bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title

Proximity-based proteomics reveals the thylakoid lumen proteome in the cyanobacterium *Synechococcus* sp. PCC 7002

Authors/Affiliations

Kelsey K. Dahlgren[1,2,3,4], Colin Gates[2], Thomas Lee[1], Jeffrey C. Cameron[1,2,5*].

[1]Department of Biochemistry, University of Colorado, Boulder, CO 80309, USA.

[2]Renewable and Sustainable Energy Institute, University of Colorado, Boulder, CO 80309, USA.

[3]BioFrontiers Institute, University of Colorado, Boulder, CO 80309, USA

[4]Interdisciplinary Quantitative Biology Program (IQ Biology), BioFrontiers Institute, University of Colorado, Boulder, CO 80309, USA

[5]National Renewable Energy Laboratory, Golden, CO 80401, USA.

Contact

*Correspondence to: jeffrey.c.cameron@colorado.edu

Running Title:

Spatial proteomics of the thylakoid lumen

Abstract

Cyanobacteria possess unique intracellular organization. Many proteomic studies have examined different features of cyanobacteria to learn about the structure-function relationships between the intracellular structures of cyanobacteria and their roles in cells. While these studies have made great progress in understanding cyanobacterial physiology, the previous fractionation methods used to purify cellular structures have limitations; specifically, certain regions of cells cannot be purified with existing fractionation methods. Proximity-based proteomics techniques were developed to overcome the limitations of biochemical fractionation for proteomics. Proximity-based proteomics relies on spatiotemporal protein labeling followed by mass spectrometry of the labeled proteins to determine the proteome of the region of interest. We have performed proximity-based proteomics in the cyanobacterium *Synechococcus sp.* PCC 7002 with the APEX2 enzyme, an engineered ascorbate peroxidase. We determined the proteome of the thylakoid lumen, a region

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of the cell that has remained challenging to study with existing methods, using a PsbU-APEX2 gene fusion. This study demonstrates the power of APEX2 as a tool to study the cell biology of intracellular features of cyanobacteria with enhanced spatiotemporal resolution.

Keywords: APEX2, proximity-based proteomics, thylakoid lumen, sub-cellular localization, cyanobacteria, photosynthesis, Photosystem II

Introduction

The intracellular spatial organization of cyanobacteria is unique among prokaryotes. As Gram-negative bacteria, cyanobacteria possess the typical inner and outer membrane systems enclosing a cell wall comprised of peptidoglycan. However, most cyanobacterial species also possess thylakoid membranes, an extra set of intracellular membranes where photosynthesis occurs, as well as carboxysomes, proteinaceous organelles used for carbon fixation. The distinctive intracellular spatial organization and protein complexes found within cyanobacteria has drawn particular interest to the cell biology of these organisms. Furthermore, cyanobacteria can also be used as a model for plant chloroplasts, as they share features and have a common evolutionary ancestor. As a result, many proteomic studies of specific cyanobacterial structures, i.e. thylakoid membranes, have been performed (Agarwal et al., 2010; Fulda et al., 2000; Huang et al., 2002, 2004, 2006; Li et al., 2012; Liberton et al., 2016; Pisareva et al., 2007, 2011; Rajalahti et al., 2007; Srivastava et al., 2005; Wang et al., 2000; Zak et al., 2001; Zhang et al., 2009). These studies have made great progress towards understanding the physiology of cyanobacteria, but lack the spatial resolution necessary to resolve the composition of many intracellular compartments resistant to traditional biochemical fractionation and purification methodologies.

Previously, proteomic studies of cyanobacterial components were limited to fractionation and separation techniques which could introduce artifacts and result in ambiguous cellular localizations. Furthermore, existing techniques are impractical for non-membrane bound regions or the thylakoid lumen. However, a technique termed proximity-based proteomics was recently developed in mammalian cells to allow for proteomic analysis of cellular regions or protein interactomes that were unable to be purified using existing techniques (Kim and Roux, 2016). Proximity-based proteomics relies on targeting a specific enzyme to a region of interest as a protein

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fusion to a full-length protein or signal sequence. The enzyme then performs chemistry in live cells to label proteins within a small radius (10-20 nm) of itself (Rhee et al., 2013). After cell lysis, the labeled proteins can then be separated from unlabeled proteins and analyzed using mass spectrometry. Several proximity-based proteomics techniques exist, but the most common use enzymes that biotinylate proteins (Kim and Roux, 2016). We chose to use APEX2, an engineered ascorbate peroxidase that catalyzes a reaction between biotin-phenol (BP) and hydrogen peroxide (H_2O_2) to create a BP radical that covalently attaches to proteins (Hung et al., 2016; Lam et al., 2015) (Figure 1A). The reactivity and short half-life of biotin-phenol gives this technique a high spatial specificity. Furthermore, APEX2 has been shown to be catalytically active in multiple cellular compartments and exhibits a short (1 minute) labeling time, allowing for high temporal specificity (Hung et al., 2016; Lam et al., 2015).

Here we demonstrate the feasibility and potential of proximity-based proteomics technique using APEX2 in *Synechococcus sp.* PCC 7002 (PCC 7002), a model cyanobacterium and promising chassis for biotechnological applications (Markley et al., 2015; Ruffing et al., 2016; Xu et al., 2011). To showcase the ability of APEX2 to interrogate regions of the cell where proteomics studies have not yet been possible due to limitations of existing biochemical methods, we targeted APEX2 to the thylakoid lumen by fusing it to PsbU, an extrinsic photosystem II (PSII) protein (Nishiyama et al., 1998), and identified the PsbU-associated proteome by mass spectrometry. Determining the thylakoid lumen proteome is vital for understanding the physiological roles of the thylakoid membrane system and the reactions of oxygenic photosynthesis.

Results and Discussion

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

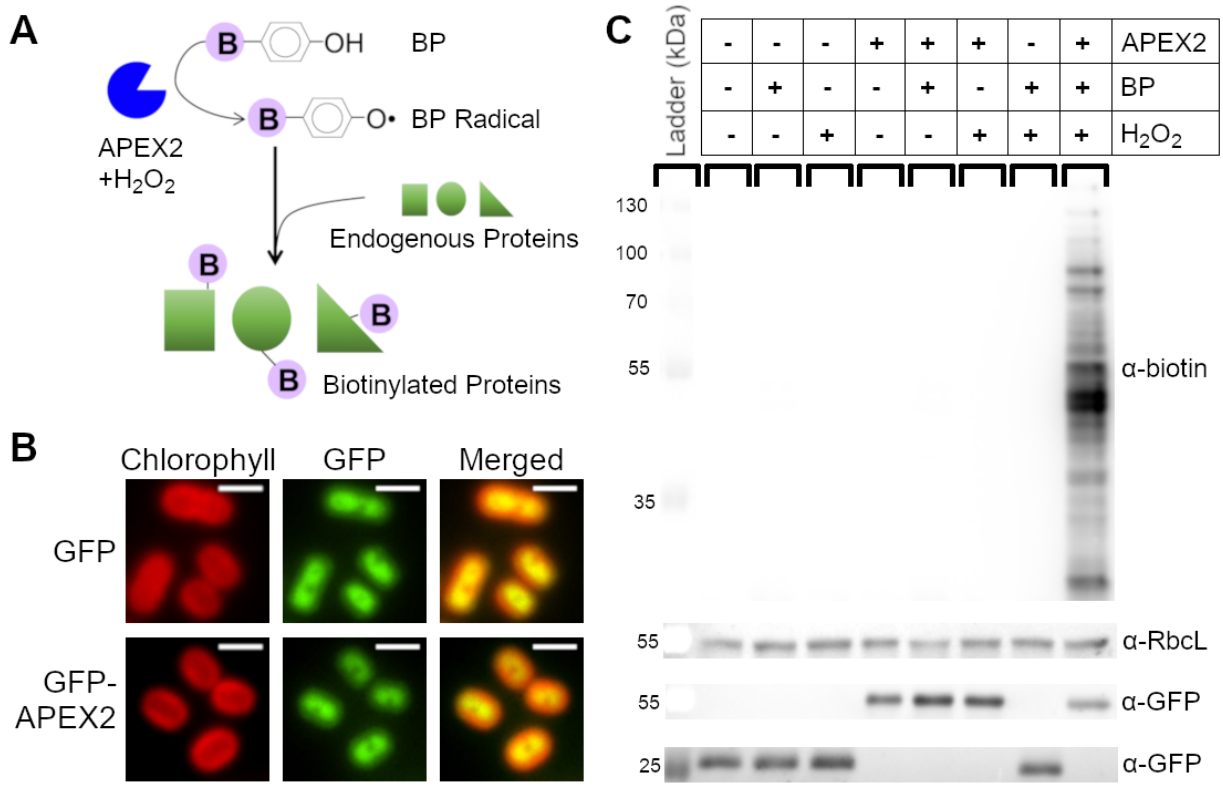


Figure 1. APEX2-dependent labeling specifically biotinylates proteins in PCC 7002.
(A) APEX2 reacts with BP in the presence of H_2O_2 to produce a BP radical. Biotinylated proteins are generated when the BP radical reacts with peptides, forming a covalent bond. **(B)** Fluorescence microscopy imagines of cells expressing GFP and GFP-APEX2 (green). Scale bars are 2 µm. Chlorophyll channel (red) indicates thylakoid membrane. **(C)** 5 µg of protein from cells expressing either GFP or GFP-APEX2 was separated by SDS-PAGE and transferred to a membrane for immunoblot analysis using α-biotin to detect APEX2 activity. α-RbcL was used as a loading control and the same membrane was stripped and re-probed with α-GFP to check for expression of GFP (28 kDa) or GFP-APEX2 (54 kDa).

