Journal of proteome research

Article

Cite This: J. Proteome Res. XXXX, XXX, XXX−XXX

pubs.acs.org/jpr

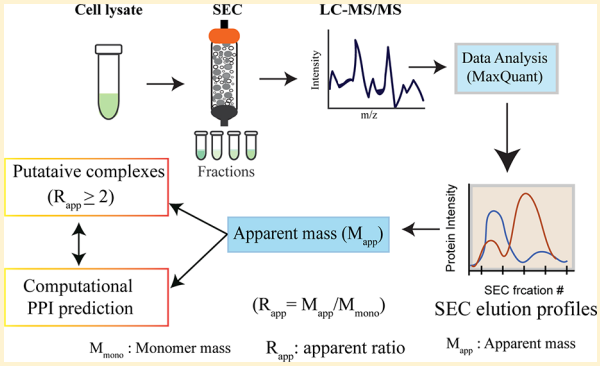
Analysis of Protein Complexes in the Unicellular Cyanobacterium *Cyanothece* ATCC 51142

Uma K. Aryal,*[,†,§] Ziyun Ding,[‡,§] Victoria Hedrick,[†] Tiago José Paschoal Sobreira,[†] Daisuke Kihara,[‡,∥] and Louis A. Sherman[∥]

[†]Bindley Bioscience Center, Discovery Park, [‡]Computer Science, and [∥]Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, United States

Supporting Information

ABSTRACT: The unicellular cyanobacterium *Cyanothece* ATCC 51142 is capable of oxygenic photosynthesis and biological N_2 fixation (BNF), a process highly sensitive to oxygen. Previous work has focused on determining protein expression levels under different growth conditions. A major gap of our knowledge is an understanding on how these expressed proteins are assembled into complexes and organized into metabolic pathways, an area that has not been thoroughly investigated. Here, we combined size-exclusion chromatography (SEC) with label-free quantitative mass spectrometry (MS) and bioinformatics to characterize many protein complexes from *Cyanothece* 51142 cells grown under a 12 h light−dark cycle. We identified 1386 proteins in duplicate biological replicates, and 64% of those proteins were identified as putative complexes. Pairwise computational prediction of protein−protein interaction (PPI) identified 74 822 putative interactions, of which 2337 interactions were highly correlated with published protein coexpressions. Many sequential glycolytic and TCA cycle enzymes were identified as putative complexes. We also identified many membrane complexes that contain cytoplasmic domains. Subunits of NDH-1 complex eluted in a fraction with an approximate mass of ∼669 kDa, and subunits composition revealed coexistence of distinct forms of NDH-1 complex subunits responsible for respiration, electron flow, and CO_2 uptake. The complex form of the phycocyanin beta subunit was nonphosphorylated, and the monomer form was phosphorylated at Ser20, suggesting phosphorylation-dependent deoligomerization of the phycocyanin beta subunit. This study provides an analytical platform for future studies to reveal how these complexes assemble and disassemble as a function of diurnal and circadian rhythms.



KEYWORDS: *Cyanothece 51142, size exclusion chromatography, proteomics, mass spectrometry, protein complexes, protein−protein interaction prediction*

INTRODUCTION

Cyanobacteria are photosynthetic organisms that have played important roles in harvesting solar energy on a global scale and in the evolution of the oxygenic atmosphere.[1] They have great potential as a platform for carbon sequestration and biological energy production.[2−5] Flexible and diverse metabolic capabilities allow them to adapt to a wide range of environments. Among them, unicellular species such as *Cyanothece* 51142 can also fix atmospheric N_2, a process highly sensitive to oxygen.[6] This ability to carry out two opposite biological processes within the same cell makes it an interesting model system to investigate the fundamental processes of photosynthesis, respiration, biological N_2 fixation, and carbon sequestration.[7,8] *Cyanothece* 51142 produces oxygen and stores photosynthetically fixed carbon in the form of glycogen granules during the day and subsequently metabolizes stored carbon to produce excess energy and to create an O_2-limited intracellular environment.[9,10] Respiratory electron transport scavenges oxygen to establish anaerobic intracellular conditions necessary for N_2 fixation. Thus, cyanobacteria are known to perform substantially different metabolic processes during the light−dark periods. The diversity of metabolic pathways allows them to succeed in a wide variety of environments and provide a wealth of targets for metabolic engineering of energy-rich biomolecules. These diverse metabolic processes are governed not only by the expression and relative abundances of proteins but also by their association, localization, modifications, as well as spatial and temporal distribution of functionally active protein complexes. Protein oligomerization is a central feature of many cellular control mechanisms, and the changes in metabolic activities of these microbes between the light and the dark periods must originate, in part, from the assembly and disassembly of protein complexes and cellular structure along the cycle. A thorough understanding of the biology of cyanobacteria requires in-depth knowledge of the composition

Received: March 14, 2018
Published: September 14, 2018

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

and dynamics of multiprotein complexes, an area that has not been thoroughly investigated.

Cyanothece 51142 show distinct circadian rhythms of photosynthesis and N_2 fixation with peaks every 24 h that are 12 h out of phase from each other. The genome indicates a wealth of metabolic potential, in addition to very active photosynthesis and CO_2 uptake mechanisms.[8] Under N_2-fixing conditions, *Cyanothece* 51142 cells become filled with large granules between the photosynthetic membranes.[9,11] These granules contain semiamylopectin and are more similar to starch than to typical bacterial glycogen. The branching pattern of this starch-like material is quite different from glycogen, and *Cyanothece* 51142 has a series of branching and debranching enzymes that might be involved. The composition and dynamics of assembly/disassembly enzyme complexes in *Cyanothece* 51142 for these glycogen granules are still outstanding. The sequencing of the genome[8] and the analysis of the transcriptome[12−14] and the proteome[7,15−17] have uncovered many diurnal- and circadian-controlled genes and protein expressions. However, the oscillation of proteins was less pronounced compared to the transcripts,[18] leaving us to speculate that the inventory of the genes and the proteins alone is not adequate to comprehend this organizational hierarchy. This led to the hypothesis that the molecular adaptation of *Cyanothece* 51142 occurs at a higher level organization of protein complexes and protein−protein interactions (PPIs).

In recent years, there have been increasing efforts directed toward generating proteome-wide maps of PPIs.[19] The most commonly used high-throughput methods for the study of protein complexes are yeast-two-hybrid (Y2H) screens[20,21] or affinity purification-mass spectrometry (AP-MS).[22−24] The Y2H screens are expensive, time consuming, and incomplete.[25] The N- or C-terminal tagging in the AP-MS method can affect the expression and interaction of endogenous proteins,[24,26] and the application of an AP-MS method is also limited by the availability of tagged constructs or antibodies.

An alternative size-based fractionation of native proteins via an SEC column combined with the high-resolution LC-MS/MS has recently been introduced.[27] SEC combined with LC-MS was applied to a non-N_2-fixing cyanobacterium *Synechococcus* elongates PCC 7942,[28] *Arabidopsis* cytosol,[19,29] and chloroplast[30] as well as human cell lysates.[31,32] In this study, we combined size fractionation of native proteins using Superdex 6 column with label-free LC-MS profiling and bioinformatic analysis to identify subunits of protein complexes in *Cyanothece* 51142. Many proteins involved in key physiological processes including the capture of sunlight to produce energy and evolve O_2, the capture of N_2 to make fixed nitrogen, the capture of CO_2 for fixed carbon, the storage of large amounts of carbohydrates that represent potential energy, and ridding the cytoplasm of toxic oxygen were identified as large protein complexes. The quality of the LC-MS profiling and complex prediction was evaluated by comparing two independent biological experiments in parallel and by the identification of previously characterized protein complexes.

2. EXPERIMENTAL METHODS

2.1. Cell Growth and Protein Extraction

Cyanothece 51142 cells were maintained as previously described[6] in ASP2 medium with $NaNO_3$ at 30 °C and continuous illumination of white light at 50 μmol of photons m^{-2} s^{-1}. Cultures for this study were also grown in the same growth medium by inoculating 1/10 volume of the stock cell cultures and maintained at 30 °C under 12 h light/dark cycle for 7 days before harvesting at 6 h into the light period. Cells were exposed to 50 μmol of photons m^{-2} s^{-1} white light during the light period. Cells were harvested by centrifugation at 14 000 rpm for 10 min at 4 °C. Pelleted cells were gently washed 2 times with ice-cold cell lysis buffer (20 mM Tris-HCl, pH 7.5, 5% glycerol, 50 mM KOAc, 2 mM $Mg(OAc)_2$, 1 mM EDTA, 1 mM EDTA, 0.5 mM DTT) followed by resuspension in 1 mL of the ice-cold lysis buffer. Cells were broken using a Precellys 24 Bead Mill Homogenizer (Bertin) at 6500 rpm for 3 cycles, each cycle lasting for 30s. Cell lysate was centrifuged at 14 000 rpm for 20 min at 4 °C, and proteins in the supernatant were separated using size exclusion chromatography (SEC). Protein concentration was measured using a bicinchoninic acid (BCA) assay (Pierce Chemical Co., Rockford, IL) before being separated in the SEC column.

2.2. Size Exclusion Chromatography

The soluble fraction (0.5 mL, ∼1 mg) was separated on a Superdex 200 10/300 GL column (GE Healthcare) using an ÄKTA FPLC system (Amersham Biosciences). Elution from the SEC column was performed with 20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, and 5% glycerol at a flow rate of 0.2 mL/min, and absorbance was monitored at 280 nm. Two biological replicates were processed identically. The column was calibrated using protein standards (MWGF1000, Sigma-Aldrich, St. Louis, MO) covering a mass range from 29 to 669 kDa. The void volume was measured with blue dextran. SEC separation was performed at 6 °C, and 20 SEC fractions of 500 μL were collected for mass spectrometry analysis as described below.

2.3. Sample Preparation for LC-MS Analysis

Sample preparation was carried out as described previously.[15] Briefly, proteins were denatured by adding 50 μL of 8 M urea for 1 h at room temperature, and the concentration in each fraction was determined by BCA assay. Proteins were reduced with 10 mM dithiothreitol (DTT); then cysteines were alkylated with IAA. Digestion was performed at 37°C overnight using mass spec-grade trypsin and Lys-C mix from Promega at a 1:25 (w/w) enzyme-to-substrate ratio. The digested peptides were desalted using Pierce C18 spin columns (Pierce Biotechnology, Rockford, IL). Peptides were eluted using 80% acetonitrile (ACN) containing 0.1% formic acid (FA) and dried in a vacuum concentrator at room temperature. Dried clean peptides were resuspended in 80 μL of the buffer containing 97% purified water, 3% ACN, and 0.1% FA. Peptides were loaded to the LC column by an equal volume (5 μL), not by an equal amount or concentration. In 80 μL of solution, the peptide concentration of the fraction that contained the highest protein amount (in this case fraction 21 in both biological replicates) was 0.2 μg/μL.

2.4. LC-MS/MS Data Acquisition

Samples were analyzed by reverse-phase HPLC-ESI-MS/MS using the Dionex UltiMate 3000 RSLC nano System coupled to the Q-Exactive High Field (HF) Hybrid Quadrupole Orbitrap Mass Spectrometer (Thermo Scientific, Waltham, MA) and a Nano-electrospray Flex ion source (Thermo Scientific). Purified peptides were loaded onto a trap column (300 μm ID × 5 mm) packed with 5 μm 100 Å PepMap C18 medium and washed using a flow rate of 5 μL/min with 98% purified water/2% ACN/0.01% FA. The trap column was then

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

switched in-line with the analytical column after 5 min. Peptides were separated using a reverse phase Acclaim PepMap RSLC C18 (75 μm × 15 cm) analytical column using a 120 min method at a flow rate of 300 nL/min. The analytical column was packed with 2 μm 100 Å PepMap C18 medium (Thermo Scientific). Mobile phase A consisted of 0.01% FA in water, and a mobile phase B consisted of 0.01% FA in 80% ACN. The linear gradient started at 5% B and reached 30% B in 80 min, 45% B in 91 min, and 100% B in 93 min. The column was held at 100% B for the next 5 min before being brought back to 5% B and held for 20 min to equilibrate the column. Sample was injected into the QE HF through the Nanospray Flex Ion Source fitted with a stainless steel emission tip from Thermo Scientific. Column temperature was maintained at 35°C. MS data was acquired with a Top20 data-dependent MS/MS scan method. The full MS spectra was collected over the 400−1650 m/z range with a maximum injection time of 100 ms, a resolution of 120 000 at 200 m/z., and AGC target of 1×10^6. Fragmentation of precursor ions was performed by high-energy C-trap dissociation (HCD) with the normalized collision energy of 27 eV. MS/MS scans were acquired at a resolution of 15 000 at m/z 200. The dynamic exclusion was set at 15 s to avoid repeated scanning of identical peptides. Instrument optimization and recalibration was carried out at the start of each batch run using the Pierce calibration solution. The sensitivity of the instrument was also monitored using an *E. coli* digest at the start of sample runs. Mass spectrometry raw data may be accessed from the MassIVE data repository (https://massive.ucsd.edu/) under MSV000082916.

