

# Profiling and Compositional Analysis of the Exoproteome of *Synechocystis* Sp. PCC 6803

**Keywords:** *Synechocystis*; Exoproteome; Glycosylation; Secretion; Extracellular Milieu

## Abstract

The exoproteome is one of the important components of the extracellular milieu of the unicellular cyanobacterium *Synechocystis* sp. PCC 6803, a photosynthetic model organism that is emerging as a potential cell factory for producing clean and renewable biofuel. Herein, we identified the exoproteome of this organism using high resolution mass spectrometry and analyzed the sources of origins of the identified extracellular proteins. In total, 201 proteins were identified, including the cell surface protein HlyA and the periplasmic protein FutA2 that are the most abundant in the exoproteome. The combined amount of the two proteins is estimated to constitute more than 50 percent of the total amount of proteins in the extracellular milieu. More than 100 proteins were predicted to be secretory or previously identified as periplasmic proteins, and the total amount of these proteins is nearly 88% of that of all identified extracellular proteins, establishing secretion as the major contributor to the protein population of the exoproteome. In addition, 2 proteins were found to be heavily glycosylated, suggesting that the exoproteome is an enriched pool of a certain subset of glycoproteins. Together, our work provided a comprehensive catalog of the *Synechocystis* exoproteome and a high valuable resource for the related research fields.

## Introduction

An exoproteome refers to all proteins in the extracellular milieu of an organism that include specifically secreted proteins, proteins exported via other mechanisms, and proteins released from cell lysis [1]. The exoproteome is one of the important components that defining the extracellular milieu of the cells. The functional importance of extracellular proteins has long been appreciated for both eukaryotes and prokaryotes. In bacteria, the major known functions of extracellular proteins include toxicity, nutrient acquisition, and cell motility [1]. The multifaceted functions of extracellular proteins endow bacterial the great potential in biotechnological applications such as production of useful protein toxins or waste water treatment [2,3]. Though the exoproteomes of many pathogenic bacteria have been the focus of proteomic studies for the identification of novel virulence determinants and for the understanding of the host-pathogen interactions [4-9], the importance of the exoproteomes of a few non-pathogenic marine bacteria has started to be realized with regard to their interaction with and the potential effect on their living environment [10,11].

Cyanobacteria are a group of photosynthetic bacteria that are widely distributed in almost every terrestrial and aquatic habitat, and are believed to play a critical role in the generation of the gas oxygen in the atmosphere and are estimated to account for more than 50 percent of the biomass on the earth [12-14]. The unicellular cyanobacterium *Synechocystis* sp. PCC 6803 (Hereafter referred to as *Synechocystis*) has been widely used as a model organism for photosynthesis



Liyan Gao<sup>1</sup>, Xiahe Huang<sup>1</sup>, Haitao Ge<sup>2</sup>, Yuanya Zhang<sup>1</sup>, Yu Kang<sup>1</sup>, Longfa Fang<sup>1</sup>, Kehui Liu<sup>1</sup> and Yingchun Wang<sup>1\*</sup>

<sup>1</sup>State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, No.1 West Beichen Rd, Beijing 100101, China

<sup>2</sup>State Key Laboratory of Microbial Technology, Shandong University, Jinan 250100, China

### \*Address for Correspondence

Dr. Yingchun Wang Ph.D., State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, No.1 West Beichen Rd, Beijing 100101, China, Tel: (010) 6480-6149; Fax: (010) 6480-6252; Email: ycwang@genetics.ac.cn

**Submission:** 25 September 2014

**Accepted:** 13 November 2014

**Published:** 15 November 2014

**Reviewed & Approved by:** Dr. Hsueh-Fen Juan is a Professor in the Department of Life Science, Institute of Molecular and Cellular Biology at National Taiwan University, Taiwan

research because of its high resemblance to the chloroplast of higher plants [15-17]. *Synechocystis* is the first cyanobacterium with completely sequenced genome [18], and has the ability to naturally take up foreign DNA and integrate into its own genome [16], making it an ideal system for genetic manipulations aiming at redirecting the metabolic flux to the pathways for biofuel production. The ability to grow photoautotrophically without exogenously supplied carbon source not only allows *Synechocystis* to be emerging as the potential cost-effective cell factory for producing clean and renewable biofuel [19-21], but also to be a promising bioreactor for producing secretory druggable peptides.

Because of its importance as a model organism in photosynthesis and the great potential in industrial application, the proteome and the subproteomes of different subcellular compartments of *Synechocystis* under different culture conditions have been extensively studied [16,17,22-25]. However, the exoproteomes of any cyanobacteria including *Synechocystis* have remained largely unexplored, and such studies are pressingly demanded for the understanding of their extracellular milieu that affect their growth, survival, toxicity, and potentially their potential as a cell factory.