Characterization of APEX2 labeling in PCC 7002

To determine if APEX2-dependent labeling of proteins was possible in cyanobacteria, GFP or GFP-APEX2 was incorporated into the genome of PCC 7002. Cytoplasmic localization of GFP and GFP-APEX2 was confirmed using fluorescence microscopy (Figure 1B). To perform APEX2-dependent biotinylation, cells were incubated with BP and then exposed to H_2O_2. After quenching

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the reaction, cells were lysed by bead beating and a streptavidin blot confirmed the ability of APEX2 to biotinylate proteins in PCC 7002 (Figure 1C). Biotin labeling only occurred in the presence of APEX2, BP, and H_2O_2, demonstrating reaction specificity in vivo. Furthermore, the rapid reaction enables precise temporal control of labeling.

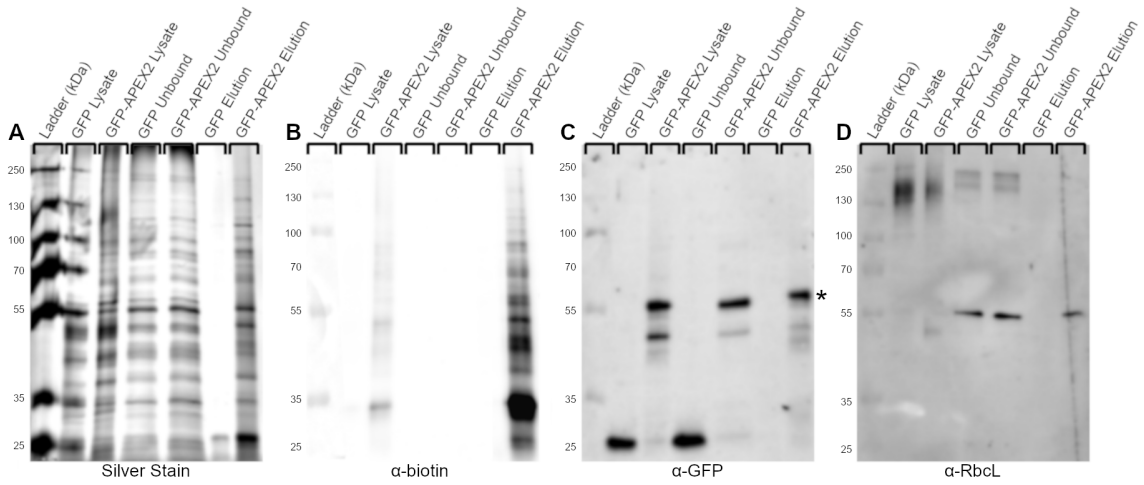


Figure 2. Enrichment of proteins biotinylated by cytoplasmic APEX2 in vivo

Cells expressing GFP or GFP-APEX2 were incubated with BP and exposed to H_2O_2. Biotinylated proteins were captured from cell lysates on streptavidin coated magnetic beads. Fractions from each enrichment step were separated by SDS-PAGE and then silver stained for contrast or transferred to a nitrocellulose membrane and probed with specific antibodies. **(A)** Silver stain of noted fractions from unlabeled (GFP) or labeled (GFP-APEX2) lysates. **(B)** Biotinylated proteins are only detected in fractions containing APEX2 and are enriched on streptavidin beads. **(C)** Expected self-labeling (biotinylation) of GFP-APEX2 (54 kDa, marked with *) is confirmed by immunoblotting against GFP. **(D)** RbcL (55 kDa), a cytoplasmic protein expected to be labeled by GFP-APEX2 was specifically captured on beads incubated with GFP-APEX2.

Purification of cytoplasmic APEX2-biotinylated proteins from PCC 7002

Proteins biotinylated in vivo were enriched for further analysis by affinity purification. APEX2-dependent biotinylation was performed in cells expressing GFP or GFP-APEX2 in the cytoplasm.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Affinity purification of biotinylated proteins was performed by incubating cellular lysates with streptavidin-coated magnetic beads. The background level of biotinylation was very low as biotinylated protein was only detected in cells expressing GFP-APEX2, but not cells expressing GFP alone (Figure 2A-B). To confirm cytoplasmic APEX2 labels cytoplasmic proteins, immunoblots using antibodies against expected cytoplasmic proteins were performed (Figure 2C-D). Since the BP radical reacts with proteins within a 10-20 nm radius of its origin, APEX2 itself is expected to be biotinylated. Biotinylated GFP-APEX2 fusion protein was detected using a α-GFP antibody, confirming the expected self-reactivity (Figure 2C). Additionally, the large subunit of rubisco (ribulose-1,5-bisphosphate carboxylase/oxygenase), RbcL, an abundant cytoplasmic protein, was only enriched on beads cells incubated with cells expressing GFP-APEX2 as detected using a specific α-RbcL antibody (Figure 2D). The high molecular weight RbcL band in lysates is likely the result of higher-order complexes formed in vivo; RbcL is associated with large protein assemblies including the carboxysome in cyanobacteria (Cameron et al., 2013). Following the more stringent enrichment and elution process, these complexes have been disrupted and RbcL migrates as expected.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

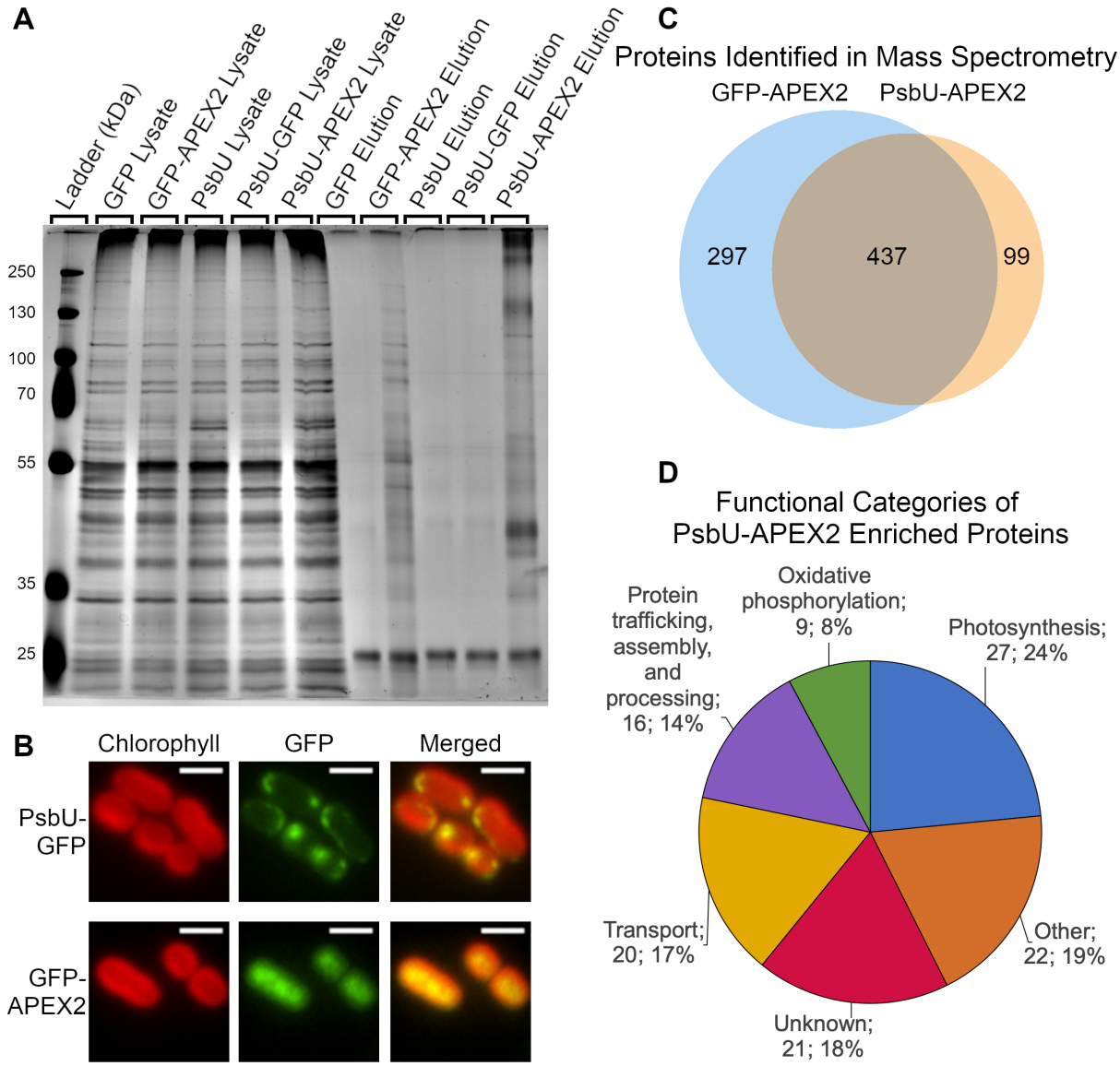


Figure 3. PsbU-APEX2 and Cytoplasmic APEX2 label different sets of proteins.

(A) Silver stain of the biotinylated protein purification from PCC 7002 expressing GFP, GFP-APEX2, PsbU, PsbU-GFP, or PsbU-APEX2 cells exposed to BP and H_2O_2 **(B)** Localization of PsbU-GFP and GFP-APEX2 were visualized with fluorescence microscopy (Green). Chlorophyll channel indicates thylakoid membrane. Scale bars are 2 µm. **(C)** Biotinylated proteins from strains expressing GFP-APEX2 and PsbU-APEX2 identified by mass spectrometry (Also see Supplemental Table 1). **(D)** Functional categories of the proteins enriched in PsbU-APEX2 samples (number of proteins; percentage of 115 total proteins). The proteins used for this analysis are listed in Table 1. (Also see Supplemental Tables 1 and 2).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

PsbU-APEX2 and Cytoplasmic APEX2 label different sets of proteins

To demonstrate the ability of proximity-based proteomics to interrogate subcellular regions that have not been successfully purified using traditional purification methods, APEX2 was fused to an extrinsic subunit of photosystem II (PSII) in an effort to identify proteins in the thylakoid lumen. A PsbU-APEX2 gene fusion expressed from a neutral site in the chromosome under a constitutive promoter was used to localize APEX2 to the thylakoid lumen of PCC 7002. APEX2-dependent labeling and biotinylated protein purification was performed with PsbU-APEX2 and GFP-APEX2. A silver stain of purified biotinylated proteins from GFP-APEX2 and PsbU-APEX2 shows different banding patterns, suggesting that a different set of proteins is labeled by the different APEX2 fusions (Figure 3A). The thylakoid localization of PsbU-GFP was confirmed using fluorescence microscopy (Figure 3B). The localization of PsbU-GFP was used as a proxy for the localization of PsbU-APEX2, since GFP and APEX2 are both C-terminal tags of a similar size. To identify the proteins labeled by the different APEX2 fusion proteins, biotinylated proteins were purified from two independent samples of both PsbU-APEX2 labeled and GFP-APEX2 labeled cells, and the resulting peptides following tryptic digestion were separated and detected using LC-MS/MS. Label-free quantitative methods based on spectral counts were used to compare samples (Old et al., 2005). 99 proteins were identified exclusively in PsbU-APEX2 samples and 279 proteins identified exclusively in GFP-APEX2 samples. 437 proteins were identified in both PsbU-APEX2 and GFP-APEX2 samples (Figure 3C and Supplemental Table 1).

