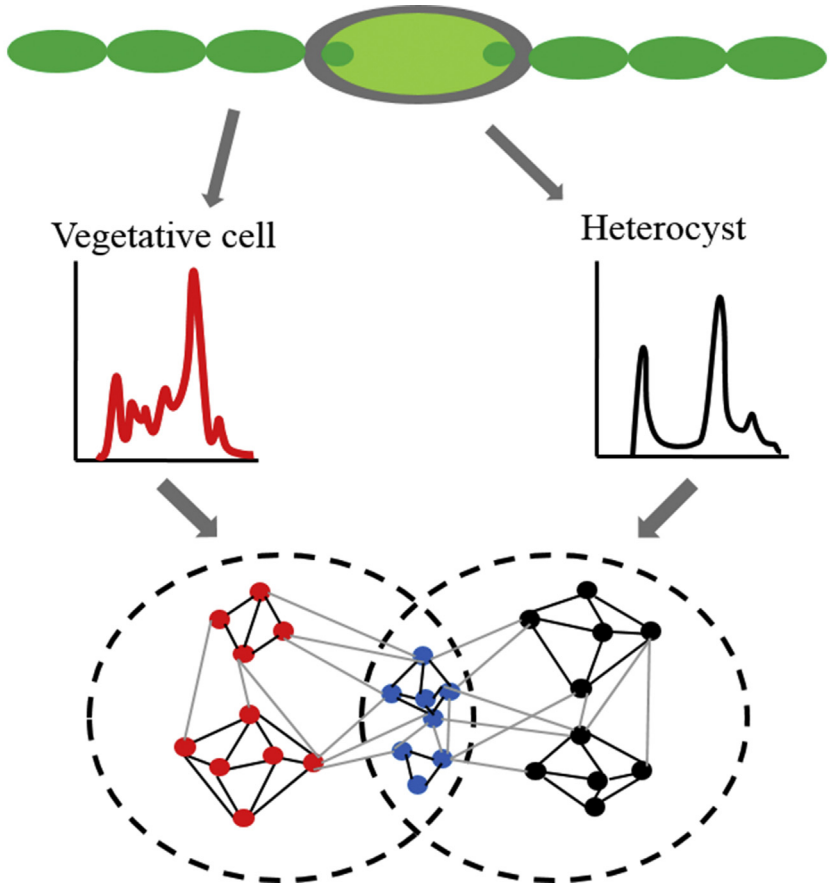
Comparative Network Biology Discovers Protein Complexes That Underline Cellular Differentiation in *Anabaena* sp.

Authors
Chen Xu, Bing Wang, Hailu Heng, Jiangmei Huang, and Cuihong Wan

Correspondence
ch_wan@ccnu.edu.cn

In Brief
We reported highly confident protein pairs in vegetative cells and heterocysts of *Anabaena* sp. PCC 7120, which is the first time such a large protein interaction dataset of *Anabaena* was generated. Comparing the protein network of the two types of cells can expand our understanding of cell differentiation. Meanwhile, protein–protein interaction data reveal new functions of proteins. We found that the hypothetical protein Alr4359 interacted with FraH and Alr4119 in heterocysts and influenced the diazotrophic growth of filaments.

Graphical Abstract



Highlights

- PPIs in two types of cells of *Anabaena* sp. 7120 were systematically identified.
- 10,302 and 8557 high-confidence PPIs were obtained and over 80% were novel.
- About 438 proteins showed significant changes in vegetative cells and heterocysts.
- Protein Alr4359 was found to influence the diazotrophic growth of filaments.

2022, Mol Cell Proteomics *21*(*4*), 100224

https://doi.org/10.1016/j.mcpro.2022.100224

Comparative Network Biology Discovers Protein Complexes That Underline Cellular Differentiation in *Anabaena* sp.

Chen Xu, Bing Wang, Hailu Heng, Jiangmei Huang, and Cuihong Wan*

The filamentous cyanobacterium *Anabaena* sp. PCC 7120 can differentiate into heterocysts to fix atmospheric nitrogen. During cell differentiation, cellular morphology and gene expression undergo a series of significant changes. To uncover the mechanisms responsible for these alterations, we built protein–protein interaction (PPI) networks for these two cell types by cofractionation coupled with mass spectrometry. We predicted 280 and 215 protein complexes, with 6322 and 2791 high-confidence PPIs in vegetative cells and heterocysts, respectively. Most of the proteins in both types of cells presented similar elution profiles, whereas the elution peaks of 438 proteins showed significant changes. We observed that some well-known complexes recruited new members in heterocysts, such as ribosomes, diflavin flavoprotein, and cytochrome *c* oxidase. Photosynthetic complexes, including photosystem I, photosystem II, and phycobilisome, remained in both vegetative cells and heterocysts for electron transfer and energy generation. Besides that, PPI data also reveal new functions of proteins. For example, the hypothetical protein Alr4359 was found to interact with FraH and Alr4119 in heterocysts and was located on heterocyst poles, thereby influencing the diazotrophic growth of filaments. The overexpression of Alr4359 suspended heterocyst formation and altered the pigment composition and filament length. This work demonstrates the differences in protein assemblies and provides insight into physiological regulation during cell differentiation.

The filamentous cyanobacterium *Anabaena* sp. PCC 7120 (hereafter *Anabaena* sp.) is a model organism for studying cell differentiation, cell–cell adhesion, and intercellular communication. In the presence of combined nitrogen, vegetative cells exhibit phenotype of the filaments and fix CO_2 through oxygenic photosynthesis. In the absence of combined nitrogen, some vegetative cells differentiate into heterocysts for nitrogen fixation (1).

In order to fix nitrogen, heterocysts need an efficient nitrogenase enzyme, which is extremely sensitive to oxygen. Therefore, heterocysts have evolved different strategies to prevent oxygen from the environments interfering with intracellular reactions (2). Polysaccharide and glycolipid layers are formed around the heterocysts to prevent gas diffusion, making them appear larger than vegetative cells under a light microscope (3). Inside cells, photosystem I (PSI) and photosystem II (PSII) change differently in terms of protein expression, and no water oxidation from PSII is detectable (4). The inner membrane structure has also been shown to form a new inner membrane close to the heterocyst poles, called "honeycomb," during cell differentiation (5). The protein FraH has been reported to influence honeycomb formation during heterocyst differentiation (6). However, the mechanisms responsible for thylakoid membrane reorganization and whether additional proteins participate in this process require further investigation.

In the diazotrophic filament, vegetative cells provide heterocysts with a carbon source in the form of sucrose, and heterocysts provide vegetative cells with combined nitrogen in the forms of glutamine and β-aspartyl-arginine (7, 8). Intercellular metabolite exchange is performed by two routes, *via* continuous periplasm or by diffusion through the septal junctions involving the septal proteins (9). The septal junctions contain the proteins SepJ, FraC, and FraD. The structure of the septal junction was recently recovered by cryo-EM as containing a cap, a plug, and tube modules, and it was found to undergo reversibly controlled material communication under stress (10). Through an analysis of cellular localization and protein–protein interactions (PPIs), it is found that the SepJ-related and FraCD-related septal junctions probably contain additional proteins (11).

Generally, intracellular biological processes rely on a series of physical associations among molecular substances, especially proteins. The PPI network is a powerful tool for exploring the fundamental metabolism of living organisms (12). The PPI network may also help us better understand the difference between vegetative cells and heterocysts and find more

From the School of Life Sciences and Hubei Key Laboratory of Genetic Regulation and Integrative Biology, Central China Normal University, Wuhan, Hubei, China

*For correspondence: Cuihong Wan, ch_wan@ccnu.edu.cn.

https://doi.org/10.1016/j.mcpro.2022.100224

functional proteins related to cell differentiation. Several approaches have been developed for the identification of PPIs at the proteome scale (13), such as yeast two-hybrid (Y2H), affinity purification followed by mass spectrometry (AP–MS), and cofractionation coupled with mass spectrometry (CoFrac–MS). The similarity of protein's elution profiles was the principal character in the CoFrac–MS experience. Different computational approaches were applied to distinguish elution profiles and generate the predicted complexes, such as hierarchical clustering and machine learning (14, 15). CoFrac–MS assay has been broadly applied to detect stable protein complexes in various organisms from prokaryotes to eukaryotes, including Trypanosoma, cyanobacterium, plants, and humans (16–21). Integrated CoFrac–MS datasets can build comprehensive protein complex datasets in each organism and help elucidate complex remodeling or evolution across different species (21–23).

In this study, we constructed PPIs of vegetative cells and heterocysts in *Anabaena* sp. using CoFrac–MS. We compared the protein interaction networks to provide a better understanding of cell differentiation between vegetative cells and heterocysts. Furthermore, we found the redistribution of the hypothetical protein Alr4359 and its function in heterocyst formation.

EXPERIMENTAL PROCEDURES

Cell Culture and Heterocyst Purification

The *Anabaena* sp. PCC 7120 was cultured in liquid BG11 (with combined nitrogen) or $BG11_0$ medium (without combined nitrogen) under continuous illumination of 30 to 40 μmol m^{-2} s^{-1} at 28 °C. To induce heterocyst, cyanobacterial cells were harvested at the exponential growth phase in BG11 and then incubated in $BG11_0$ medium for 48 h after being washed twice using $BG11_0$ medium. To purify the heterocysts, the filaments containing heterocysts were harvested by centrifugation (3000*g* at 4 °C for 5 min) after 48 h of induction in $BG11_0$ medium. The isolation process was based on previous studies (24). The pellet was resuspended in 8% sucrose, 5% Triton X-100, 50 mM EDTA, pH = 8.0, 50 mM Tris–HCl, and pH = 8.0 containing lysozyme (1 mg/ml) at 4 °C and vortexed vigorously for 2 to 3 min at room temperature. The suspension was mildly sonicated for approximately 2 min on ice to break the vegetative cells, whereas the heterocyst remained intact during this procedure. The heterocysts were collected by centrifugation at 3000*g* for 5 min at 4 °C and washed twice in 8% sucrose, 50 mM EDTA, pH = 8.0, 50 mM Tris–HCl, and pH = 8.0 at 4 °C. The isolated heterocysts were assessed by microscopy, and contamination by vegetative cell contents was tested by measuring the concentration of ribulose-1,5-bisphophate carboxylase/oxygenase.

Protein Extraction and Separation

Cells from vegetative cells or heterocysts were suspended in lysis buffer containing 20 mM Tris-Cl (pH = 7.5), 150 mM NaCl, 1% dodecylmaltoside, and Complete Protease Inhibitors EDTA-free (Roche), and sonicated on ice with an output of 135 W. The whole-cell lysate was centrifuged (10,000*g* at 4 °C for 10 min) to remove the cell debris. The protein concentration was determined using the Bradford assay. Then, 300 μg of protein mixtures of vegetative cells and heterocysts were individually fractionated by size-exclusion chromatography (SEC) using a Thermo Scientific Ultimate 3000 HPLC system. The lysates separated by SEC were injected into a Superose 6 10/300GL column (GE Life Sciences) equilibrated with PBS (pH = 7.2) and exposed to 120 min of isocratic elution. The total collection time was 55 min, as the first fraction was collected at 20 min and the last fraction finished at 75 min. In total, 55,300 μl fractions were collected, with a flow rate of 0.3 ml min^{-1}.

