Global Proteomic Analysis Reveals an Exclusive Role of Thylakoid Membranes in Bioenergetics of a Model Cyanobacterium*⒮

Michelle Liberton‡‖, Rajib Saha‡‖, Jon M. Jacobs§‖, Amelia Y. Nguyen‡,
Marina A. Gritsenko§, Richard D. Smith§, David W. Koppenaal§,
and Himadri B. Pakrasi‡¶

Cyanobacteria are photosynthetic microbes with highly differentiated membrane systems. These organisms contain an outer membrane, plasma membrane, and an internal system of thylakoid membranes where the photosynthetic and respiratory machinery are found. This existence of compartmentalization and differentiation of membrane systems poses a number of challenges for cyanobacterial cells in terms of organization and distribution of proteins to the correct membrane system. Proteomics studies have long sought to identify the components of the different membrane systems in cyanobacteria, and to date about 450 different proteins have been attributed to either the plasma membrane or thylakoid membrane. Given the complexity of these membranes, many more proteins remain to be identified, and a comprehensive catalogue of plasma membrane and thylakoid membrane proteins is needed. Here we describe the identification of 635 differentially localized proteins in *Synechocystis* sp. PCC 6803 by quantitative iTRAQ isobaric labeling; of these, 459 proteins were localized to the plasma membrane and 176 were localized to the thylakoid membrane. Surprisingly, we found over 2.5 times the number of unique proteins identified in the plasma membrane compared with the thylakoid membrane. This suggests that the protein composition of the thylakoid membrane is more homogeneous than the plasma membrane, consistent with the role of the plasma membrane in diverse cellular processes including protein trafficking and nutrient import, compared with a more specialized role for the thylakoid membrane in cellular energetics. Thus, our data clearly define the two membrane systems with distinct functions. Overall, the protein compositions of the *Synechocystis* 6803 plasma membrane and thylakoid membrane are quite similar to that of the plasma membrane of *Escherichia coli* and thylakoid membrane of Arabidopsis chloroplasts, respectively. *Synechocystis* 6803 can therefore be described as a Gram-negative bacterium with an additional internal membrane system that fulfills the energetic requirements of the cell.   *Molecular & Cellular Proteomics 15: 10.1074/mcp.M115.057240, 2021–2032, 2016.*

Photosynthetic microbes such as the cyanobacterium *Synechocystis* sp. PCC 6803 convert light to cellular energy, an ability that makes these organisms of particular interest in renewable energy studies. Cyanobacteria typically have a Gram-negative-type cell envelope consisting of a plasma membrane (PM)[1], peptidoglycan layer, and outer membrane. These microbes also have an internal thylakoid membrane (TM) system where the protein complexes of the photosynthetic and respiratory electron transfer chains function. The presence of these differentiated membrane systems makes cyanobacteria more complex than other Gram-negative bacteria. There is considerable interest in understanding the roles of the membrane systems and their relation with each other. Our studies using electron tomography revealed that the TM in the cyanobacterium *Cyanothece* sp. ATCC 51142 forms a complicated network of membranes that enclose a single lumenal space (1). Several studies have probed the question of whether the PM and TM are contiguous, or if these two systems are physically independent (2–4). Recent reports have proposed the existence of sites of "hemifusion" between PM and TM, which can be analyzed as a subfraction of the PM and used to further clarify the targeting pathways between the membrane systems (5). Similarly, the existence of a membrane subfraction that associates with both PM and TM has been proposed (6, 7). Thus, identifying the protein composition of the different membrane systems is of considerable

[1] The abbreviations used are: PM, plasma membrane; TM, thylakoid membrane; PSI, photosystem I; PSII, photosystem II; TMH, transmembrane helix.

interest in understanding the form and function of cyanobacterial membranes.

Several previous studies have begun to catalogue the protein complement of the cyanobacterial membrane systems. One study of the PM proteome used two-dimensional gel electrophoresis coupled with mass spectrometry (MS) to identify 57 proteins, of which 17 are integral membrane proteins and 40 are peripheral proteins (8). Another study identified 51 integral PM proteins by peptide mass fingerprinting (9). Isolated TM samples were used to identify 76 proteins from 1- and 2-D gels by MALDI-TOF MS (10). A study of both isolated PM and TM samples probed by nano-LC separation and MS/MS identified 379 different proteins (5), of which 237 were uniquely localized to either PM or TM. However, all together to date only about 450 different proteins have been identified as localized to the PM or TM.

In order to comprehensively detect and identify proteins localized to the PM and TM, we applied a sensitive LC-MS/MS based analysis pipeline for the identification and quantification of this protein complement. This resulted in the identification of 635 proteins observed with significantly different localizations across PM and TM from purified membrane samples isolated from *Synechocystis* sp. PCC 6803 (hereafter, *Synechocystis* 6803). This is a large increase in the number of differentially abundant proteins compared with previous studies and offers considerable insight into the composition of PM and TM. Our study found a larger number of proteins uniquely localized in PM (459) compared with TM (176). The overall protein composition of PM was characterized by proteins involved in transport, secretion, and trafficking, whereas the TM protein composition described a specialized membrane system dedicated to the energetics of electron transport, highlighting the very different roles these membrane systems have in cyanobacterial cellular metabolism. Comparison of the *Synechocystis* 6803 membrane systems with the *E. coli* PM and Arabidopsis TM showed how an oxygenic phototrophic bacterium modified the Gram-negative PM for specific purposes while creating a specialized internal membrane compartment for photosynthetic electron transfer.

EXPERIMENTAL PROCEDURES

Cell Growth and Sample Preparation—*Synechocystis* 6803 cells were grown in BG11 at 30 °C under 30 μmole photons·m^{-2}·s^{-1} white light. Membrane isolation and two-phase partitioning were performed as described (11). Two-phase systems were prepared from stock solutions of 20% (w/w) Dextran T-500 and 40% (w/w) polyethylene glycol 3350.

Protein samples were incubated in 8 M urea, 100 mM ammonium bicarbonate, pH 8.0 solution containing 5 mM dithiothreitol at 56 °C for 45 min with constant shaking at 800 rpm in Thermomixer R (Eppendorf, Hauppauge, NY). Alkylation was performed with 20 mM iodoacetamide at 37 °C in the dark with constant shaking (800 rpm in Thermomixer), followed by an eightfold dilution with 25 mM ammonium bicarbonate, pH 8.0 containing 1 mM CaCl$_2$. Tryptic digestion with Sequencing Grade Modified Trypsin (Promega, Madison, WI) was performe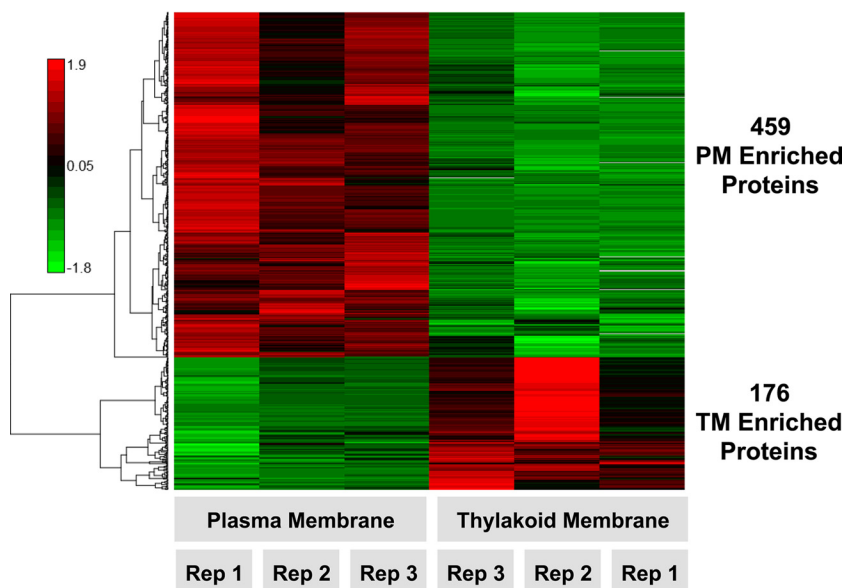d at 1:50 enzyme-to-substrate ratio for 4 h at 37 °C. The digested samples were then acidified with 10% trifluoroacetic acid to ~pH 3 and 5% acetonitrile was added to the digested samples prior to desalting. SPE C-18 columns (SUPELCO Discovery) were used for cleanup of the resultant peptide mixture, and samples were concentrated in a SpeedVac SC250 Express (Thermo Fisher Scientific, Waltham, MA) followed by BCA assay to determine final peptide concentration.