Biotinylated proteins enriched in PsbU-APEX2 samples

Mass spectrometry data was further analyzed to determine what proteins were labeled by PsbU-APEX2. PsbU is a luminal extrinsic subunit of photosystem II and therefore the majority of proteins are expected to be localized to the thylakoid membrane or lumen. However, because PsbU-APEX2 is translated in the cytoplasm and then translocated to its final localization in the lumen, we also expected that a small population of PsbU-APEX2 could be present in the cytoplasm, resulting in labeling of cytoplasmic proteins. Therefore GFP-APEX2 was used as a control instead of a sample lacking APEX2/BP/H_2O_2, since it would control for the small cytoplasmic population of PsbU-APEX2 in addition to proteins nonspecifically bound to the streptavidin beads.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MaxQuant Label Free Quantitation (LFQ) intensities and normalized spectral counts were used to determine the identity of proteins specifically enriched with PsbU-APEX2 compared to the GFP-APEX2 control. Proteins were organized by descending enrichment ($\log_2$(PsbU-APEX2 LFQ intensity or spectral counts/GFP-APEX2 LFQ intensity or spectral counts). A true positive list, which includes proteins with evidence for thylakoid lumen or thylakoid membrane localization, was constructed from proteomic studies including localization data in *Synechocystis sp.* PCC 6803 (PCC 6803) (Agarwal et al., 2010; Fulda et al., 2000; Huang et al., 2002, 2004, 2006; Li et al., 2012; Liberton et al., 2016; Pisareva et al., 2011; Rajalahti et al., 2007, 2007; Srivastava et al., 2005; Wang et al., 2000; Zak et al., 2001; Zhang et al., 2009). A false positive list was constructed from proteins annotated as involved in DNA replication, transcription, or translation, because those proteins are expected to be localized to the cytoplasm (McClure et al., 2016). As expected, proteins from the true positive list have significantly higher enrichment values than proteins from the false positive list (Supplemental Figure 2). Using the true and false positive list, we identified a cutoff to discriminate between enriched proteins and those that bound to the beads non-specifically. This analysis was performed using enrichment values from both PsbU-APEX2 samples and was repeated using enrichment values calculated from normalized spectral counts and from LFQ intensities. The proteins above the cutoff in all four analyses are listed in Table 1. The major functions of enriched proteins are shown in Figure 3D.

The list of PsbU-APEX2 enriched proteins includes many proteins expected to be present within the thylakoid lumen. For example, the extrinsic subunits of the PSII water oxidizing complex (PsbU, PsbQ, PsbO, and PsbV) and Cyt c_6 (PetJ1) were enriched by PsbU-APEX2. Many integral membrane proteins from PSII, photosystem I (PSI), and cytochrome b_6f were also enriched. Furthermore, several factors involved in PSII and PSI assembly and PSII repair were also enriched (PsbP, Ycf48, CtpA, PratA, Psb32, FtsH2, YidC, SecY, Ycf4, SecD, and A1897). Interestingly, PSI extrinsic subunits (PsaC and PsaD) on the cytoplasmic side of PSI were also enriched, although these proteins were also identified in the cytoplasmic GFP-APEX2 samples. Several ATP synthase subunits were also enriched in the PsbU-APEX2 samples. In addition to proteins involved in photosynthetic electron transport, many proteins involved in cellular respiration, metabolite transport, redox regulation, and protein trafficking, processing, and assembly were enriched in PsbU-APEX2 samples.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Many proteins enriched in PsbU-APEX2 samples have previously been found in the periplasm or inner membrane (Table 1). It is unclear what biological relevance this may have. It is possibly an artifact of overexpression of PsbU-APEX2. However, the cyanobacterium *Gloeobacter violaceus* does not contain a thylakoid membrane (Mareš et al., 2013) and instead performs oxygenic photosynthesis in the inner membrane. If the thylakoid membrane and lumen originated from the plasma membrane and periplasmic space, respectively, perhaps it is not surprising that some proteins are found in both cellular fractions. Furthermore, ultrastructural studies of *Synechocystis* sp. PCC 6803 using cryo-electron tomography identified sites of contact between the thylakoid and plasma membrane (Rast et al., 2019). Additional possibilities include dual localization of proteins, low fidelity of the sorting mechanism of translocated proteins into the lumen and the periplasm, and post-translocation sorting of proteins into their final localization. Further experiments are needed to determine the biological relevance of the periplasmic and inner membrane proteins observed. There were several proteins enriched by PsbU-APEX2 that have not been previously localized or have unknown functions. These proteins could be the subject of future research.

The experiments performed here demonstrate the potential of APEX2 to interrogate the proteome of regions of cyanobacteria that have not been previously biochemically purified, like the thylakoid lumen. In the future, this technique can be used to monitor the proteomes of other regions of the cell under different environmental conditions. Additionally, APEX2 can be used to determine the topology of membrane proteins and identify candidates for protein-protein interactions (Lee et al., 2016; Lobingier et al., 2017; Mavylutov et al., 2018; Paek et al., 2017). Proximity-based proteomics using APEX2 has the potential to be a powerful tool in the pursuit of understanding the physiology of photosynthetic organisms.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

244 **Table 1: Proteins Enriched by PsbU-APEX2**

Soluble Proteins Enriched by PsbU-APEX2 Labeling:					
Gene locus	**Gene name**	**PCC 6803 homolog (t)**	**Identified in GFP-APEX2 samples? (u)**	**Function**	**Localizations in PCC 6803**
A0322	*psbU*	sll1194	One	PSII water-oxidizing complex	Thylakoid membrane associated (a, b, c, d) Inner membrane associated (c)
A1605	*psbQ*	sll1638	No	PSII water-oxidizing complex	Thylakoid membrane associated (b, c, e, f) Inner membrane associated (c, g, h) Outer membrane associated (i)
A2695		slr1796	No	Thylakoid-specific thioredoxin	Thylakoid membrane associated (b, f)
A0269	*psbO*	sll0427	Both	PSII water-oxidizing complex	Thylakoid membrane associated (a, b, c, d, f, j) Inner membrane associated (c, g, h, j, k)
A2730	*rfrC*	sll0274	No	Manganese uptake-type porter	Thylakoid membrane associated (b)
A0682	*psaD*	slr0737	Both	PSI	Thylakoid membrane associated (a, b, c, d, e, f, j) Inner membrane associated (c, h, j)
A1589	*psaC*	ssl0563	Both	PSI	Thylakoid membrane associated (a, b, c, f, j) Inner membrane associated (c, h, j)
A2294	*ppiB*	sll0408	No	Putative peptidyl-prolyl cis-trans isomerase	Thylakoid membrane associated (b, f)
A1303	*psbP*	sll1418	One	PSII assembly	Thylakoid membrane associated (b, c, e)
A0112	*psbV*	sll0258	One	PSII water-oxidizing complex	Thylakoid membrane associated (a, b, c) Inner membrane associated (c)
A0167	*petJ1*	sll1796	No	Cytochrome 6	No localization data
A1987	*ytfC*	slr1761	Both	FKBP-type peptidyl-prolyl cis-trans isomerase	Periplasmic (l, m)
A1522		slr1273	No	Biotin carboxylase	Thylakoid membrane associated (b)
A0783		sll0301	No	rfrA-like protein	Thylakoid lumen (l)
A2577	*pqqE*	sll0915	No	Peptidase, insulinase	Thylakoid membrane associated (b) Periplasmic (m), Outer membrane associated (i)
A0229	*ycf48*	slr2034	No	PSII assembly	Thylakoid membrane associated (a, b, c, e) Inner membrane associated (c)

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A1909	*petC*	sll1316	One	Cytochrome b6-f complex iron-sulfur subunit	Thylakoid membrane associated (b, c, d, e, f) Inner membrane associated (c)
A0477	*ymxG*	slr1331	No	Zn-dependent peptidase	Periplasmic (l, m)
A1950		slr0695	No	Unknown	Inner membrane associated (c, h, m) Outer membrane associated (i, m)
A0843	*ctpA*	slr0008	No	PSII assembly; carboxy-terminal-processing protease	Inner membrane associated (b, j)
A0109	*bcp*	sll0221	No	ROS scavenging; peroxiredoxin	No localization data
A1952		slr1087	No	Endonuclease; exonuclease; phosphatase or phytase	Thylakoid membrane associated (b)
A2166		slr1260	No	Unknown	No localization data
A0806		slr1579	One	Putative metalloendopeptidase	No localization data
A1899		sll0886	No	TPR domain protein of unknown function	Thylakoid membrane associated (b)
A1092	*ppib*	sll0227	No	Putative peptidyl-prolyl cis-trans isomerase	Thylakoid membrane associated (b, c) Inner membrane associated (c, g, l) Periplasmic (l, m)
A0948		sll0066	No	Unknown	No localization data
A0775			No	Histidinol phosphatase	No homolog
A2165		sll5034	No	Unknown	Thylakoid membrane associated (b)
A2202	*rfrO*	sll0577	No	Metal porter; pentapeptide repeat protein	No localization data
A0101		sll0148	No	ICE-like protease (Caspase) p20 domain protein	Thylakoid membrane associated (b)
A0086			No	Oxidoreductase with molybdopterin binding domain, oxidizes thioredoxin	No homolog
A1127			No	Unknown	No homolog
A0398	*livK*	slr0447	One	periplasmic-binding protein of a branched-chain amino acid ABC transporter	Thylakoid membrane associated (c) Inner membrane associated (b, c, e, ,g, h, l, n, o) Periplasmic (l, m) Outer membrane associated (l)