Protein Digestion and Desalting

The proteins from all the HPLC fractions were denatured at 95 °C for 10 min. For the LC–MS/MS analysis, each fraction was reduced with 10 mM freshly prepared DTT at 37 °C for 45 min, and the cysteines were alkylated with 15 mM iodoacetamide at 37 °C in the dark for 30 min. Trypsin (Promega) was added, and the samples were incubated overnight at 37 °C. Each fraction was desalted using ZipTip C18 plates (Millipore) following the manufacturer's protocol. Peptides were dried using a LABCONCO evaporate and resuspended in 0.1% formic acid.

LC–MS/MS Analysis and Data Processing

The peptides were analyzed by online nanoflow LC–MS/MS using an Easy-nLC 1200 system connected to a Q-Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). Peptides in 0.1% formic acid were injected onto a C18 column (75 μm × 15 cm, 3 μm, 100 Å) and eluted at a flow rate of 300 nl/min with a 100 min gradient from 10% solvent B (90% acetonitrile/0.1% formic acid, v/v) to 80% solvent B. The peptides were ionized by nanoelectrospray at 2.0 kV and analyzed with higher-energy collisional dissociation fragmentation. The MS/MS spectra of the top 20 most-abundant precursor ions were acquired using a data-dependent method. The dynamic exclusion duration was 40 s with a repeat count of 1 and a ±10 ppm exclusion window. Automatic gain control was used to prevent the overfilling of the ion trap, and 5×10^4 ions were accumulated for the generation of MS/MS spectra.

The RAW data files were analyzed using Proteome Discoverer 2.1 (Thermo Fisher Scientific). The database containing 6175 entries supplied with the *Anabaena* sp. PCC 7120 proteome (UP000002483) was from the UniProt database. The mass tolerance of the precursor ions was set to 10 ppm, and the MS/MS mass tolerance was set at 0.02 Da. The enzyme was set as trypsin, allowing up to two missed cleavages. The carbamidomethyl modification of cysteines was set as a fixed modification, and methionine oxidation and protein N-terminal acetylation were set as variable modifications. The false discovery rate for protein-level and peptide-level identification was set at 1%, using a target-decoy–based strategy.

Gene Ontology Enrichment Analysis

The annotations of *Anabaena* sp. were taken from the Gene Ontology (GO) Annotation Database (https://www.ebi.ac.uk/GOA/). The annotation enrichment was calculated using the ClusterProfiler software (http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) using the built-in "enricher" module, with the *p* value cutoff and *q* value cutoff set as 0.05.

Machine Learning

Machine learning was performed using EPIC (elution profile–based inference of complexEPIC) following the manufacturer's instructions. The command line is: python/EPIC/src/main.py -s 1110011 -c ../complexes.txt. -r 0.75 -R 0.5 -e 3 -E 5 ../InputFolder/ ../OutputFolder/. The interpretation of each parameter can be found at https://github.com/BaderLab/EPIC. The standard true-positive protein complexes (supplemental Table S3) were manually collected using the

experimental data in the STRING database. The input data are mass spectrometry (MS) results in supplemental Table S1. We created candidate protein-pair datasets in which only the protein pairs with at least one elution profile similarity score greater than 0.75 were retained. Then finally, PPIs (supplemental Table S4) were generated from these candidate protein-pair datasets using the random-forest machine-learning classifier. ClusterONE served as the clustering algorithm to predict the final protein complexes (supplemental Table S5). The visualization of protein complexes was built by Cytoscape 3.8 (https://cytoscape.org/).

Construction of Overexpression Strains

The full-length target genes were amplified by PCR using *Anabaena* sp. chromosome DNA as the template, and they were cloned into the plasmid pRL25N at the SmaI site, which contains the Cu^{2+}-inducible promotor *petE* and GFP ORFs. The plasmid pRL25N was donated by Professor Xiang Gao. The resultant plasmids were validated by PCR and sequencing analysis and then introduced into the *Anabaena* sp. *via* conjugative transformation from an HB101 carrying the cargo plasmid and the helper plasmid.

The primers used in this study are as follow:
pRL25N-up TACTGAGTACACAGCTAATAAAATTG;
pRL25N-down TCATATGATCTGGGTATCTCGCAAAG;
all1475-up ATGTTTGAAATCTTTAAAAATCG;
all1475-down GTGAGAAGAAGTTGCTGTAGTG;
all7197-up ATGAATAGAAAAAGCAACTCGG;
all7197-down ATATAACTTCTTGCTAGTTAAATATCC;
alr4359-up ATGAAAGTTAATTTGCAGCCTGTCCTCAATG;
alr4359-down TTTAGGAGGGGTATCTTGCAAGACGGTTTTG;
fraH-up ATGATCGTCTGTCCAAATTGCAACC;
fraH-down AGCGAGTTTAAAGAGGAAAGTTACC.

AP–MS

The cell lysates of the GFP-tag strains were subjected to affinity purification using an anti-GFP antibody (catalog no.: ab290; Abcam). The antibody purification was performed using Protein A MagBeads following the manufacturer's instructions (GenScript). In brief, 2 μl GFP antibody was first incubated with 100 μl magnetic beads for 30 min. The magnetic beads were washed by 1 ml binding buffer (20 mM Na_2HPO_4, 0.15 M NaCl, pH = 7.0) twice to remove the free antibody. The cell lysate that contained about 500 μg proteins was mixed with magnetic beads binding with GFP antibody and was incubated for 1 h at room temperature. After removing the dissociated proteins from the solution and washing three times with PBS, the target protein and its interaction partners were released using 0.1 M glycine solution (pH = 2–3). The neutralization buffer (1 M Tris, pH = 8.5) was added to each eluate to neutralize the pH at the ratio of 1:10. Finally, the sample was denatured at 95 °C for 10 min, digested with 0.3 μg trypsin, and analyzed by MS. The LC/MS/MS and database search were the same as the fractionation samples mentioned previously, except that the phosphorylation modifications of serine, threonine, and tyrosine were set as variable modifications. The SAINT software (https://sourceforge.net/projects/saint-apms/files/) analyzed all the results with default settings, and the command line is SAINTexpress-spc inter.dat prey.dat bait.dat.

Yeast Two-Hybrid Assays

Full-length *fraH* and *alr4359* were amplified from *Anabaena*'s genome by PCR with the NdeI restriction site at their N-terminal region, the XhoI restriction site at the *fraH* C-terminal region, and the PstI restriction site at the *alr4359* C-terminal region. Then, the gene fragments of *fraH* and *alr4359* were cloned into the AD-vector pGADT7 and the BD-vector pGBKT7 to generate pGADT7-*fraH* and pGBKT7-*alr4359*, respectively. The plasmids of pGADT7-T and pGBKT7-53 were used as positive controls. pGADT7-T and pGBKT7-lam were negative controls. The recombinant plasmids were cotransformed into *Saccharomyces cerevisiae* AH109 and grown on Synthetic Dropout Medium/-Trp-Leu-His agar plates at 28 °C for 3 days. This experiment was repeated twice. The primers used in this experiment are as follows:

AD-*frah*-up: CTTCATATGATCGTCTGTCCAAATTGCAACC;
AD-*frah*-down: CTTCTCGAGTTAAGCGAGTTTAAAGAGGAAAG;
BD-*alr4359*-up: CTTCATATGAAAGTTAATTTGCAGCCTGTCCT;
BD-*alr4359*-down: CTTCTGCAGTCATTTAGGAGGGGTATCTTGC.

Confocal Microscopy and Electron Microscopy

Anabaena cells were visualized with a Leica SP8 confocal microscope. GFP was excited using 488 nm laser irradiation. The fluorescent emission was monitored by collection across windows of 498 to 541 nm for GFP imaging and 630 to 700 nm for cyanobacterial autofluorescence. For electron microscopy imaging, the Alr4359-GFP overexpression strain cultured in BG11 was harvested, washed twice with fresh BG11, and prepared by the method presented by Merino-Puerto *et al.* (6). The samples were examined with a Hitachi HT-7700 electron microscope at 120 kV.

Experimental Design and Statistical Rationale

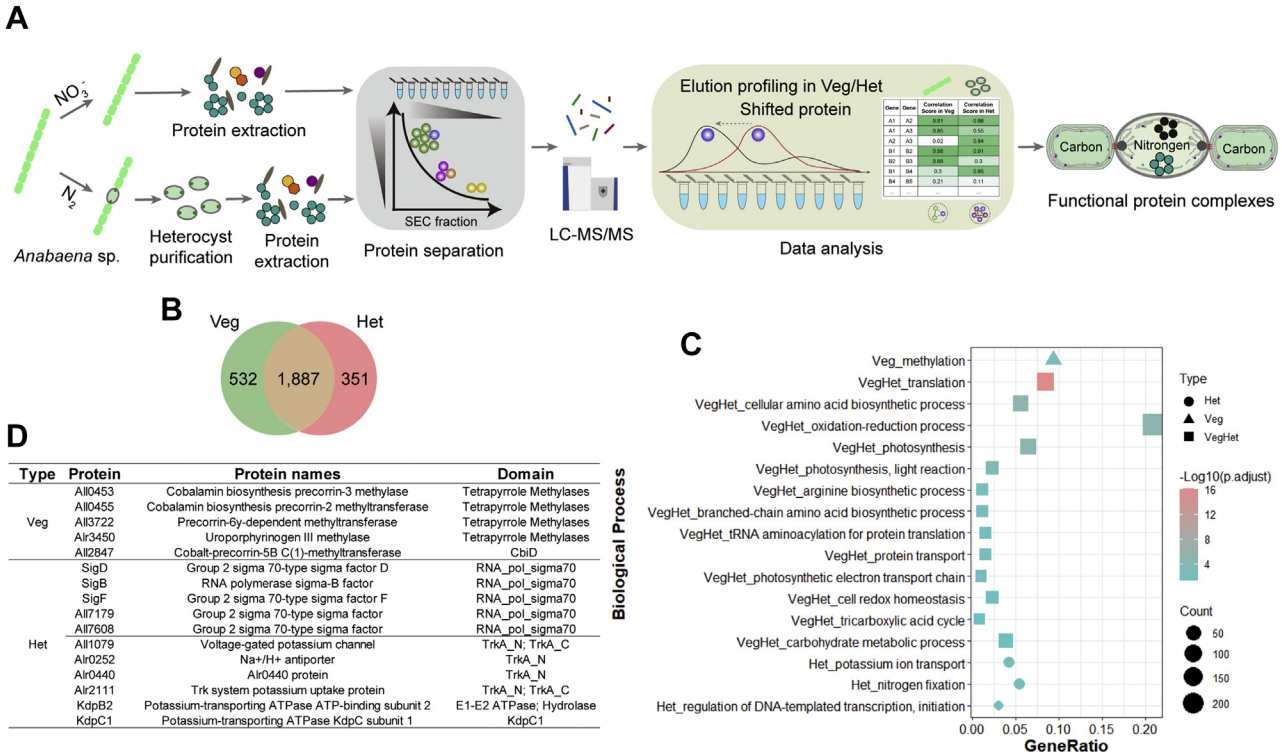
Three independent SEC separation experiments were conducted in vegetative cells and heterocyst. In each experiment, 300 μg of protein mixtures were injected into the SEC column, and a total of 55 fractions were collected for one fraction per minute. The SEC reproducibility of two cell types was evaluated based on the Pearson correlation coefficient. The false discovery rate for protein-level and peptide-level identification was set at 1%. After LC/MS/MS analysis and database search, the results for each experiment were summarized into one matrix, where each row represents a protein and each column contains the spectral counts of proteins for the corresponding fraction. The peak shift of proteins was determined by fold change >1.5 with *t* test p value <0.05, which reflected shifting of average elution peak position in heterocyst to vegetative in three replicates. The protein elution profiles based on spectra counts were put together and run in EPIC software for protein complex prediction. The final protein complex datasets were generated by combining three independent experiments based on the in-build algorithm.