iTRAQ Labeling and HPLC Fractionation—Isobaric labeling of peptides using two separate four-plex iTRAQ™ reagents was performed according to the manufacturer's instructions (AB Sciex, Foster City, CA) and as previously described (12). Labeled peptide samples were separated at a flow rate of 0.5 ml/min on a reverse phase Waters XBridge C18 column (250 mm × 4.6 mm column containing 5 μm particles, and a 4.6 mm × 20 mm guard column) using an Agilent 1200 HPLC System equipped with a quaternary pump, degasser, diode array detector, Peltier-cooled auto-sampler and fraction collector (both set at 4 °C). Approximately 120 μg of labeled tryptic peptides was suspended in buffer A (10 mM triethylammonium bicarbonate, pH 7.5) and loaded onto the column. After the sample loading, the C18 column was washed for 35 min with solvent A, before applying the LC gradient. The LC gradient started with a linear increase of solvent A to 10% B (10 mM triethylammonium bicarbonate, pH 7.5, 90% acetonitrile) for 10 min, then linearly increased at 15 min to 20% B, 30 min to 30% B, 15 min to 35% B, 10 min to 45% B and another 10 min to 100% solvent B. Using an automated fraction collector, 96 fractions were collected for each sample, lyophilized and reconstituted into 12 fractions prior to LC-MS/MS analysis.

LC-MS/MS Analysis—All iTRAQ™-labeled fractions were analyzed by LC-MS/MS. Each sample was loaded onto a homemade 65 cm × 75 mm i.d. reversed-phase capillary column using 3 mm C18 particles (Phenomenex, Torrance, CA). The HPLC system consisted of a custom configuration of 100 ml Isco Model 100DM syringe pumps (Isco, Lincoln, NE), two-position Valco valves (Valco Instruments Co., Houston, TX), and a PAL autosampler (Leap Technologies, Carrboro, NC) that allowed fully automated sample analysis across four HPLC columns (13). The system was operated at a constant pressure of 10,000 psi over 3 h with an exponential gradient starting with 100% of mobile phase A (0.1% (v/v) formic acid in water) to 60% (v/v) of mobile phase B (0.1% (v/v) formic acid in acetonitrile). MS analysis was performed on a Thermo Scientific LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific, San Jose, CA) coupled with an electrospray ionization interface using homemade 150-mm o.d. × 20-mm i.d. chemically etched electrospray emitters (14). Full MS spectra were recorded at resolution of 100 K (*m/z* 400) over the range of *m/z* 400–2000 with an automated gain control (AGC) value of 1 × 10^6. MS/MS was performed in the data-dependent mode with an AGC target value of 3 × 10^4. The most abundant 10 parent ions were selected for MS/MS using high-energy collision dissociation with a normalized collision energy setting of 45. Precursor ion activation was performed with an isolation width of 2 Da, a minimal intensity of 500 counts, and an activation time of 10 ms.

Data Analysis—LC-MS/MS raw data were converted into dta files using Bioworks Cluster 3.2 (Thermo Fisher Scientific, Cambridge, MA), and MSGF+ algorithm (15) was used to search MS/MS spectra against *Synechocystis* PCC 6803 (NCBI 2011–02-28, 3672 entries). The key search parameters used were 20 ppm tolerance for precursor ion masses, +0.5 Da and −0.5 Da window on fragment ion mass tolerances (16), no limit on missed cleavages, partial tryptic search, no exclusion of contaminants, dynamic oxidation of methionine (15.9949 Da), static IAA alkylation on cysteine (57.0215 Da), and static iTRAQ modification of lysine and N termini (+144.1021 Da). No additional mass shifts were performed on the data. The decoy data base searching methodology (17, 18) was used to control the false discovery rate at the unique peptide level to ~0.1% (15). Only proteins

Fig. 1. **Heat map of differentially abundant membrane proteins.** LC-MS/MS quantification of >1400 proteins resulted in the identification of 459 PM and 176 TM proteins observed with significantly different membrane localization after direct comparison, *p* value <0.05, between plasma and thylakoid membrane sample preparations. Proteins were hierarchically clustered using a Pearson correlation distance metric. Three biological replicates of PM and TM samples were analyzed (as shown), which included the additional averaged technical replicate values for TM rep 2 and TM rep 3. The red-green color scale depicts normalized $\log_2$ ratio values of PM to TM abundance of any individual membrane protein (Red: high, Green: low).

459
PM Enriched
Proteins

176
TM Enriched
Proteins

Plasma Membrane   Thylakoid Membrane

Rep 1 | Rep 2 | Rep 3 | Rep 3 | Rep 2 | Rep 1

containing multiple peptide identifications with reporter ion intensities were quantified. The ppm distribution of identified spectra is shown is supplemental Fig. S1 and protein and peptide data can be found in supplemental Table S3. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (19) via the PRIDE partner repository with the data set identifier PXD003079 and 10.6019/PXD003079.

Experimental Design and Statistical Rationale—The experimental design incorporated three biological replicates of PM and TM samples, across two independent four-plex iTRAQ experiments to cover all six samples, where final quantitative comparisons included the averaged technical replicate values for TM-2 and TM-3. For quantification purposes, peptide reporter ion intensities were captured across all channels and compared by calculating the summed peptide intensity values for TM and PM samples. Summed protein values were then scaled within each experiment and then central tendency normalized and statistically compared across biological replicates using ANOVA with membrane type as a fixed effect and using the program DAnTE (20) for final comparisons.

Computational Web-based Tools—We used SignalP 4.1 server (21) (http://www.cbs.dtu.dk/services/SignalP/) for identifying putative signal peptides as well as their cleavage sites. In addition, we utilized THHMM server 2.0 (22) (http://www.cbs.dtu.dk/services/TMHMM/) and LipoP 1.0 server (23) (http://www.cbs.dtu.dk/services/LipoP/) to identify TMHs and lipoproteins, respectively, in identified proteins.

RESULTS

Identification of PM and TM Proteins—In order to analyze isolated plasma membrane (PM) and thylakoid membrane (TM) samples, we previously devised a 2D protocol for membrane isolation with a short preparation time (11). This procedure used multiple rounds of a polymer two-phase isolation step, which is a fast procedure that maintained the photochemical activity of the isolated membranes. As markers for the purity of the isolated membranes, antibodies raised against the NrtA and CP47 proteins were used that were found exclusively in PM and TM, respectively, and immunoblotting results indicated that membrane fractions isolated using this procedure are highly purified (11).

PM and TM samples prepared using this procedure were analyzed by a sensitive LC-MS/MS-based pipeline for the identification and quantification of proteins. Quantitative iTRAQ 4-plex isobaric labeling coupled with pre-MS high pH reversed phase peptide separations were used to quantitatively capture a triplicate comparison of PM and TM samples (see Experimental Procedures for details). This resulted in the identification of 1496 proteins with appropriate quantitative values for statistical comparison (supplemental Table S1). Of these 1496 proteins, 635 proteins were observed with significantly different localizations across PM and TM (*p* value <0.05) (Fig. 1, supplemental Table S1, supplemental Fig. 2), and 861 proteins were found to be present in both PM and TM (supplemental Table S1).

Of the 861 proteins found in both PM and TM, 257 of these are predicted to be hypothetical and 88 are unknown proteins (supplemental Table S1). Of the remaining 516 proteins, many of these are known soluble proteins of high cellular abundance, such as phycobilisome light-harvesting antenna subunits, and are present in both isolated PM and TM membrane samples. Interestingly, a number of photosystem I (PSI) and photosystem II (PSII) proteins (*e.g.* the PSII D1 protein and PSI proteins PsaB and PsaK) are present in both PM and TM. Another major group (*i.e.* 91) of such proteins belong to translational processes such as aminoacyl tRNA synthetase and tRNA modification, protein modification and degradation, and ribosomal protein synthesis and modification. The remaining proteins are distributed across several pathways/subsystems including biosynthesis of amino acids and cofactors, energy metabolism, regulatory and transport processes.