In the current work, we isolated the total extracellular proteins from a photomixotrophically-growing *Synechocystis* culture that was supplemented with glucose, and then systematically identified the extracellular proteins using high resolution mass spectrometry. We further performed in-depth analyses to rank the relative abundance of the extracellular proteins and to estimate their sources of origins. The results obtained will be expected to build a useful resource with a wealth of information regarding protein expression, modification, source of origin, and abundance in the extracellular milieu of the important photosynthetic organism.

## Materials and Methods

### Preparation of the Total and the Extracellular Proteins of *Synechocystis*

The wild type *Synechocystis* was cultured in liquid BG11 medium supplemented with 5 mM glucose at 30°C under medium light intensity (50  $\mu\text{mol m}^{-2} \text{s}^{-1}$  photons). The concentration of cells in the liquid culture was estimated from the optical density at 730 nm ( $\text{OD}_{730}$ ). About 300 ml liquid culture ( $\text{OD}_{730}=1$ ) was centrifuged at 4,000 rpm for 10 min at 4°C to pellet the cells. The supernatant containing the extracellular proteins was collected and then filtered through 0.45  $\mu\text{m}$  pore size filters to further remove any insoluble cell debris. The proteins in the filtered supernatant were precipitated by 10% TCA at 4°C overnight, washed with ice-cold acetone twice and resuspended in 200  $\mu\text{l}$  2 M urea, 50 mM  $\text{NH}_4\text{HCO}_2$  for further analysis.

To prepare the total proteins from the whole cell lysate, the harvested *Synechocystis* cells were resuspended in the SMN buffer containing 0.4 M sucrose, 50 mM 3-(N-morpholino) propanesulfonic acid, pH 7.0, 10 mM NaCl, 5 mM EDTA, and 0.5 mM PMSF, and broken by vortexing with glass beads at 4°C. The glass beads and unbroken cells were removed by centrifugation at 5,000 x g for 30 min. The supernatant was collected as the total proteins.

### SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Samples for electrophoresis were prepared by mixing the respective samples with 6x loading buffer in a 5:1 ratio. The samples were then heated at 95°C for 5min, cooled to room temperature and loaded in the wells of the SDS-PAGE gel. The electrophoresis was then performed with constant power of 2 W per gel and stopped when the bromophenol blue reached the bottom of the gel. The proteins on the gel were visualized with Coomassie blue staining.

### Protein Digestion and Mass Spectrometry Analysis

The extracellular proteins resuspended in 200  $\mu\text{l}$  2M urea, 50 mM  $\text{NH}_4\text{HCO}_3$  were reduced with 10 mM dithiothreitol at 56°C for 1 h and followed by alkylation with 50 mM iodoacetamide for 45 min in dark at room temperature. The denatured proteins were then digested with sequencing grade trypsin (Promega, Madison, WI) at 37°C overnight. The resulting tryptic peptides were then desalted with C18-strong cation exchange stop-and-go extraction tips (StageTips) [26] and dried with a SpeedVac. The peptides were resuspended in 0.1% formic acid and analyzed by LTQ Orbitrap Elite mass spectrometer (ThermoFisher Scientific) coupled online to an Easy-nLC 1000 (Thermo Fisher Scientific) in a data-dependent mode. Briefly, the peptides were separated by reverse phase LC with a 75  $\mu\text{m}$  (ID)  $\times$  250 mm (length) analytical column packed with C18 particles of 5  $\mu\text{m}$  diameter. The mobile phases for the LC contains buffer A (2 % acetonitrile (ACN), 0.1 % formic acid (FA)) and buffer B (98 % ACN, 0.1 % FA), a linear gradient of buffer B from 3%-30% for 90 min was used for the separation. All MS measurements were performed in the positive ion mode and acquired across the mass range 300-1800 m/z. The fifteen most intense ions from each MS scan were isolated and fragmented by HCD.

