGK KHPQ HGF FYG EYY DVED
56_XP_008900952.1Phytophthora_parasitica QHARQL STI TG DPTCW TGK KHPQ HGF FYG EYY DVED
59_CCD28267.Plasmopara_viticola LHARTL STI TG DPTCW TGK KHPQ HNF SYG EYY DVED
57_GAY06784.1Pythium_insidiosum VHAREL ATR TG DPTCM VGR RHPQ HGF YYG EYY DVED
60_XP_002896079.1Phytophthora_infestans KHARML ATT TG DPTCW TGR KHPQ HDF FYG DYY DAED
61_ETI53266.1Phytophthora_parasitica KHARML ATT TG DPTCW TGR KHPQ HDF FYG DYY DAED
55_XP_009528543.1Phytophthora_sojae KHARML ATT TG DPTCW TGR KHPQ HGF FYG DYY DPED
62_OQR82299.1Achlya_hypogyna RVARWI ATT TG DPTCW SGN KHPQ HSF FYG AYY DPQD
63_XP_012209989.1Saprolegnia_parasitica RIARWI ATT TG DPTCW SGT KHPQ HDF FYG TYY DPQD
chlorophytes
CHLOROPHYTES_Pterosperma_CAMPEP_0197852200 KMARWL ATT TG SPLCA SGD THPA HDF MFG DYL DPES
Chlorella_281865 KMARWL STV TG DPACA VGR RHPA HQF WYG EYY DPED
Dunaliella_CAMPEP_0202349486 QMARWM STA SP DPACA LGS RHPA HHF WYG DYY DPED
Pyramimonas_CAMPEP_0191463466 RFARWL STT TG DPNCS SGT RHPS HGW HYG EYF DAED
Crustomastix_CAMPEP_0183812746 ------ --- TG DPNCA SGQ RHPE HGW EYG EYF PAED
Dolichomastix_CAMPEP_0170139300 RVARWL STT TG DPPCA TGK RFPD HEF FYG DYF YAED
Pycnococcus_CAMPEP_0205957742 RRARWL GTT TG DPTCA TGP RHPA HHF MYG EYW CAED
plants
PLANTS_7_PPS00070.1Gossypium_barbadense AYARWL TTL TG SPVCA TGR KHPE HDF WFG QYL DPES
9_XP_022740311.1Durio_zibethinus AYARWL NTI TG SPVCA TGR KHPE HDF WFG QYL DPES
8_XP_022751888.1Durio_zibethinus GYARWL NTI SG SPVCA TGK KHPE HNF WFG QYL DPES
11_XP_007031353.2Theobroma_cacao AYARWL NTL TG SPVCA TGS KHPE HNF WFG QYL DPES
20_XP_015063292.1Solanum_pennellii LFARWL NTI RG NPSCA VGK KHPE HDF WFG QYL DPEN
21_XP_006338517.1Solanum_tuberosum LFARWL NTI RG NPSCA VGK KHPE HDF WFG QYL DPEN
17_XP_009793787.1Nicotiana_sylvestris LFARWL NTI RG NPSCA IGK KHPE HNF WFG QYL DPEN
15_XP_009615033.1Nicotiana_tomentosiformis LFARWL NTI RG NPSCA IGK KHPE HNF WFG QYL DPEN
Populus_177625 ASARWL NTI SG NPSCA TGK KHPE HNF WFG QYL DPEN
10_XP_002517410.2Ricinus_communis ATARWL NTI SG NPTCA TGK RHQE HNF WFG EYF DPEN
13_XP_018478583.1Raphanus_sativus ASARWL STL KG DPTCS TGK KHPE HEF WYG EYL DPMD
19_VDD12421.1Brassica_rapa ASARWL STL KG DPTCS TGK KHPE HEF WYG EYL DPMD
5_OAE27149.1Marchantia_polymorpha AMARWL STL TG DPTCA SGK KHPE HDW FYG DYY DVED
12_XP_024380128.1Physcomitrella_patens ATARWL STI GG NPTCA SGK KHPE HKW NYG EYY DPEN
HAPTOPHYTE_Pavlova-sp-CCMP459_CAMPEP_0190508168 GTARWL STS TG DPRCS HGR VHPA HGF DYG QYV DPED
SLIME_MOLD_Thraustochytrium-sp-LLF1b_CAMPEP_0193130544 EAARFV ATG TG DPTCW TGP KHPE HGF FYG EYF DAMD
fungi
FUNGI_47_XP_020065698.1Suhomyces_tanzawaensis[fungi] TVARTL NTI DG PAGSP SGR RHPD HSS GFG LYH EQSP
48_XP_015466004.1Debaryomyces_fabryi RVARTL NTL NG PAGSP MGK RHPD HKS GFG IYH KHSP
49_XP_001383610.2Scheffersomyces_stipitis SVARTL NTI SG PAGSP SGR RHPD HSS GFG LYH AASP
50_ORX95428.1Basidiobolus_meristosporus ------ IT- NG PMPHA MGE KHPE HDF GFG VYH SQSP
52_RCI02963.1Rhizopus_stolonifer HLTRRI LTL TG PVQQP FGH THPE HDF GFG LYR RSTP
51_ORX88516.1Basidiobolus_meristosporus -----M IT- SG PMPHA MGE KHPE PDF GFG IYH RQSP
bacteria
BACTERIA_Pleomorphomonas_WP_100080278 DMARRI GTL DG PLTGA VGR RHPE GDF GFG AFV GDDP
Pleomorphomonas_WP_101290810 DMTRRI GTL DG PLTGA IGR RHPE ADF GFG AFA DEDP
Sinobacteraceae_WP_123212147 PLARAL STI QG AQAAG LGD RYPQ HDF GFG FAG GDDA
Sinobacteraceae_TAJ53459 IHARAL STI QG AQAAG LGD RYPQ HDF GFG FAG GEDA
Betaproteobacteria_PKO45672 --ARLF STH AC LQANA LGE YLPQ HDF GFG LAS DGDL
Alteromonas_WP_015067981 FQAKQL GTH AG SQANG MGS LFPQ HDF GFG DWH EDDS
Alteromonas_WP_071815886 FQAKQL GTH AG SQANG MGS LFPQ HDF GFG DWH EDDS
Consensus RTARWL STI TG DPRCA SGK RHPE HNF FYG EYY DPED
100%
Conservation
0%
1 2 3 4 5 6 7 8 9 loop
0% 100%