This identification of 635 proteins with significantly different localizations across PM and TM provides insight into the unique properties and characteristics of these two membrane systems. In addition, this study represents a more compre-

FIG. 2. **Functional categorization distribution of proteins between PM and TM analyzed preparations.** Shown is the most abundant and relevant functional categorization populated by proteins enriched in either PM or TM preparations, demonstrating clear segregation of cellular processes. Categorization is based upon CyanoBase annotation.
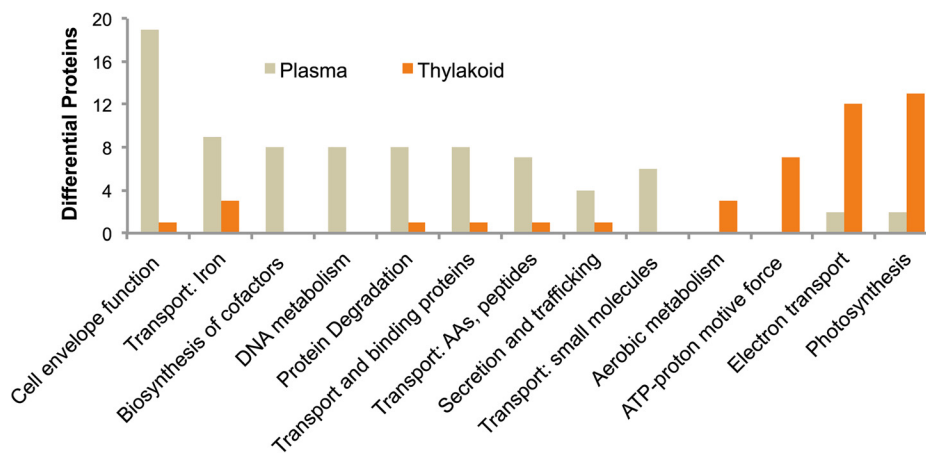
hensive investigation, in terms of both depth of coverage and quantitation, compared with the most recent study of isolated PM and TM samples (5). This previous study, which also employed LC-MS/MS analysis but utilized spectral count approaches, identified a total of 379 proteins, 237 of which were designated as differentially localized to PM or TM (5). Note that this comparison considers only the number of proteins identified in either PM or TM, but not in both, in these studies. As iTRAQ labeling and quantitative comparison were used in the current study, we focused on the differential abundance of proteins to determine membrane localization, recognizing that low abundance signals can be generated by iTRAQ in both TM and PM preparations, and it is likely that the sensitivity of the analyses coupled with known co-fragmentation and isolation issues with iTRAQ labeling (24) contribute to this overlap.

Of these 635 differentially localized proteins identified in our study, 459 and 176 proteins were localized to PM and TM, respectively (supplemental Table S1 and supplemental Fig. S2). This was a surprising result, with over 2.5 times the number of proteins identified in PM compared with TM, suggesting that the protein composition of TM is more limited and specialized than PM. Proteins involved in protein trafficking, nutrient transport, cofactor biosynthesis, cell envelope function, secretion, and small molecule transport were found predominately in the PM (Fig. 2). In comparison, TM was characterized by proteins involved in metabolism, ATP generation, electron transport, and photosynthesis (Fig. 2).

Another possible rationale for this finding could be that more abundant proteins in TM, such as PSI and PSII components, bias against peptide/protein identification compared with PM. However, considering the sensitivity of the analyses and the large number of overall proteins identified, it is unlikely that there are a significant number of proteins specific to TM that have been missed. An additional consideration is the overall abundance of PM *versus* TM in the cell. We chose to compare similar peptide/protein amounts between PM and TM for optimizing coverage and comparison. This could introduce a bias for greater PM protein coverage; however,

when viewing the volcano plot (supplemental Fig. S2), the right side (PM specific) is more consistently populated at higher *p* values and fold changes compared with the left side (TM specific), which interestingly appears more variable in its protein distribution, regardless of which cutoff value is used. Furthermore, annotation of the findings is consistent with the diversity of roles of PM in cellular processes. Though there may be limited biases in the protein distributions, overall our results describe independent membrane systems that have unique, highly specific cellular roles.

Characteristics of Identified Proteins—Important characteristics of membrane proteins describe their types (*e.g.* integral membrane proteins or soluble proteins) and provide other relevant information such as the number of transmembrane helices (TMH) and the protein orientation across the membrane (for transmembrane proteins) or presence/absence of a signal peptide (for soluble proteins). From the 176 TM proteins identified in the current study, the number of TMH was predicted using TMHMM, and lipoprotein N-terminal signals were predicted using SignalP/LipoP. Based on this analysis, Fig. 3*A* shows the predicted topologies of identified TM proteins. Eighty-seven (of 176) TM proteins (49%) were identified as integral membrane proteins. About 45% of these integral membrane proteins have only one TMH, whereas about 20% have more than five. Seven (of 87) integral membrane proteins have N-terminal signal peptides as identified by SignalP 4.1; these include three PS II proteins (Sll1418 (PsbP), Sll1194 (PsbU), and Sll0427 (PsbO)), the PS I reaction center subunit III precursor/plastocyanin docking protein PsaF (Sll0819), the cytochrome c550 protein (Sll0258), a H^+/Ca^{+2} exchanger protein (Slr1336), and a hypothetical protein (Slr1273). Five of the identified nonintegral TM proteins contain an N terminus with a consensus pattern for lipoproteins. There are five proteins with a Sec signal among the rest of the proteins (*i.e.* soluble proteins). Finally, 79 of the 84 soluble proteins do not have any N-terminal signal peptide and hence are considered as peripheral proteins on the cytoplasmic side of the thylakoid membrane.
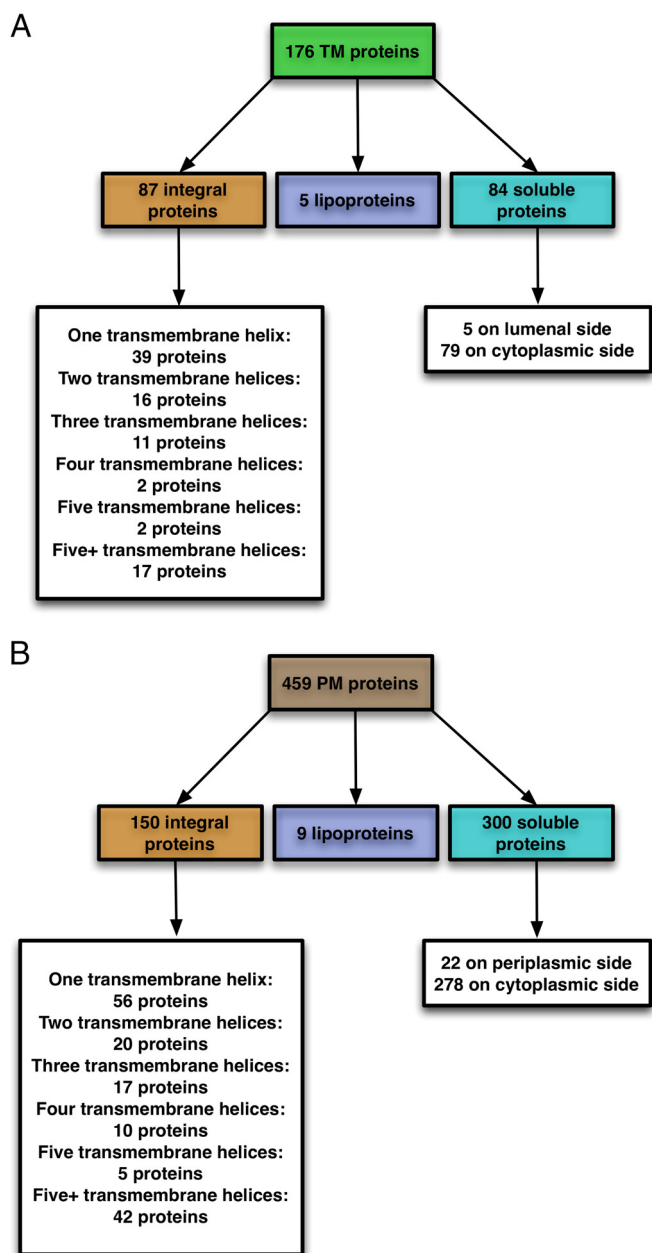
A