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A0190		sll0325	No	Periplasmic component of the Tol biopolymer transport system	Inner membrane associated (b, e, g, l)
A2605		slr1704	No	S-layer protein	Inner membrane associated (b)
A0425	*pratA*	slr2048	No	PSII assembly	Periplasmic (l, m)
A1248	*crtU*	sll0254	No	Beta-carotene desaturase	Thylakoid membrane associated (e)
A1244	*trxA*	sll1980	No	Thioredoxin-like protein	Thylakoid membrane associated (e, f)
A1431		slr1173	No	Unknown	Thylakoid membrane associated (b)
A0213	*prc*	slr1751	No	Carboxyl-terminal protease	Thylakoid membrane associated (f) Inner membrane associated (h) Periplasmic (m) Outer membrane associated (i)
A1207			No	Unknown	No homolog
A2556		sll1488	No	Peptidase	No localization data
A1654	*hhoA/ degQ*	sll1679	No	Serine protease	Inner membrane associated (e, g, l) Periplasmic (l, m)
A1574		sll0180	No	Efflux transporter	Thylakoid membrane associated (c) Inner membrane associated (c, e, g, h, k, l, n, o) Outer membrane associated (i, l)
A1557		slr1624	No	FHA domain protein	Thylakoid membrane associated (b, e) Inner membrane associated (h, k, l)
A2116			No	Tfp pilus assembly protein FimV	No homolog
A0504	*gspD*	slr1277	No	General secretion pathway protein D	Thylakoid membrane associated (d) Inner membrane associated (b, k) Outer membrane associated (i, l)
A1076		slr0971	No	Unknown	Inner membrane associated (b, e)
A0578	*amiC*	slr1744	No	N-acetylmuramoyl-L-alanine amidase	Inner membrane associated (b, e) Periplasmic (l, m)
A2573			No	Alpha-2-macroglobulin; protease inhibitor	No homolog
A1664		sll1573	No	Unknown	No localization data
A2578			One	Unknown	No homolog
G0157			No	Unknown	No homolog
A0339	*pbpB*		No	Penicillin-binding protein 2; important for heterocyst differentiation	No homolog
A1638		slr0191	No	Cell wall hydrolase, SpoIID-like protein	Periplasmic (l, m)

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A2019	*oppA*	sll1699	No	ABC-type dipeptide transport system, periplasmic component	Inner membrane associated (b, c, e, g, h, l, o)
A2507	*sufA*	slr1295	No	ABC-type Fe3+ transport system, periplasmic component (A1 iron uptake protein)	Thylakoid membrane associated (c, f) Inner membrane associated (c, e, g, h, k, n, o) Outer membrane associated (i)
A0064	*ushA*	slr0306	No	Bifunctional metallophosphatase; 5'-nucleotidase	Inner membrane associated (b) Outer membrane associated (l)
A2847			No	Unknown	No homolog
A2056		sll0854	One	Serine hydrolase	No localization data
A1020		slr1431	No	S-layer protein	Inner membrane associated (e)
A0445		slr1740	No	Extracellular solute binding protein specific for oligopeptides	Thylakoid membrane associated (c) Inner membrane associated (b, c, e, g, k, l)
A2324	*bvdR*	slr1784	No	Biliverdin reductase	No localization data
A2554		slr1224	No	Sugar ABC transporter ATP-binding protein	Inner membrane associated (b, c, e)
Membrane Proteins Enriched by PsbU-APEX2 Labeling:					
Gene locus	**Gene name**	**PCC 6803 homolog(t)**	**Identified in GFP-APEX2 samples? (u)**	**Function**	**Localizations in PCC 6803**
A1961	*psaA*	slr1834	Both	PSI core	Thylakoid membrane (a, b, c, e, j) Inner membrane (c, j)
A1330		slr0404	Both	Glycine-rich membrane protein	membrane protein (p)
A1910	*petA*	sll1317	No	Apocytochrome f precursor	Thylakoid membrane (a, b, c, e) Inner membrane (c)
A1962	*psaB*	slr1835	Both	PSI core	Thylakoid membrane (a, b, c, e, f, j) Inner membrane (b, c, j)
A2620	*psaL*	slr1655	Both	PSI integral	Thylakoid membrane (b, c, e, j) Inner membrane (c)
A2619	*psb32*	sll1390	No	PSII assembly factor	Thylakoid membrane (b, e, q) Inner membrane (q)
A2606		slr1106	One	Prohibitin	Thylakoid membrane (b, e) Inner membrane (g, h, k)
A0842	*petB*	slr0342	Both	Cytochrome b6	Thylakoid membrane (c, e, f)

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A1008	*psaF*	sll0819	One	PSI integral	Thylakoid membrane (a, b, c, e, f, j) Inner membrane (c, g, h, o)
A2533	*psbZ*	slr1645	No	PSII integral	Thylakoid membrane (a, b, c, e) Inner membrane (c)
A0349	*ftsH2*	slr0228	Both	PSII repair	Thylakoid membrane (b, e, r)
A1405	*nblS*	sll0698	Both	2-component system sensor for light stress/nutrient stress	Thylakoid membrane (b, e)
A1418; A2164; A0157	*psbA; psbA*	slr1311 ; slr1181	Both	PSII core	Thylakoid membrane (b, c, e, f, j) Inner membrane (b, c, j)
A0381		sll1071	One	Methanol dehydrogenase	Thylakoid membrane (e)
A1759	*psbB*	slr0906	Both	PSII core	Thylakoid membrane (a, b, c, e, j) Inner membrane (b, c)
A0604	*ycf51*	sll1702	No	Unknown	Thylakoid membrane (b)
A0737	*atpG*	sll1323	No	ATP synthase	Thylakoid membrane (b, c, e, f) Inner membrane (c, g, h)
A0776			No	Histidinol phosphatase	No homolog
A1559	*psbC*	sll0851	Both	PSII core	Thylakoid membrane (a, b, c, e, j) Inner membrane (c)
A2400		slr0575	One	Unknown	Thylakoid membrane (b, e, f)
A0736	*atpF*	sll1324	One	ATP synthase	Thylakoid membrane (a, b, c, e, f) Inner membrane (c, g, h)
A0230	*psbE*	ssr345 1	No	PSII integral	Thylakoid membrane (a, b, c, e, j) Inner membrane (j)
A0991	*yidC*	slr1471	Both	YidC/OxaA	Thylakoid membrane (a, b, c, e)
A1433		slr0431	No	Unknown	Thylakoid membrane (d) Inner membrane (h, l) Outer membrane (i, l)
A0727	*ctaCII*	sll0813	One	Cytochrome C oxidase	Inner membrane (b, c, e, g, h, l, n, o)
A1047	*secY*	sll1814	Both	PSII assembly factor	Thylakoid membrane (e, s) (SecY in *Synechococcus* sp. PCC 7942); Inner membrane (s) (SecY in *Synechococcus* sp. PCC 7942)
A0926	*ndhA*	sll0519	Both	NDH complex	Thylakoid membrane (a, b, c, e)
A0238			No	Purine nucleoside permease	No homolog
A1093	*ycf4*	sll0226	One	PSI assembly factor	Thylakoid membrane (b, c, e) Inner membrane (c, j)
A0656	*secD*	slr0774	Both	PSI assembly	Thylakoid membrane (c)
A2372	*napA*		No	Thylakoid sodium-proton antiporter	No homolog
A0728		sll1486	No	Na+-dependent transporters of the SNF family	Inner membrane (b, e)

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A1956		slr0589	No	Cofactor assembly of complex C subunit B	No localization data
A0767	*lepB*	sll0716	No	Signal peptidase I	Thylakoid membrane (b, f) Inner membrane (e)
A1088	*ccs1*	slr2087	One	Cytochrome c biogenesis protein CcsB	Thylakoid membrane (b)
A0189		slr1721	No	Periplasmic Component of the Tol biopolymer transport system	Inner membrane (b, e, n)
A0854	*ndhF*	slr0844	Both	Cytochrome C assembly	Thylakoid membrane (b, e)
A0275		slr0959	No	CAAX amino terminal protease family membrane protein	Thylakoid membrane (e)
A1973	*ndhD 2*		No	NDH complex	No homolog
A1363	*manS*	slr0640	No	Two-component sensor histidine kinase	No localization data
A2176	*ycf37*	slr0171	No	PSI assembly; Tetratricopeptide repeat-containing protein	Thylakoid membrane (b, c, e)
A1897		slr1177	No	PSI assembly factor	Inner membrane (e)
A1231			No	OmpA/MotB outer membrane porin	No homolog
A2547	*ndhB*	sll0223	One	NDH complex	Thylakoid membrane (b, e)
A0465		sll1024	No	Unknown	No localization data
A0794	*sac1*	sll0640	No	Sodium/sulfate symporter	Inner membrane (b, e)
A2000	*ndhD 1*	slr0331	No	NDH complex	Thylakoid membrane (e)
A1964			No	Permease	No homolog
A2551		sll0283	No	Unknown	Inner membrane (b, e)
A1379	*spr*	slr0535	No	Subtilisin-like serine protease	Inner membrane (b, e)

Homologs of PCC 7002 proteins from PCC 6803 (unless noted otherwise) have been localized in previous studies: (a) Agarwal et al., 2010; (b) Liberton et al., 2016; (c) Pisareva et al., 2011; (d) Wang et al., 2000; (e) Baers et al., 2019; (f) Srivastava et al., 2005; (g) Huang et al., 2006; (h) Huang et al., 2002; (i) Huang et al., 2004; (j) Zak et al., 2001; (k) Li et al., 2012; (l) Rajalahti et

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

al., 2007; (m) Fulda et al., 2000; (n) Pisareva et al., 2007; (o) Zhang et al., 2009; (p) Kwon et al., 2010; (q) Wegener et al., 2011; (r) Sacharz et al., 2015; (s) Nakai et al., 1993. Homologs are best reciprocal BLAST hit protein pairs between PCC 7002 and PCC 6803 (t). Number of GFP-APEX2 samples (2 total) that identified this protein (u). All proteins in this list were identified in both PsbU-APEX2 samples.