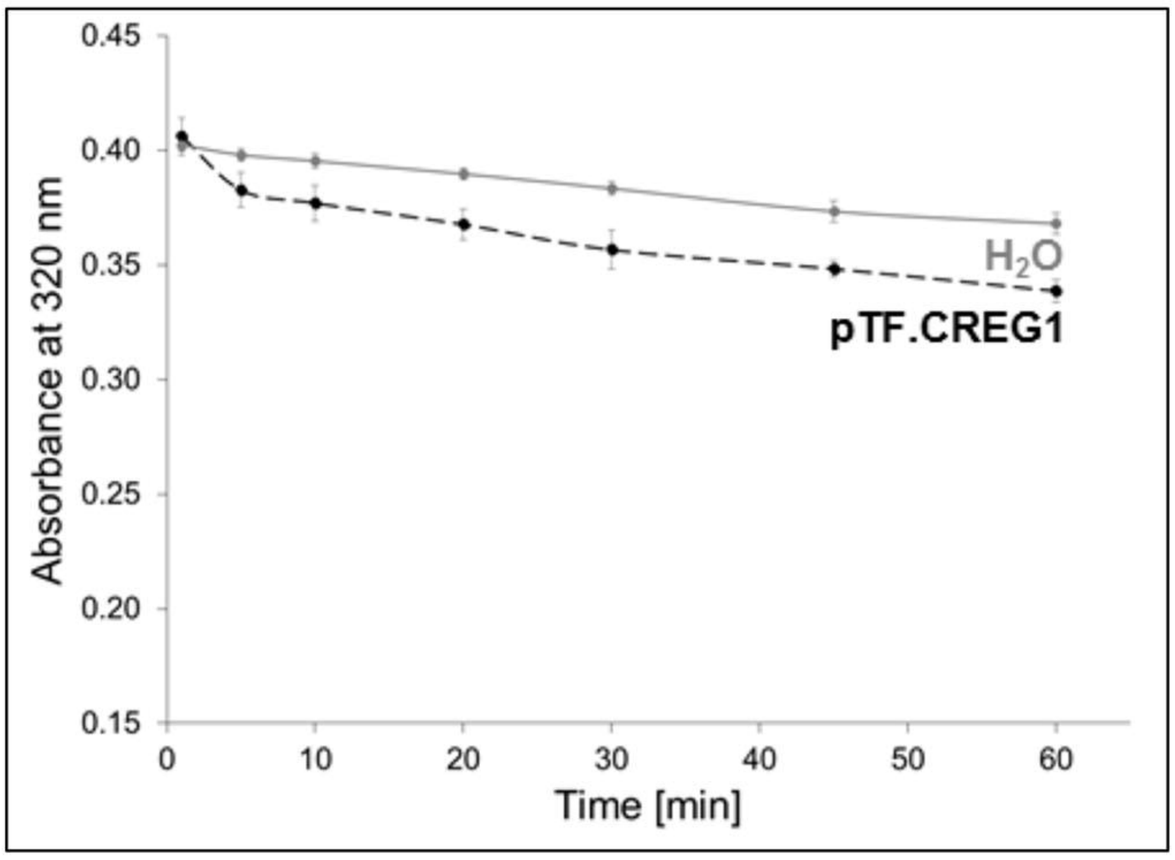
A

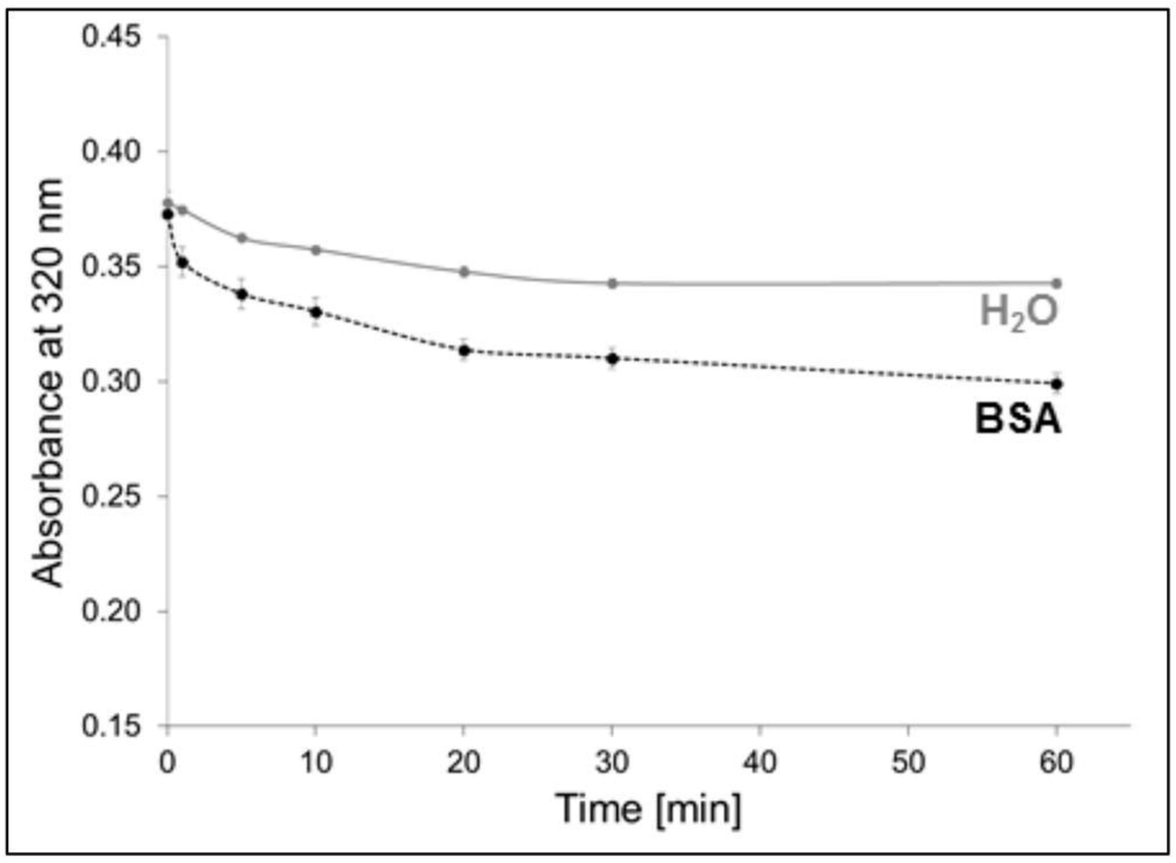

CREG1_H_sapiens MAGLSRGSAR ALLAALLAST LLALLVSPAR GRGGRDHGDW DEASRLPPLP PREDAARVAR FVTHVSDWGA LATISTLEAV 80
pTF.CREG1_P_tricornutum MF---RFSAV ILYAMLTLSG AYPSLRHPFE LDKQSDECPC EPLQR----P DRFNKEELAR WMVHSMDWGV LTTIST-RLP 72
CREG1_H_sapiens RGRPFADVLS LSDGPPGAGS GVPYFYLSPL QLSVSNLQEN PYATLTMTLA QT-NFCKK-- ---------H GFDPQSPLCV 148
pTF.CREG1_P_tricornutum DGQPFGNVYS FVDGPCGKGM GTPYFYGTYM DQSFHDSKQN DKVSFTLTEA SLPSVCHAGA SKDCSISHAN AGDPESPVCA 152
CREG1_H_sapiens HIMLSGTVTK VN--ETEMDI AKHSLFIRHP EMKTWPSSHN WFFAKLNITN IWVLDYFGGP KIVTPEEYYN VTVQ------ 220
pTF.CREG1_P_tricornutum RLTLSGKLVE VNPDSEEYTR AQAAFFQRHP QMATWPSEHN WIIAKLEIED LWLINFYGGA AILSIEEYFS AKLSPDMQLE 232
CREG1_H_sapiens -- 220
pTF.CREG1_P_tricornutum RL 234
0%
100%