For AP–MS experiments, three independent replicates for one bait protein were performed in vegetative cells and heterocyst. The empty vector served as a negative control was also treated three times in each cell type. The prey proteins in vegetative cell's experiment or heterocyst's experiment were analyzed by SAINT software, which prey proteins with SAINTscore more than 0.6 were reserved and used for further analysis

RESULTS

Identification of Proteins in Vegetative Cells and Heterocysts

The filaments of *Anabaena* sp. consist of vegetative cells, which specialize into heterocysts with no red spontaneous fluorescent signals in the absence of combined nitrogen (supplemental Fig. S1). Heterocysts were purified from the filaments without vegetative cell contamination, as verified by fluorescence imaging (supplemental Fig. S2*A*) and Western immunoblotting (supplemental Fig. S2*B*). Cellular extracts of vegetative cells or heterocysts containing native proteins and protein complexes were fractionated by SEC and quantified



D

Type	Protein	Protein names	Domain
Veg	All0453	Cobalamin biosynthesis precorrin-3 methylase	Tetrapyrrole Methylases
	All0455	Cobalamin biosynthesis precorrin-2 methyltransferase	Tetrapyrrole Methylases
	All3722	Precorrin-6y-dependent methyltransferase	Tetrapyrrole Methylases
	Alr3450	Uroporphyrinogen III methylase	Tetrapyrrole Methylases
	All2847	Cobalt-precorrin-5B C(1)-methyltransferase	CbiD
Het	SigD	Group 2 sigma 70-type sigma factor D	RNA_pol_sigma70
	SigB	RNA polymerase sigma-B factor	RNA_pol_sigma70
	SigF	Group 2 sigma 70-type sigma factor F	RNA_pol_sigma70
	All7179	Group 2 sigma 70-type sigma factor	RNA_pol_sigma70
	All7608	Group 2 sigma 70-type sigma factor	RNA_pol_sigma70
	All1079	Voltage-gated potassium channel	TrkA_N; TrkA_C
	Alr0252	Na+/H+ antiporter	TrkA_N
	Alr0440	Alr0440 protein	TrkA_N
	Alr2111	Trk system potassium uptake protein	TrkA_N; TrkA_C
	KdpB2	Potassium-transporting ATPase ATP-binding subunit 2	E1-E2 ATPase; Hydrolase
	KdpC1	Potassium-transporting ATPase KdpC subunit 1	KdpC1

FIG. 1. **Characteristic analysis of proteins in vegetative cells and heterocysts.** *A*, experimental workflow used for protein coelution profiling. The cell-lysis material from vegetative cells (Veg) or heterocysts (Het) was separated by SEC and analyzed by LC–MS/MS. The proteins with significantly different elution profiling between the two types of cells were selected for subsequent analysis. *B*, the overlap of total identified proteins between Veg and Het. *C*, Gene Ontology analysis of proteins specifically identified in Veg and Het or both using the ClusterProfiler software. GeneRatio represents the ratio of candidate genes compared with whole genes in each catalog. *D*, the proteins related to enriched GO function in *C*, including methylation, transcription, and potassium-ion transport. GO, gene Ontology; SEC, size-exclusion chromatography.

by LC–MS/MS (Fig. 1*A*). Most of the fractions had a decent reproducibility between biological replicates in two cell types with Pearson correlation coefficients over 0.85 (supplemental Fig. S3, *A* and *B*). The correlation coefficients of fractions between two types of cells were slightly lower than in the same cell type (supplemental Fig. S3*C*). Besides, the correlation coefficients of each protein in two replicates were also calculated (supplemental Fig. S4). High abundant proteins presented higher correlation coefficients in both types of cells.

Finally, a total of 2770 *Anabaena* sp. proteins were identified in both cell types with the peptide-spectrum matches ≥2. Of these, 532 proteins were specific to vegetative cells, and 351 proteins were specific to heterocysts (Fig. 1*B* and supplemental Table S1). We analyzed the over-representation in the GO annotations of these proteins. For the proteins identified in the vegetative cells, only one biological process—methylation—was enriched, with 21 proteins in this term (Fig. 1*C*). We note that five of these 21 proteins contained a predicted tetrapyrrole methylase domain that was able to catalyze the methylation of different porphyrin compounds and participate in chlorophyll metabolism in cyanobacteria (Fig. 1*D*). These results indicate that protein methylation plays an important role in regulating the pigment composition in vegetative cells. The nitrogen fixation process was significantly enriched for the 351 proteins identified in heterocyst (Fig. 1*C*). The biological process of DNA-templated transcription regulation was also found in heterocysts. By analyzing the protein compositions associated with the term, we found five proteins belonging to group 2 sigma factor of RNA polymerase, containing the RNA_pol_sigma70 domain (Fig. 1*D*). The group 2 sigma factors are known as alternative sigma factors, and it has been reported that they are required for normal growth under nitrogen stress (25–28).

In addition, we observed that high-affinity potassium ion transport proteins, such as the Ktr/Trk-type K^+ uptake transporter families (Ktr/Trk) and ATP-dependent transporters specific for K^+ families (Kdp), were activated in heterocysts (Fig. 1, *C* and *D*). Potassium ions are the dominant intracellular cations and play essential roles in turgor homeostasis and pH regulation (29). Potassium deficiency can cause multiple metabolic impairments and influence photosynthetic functions and nitrogenase activity in cyanobacteria (30). The Kdp family is considered a high-affinity system and is usually activated when the level of K^+ is low and cannot be maintained by other constitutive systems, such as the Ktr/Trk families (31). The extra expression of K^+ uptake transporters indicated a

massive demand for K^+ in heterocysts, but the biological function associated with this has not been elucidated.

Comparison of Protein Elution Profiles Between Vegetative Cells and Heterocysts

To thoroughly compare the elution profiles of the proteins in the two cell types, we performed an unsupervised hierarchical clustering analysis of the average protein intensities of relatively highly abundant proteins (peptide-spectrum matches ≥3) (Fig. 2*A*). Generally, most of the proteins in both cell types showed similar elution profiles, but some proteins showed a distinct peak shift. For example, most chaperone proteins or 30S ribosomal proteins were found in two types of cells with similar elution profiles, and the elution peak remained unchanged, but proteins such as DnaJ or uS10 (also known as RpsJ) (32) had a distinct peak shift (Fig. 2*B*).

We assume that proteins performing different functions in the two types of cells will show differences in protein interaction and chromatography behavior (14). To identify elements related to variations in function, we searched for protein inconsistencies between vegetative cells and heterocysts. After calculating the peak shifts, we detected 438 proteins that showed significant peak changes in two cell types (Fig. 2*C* and supplemental Table S2). GO analysis revealed that these proteins were mainly involved in transcription and translation (Fig. 2*D*). DNA-directed RNA polymerase complexes are