2.5. Data Analysis

All LC-MS/MS data were analyzed using MaxQuant software (v. 1.5.3.28)[33−35] against the *Cyanothece* 51142 genome (http://img.jgi.doe.gov/cgi-bin/w/main.cgi) that contained 5300 nonredundant protein sequences. MaxQuant includes common contaminants as a default. No external contaminants were added to the database. The minimal length of six amino acids was required in the database search. The database search was performed with the precursor mass tolerance set to 10 ppm, and MS/MS fragment ions tolerance was set to 20 ppm. The database search was performed with the enzyme specificity for trypsin/Lys-C, allowing up to two missed cleavages. Oxidation of methionine (M) and phosphorylation of STY (pSTY) was defined as variable modifications, and carbamidomethylation of cysteine was defined as a fixed modification. MaxQuant search was performed as target decoy, and the false discovery rate (FDR) of peptide spectral match (PSM) and protein identification was set at 0.01. After the search peptides without any identifiable peak (0 intensity) and with no MS/MS counts were removed from consideration. At the protein level, proteins with 0 intensity and with 1 MS/MS counts were also removed from consideration. The "unique plus razor peptides" were used for peptide quantitation. Razor peptides are the nonunique peptides shared between the protein groups with the most other peptides. To increase the number of peptides that can be used for protein quantification and relative abundance profiling across SEC fractions, we enabled the "match between runs" function with a maximum retention time window of 1 min. This "match between runs" allows the transfer of peptide identification between fractions in the absence of peptide sequencing by MS/MS spectra, utilizing their accurate mass and aligned retention time.[33] The identified peptides and protein groups with their raw intensities were exported to Microsoft Access 2010 to perform subsequent analyses. The correlation coefficients between SEC fractions were calculated using the Data Analysis and Extension Tool (DAnTE).[36]

2.6. Data Normalization and Clustering of Protein Profiles

In a protein elution profile, the peak is defined as the elution fraction with the largest abundance among all fractions in each SEC experiment. Since the SEC experiment was repeated twice independently, the two independent experiments should generate similar elution profiles for the same protein. To ensure the quality of the elution profiles, the difference of the index of peak fraction was checked between the two SEC experiments, and only proteins with a peak index shift within 2 fractions were selected for clustering analysis, which indicates the SEC experimental results are consistent. Since the experiments were performed independently, the elution profiles generated by the two independent experiments were normalized independently by dividing the corresponding maximum intensity among each experiment. The elution profiles of Bio1 and Bio2 were normalized separately by dividing the LFQ intensities by the maximum intensity among the 20 fractions. The normalized 20 fractions from Bio1 and 20 fractions from Bio2 were concatenated into 40 fractions and clustered using the Euclidean distance measurement and the different combination of hierarchical methods such as average, complete, mcquitty, and ward. For each clustering method, different numbers of clusters were applied by cutting the dendrogram tree at different distances to determine the optimum number of clusters. Clustering results were compared with some known protein complexes to determine the cluster quality and the optimal cluster numbers.

2.7. Sequence-Based PPI Prediction

For sequence-based pairwise PPI prediction,[37] the amino acid sequences of *Cyanothece* 51142 proteins were downloaded from CyanoBase (http://genome.annotation.jp/CyanoBase).[38] The experimental results contained GeneBank protein IDs starting with "gi" and were converted into RefSeq ID following instructions on the GenBank webpage and the UniProt database.[39] For predicting PPI based on sequence information, we considered seven physiochemical properties including hydrophobicity, hydrophilicity, volumes of side chains of amino acids, polarity, polarizability, solvent-accessible surface area (SASA), and net charge index (NCI) of side chains of amino acids. The protein sequences were then represented as periodicity of each physicochemical property (eq 1)

$$\mathrm{AC}(lag, j) = \frac{\sum_{i=1}^{L-lag}\left(P_{i,j} - \frac{1}{L}\sum_{i=1}^{L} P_{i,j}\right) \times \left(P_{(i+lag),j} - \frac{1}{L}\sum_{i=1}^{L} P_{ij}\right)}{L - lag} \quad (1)$$

where lag is the distance between covariant residues to consider, which ranges from 1 to 30, j is the jth physiochemical descriptor, i is the position in the sequence, and L is the length of sequence. Each protein pair was transformed into 420 dimensional vectors.[40] Then support vector machine (SVM) (the software libsvm 2.84 http://www.csie.ntu.edu.tw/cjlin/libsvm/http://www.csie.ntu.edu.tw/cjlin/libsvm/)[41] was used to predict PPIs. SVM is a supervised learning method which

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

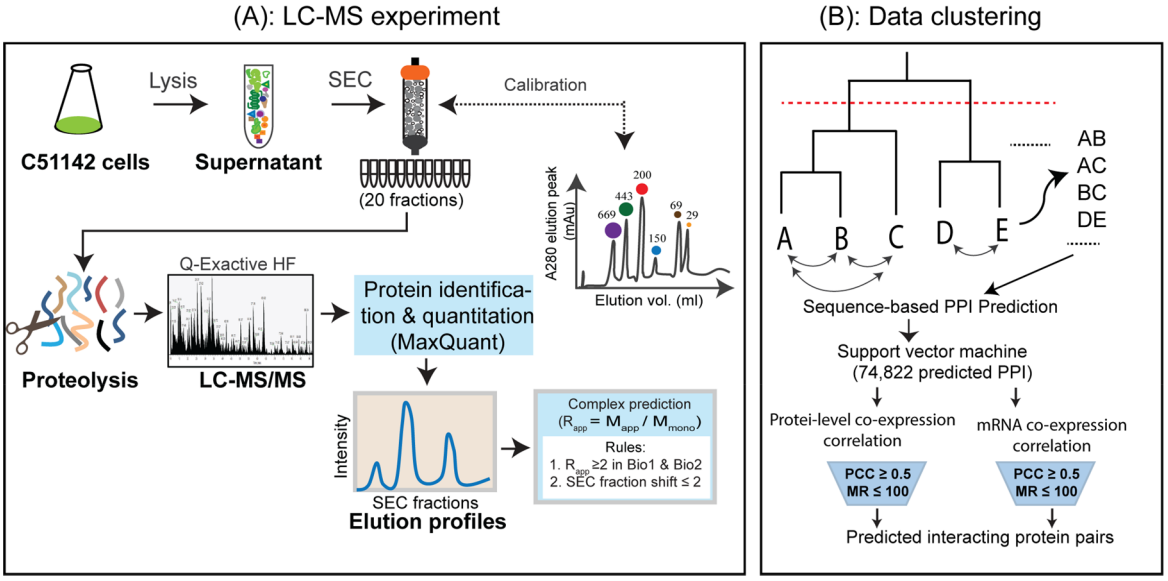


Figure 1. Experimental workflow. (A) Proteins extracted under native condition were fractionated by SEC and analyzed by a Q Exactive Orbitrap HF mass spectrometer. Data were analyzed using MaxQuant[33−35] for protein identification and label-free MS1 quantitation. Peak elution fraction of each identified protein, M_{app}, and R_{app} were determined as described previously.[19] M_{app}, apparent molecular mass. M_{mono}, predicted molecular mass of monomer. R_{app}, ratio of M_{app} to M_{mono} (M_{app}/M_{mono}). Proteins with an $R_{app} \geq 2$ in both replicates were considered to be in a complex.

uses the kernel function to transform the nonlinear features into linearly separable data. A total of 4908 experimentally verified nonredundant protein interactions in *Arabidopsis* were used as a training data set for the SVM. A radial basis function (RBF) was chosen as the kernel function with regularization parameters C and kernel parameter γ optimized as 32 and 0.5 because of the highest cross validation accuracy.

2.8. Calculation of Gene Coexpression Pearson's Correlation Coefficient and Mutual Ranks

Next, we compared the current protein complex profiles and the computationally predicted PPIsto previously published gene[14] and protein expression[16] data sets. The mRNA gene expression data set by Stockel et al.[14] includes 1443 genes of *Cyanothece* 51124. Five hundred seventy two out of 1443 proteins overlap with our experiment. The protein expression data set by Aryal et al.[16] was collected under day and night period and includes 976 proteins. Five hundred sixty one out of 976 proteins overlap with our experiment. Coexpression level of protein pairs was evaluated by the Pearson's correlation coefficient (PCC) (eq 2)

$$\mathrm{PCC} = \frac{\mathrm{cov}(A, B)}{\sigma_A \sigma_B} \quad (2)$$

where COV(A,B) is a covariance of protein A and B and σ_A and σ_B are the standard deviation of protein A and B, respectively. In Table S3 we provide the PCC of the day, night expression and the average of the two (overall PCC). For the protein coexpression data,[16] we also computed the mutual rank (MR) of coexpression strength

$$\mathrm{MR} = \sqrt{R_{A\to B} \times R_{B\to A}} \quad (3)$$

which is the geometric mean of the correlation rank of gene A to gene B ($R_{A\to B}$) and of gene B to gene A ($R_{B\to A}$) (eq 3). A small MR correlates to a stronger coexpression of the gene. MR is useful in evaluating coexpression when some genes weakly coexpressed with all other genes and have spurious PCC values. In Table S3 we provide PCC and MR for the protein expression data.[16]

3. RESULTS AND DISCUSSION

3.1. SEC Fractionation and LC-MS Reproducibility

Native *Cyanothece* 51142 proteins were separated into 20 SEC fractions (Figure 1, Supporting Information Figure S1). The void volume was determined based on the elution peak of blue dextran (Supporting Information Figure S1A). The molecular weight of proteins eluting in each SEC fraction was determined based on the calibration curve (Supporting Information Figure S1B). We performed two independent SEC fractionations (Supporting Information Figures S1C and S1D). Accuracy of label-free protein quantitation is limited if peptide intensity measurement is inconsistent. We tested the reproducibility of peptide signal intensity and peptide retention time on the LC column by analyzing three technical replicates from one of the fractions (F9 of Bio2). Of the total 1170 peptides and 335 proteins, 971 (83%) peptides and 298 (89%) proteins overlapped in all of the 3 technical replicates (Supporting Information Figures S2A and S2B), which is a good indication of LC-MS reproducibility for protein identification. The average coefficient of variation (CV) of MS1 intensity was ~15.1%, and the CV of the peptide retention time was <1.0% (Supporting Information Figures S2C and S2D), which also indicated good reproducibility for intensity-based label-free quantitation.

3.2. Global Analysis of the Expressed Proteome

In total, we identified 1567 proteins in Bio1 and 1436 proteins in Bio2, of which 1386 proteins (88% of Bio1 and 96% of Bio2) were common (Figure 2A). Pearson's correlation coefficient of 1386 protein intensities as a function of SEC fraction numbers (Figure 2B) showed the highest correlation coefficients along the diagonal, which indicated that protein elution peaks were reproducible between the biological replicates. However, the high correlation of signal intensities expanded to several adjacent fractions for high molecular

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

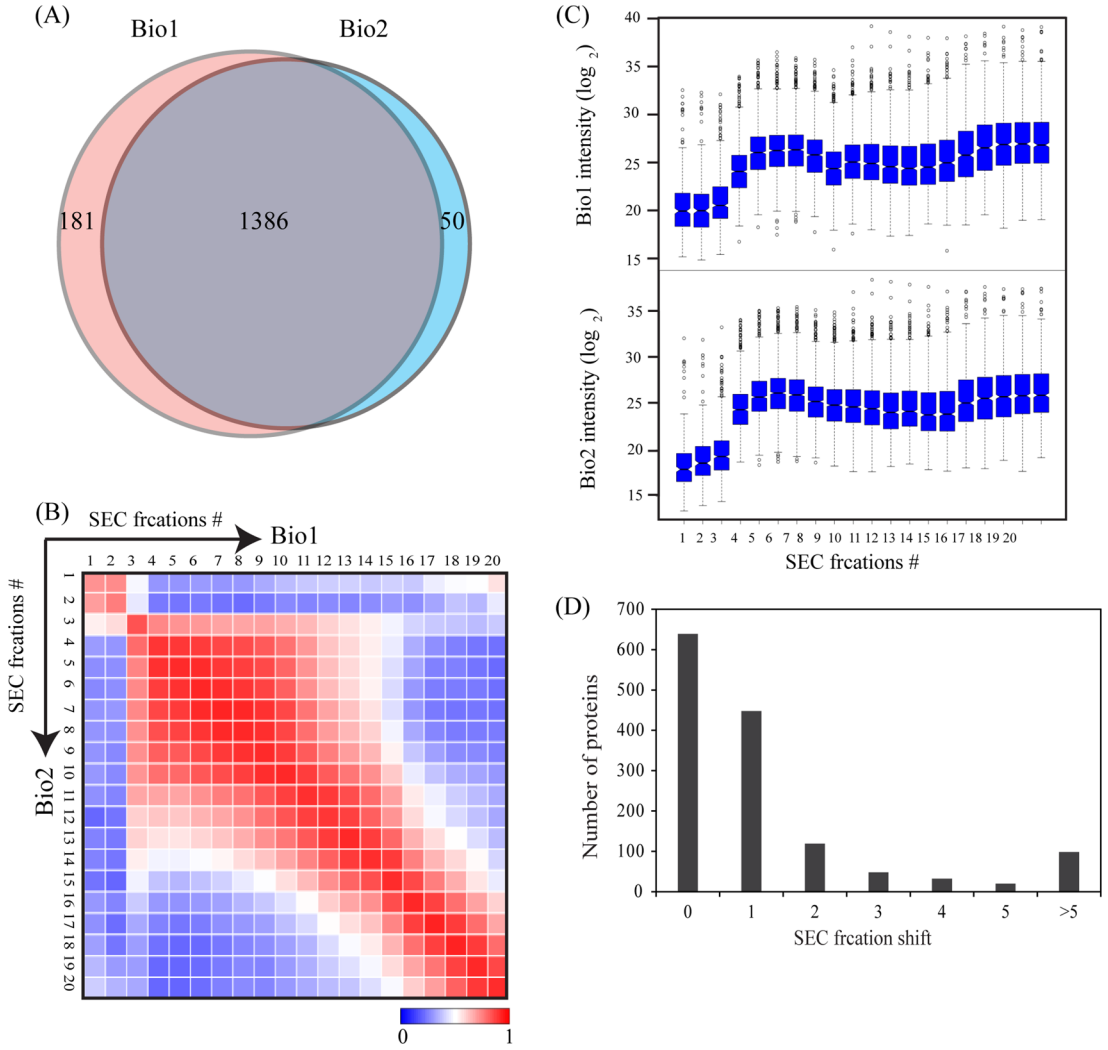


Figure 2. LC-MS reproducibility. (A) Venn diagram showing the overlap of proteins identified between two biological replicates. (B) Heat map showing the Pearson's correlation coefficients (PCC) of protein abundances (MS1 intensity) across SEC fractions. Correlation coefficients were calculated using the Data Analysis and Extension Tool (DAnTE).[36] (C) Box plot showing the median distribution of protein intensities. (D) Shift in peak elution fraction of proteins in two SEC separations; ~90% proteins were identified within 0−2 fractions shift, indicating good SEC reproducibility.