### Data Analysis

The raw MS files were analyzed by the software MaxQuant

version 1.4.1.2 integrated with the search engine Andromeda [27,28]. MS/MS spectra were searched against the decoy *Synechocystis* proteome sequences downloaded from CyanoBase (genome. microbedb.jp/cyanobase/*Synechocystis*) [18], which includes 3,672 forward protein sequences concatenated with corresponding reverse sequences and 248 common contaminants. The precursor mass tolerance was set at 20 ppm for the initial search that was performed for mass recalibration. The mass tolerance was set at 4.5 ppm and 20 ppm for precursor and fragment ions, respectively, in the main search. The variable modifications include N-terminal acetylation and methionine oxidation, and the fix modification includes cysteine carbamidomethylation. For the identification of glycosylation sites, variable modification groups including pentose (mass difference: 132.0422), fucose (146.0579), fucose-fucose (292.1158), and fucose-pentose (278.1002) on serine, threonine, tyrosine, and asparagine residues respectively, were included for database search. A maximum of 2 miscleavages and a minimal peptide length of 7 amino acids were allowed. The false discovery rates (FDR) for both peptide and protein identifications were set to 0.01. To report the search results, multiple proteins sharing a subset of identified peptides were combined as a single protein group.

### Bioinformatics and Statistics

Bioinformatic and statistical analyses were mainly performed using the software Perseus version 1.4.0.17 [29]. The potential secretory proteins were predicted using the combination of the softwares SignalP [30], LipoP [31], and SecretomeP [32]. The Gene Ontology (GO) annotation was performed by Blast2GO using NCBI non-redundant proteome database [33], the e-value for the blast analysis was set to 1.0E-25. For the Fisher's exact test, the GO terms with p-value < 0.01 were further filtered with false discovery rate (FDR) that was set to 0.01.

## Results

### Profiling the Exoproteome of *Synechocystis*

The proteins in the extracellular milieu of *Synechocystis* are potentially originated from the shedding of cell surface, the leakage of periplasm, and the lysis of dead cells. To minimize the identification of proteins from the lysis of dead cells, we harvested the *Synechocystis* culture at an exponentially growing phase with minimal cell death. The total proteins in the culture medium were precipitated and resuspended in a SDS buffer. An aliquot of the sample, together with the proteins in the whole cell lysate, was separated by SDS-PAGE and visualized with Coomassie blue staining. The patterns of the protein profile in the two samples are drastically different, as indicated by the absence of many major protein bands in the lane for the extracellular protein that are present in the whole cell lysate. This observation suggests that lysis of dead cells is the minor source of origin for proteins in the extracellular milieu. Otherwise, the protein profiles of the two samples should be identical or similar. Remarkably, a dominant band at about 35kD is only observed in the lane for extracellular proteins but not for the whole cell lysate, suggesting that some proteins are specifically enriched in the extracellular milieu, probably due to the shedding of cell surface proteins or protein secretion.

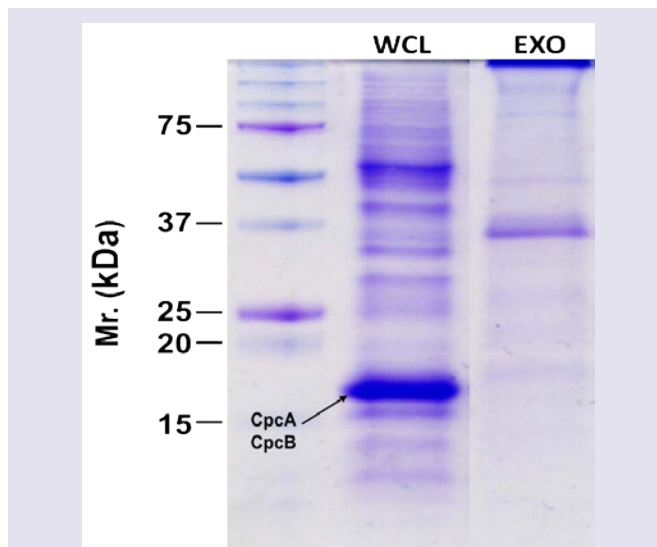
The isolated extracellular proteins were subsequently identified with LC-MS/MS. In total, 22,779 spectra were collected resulting in

the identification of 685 peptides from 201 proteins with FDR 1%, including 118 proteins with at least 2 peptides identified. Because many extracellular proteins are low abundant and because we used high resolution mass spectrometry and high stringent criteria (FDR 1%) for protein identification, we included all 1-peptide matching whose MS/MS spectra were manually examined to ensure high confident identification as positive identification for further analysis (Supplemental Figure 1). Compared with the result in the only previous report for the analysis of *Synechocystis* exoproteome, in which 7 extracellular proteins were identified including 5 proteins identified in the current study [34], the current dataset represents the first and the most comprehensive catalog of the exoproteome for this photosynthetic model organism.