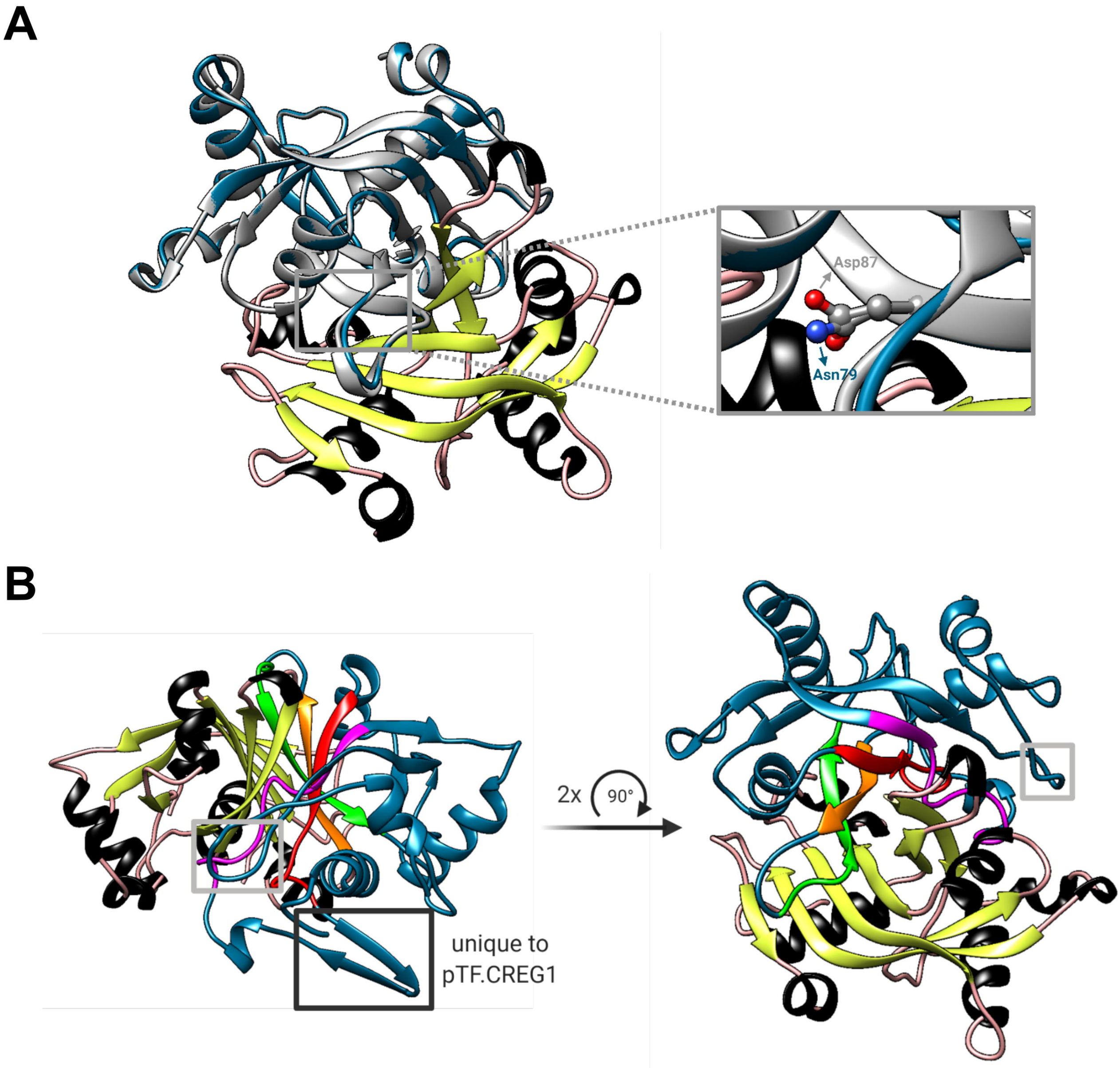
B

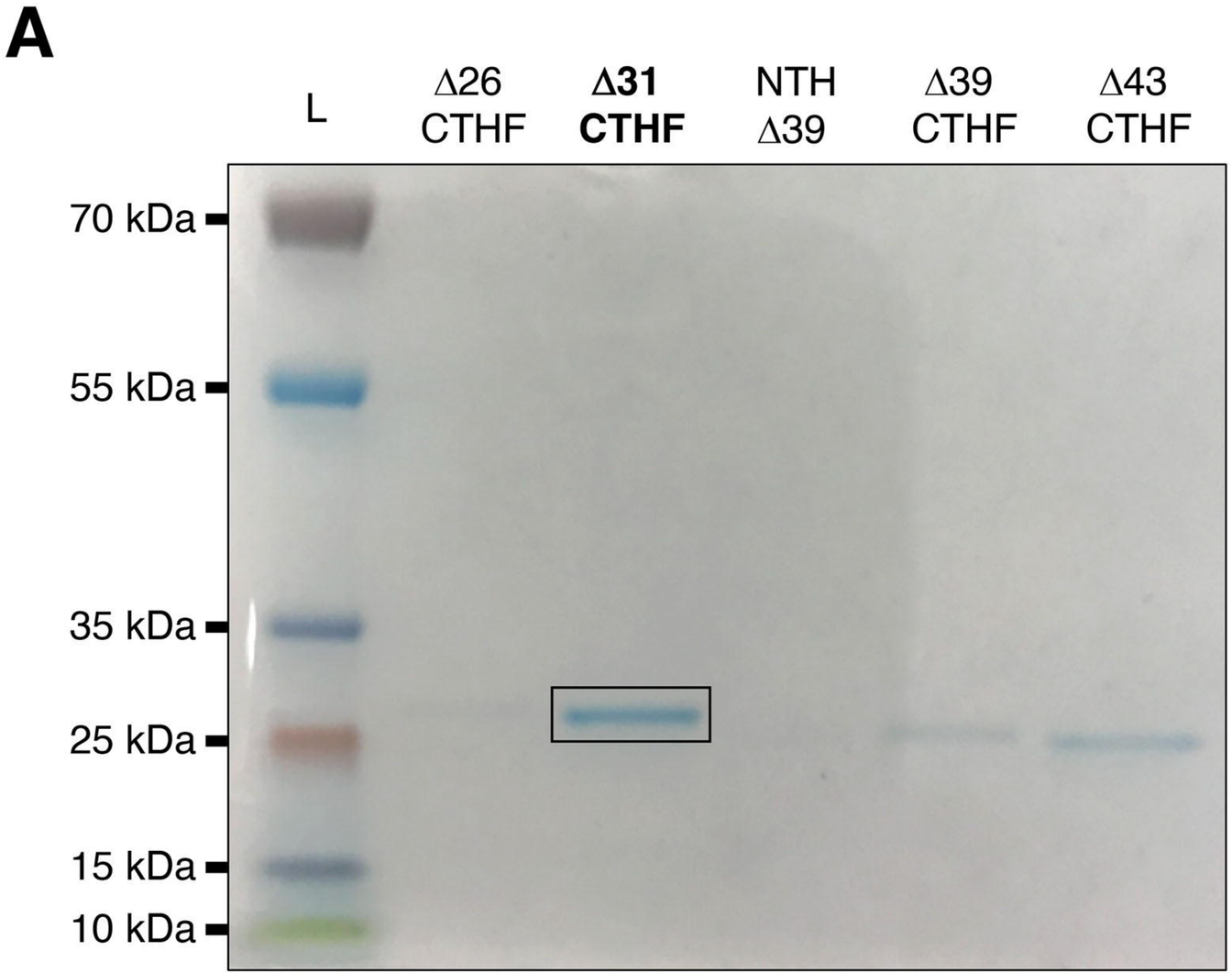

C
C
N
N
N
C
Pro146
Pro142
Glu147
Gln143

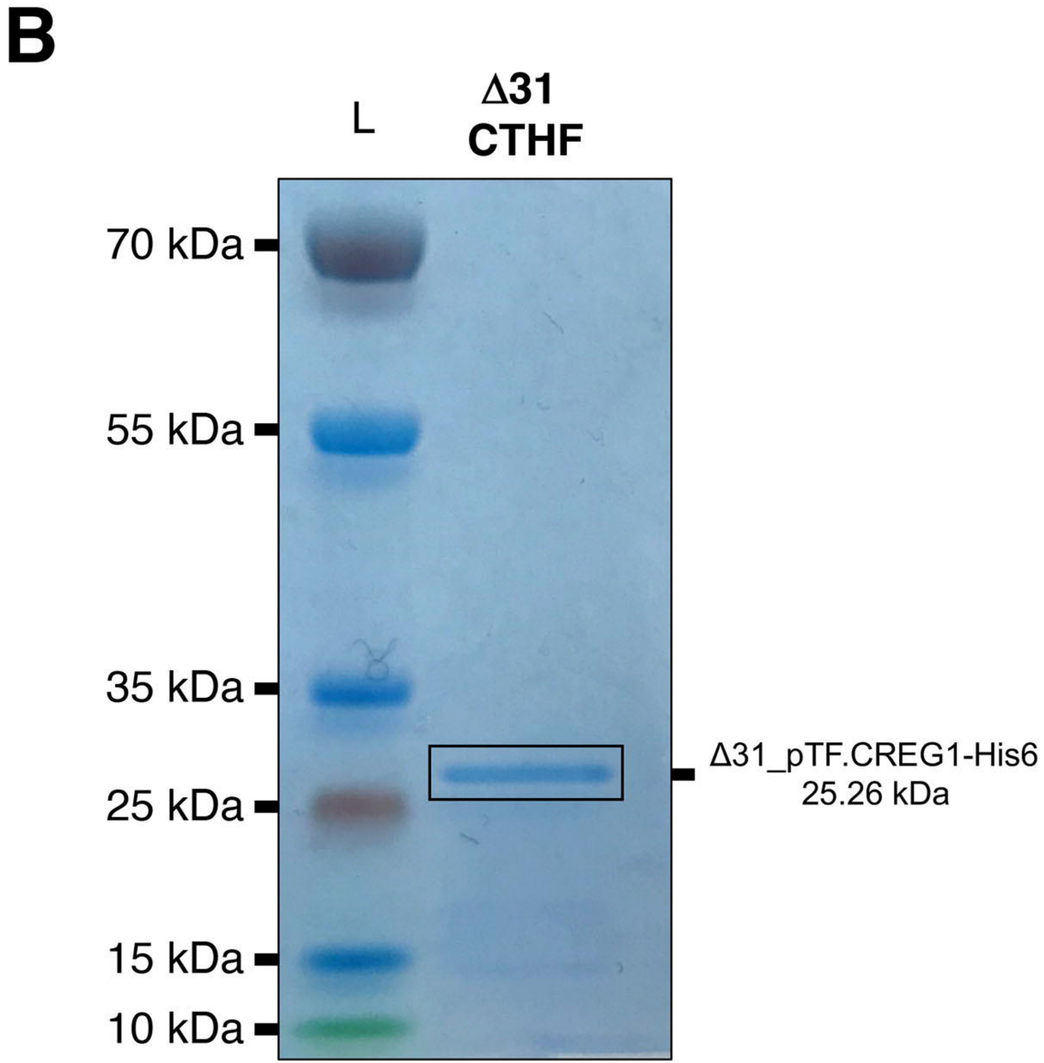
C


Absorbance at 320 nm
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0