Methods

Creation of PCC 7002 Strains:

The *psbU* gene (SynPCC7002_A0322) was amplified from PCC 7002 while APEX2 was amplified from a plasmid gifted to us by Alice Ting (Addgene plasmid # 72558 ; http://n2t.net/addgene:72558; RRID:Addgene_72558). Plasmids were assembled using Gibson Assembly (Gibson et al., 2009) with NS1 as the homology arms, p_{ccmK2} as the promoter (Cameron et al., 2013), and kanamycin resistance for selection. The Gibson reactions were transformed into DH5α *E. coli*, and minipreps of liquid cultures started from single colonies were performed to collect plasmid. Plasmid was transformed into PCC 7002 (Stevens and Porter, 1980) and colonies containing the desired insert were serially passaged in the presence of antibiotic until segregated.

Biotinylation of Proteins by APEX2 in PCC 7002:

Biotinylation of proteins was performed using a modified protocol from Hung et al. and Hwang and Espenshade (2016; 2016). Briefly, 50 mL cultures of PCC 7002 strains were grown in A+ media (Stevens et al., 1973) in air at 37˚C with a light intensity of 185 µmol photons • m^{-2} • s^{-1} for 2 days to an OD_{730} of about 0.5. Several µL of culture were saved to image on the microscope. The culture was pelleted at 4300 x *g* for 10 minutes at 4˚C. The supernatant was poured off and cells were resuspended in 4 mL A+ with 2.5 mM BP and transferred to a six-well plate. Six-well plates were incubated shaking in air at 37˚C with a light intensity of 185 µmol photons • m^{-2} • s^{-1} for 30 minutes. Samples were then pelleted in a 1.5 mL tube and resuspended in 1 mL phosphate buffered saline pH 7.8 (Bio-Rad) (PBS). 10 µL of 100 mM H_2O_2 was added and cells were inverted for 30 seconds before pelleting for 30 s. Supernatant was removed and cells were resuspended in quencher solution (PBS with 10 mM sodium ascorbate, 5 mM Trolox and 10 mM sodium azide)

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and pelleted. This step was repeated two additional times. The supernatant was removed and the cell pellets were frozen at -80˚C for storage and to facilitate cell lysis.

Cell Lysis:

The cell pellet was resuspended in RIPA lysis buffer with quenchers (50 mM Tris pH 7.4, 150 mM NaCl, 0.1% (w/v) SDS, 0.5% (w/v) sodium deoxycholate, 1% (v/v) Triton X-100, 10 mM sodium ascorbate, 5 mM Trolox, 10 mM sodium azide, 1 mM PMSF). Cells were lysed using bead beating, with 30 cycles of 20 seconds on and 20 seconds off on ice. The lysate and beads were pelleted at 5000 x g in a microfuge and the supernatant was collected. The supernatant was clarified to remove debris and unbroken cells following centrifugation at 21,100 x g for 5 min at 4˚C.

Protein Concentration Measurement:

The protein concentration of cell lysate was quantified using the Pierce 660 nm Protein Assay (Thermo Fisher).

Purification of Biotinylated Proteins:

Streptavidin magnetic beads (Pierce) were washed twice in RIPA lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 0.1% (w/v) SDS, 0.5% (w/v) sodium deoxycholate, 1% (v/v) Triton X-100) and the supernatant was removed. 800 µL of RIPA lysis buffer with quenchers containing 50 µg of protein for every 50 µL of streptavidin magnetic beads was added. Beads were incubated with protein for 1 hour at room temperature on a rotator. The beads were then washed twice with RIPA lysis buffer, once with 1M KCl, once with 0.1 M Na_2CO_3, once with 8 M urea in 10 mM Tris pH 7.5, and once again with RIPA lysis buffer.

Elution of Biotinylated Proteins for gels and blots:

Beads were boiled in 30 µL of elution buffer (3X Laemmli buffer, 2 mM biotin, 20 mM DTT) to elute biotinylated proteins. The eluate was collected and diluted with 60 µL of water to run on gels.

Preparation for Mass Spectrometry:

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Beads were washed an additional 5 times with 50 mM NH_4HCO_3 containing 0.2% (w/v) sodium deoxycholate. The supernatant was removed and beads were resuspended in 50 µL 10 mM TCEP and 40 mM chloroacetamide and incubated at 37˚C for 30 minutes to reduce and alkylate the proteins. 150 µL water containing 0.225% (w/v) sodium deoxycholate and 0.2 µg Promega sequencing grade modified trypsin was added. An on-bead digestion was performed overnight on a rotator at 37˚C. Beads were pelleted and the supernatant was collected. Formic acid was added to 2% (w/v) to stop digestion. Sodium deoxycholate was removed using 3 phase transfers with ethyl acetate. The samples were desalted using in-house STAGE tips with 3M Empore SDB-RPS membrane and dried using a vacuum centrifugation.

LC-MS/MS:

The tryptic peptides were resolved using an UltiMate 3000 UHPLC system (Thermo Fisher) in a direct injection mode. Peptides were reconstituted in Buffer A (0.1% formic acid in water), and peptide concentration was measured using Fluoraldehyde o-Phthaldialdehyde Reagent (Thermo Fisher). For each sample, 250 ng (5 µL) of the peptides were loaded onto a Waters BEH C18 column (130 Å, 1.7 µm × 75 µm × 250 mm) with 98.4% Buffer A and 1.6% Buffer B (0.1% formic acid in acetonitrile) at 0.4 µL/min for 16.67 min. Peptides were resolved and eluted using a gradient of 1.6 to 8% B (0-8 min), 8-20% B (8-140 min), and 20-32% B (140-160 min) at 0.3 µL/min. MS/MS was performed on a Q-Exactive HF-X mass spectrometer (Thermo Fisher), scanning precursor ions between 380-1580 m/z (60,000 resolution, 3 x 10^6 ions AGC target, 45 msec maximum ion fill time), and selecting the 12 most intense ions for MS/MS (15,000 resolution, 1 x 10^5 ions AGC target, 150 ms maximum ion fill time, 1.4 m/z isolation window, 27 NCE, 30 s dynamic exclusion). Ions with unassigned charge state, +1, and > +7 were excluded from the MS/MS.

Silver Stain Protocol:

Proteins were separated on a 10% SDS-PAGE gel and stained using the short silver nitrate staining protocol described in by Chevallet et al. (2006).

Immunoblotting:

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Proteins were separated on a 10% SDS-PAGE gel and immunoblots were performed following the protocol from Green and Sambrook (2012). Protein was transferred to a nitrocellulose membrane, or a polyvinylidene fluoride (PVDF) membrane if fluorescent secondary antibodies were used. After blocking membranes overnight, membranes were incubated with GFP (Invitrogen, cat. no. A6455) or RbcL (Agrisera, cat. no. AS03037) antibodies, or streptavidin-HRP (Life Technologies, cat. no. R960-25). Membranes probed for GFP or RbcL were then incubated with a secondary antibody conjugated to HRP or AlexaFluor 488 (Thermo Fisher, cat. no. A-11008 or cat. no. 31460). Membranes were visualized using chemiluminescence after exposure to the Clarity Western ECL substrate (Bio-Rad) or fluorescence. If necessary, blots were stripped using ReBlot Plus Mild Solution (Millipore).

Fluorescence Microscopy:

Cells were spotted onto an agar pad (A+ with 1% agar) and placed onto a microscope slide. Cells were imaged on a customized Nikon TiE inverted wide-field microscope with a Near-IR-based Perfect Focus system. Images were acquired with an ORCA Flash4.0 V2+ Digital sCMOS camera (Hamamatsu) using a Nikon CF160 Plan Apochromat Lambda 100x oil immersion objective (1.45 N.A.). Chlorophyll fluorescence of thylakoid membranes was imaged using a 640 nm LED light source (SpectraX) for excitation and a standard Cy5 emission filter. GFP localization was imaged using a 470 nm LED light source (SpectraX) for excitation and a standard GFP emission filter.

LC-MS/MS Data Analysis:

Only proteins with at least two unique peptides and two spectral counts were considered identified. Proteins identified in both GFP-APEX2 replicates and/or both PsbU-APEX2 replicates were retained for further analysis. A presence/absence Venn diagram was constructed (Figure 3C). A protein must be identified in both replicates of a sample to appear in the Venn diagram. Proteins identified in both replicates of a sample and only one replicate of the other sample (176 proteins) were not added to the Venn diagram as their localization was unclear.