weight protein complexes. This is because the molecular weight of these proteins was beyond the size limit of the SEC column. The box plots in Figure 2C further confirmed that quantitation was consistent across column fractions. Reproducibility of protein elution peaks in the SEC column between the replicates is important to predict protein complexes based on their apparent mass (size). To check the reproducibility, we compared the shift in the elution peak fraction (global maximum) of all of the identified proteins between Bio1 and Bio2 (Figure 2D). Fifty-five percent of the proteins were identified without any peak shift (0 fraction shift), and >90% of the proteins were identified within a 0−2 fraction shift, confirming good SEC reproducibility.

3.3. Hierarchical Clustering of Protein Elution Profiles

Proteins with a similar elution profile were clustered and further subjected to bioinformatics predictions of PPI. Proteins interacting within complexes should display similar SEC elution profiles and belong to the same cluster. The results of different clustering methods (see Experimental Methods for details) were compared using several known protein complexes such as PSI, PSII, light-harvesting complex, ribosomal proteins, and others, and the method which assigned most of the known protein complex subunits within the same cluster was selected. Since these known protein complexes stably exist in the *Cyanothece* 51142, the clustering results did not differ much with different combinations of clustering methods. Because the computational method was used to further filter out the false interacting pairs with similar elution profiles, the smaller number of clusters with more proteins within each cluster was adopted in order to generate more protein pairs subject to prediction within the same cluster. The average linkage hierarchical clustering method with 30 clusters was used. The heat map of the Euclidean distance of elution profiles is plotted in Figure 3A. The heat map of elution throughout the SEC fractions shows that a significant number of proteins peaked at the high molecular weight fractions, which indicates that many proteins are migrating through the SEC column as complexes. To roughly estimate the proportion of proteins that migrate as stable complexes, we determined the peak elution fraction (global max) of each protein and used that global peak fraction to estimate the size or apparent (native) molecular

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

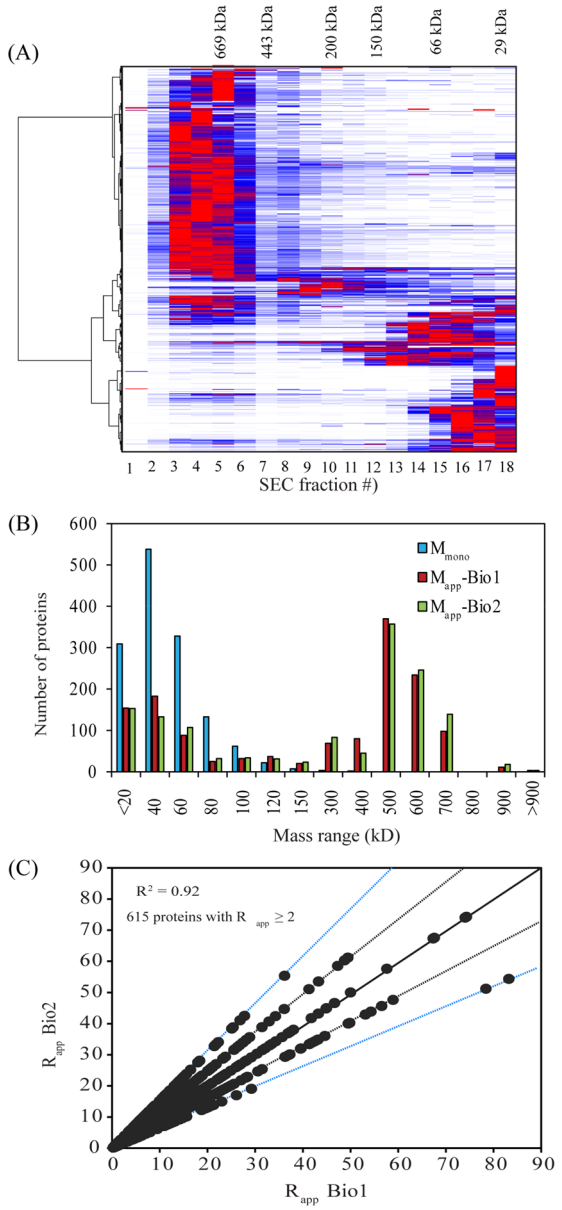


Figure 3. Determination of protein oligomerization states. (A) Hierarchical clustering of protein elution profiles. Proteins were clustered using the Euclidean distance and average linkage hierarchical clustering method. In this plot, each row represents a protein and each column represents the index of protein elution fraction. Numbers on the top show molecular masses of protein standards, and peak elution fraction for each of the standards was used to determine the M_{app} of proteins. (B) Histogram showing the distribution of the monomeric (blue) and experimentally determined apparent masses (green and red) of proteins that were identified in both biological replicates. (C) Scatter plots showing the R_{app} distribution of proteins between the two biological replicates. Each circle represents R_{app} values for Bio1 and Bio 2. Circles along the black solid line represent proteins without any fraction shift in elution peak (same R_{app} values) in both replicates. Circles along the black dotted lines represent proteins with 1 fraction shift, and circles along the blue dotted lines represent proteins with 2 fraction shifts between the replicates. Bio1, biological replicate 1. Bio2, biological replicate 2.

weight (M_{app}). Many proteins eluted in high-mass fractions, suggesting that they remained intact during SEC separation.

3.4. Determination of Protein Complexes

Figure 3B shows the distribution of the monomer (M_{mono}) and the M_{app} of proteins. The M_{mono} is concentrated in the lower molecular weight ranges, and M_{app} is concentrated in high molecular weight ranges. Previously, we used R_{app} (apparent ratio = M_{app} divided by M_{mono})[19,42,43] to define a protein complex as those having an R_{app} value of 2 or higher in both biological replicates. Despite several limitations, R_{app} is a useful metric to globally predict putative protein complexes. Figure 3C shows the R_{app} distribution of proteins in the two biological replicates. The circles along the solid line represent proteins eluting in exactly the same fraction, thus the same R_{app} values, in both biological replicates (0 fraction shift in elution peak); ∼55% of the proteins fell in this category. Circles along the dotted lines indicate proteins with 1 fraction shift, and ∼30% of the proteins had 1 fraction shift in their elution peaks. Our R_{app} predictions agreed well with the oligomerization state of several known protein complexes. For example, the M_{app} of PSI complex subunits ranged from ∼376 to 550 kDa (Supporting Information Table S2, rows 565−578), in agreement with the previous report.[44] Enolase peaked in fraction 11 with an M_{app} of ∼105 kDa, close to the known dimeric structure.[45] Enolase also peaked in a fraction with M_{app} of ∼105 kDa in our previous analysis using *Arabidopsis*.[19] Another glycolytic enzyme, phosphoenolpyruvate carboxylase (Ppc; cce_3822), was identified with R_{app} of 4.6 in both the replicates, close to the known tetrameric structure of this enzyme.[46] *Arabidopsis* PEPC (PEPC1 and PEPC2) were also detected with an R_{app} of ∼4 in our previous study.[19,29] Using R_{app} values we found that 64% (946 out of 1386) of the proteins detected in both biological replicates were predicted as complexes. The protein complexes were functionally diverse including those involved in translation, carbohydrate metabolism, photosynthesis, respiration, ion transport, folding, and ATP and metal ion binding (Supporting Information Figures S3A and S3B). Despite our mild lysis buffer, the protein list included both cytosolic and membrane proteins (Supporting Information Figure S3C). Our membrane protein list included many cytoplasmic and thylakoid membrane proteins and both cytoplasmic (hydrophilic) and membrane (hydrophobic) domain proteins. However, cytoplasmic domain proteins were detected with higher relative abundances than membrane domain proteins, indicating that they are more accessible for solubilization during extraction. It is important to mention here that we detected PsaA, PsaB, PsaC, PsbB, PsbC, and PsbA2, and all are known to be hydrophobic.

Protein sizes were also diverse ranging from ∼20 to ∼800 kDa. About 50% of those putative complexes eluted in either the void or high molecular weight (>600 kDa) fractions, including many 30S and 50S ribosomal proteins, PSI and PSII proteins (Supporting Information Figure S4), phycobilisomes, thioredoxins, ferroredoxins, glutaredoxins, NDH-1 complex (Figure 5), elongation factors, and many unknown or hypothetical proteins (Supporting Information Table S2). One-third of the proteins eluting in the void were unknown or hypothetical proteins. Many of these unknown proteins showed highly correlated elution profiles with other known protein complexes and also were predicted as interacting pairs by the computational method. For example, unknown protein cce_4744 showed a correlated elution profile with cytochrome f (PetC1; cce_2958) (Figure S5A); another unknown protein cce_0494 coeluted with PSII reaction center protein PsbB (cce_1837) and PsbC (cce_0659) (Figure S5B). In addition,

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

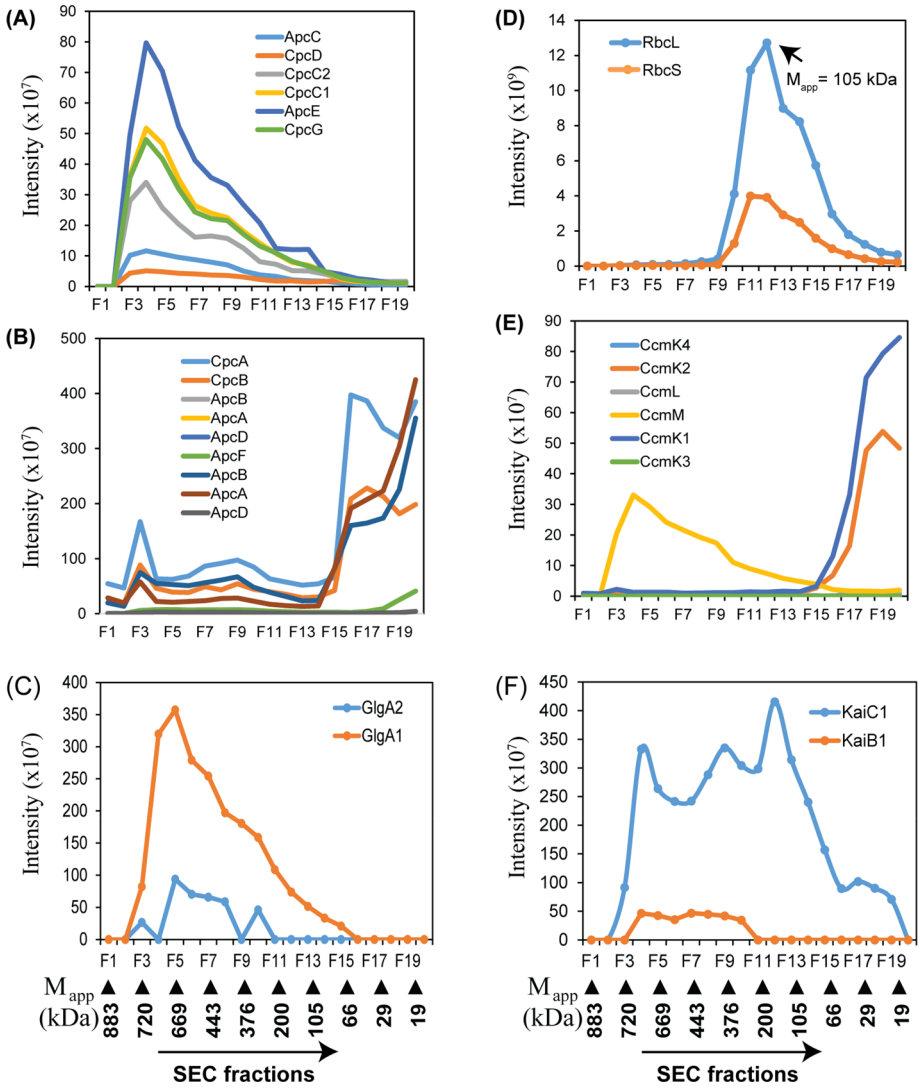


Figure 4. Elution profiles of phycobilisomes (PBS) and other complexes. (A, B) Elution profiles of phycocyanin (Cpc) and allophycocyanin (Apc) subunits. Elution profiles varied among the individual polypeptide. (C) Elution profiles of Rubisco large (RbcL) and small (RbcS) subunits. Both RbcL and RbcS peaked at fraction 12 with calculated M_{app} of 105 kDa. (D) Elution profiles of CO_2 concentrating mechanism (Ccm) proteins. CcmM showed a major elution peak as a complex, while others showed major peaks as monomers.

uncharacterized proteins cce_1749, cce_3678, and cce_3430 have highly correlated elution profiles with the protein involved in disulfide bond formation (cce_1972) (Figure S5C), and their protein-level expression is highly correlated (Supporting Information Table S3). These and several other evidence (Supporting Information Table S2) suggest that we may have uncovered many novel and apparently large protein complexes that are currently characterized as unknown.