B

FIG. 3. **Predicted topology of identified (*A*) TM and (*B*) PM proteins in *Synechocystis* 6803.**

Based on similar analysis as that for TM proteins, Fig. 3*B* shows the predicted topologies of identified PM proteins. Of 459 PM proteins, 150 (33%) were identified as integral membrane proteins. About 37% of these 150 integral membrane proteins have only one TMH whereas about 28% have more than five. Eight (of 150) integral membrane proteins have N-terminal signal peptides, including three proteins (Slr1744, Slr0089, and Slr1897) from cell envelope, fatty acid, and transport/binding processes, respectively, and five other hypothetical/unknown proteins. Nine of the identified nonintegral PM proteins contain an N terminus with a consensus pattern for lipoproteins, whereas 22 others were detected as

soluble proteins with a Sec signal on the periplasmic side. Because the remaining 278 of the soluble proteins do not have any N-terminal signal peptide, they are considered as peripheral proteins on the cytoplasmic side of the plasma membrane.

Synechocystis 6803 is predicted to have more than 800 membrane proteins (25, 26), or ~22% of the predicted 3672 open reading frames. In this study, we have identified a total of 237 integral membrane proteins differentially localized to TM or PM (87 in TM and 150 in PM). The remaining set of integral membrane proteins (~500) are proteins that are potentially found in OM, in both PM and TM, or were not identified in our analysis. Our strategy was successful in identifying a substantial number of integral membrane proteins with more than five TMHs (17 in TM and 42 in PM) (Fig. 3), demonstrating the ability of this approach to identify highly hydrophobic proteins.

Photosystems and Respiratory Proteins—The current work identified a majority of the subunits of important photosynthetic complexes localized to the TM, including PSI, PSII, cytochrome b_6f, and ATP synthase. Similarly, major respiratory complexes (*e.g.* NADH dehydrogenase and cytochrome b_6f) are also mostly located in the TM. However, of the two subunits of cytochrome oxidase, one is located in the TM (Slr1136) and the other in the PM (Sll0813).

Table I shows the most abundant integral TM proteins and includes many important proteins from photosynthetic and respiratory metabolism. Included in this set are four PSI proteins (Slr1834 (PsaA), Sll0819 (PsaF), Slr1655 (PsaL), and the PSI assembly related protein Sll0226), all which have one or more TM TMH. Note that PsaF is the only PSI protein having N-terminal signal peptides (see above). There are five other PSI-related proteins including Sll0563 (PsaC), Slr0737 (PsaD), and Ssr2831 (PsaE) identified as peripheral TM proteins on the cytoplasmic side (Fig. 3).

Out of eight PSII proteins (Psb subunits with one or more TMH), four proteins, Sll0427 (PsbO), Sll1418 (PsbP), Sll1194 (PsbU), and Sll0258 (PsbV), have a single TMH each with N-terminal signal peptides. Of these, PsbO, PsbU, and PsbV are included with the most abundant proteins in Table I. Three other PSII-related proteins were detected as soluble proteins: Slr2034 (Ycf48) on the lumenal side and two Psb28 subunits (Slr1739 and Sll1398) on the cytoplasmic side. This localization of Psb28 is consistent with previous reports (27). Consistent with a recent report (28), the D1 protein processing protease, CtpA (Slr0008) was found as an integral TM protein (supplemental Table S1).

Of seven subunits of ATP synthase located in TM, Sll1323 (AtpG(β)) and Sll1322 (AtpI(α)) (Table I) have one and five TM TMH, respectively, with the remaining being soluble proteins on the cytoplasmic side. There are 9 NADH dehydrogenase subunits identified in TM, five (NdhA, NdhB, NdhE, NdhF1 and NdhF4) as integral and four others as soluble proteins. However, one of the two Type 2 dehydrogenase subunits (Sll1484)

TABLE I
Distribution of TM integral proteins with Log$_2$ ratio of -2 or less across different pathways/subsystems

Proteins	Function	Subsystem/Pathway	Log$_2$ ratio (PM/TM)	No of TMH[a]
Sll0026	NADH dehydrogenase subunit 5 (involved in constitutive, low affinity CO$_2$ uptake)	Photosynthesis and respiration	-3.65	16
Sll0223	NADH dehydrogenase subunit 2	Photosynthesis and respiration	-4.25	14
Sll0226	Photosystem I assembly related protein	Photosynthesis and respiration	-2.01	2
Sll0258	Photosystem II PsbV protein	Photosynthesis and respiration	-3.50	1
Sll0262	Acyl-lipid desaturase (delta 6)	Fatty acid, phospholipid and sterol metabolism	-2.16	5
Sll0408	Peptidyl-prolyl cis-trans isomerase	Translation	-2.93	1
Sll0427	Photosystem II PsbO protein	Photosynthesis and respiration	-4.85	1
Sll0450	Cytochrome b subunit of nitric oxide reductase	Amino acid biosynthesis	-3.38	13
Sll0522	NADH dehydrogenase subunit 4L	Photosynthesis and respiration	-2.85	3
Sll0556	Na$^+$/H+ antiporter	Transport and binding proteins	-2.36	12
Sll0819	Photosystem I PsaF protein	Photosynthesis and respiration	-2.40	2
Sll0851	Photosystem II CP43 protein	Photosynthesis and respiration	-2.59	7
Sll1147	Glutathione S-transferase	Biosynthesis of cofactors, prosthetic groups, and carriers	-3.05	3
Sll1194	Photosystem II PsbU protein	Photosynthesis and respiration	-3.50	1
Sll1316	cytochrome b_6-f complex iron-sulfur subunit (Rieske iron sulfur protein)	Photosynthesis and respiration	-4.50	1
Sll1322	ATP synthase subunit a	Photosynthesis and respiration	-2.02	5
Sll1323	ATP synthase subunit b'	Photosynthesis and respiration	-2.84	1
Sll1471	Phycobilisome rod-core linker polypeptide	Photosynthesis and respiration	-2.78	1
Sll1513	C-type cytochrome synthesis protein	Other categories	-3.81	8
Sll1694	Pilin polypeptide PilA1	Cellular processes	-2.28	1
Sll1702	Hypothetical protein YCF51	Hypothetical	-3.34	2
Sll1784	Periplasmic protein, function unknown	Unknown	-2.78	1
Slr0228	Cell division protein FtsH	Cellular processes	-2.72	2
Slr1336	H+/Ca^{2+} exchanger	Transport and binding proteins	-3.96	11
Slr1434	Pyridine nucleotide transhydrogenase beta subunit	Biosynthesis of cofactors, prosthetic groups, and carriers	-2.44	9
Slr1645	Photosystem II Psb27 protein	Photosynthesis and respiration	-2.15	1
Slr1655	Photosystem I subunit XI	Photosynthesis and respiration	-2.61	2
Slr1834	Photosystem I PsaA protein	Photosynthesis and respiration	-2.65	9
Ssr2087	C-type cytochrome biogenesis protein Ccs1	Other categories	-2.27	3
Ssr3451	Photosystem II PsbE protein	Photosynthesis and respiration	-4.15	1
Sll0862	Hypothetical protein	Hypothetical	-2.84	11
Sll5034	Hypothetical protein	Hypothetical	-2.96	1
Slr0637	Hypothetical protein	Hypothetical	-3.68	1
Slr0813	Hypothetical protein	Hypothetical	-2.54	3
Slr0962	Unknown protein	Unknown	-4.26	3
Slr1261	Hypothetical protein	Hypothetical	-2.08	2
Slr1273	Hypothetical protein	Hypothetical	-2.22	1
Slr1470	Hypothetical protein	Hypothetical	-2.70	1
Slr1471	Hypothetical protein	Hypothetical	-3.39	3
Slr1624	Hypothetical protein	Hypothetical	-2.71	1
Ssl0410	Unknown protein	Unknown	-2.84	1
Ssl1328	Hypothetical protein	Hypothetical	-3.92	1
Ssl5113	Unknown protein	Unknown	-2.47	1
Ssr2406	Unknown protein	Unknown	-5.86	2

[a] Number of TMH was predicted using TMHMM-v2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0).

is found as an integral PM protein, whereas the other (Slr1743) is a soluble TM protein on the cytoplasmic side. Among the remaining proteins, two cytochrome b_6f complex proteins (PetA and PetC) were identified in the TM. Overall, almost the entire electron transport machinery functions in the TM and these comprise a large portion of the proteins listed in Table I.