10
20
30
40
50
60
Time [min]
H2O
pTF.CREG1

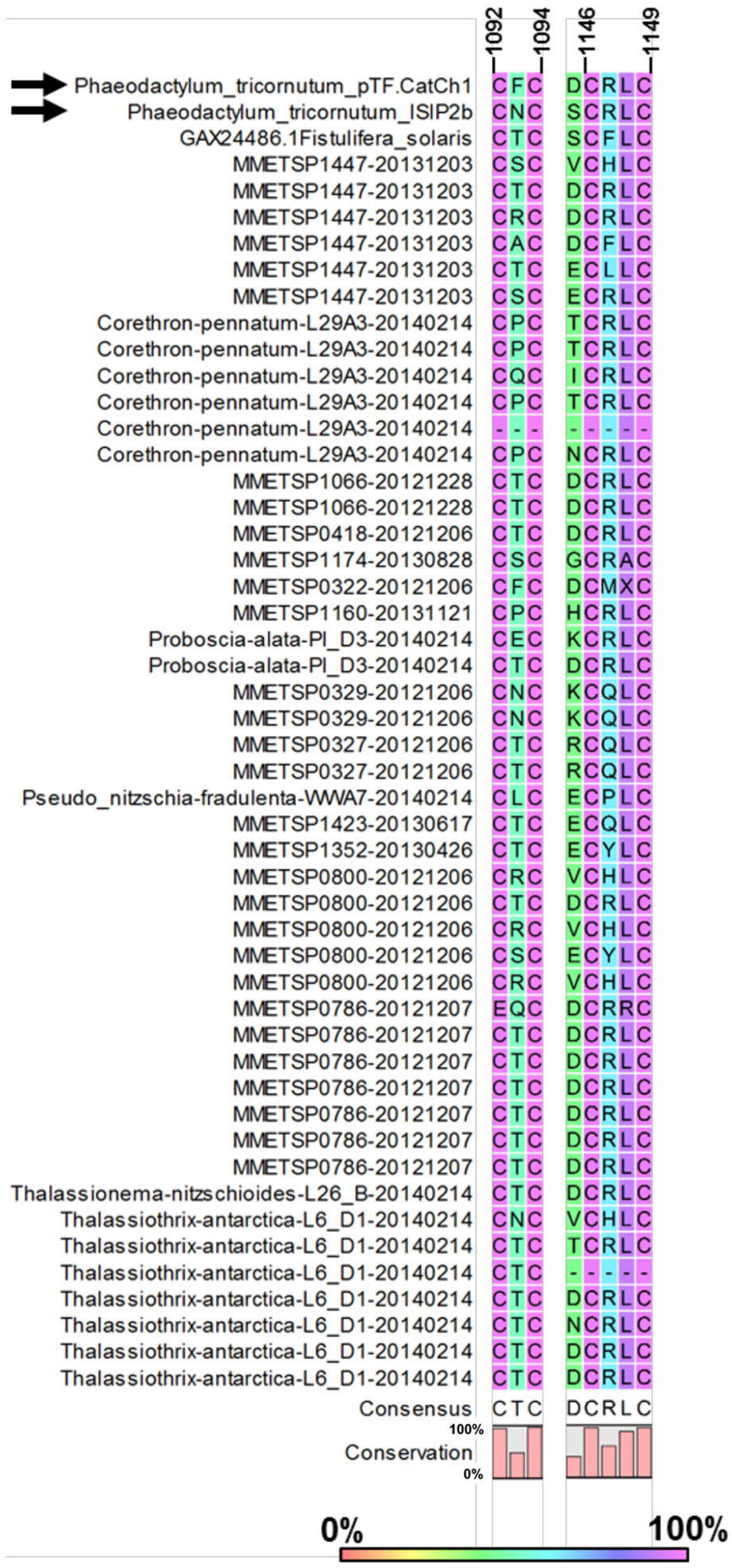

Absorbance at 320 nm
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0
10
20
30
40
50
60
Time [min]
H2O
BSA