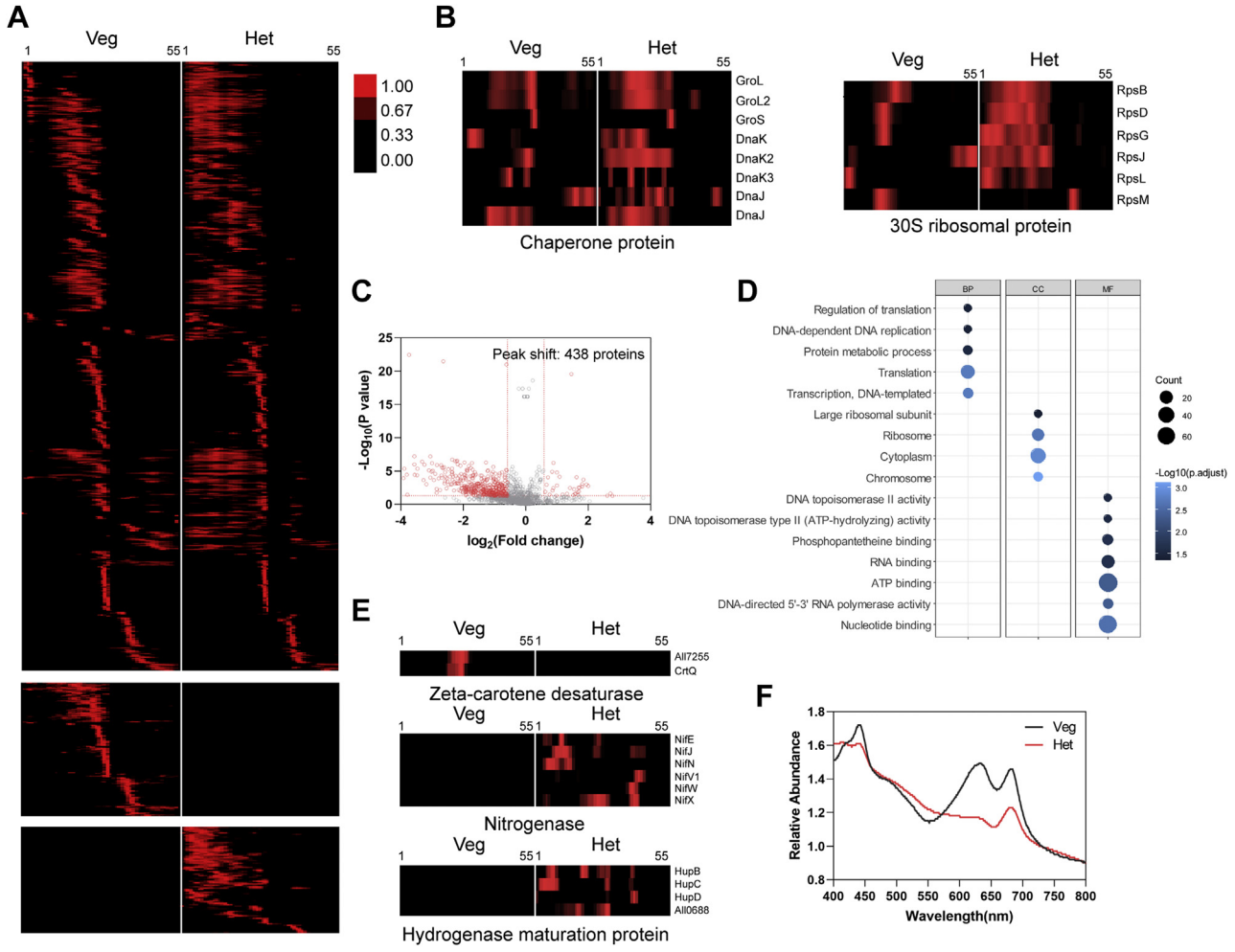


FIG. 2. **Classification of SEC elution-profile changes and relations to functional differences.** *A*, hierarchical clustering of protein elution profiles in SEC presented by Java TreeView 1.2. Each *row* represents a protein, and each *column* represents the protein-elution fraction index. *B*, detailed heatmap of Chaperone protein and 30S ribosomal protein. *C*, the distribution of shifted proteins. The plots represent the distribution of proteins with elution peaks showing significant shifts in their elution patterns. *t* test, *p* value ≤0.05; elution-peak fold change ≥1.5-fold. *D*, the Gene Ontology analysis of the whole set of shifting proteins suggests that activity occurred in processes related to cell differentiation. *E*, detailed heatmap of the proteins that were only found to be expressed in vegetative cells or heterocysts. *F*, the absorption spectra analysis of vegetative cell and isolated heterocyst measured by an ultraviolet spectrophotometer from 400 to 800 nm. The curves were normalized according to the absorption at 730 nm, and the change of absorption peak at 682 nm was checked by *t* test with the *p* value <0.01. Carotenoids, with an absorption peak at 495 nm; chlorophyll a, with two absorption peaks at 440 and 680 nm; and phycocyanobilin, with an absorption peak at 630 nm. SEC, size-exclusion chromatography.

stable and tightly linked in both vegetative cells and heterocysts (supplemental Fig. S5*A*). Their elution peaks shifted toward a higher molecular weight range, and thus they may bind with new components in heterocysts. For translation processes, the main peak-shifted proteins were a series of amino acid-tRNA ligases (supplemental Fig. S5*B*), such as AlaS, GlyS, and ValS. It suggests that the regulation of amino acid-tRNA ligases is necessary for the differences in protein synthesis between the two cell types.

In addition to proteins with different elution profiles between the two cell types, we also observed that certain protein complexes were only identified in either vegetative cells or heterocysts (Fig. 2*E*). In vegetative cells, two zeta-carotene desaturases, All7255 and CrtQ, were identified and presented similar elution profiles that regulated the carotene biosynthetic pathway (33) (Fig. 2*E*). The tetrapyrrole methylase and zeta-carotene desaturase in vegetative cells provide a reasonable explanation for the differences in the pigment composition between the two cell types (34–37). An analysis of the whole-cell absorption spectra of vegetative cells and heterocysts shown that the ratio of carotenoids to chlorophyll differed significantly, and the absorption peak of phycobilin almost disappeared in heterocysts (Fig. 2*F*). Nitrogenases and hydrogenase formation/maturation proteins were identified in heterocysts (Fig. 2*E*). Most of the components of the nitrogenase or Hup families had similar elution profiles, but several proteins, such as NifV1 and HupD, differed from others, indicating loose and weak bonds with the complex. The hydrogenase expression/formation/maturation proteins are usually required for the maturation of hydrogenase, a nickel metalloenzyme that catalyzes the reversible oxidation of molecular hydrogen (38). These results confirm that heterocysts require large amounts of energy to meet the requirements of the biological nitrogen-fixation process *via* various strategies (39).

Variation in Protein Complexes After Cell Differentiation

We predicted PPIs in two cell types using the EPIC software (40) based on the elution profile dataset. The training set of "gold standard" protein complexes was derived from experimental data in the STRING database, together with information in the literature (supplemental Table S3). In total, 10,302 and 8557 high-confidence protein pairs were obtained in vegetative cells and heterocysts, respectively. Among them, 1429 protein pairs containing 808 proteins were present in both types of cells (supplemental Fig. S6*A* and supplemental Table S4). These 808 proteins had a higher level of degree than the proteins that only existed in the protein interaction network of one type of cell (supplemental Fig. S6*B*). High-degree proteins can interact with many proteins and play an important role in the protein network (41).

Although most proteins and complexes are found in both cell types, some well-known protein complexes showed differences between vegetative cells and heterocysts (Fig. 3*A*). For example, the 50S and 30S ribosomes in heterocysts contain more heterocyst-specific PPIs and components such as uL2, uL23, uS2, uS5 (also known as RplB, RplW, RpsB, and RpsE), which are distinct peak-shifted proteins. This result is consistent with previous studies, that ribosome is susceptible to environmental fluctuations with variations in structure and components (42). We also observed that most photosynthetic complexes, including PSI, PSII, and phycobilisomes, might remain intact in both cell types, indicating that the photosystems have the ability to absorb light energy and perform electron transfer (43, 44). However, some components, such as PsaL in PSI, PsbA1 in PSII, and PecC in phycobilisomes, presented distinct elution peak shifts and participated in different binary interactions between two types of cells (supplemental Fig. S7 and Fig. 3*A*). In particular, most phycobilisome components remained after cell differentiation, and the elution peak had no obvious shift (supplemental Fig. S8). However, the absorption spectrum of phycocyanobilin declined dramatically at 630 nm in heterocysts (Fig. 2*F*). The light absorption of phycobiliprotein depends on phycobilin (34, 45), so we propose that phycocyanobilin may not covalently couple with phycobiliproteins, or its number may be reduced in heterocysts.

We also observed that other known complexes recruited new members in heterocysts, such as the diflavin flavoprotein and cytochrome *c* oxidase (Cox) (Fig. 3*A*). The transcriptions and quantitative proteomics results have shown that these complexes were dramatically upregulated, and some specific components were induced, leading to efficient electron transport and energy generation (46–48). The flavoproteins served as electron carriers by noncovalently binding with two cofactors, FAD or FMN, catalyzing the transfer of the reducing equivalents along the electron-transport chain (49, 50). The diflavin flavoprotein complex recruited more components to participate in the electron transfer process in heterocysts (Fig. 3*B*), meeting the enormous energy requirement. In addition, the elimination of oxygen is indispensable for the proper functioning of nitrogenase, and several mechanisms of elimination to prevent oxygen's detrimental effects in heterocysts have arisen (1). Cox serves as the terminal constituent of respiration by reducing oxygen to form water (51). The elution profile of Cox was similar in the two types of cells. However, CoxB shifted toward the higher molecular weight range (Fig. 3*C*), indicating the complex became larger in heterocysts.

Continuous Genomic Regions Tend to Be Induced and Expressed as Protein Complexes

The 10,302 and 8557 high-confidence protein pairs were clustered into 280 and 215 protein complexes of vegetative cells and heterocysts, respectively, which removed interactions among complexes and remained 6322 and 2791 PPIs (Fig. 4*A* and supplemental Table S5). Among these, we found well-known protein complexes, such as ribosomes and phycobilisomes. Except for conserved PPIs, nearly 88% of the

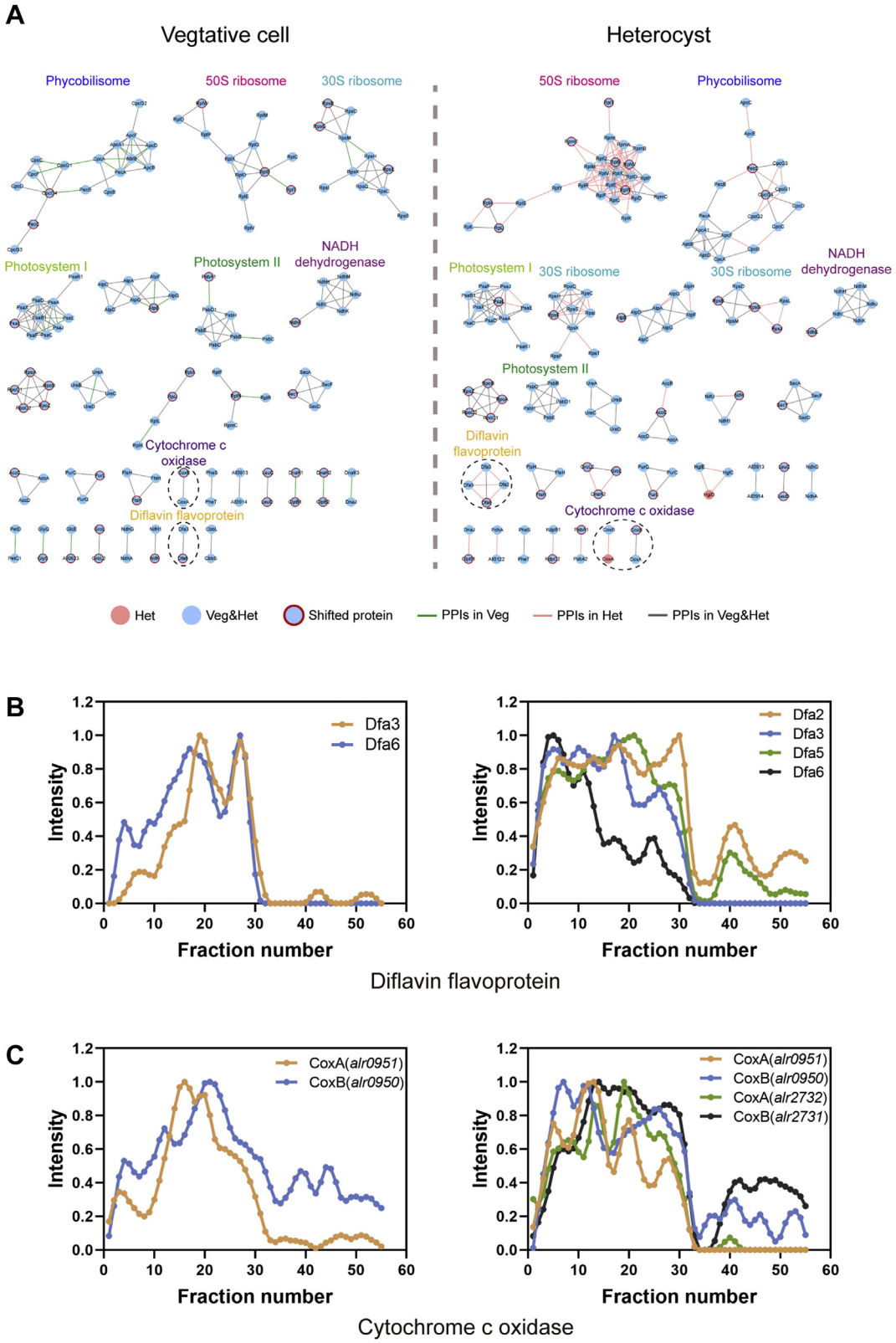


FIG. 3. **Variation of known complexes in vegetative cells and heterocysts.** *A*, the establishment of known complexes in vegetative cells and heterocysts. *B*, elution profiling of diflavin flavoprotein. *Left*, vegetative cell. *Right*, heterocyst. The *y*-axis is normalized to the LFQ intensity. The *x*-axis shows the elution fraction. *C*, elution profiling of cytochrome *c* oxidase. LFQ, label-free quantification.