Pigment Biosynthesis and Transport Proteins—Our analysis revealed the location of 11 proteins from pigment biosynthesis (three from carotenoid and eight from cobalamin, heme, phycobilin and porphyrin metabolism). All but ferrochelatase (HemH/ScpA) were identified as PM soluble proteins on the cytoplasmic side.

A total of 62 transport (and binding) proteins were identified in the PM. Among these proteins, 34 proteins were PM integral proteins, including components of ABC transporters, permease proteins, biopolymer transporter system, P-type ATPase, and metal ion/cation transport system proteins. Of the 35 most abundant integral PM proteins identified (Table II), half were classified as hypothetical/unknown, and the majority of the categorized proteins were transport/binding proteins.

The periplasmic sugar-binding protein of an ABC transporter, Slr1897, is the only protein with a single transmembrane span and N-terminal signal peptides. Seven PM proteins including Slr0040 (CmpA), Sll1450 (NrtA), and Slr0447 (UrtA) were detected as lipoproteins, whereas two others (NatB and Slr2043) as soluble periplasmic proteins. The remaining 19 proteins (mostly ATP binding proteins of ABC/urea transporter) are soluble proteins on the PM cytoplasmic side. Compared with the large number of transport proteins located in PM, only seven such proteins are located in the TM. Five of these that involve metal (*i.e.* Ca^{+2} and Na^+) and urea transport systems were found to be integral TM proteins, whereas (similar to PM) the remaining two were ATP binding proteins located on the cytoplasmic side of the membrane.

Proteins Involved in Other Important Cellular Processes—Of the 50 proteins from cell envelope and cellular processes that were identified in the current study, 10 proteins, including five putative porins and PilQ, are soluble PM proteins on the periplasmic side, whereas solute-binding protein Slr1962 and putative endoglucanase Slr0897 were detected as a PM lipoprotein. Note the remaining proteins including pilus biogenesis protein Slr0063 were detected as soluble PM proteins on the cytoplasmic side. Fourteen other proteins including MurC, PilC, PilA2, Ctr1, and TaxD1 are integral PM proteins. Among these proteins, N-acetylmuramoyl-L-alanine amidase (Slr1744) has a single transmembrane span with N-terminal signal peptides. In addition, five proteins (*e.g.* PilA1) were identified as integral TM proteins. Only one protein from the amino acid metabolism, cytochrome *b* subunit of nitric oxide reductase (NorB), and two proteins (DesD and Sll0418) from fatty acid metabolism were identified as integral TM proteins, whereas just one protein, gamma-tocopherol methyltrans-

ferase (Slr0089) from fatty acid metabolism, was identified as a PM protein.

Compared with proteins from metabolic/transport processes, only a handful of proteins involved in translation and regulation were identified in PM/TM by our current work. These include a protease (Slr0535) as an integral PM protein, and the tRNA synthetase AsnS, and the cis-trans isomerase Sll0408 as integral TM proteins. From regulatory mechanism, IcfG (Carbon metabolism regulatory protein), Slr1225 and Slr0599 (Serine/Threonine kinase), and two component system (Hik6/10/12/21/31) proteins were detected as integral proteins in the PM.

Hypothetical and Unknown Proteins—Hypothetical proteins are predicted from nucleic acid sequences without experimental evidence. Oftentimes, these proteins are associated with low identity to known/annotated proteins (29). Unknown proteins are also predicted by bioinformatics tools, but unlike hypothetical proteins, these are experimentally proven to exist without any biochemical characterization (29). Based on the annotations provided on the Cyanobase database (30), the *Synechocystis* genome contains about 33% hypothetical and 18% unknown ORFs. Our current study reveals about 25 and 13% of PM proteins as hypothetical and unknown, respectively. The corresponding percentages are 32% (hypothetical) and 10% (unknown) for TM proteins.

Of 57 hypothetical TM proteins, 27 are integral, 11 have one predicted TMH, 7 have two, and the remaining 9 have three or more helices. One of these proteins (Slr1273) has an N-terminal signal peptide. In addition, two of the identified nonintegral hypothetical TM proteins contain an N terminus with a consensus pattern for lipoproteins, whereas 3 others are soluble proteins on the lumen side. Finally, the remaining 25 proteins are peripheral proteins on the cytoplasmic side of TM. Among 17 unknown TM proteins, 10 were identified as integral, seven have one predicted TMH, and the remaining three have four or more helices. Sll0022 was identified as a nonintegral lipoprotein, whereas the remaining six are soluble proteins on the cytoplasmic side.

Of 117 hypothetical PM proteins, 51 are integral, 19 have one predicted TMH, 18 others (three sets of six proteins) have two, three, and four helices, respectively, and 14 have five or more helices. Two such proteins (Slr0200 and Slr0765) were found to have N-terminal signal peptides. Four other proteins were identified as soluble proteins on the periplasmic side, whereas the remaining 62 are peripheral proteins on the cytoplasmic side of PM. Among these 62 unknown PM proteins, 25 were identified as integral, 12 have one predicted TMH, 3 have two, 3 others have three, and 7 have four or more helices. Of these 25, Slr01257 and Ssr0693 have N-terminal signal peptides. The remaining proteins include 2 nonintegral lipoproteins, 3 soluble PM proteins on the periplasmic side, and 32 peripheral proteins on the cytoplasmic side.

TABLE II

Distribution of PM integral proteins with Log$_2$ ratio of 2 or more across different pathways/subsystems

Proteins	Function	Subsystem/Pathway	Log$_2$ ratio (PM/TM)	No of TMH[a]
Sll0108	Ammonium/methylammonium permease	Transport and binding proteins	3.01	12
Sll0169	Cell division protein Ftn2 homolog	Cellular processes	2.04	1
Sll0574	Probable permease protein of lipopolysaccharide ABC transporter	Transport and binding proteins	3.08	6
Sll0855	Putative channel transporter	Transport and binding proteins	2.01	10
Sll0993	Potassium channel	Transport and binding proteins	2.67	3
Sll1276	ATP-binding protein of ABC transporter	Transport and binding proteins	2.09	3
Sll1695	Pilin polypeptide PilA2	Cellular processes	3.19	1
Slr0114	Putative PP2C-type protein phosphatase	Unknown	2.71	2
Slr0369	RND multidrug efflux transporter	Transport and binding proteins	2.05	11
Slr0593	cAMP binding membrane protein	Unknown	2.11	6
Slr0615	ATP-binding protein of ABC transporter	Transport and binding proteins	2.53	6
Slr0678	Biopolymer transport ExbD like protein	Transport and binding proteins	2.56	1
Slr0798	Zinc-transporting P-type ATPase (zinc efflux pump) involved in zinc tolerance	Transport and binding proteins	2.27	5
Slr1149	ATP-binding protein of ABC transporter	Transport and binding proteins	2.66	5
Slr1216	Mg^{2+} transport protein	Transport and binding proteins	2.12	3
Slr1423	UDP-N-acetylmuramate–alanine ligase	Cell envelope	2.91	1
Slr1515	Putative membrane protein required for bicarbonate uptake	Hypothetical	2.54	12
Slr1575	Probable potassium efflux system	Transport and binding proteins	2.28	2
Sll0267	Unknown protein	Unknown	2.14	3
Sll0283	Hypothetical protein	Hypothetical	2.31	4
Sll0384	Unknown protein	Unknown	2.01	4
Sll0505	Hypothetical protein	Hypothetical	2.15	2
Sll0602	Hypothetical protein	Hypothetical	2.25	1
Slr0727	Hypothetical protein	Hypothetical	2.21	11
Sll1608	Hypothetical protein	Hypothetical	2.09	7
Sll2003	Hypothetical protein	Hypothetical	2.02	11
Slr0060	Unknown protein	Unknown	2.09	1
Slr0594	Hypothetical protein	Hypothetical	2.65	4
Slr0625	Hypothetical protein	Hypothetical	2.59	11
Slr0818	Hypothetical protein	Hypothetical	2.02	1
Slr0960	Unknown protein	Unknown	2.31	3
Slr1257	Unknown protein	Unknown	2.05	2
Slr1875	Hypothetical protein	Hypothetical	2.05	4
Slr1927	Hypothetical protein	Hypothetical	2.16	2
Slr2011	Hypothetical protein	Hypothetical	2.24	3