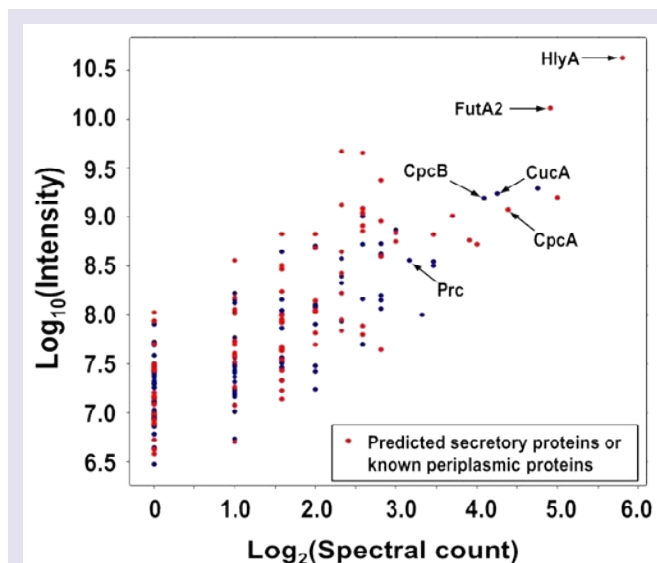
**Ranking the Relative Abundance of Extracellular Proteins**

To uncover the major components of the *Synechocystis* exoproteome, we ranked the identified extracellular proteins according to their abundances as estimated with peptide spectral counts and peak intensities (Supplemental Table 1 and Figure 2). The amount of the two most abundant proteins HlyA and FutA2, which are encoded by sl11951 and slr0513, respectively, is about 53% of the total amount of extracellular proteins if estimated by the measured peptide peak intensities. HlyA is a known S-layer RTX (repeat in toxin) motif-containing hemolysin like protein that is important for the maintenance of the cell integrity [35]. The localization of HlyA on the S-layer is mediated by type I secretion pathway that transports target proteins directly from the cytoplasm to the cell surface [36,37]. The high abundance of HlyA in the cell culture, which is consistent with a previous observation that bacterial cell surface proteins tend to accumulate in the exoproteome [38], is probably the result of improper incorporation into or the shedding of the protein from the S-layer. FutA2, a putative secretory protein with a predicted signal peptidase I (SPaseI) cleavage site at the N-terminal, is the most abundant protein in the periplasm of *Synechocystis* [22]. FutA2 is an iron transport system substrate-binding protein that facilitates iron transport into cytoplasm across the plasma membrane. The high abundance of this protein in the culture medium either suggests that FutA2 may have a moonlighting function in the extracellular milieu, or is simply a concomitant result of the leakage of periplasmic proteins that is usually observed in gram-negative bacteria [39,40]. The moonlighting function can be partially supported by the exported protein FagB, a component of iron transport system and one of the major virulence-associated factors in pathogenic bacterium *Corynebacterium pseudotuberculosis* [41]. Taken together, the observations suggest that a limited number of cell surface or periplasmic proteins are dominant in the extracellular milieu of *Synechocystis* under the current culture condition, and the extent of the dominance can be estimated that there is almost a HlyA or FutA2 out of every two protein molecules in the extracellular milieu.

The two phycobilisomal proteins phycocyanin alpha subunit (CpcA) and beta subunit (CpcB), which are the most abundant proteins in *Synechocystis* and were estimated to constitute 10-20% of the dry weight of *Synechocystis* [42], were also identified as the component of the exoproteome (Figure 1 and Figure 2). The two proteins are well known to function mainly on the thylakoid membrane, and their presence in the cell culture is more likely due to the lysis of dead cells



**Figure 1: The protein profiles of the whole cell lysate (WCL) and the extracellular proteins (EXO) of *Synechocystis* displayed by SDS-PAGE and Coomassie blue staining.** The amount of proteins loaded for the lane EXO is nearly one third of that of the WCL because of the difficulty in preparing large amount of extracellular proteins. The most dominant band indicated in the WCL lane contains the two most abundant proteins with nearly identical apparent molecular weight in *Synechocystis*, which are phycobilisome subunits CpcA and CpcB. Note that the other non-relevant lanes between WCL and EXO were cropped out.



**Figure 2: Ranking the relative abundance and the compositional analysis of the exoproteome.** The scatter plot shows the relative abundances of all identified extracellular proteins as represented by both logarithm-transformed spectral counts (x-axis) and peptide peak intensities. The predicted secretory or known periplasmic proteins are indicated as shown.

though CpcA is predicted to be secretable. Nevertheless, the relative abundances of the two proteins in the extracellular milieu is very low compared with those in the whole cell lysate, as the amount of the two proteins are only 1.2% and 1.5%, respectively, of the total amount of extracellular proteins estimated by the peptide peak intensities (Supplemental Table 1). The observation suggests that

the contribution of cell lysis to the exoproteome is minimal, though unavoidable.