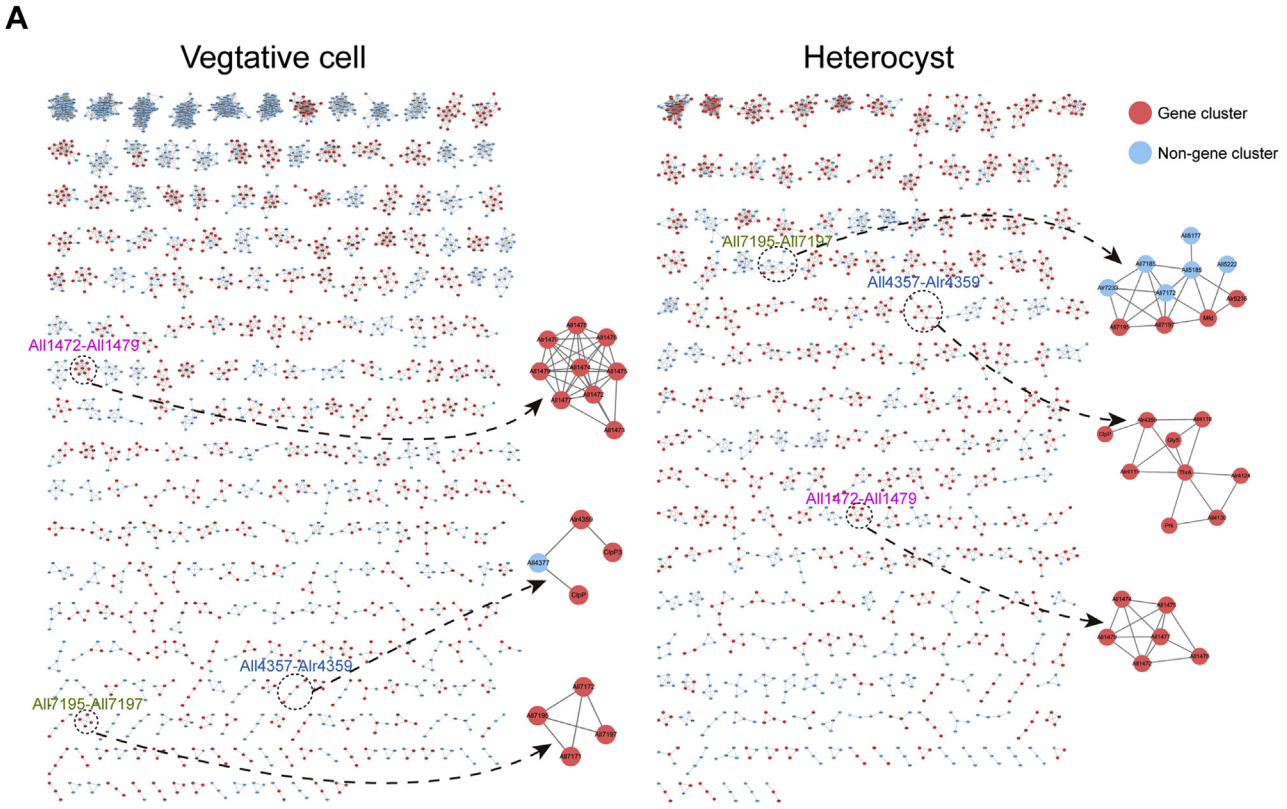


FIG. 4. **The characteristics of predicted protein complexes.** *A*, schematic diagram of the inferred *Anabaena* sp. protein complexes with representative examples. The proteins localized in the continuous genomic region are marked in *red*. *B*, the distribution of genes on chromosomes or plasmids. The genes are shown in order from *all0001* at the *top* to *alr9505* at the *bottom*. Clusters of genes found to participate in protein complexes are marked with *red lines*. *C*, the protein complexes with continuous gene clusters was confirmed by AP–MS. The bait proteins All1475, All7197, and Alr4359 were fused with the C-terminal GFP-tagged region and coimmunoprecipitated with a GFP antibody. AP–MS, affinity purification followed by mass spectrometry.

identified PPIs were novel. To our knowledge, this is the first study to obtain a large protein complex dataset for *Anabaena* sp.

By analyzing the annotations of the protein complexes, we found that most of the components in the predicted complexes were localized in continuous genomic regions of chromosomes or plasmids. It has been reported that gene transcription is physically clustered to form "expressed islands," and the structure of chromatin can change and be remodeled during heterocyst differentiation in *Anabaena* sp (47). We observed more continuous genomic regions in vegetative cells than in heterocysts (Fig. 4*B*). This difference

became apparent when observing the protein identified in the two types of cells. For example, the continuous DNA region, *alr3602–alr3071*, encodes a probable glycosyl transferase, which was only identified and clustered into a complex in vegetative cells. Furthermore, the DNA region, *alr5340–alr5355*, which encodes a glycolipid synthase, was only expressed in heterocysts. That protein complex is essential for heterocyst maturation (52, 53). We also found that some continuous DNA regions, such as the 50S ribosome, existed in both cell types, indicating conserved and indispensable functions (Fig. 4*B*).

We selected several hypothetical protein complexes from our dataset for AP–MS experimental validation to verify the protein complexes from continuous DNA regions. The hypothetical proteins of All1475, All7197, and Alr4359 were amplified from the *Anabaena* sp. genome and cloned into plasmid pRL25N for overexpression, which contained a Cu^{2+}-induced *petE* promotor and a GFP ORF on the C-terminal region of the target gene (54, 55). From AP–MS results, it was clear that their interacting partners were located upstream and downstream of the genes (Fig. 4*C*). It is worth noting that some gene segments, including ORFs with the same or reverse origin, could form protein complexes, such as Alr4359, which could interact with All4357 and All4358 (Fig. 4*C*), which are ATP-dependent Clp proteases that regulate protein degradation within the cell (56).

Alr4359 Is Located on Cellular Poles, and Its Overproduction can Influence the Diazotrophic Growth of Filaments

We observed that the hypothetical protein, Alr4359, interacts with proteins from the continuous DNA region *all4118–all4120* in heterocysts (Fig. 4*A*). The elution profiles of Alr4359 and Alr4119 were more similar in heterocysts than in vegetative cells (Fig. 5*A*). The strong interaction between Alr4359 and Alr4119 in heterocysts was validated by AP–MS (supplemental Table S6). In addition, we found that Alr4359 interacted with FraH (Fig. 5*A*), which was validated by Y2H (supplemental Fig. S9*A*). Previously, an interaction between these two proteins was also found by Y2H in *Synechocystis* sp. PCC 6803 (57). The AP–MS experiment further confirmed that FraH–Alr4359–Alr4119 protein complex exists in heterocysts (supplemental Fig. S9*B*). Alr4119 belongs to the CURVATURE THYLAKOID 1 family and contains the homologous CAAD domain, which has the membrane-bending capacity and influences thylakoid organization (58). The inactivation of *fraH* can cause filament fragmentation and lead to defects in the intracellular membrane structure close to the heterocyst poles (6). Interestingly, both FraH and Alr4119 have been shown to be dynamically located, changing from peripheral localization in the vegetative cells of nitrate-grown filaments to the heterocyst poles in diazotrophic filaments (58).

We used the Alr4359–GFP overexpression strain to trace the intracellular location of the Alr4359 protein and determine whether it had the same location pattern as FraH and Alr4119 (Fig. 5*B*). The fluorescence intensity of Alr4359 was obviously less than that of the control and FraH, indicating that the level of Alr4359 is strictly regulated. Alr4359 also interacts with ATP-dependent Clp proteases All4357 and All4358, which might regulate the rapid turnover of Alr4359. However, similarly to FraH, Alr4359 was also clearly located on the cell membrane and vegetative cell poles after nitrogen deprivation for 24 h. It is worth noting that no heterocysts were formed in the Alr4359 overexpression strain after nitrogen deprivation for 24 h or more (Fig. 5*B*). To observe the localization of Alr4359 in heterocysts, we regulated the expression of *alr4359* by controlling the copper concentration in BG11 medium to reduce the activity of the *petE* promoter in the pRL25N plasma. When the filaments were cultured in BG11 without copper sulfate, the *petE* promoter's transcription was dramatically reduced, and GFP fluorescence was eliminated in the filaments (supplemental Fig. S10). The filaments were then transferred to BG11 medium without soluble copper for 24 h, and heterocyst differentiation was induced without any observation of GFP fluorescence. After returning the copper levels in the BG11 medium to normal levels for 12 h, we observed that Alr4359 was located on heterocyst poles (supplemental Fig. S10).

We observed the morphological characteristics of strains grown with or without combined nitrogen resources (Fig. 6*A*). Interestingly, thylakoid membranes were dramatically condensed and aggregated in both types of cells in FraH overexpression strain under nitrogen-deficient conditions. Unlike the supercompression of the thylakoid membrane in the FraH overexpressing strain, red spontaneous fluorescence was eliminated along the filaments in the Alr4359 overexpression strain (Fig. 6*A*). The diazotrophic growth of the Alr4359 overexpression strain was also seriously decreased, as shown by the significant change in the pigment composition (Fig. 6*B*). In addition, the filament length of the Alr4359 overexpression strain was longer than that of the control under nitrogen-deprivation conditions, even when no heterocysts were formed (Fig. 6*C*). Our results verify that the Alr4359 was mainly localized on the cell's poles, and its overexpression can suppress heterocyst development and increase the length of filaments.