The $\log_2$ ratio of the MAXQUANT LFQ intensities and the $\log_2$ ratio of normalized spectral counts were used as metrics to determine enrichment in the PsbU-APEX2 A and PsbU-APEX2 B samples over the GFP-APEX2 B sample ($\log_2$(U/G)) (Old et al., 2005). If a protein was not identified in a

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sample, the LFQ intensity was set to zero. To determine the cutoff for proteins enriched in PsbU-APEX2 samples, identified proteins were cross-referenced with true positive (TP) or false positive (FP) lists. The TP lists were assembled using localization data from proteomic studies in PCC 6803 (Agarwal et al., 2010; Fulda et al., 2000; Huang et al., 2002, 2004, 2006; Li et al., 2012; Liberton et al., 2016; Pisareva et al., 2007, 2011; Rajalahti et al., 2007; Srivastava et al., 2005; Wang et al., 2000; Zak et al., 2001; Zhang et al., 2009). The FP list contained proteins annotated as DNA binding or involved in transcription or translation by McClure et al. (2016). TP proteins included all thylakoid lumen protein homologs from Rajalahti et al. (2007). PCC 7002 homologs of proteins found associated with the thylakoid membrane by Pisareva et al. (2011) and at least one other study were included in the TP list, otherwise TP protein homologs were found in at least three thylakoid membrane studies. TP proteins were required to possess a secretion signal or at least one transmembrane helix to be retained in the TP list and had to be found in equal to or more thylakoid proteomics studies than inner membrane/outer membrane/periplasm studies (Supplemental Table 3).

A total of four analyses were performed, one for each enrichment metric in each PsbU-APEX2 sample. For each protein in every analysis, the true positive rate (TPR) and the false positive rate (FPR) were calculated using only the proteins with enrichment equal to or greater than the protein of interest. The TPR was the number of TP proteins found at or above the cutoff divided by the total number of TP proteins found in the experiment. The FPR was the number of FP proteins found at or above the cutoff divided by the total number of FP proteins. The cutoff for each sample was the enrichment with the greatest TPR-FPR value. The proteins above the cut-off of the in all 4 analyses are reported in Table 1.

Signal Sequence Prediction:

To predict if a protein had a signal sequence and the cut site to the remove the signal sequence, all proteins in the UniProt reference proteome for PCC 7002 were analyzed with SignalP-5.0 using both the Gram-positive and Gram-negative bacterial options.

Transmembrane Helices Prediction

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To predict if a protein had transmembrane helices, all proteins in the UniProt reference proteome for PCC 7002 were analyzed using the TMHMM Server v. 2.0.

Author Contributions

K.K.D., T.L., and J.C.C. designed experiments. K.K.D. and T.L. performed experiments. K.K.D. and C.G. analyzed data. K.K.D. and J.C.C. wrote the manuscript.

Acknowledgements

We would like to thank Patrick Thomas for the original design of plasmid constructs and Kristin A. Moore for critical analysis of manuscript. We thank all members of the Cameron laboratory for helpful discussions when designing experiments and analyzing data. This material is based upon work supported in part by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0019306 to J.C.C. This work was supported in part by the Interdisciplinary Quantitative Biology (IQ Biology) program at the BioFrontiers Institute, University of Colorado, Boulder, NSF IGERT grant number 1144807. Financial support for this study was also provided by startup funds from the University of Colorado-Boulder to J.C.C. The authors declare no conflicts of interests.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Agarwal, R., Matros, A., Melzer, M., Mock, H.-P., and Sainis, J.K. (2010). Heterogeneity in thylakoid membrane proteome of Synechocystis 6803. Journal of Proteomics *73*, 976–991.

Baers, L.L., Breckels, L.M., Mills, L.A., Gatto, L., Deery, M.J., Stevens, T.J., Howe, C.J., Lilley, K.S., and Lea-Smith, D.J. (2019). Proteome Mapping of a Cyanobacterium Reveals Distinct Compartment Organization and Cell-Dispersed Metabolism. Plant Physiology *181*, 1721–1738.

Cameron, J.C., Wilson, S.C., Bernstein, S.L., and Kerfeld, C.A. (2013). Biogenesis of a bacterial organelle: the carboxysome assembly pathway. Cell *155*, 1131–1140.

Chevallet, M., Luche, S., and Rabilloud, T. (2006). Silver staining of proteins in polyacrylamide gels. Nat Protoc *1*, 1852–1858.

Fulda, S., Huang, F., Nilsson, F., Hagemann, M., and Norling, B. (2000). Proteomics of Synechocystis sp. strain PCC 6803. European Journal of Biochemistry *267*, 5900–5907.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison Iii, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods *6*, 343–345.

Green, M.R., and Sambrook, J. (2012). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press).

Huang, F., Parmryd, I., Nilsson, F., Persson, A.L., Pakrasi, H.B., Andersson, B., and Norling, B. (2002). Proteomics of Synechocystis sp. Strain PCC 6803: Identification of Plasma Membrane Proteins. Molecular & Cellular Proteomics *1*, 956–966.

Huang, F., Hedman, E., Funk, C., Kieselbach, T., Schröder, W.P., and Norling, B. (2004). Isolation of Outer Membrane of Synechocystis sp. PCC 6803 and Its Proteomic Characterization. Molecular & Cellular Proteomics *3*, 586–595.

Huang, F., Fulda, S., Hagemann, M., and Norling, B. (2006). Proteomic screening of salt-stress-induced changes in plasma membranes of Synechocystis sp. strain PCC 6803. PROTEOMICS *6*, 910–920.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

457 Hung, V., Udeshi, N.D., Lam, S.S., Loh, K.H., Cox, K.J., Pedram, K., Carr, S.A., and Ting, A.Y.
458 (2016). Spatially resolved proteomic mapping in living cells with the engineered peroxidase
459 APEX2. Nat Protoc *11*, 456–475.

460 Hwang, J., and Espenshade, P.J. (2016). Proximity-dependent biotin labelling in yeast using the
461 engineered ascorbate peroxidase APEX2. Biochemical Journal *473*, 2463–2469.

462 Kim, D.I., and Roux, K.J. (2016). Filling the Void: Proximity-Based Labeling of Proteins in Living
463 Cells. Trends in Cell Biology *26*, 804–817.

464 Kwon, J., Oh, J., Park, C., Cho, K., Kim, S.I., Kim, S., Lee, S., Bhak, J., Norling, B., and Choi, J.-
465 S. (2010). Systematic cyanobacterial membrane proteome analysis by combining acid hydrolysis
466 and digestive enzymes with nano-liquid chromatography–Fourier transform mass spectrometry.
467 Journal of Chromatography A *1217*, 285–293.

468 Lam, S.S., Martell, J.D., Kamer, K.J., Deerinck, T.J., Ellisman, M.H., Mootha, V.K., and Ting,
469 A.Y. (2015). Directed evolution of APEX2 for electron microscopy and proximity labeling. Nat
470 Meth *12*, 51–54.

471 Lee, S.-Y., Kang, M.-G., Park, J.-S., Lee, G., Ting, A.Y., and Rhee, H.-W. (2016). APEX
472 Fingerprinting Reveals the Subcellular Localization of Proteins of Interest. Cell Reports *15*, 1837–
473 1847.

474 Li, T., Yang, H.-M., Cui, S.-X., Suzuki, I., Zhang, L.-F., Li, L., Bo, T.-T., Wang, J., Murata, N.,
475 and Huang, F. (2012). Proteomic Study of the Impact of Hik33 Mutation in Synechocystis sp. PCC
476 6803 under Normal and Salt Stress Conditions. J. Proteome Res. *11*, 502–514.

477 Liberton, M., Saha, R., Jacobs, J.M., Nguyen, A.Y., Gritsenko, M.A., Smith, R.D., Koppenaal,
478 D.W., and Pakrasi, H.B. (2016). Global Proteomic Analysis Reveals an Exclusive Role of
479 Thylakoid Membranes in Bioenergetics of a Model Cyanobacterium. Mol Cell Proteomics *15*,
480 2021–2032.

481 Lobingier, B.T., Hüttenhain, R., Eichel, K., Miller, K.B., Ting, A.Y., von Zastrow, M., and
482 Krogan, N.J. (2017). An Approach to Spatiotemporally Resolve Protein Interaction Networks in
483 Living Cells. Cell *169*, 350-360.e12.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mareš, J., Hrouzek, P., Kaňa, R., Ventura, S., Strunecký, O., and Komárek, J. (2013). The Primitive Thylakoid-Less Cyanobacterium Gloeobacter Is a Common Rock-Dwelling Organism. PLOS ONE *8*, e66323.

Markley, A.L., Begemann, M.B., Clarke, R.E., Gordon, G.C., and Pfleger, B.F. (2015). Synthetic Biology Toolbox for Controlling Gene Expression in the Cyanobacterium Synechococcus sp. strain PCC 7002. ACS Synth. Biol. *4*, 595–603.

Mavylutov, T., Chen, X., Guo, L., and Yang, J. (2018). APEX2- tagging of Sigma 1-receptor indicates subcellular protein topology with cytosolic N-terminus and ER luminal C-terminus. Protein Cell *9*, 733–737.

McClure, R.S., Overall, C.C., McDermott, J.E., Hill, E.A., Markillie, L.M., McCue, L.A., Taylor, R.C., Ludwig, M., Bryant, D.A., and Beliaev, A.S. (2016). Network analysis of transcriptomics expands regulatory landscapes in Synechococcus sp. PCC 7002. Nucleic Acids Res *44*, 8810–8825.

Nakai, M., Sugita, D., Omata, T., and Endo, T. (1993). Sec-Y Protein Is Localized in Both the Cytoplasmic and Thylakoid Membranes in the Cyanobacterium Synechococcus PCC7942. Biochemical and Biophysical Research Communications *193*, 228–234.

Nishiyama, Y., Los, D.A., and Murata, N. (1998). Role of Psbu, an Extrinsic Protein of Photosystem II, In the Acquisition of Thermotolerance in Synechococcus sp. PCC 7002. In Photosynthesis: Mechanisms and Effects: Volume I–V: Proceedings of the XIth International Congress on Photosynthesis, Budapest, Hungary, August 17–22, 1998, G. Garab, ed. (Dordrecht: Springer Netherlands), pp. 2449–2452.

Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A., and Ahn, N.G. (2005). Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics. Molecular & Cellular Proteomics *4*, 1487–1502.

Paek, J., Kalocsay, M., Staus, D.P., Wingler, L., Pascolutti, R., Paulo, J.A., Gygi, S.P., and Kruse, A.C. (2017). Multidimensional Tracking of GPCR Signaling via Peroxidase-Catalyzed Proximity Labeling. Cell *169*, 338-349.e11.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

511 Pisareva, T., Shumskaya, M., Maddalo, G., Ilag, L., and Norling, B. (2007). Proteomics of
512 Synechocystis sp. PCC 6803. The FEBS Journal *274*, 791–804.