A
Asp87
Asn79
B
unique to
pTF.CREG1
2x
90°

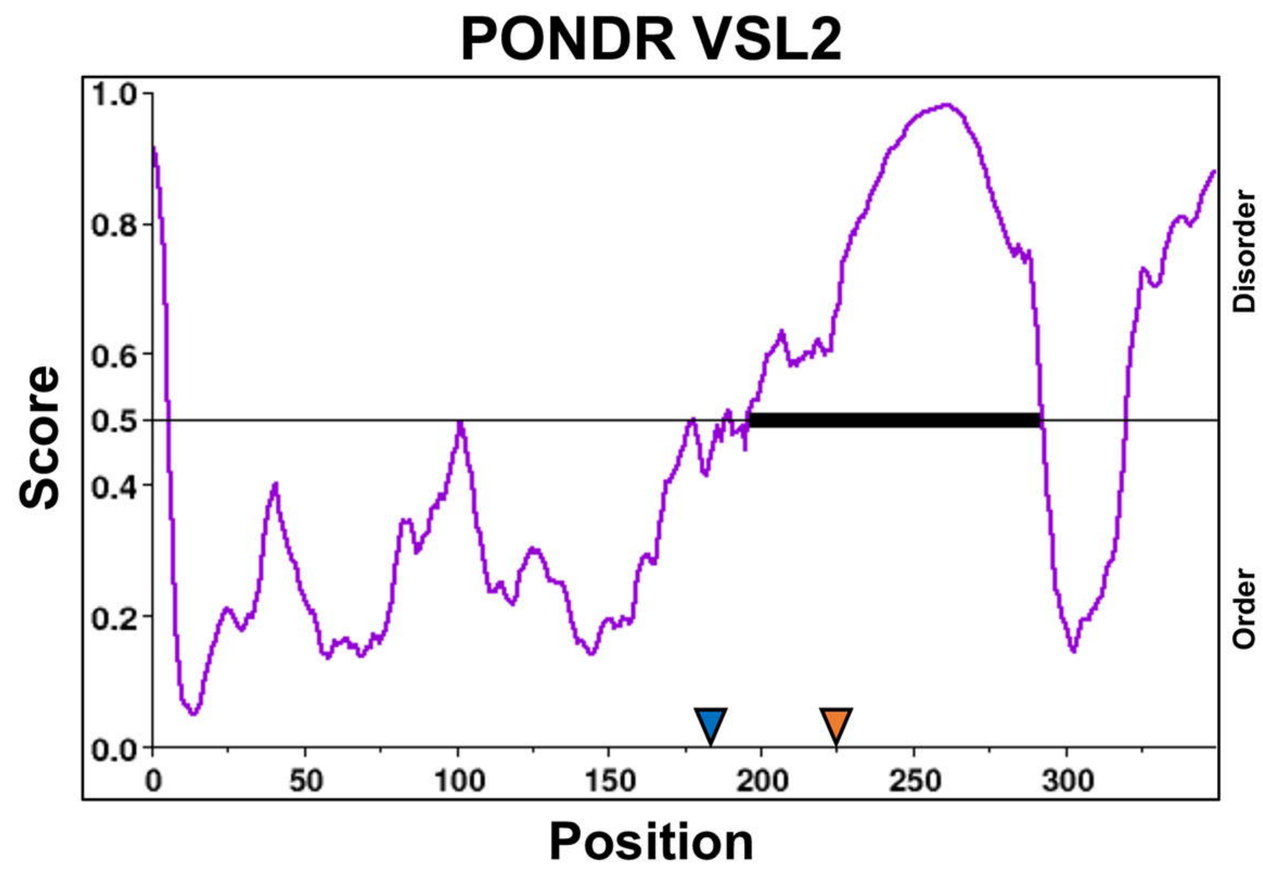

A
L
Δ26 CTHF
Δ31 CTHF
NTH Δ39
Δ39 CTHF
Δ43 CTHF
70 kDa
55 kDa
35 kDa
25 kDa
15 kDa
10 kDa

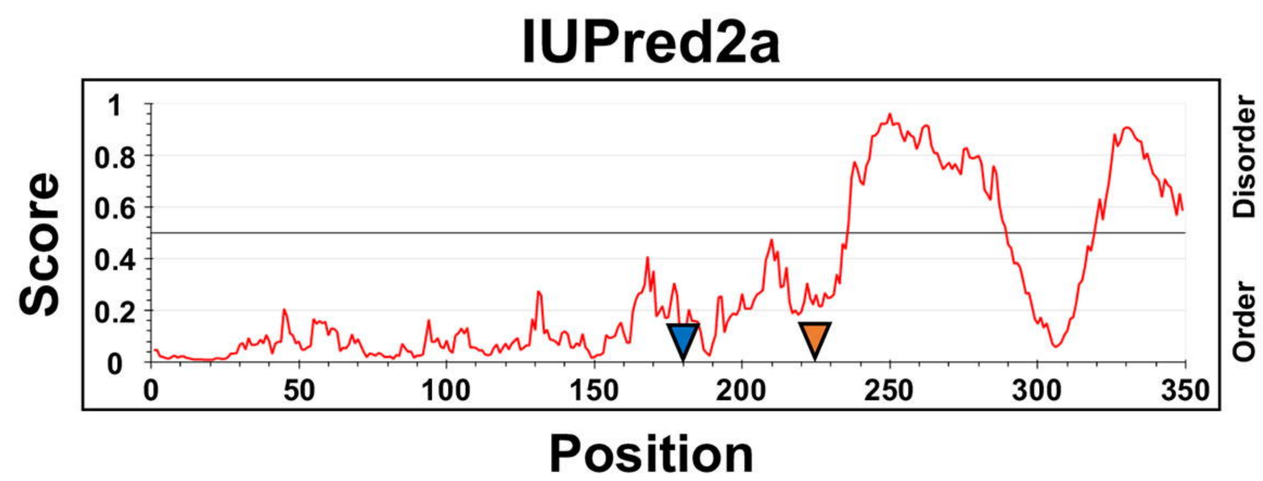

B
L
Δ31 CTHF
70 kDa
55 kDa
35 kDa
25 kDa
15 kDa
10 kDa
Δ31_pTF.CREG1-His6
25.26 kDa