Of the 1386 proteins, ~400 proteins were annotated as unknown and ~70 proteins were classified as hypothetical proteins. Two-thirds of these proteins (~300) have $R_{app} \geq 2$ in both biological replicates (Supporting Information Table S2). This suggests that we have detected many protein complexes whose function is currently unknown, and highlights the significant challenge ahead for functional characterization of these unknown proteins, as in general >40% of the proteome in prokaryotes and >50% in eukaryotes are not characterized.[47]

Our experimental system also detected proteins that are partitioned between the cytosol and the cytoplasmic and/or thylakoid membrane; indeed, there are a number of proteins with known membrane localization that were detected as apparent subunits of large complexes. Most of those detected membrane proteins are abundant proteins such as light-harvesting phycobilisomes proteins (Figure 4A and 4B), subunits of NDH-1 complex (Figure 5), PSI and PSII complexes (Supporting Information Figure S4), and the ATP synthases (Supporting Information Figure S6). It appears that subunits of these complexes are easily accessible for solublization during cell lysis due to cytoplasmic domain localization.

Key enzymes of glycolysis (GlgP1, Pgi1, Pgi2, PfkA1, Fda, Gap, Pgk, Eno1, Eno2, Ppc), TCA cycle (GltA, AcnB, SucC, SdhB, FumC), pentose phosphate (PP) pathways (Zwf, Gnd, TalA, Rpe, Pkt), and amino acid biosynthesis (AroQ, IlvN, TrpD, AroK, CysK, LeuB) (Supporting Information Table S2) eluted as stable complexes. Proteins involved in glycogen synthesis, GlgA1 (cce_3396) and GlgA2 (cce_0890), were identified as large protein complexes with M_{app} of 466 kDa and R_{app} > 5 in both replicates (Figure 4C). Of the three circadian clock (Kai) proteins, we identified KaiB (cce_0423) and KaiC (cce_0422; cce_4716) and eluted with multiple but consistent elution peaks in Bio1 and Bio2 (Supporting Information Table

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

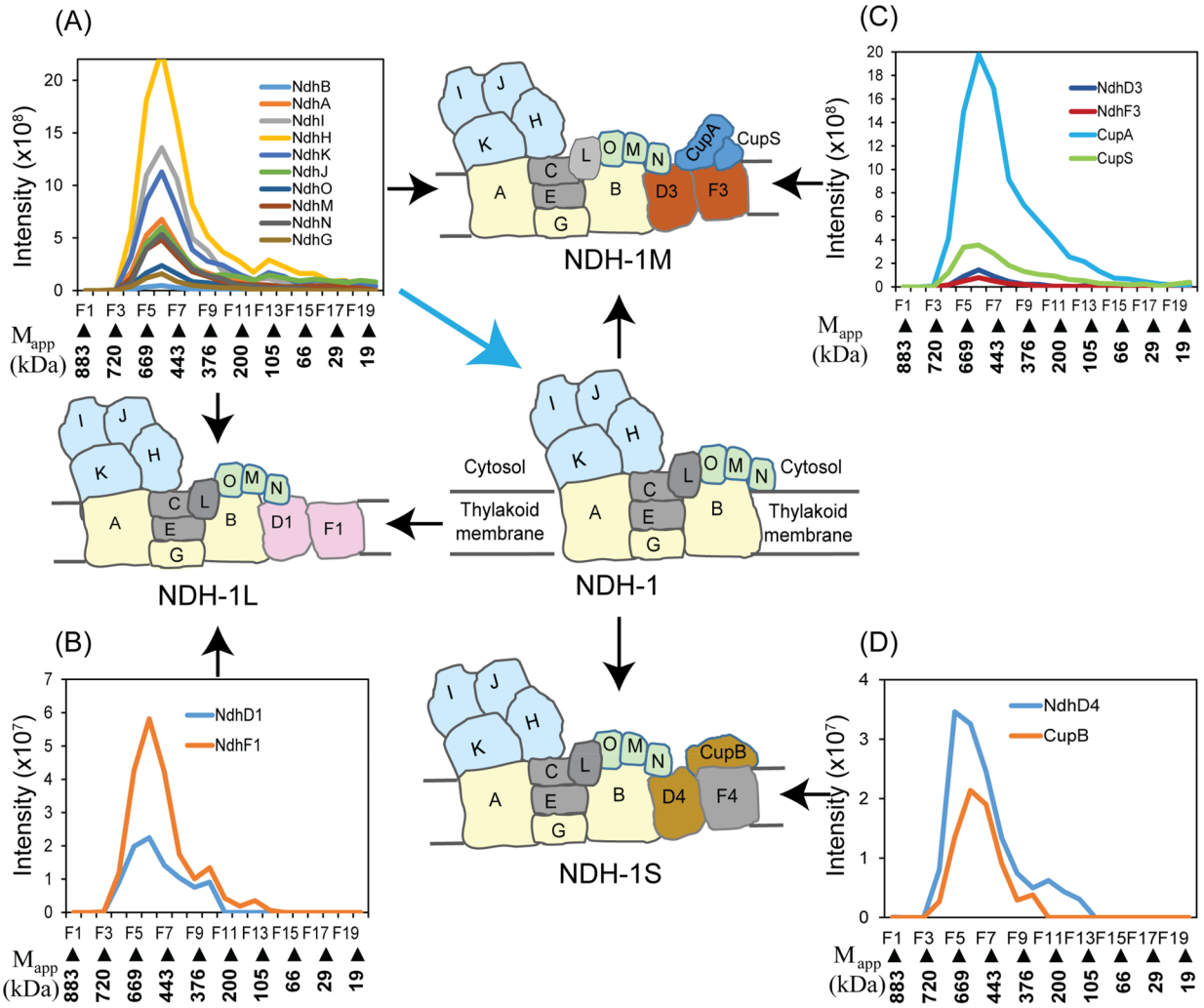


Figure 5. NDH-1 complex. (A–D) Elution profiles and structure of multiple forms of NDH-1 complex subunits. All subunits eluted in a high molecular weight (669 kDa) fraction. Existence of NDH-1L (respiratory) and NDH-1MS and NDH-1MS′ (CO_2 uptake) forms of NDH-1 complexes were determined by comparing SEC coelution profiles and known functional and structural multiplicity in the literature.[64,65] Hydrophilic domain subunits showed higher abundance than the membrane domain subunits. We identified both hydrophilic (I, J, K, H) and hydrophobic domain subunits (A, B, C, D1, F1, D3, F3, D4) as well as Oxygenic-Photosynthesis-Specific (OPS) domain subunits (O, M, N). Results show the existence of functional multiplicity of NDH-1 complexes in *Cyanothece* 51142 cells that are responsible for a variety of functions including respiration, cyclic electron flow, and CO_2 uptake.

S2, rows 236–238). The first elution peak corresponding to fraction 5 represents approximately 466 kDa in both replicates (Figure 4F). In cyanobacteria, KaiA and KaiB work together to modulate the activity of KaiC in a phosphorylation-dependent manner.[48] The link between metabolic activity and the circadian behavior has previously been reported,[48] and this link might be important in *Cyanothece* 51142 as these microbes are typically dependent on photosynthesis as an energy source.

3.5. Computational Protein–Protein Interaction Prediction

Pairwise sequence-based PPI prediction[49] identified 74 822 putative PPI pairs among all of the 1386 proteins, of which 561 proteins have been found in the previously published protein expression data by Aryal et al.,[16] and 572 genes overlap with mRNA expression data by Stockel et al..[14] To further select predicted PPI pairs with high confidence, we referred to these protein-level and mRNA-level coexpression information. In Table S3, predicted PPI pairs among the 561 proteins are selected that have a protein coexpression correlation[16] above 0 (Supporting Information Table S3, column F) or mutual rank below 100 (Supporting Information Table S3, column G) and with at least one mRNA coexpression correlation[14] above 0 (Supporting Information Table S3, column C)). There were in total 2461 such protein pairs. These proteins are plotted in Figure 6 with the Euclidean distance of protein elution profiles and Pearson's correlation coefficient of the mRNA-level coexpression information. If protein pairs are both annotated, the number of common GO terms of the protein pairs is indicated in a color scheme with a darker color for stronger function similarity. The figure shows such pairs with functional similarity mainly locate at the top left of the plot, which indicates that they have a higher coexpression correlation and similar elution profiles with each other. Thus, the plot implies that the similarity of elution profile and high expression correlation indeed capture physically interacting protein pairs.

The predicted PPI list (Table S3, Figure S7) includes pairs of obviously similar function such as PSI and PSII proteins, ribosomal proteins, cytochrome b6f complex, ATP synthases, NADPH-related proteins, chaperones, amino acid synthesis, and carbohydrate metabolism. Figure S7 visualizes the

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX–XXX

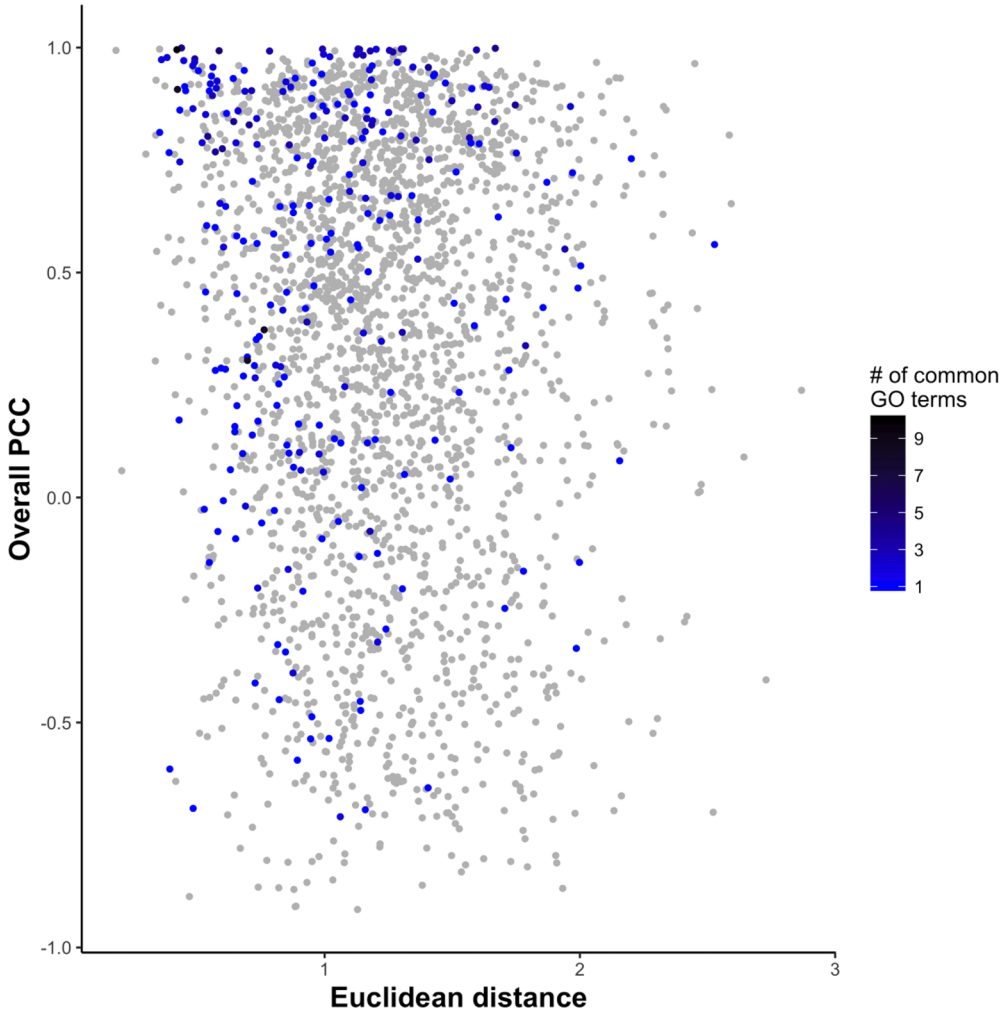


Figure 6. Plot of the Euclidean distance of protein elution profiles vs Pearson's correlation coefficient of the mRNA-level coexpression information.[14] Dots are colored based on the number of common GO terms. Gray indicates no common GO terms. Blue to black color indicates the number of common GO terms is from 1 to 10.

interaction network of the pairs using Cytoscape.[50] For example, PBS complex subunits ApcA and ApcB, which coelute together (Figure 4A), have a very high coexpression correlation at both protein and mRNA levels, share four GO terms (Supporting Information Table S3), and were predicted as interacting proteins (Figure S7A). CcmK1 and CcmK2 were also predicted as interacting pairs with a very high protein level and mRNA level coexpression (Supporting Information, Figure S7B) as well as AtpE and AtpB1 proteins (Supporting Information Table S3) and PSI and PSII polypeptides (Supporting Information Table S3, Figures S7C and S7D). PsaB and PsaA (MR = 2.83, PCC = 0.98) and CpcG, ApcE, and CpcC2 (Figure 4A) were predicted as interacting pairs with strong coexpression (Supporting Information Table S3). The NDH-1 complex subunits NdhO and NdhM (Figure 5) were predicted as interacting pairs with high protein coexpression score (MR, 39.87; PCC, 0.81) and 3 common GO terms (Supporting Information Table S3).