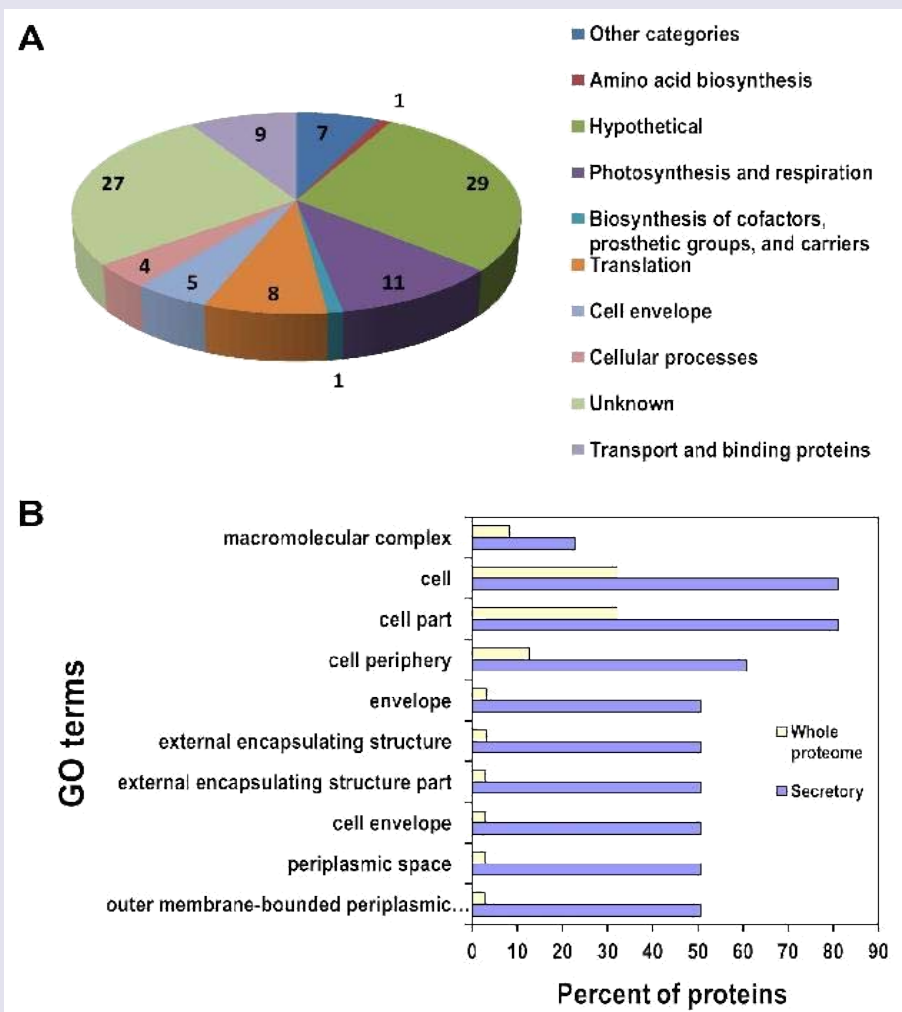
### Large Scale Analysis of the Sources of Origins of the Extracellular Proteins

In addition to cell lysis during culturing or harvesting, shedding from cell surface and leakage of periplasmic proteins are intuitively expected to be the two major sources for extracellular proteins. To reach the cell surface or the periplasm, proteins must be transported across the inner cell membrane via different secretion pathways [32,43,44]. Thus, we used multiple secretion-predicting algorithms, including SignalP [30], LipoP [31], and SecretomeP [32], to predict proteins that are potentially secreted via distinct classic or non-classic secretion pathways. In total, 34, 57, and 79 proteins were predicted to be secretory proteins by the three algorithms, respectively. In addition, the periplasm proteome was also previously identified [22], including 38 proteins identified in the current study. In total, after removal of the redundancy of the overlapping proteins among different predictions

and the previous identification, 102 proteins identified in the current study, as displayed in the scatter plot showing both peptide spectral counts and peak intensities (Figure 2 and Supplemental Table 1), can be considered as the periplasmic or cell surface proteins based on their predicted ability in secretion or known localization in periplasm. The sum of the peptide peak intensities for such proteins is about 88% of the total peak intensities for all proteins, confirming that the major compositional proteins in the culture medium are originated directly or indirectly from secretion. The sum of peptide peak intensities for all non-secretory proteins is only about 12%. These proteins are more likely originated from lysis of dead cells or excretion via unknown pathways.