A

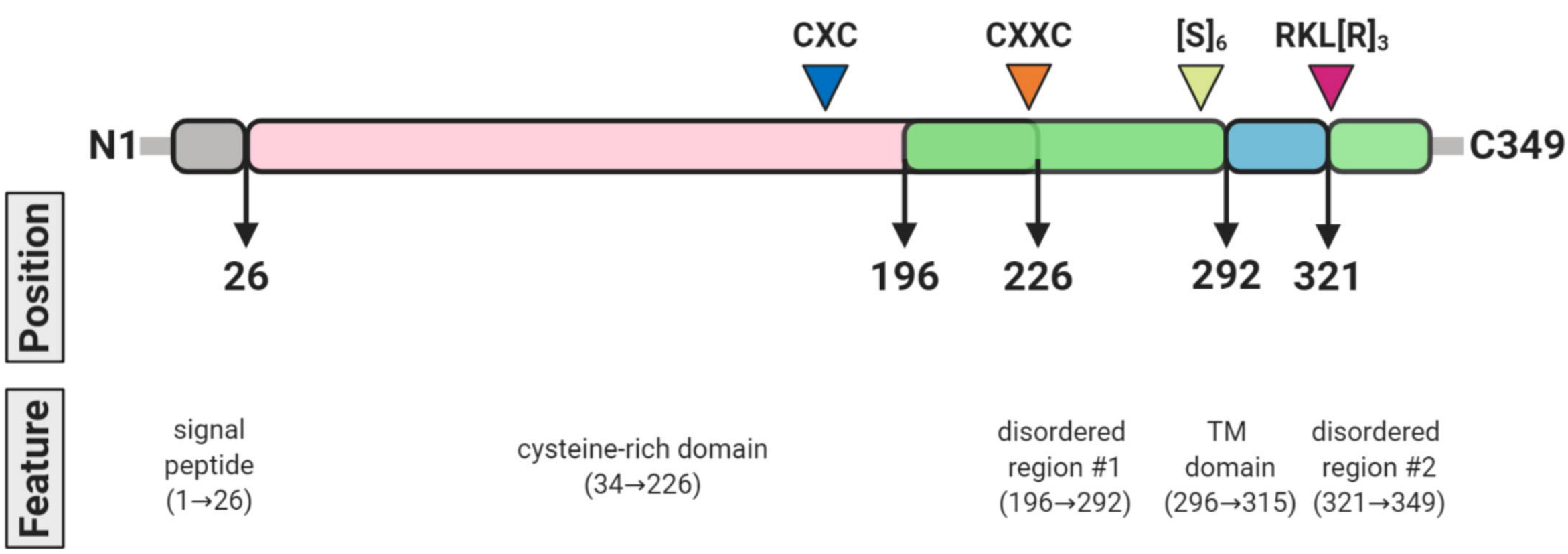

1092
1094
1146
1149
Phaeodactylum_tricornutum_pTF.CatCh1 CFC DCRLC
Phaeodactylum_tricornutum_ISIP2b CNC SCRLC
GAX24486.1Fistulifera_solaris CTC SCFLC
MMETSP1447-20131203 CSC VCHLC
MMETSP1447-20131203 CTC DCRLC
MMETSP1447-20131203 CRC DCRLC
MMETSP1447-20131203 CAC DCFLC
MMETSP1447-20131203 CTC ECLLC
MMETSP1447-20131203 CSC ECRLC
Corethron-pennatum-L29A3-20140214 CPC TCRLC
Corethron-pennatum-L29A3-20140214 CPC TCRLC
Corethron-pennatum-L29A3-20140214 CQC ICRLC
Corethron-pennatum-L29A3-20140214 CPC TCRLC
Corethron-pennatum-L29A3-20140214 - - - - - - - -
Corethron-pennatum-L29A3-20140214 CPC NCRLC
MMETSP1066-20121228 CTC DCRLC
MMETSP1066-20121228 CTC DCRLC
MMETSP0418-20121206 CTC DCRLC
MMETSP1174-20130828 CSC GCRAC
MMETSP0322-20121206 CFC DCMXC
MMETSP1160-20131121 CPC HCRLC
Proboscia-alata-PI_D3-20140214 CEC KCRLC
Proboscia-alata-PI_D3-20140214 CTC DCRLC
MMETSP0329-20121206 CNC KCQLC
MMETSP0329-20121206 CNC KCQLC
MMETSP0327-20121206 CTC RCQLC
MMETSP0327-20121206 CTC RCQLC
Pseudo_nitzschia-fradulenta-WWA7-20140214 CLC ECPLC
MMETSP1423-20130617 CTC ECQLC
MMETSP1352-20130426 CTC ECYLC
MMETSP0800-20121206 CRC VCHLC
MMETSP0800-20121206 CTC DCRLC
MMETSP0800-20121206 CRC VCHLC
MMETSP0800-20121206 CSC ECYLC
MMETSP0800-20121206 CRC VCHLC
MMETSP0786-20121207 EQC DCRRC
MMETSP0786-20121207 CTC DCRLC
MMETSP0786-20121207 CTC DCRLC
MMETSP0786-20121207 CTC DCRLC
MMETSP0786-20121207 CTC DCRLC
MMETSP0786-20121207 CTC DCRLC
MMETSP0786-20121207 CTC DCRLC
Thalassionema-nitzschioides-L26_B-20140214 CTC DCRLC
Thalassiothrix-antarctica-L6_D1-20140214 CNC VCHLC
Thalassiothrix-antarctica-L6_D1-20140214 CTC TCRLC
Thalassiothrix-antarctica-L6_D1-20140214 CTC - - - - - - - -
Thalassiothrix-antarctica-L6_D1-20140214 CTC DCRLC
Thalassiothrix-antarctica-L6_D1-20140214 CTC NCRLC
Thalassiothrix-antarctica-L6_D1-20140214 CTC DCRLC
Thalassiothrix-antarctica-L6_D1-20140214 CTC DCRLC
Consensus CTC DCRLC
100%
Conservation
0%
0%
100%