DISCUSSION

Filamentous nitrogen-fixing cyanobacteria served as a model organism to study cell differentiation and nitrogen fixation. It is necessary to explore PPI networks to understand better the molecular mechanisms underlying these processes. We constructed the PPIs of *Anabaena* sp. PCC 7120 by CoFrac–MS and machine learning, which is the largest protein interaction dataset of *Anabaena* sp. so far. We also generated a protein interaction map of *Synechocystis* sp. PCC 6803 recently (18). However, presenting protein interaction is not

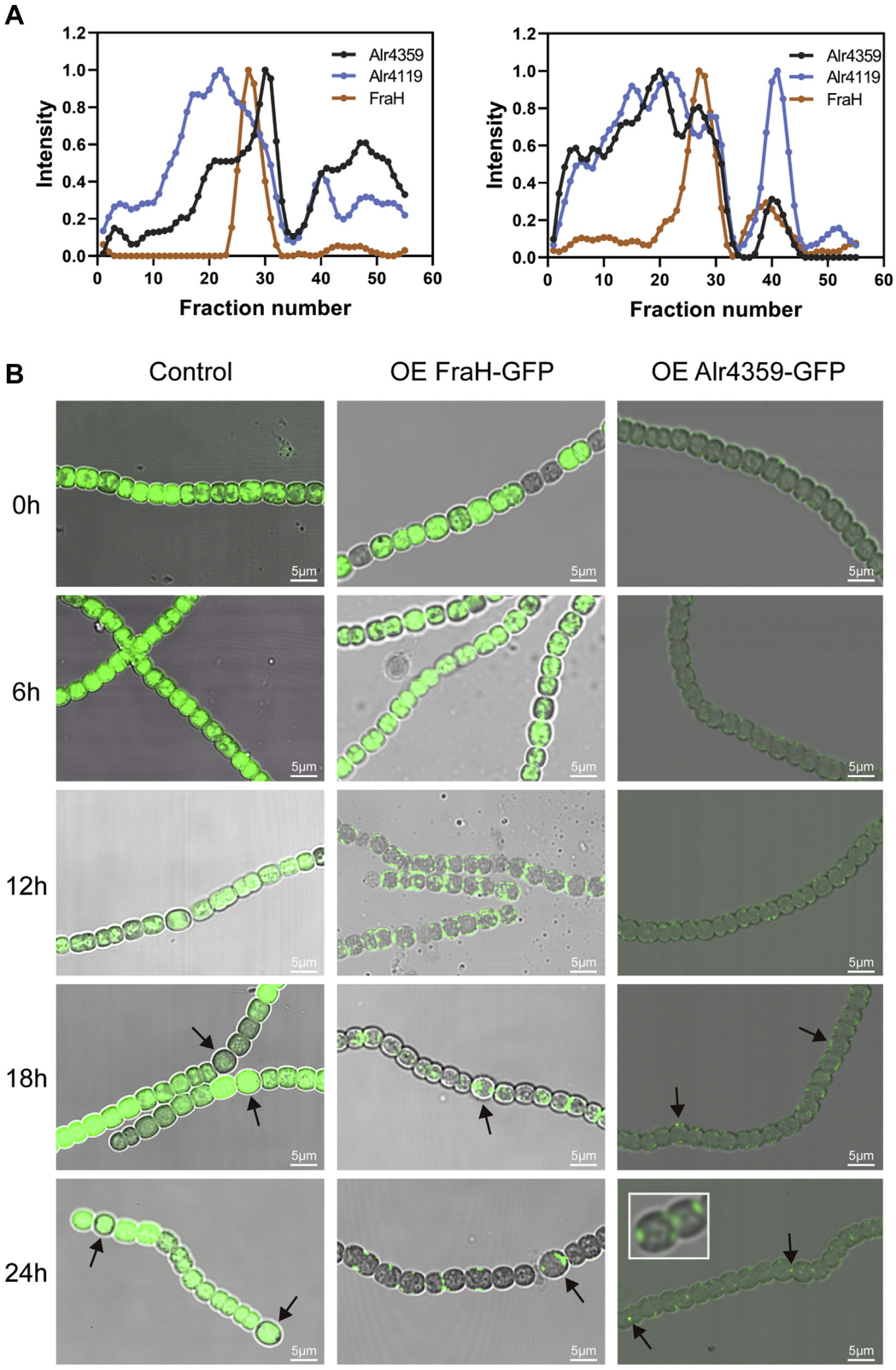


FIG. 5. **The localization of the hypothetical protein Alr4359 and its interacting partners.** *A*, the elution profiles of Alr4119–Alr4359–FraH in vegetative cells and heterocysts. *B*, the overexpression strains grown with nitrate were subjected to nitrogen step down and visualized by confocal microscopy at multiple time points. The bar represents 5 μm.

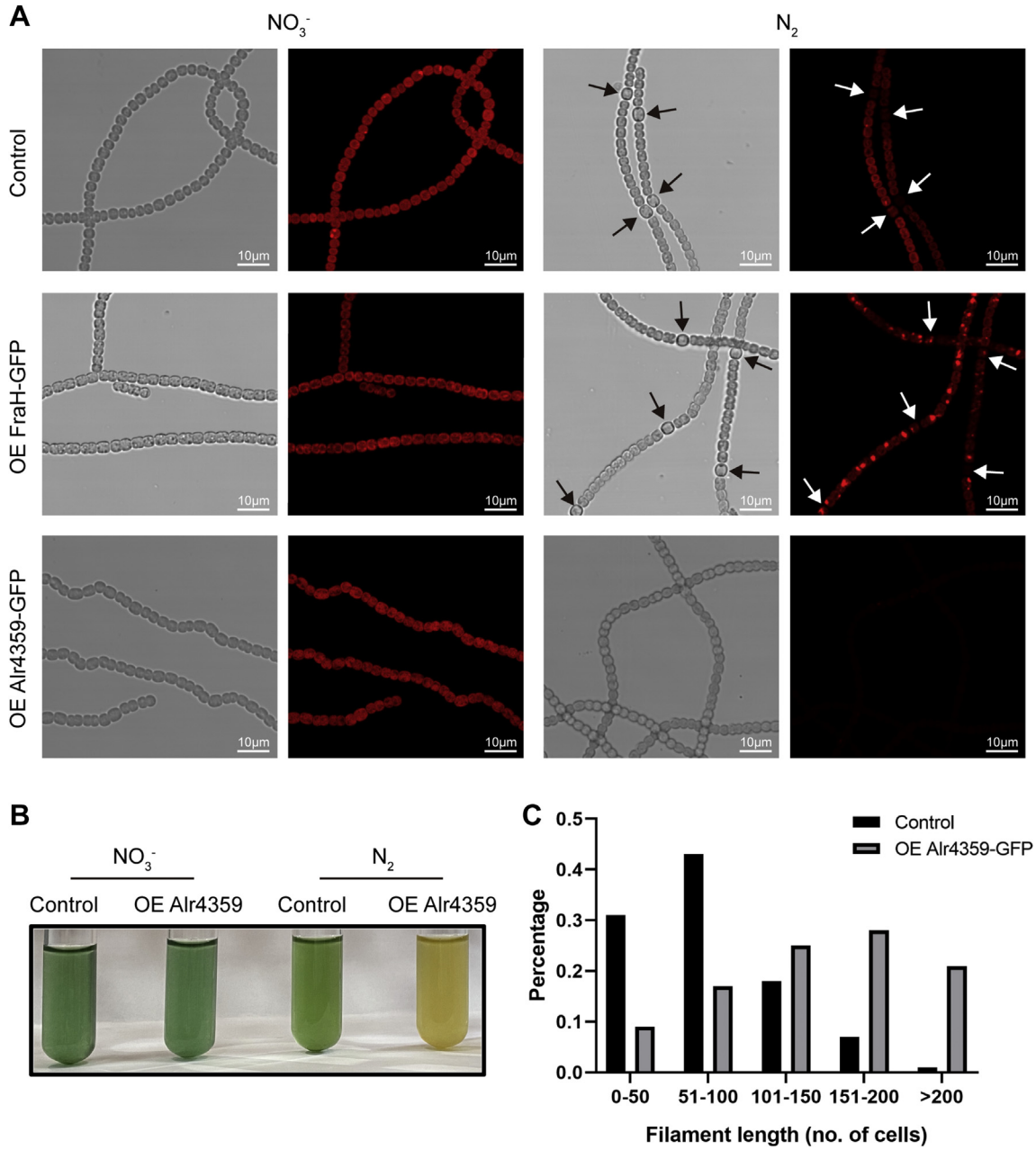


FIG. 6. **Diazotrophic growth and thylakoid membrane organization in the Alr4359 overexpression strain.** *A*, the morphological characteristics and distribution of *red spontaneous fluorescence* in different overexpression strains with or without combined nitrogen. *Arrows* mark the heterocyst. The bar represents 10 μm. *B*, the growth conditions of the control and Alr4359 overexpression strains under nitrogen-sufficient and nitrogen-deficient conditions. *C*, the distribution of the filament length in strains cultured without combined nitrogen.

the only purpose of this work. Most important is to figure out the variation of PPIs between two cell types and then reveal the function variation of proteins.

Generally, oxygenic photosynthesis and CO_2 fixation are performed in vegetative cells, whereas nitrogen fixation occurs in heterocysts (8). To date, no photosynthetic oxygen evolution activity has been detected in heterocysts. Thus, it has been misunderstood that the PSII structure is degraded in heterocysts (4). High-resolution MS was used to predict *Anabaena* sp. PPIs, and we observed most of the PSII proteins and their interaction with other proteins. In addition, phycobilisomes remained present, and the complex structure did not change dramatically, as most of the components did not shift on chromatography. However, the absorption peak of phycocyanobilin at 630 nm was diminished in heterocysts. Meanwhile, respiration was enhanced in heterocysts, maybe

because of energy consumption during nitrogen fixation. In line with these observations, PSI components that serve as part of the electron-transport chain were upregulated in heterocysts (48). In recent years, some PSI proteins have been identified in heterocysts; for example, PsaB2 affects the electron transfer properties of PSI in heterocysts (59). The results presented here indicate that intact and functional photosystems, including PSI, PSII, and phycobilisomes, exist in both cell types. However, the pigment components associated with them change during heterocyst differentiation.

Interestingly, the thylakoid membrane structure was reorganized during heterocyst differentiation. FraH and Alr4119 can influence honeycomb formation and protein redistribution in the thylakoid membrane during heterocyst differentiation (6, 58). Here, we found that Alr4359 was a new factor that influences heterocyst differentiation and the diazotrophic growth of filaments. Under nitrogen-deficient conditions, we observed that the Alr4359 overexpression strain could not form heterocysts, and the pigment composition and filament length were also altered in this line. It may cause by twisted and loose thylakoid membrane structure, as shown by scanning electron microscopy (supplemental Fig. S11). A further experiment is needed to confirm the thylakoid membrane structure because of the technical limitations in this experiment. We also found that in heterocysts, FraH can interact with SepJ, which is located in the cell poles to control material communication between adjacent cells (60). Alr4359 was not found to interact with SepJ directly, but it interacts with Alr2947, an interaction partner of SepJ (supplemental Table S6). It is reasonable to speculate that the Alr4359–FraH–Alr4119 complex, located on heterocyst poles, influences material exchange by interacting with SepJ and its partner Alr2947.