### Functional Categorization of the Exoproteome of *Synechocystis*

To investigate the functional composition of the identified exoproteome of *Synechocystis*, we grouped all predicted secretory proteins according to their functions annotated by the CyanoBase



**Figure 3: Functional analysis of predicted secretory or known periplasmic proteins.** (A) The pie chart represents the functional categorization of all predicted secretory and known or known periplasmic proteins according to the functions annotated by the CyanoBase. Shown is also the number of proteins in each category. (B) The bar chart shows the enriched GO terms in cellular components for the predicted or unknown periplasmic proteins. The enrichment analysis was performed with Fisher's exact test.



(Figure 3A) [18]. The top two functional categories with the greatest number of proteins identified are hypothetical, unknown and photosynthesis and respiration. Hypothetical and unknown proteins constitute nearly 50 percent of the total proteome of *Synechocystis*, thus it is not surprising that the two categories contain the largest number of identified proteins. To gain more functional information for the predicted secretory proteins, especially for the hypothetical and unknown proteins, we blasted the whole *Synechocystis* proteome against the NCBI non-redundant proteome database using Blast2GO and mapped the blast hits with GO terms and InterPro domains [33]. Functional enrichment analysis using Fisher's exact test revealed that there is no GO term in biological processes or molecular functions enriched in the predicted secretory proteins. However, a number of GO terms in cellular components including periplasmic space, cell envelope, and cell periphery are significantly enriched (Figure 3B), further confirming that the exoproteome are highly enriched with proteins from the cell surface or periplasmic space. It is notable that some proteins in photosynthesis are also predicted to be secretory including PsbU and PsbV (Figure 3A). PsbU and PsbV are known to contain signal peptide that facilitates the transport of the host protein to the luminal side of the thylakoid, the major subcellular compartment where they execute functions. The presence of the two proteins and the other photosynthetic proteins in the extracellular milieu suggest that they may have some unknown moonlighting functions, or simply because of cell lysis.