513 Pisareva, T., Kwon, J., Oh, J., Kim, S., Ge, C., Wieslander, Å., Choi, J.-S., and Norling, B. (2011).
514 Model for Membrane Organization and Protein Sorting in the Cyanobacterium Synechocystis sp.
515 PCC 6803 Inferred from Proteomics and Multivariate Sequence Analyses. J. Proteome Res. *10*,
516 3617–3631.

517 Rajalahti, T., Huang, F., Rosén Klement, M., Pisareva, T., Edman, M., Sjöström, M., Wieslander,
518 Å., and Norling, B. (2007). Proteins in Different Synechocystis Compartments Have
519 Distinguishing N-Terminal Features: A Combined Proteomics and Multivariate Sequence
520 Analysis. J. Proteome Res. *6*, 2420–2434.

521 Rast, A., Schaffer, M., Albert, S., Wan, W., Pfeffer, S., Beck, F., Plitzko, J.M., Nickelsen, J., and
522 Engel, B.D. (2019). Biogenic regions of cyanobacterial thylakoids form contact sites with the
523 plasma membrane. Nat. Plants *5*, 436–446.

524 Rhee, H.-W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A., and Ting, A.Y.
525 (2013). Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic
526 Tagging. Science *339*, 1328–1331.

527 Ruffing, A.M., Jensen, T.J., and Strickland, L.M. (2016). Genetic tools for advancement of
528 Synechococcus sp. PCC 7002 as a cyanobacterial chassis. Microbial Cell Factories *15*, 190.

529 Sacharz, J., Bryan, S.J., Yu, J., Burroughs, N.J., Spence, E.M., Nixon, P.J., and Mullineaux, C.W.
530 (2015). Sub-cellular location of FtsH proteases in the cyanobacterium S ynechocystis sp. PCC
531 6803 suggests localised PSII repair zones in the thylakoid membranes. Mol Microbiol *96*, 448–
532 462.

533 Srivastava, R., Pisareva, T., and Norling, B. (2005). Proteomic studies of the thylakoid membrane
534 of Synechocystis sp. PCC 6803. PROTEOMICS *5*, 4905–4916.

535 Stevens, S.E., and Porter, R.D. (1980). Transformation in Agmenellum quadruplicatum. PNAS *77*,
536 6052–6056.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Stevens, S.E., Patterson, C.O.P., and Myers, J. (1973). The Production of Hydrogen Peroxide by Blue-Green Algae: A Survey1. Journal of Phycology *9*, 427–430.

Wang, Y., Sun, J., and Chitnis, P.R. (2000). Proteomic study of the peripheral proteins from thylakoid membranes of the cyanobacterium Synechocystis sp. PCC 6803. ELECTROPHORESIS *21*, 1746–1754.

Wegener, K.M., Bennewitz, S., Oelmüller, R., and Pakrasi, H.B. (2011). The Psb32 Protein Aids in Repairing Photodamaged Photosystem II in the Cyanobacterium Synechocystis 6803. Molecular Plant *4*, 1052–1061.

Xu, Y., Alvey, R.M., Byrne, P.O., Graham, J.E., Shen, G., and Bryant, D.A. (2011). Expression of Genes in Cyanobacteria: Adaptation of Endogenous Plasmids as Platforms for High-Level Gene Expression in Synechococcus sp. PCC 7002. In Photosynthesis Research Protocols, R. Carpentier, ed. (Totowa, NJ: Humana Press), pp. 273–293.

Zak, E., Norling, B., Maitra, R., Huang, F., Andersson, B., and Pakrasi, H.B. (2001). The initial steps of biogenesis of cyanobacterial photosystems occur in plasma membranes. PNAS *98*, 13443–13448.

Zhang, L.-F., Yang, H.-M., Cui, S.-X., Hu, J., Wang, J., Kuang, T.-Y., Norling, B., and Huang, F. (2009). Proteomic Analysis of Plasma Membranes of Cyanobacterium Synechocystis sp. Strain PCC 6803 in Response to High pH Stress. J. Proteome Res. *8*, 2892–2902.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends

Figure 1. APEX2-dependent labeling specifically biotinylates proteins in PCC 7002.
(A) APEX2 reacts with BP in the presence of H_2O_2 to produce a BP radical. Biotinylated proteins are generated when the BP radical reacts with peptides, forming a covalent bond. **(B)** Fluorescent microscopy imagines of cells expressing GFP and GFP-APEX2 (green). Scale bars are 2 µm. Chlorophyll channel (red) indicates thylakoid membrane. **(C)** 5 µg of protein from cells expressing either GFP or GFP-APEX2 was separated by SDS-PAGE and transferred to a membrane for immunoblot analysis using α-biotin to detect for APEX2 activity. α-RbcL was used as a loading control and the same membrane was stripped and re-probed with α-GFP to check for expression of GFP (28 kDa) or GFP-APEX2 (54 kDa).

Figure 2. Enrichment of proteins biotinylated by cytoplasmic APEX2 in vivo
Cells expressing GFP or GFP-APEX2 were incubated with BP and exposed to H_2O_2. Biotinylated proteins were captured from cell lysates on streptavidin coated magnetic beads. Fractions from each enrichment step were separated by SDS-PAGE and then silver stained for contrast or transferred to a nitrocellulose membrane and probed with specific antibodies. **(A)** Silver stain of noted fractions from unlabeled (GFP) or labeled (GFP-APEX2) lysates. **(B)** Biotinylated proteins are only detected in fractions containing APEX2 and are enriched on streptavidin beads. **(C)** Expected self-labeling (biotinylation) of GFP-APEX2 (54 kDa, marked with *) is confirmed by immunoblotting against GFP. **(D)** RbcL (55 kDa), a cytoplasmic protein expected to be labeled by GFP-APEX2 was specifically captured on beads incubated with GFP-APEX2.

Figure 3. PsbU-APEX2 and Cytoplasmic APEX2 label different sets of proteins.
(A) Silver stain of the biotinylated protein purification from PCC 7002 expressing GFP, GFP-APEX2, PsbU, PsbU-GFP, or PsbU-APEX2. **(B)** Localization of PsbU-GFP and GFP-APEX2 were visualized with fluorescence microscopy (Green). Chlorophyll channel indicates thylakoid membrane. Scale bars are 2 µm. **(C)** Biotinylated proteins from strains expressing GFP-APEX2 and PsbU-APEX2 identified by mass spectrometry. **(D)** Functional categories of the proteins enriched in PsbU-APEX2 samples (number of proteins; percentage of 115 total proteins). The proteins used for this analysis are listed in Table 1. (Also see Supplemental Tables 1 and 2).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SUPPLEMENTAL INFORMATION

This manuscript contains the following supplemental information:

Supplemental Data Analysis

Supplemental Figure 1. Comparison of enrichment values between samples.

Supplemental Figure 2. Comparison of FP and TP enrichment values.

Supplemental Figure 3. ROC curves and enrichment cutoffs.

Supplemental Table 1. Proteomics data and analysis

Supplemental Table 2. Additional information about enriched proteins

Supplemental Table 3. List of true and false positive proteins for analysis

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

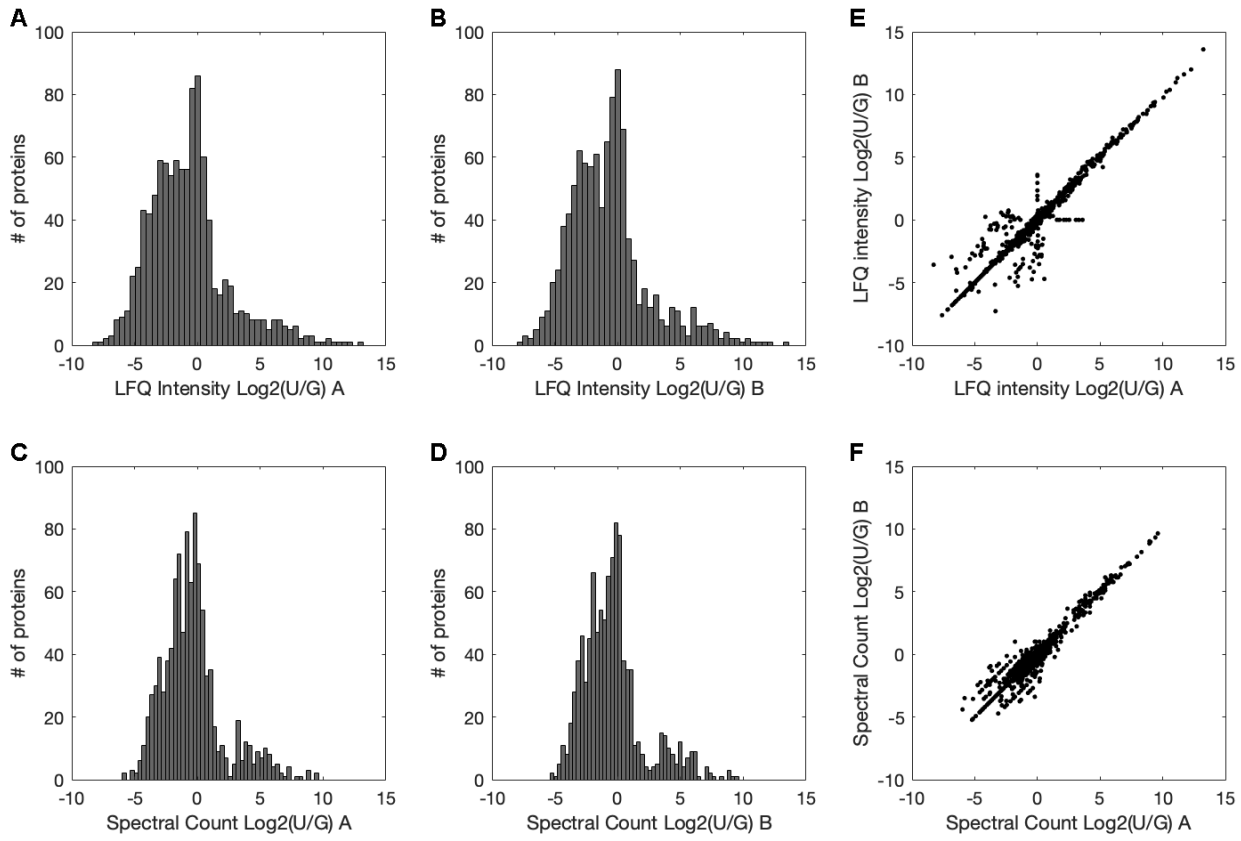
Supplemental analysis of LC/MS-MS data:

After calculating the enrichment metrics for each analyzed protein, these values were plotted in histograms to check for normal distributions (Supplemental Figure 1A-D). Furthermore, the enrichment ratios from each sample for the same sample were compared against each other to check for correlation between samples (Supplemental Figure 1E-F). Pearson's correlation coefficient was calculated for the enrichment values of samples A and B using both LFQ intensities and spectral counts. For both the LFQ intensities and spectral counts, the enrichment values had a Pearson's correlation coefficient of 0.96, indicating a strong positive correlation between samples.