[a] Number of TMH was predicted using TMHMM-v2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0).
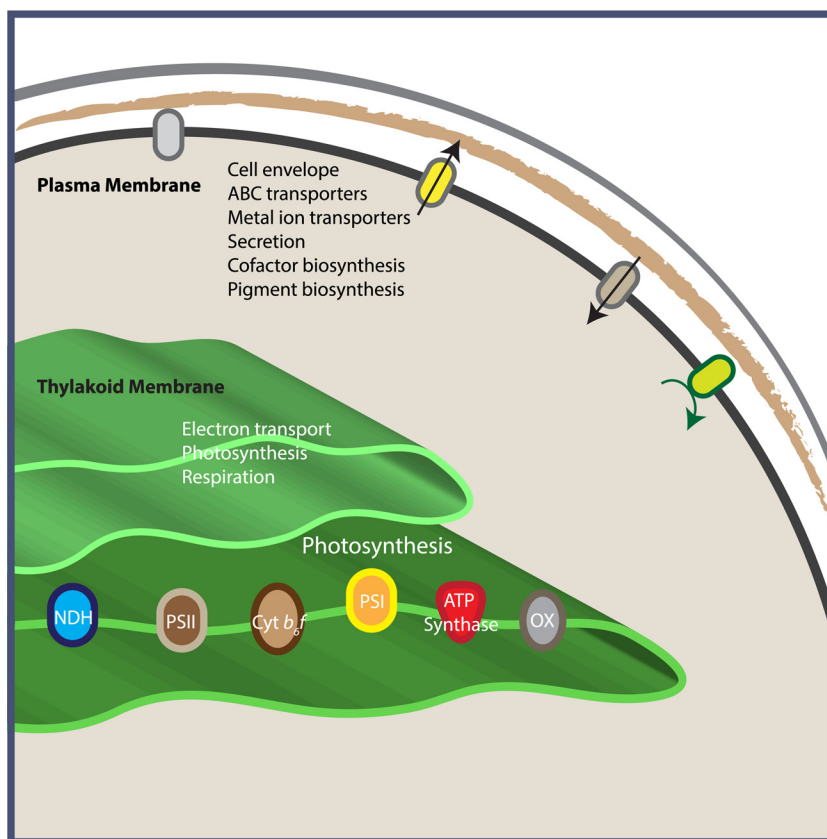
Fɪɢ. 4. **Schematic drawing showing distribution of functional roles between PM and TM.**

DISCUSSION

Of the 635 differentially abundant proteins identified in our study, 459 proteins were localized to PM and 176 were localized to TM. Fig. 1 shows the clear quantitative differentiation between these two localized protein groups; however, interestingly, we also observed more quantitative variation within the biological replicates of TM or PM protein groups than anticipated (more so for the TM preparation, as also seen in supplemental Fig. S1), likely hinting at a more dynamic localization of proteins between biological replicates. This is difficult to determine, however, as the scale and ratio type data of iTRAQ quantification is somewhat limited for this type of comparison. Regardless, the fact that over 2.5 times the number of proteins appear localized to PM compared with TM suggests that the protein composition of TM is more limited and specialized compared with PM. Functional annotation of these differential proteins also confirms, that as a whole, our data describe two membrane systems that have very different roles: PM is involved in transport, secretion, trafficking, and general "gatekeeping" functions, whereas TM is devoted to the energetics of electron transport and cellular metabolism (Fig. 2). Based on our analysis, it is evident that a higher percentage of PM proteins are soluble proteins (*i.e.* 65% PM soluble proteins *versus* 48% TM soluble proteins), whereas the opposite is the case with integral membrane proteins (*i.e.* 32% PM integral proteins *versus* 49% TM integral pro-

teins). Therefore, these topological/functional variations are well correlated with their stated roles. An overview of the distribution of cellular processes between PM and TM is depicted in Fig. 4.

The association between PM and TM in cyanobacteria has long been a topic of interest, and even though we technically identified the majority of proteins across both membranes, we assert that these data provide further evidence that PM and TM are separate membrane systems, because given the differences in protein composition described here, it is unlikely that PM and TM are contiguous. Other studies have investigated a membrane subfraction existing between PM and TM, but our purification procedure does not result in the isolation of this subfraction. In fact, we did not identify the PratA protein (Slr2048), a marker for this subfraction, in our analysis. The PSII biogenesis and repair cycle has recently been explored by a combination of aqueous two-phase partitioning, epitope tagging, and radioactive pulse-chase, and it was determined that processing of the D1 protein occurs in the TM (28), a finding that is consistent with our localization of the CtpA processing protease in the TM (supplemental Table S1). However, this study found the vesicle-inducing protein 1 (Vipp1, Sll0617) predominantly in the PM (28), with a weaker signal found in the TM, whereas our analysis identified Vipp1 as a TM protein (supplemental Table S1). Interestingly, we identified several photosystem-related proteins in the set of
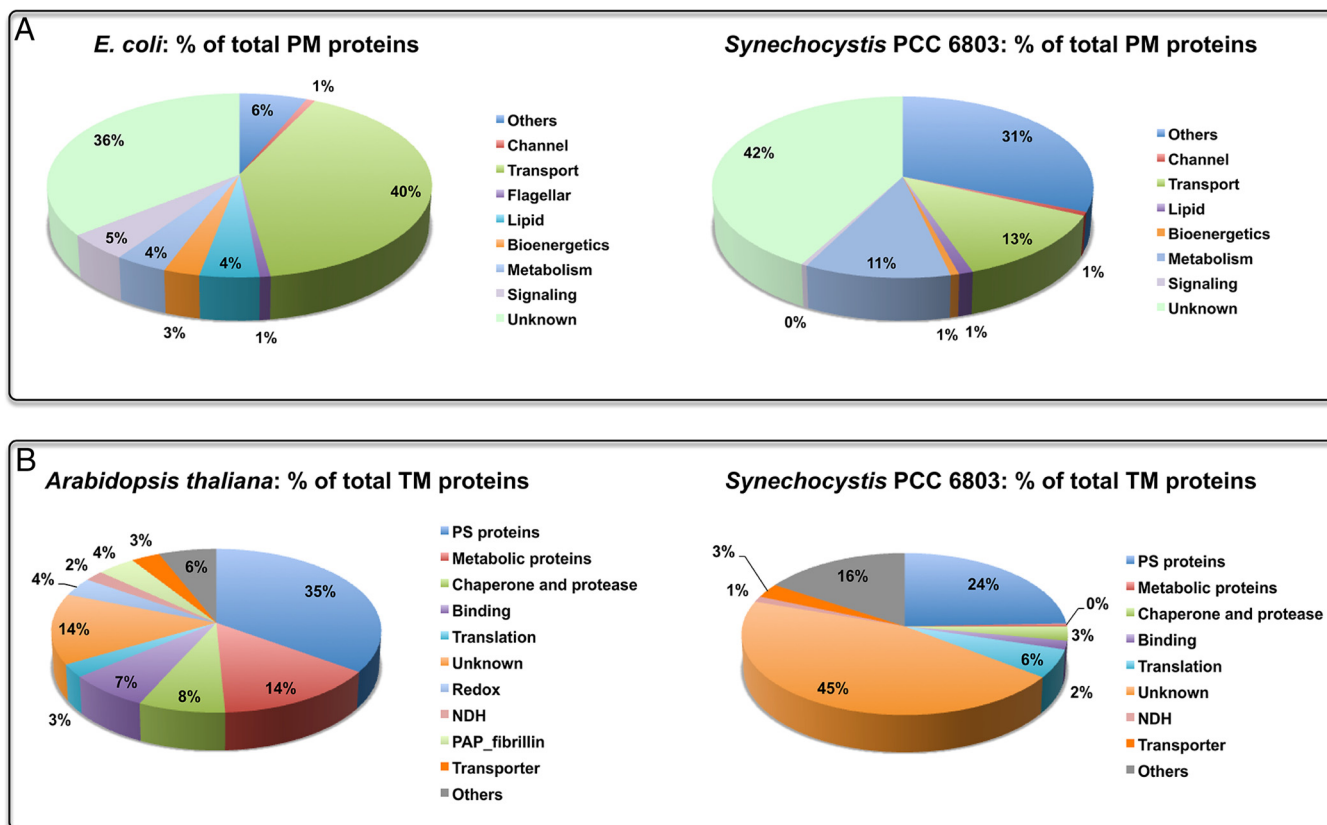
FIG. 5. **Comparative analysis of membrane proteins: (A) E. coli PM versus Synechocystis PM, and (B) Arabidopsis TM versus Synechocystis TM.**

proteins shared between PM and TM, including PsbA3 (Sll1867), PsbA1 (Slr1181), PsaB (Slr1835), and PsbB (Slr0906).