Additionally, we referred to the STRING score for the predicted protein pairs in Supporting Information Table S3. STRING is a database which provides various data that indicate functional and physical interactions of protein pairs in over 2000 organisms.[51] The plausibility of interactions is indicated by a score, which ranges from 0 to 1000, with 1000 for the most confident interaction. Thus, STRING provides further additional support of identified interacting protein pairs. Among the predicted PPIs with STRING combined score above 900, we discuss four interesting examples.

Putative homologues of glucose-1-phosphate adenylyltransferase (GlgC2; cce_2658) and phosphoglucomutase (cce_0770) are both involved in the glycogen biosynthesis pathway in 10 other organisms. Since proteins involved in the same pathway have a higher probability to interact, it is highly possible that these two proteins interact. Another example is the type IV pilus assembly protein (PilM; cce_1578) and hypothetical protein (cce_1579). Their genes are coded in the vicinity on the *Cyanothece* genome within only 4 bp intergenic distance, and they also co-occur across multiple organisms. Studies of protein interactions show that genes-encoding interacting proteins are kept close to each other on the genome,[52,53] and thus, these two proteins have high probability of interacting. The third example is uroprophyrinogen decarboxylase (HemE; cce_2966) and corproporphyrinogen III oxidase (HemF; cce_3201). They are both involved in the heme biosynthesis pathway and porphyrin chlorophyll metabolism pathway not only in *Cyanothece* 51142 but also in other 4 *Cyanothece* strains. Also, their putative homologous proteins are found to have correlated expression patterns in

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

other organism. The fourth example is the pyrroline-5-carboxylate reductase (ProC; cce_2615) and bifunctional proline dehydrogenase (PutA; cce_1595). They are both involved in the arginine and proline metabolism pathway. Furthermore, it has been shown in *Thermus thermophilus* HB27 that PutA catalyzes the conversion of proline to pyroline-5-carboxylate, which is the target of ProC.[54] Therefore, it is highly possible that these two homologous proteins also interact in *Cyanothece* 51142. Overall, we identified many large and apparently novel protein complexes in *Cyanothece* 51142 and further discuss several more complexes below, which are highlighted in yellow in Supporting Information Table S3. The high resolution and searchable cluster heat maps of all of the identified proteins are shown in the Supporting Information Figure S8 and S9 for Bio1 and Bio2, respectively.

3.6. Phycobilisome (PBS) Complex Assembly

The PBS assembly consists of rod and core complexes, which are connected by several nonpigment linker polypeptides.[55] Phycocyanin (PC) is the major phycobiliprotein of the rod, and the allophycocyanin (APC) is the major phycobiliprotein of the core cylinder. The PC rod−core linker polypeptide (CpcG) connects the rod to the core and plays a key role in the assembly of the PBS. Our experiments identified all major APC and PC polypeptides, and their elution profiles were remarkably consistent between Bio1 and Bio2 (Figure 4A and 4B). However, the elution profiles and the abundances varied among the individual polypeptides. For example, the ApcC, ApcE, CpcC1, CpcC2, CpcD, and CpcG eluted as a single peak at fraction 4 (M_{app} = 577 kDa) or 5 (M_{app} = 466 kDa) (Figures 4A) likely due to their stable association with the PBS assembly. In contrast, CpcA, CpcB, ApcA, and ApcB showed multiple peaks and mostly eluted as monomers (Figure 4B), suggesting variation in the stability and dissociation of different PBS polypeptides. The relative abundances of ApcA, ApcB, CpcA, and CpcB were higher than the other PBS polypeptides.

Grant and Lipschultz[56] reported that the high molarity of phosphate buffer (up to 1 M) was required to maintain the stability of the PBS assembly and suggested that PBS dissociated when exposed to cold temperature. Cell lysis in cold (4 °C) in the absence of a high salt (anion) concentration might have dissociated some of the PBS polypeptides in this study. Most recently, Zhang and co-workers[57] also used 0.65 M Na/K−PO_4 buffer to purify intact PBS complex from red alga, *Griffithsia pacifica*. We argue that our SEC-based profiling approach provides useful information to globally test the stability and dissociation of many protein complexes and can be a valuable source to develop protocols to isolate individual intact proteins or protein complexes.

3.7. Protein Complexes Associated with Carboxysomes

Carbon fixation in cyanobacteria occurs in carboxysomes through the compartmentalization of enzymatic reactions.[58] The carboxysomal beta-carbonic anhydrases, IcfA1 (cce_2257) and IcfA2 (cce_0871), showed a broad range of elution profiles but mostly eluted as complexes (Supporting Information Table S2, rows 212 and 213). The ribulose-1,5-bisphosphate carboxylase−oxygenase (RuBisCo) large (RbcL; cce_3166) and small (RbcS; cce_3164) subunits showed very similar elution profiles, and both peaked in fraction 12 with M_{app} of 105 kDa (Figure 4D). The elution profiles of CO_2 concentration mechanism proteins (CcmM, CcmL, CcmK1, CcmK2, CcmK3, CcmK4) were also highly consistent in Bio1 and Bio2 (Figure 4E), and all but CcmK2 and CcmK4 eluted in high-mass SEC fractions. Ccm protein profiles suggest variations in their stability, abundances, and complex association.

3.8. Enzymes Involved in Glycogen Synthesis and Metabolism

The glycogen granules in *Cyanothece* 51142 are formed via photosynthesis and metabolized in the dark as a substrate for respiration to make ATP and to reduce intracellular oxygen to protect nitrogenase.[9,11] The storage of α-glucan occurs through the sequential actions of ADP-glucose pyrophosphorylase (AGPase), glycogen/starch synthase (GS/SS), and a brancing enzyme (BE).[59] *Cyanothece* 51142 genome contains two genes for GS/SS (GlgA1; cce_0890 and GlgA2; cce_3396), two AGPase genes (GlgC1; cce_0987 and GlgC2; cce_2658), and three BE genes (cce_1806, GlgB1; cce_2248, and GlgB2; cce_4595). We identified all of the enzymes encoded by these genes (Supporting Information Table S2). The GlgA1, GlgA2 (Figure 4C), and cce_1806 (1,4-alpha-glucan branching enzyme) (Supporting Information Table S2, row 9) were identified as putative large complexes with R_{app} > 5 in both replicates. In contrast, the GlgC1 (cce_0987) and GlgC2 (cce_2658) eluted mostly as monomers (Supporting Information Table S2, rows 352 and 353). The two BE enzymes GlgB1 (cce_2248) and GlgB2 (cce_4595) were also detected as monomers (Supporting Information Table S2, rows 10 and 11). These varied elution profiles suggest that different enzymes involved in the same biochemical processes can have different complex stabilities. Enzymes responsible for glycogen metabolism (GlgP1; cce_1619, GlgP2; cce_5186 and GlgP3; cce_1603) showed broad elution profiles and eluted both as a complex as well as monomers (Supporting Information Table S2, rows 131, 130, and 129, respectively).

3.9. Photosynthesis, ATP Synthase, and Respiratory Complex Assembly

3.9.1. PSI and PSII. The thylakoid membrane of cyanobacteria contains PSI, PSII, the cytochrome b6f complex, and the ATP synthase. Both PSI (PsaA, PsaB, PsaC, PsaD, PsaE, PsaF, PsaK1, PsaK2, PsaL, PsaL2) and PSII polypeptides (PsbA1, PsbA2, PsbA3, PsbB, PsbC, PsbD1, PsbE, PsbF, PsbH, PsbL, PsbP, PsbQ, Psb27, Psb28, Psb28-2, and PsbV) eluted in a high-mass SEC fraction with calculated M_{app} from ∼500 to ∼400 kDa (Supporting Information Figure S4). Second minor elution peaks were observed in fraction 10 or 11 with a M_{app} of ∼160 or 120 kDa, but we did not detect any major peaks eluting as monomers. PSI assembly proteins Ycf3 (cce_0285), Ycf4 (cce_2172), and Ycf37 (cce_0285) also coeluted with other PSI proteins (Supporting Information Figure S4A). The PSI biogenesis protein (BtpA; cce_1973) was identified as a complex with R_{app} of 3.5 in Bio1 and 2.8 in Bio2 (Supporting Information Table S2, row 568). BtpA is an extrinsic thylakoid membrane protein and was recently found to be a necessary regulatory factor for the stabilization of the PsaA and PsaB proteins in *Synechocystis* sp. 6803.[60] The exact molecular mechanism of BtpA is currently unknown, but in *Synechocystis* sp. 6803, it was predicted to likely function as a chaperone, directly interacting with PsaA and PsaB proteins.[60] In our experiment, BtpA did not coelute with PsaA and PsaB, indicating that it may not directly interact with PsaA and PsaB.

Photosystem II eluted in fractions F5 and F6 with an approximate molecular weight of 500−600 kDa, which would be consistent with a PSII dimer.[61] As shown in Supporting Information Figure S4B, all of the major PSII proteins were

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

found in this fraction, including the oxygen-evolving complex (OEC) proteins PsbO and PsbQ and assembly proteins such as Psb28 and Psb27. On the other hand, two other proteins associated with the OEC, PsbU and PsbV, are found mostly as monomers in fractions F17−F19. Importantly, the D1 proteins encoded by the *psbA* genes and are not found in stoichiometric levels relative to components PsbB, PsbC, and PsbD. There are also lesser quantities of PsbA2 and PsbA3 in these fractions. This can be attributed to the repair and replacement of D1 proteins in the light.

The D1 protein of PSII, encoded by the *psbA* gene, is a key member of the PSII enzyme complex and provides multiple cofactors necessary to mediate light-induced oxidation of water to molecular oxygen.[62] The *Cyanothece* 51142 genome contains 5 *psbA* gene copies that encode 4 D1 protein isoforms. The D1 protein isoforms encoded by *psbA1* (cce_3501) and *psbA5* (cce_0636) are highly identical in amino acid sequences followed by the D1 isoforms encoded by *psbA3* (cce_0267) and *psbA2* (cce_3411) genes, respectively. The most divergent of the four *Cyanothece* 51142 D1 orthologs is the D1 isoform encoded by *psbA4* (cce_3477).[62] We identified D1 isoforms encoded by *psbA1*, *psbA5*, *psbA2*, and *psbA3* genes, and all coeluted in high molecular weight fractions (Supporting Information Figure S4B). The D1 isoforms encoded by *psbA1* and *psbA5* were grouped together (Supporting Information Table S2, row 585) due to their high amino acid similarity. We did not detect the most divergent D1 isoform encoded by *psbA4* gene. We previously demonstrated that PsbA4 would appear to be unable to bind the Mn cluster and is only expressed in the dark under N_2-fixing condition.[63] This may represent an evolutionary adaptation so that less O_2 can be evolved under such conditions. We will plan to study the differences in protein complexes in the light and dark in the near future.

3.9.2. ATP Synthase. We also detected many ATP synthase subunits (AtpA1, ATPA2, AtpB1, ATPB2, AtpC, AtpD, AtpE, AtpF1, AtpF2) with two clearly distinct elution peaks (Supporting Information Figure S6). AtpA1 and AtpB1 subunits were detected with higher abundances due to their accessibility to the cytoplasmic domain (Supporting Information Figure. S6). About one-third of the ATP synthase was recovered in fractions F5 and F6, consistent with the ATP synthase holoenzyme with most of the subunits, including at least some of the a and c proteins that are located entirely within the membrane. The other two-thirds of the ATP synthase proteins were found around fraction F12, which would be consistent with the cytoplasmic components, especially the α and β subunits, represented as $\alpha_2\beta_2$, α_3, and β_3 with additional proteins such as b, b′, δ, γ, and ε. Almost identical patterns were found in both biological replicates. The results with PSI, PSII, and the ATP synthase indicate that our isolation procedure in *Cyanothece* 51142 allowed full complexes to be extracted from the membrane, as long as a major component of the complex was in the cytoplasm.