### Identification of Glycosylated Proteins

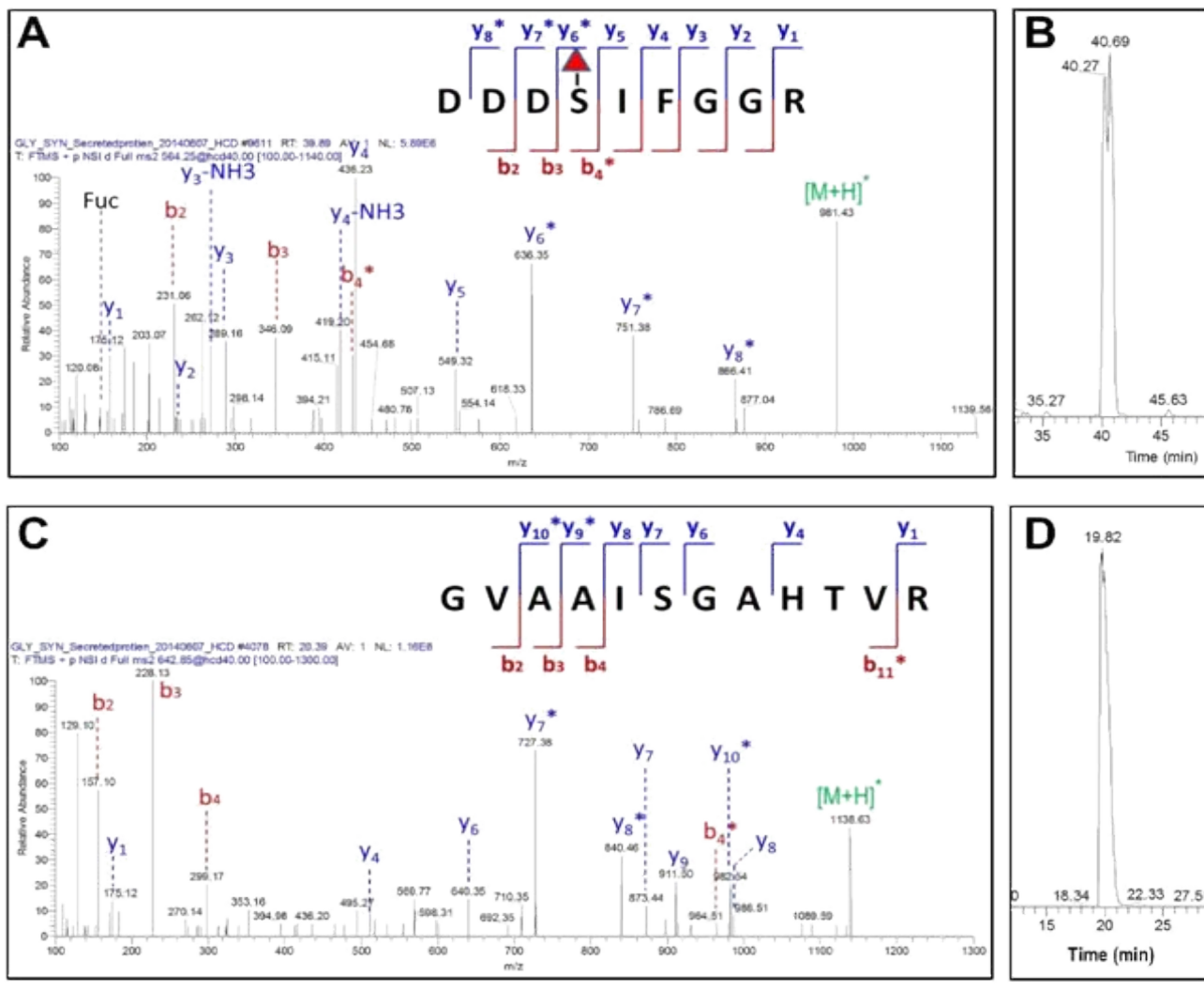
Glycosylation is a universal protein modification with important functions in eukaryotes and prokaryotes. The cyanobacteria are also known to contain glycosylated proteins including the pilin and the cell surface HlyA [35,45,46]. However, no glycosylation site of cyanobacterial proteins has been identified probably due to their overall low abundance in the whole cell lysate that is typically used for previous proteomic analysis. Bacterial cell surface proteins are usually the targets of glycosylation [47]. The specific enrichment of cell surface proteins in the exoproteome, as demonstrated by the high enrichment of HlyA, may facilitate the identification of novel glycoproteins and glycosylation sites without affinity-based pre-enrichment of glycoproteins. A fucose synthase Sll1213 was previously shown to be essential for the glycosylation of HlyA [48], implicating fucose as a potential sugar unit of modifying glycan. Also, Pilin, the major components of *Synechocystis* pilus, was also previously shown to be modified with a glycan chain composed of pentose and fucose [45]. Therefore, we re-searched the raw mass spectrometry data against the *Synechocystis* proteome database for the identification of potential glycosylated proteins using fucose and pentose as the modification group. As expected, two glycosylated proteins were identified with high confidence, including HlyA with an unambiguously assigned glycosylation site at S1196, which was conjugated with a fucose, as demonstrated by the annotated spectrum of the peptides bearing the modifying sugar (Figure 4A). Two other glycopeptides from HlyA and Sll1009 were also identified with high confidence as being conjugated with a fucose and a fucose-fucose, respectively, though the exact conjugation sites could not be unambiguously assigned (Figure 4C and Supplemental Figure 2). The observation suggests that fucose is an important building block of the glycan chains that modify the extracellular proteins of *Synechocystis*, at least for HlyA. Other

monomeric sugars, including pentose and hexose were not found to be conjugated directly to any extracellular proteins by the current database search. Intriguingly, the corresponding non-glycosylated forms of all identified glycopeptides in the current research were not identified, indicating that the glycosylated forms are predominant in the extracellular milieu. Indeed, the dominant peaks in the extracted ion chromatograms for the two glycopeptides from HlyA suggest that they are highly abundant (Figure 4, B and D).

### Discussion

The exoproteome of *Synechocystis* is one of the major components of the extracellular milieu of the photosynthetic organism, and it is expected to be functionally involved in the growth, survival, motility, toxicity, and potentially the direct or indirect regulation of photosynthetic activity of the cyanobacterium [11,49]. The proteome of *Synechocystis* has been extensively studied leading to the identification of more than 2,400 proteins under different growth conditions [16,17,23]. However, the study of the exoproteome of *Synechocystis* has been significantly lagged behind. Here, we systematically identified the exoproteome of the mixotrophically growing *Synechocystis* using high resolution mass spectrometry, and provided, for the first time, the general overview of the protein profile in the extracellular milieu of *Synechocystis*. The most dominant proteins in the exoproteome are predicted to be secretory proteins, whose promoters and the other cis or trans elements involved in the regulation of expression and secretion, could be potentially used in synthetic biology aiming at the generation of genetically modified *Synechocystis* strains to produce useful protein products such as peptide drugs with high efficiency and less cost. It is worth noting that the abundances of proteins in the exoproteome could be varied in different culture conditions, which may require quantitative mass spectrometry to measure the relative abundance of extracellular proteins across different culture conditions. This type of work may facilitate the identification of novel mechanisms that regulate the expression or secretion of the extracellular proteins.