One of the goals of this study was to generate a comprehensive list of uniquely localized proteins for both PM and TM (see Supplemental Table S1). We therefore examined the scope of the current work in comparison to the most comprehensive previous study, that of Pisareva *et al.* (5). Of the 176 proteins we identified localized to TM, 149 of these were uniquely identified in our analysis, whereas 27 proteins were also identified by Pisareva *et al.* (5). However, an additional 73 proteins were uniquely localized to the TM by Pisareva *et al.* (5), but not in the current study. Similarly, of the 459 proteins we identified as localized to PM, 71 of these were also identified by Pisareva *et al.* in the PM, with 86 additional proteins uniquely localized to the PM by Pisareva *et al.* Supplemental Table S2 summarizes the proteins uniquely identified by Pisareva *et al.* in both PM and TM. It is interesting to note that the vast majority (84%) of these 73 proteins uniquely localized to the TM by Pisareva *et al.* were identified/quantified with one spectra (61 out of 73), from which it is difficult to accurately determine or quantify appropriate localization. This most likely explains the limited overlap between the two studies, and in this regard, our current, more in-depth study, which included pre-MS fractionation along with quantitative iTRAQ labeling

performed in biological triplicate, provides the necessary quantitative information to inform upon the previously limited protein identifications. This more comprehensive survey of TM components included key photosynthetic and electron transport components, *i.e.* seven different NADH dehydrogenase subunits (Subunits 2 and 5 as integral and the rest as soluble proteins), seven photosystem I proteins (three integral and four soluble), nine photosystem II proteins (five integral and four soluble), and five ATP synthase subunits (two integral and three soluble).

A similar comparison in terms of PM proteins across these two data sets revealed overlap of 71 proteins, whereas 388 proteins were identified as PM specific in the current study and 86 identified as PM specific by Pisareva *et al.* (5) (Fig. 5B). Among these unique proteins, 160 (out of 388) and 49 (out of 86) are either hypothetical or have unknown functions. Of the remaining 228 PM proteins detected in the current analysis, 53 are integral membrane proteins (such as two-component hybrid system proteins, ATP binding proteins on ABC transporters, and Na^+/H^+ antiporter), seven are substrate-binding lipoproteins, 11 are soluble proteins on the periplasmic end (*e.g.* putative porins), and the remaining ones are peripheral proteins on the cytoplasmic side.

Cross-comparison between the current data set and that in Pisareva *et al.* (5) (see supplemental Table S2) also revealed

multiple instances where proteins previously deemed localized to TM or PM were in fact found quantitatively enriched in the other membrane. Seven TM proteins from the current study were identified in the PM by Pisareva *et al.*, and 16 PM proteins from the current study were found in the TM by Pisareva *et al.* Excluding hypothetical/unknown proteins, three TM proteins identified in the current study included two (Sll0897 and Sll1260) peripheral proteins on the cytoplasmic end, plus lipoprotein Sll0915. Out of 10 PM proteins, four are integral membrane proteins (*e.g.* two-component regulatory proteins and binding proteins), two are lipoproteins (GgtB and Slr0804) with the remaining being soluble proteins. The remainder of unique proteins found by Pisareva *et al.* (5) (57 TM and 79 PM proteins) were either not detected or excluded through our analysis pipeline requirement of multiple peptide identifications per protein, (27 TM and 38 PM proteins), or quantitatively rejected because of *p* value criterion (30 TM and 96 PM proteins). Overall, these proteins are distributed across different functional categories; however, a significant portion of these has photosynthesis/respiratory function or are hypothetical/unknown proteins. A detailed comparison of TM/PM proteins between these two data sets is included in supplemental Table S2.

Synechocystis 6803 is a Gram-negative photosynthetic bacterium, and therefore, the general PM features are predicted to resemble those of a Gram-negative bacterium such as *E. coli*. In addition, because a cyanobacterial ancestor was the progenitor of plant chloroplasts, the *Synechocystis* 6803 TM membrane properties might also be correlated to that of a plant species such as the model plant *Arabidopsis thaliana*. Based on this premise, we carried out a comparative study by using literature data available on topologies of *E. coli* PM (31) and Arabidopsis TM (32) (Fig. 5). As shown in Fig. 5*A*, both *E. coli* and *Synechocystis* 6803 have an almost similar amount of PM proteins (17% *versus* 13% of total proteins in *E. coli* (33) and *Synechocystis* 6803 (34), respectively), but two categories (namely, transport proteins and proteins involved in cellular processes such as biogenesis, cell envelope development, DNA replication/repair, and regulation) differ considerably. In contrast to 13% of PM proteins involved in transport in *Synechocystis* 6803, *E. coli* has 40% proteins involved in various transport activities (*e.g.* export/import, active/passive transport). Interestingly, a significant portion of *Synechocystis* 6803 transport proteins are generic ABC transporters, whereas the majority of *E. coli* transporter proteins are substrate specific (*e.g.* amino acid transporters). This difference might be primarily because of their physiological differences: whereas *Synechocystis* 6803, as a photosynthetic bacterium, is able to synthesize all essential amino acids, *E. coli* is equipped to import these when available, without expending energy on their production (35). Compared with a mere 6% of *E. coli* PM proteins, a staggering 31% *Synechocystis* 6803 PM proteins are involved in various cellular processes such as cell envelope development, DNA modification, restriction,

modification, recombination, and repair, regulatory functions, transcription, and translation. Another interesting difference is the localization of electron transfer chain: whereas in *E. coli* the entire electron transfer chain is in the PM, in *Synechocystis* 6803 it mostly functions in the TM. It is interesting to note that the evolutionary line of bacteria involves the transition from anoxygenic to oxygenic photosynthesis, and in purple anoxygenic photosynthetic bacteria, reaction centers are located in a structure called the intracytoplasmic membrane, which is a specialized invagination of the plasma membrane (36). This is in contrast to the autonomous nature of the PM and TM in cyanobacteria.

Fig. 5*B* shows the comparison between Arabidopsis and *Synechocystis* 6803 TM proteins. Although proteins from the majority of categories, such as photosystems, binding, and NDH, have similar contributions across these two photosynthetic species, the main differences lie in the corresponding contributions of metabolic and unknown proteins. About 44% of *Synechocystis* 6803 TM proteins have unknown functions, which might be because of differences in the levels of annotation of these two species. In contrast, compared with less than 1% of *Synechocystis* 6803 TM metabolic proteins, about 14% of Arabidopsis TM proteins function as metabolic proteins mainly in vitamin, pigment, and lipid metabolism. Given that a cyanobacterial ancestor was the progenitor of chloroplasts via an endosymbiotic event, these differences point to the changes that have occurred during the evolution of modern-day chloroplasts.

To summarize, the *Synechocystis* 6803 PM functions similarly to a Gram-negative bacterial PM with additional activities across different cellular processes, whereas the TM behaves in a slightly different way compared a plant TM, with lesser activities in metabolic processes. Overall, the *Synechocystis* 6803 membrane systems can be considered as a Gram-negative bacterial PM membrane system with bioenergetic electron transfer functioning in the TM.