3.9.3. NDH-1 Complex. The type 1 NAD(P)H: quinone oxidoreductase (NDH-1) is a membrane complex involved in diverse physiological functions.[64,65] Cyanobacterial NDH-1 consists of at least 17 subunits,[66] and we identified a total of 15 subunits, and all identified subunits eluted in a high molecular weight SEC fraction with $M_{app} \approx 669$ kDa (Figure 5). NDH-1 subunits are exposed to both cytoplasmic and membrane domains. While we identified both the hydrophilic (I, J, K, H) and the hydrophobic domain subunits (A, B, C, D1, F1, D3, F3, D4), the relative abundances of the hydrophilic domain subunits were clearly higher compared to the membrane domain subunits (Figure 5). The NDH-1 complexes in cyanobacteria share a common NDH-1 M core complex and differ in composition of the distal membrane domain comprised of specific NdhD and NdhF subunits and the hydrophilic carbon uptake (Cup) domain.[64] We identified NdhD1, NdhD3, NdhD4, NdhF1, and NdhF3 subunits of the membrane domain and CupA, CupB, and CupS subunits of the hydrophilic domain (Figure 5). Only NdhD2 and NdhF4 were not detected. Our analysis suggests the existence of 3 NDH-1 complex forms including NDH-1L, NDH-1MS, and NDH-1MS′ (Figure 5). NDH-1L functions in cyclic electron transport and respiration, and the NDH-1MS and NDH-1MS′ function in CO_2 uptake.[64] Many NDH subunits function to stabilize NDH-1 including NdhP and NdhQ. NdhP and NdhQ are also involved in respiratory and cyclic electron flow.[67] The NdhD3, NdhF3, and CupA have higher uptake affinity for CO_2 and function by the NdhD3/NdhF3/CupA/CupS system which forms a complex with a NDH-1MS. CupB, on the other hand, is involved in a constitutive CO_2 uptake system encoded by NdhD4/NdhF4/CupB which forms a complex NDH-1MS′.[65] Our results show the coexistence of multiple forms of NDH-1 complex in *Cyanothece* 51142 that are responsible for respiration, cyclic electron transport, and CO_2 uptake.

3.10. Phosphorylation-Dependent Protein Oligomerization

There is evidence suggesting that protein phosphorylation plays a key role in protein oligomerization.[68] Hence, there is growing interest to reveal the correlation between protein phosphorylation and protein complex formation. This requires the extraction of proteins under denaturing conditions, and our method described in this paper has advantages for studying phosphorylation-mediated protein oligomerization. Although we did not purify phosphorylated peptides, we searched out data using pSTY as variable modification and, as expected, identified a small number of phosphorylated proteins. The phycocyanin α (CpcA) and β (CpcB) subunits eluted as a complex as well as monomers, but the monomers showed the highest elution peaks (Figure 7A). The major complex peak had an approximate molecular weight of ∼304 kDa. Interestingly, the complex form (shaded blue) was non-phosphorylated, whereas the monomer (shaded orange) was phosphorylated at Ser20 (Figure 7A). Figure 7B shows the MS/MS spectra of the phosphorylated peptide GEFLSTSpQL-DALSQMVAESNKR mapped to CpcB, and Figure 7C shows the structure of α and β subunits showing the phosphorylated S20 site.

Cyanobacteria are known to utilize a two-component regulatory system for signal transduction to cope with changes in internal and external environment[69] and use a sensor kinase to transfer phosphate from a histidine residue on the enzyme to an aspartate residue on the response regulator.[70] In contrast, Ser/Thr/Tyr kinases (STKs) are known to be involved in eukaryotic signal transduction networks; however, recent functional genomic analyses have shown a wide distribution of STKs in prokaryotes-signaling networks as well.[71] Previously, functional analysis using *Synechocystis* 6803 mutants revealed that phosphorylation of the β subunits of phycocyanin is involved in the perception of high light and energy transfer, which affects state transitions.[69] Our results of the phosphorylation status of CpcA and CpcB monomers and non-

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

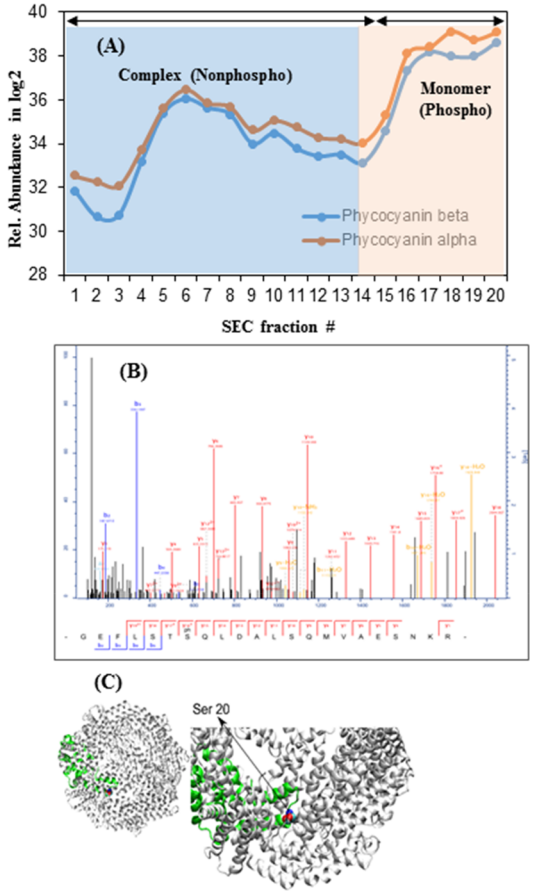


Figure 7. (A) Elution profiles of phycocyanin α and β subunits as a complex (blue) and as a monomer (orange). Proteins eluting as a complex were nonphosphorylated, and proteins eluting as a monomer were phosphorylated. (B) MS/MS spectra showing the phosphorylated peptide mapped to β subunit. (C) Structure of α and β subunits showing the phosphorylated S20 site. Results indicate phosphorylation-dependent deoligomerization of phycocyanin.

phosphorylation status of the complex provides new information about the novel role of protein phosphorylation in deoligomerization of phycocanin α and β subunits. Other proteins such as 2-phosphosulfolacetate phosphatase (ComB, cce_1018), 50S ribosomal protein (Rpl14, cce_4024), phosphatase ABC transporter (PstB1, cce_0883), and PSII protein Q (PsbQ, cce_0776) were also phosphorylated. Results indicate that identification of phosphorylated proteins were limited to abundant proteins likely due to the low stoichiometry of phosphorylation and no phosphopeptide enrichment.

3.11. Central Metabolism (Glycolysis, TCA Cycle, and PP Pathway)

One application of our method is to analyze the oligomerization state of enzymes that functions within a specific and well-characterized metabolic pathways such as glycolysis, TCA cycle, and pentose phosphate pathway (PPP).[19] Many glycolytic enzymes showed broad and multiple elution peaks (Figure 8). There is experimental evidence in plant and microbial systems for higher order physical associations of glycolytic enzymes that could mediate substrate channeling or the efficient delivery of pyruvate to the TCA cycle.[72,73] Hypothetically, if this is the case one would expect to detect coelution of sequential enzymes in *Cyanothece* 51142. Although this very simple regulatory scheme was difficult to explain by our elution patterns, we indeed observed overlapping elution peaks of many of the glycolytic enzymes (except ENOLASE and ALDOLASE) in high-mass SEC fractions (Figure 8A). This might be due to the existence of some stable and higher order organization of these glycolytic enzymes; however, the broad elution profiles were complex, and further investigation is needed. The majority of glycolytic enzymes except GlgP1, GpmA, and GpmB eluted as complexes. We identified all of the 3 GlgP isoforms (GlgP1, GlgP2, and GlgP3) which peaked elution at the low-mass (monomer) fractions, but they also have peaks at the high-mass SEC fractions, suggesting that a fraction of these enzymes also exists as a complex (Figure 8A). Similarly, GpmA and GpmB also showed the largest peak as monomer but also showed peaks as complexes (Figure 8A).

Many TCA cycles and the Oxidative PP pathway enzymes were also identified as complexes (Figure 8B). The citrate synthase, GltA (cce_1900), succinate dehydrogenase iron–sulfur protein subunit, SdhB (cce_3244), aconitase, AcnB (cce_3280), isocitrate dehydrogenase, Icsd (cce_3202), fumarate hydratase, FumC (cce_0396), and malate dehydrogenase, Mdh (cce_1850), and all but Icd, GabD, and Mdh were identified as putative complexes (Figure 8B). GltA eluted as a large complex with the R_{app} of 11 in both replicates. The R_{app} of SdhB was 6.6 in Bio1 and 8.2 in Bio2 (Figure 8B, Supporting Information Table S2, row 852). In contrast, SdhA (cce_0663) was identified with $R_{app} \approx 0.3$ in both replicates (Supporting Information Table S2, row 851). Oxidative PP enzymes including glucose 6-phosphate dehydrogenase, Zwf (cce_2536), 6-phosphogluconate dehydrogenase, Gnd (cce_3746), 6-phosphogluconolactonase, Pgl (cce_4743), ribulose-phosphate 3-epimerase, Rpe (cce_0798), and transaldolases, TalC (cce_4208) and TalA (cce_4687), were identified as complexes. For example, Zwf, a branching enzyme of the PP pathway, was identified with $R_{app} > 6$. In contrast, transketolase, TktA (cce_4627), and ribose 5-phosphate isomerase, RpiA (cce_0103), were identified as monomers (Figure 8B).

4. SUMMARY

The physiology of unicellular *Cyanothece* 51142 is diverse. An understanding of its physiology requires analysis of the full complement of proteins and the way they are organized and regulated in the cell. We started this by analyzing the *Cyanothece* 51142 protein complexes using the combination of SEC fractionation and quantitative LC-MS/MS profiling. This technique opens up the possibility for systems-wide studies of protein complex dynamics and interactions in cyanobacteria under various physiological conditions. We note that while this technique is very suitable for mapping stable complexes, transient or weak complexes have a higher chance to dissociate during lysis (and dilution) and SEC fractionation and consequently missed from the detection. Therefore, there is still a great need to develop a method that can better discover transient PPIs. Nonetheless, we successfully identified a number of protein complexes that are involved in key metabolic processes, which indicates the validity of our approach, and furthermore, many other known and unknown interacting pairs were identified (Supporting Information Table S3), which can serve as a valuable reference for future biological works.