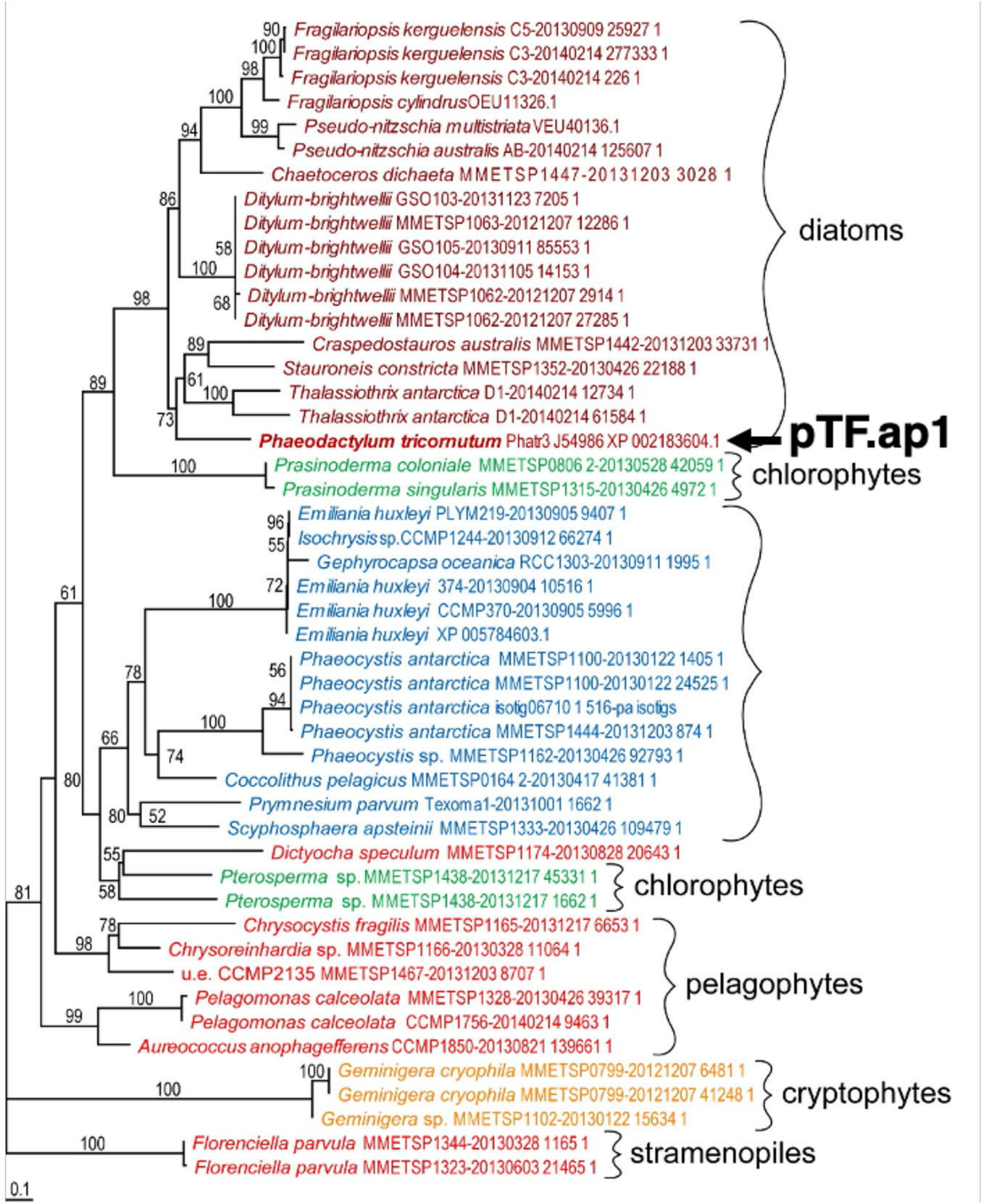
B

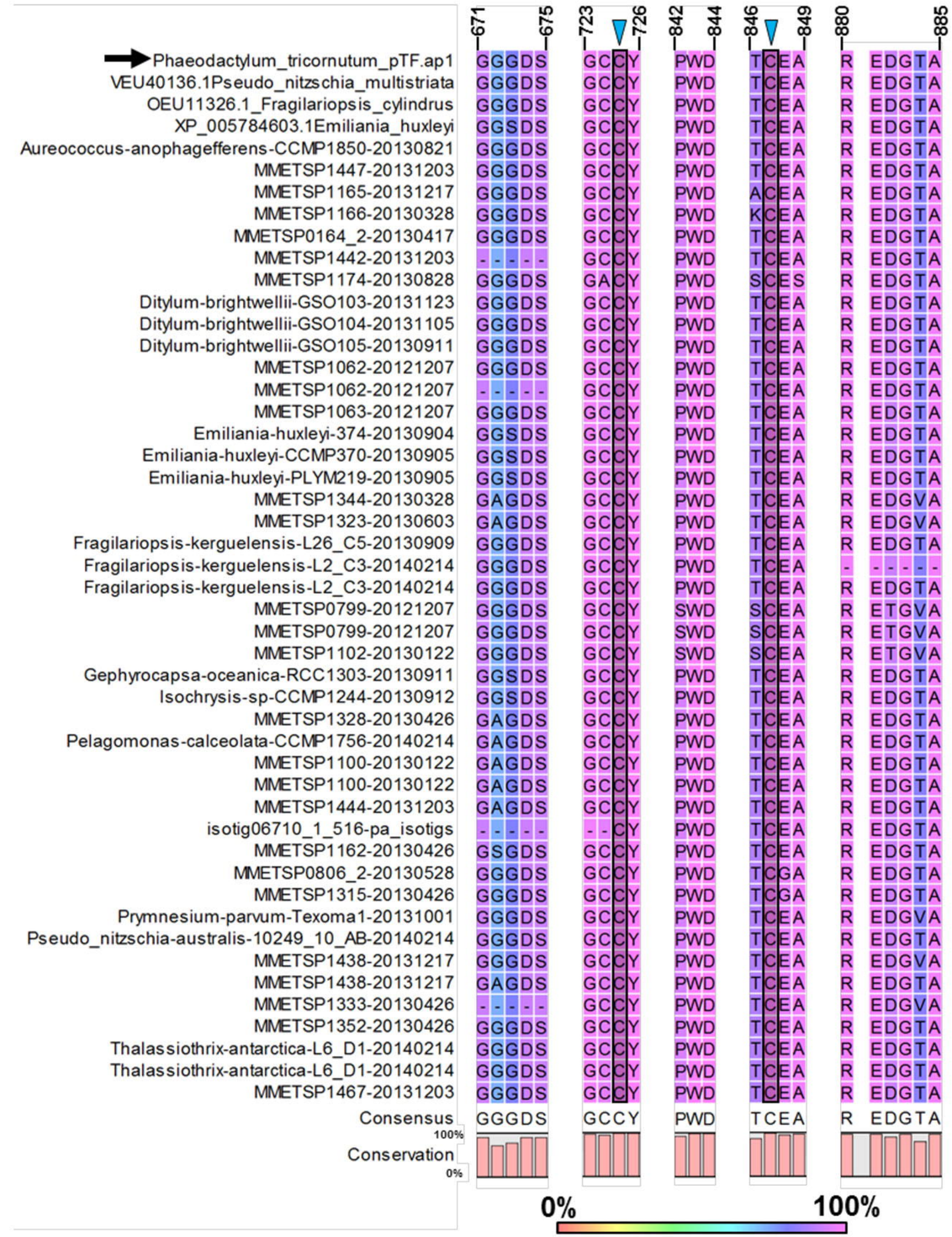

PONDR VSL2
Score
1.0
0.8
0.6
0.5
0.4
0.2
0.0
0
50
100
150
200
250
300
Position
Disorder
Order

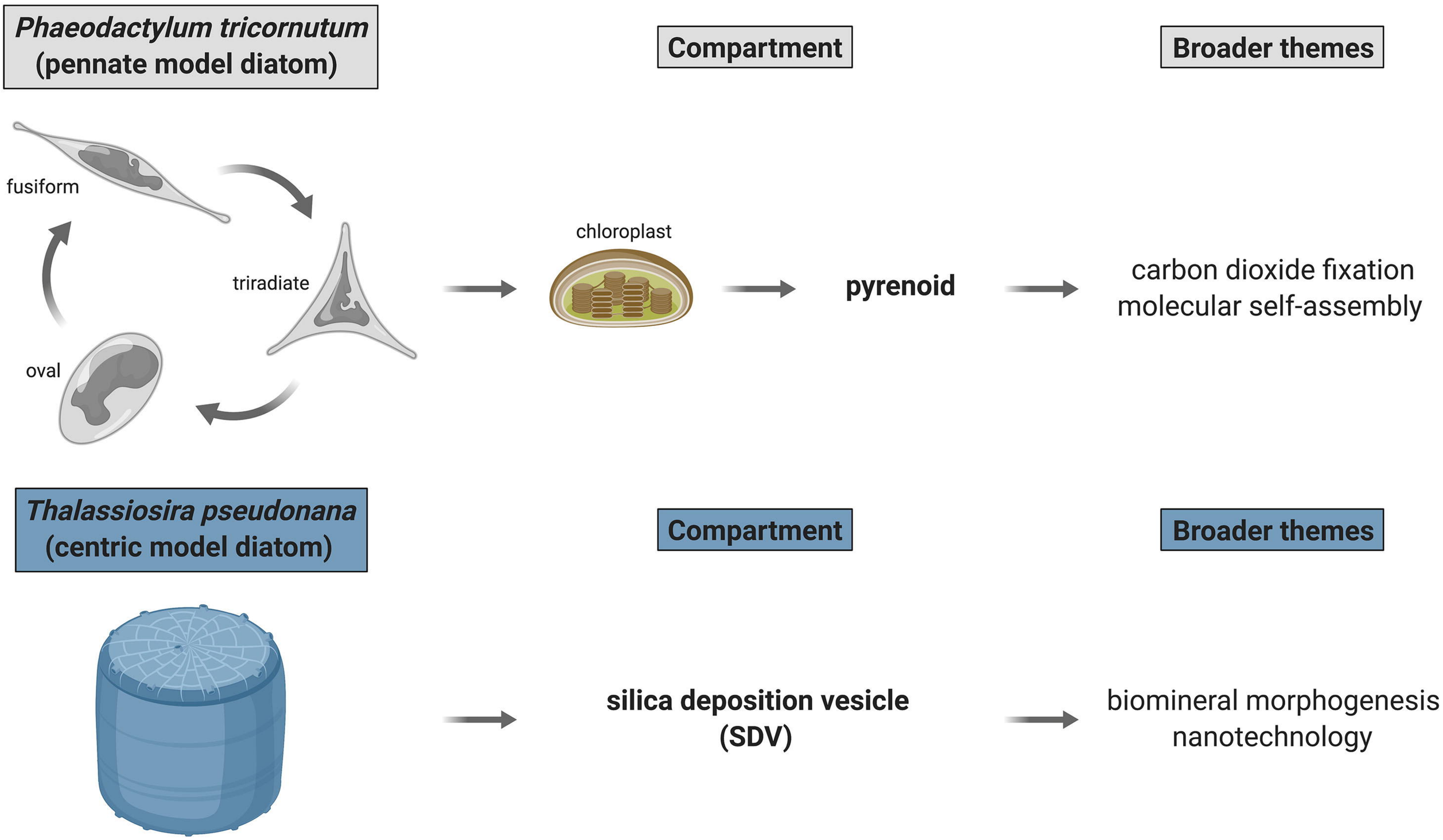

IUPred2a
Score
1
0.8
0.6
0.4
0.2
0
0
50
100
150
200
250
300
350
Position
Disorder
Order

C


CXC
CXXC
[S]6
RKL[R]3
N1
C349
Position
26
196
226
292
321
Feature
signal peptide (1→26)
cysteine-rich domain (34→226)
disordered region #1 (196→292)
TM domain (296→315)
disordered region #2 (321→349)