Ⓢ This article contains supplemental material.

¶ To whom correspondence should be addressed: Himadri B. Pakrasi, Department of Biology, CB1137, Washington University, One Brookings Drive, St. Louis, MO 63130. Tel.: 314-935-6853; E-mail: pakrasi@wustl.edu.

‖ These authors contributed equally to this work.

Data have been deposited to the ProteomeXchange Consortium with dataset identifier PXD003079.

REFERENCES

1. Liberton, M., Austin, J. R., 2nd, Berg, R. H., and Pakrasi, H. B. (2011) Unique thylakoid membrane architecture of a unicellular N_2-fixing cyanobacterium revealed by electron tomography. *Plant Physiol.* **155,** 1656–1666

2. Liberton, M., Howard Berg, R., Heuser, J., Roth, R., and Pakrasi, H. B. (2006) Ultrastructure of the membrane systems in the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. *Protoplasma* **227,** 129–138

3. van de Meene, A. M., Hohmann-Marriott, M. F., Vermaas, W. F., and Roberson, R. W. (2006) The three-dimensional structure of the cyanobacterium *Synechocystis* sp. PCC 6803. *Arch. Microbiol.* **184,** 259–270

4. Schneider, D., Fuhrmann, E., Scholz, I., Hess, W. R., and Graumann, P. L. (2007) Fluorescence staining of live cyanobacterial cells suggest non-stringent chromosome segregation and absence of a connection between cytoplasmic and thylakoid membranes. *BMC Cell Biol.* **8,** 39

5. Pisareva, T., Kwon, J., Oh, J., Kim, S., Ge, C., Wieslander, A., Choi, J. S., and Norling, B. (2011) Model for membrane organization and protein sorting in the cyanobacterium *Synechocystis* sp. PCC 6803 inferred from proteomics and multivariate sequence analyses. *J. Proteome Res.* **10,** 3617–3631

6. Nickelsen, J., Rengstl, B., Stengel, A., Schottkowski, M., Soll, J., and Ankele, E. (2011) Biogenesis of the cyanobacterial thylakoid membrane system–an update. *FEMS Microbiol. Lett.* **315,** 1–5

7. Rengstl, B., Oster, U., Stengel, A., and Nickelsen, J. (2011) An intermediate membrane subfraction in cyanobacteria is involved in an assembly network for Photosystem II biogenesis. *J. Biol. Chem.* **286,** 21944–21951

8. 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. *Mol. Cell. Proteomics* **1,** 956–966

9. Pisareva, T., Shumskaya, M., Maddalo, G., Ilag, L., and Norling, B. (2007) Identification of novel integral plasma membrane proteins. *FEBS J.* **274,** 791–804

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

11. Keren, N., Liberton, M., and Pakrasi, H. B. (2005) Photochemical competence of assembled photosystem II core complex in cyanobacterial plasma membrane. *J. Biol. Chem.* **280,** 6548–6553

12. Welkie, D., Zhang, X., Markillie, M. L., Taylor, R., Orr, G., Jacobs, J., Bhide, K., Thimmapuram, J., Gritsenko, M., Mitchell, H., Smith, R. D., and Sherman, L. A. (2014) Transcriptomic and proteomic dynamics in the metabolism of a diazotrophic cyanobacterium, Cyanothece sp. PCC 7822 during a diurnal light-dark cycle. *BMC genomics* **15,** 1185

13. Livesay, E. A., Tang, K., Taylor, B. K., Buschbach, M. A., Hopkins, D. F., LaMarche, B. L., Zhao, R., Shen, Y., Orton, D. J., Moore, R. J., Kelly, R. T., Udseth, H. R., and Smith, R. D. (2008) Fully automated four-column capillary LC-MS system for maximizing throughput in proteomic analyses. *Anal. Chem.* **80,** 294–302

14. Kelly, R. T., Page, J. S., Luo, Q., Moore, R. J., Orton, D. J., Tang, K., and Smith, R. D. (2006) Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry. *Anal. Chem.* **78,** 7796–7801

15. Kim, S., Gupta, N., and Pevzner, P. A. (2008) Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J. Proteome Res.* **7,** 3354–3363

16. Kim, S., Mischerikow, N., Bandeira, N., Navarro, J. D., Wich, L., Mohammed, S., Heck, A. J., and Pevzner, P. A. (2010) The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. *Mol. Cell. Proteomics* **9,** 2840–2852

17. Elias, J. E., and Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4,** 207–214

18. Qian, W. J., Liu, T., Monroe, M. E., Strittmatter, E. F., Jacobs, J. M., Kangas, L. J., Petritis, K., Camp, D. G., 2nd, and Smith, R. D. (2005) Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. *J. Proteome Res.* 4, 53–62

19. Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Rios, D., Dianes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P. A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H. J., Albar, J. P., Martinez-Bartolome, S., Apweiler, R., Omenn, G. S., Martens, L., Jones, A. R., and Hermjakob, H. (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32,** 223–226

20. Polpitiya, A. D., Qian, W. J., Jaitly, N., Petyuk, V. A., Adkins, J. N., Camp, D. G., 2nd, Anderson, G. A., and Smith, R. D. (2008) DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* **24,** 1556–1558

21. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8,** 785–786

22. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305,** 567–580

23. Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H., and Krogh, A. (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* **12,** 1652–1662

24. Wuhr, M., Haas, W., McAlister, G. C., Peshkin, L., Rad, R., Kirschner, M. W., and Gygi, S. P. (2012) Accurate multiplexed proteomics at the MS2 level using the complement reporter ion cluster. *Anal. Chem.* **84,** 9214–9221

25. Wang, Y., Xu, W., and Chitnis, P. R. (2009) Identification and bioinformatic analysis of the membrane proteins of *Synechocystis* sp. PCC 6803. *Proteome Sci.* **7,** 11

26. Gao, L., Wang, J., Ge, H., Fang, L., Zhang, Y., Huang, X., and Wang, Y. (2015) Toward the complete proteome of *Synechocystis* sp. PCC 6803. *Photosynth Res.* **126,** 203–219

27. Dobakova, M., Sobotka, R., Tichy, M., and Komenda, J. (2009) Psb28 protein is involved in the biogenesis of the photosystem II inner antenna CP47 (PsbB) in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol.* **149,** 1076–1086

28. Selao, T. T., Zhang, L., Knoppova, J., Komenda, J., and Norling, B. (2016) Photosystem II assembly steps take place in the thylakoid membrane of the cyanobacterium *Synechocystis* sp. PCC6803. *Plant Cell Physiol.* **57,** 95–104

29. Lubec, G., Afjehi-Sadat, L., Yang, J. W., and John, J. P. (2005) Searching for hypothetical proteins: theory and practice based upon original data and literature. *Progress Neurobiol.* **77,** 90–127

30. Nakao, M., Okamoto, S., Kohara, M., Fujishiro, T., Fujisawa, T., Sato, S., Tabata, S., Kaneko, T., and Nakamura, Y. (2010) CyanoBase: the cyanobacteria genome database update 2010. *Nucleic Acids Res.* **38,** D379–D381

31. von Heijne, G. (2006) Membrane-protein topology. *Nat. Rev. Mol. Cell Biol.* **7,** 909–918

32. Ferro, M., Brugiere, S., Salvi, D., Seigneurin-Berny, D., Court, M., Moyet, L., Ramus, C., Miras, S., Mellal, M., Le Gall, S., Kieffer-Jaquinod, S., Bruley, C., Garin, J., Joyard, J., Masselon, C., and Rolland, N. (2010) AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. *Mol. Cell. Proteomics* **9,** 1063–1084

33. Tatusova, T., Ciufo, S., Fedorov, B., O'Neill, K., and Tolstoy, I. (2015) RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res.* **43,** 3872

34. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.* **43,** D204–D212

35. Shimizu, K. (2013) Regulation Systems of bacteria such as *Escherichia coli* in response to nutrient limitation and environmental stress. *Metabolites* **4,** 1–35

36. Sener, M. K., Olsen, J. D., Hunter, C. N., and Schulten, K. (2007) Atomic-level structural and functional model of a bacterial photosynthetic membrane vesicle. *Proc. Natl. Acad. Sci. U.S.A.* 15723–15728