CoFrac–MS is a powerful tool to monitor dynamic protein changes and uncover important functional factors when cells are in different life cycles or exposed to different environmental conditions (42, 61). However, it also has some disadvantages, such as cannot distinguish protein aggregation. It is worth noting that both AP–MS and CoFrac–MS tended to identify tight and stable protein complexes in the cell since the weak physical interaction always depolymerized during the protein extraction process. Combining these methods with crosslinking, the weak and instantaneous interactions that have been ignored can be captured easily. Crosslinking combined with CoFrac–MS was also applied to generate protein correlation profiling of global membrane proteins in humans (62). In addition, the interaction between Psb28 and cytochrome b559 in PSII was found by crosslinking combined with MS in *Synechocystis* sp. PCC 6803 (63). Over the years, the combination of cryo-EM with CoFrac–MS has facilitated the development of systems' structural proteomics by reducing the requirement for a pure and homogeneous sample (64, 65). A fully assembled structural information of a certain protein complex can be obtained by utilizing the cryo-EM, crosslinking, and MS in native cell extracts, such as pyruvate dehydrogenase complex (66, 67). Combined with these assays, the active nanostructures of protein complexes involved in heterocyst differentiation can be further investigated in the future.

In addition to protein expression and interaction, post-translational modifications can act as another factor of protein regulation. Spectra of AP–MS revealed that Alr4359, FraH, and their interaction partner, Alr4119, all have phosphorylation modifications. The T362 and T410 phosphorylation on Alr4359 and T177 phosphorylation on FraH are shown in supplemental Fig. S12. Alr4119 has four phosphorylation modification sites on S32, T25, T27, and T29. The phosphokinase Alr0548 and the phosphorylase Alr0547 were also found in the prey proteins of FraH AP–MS result (supplemental Fig. S9*B*). FraH interacts with phosphokinase Alr0548 in vegetative cells, and it interacts with phosphorylase Alr0547 in heterocysts. Alr0548 was also found in the prey proteins of Alr4359 AP–MS result (supplemental Table S6). The results indicate that phosphorylation modification may affect the function of the Alr4359–FraH–Alr4119 complex during heterocyst differentiation.

There is still follow-up work to be done. The relationships across protein complexes, genomic regions, and upregulation or downregulation in different *Anabaena* states need to be explored. More experiments are needed to determine the specific function of each protein. However, our dataset provides valuable candidate proteins for further research. Knowledge of dynamic protein interactions can help to explain the functional differences between vegetative cells and heterocysts.

DATA AVAILABILITY

All the LC/MS/MS raw files have been deposited in the iProX database and can be accessed with ID IPX0002954000 (https://www.iprox.org//page/project.html?id=IPX0002954000) or PXD025312 (http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD025312).

Supplemental data—This article contains supplemental data.

Acknowledgments—This work was supported by the National Natural Science Foundation of China (grant nos.: 31800647 and 91951210). We thank Dr Juyuan Zhang at the Institute of Hydrobiology Chinese Academy of Sciences for valuable discussions. We also thank Prof Xiang Gao at the Central China Normal University for providing algal, cloning vectors, and technical guidance.

Author contributions—C. W. conceptualization; C. X. formal analysis; C. X., B. W., H. H., and J. H. investigation; C. X. data curation; C. X. writing–original draft; C. W. writing–review & editing; C. W. supervision.

Conflict of interest—The authors declare no competing interests.

Abbreviations—The abbreviations used are: AP–MS, affinity purification followed by mass spectrometry; CoFrac–MS, cofractionation coupled with mass spectrometry; Cox, cytochrome *c* oxidase; EPIC, elution profile–based inference of complex; GO, Gene Ontology; MS, mass spectrometry; PPI, protein–protein interaction; PSI, photosystem I; PSII, photosystem II; SEC, size-exclusion chromatography; Y2H, yeast two-hybrid.

Received April 15, 2021, and in revised from, February 20, 2022
Published, MCPRO Papers in Press, March 11, 2022, https://doi.org/10.1016/j.mcpro.2022.100224

REFERENCES

1. Flores, E., Picossi, S., Valladares, A., and Herrero, A. (2019) Transcriptional regulation of development in heterocyst-forming cyanobacteria. *Biochim. Biophys. Acta Gene Regul. Mech.* **1862**, 673–684
2. Kumar, K., Mella-Herrera, R. A., and Golden, J. W. (2010) Cyanobacterial heterocysts. *Cold Spring Harb. Perspect. Biol.* **2**, a000315
3. Nicolaisen, K., Hahn, A., and Schleiff, E. (2009) The cell wall in heterocyst formation by *Anabaena* sp. PCC 7120. *J. Basic Microbiol.* **49**, 5–24
4. Magnuson, A. (2019) Heterocyst thylakoid bioenergetics. *Life* **9**, 13
5. Muro-Pastor, A. M., and Hess, W. R. (2012) Heterocyst differentiation: From single mutants to global approaches. *Trends Microbiol.* **20**, 548–557
6. Merino-Puerto, V., Mariscal, V., Schwarz, H., Maldener, I., Mullineaux, C. W., Herrero, A., and Flores, E. (2011) FraH is required for reorganization of intracellular membranes during heterocyst differentiation in *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **193**, 6815–6823
7. Nürnberg, D. J., Mariscal, V., Bornikoel, J., Nieves-Morión, M., Krauß, N., Herrero, A., Maldener, I., Flores, E., and Mullineaux, C. W. (2015) Intercellular diffusion of a fluorescent sucrose analog via the septal junctions in a filamentous cyanobacterium. *mBio* **6**, e02109
8. Herrero, A., and Flores, E. (2019) Genetic responses to carbon and nitrogen availability in *Anabaena*. *Environ. Microbiol.* **21**, 1–17
9. Flores, E., Nieves-Morión, M., and Mullineaux, C. W. (2019) Cyanobacterial septal junctions: Properties and regulation. *Life (Basel)* **9**, 1
10. Weiss, G. L., Kieninger, A.-K., Maldener, I., Forchhammer, K., and Pilhofer, M. (2019) Structure and function of a bacterial gap junction analog. *Cell* **178**, 374–384.e15
11. Omairi-Nasser, A., Mariscal, V., Austin, J. R., and Haselkorn, R. (2015) Requirement of Fra proteins for communication channels between cells in the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. PCC 7120. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E4458–E4464
12. Zhong, Q., Pevzner, S. J., Hao, T., Wang, Y., Mosca, R., Menche, J., Taipale, M., Tasan, M., Fan, C., Yang, X., Haley, P., Murray, R. R., Mer, F., Gebreab, F., Tam, S., *et al*. (2016) An inter-species protein-protein interaction network across vast evolutionary distance. *Mol. Syst. Biol.* **12**, 865
13. Luck, K., Sheynkman, G. M., Zhang, I., and Vidal, M. (2017) Proteome-scale human interactomics. *Trends Biochem. Sci.* **42**, 342–354
14. Kristensen, A. R., Gsponer, J., and Foster, L. J. (2012) A high-throughput approach for measuring temporal changes in the interactome. *Nat. Methods* **9**, 907–909
15. Havugimana, P. C., Hart, G. T., Nepusz, T., Yang, H. X., Turinsky, A. L., Li, Z. H., Wang, P. I., Boutz, D. R., Fong, V., Phanse, S., Babu, M., Craig, S. A., Hu, P. Z., Wan, C. H., Vlasblom, J., *et al*. (2012) A census of human soluble protein complexes. *Cell* **150**, 1068–1081
16. Kirkwood, K. J., Ahmad, Y., Larance, M., and Lamond, A. I. (2013) Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics. *Mol. Cell. Proteomics* **12**, 3851–3873
17. Gilbert, M., and Schulze, W. X. (2019) Global identification of protein complexes within the membrane proteome of *Arabidopsis* roots using a SEC-MS approach. *J. Proteome Res.* **18**, 107–119
18. Xu, C., Wang, B., Yang, L., Zhongming Hu, L., Yi, L., Wang, Y., Chen, S., Emili, A., and Wan, C. (2021) Global landscape of native protein complexes in *Synechocystis* sp. PCC 6803. *Genomics Proteomics Bioinformatics*. https://doi.org/10.1016/j.gpb.2020.06.020
19. Aryal, U. K., Xiong, Y., McBride, Z., Kihara, D., Xie, J., Hall, M. C., and Szymanski, D. B. (2014) A proteomic strategy for global analysis of plant protein complexes. *Plant Cell* **26**, 3867–3882
20. Crozier, T. W. M., Tinti, M., Larance, M., Lamond, A. I., and Ferguson, M. A. J. (2017) Prediction of protein complexes in *Trypanosoma brucei* by protein correlation profiling mass spectrometry and machine learning. *Mol. Cell. Proteomics* **16**, 2254–2267
21. Wan, C., Borgeson, B., Phanse, S., Tu, F., Drew, K., Clark, G., Xiong, X., Kagan, O., Kwan, J., Bezginov, A., Chessman, K., Pal, S., Cromar, G., Papoulas, O., Ni, Z., *et al*. (2015) Panorama of ancient metazoan macromolecular complexes. *Nature* **525**, 339–344
22. McWhite, C. D., Papoulas, O., Drew, K., Cox, R. M., June, V., Dong, O. X., Kwon, T., Wan, C., Salmi, M. L., Roux, S. J., Browning, K. S., Chen, Z. J., Ronald, P. C., and Marcotte, E. M. (2020) A pan-plant protein complex map reveals deep conservation and novel assemblies. *Cell* **181**, 460–474 e414
23. Drew, K., Lee, C., Huizar, R. L., Tu, F., Borgeson, B., McWhite, C. D., Ma, Y., Wallingford, J. B., and Marcotte, E. M. (2017) Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes. *Mol. Syst. Biol.* **13**, 932
24. Razquin, P., Fillat, M. F., Schmitz, S., Stricker, O., Bohme, H., Gomez-Moreno, C., and Peleato, M. L. (1996) Expression of ferredoxin-NADP+ reductase in heterocysts from *Anabaena* sp. *Biochem. J.* **316**, 157–160
25. Lemeille, S., Geiselmann, J., and Latifi, A. (2005) Crosstalk regulation among group 2-sigma factors in *Synechocystis* PCC 6803. *BMC Microbiol.* **5**, 18
26. Yoshimura, T., Imamura, S., Tanaka, K., Shirai, M., and Asayama, M. (2007) Cooperation of group 2 sigma factors, SigD and SigE for light-induced transcription in the cyanobacterium *Synechocystis* sp. PCC 6803. *FEBS Lett.* **581**, 1495–1500
27. Muro-Pastor, A. M., Herrero, A., and Flores, E. (2001) Nitrogen-regulated group 2 sigma factor from *Synechocystis* sp. strain PCC 6803 involved in survival under nitrogen stress. *J. Bacteriol.* **183**, 1090–1095
28. Imamura, S., Tanaka, K., Shirai, M., and Asayama, M. (2006) Growth phase-dependent activation of nitrogen-related genes by a control network of group 1 and group 2 sigma factors in a cyanobacterium. *J. Biol. Chem.* **281**, 2668–2675
29. Ballal, A., Basu, B., and Apte, S. K. (2007) The Kdp-ATPase system and its regulation. *J. Biosci.* **32**, 559–568
30. Alahari, A., and Apte, S. K. (1998) Pleiotropic effects of potassium deficiency in a heterocystous, nitrogen-fixing cyanobacterium, *Anabaena* torulosa. *Microbiology* **144**, 1557–1563
31. Nanatani, K., Shijuku, T., Takano, Y., Zulkifli, L., Yamazaki, T., Tominaga, A., Souma, S., Onai, K., Morishita, M., and Ishiura, M. (2015) Comparative analysis of kdp and ktr mutants reveals distinct roles of the potassium transporters in the model cyanobacterium *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* **197**, 676–687
32. Ban, N., Beckmann, R., Cate, J. H., Dinman, J. D., Dragon, F., Ellis, S. R., Lafontaine, D. L., Lindahl, L., Liljas, A., Lipton, J. M., McAlear, M. A., Moore, P. B., Noller, H. F., Ortega, J., Panse, V. G., *et al*. (2014) A new system for naming ribosomal proteins. *Curr. Opin. Struct. Biol.* **24**, 165–169
33. Araya-Garay, J., Feijoo-Siota, L., Veiga-Crespo, P., Sanchez-Perez, A., and González Villa, T. (2014) Cloning and functional expression of zeta-carotene desaturase, a novel carotenoid biosynthesis gene from Ficus carica. *Int. J. Microbiol. Adv. Immunol.* **2**, 32–40
34. Cardona, T., and Magnuson, A. (2010) Excitation energy transfer to photosystem I in filaments and heterocysts of Nostoc punctiforme. *Biochim. Biophys. Acta* **1797**, 425–433
35. Cardona, T., Battchikova, N., Zhang, P. P., Stensjo, K., Aro, E. M., Lindblad, P., and Magnuson, A. (2009) Electron transfer protein complexes in the thylakoid membranes of heterocysts from the cyanobacterium Nostoc punctiforme. *Biochim. Biophys. Acta* **1787**, 252–263
36. Magnuson, A., and Cardona, T. (2016) Thylakoid membrane function in heterocysts. *Biochim. Biophys. Acta* **1857**, 309–319
37. Watanabe, M., Semchonok, D. A., Webber-Birungi, M. T., Ehira, S., Kondo, K., Narikawa, R., Ohmori, M., Boekema, E. J., and Ikeuchi, M. (2014) Attachment of phycobilisomes in an antenna-photosystem I supercomplex of cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2512–2517