A
Fragilariopsis kerguelensis C5-20130909 25927 1
Fragilariopsis kerguelensis C3-20140214 277333 1
Fragilariopsis kerguelensis C3-20140214 226 1
Fragilariopsis cylindrusOEU11326.1
Pseudo-nitzschia multistriata VEU40136.1
Pseudo-nitzschia australis AB-20140214 125607 1
Chaetoceros dichaeta MMETSP1447-20131203 3028 1
Ditylum-brightwellii GSO103-20131123 7205 1
Ditylum-brightwellii MMETSP1063-20121207 12286 1
Ditylum-brightwellii GSO105-20130911 85553 1
Ditylum-brightwellii GSO104-20131105 14153 1
Ditylum-brightwellii MMETSP1062-20121207 2914 1
Ditylum-brightwellii MMETSP1062-20121207 27285 1
Craspedostauros australis MMETSP1442-20131203 33731 1
Stauroneis constricta MMETSP1352-20130426 22188 1
Thalassiothrix antarctica D1-20140214 12734 1
Thalassiothrix antarctica D1-20140214 61584 1
Phaeodactylum tricornutum Phatr3 J54986 XP 002183604.1
pTF.ap1
diatoms
Prasinoderma coloniale MMETSP0806 2-20130528 42059 1
Prasinoderma singularis MMETSP1315-20130426 4972 1
chlorophytes
Emiliania huxleyi PLYM219-20130905 9407 1
Isochrysis sp.CCMP1244-20130912 66274 1
Gephyrocapsa oceanica RCC1303-20130911 1995 1
Emiliania huxleyi 374-20130904 10516 1
Emiliania huxleyi CCMP370-20130905 5996 1
Emiliania huxleyi XP 005784603.1
Phaeocystis antarctica MMETSP1100-20130122 1405 1
Phaeocystis antarctica MMETSP1100-20130122 24525 1
Phaeocystis antarctica isotig06710 1 516-pa isotigs
Phaeocystis antarctica MMETSP1444-20131203 874 1
Phaeocystis sp. MMETSP1162-20130426 92793 1
Coccolithus pelagicus MMETSP0164 2-20130417 41381 1
Prymnesium parvum Texoma1-20131001 1662 1
Scyphosphaera apsteinii MMETSP1333-20130426 109479 1
Dictyocha speculum MMETSP1174-20130828 20643 1
Pterosperma sp. MMETSP1438-20131217 45331 1
Pterosperma sp. MMETSP1438-20131217 1662 1
chlorophytes
Chrysocystis fragilis MMETSP1165-20131217 6653 1
Chrysoreinhardia sp. MMETSP1166-20130328 11064 1
u.e. CCMP2135 MMETSP1467-20131203 8707 1
Pelagomonas calceolata MMETSP1328-20130426 39317 1
Pelagomonas calceolata CCMP1756-20140214 9463 1
Aureococcus anophagefferens CCMP1850-20130821 139661 1
pelagophytes
Geminigera cryophila MMETSP0799-20121207 6481 1
Geminigera cryophila MMETSP0799-20121207 41248 1
Geminigera sp. MMETSP1102-20130122 15634 1
cryptophytes
Florenciella parvula MMETSP1344-20130328 1165 1
Florenciella parvula MMETSP1323-20130603 21465 1
stramenopiles
0.1


B
671 675 723 726 842 844 846 849 880 885
Phaeodactylum_tricornutum_pTF.ap1 GGGDS GCCY PWD TCEA R EDGTA
VEU40136.1Pseudo_nitzschia_multistriata GGGDS GCCY PWD TCEA R EDGTA
OEU11326.1_Fragilariopsis_cylindrus GGGDS GCCY PWD TCEA R EDGTA
XP_005784603.1Emiliania_huxleyi GGSDS GCCY PWD TCEA R EDGTA
Aureococcus-anophagefferens-CCMP1850-20130821 GGGDS GCCY PWD TCEA R EDGTA
MMETSP1447-20131203 GGGDS GCCY PWD TCEA R EDGTA
MMETSP1165-20131217 GGGDS GCCY PWD ACEA R EDGTA
MMETSP1166-20130328 GGGDS GCCY PWD KCEA R EDGTA
MMETSP0164_2-20130417 GGGDS GCCY PWD TCEA R EDGTA
MMETSP1442-20131203 - - - - GCCY PWD TCEA R EDGTA
MMETSP1174-20130828 GGGDS GACY PWD SCES R EDGTA
Ditylum-brightwellii-GSO103-20131123 GGGDS GCCY PWD TCEA R EDGTA
Ditylum-brightwellii-GSO104-20131105 GGGDS GCCY PWD TCEA R EDGTA
Ditylum-brightwellii-GSO105-20130911 GGGDS GCCY PWD TCEA R EDGTA
MMETSP1062-20121207 GGGDS GCCY PWD TCEA R EDGTA
MMETSP1062-20121207 - - - - GCCY PWD TCEA R EDGTA
MMETSP1063-20121207 GGGDS GCCY PWD TCEA R EDGTA
Emiliania-huxleyi-374-20130904 GGSDS GCCY PWD TCEA R EDGTA
Emiliania-huxleyi-CCMP370-20130905 GGSDS GCCY PWD TCEA R EDGTA
Emiliania-huxleyi-PLYM219-20130905 GGSDS GCCY PWD TCEA R EDGTA
MMETSP1344-20130328 GAGDS GCCY PWD TCEA R EDGVA
MMETSP1323-20130603 GAGDS GCCY PWD TCEA R EDGVA
Fragilariopsis-kerguelensis-L26_C5-20130909 GGGDS GCCY PWD TCEA R EDGTA
Fragilariopsis-kerguelensis-L2_C3-20140214 GGGDS GCCY PWD TCEA - - - - - -
Fragilariopsis-kerguelensis-L2_C3-20140214 GGGDS GCCY PWD TCEA R EDGTA
MMETSP0799-20121207 GGGDS GCCY SWD SCEA R ETGVA
MMETSP0799-20121207 GGGDS GCCY SWD SCEA R ETGVA
MMETSP1102-20130122 GGGDS GCCY SWD SCEA R ETGVA
Gephyrocapsa-oceanica-RCC1303-20130911 GGSDS GCCY PWD TCEA R EDGTA
Isochrysis-sp-CCMP1244-20130912 GGSDS GCCY PWD TCEA R EDGTA
MMETSP1328-20130426 GAGDS GCCY PWD TCEA R EDGTA
Pelagomonas-calceolata-CCMP1756-20140214 GAGDS GCCY PWD TCEA R EDGTA
MMETSP1100-20130122 GAGDS GCCY PWD TCEA R EDGTA
MMETSP1100-20130122 GAGDS GCCY PWD TCEA R EDGTA
MMETSP1444-20131203 GAGDS GCCY PWD TCEA R EDGTA
isotig06710_1_516-pa_isotigs - - - - - - CY PWD TCEA R EDGTA
MMETSP1162-20130426 GSGDS GCCY PWD TCEA R EDGTA
MMETSP0806_2-20130528 GGGDS GCCY PWD TCGA R EDGTA
MMETSP1315-20130426 GGGDS GCCY PWD TCGA R EDGTA
Prymnesium-parvum-Texoma1-20131001 GGGDS GCCY PWD TCEA R EDGVA
Pseudo_nitzschia-australis-10249_10_AB-20140214 GGGDS GCCY PWD TCEA R EDGTA
MMETSP1438-20131217 GGGDS GCCY PWD TCEA R EDGVA
MMETSP1438-20131217 GAGDS GCCY PWD TCEA R EDGTA
MMETSP1333-20130426 - - - - - GCCY PWD TCEA R EDGVA
MMETSP1352-20130426 GGGDS GCCY PWD TCEA R EDGTA
Thalassiothrix-antarctica-L6_D1-20140214 GGGDS GCCY PWD TCEA R EDGTA
Thalassiothrix-antarctica-L6_D1-20140214 GGGDS GCCY PWD TCEA R EDGTA
MMETSP1467-20131203 GGGDS GCCY PWD TCEA R EDGTA
Consensus GGGDS GCCY PWD TCEA R EDGTA
100%
Conservation
0%
0%
100%


Phaeodactylum tricornutum
(pennate model diatom)
Compartment
Broader themes
fusiform
triradiate
oval
chloroplast
pyrenoid
carbon dioxide fixation
molecular self-assembly
Thalassiosira pseudonana
(centric model diatom)
Compartment
Broader themes
silica deposition vesicle
(SDV)
biomineral morphogenesis
nanotechnology