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

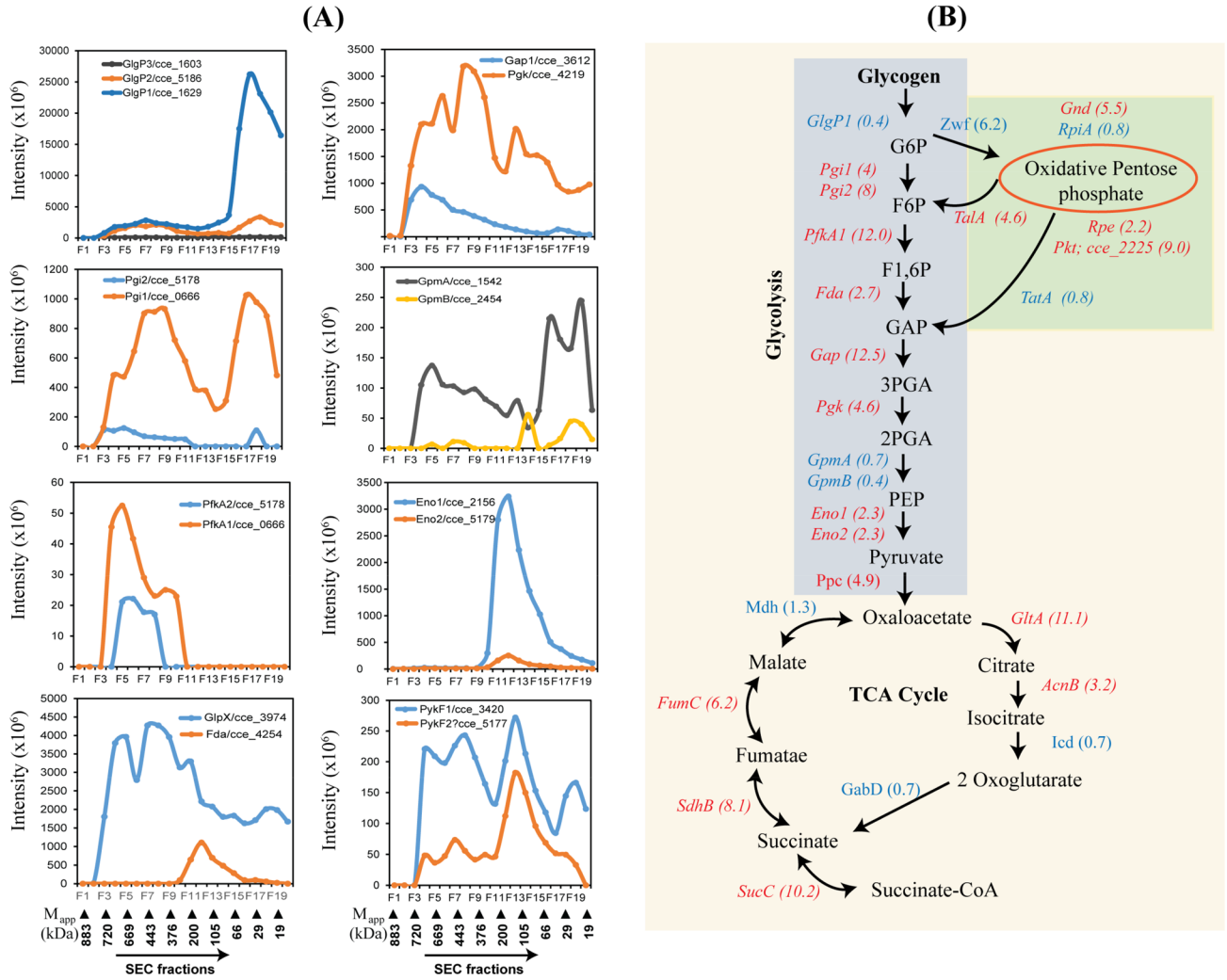


Figure 8. (A) Profiles of glycolytic enzymes organized to reflect their order in the pathway. (B) Biochemical pathways and enzymes involved in carbon metabolism. Pathways were generated by mapping proteins onto known pathways. Each arrow indicates the direction of the reaction. Symbols in red indicate proteins identified as putative complexes ($R_{app} \geq 2$) and in blue as monomers ($R_{app} \leq 1$). Numbers in parentheses correspond to the R_{app} value. GlgP1, glycogen phosphorylase (cce_1269); Zwf, glucose 6-phosphate dehydrogenase (cce_2536); Pgi; glucose-6-phosphate isomerase (Pgi1, cce_0666; Pgi2, cce_5178); PfkA1, 6-phosphofructokinase (cce_0669); Fda, fructose-bisphosphate aldolase class I (cce_4254); Gap, glyceraldehyde-3-phosphate dehydrogenase (cce_3612); Pgk, phosphoglycerate kinase (cce_4219); Gpm, phosphoglycerate mutase (GpmA, cce_1542; GpmB, cce_2454); Eno, enolase (Eno1, cce_2156; Eno2, cce_5179); Ppc, phosphoenolpyruvate carboxylase (cce_3822); Gnd, 6-phosphogluconate dehydrogenase (cce_3746); RpiA, ribose 5-phosphate isomerase (cce_0103); Rpe, ribulose-phosphate 3-epimerase (cce_0798); TalA, transaldolase (cce_4687); TktA, transketolase (cce_4627); Pkt, phosphoketolase (cce_2225); GltA, citrate synthase (cce_1900); AcnB, aconitate hydratase 2 (cce_3280); Icd, isocitrate dehydrogenase (cce_3202); GabD, succinate-semialdehyde dehydrogenase (cce_4228); SucC, succinyl-CoA synthetase (cce_2357); SdhB, succinate dehydrogenase iron−sulfur protein subunit (cce_3244); FumC, fumarate hydratase (cce_0396); Mdh, malate dehydrogenase (cce_1850); G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; F1,6P, fructose 1,6-bisphosphate; Gap, glyceraldehyde-3-phosphate; 3PGA, 3-phosphoglycerate; 2PGA, 2-phosphoglycerate; PEP, phosphoenolpyruvate.

In this work we used bioinformatics analysis to follow up the experiments to provide further supporting evidence of detected PPIs. Since the protein clusters with their elution profiles from the SEC fractionation only provide sets of proteins that have similar profiles, bioinformatics analysis is necessary to actually identify interacting pairs. As a future direction of this work, we can further construct a tertiary structure of protein complexes using protein-docking programs[74−76] to provide residue- and atom-level information on protein complexes.

In conclusion, this work represents the first comprehensive analysis on a large-scale protein complex study in *Cyanothece* 51142. Thus far, differential and quantitative proteomic analysis of soluble and membrane proteins of this strain has been well established, and a wealth of information on proteins involved in major metabolic pathways is known. However, how these proteins assemble into complexes and function was largely unknown. Here, we were able to add protein complex information with other qualitative and quantitative information and established an isolation procedure and analytical platform for future studies to reveal how these protein complexes assemble and disassemble as a function of diurnal and circadian rhythms.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.8b00170.

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

List of peptides commonly identified in duplicate biological runs; list of proteins commonly identified in duplicate biological runs; list of computationally predicted protein−protein interactions (XLSX)

SEC elution profiles of protein standards used to calibrate the column and *Cyanothece* 52241 proteins; overlaps of peptides and proteins identification and coefficient of variation (CV) in technical replicates; distribution of proteins into biological processes, molecular functions, and cellular components; elution profiles and subunit stoichiometry of PSI and PSII polypeptides; correlated elution of unknown proteins with known protein complexes; elution profiles and subunit stoichiometry of ATP synthases; sequence-based prediction of protein−protein interaction network in Cytoscape; searchable heat map of proteins in biological replicate 1; searchable heat map of proteins in biological replicate 2 (PDF)

AUTHOR INFORMATION

Corresponding Author

*Phone: (765) 494-4960. E-mail: uaryal@purdue.edu.

ORCID

Uma K. Aryal: 0000-0003-4543-1536

Daisuke Kihara: 0000-0003-4091-6614

Author Contributions

§U.K.A. and Z.D.: Co-first authors.

Author Contributions

U.K.A. and L.A.S. designed the research, U.K.A. and V.H. performed proteomic experiments, U.K.A., Z.D., T.J.P.S., D.K., and L.A.S. analyzed the data, U.K.A., Z.D., L.A.S., and D.K. wrote the paper. All authors read, edited, and approved the contents of the paper.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

All of the LC-MS/MS data were collected at the Purdue Proteomics Facility. This work was supported in part by a grant from the DOE Genomics GTL program (DE 09-19 PO 2905402N; H.P., Principle Investigator; L.A.S., Co-principle Investigator). D.K. also acknowledges the funding support from the National Institutes of Health (R01GM123055) and the National Science Foundation (IOS1127027, DMS1614777). Z.D. is partly supported by the Purdue Research Foundation.

REFERENCES

(1) Elvitigala, T.; Stockel, J.; Ghosh, B. K.; Pakrasi, H. B. Effect of continuous light on diurnal rhythms in Cyanothece sp. ATCC 51142. *BMC Genomics* **2009**, *10*, 226.

(2) Dutta, D.; De, D.; Chaudhuri, S.; Bhattacharya, S. K. Hydrogen production by Cyanobacteria. *Microb. Cell Fact.* **2005**, *4*, 36.

(3) Ghirardi, M. L.; Zhang, L.; Lee, J. W.; Flynn, T.; Seibert, M.; Greenbaum, E.; Melis, A. Microalgae: a green source of renewable H(2). *Trends Biotechnol.* **2000**, *18* (12), 506−511.

(4) Nozzi, N. E.; Oliver, J. W.; Atsumi, S. Cyanobacteria as a Platform for Biofuel Production. *Front. Bioeng. Biotechnol.* **2013**, *1*, 7.

(5) Zehr, J. P.; Waterbury, J. B.; Turner, P. J.; Montoya, J. P.; Omoregie, E.; Steward, G. F.; Hansen, A.; Karl, D. M. Unicellular cyanobacteria fix N2 in the subtropical North Pacific Ocean. *Nature* **2001**, *412* (6847), 635−8.

(6) Reddy, K. J.; Haskell, J. B.; Sherman, D. M.; Sherman, L. A. Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus Cyanothece. *J. Bacteriol.* **1993**, *175* (5), 1284−92.

(7) Stockel, J.; Jacobs, J. M.; Elvitigala, T. R.; Liberton, M.; Welsh, E. A.; Polpitiya, A. D.; Gritsenko, M. A.; Nicora, C. D.; Koppenaal, D. W.; Smith, R. D.; Pakrasi, H. B. Diurnal rhythms result in significant changes in the cellular protein complement in the cyanobacterium Cyanothece 51142. *PLoS One* **2011**, *6* (2), e16680.

(8) Welsh, E. A.; Liberton, M.; Stockel, J.; Loh, T.; Elvitigala, T.; Wang, C.; Wollam, A.; Fulton, R. S.; Clifton, S. W.; Jacobs, J. M.; Aurora, R.; Ghosh, B. K.; Sherman, L. A.; Smith, R. D.; Wilson, R. K.; Pakrasi, H. B. The genome of Cyanothece 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (39), 15094−9.

(9) Schneegurt, M. A.; Sherman, D. M.; Sherman, L. A. Composition of the carbohydrate granules of the cyanobacterium, Cyanothece sp. strain ATCC 51142. *Arch. Microbiol.* **1997**, *167* (2−3), 89−98.

(10) Sherman, L. A.; Meunier, P.; Colon-Lopez, M. S. Diurnal rhythms in metabolism: A day in the life of a unicellular, diazotrophic cyanobacterium. *Photosynth. Res.* **1998**, *58* (1), 25−42.

(11) Schneegurt, M. A.; Sherman, D. M.; Nayar, S.; Sherman, L. A. Oscillating behavior of carbohydrate granule formation and dinitrogen fixation in the cyanobacterium Cyanothece sp. strain ATCC 51142. *J. Bacteriol.* **1994**, *176* (6), 1586−97.

(12) Bandyopadhyay, A.; Elvitigala, T.; Welsh, E.; Stockel, J.; Liberton, M.; Min, H.; Sherman, L. A.; Pakrasi, H. B. Novel metabolic attributes of the genus cyanothece, comprising a group of unicellular nitrogen-fixing Cyanothece. *mBio* **2011**, *2* (5), 00214-11.

(13) Toepel, J.; Welsh, E.; Summerfield, T. C.; Pakrasi, H. B.; Sherman, L. A. Differential transcriptional analysis of the cyanobacterium Cyanothece sp. strain ATCC 51142 during light-dark and continuous-light growth. *J. Bacteriol.* **2008**, *190* (11), 3904−13.

(14) Stockel, J.; Welsh, E. A.; Liberton, M.; Kunnvakkam, R.; Aurora, R.; Pakrasi, H. B. Global transcriptomic analysis of Cyanothece 51142 reveals robust diurnal oscillation of central metabolic processes. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (16), 6156−61.

(15) Aryal, U. K.; Callister, S. J.; McMahon, B. H.; McCue, L. A.; Brown, J.; Stockel, J.; Liberton, M.; Mishra, S.; Zhang, X.; Nicora, C. D.; Angel, T. E.; Koppenaal, D. W.; Smith, R. D.; Pakrasi, H. B.; Sherman, L. A. Proteomic profiles of five strains of oxygenic photosynthetic cyanobacteria of the genus Cyanothece. *J. Proteome Res.* **2014**, *13* (7), 3262−76.

(16) Aryal, U. K.; Stockel, J.; Krovvidi, R. K.; Gritsenko, M. A.; Monroe, M. E.; Moore, R. J.; Koppenaal, D. W.; Smith, R. D.; Pakrasi, H. B.; Jacobs, J. M. Dynamic proteomic profiling of a unicellular cyanobacterium Cyanothece ATCC51142 across light-dark diurnal cycles. *BMC Syst. Biol.* **2011**, *5*, 194.

(17) Aryal, U. K.; Stockel, J.; Welsh, E. A.; Gritsenko, M. A.; Nicora, C. D.; Koppenaal, D. W.; Smith, R. D.; Pakrasi, H. B.; Jacobs, J. M. Dynamic proteome analysis of Cyanothece sp. ATCC 51142 under constant light. *J. Proteome Res.* **2012**, *11* (2), 609−19.

(18) McDermott, J. E.; Archuleta, M.; Stevens, S. L.; Stenzel-Poore, M. P.; Sanfilippo, A. Defining the players in higher-order networks: predictive modeling for reverse engineering functional influence networks. *Pac. Symp. Biocomput.* **2011**, 314−325.

(19) Aryal, U. K.; Xiong, Y.; McBride, Z.; Kihara, D.; Xie, J.; Hall, M. C.; Szymanski, D. B. A proteomic strategy for global analysis of plant protein complexes. *Plant Cell* **2014**, *26* (10), 3867−82.

(20) Rolland, T.; Tasan, M.; Charloteaux, B.; Pevzner, S. J.; Zhong, Q.; Sahni, N.; Yi, S.; Lemmens, I.; Fontanillo, C.; Mosca, R.; Kamburov, A.; Ghiassian, S. D.; Yang, X.; Ghamsari, L.; Balcha, D.; Begg, B. E.; Braun, P.; Brehme, M.; Broly, M. P.; Carvunis, A. R.; Convery-Zupan, D.; Corominas, R.; Coulombe-Huntington, J.; Dann, E.; Dreze, M.; Dricot, A.; Fan, C.; Franzosa, E.; Gebreab, F.; Gutierrez, B. J.; Hardy, M. F.; Jin, M.; Kang, S.; Kiros, R.; Lin, G. N.;

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

Luck, K.; MacWilliams, A.; Menche, J.; Murray, R. R.; Palagi, A.; Poulin, M. M.; Rambout, X.; Rasla, J.; Reichert, P.; Romero, V.; Ruyssinck, E.; Sahalie, J. M.; Scholz, A.; Shah, A. A.; Sharma, A.; Shen, Y.; Spirohn, K.; Tam, S.; Tejeda, A. O.; Trigg, S. A.; Twizere, J. C.; Vega, K.; Walsh, J.; Cusick, M. E.; Xia, Y.; Barabasi, A. L.; Iakoucheva, L. M.; Aloy, P.; De Las Rivas, J.; Tavernier, J.; Calderwood, M. A.; Hill, D. E.; Hao, T.; Roth, F. P.; Vidal, M. A proteome-scale map of the human interactome network. *Cell* **2014**, *159* (5), 1212−26.