38. Ghirardi, M. L., Posewitz, M. C., Maness, P.-C., Dubini, A., Yu, J., and Seibert, M. (2007) Hydrogenases and hydrogen photoproduction in oxygenic photosynthetic organisms. *Annu. Rev. Plant Biol.* **58**, 71–91
39. Einsle, O., and Rees, D. C. (2020) Structural enzymology of nitrogenase enzymes. *Chem. Rev.* **120**, 4969–5004
40. Hu, L. Z., Goebels, F., Tan, J. H., Wolf, E., Kuzmanov, U., Wan, C., Phanse, S., Xu, C., Schertzberg, M., Fraser, A. G., Bader, G. D., and Emili, A. (2019) EPIC: Software toolkit for elution profile-based inference of protein complexes. *Nat. Methods* **16**, 737–742
41. Ouma, W. Z., Pogacar, K., and Grotewold, E. (2018) Topological and statistical analyses of gene regulatory networks reveal unifying yet quantitatively different emergent properties. *PLoS Comput. Biol.* **14**, 1–17
42. Guerreiro, A. C. L., Penning, R., Raaijmakers, L. M., Axman, I. M., Heck, A. J. R., and Altelaar, A. F. M. (2016) Monitoring light/dark association dynamics of multi-protein complexes in cyanobacteria using size exclusion chromatography-based proteomics. *J. Proteomics* **142**, 33–44
43. Kumazaki, S., Akari, M., and Hasegawa, M. (2013) Transformation of thylakoid membranes during differentiation from vegetative cell into heterocyst visualized by microscopic spectral imaging. *Plant Physiol.* **161**, 1321–1333
44. Sugiura, K., and Itoh, S. (2012) Single-cell confocal spectrometry of a filamentous cyanobacterium *Nostoc* at room and cryogenic temperature. Diversity and differentiation of pigment systems in 311 cells. *Plant Cell Physiol.* **53**, 1492–1506
45. MacColl, R. (1998) Cyanobacterial phycobilisomes. *J. Struct. Biol.* **124**, 311–334
46. Flaherty, B. L., Van Nieuwerburgh, F., Head, S. R., and Golden, J. W. (2011) Directional RNA deep sequencing sheds new light on the transcriptional response of *Anabaena* sp. strain PCC 7120 to combined-nitrogen deprivation. *BMC Genomics* **12**, 1–10
47. Ehira, S., Ohmori, M., and Sato, N. (2003) Genome-wide expression analysis of the responses to nitrogen deprivation in the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res.* **10**, 97–113
48. Ow, S. Y., Cardona, T., Taton, A., Magnuson, A., Lindblad, P., Stensjö, K., and Wright, P. C. (2008) Quantitative shotgun proteomics of enriched heterocysts from *Nostoc* sp. PCC 7120 using 8-plex isobaric peptide tags. *J. Proteome Res.* **7**, 1615–1628
49. Murataliev, M. B., Feyereisen, R., and Walker, F. A. (2004) Electron transfer by diflavin reductases. *Biochim. Biophys. Acta* **1698**, 1–26
50. Mellor, S. B., Vavitsas, K., Nielsen, A. Z., and Jensen, P. E. (2017) Photosynthetic fuel for heterologous enzymes: The role of electron carrier proteins. *Photosynth. Res.* **134**, 329–342
51. Valladares, A., Maldener, I., Muro-Pastor, A. M., Flores, E., and Herrero, A. (2007) Heterocyst development and diazotrophic metabolism in terminal respiratory oxidase mutants of the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **189**, 4425–4430
52. Awai, K., and Wolk, C. P. (2007) Identification of the glycosyl transferase required for synthesis of the principal glycolipid characteristic of heterocysts of *Anabaena* sp. strain PCC 7120. *FEMS Microbiol. Lett.* **266**, 98–102
53. Fan, Q., Huang, G., Lechno-Yossef, S., Wolk, C. P., Kaneko, T., and Tabata, S. (2005) Clustered genes required for synthesis and deposition of envelope glycolipids in *Anabaena* sp. strain PCC 7120. *Mol. Microbiol.* **58**, 227–243
54. Zhang, S. R., Lin, G. M., Chen, W. L., Wang, L., and Zhang, C. C. (2013) ppGpp metabolism is involved in heterocyst development in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **195**, 4536–4544
55. Wolk, C. P., Cai, Y., Cardemil, L., Flores, E., Hohn, B., Murry, M., Schmetterer, G., Schrautemeier, B., and Wilson, R. (1988) Isolation and complementation of mutants of *Anabaena* sp. strain PCC 7120 unable to grow aerobically on dinitrogen. *J. Bacteriol.* **170**, 1239–1244
56. Olinares, P. D. B., Kim, J., and van Wijk, K. J. (2011) The Clp protease system; a central component of the chloroplast protease network. *Biochim. Biophys. Acta* **1807**, 999–1011
57. Sato, S., Shimoda, Y., Muraki, A., Kohara, M., Nakamura, Y., and Tabata, S. (2007) A large-scale protein–protein interaction analysis in *Synechocystis* sp. PCC6803. *DNA Res.* **14**, 207–216
58. Santamaria-Gomez, J., Mariscal, V., and Luque, I. (2018) Mechanisms for protein redistribution in thylakoids of *Anabaena* during cell differentiation. *Plant Cell Physiol.* **59**, 1860–1873
59. Magnuson, A., Krassen, H., Stensjö, K., Ho, F. M., and Styring, S. (2011) Modeling photosystem I with the alternative reaction center protein PsaB2 in the nitrogen fixing cyanobacterium *Nostoc punctiforme*. *Biochim. Biophys. Acta* **1807**, 1152–1161
60. Flores, E., Pernil, R., Muro-Pastor, A. M., Mariscal, V., Maldener, I., Lechno-Yossef, S., Fan, Q., Wolk, C. P., and Herrero, A. (2007) Septum-localized protein required for filament integrity and diazotrophy in the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **189**, 3884–3890
61. Heusel, M., Frank, M., Kohler, M., Amon, S., Frommelt, F., Rosenberger, G., Bludau, I., Aulakh, S., Linder, M. I., Liu, Y., Collins, B. C., Gstaiger, M., Kutay, U., and Aebersold, R. (2020) A global screen for assembly state changes of the mitotic proteome by SEC-SWATH-MS. *Cell Syst.* **10**, 133–155.e6
62. Larance, M., Kirkwood, K. J., Tinti, M., Murillo, A. B., Ferguson, M. A. J., and Lamond, A. I. (2016) Global membrane protein interactome analysis using *in vivo* crosslinking and mass spectrometry-based protein correlation profiling. *Mol. Cell. Proteomics* **15**, 2476–2490
63. Weisz, D. A., Liu, H., Zhang, H., Thangapandian, S., Tajkhorshid, E., Gross, M. L., and Pakrasi, H. B. (2017) Mass spectrometry-based cross-linking study shows that the Psb28 protein binds to cytochrome b 559 in photosystem II. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 2224–2229
64. Su, C. C., Lyu, M., Morgan, C. E., Bolla, J. R., Robinson, C. V., and Yu, E. W. (2021) A 'Build and Retrieve' methodology to simultaneously solve cryo-EM structures of membrane proteins. *Nat. Methods* **18**, 69
65. Kastritis, P. L., O'Reilly, F. J., Bock, T., Li, Y. Y., Rogon, M. Z., Buczak, K., Romanov, N., Betts, M. J., Bui, K. H., Hagen, W. J., Hennrich, M. L., Mackmull, M. T., Rappsilber, J., Russell, R. B., Bork, P., *et al*. (2017) Capturing protein communities by structural proteomics in a thermophilic eukaryote. *Mol. Syst. Biol.* **13**, 936
66. Tuting, C., Kyrilis, F. L., Muller, J., Sorokina, M., Skalidis, I., Hamdi, F., Sadian, Y., and Kastritis, P. L. (2021) Cryo-EM snapshots of a native lysate provide structural insights into a metabolon-embedded transacetylase reaction. *Nat. Commun.* **12**, 6933
67. Kyrilis, F. L., Semchonok, D. A., Skalidis, I., Tuting, C., Hamdi, F., O'Reilly, F. J., Rappsilber, J., and Kastritis, P. L. (2021) Integrative structure of a 10-megadalton eukaryotic pyruvate dehydrogenase complex from native cell extracts. *Cell Rep.* **34**, 108727