Next, box plots were made comparing the enrichment distribution of all proteins, TP proteins, and FP proteins for each sample and each enrichment metric (Supplemental Figure 2A-D). In each comparison, the enrichment of TP proteins had a greater mean than the FP proteins. A student t-test found significant ($p<0.5$) differences between the enrichment values of TP and FP proteins in both enrichment metrics in each sample, supporting the fact that the experiment did enrich for proteins biotinylated by PsbU-APEX2 localized in the thylakoid lumen.

After confirming that thylakoid lumen and thylakoid membrane proteins were enriched in the experiments, Receiver Operating Curve (ROC) were plotted for each sample. For this analysis, proteins were organized in descending enrichment value for both enrichment metrics in each sample (Supplemental Figure 3A-D). The True Positive Rate (TPR), was calculated for each protein as the % of all identified TP proteins with an enrichment greater than or equal to itself. Conversely, the False Positive Rate (FPR) was calculated for each protein as the % of all identified FP proteins with an enrichment greater than or equal to itself. In ROCs, the TPR is plotted against FPR for each individual protein and the graph is examined to determine if the line arcs over a TPR = FPR line, which demonstrates that TPs are enriched in the experiment. For each enrichment metric in each sample, the ROC curve arced over the TPR = FPR line, indicating that TP proteins were enriched in the experiment. Furthermore, the cutoff for enriched proteins was placed at the enrichment value were the TPR-FPR was at a maximum. To visualize this, the TPR - FPR was plotted against enrichment in each sample (Supplemental Figure 3E-H).

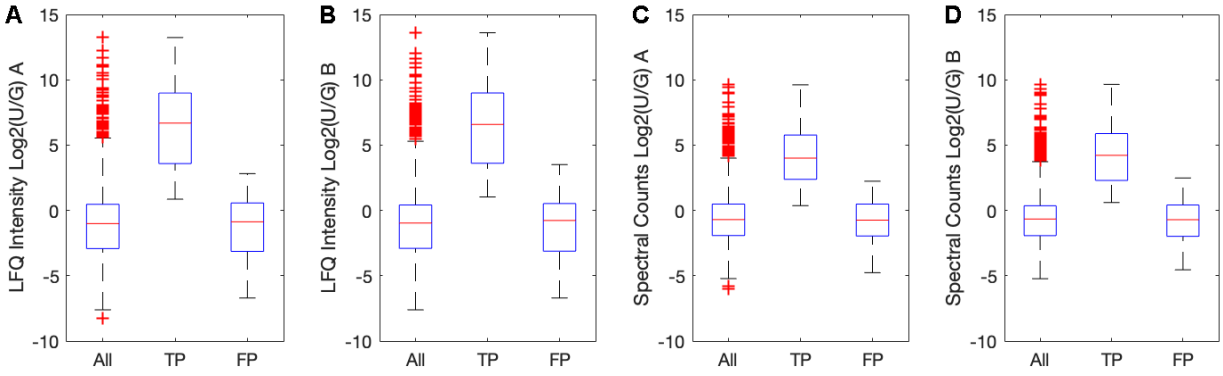
bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplemental Figure 1. Comparison of enrichment values between samples.

(A) Histogram of enrichment using LFQ intensities from PsbU-APEX2 sample A. **(B)** Histogram of enrichment using LFQ intensities from PsbU-APEX2 sample B. **(C)** Histogram of enrichment using spectral counts from PsbU-APEX2 sample A. **(D)** Histogram of enrichment using spectral counts from PsbU-APEX2 sample B. **(E)** Comparison of LFQ intensity enrichment values for each analyzed protein between PsbU-APEX2 samples A and B. PCC is 0.96. **(F)** Comparison of spectral count enrichment values for each analyzed protein between PsbU-APEX2 samples A and B. PCC is 0.96.

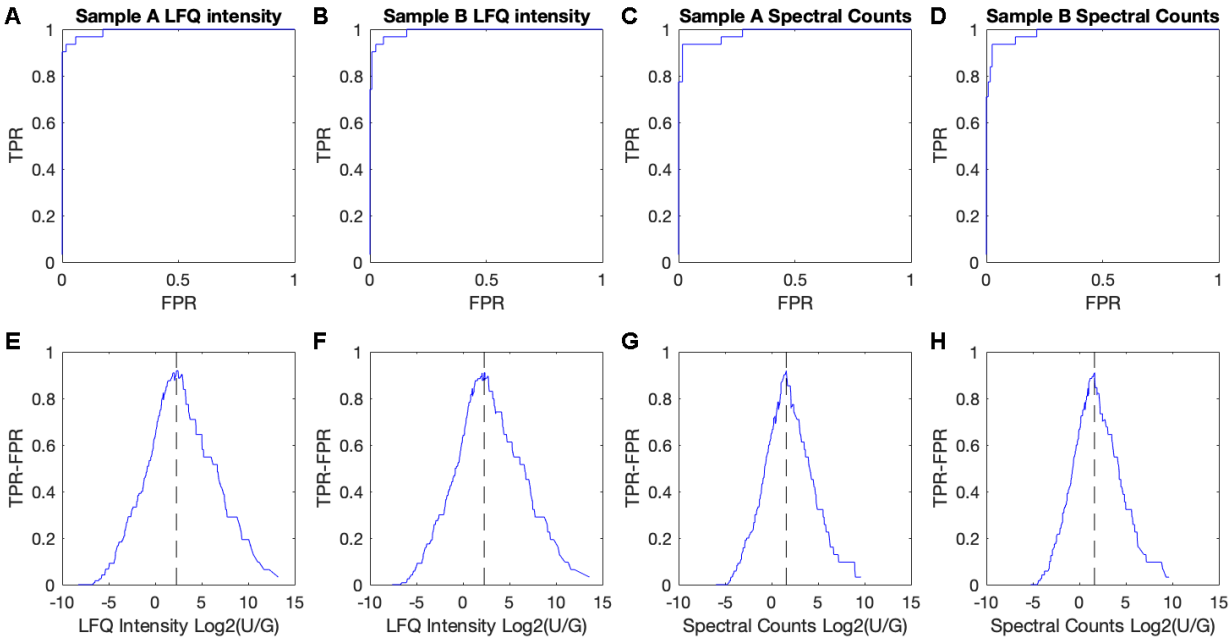
bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplemental Figure 2. Comparison of FP and TP enrichment values.

(A) Distributions of enrichment values calculated using LFQ intensity in sample A. **(B)** Distributions of enrichment values calculated using LFQ intensity in sample B. **(C)** Distributions of enrichment values calculated using spectral counts in sample A. **(D)** Distributions of enrichment values calculated using spectral counts in sample B. All panels show the distribution enrichment values for all proteins, True Positive (TP) proteins, and False Positive (FP) proteins in a sample in box and whisker plots. In each sample the TP enrichment values are shifted up compared to the FP enrichment values. A Student's t-test found significant differences ($p < 0.05$) between the TP and FP proteins in each sample (A, $p = 4 \times 10^{-15}$; B, $p = 7 \times 10^{-15}$; C, $p = 6 \times 10^{-14}$; D, $p = 5 \times 10^{-14}$). Red crosses signify outliers, which are greater than 75^{th} percentile + 2.97 • standard deviation • (75^{th} percentile-25^{th} percentile) or less than 25^{th} percentile – 2.97 • standard deviation • (75^{th} percentile-25^{th} percentile).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.030940; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplemental Figure 3. ROC curves and enrichment cutoffs.

The top row of samples shows ROC curves **(A-D)**, which plot TPR vs FPR to determine if the curve arcs over the TPR=FPR line. In all cases the ROC arcs over the TPR=FPR line, indicating that TP proteins are enriched in the experiment. The bottom row of samples shows TPR-FPR vs enrichment metrics **(E-H)**. The cutoff was drawn at the enrichment with the greatest TPR-FPR and is indicated by the dashed line. **(A)** ROC curve made using the enrichment values calculated using LFQ intensity in sample A. **(B)** ROC curve made using the enrichment values calculated using LFQ intensity in sample B. **(C)** ROC curve made using the enrichment values calculated using spectral counts in sample A. **(D)** ROC curve made using the enrichment values calculated using spectral counts in sample B. **(E)** TPR-FPR vs enrichment in sample A calculated using LFQ intensity. Cutoff is at 2.28. **(F)** TPR-FPR vs enrichment in sample B calculated using LFQ intensity. Cutoff is at 2.28. **(G)** TPR-FPR vs enrichment in sample A calculated using spectral counts. Cutoff is at 1.58. **(H)** TPR-FPR vs enrichment in sample B calculated using spectral counts. Cutoff is at 1.62.