(21) Jansen, R.; Yu, H.; Greenbaum, D.; Kluger, Y.; Krogan, N. J.; Chung, S.; Emili, A.; Snyder, M.; Greenblatt, J. F.; Gerstein, M. A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science* **2003**, *302* (5644), 449−453.

(22) Dunham, W. H.; Mullin, M.; Gingras, A. C. Affinity-purification coupled to mass spectrometry: basic principles and strategies. *Proteomics* **2012**, *12* (10), 1576−90.

(23) Altelaar, A. F.; Munoz, J.; Heck, A. J. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **2013**, *14* (1), 35−48.

(24) Rigaut, G.; Shevchenko, A.; Rutz, B.; Wilm, M.; Mann, M.; Seraphin, B. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **1999**, *17* (10), 1030−2.

(25) Du, C.; Reade, J. P.; Rogers, L. J.; Gallon, J. R. Dinitrogenase reductase ADP-ribosyl transferase and dinitrogenase reductase activating glycohydrolase in Gloeothece. *Biochem. Soc. Trans.* **1994**, *22* (3), 332S.

(26) Wodak, S. J.; Pu, S.; Vlasblom, J.; Seraphin, B. Challenges and rewards of interaction proteomics. *Mol. Cell. Proteomics* **2009**, *8* (1), 3−18.

(27) Dong, M.; Yang, L. L.; Williams, K.; Fisher, S. J.; Hall, S. C.; Biggin, M. D.; Jin, J.; Witkowska, H. E. A "tagless" strategy for identification of stable protein complexes genome-wide by multidimensional orthogonal chromatographic separation and iTRAQ reagent tracking. *J. Proteome Res.* **2008**, *7* (5), 1836−49.

(28) Guerreiro, A. C.; Penning, R.; Raaijmakers, L. M.; Axman, I. M.; Heck, A. J.; Altelaar, A. F. Monitoring light/dark association dynamics of multi-protein complexes in cyanobacteria using size exclusion chromatography-based proteomics. *J. Proteomics* **2016**, *142*, 33−44.

(29) Aryal, U. K.; McBride, Z.; Chen, D.; Xie, J.; Szymanski, D. B. Analysis of protein complexes in Arabidopsis leaves using size exclusion chromatography and label-free protein correlation profiling. *J. Proteomics* **2017**, *166*, 8−18.

(30) Olinares, P. D.; Ponnala, L.; van Wijk, K. J. Megadalton complexes in the chloroplast stroma of Arabidopsis thaliana characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering. *Mol. Cell. Proteomics* **2010**, *9* (7), 1594−615.

(31) Kirkwood, K. J.; Ahmad, Y.; Larance, M.; Lamond, A. I. Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics. *Mol. Cell. Proteomics* **2013**, *12* (12), 3851−73.

(32) Kristensen, A. R.; Gsponer, J.; Foster, L. J. A high-throughput approach for measuring temporal changes in the interactome. *Nat. Methods* **2012**, *9* (9), 907−9.

(33) Cox, J.; Hein, M. Y.; Luber, C. A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **2014**, *13* (9), 2513−26.

(34) Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26* (12), 1367−72.

(35) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **2011**, *10* (4), 1794−805.

(36) Polpitiya, A. D.; Qian, W. J.; Jaitly, N.; Petyuk, V. A.; Adkins, J. N.; Camp, D. G., 2nd; Anderson, G. A.; Smith, R. D. DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* **2008**, *24* (13), 1556−8.

(37) Ding, Z.; Kihara, D. Computational Methods for Predicting Protein-Protein Interactions Using Various Protein Features. *Current Protocols in Protein Science* **2018**, *93*, e62.

(38) Fujisawa, T.; Narikawa, R.; Maeda, S. I.; Watanabe, S.; Kanesaki, Y.; Kobayashi, K.; Nomata, J.; Hanaoka, M.; Watanabe, M.; Ehira, S.; Suzuki, E.; Awai, K.; Nakamura, Y. CyanoBase: a large-scale update on its 20th anniversary. *Nucleic Acids Res.* **2017**, *45* (D1), D551−D554.

(39) Benson, D. A.; Karsch-Mizrachi, I.; Clark, K.; Lipman, D. J.; Ostell, J.; Sayers, E. W. GenBank. *Nucleic Acids Res.* **2012**, *40* (D1), D48−D53.

(40) Guo, Y.; Yu, L.; Wen, Z.; Li, M. Using support vector machine combined with auto covariance to predict protein-protein interactions from protein sequences. *Nucleic Acids Res.* **2008**, *36* (9), 3025−30.

(41) Chang, C. C.; Lin, C. J. LIBSVM: A Library for Support Vector Machines. *Acm Transactions on Intelligent Systems and Technology* **2011**, *2* (3), 1.

(42) Liu, X. P.; Yang, W. C.; Gao, Q.; Regnier, F. Toward chromatographic analysis of interacting protein networks. *Journal of Chromatography A* **2008**, *1178* (1−2), 24−32.

(43) Gao, Q.; Madian, A. G.; Liu, X.; Adamec, J.; Regnier, F. E. Coupling protein complex analysis to peptide based proteomics. *J. Chromatogr A* **2010**, *1217* (49), 7661−8.

(44) Tucker, D. L.; Sherman, L. A. Analysis of chlorophyll-protein complexes from the cyanobacterium Cyanothece sp. ATCC 51142 by non-denaturing gel electrophoresis. *Biochim. Biophys. Acta, Biomembr.* **2000**, *1468* (1−2), 150−60.

(45) Pancholi, V. Multifunctional alpha-enolase: its role in diseases. *Cell. Mol. Life Sci.* **2001**, *58* (7), 902−20.

(46) O'Leary, B.; Rao, S. K.; Kim, J.; Plaxton, W. C. Bacterial-type phosphoenolpyruvate carboxylase (PEPC) functions as a catalytic and regulatory subunit of the novel class-2 PEPC complex of vascular plants. *J. Biol. Chem.* **2009**, *284* (37), 24797−805.

(47) Dhanyalakshmi, K. H.; Naika, M. B.; Sajeevan, R. S.; Mathew, O. K.; Shafi, K. M.; Sowdhamini, R.; Nataraja, K. N. An Approach to Function Annotation for Proteins of Unknown Function (PUFs) in the Transcriptome of Indian Mulberry. *PLoS One* **2016**, *11* (3), e0151323.

(48) Rust, M. J.; Golden, S. S.; O'Shea, E. K. Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator. *Science* **2011**, *331* (6014), 220−3.

(49) Guo, Y.; Yu, L.; Wen, Z.; Li, M. Using support vector machine combined with auto covariance to predict protein−protein interactions from protein sequences. *Nucleic Acids Res.* **2008**, *36* (9), 3025−3030.

(50) Smoot, M. E.; Ono, K.; Ruscheinski, J.; Wang, P.-L.; Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **2011**, *27* (3), 431−432.

(51) Szklarczyk, D.; Morris, J. H.; Cook, H.; Kuhn, M.; Wyder, S.; Simonovic, M.; Santos, A.; Doncheva, N. T.; Roth, A.; Bork, P.; Jensen, L. J.; von Mering, C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **2017**, *45* (D1), D362−D368.

(52) Tamames, J.; Casari, G.; Ouzounis, C.; Valencia, A. Conserved clusters of functionally related genes in two bacterial genomes. *J. Mol. Evol.* **1997**, *44* (1), 66−73.

(53) Dandekar, T.; Snel, B.; Huynen, M.; Bork, P. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* **1998**, *23* (9), 324−328.

(54) Kosuge, T.; Hoshino, T. Construction of a proline-producing mutant of the extremely thermophilic eubacterium Thermus thermophilus HB27. *Appl. Environ. Microbiol.* **1998**, *64* (11), 4328−4332.

(55) Liu, L. N.; Chen, X. L.; Zhang, Y. Z.; Zhou, B. C. Characterization, structure and function of linker polypeptides in phycobilisomes of cyanobacteria and red algae: an overview. *Biochim. Biophys. Acta, Bioenerg.* **2005**, *1708* (2), 133−42.

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

(56) Gantt, E.; Lipschultz, C. A. Phycobilisomes of Porphyridium cruentum. I. Isolation. *J. Cell Biol.* **1972**, *54* (2), 313−324.
(57) Zhang, J.; Ma, J.; Liu, D.; Qin, S.; Sun, S.; Zhao, J.; Sui, S. F. Structure of phycobilisome from the red alga Griffithsia pacifica. *Nature* **2017**, *551* (7678), 57−63.
(58) Savage, D. F.; Afonso, B.; Chen, A. H.; Silver, P. A. Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science* **2010**, *327* (5970), 1258−61.
(59) Preiss, J. Bacterial glycogen synthesis and its regulation. *Annu. Rev. Microbiol.* **1984**, *38*, 419−58.
(60) Zak, E.; Pakrasi, H. B. The BtpA protein stabilizes the reaction center proteins of photosystem I in the cyanobacterium Synechocystis sp. PCC 6803 at low temperature. *Plant Physiol.* **2000**, *123* (1), 215−22.
(61) Umena, Y.; Kawakami, K.; Shen, J. R.; Kamiya, N. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 A. *Nature* **2011**, *473* (7345), 55−60.
(62) Wegener, K. M.; Nagarajan, A.; Pakrasi, H. B. An atypical psbA gene encodes a sentinel D1 protein to form a physiologically relevant inactive photosystem II complex in cyanobacteria. *J. Biol. Chem.* **2015**, *290* (6), 3764−74.
(63) Zhang, X.; Sherman, L. A. Alternate copies of D1 are used by cyanobacteria under different environmental conditions. *Photosynth. Res.* **2012**, *114* (2), 133−5.
(64) Battchikova, N.; Aro, E. M. Cyanobacterial NDH-1 complexes: multiplicity in function and subunit composition. *Physiol. Plant.* **2007**, *131* (1), 22−32.
(65) Battchikova, N.; Eisenhut, M.; Aro, E. M. Cyanobacterial NDH-1 complexes: novel insights and remaining puzzles. *Biochim. Biophys. Acta, Bioenerg.* **2011**, *1807* (8), 935−44.
(66) Chen, X.; He, Z.; Xu, M.; Peng, L.; Mi, H. NdhV subunit regulates the activity of type-1 NAD(P)H dehydrogenase under high light conditions in cyanobacterium Synechocystis sp. PCC 6803. *Sci. Rep.* **2016**, *6*, 28361.
(67) Schwarz, D.; Schubert, H.; Georg, J.; Hess, W. R.; Hagemann, M. The gene sml0013 of Synechocystis species strain PCC 6803 encodes for a novel subunit of the NAD(P)H oxidoreductase or complex I that is ubiquitously distributed among Cyanobacteria. *Plant Physiol.* **2013**, *163* (3), 1191−202.
(68) Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Muzio, L. L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int. J. Mol. Med.* **2017**, *40* (2), 271−280.
(69) Chen, Z.; Zhan, J.; Chen, Y.; Yang, M.; He, C.; Ge, F.; Wang, Q. Effects of Phosphorylation of beta Subunits of Phycocyanins on State Transition in the Model Cyanobacterium Synechocystis sp. PCC 6803. *Plant Cell Physiol.* **2015**, *56* (10), 1997−2013.
(70) Zhang, C. C.; Jang, J.; Sakr, S.; Wang, L. Protein phosphorylation on Ser, Thr and Tyr residues in cyanobacteria. *J. Mol. Microbiol Biotechnol* **2006**, *9* (3−4), 154−66.
(71) Perez, J.; Castaneda-Garcia, A.; Jenke-Kodama, H.; Muller, R.; Munoz-Dorado, J. Eukaryotic-like protein kinases in the prokaryotes and the myxobacterial kinome. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (41), 15950−5.
(72) Brandina, I.; Graham, J.; Lemaitre-Guillier, C.; Entelis, N.; Krasheninnikov, I.; Sweetlove, L.; Tarassov, I.; Martin, R. P. Enolase takes part in a macromolecular complex associated to mitochondria in yeast. *Biochim. Biophys. Acta, Bioenerg.* **2006**, *1757* (9−10), 1217−28.
(73) Gavin, A. C.; Bosche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A. M.; Cruciat, C. M.; Remor, M.; Hofert, C.; Schelder, M.; Brajenovic, M.; Ruffner, H.; Merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.; Gnau, V.; Bauch, A.; Bastuck, S.; Huhse, B.; Leutwein, C.; Heurtier, M. A.; Copley, R. R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G.; Superti-Furga, G. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **2002**, *415* (6868), 141.
(74) Peterson, L. X.; Roy, A.; Christoffer, C.; Terashi, G.; Kihara, D. Modeling disordered protein interactions from biophysical principles. *PLoS Comput. Biol.* **2017**, *13* (4), e1005485.
(75) Esquivel-Rodriguez, J.; Yang, Y. D.; Kihara, D. Multi-LZerD: multiple protein docking for asymmetric complexes. *Proteins: Struct., Funct., Genet.* **2012**, *80* (7), 1818−1833.
(76) Venkatraman, V.; Yang, Y. D.; Sael, L.; Kihara, D. Protein-protein docking using region-based 3D Zernike descriptors. *BMC Bioinf.* **2009**, *10*, 407.

DOI: 10.1021/acs.jproteome.8b00170
J. Proteome Res. XXXX, XXX, XXX−XXX