In addition to the cell surface, the periplasm is another important pool that supplies proteins to the extracellular milieu, as supported by the identifications of 38 proteins that were previously identified in a large scale analysis of the isolated periplasmic proteins in which 57 most abundant proteins were identified [22]. Notably, the high abundant proteins in the periplasm including FutA2, CucA, and Prc are also high abundant in the extracellular milieu (Figure 2). This observation raises an outstanding question as to how the periplasmic proteins are transported to the extracellular milieu across the outer membrane of the gram negative cyanobacterium. Is the periplasmic leakage, a random process, taking the major role in transporting these proteins or are they transported via some specific export pathways? The current observation may favor the random periplasmic leakage as the major mechanism, because the simultaneous high abundance of these proteins in periplasm and the extracellular milieu can be more easily achieved by the random periplasmic leakage. If this is the case, we may expect that any environment changes that affect the abundances of periplasmic proteins, as demonstrated by the induced changes in abundances of periplasmic proteins in *Synechocystis* grown in high salts [22], can similarly affect the abundances of extracellular proteins. Nevertheless, this speculation need to be further proved experimentally, and detailed investigations are still necessary to study the transport mechanism for each individual protein.



**Figure 4: Identification of two glycopeptides from HlyA.** Shown are the annotated mass spectrum of the two glycopeptides (A and C) and the corresponding extracted ion chromatograms (B and D). The peaks labeled in green represent the precursor ions with a neutral loss of a fucose. Fuc indicates the diagnostic peak of fucose, which is also represented by the red triangle in the annotated peptide sequences. \* indicates fragment ions with a neutral loss of a fucose.

One of the interesting observations in the current study is the extremely low identification rate of the tandem mass spectra, as indicated by the 3.6% identification rate of the 22,779 high resolution MS/MS spectra acquired in the current study versus the more than 30% identification rate for the whole cell lysate of *Synechocystis* that we usually observed. Using the alternative database search engine such as SEQUEST did not significantly improve the annotation of the spectra (Data not shown). Two major factors may account for the low identification rate. First, proteins in the extracellular milieu may be partially digested by some extracellular proteases, this can be supported by the existence of 5 proteins with proteolytic activity in the extracellular milieu, including Sll1751, Sll2001, Sll0915, Slr0659, and Slr1331 (Supplemental Table 1). Randomly and partially degraded proteins may generate non-typical tryptic peptides with MS/MS spectra difficult to be identified by the current searching algorithm. Second, the extracellular proteins may be heavily post-translationally modified that are difficult to be identified. This can also be supported

by the observation that HlyA is heavily glycosylated. Unfortunately, we currently have very limited information regarding the type of modifications that could occur on the extracellular proteins of *Synechocystis*, which prevents the identification of the modified peptides.

The other important finding is the high enrichment of glycosylated proteins in the extracellular milieu, as demonstrated by the heavy glycosylation of HlyA. We searched many raw mass spectrometry data files previously generated from the whole cell lysate of *Synechocystis* for the identification of glycosylated HlyA, but without a success, suggesting that the exoproteome is a better target for the identification of a certain subset of glycosylated proteins. Further studies including enrichment of glycosylated proteins with affinity purification may be necessary for the systematical identification of the glycosylated extracellular proteins.

In conclusion, our current work, for the first time, provided a

comprehensive catalog of the *Synechocystis* exoproteome with in-depth analysis of its composition. The identified extracellular proteins, glycosylated peptides, and their estimated relative abundance in the extracellular milieu can be served as a useful resource that will benefit the research communities in the fields of photosynthesis, environment sciences, and biotechnologies using cyanobacteria as the model or potential bioreactor, or in synthetic biology taking advantage of the regulatory elements of high abundant extracellular proteins for the expression and secretion of druggable peptides.

## References

1. Armengaud J, Christie-Oleza JA, Clair G, Malard V, Duport C (2012) Exoproteomics: exploring the world around biological systems. *Expert Rev Proteomics* 9: 561-575.
2. Montecucco C, Molgo J (2005) Botulinum neurotoxins: revival of an old killer. *Curr Opin Pharmacol* 5: 274-279.
3. Gallert C, Winter J (2005) Bacterial Metabolism in Wastewater Treatment Systems. In *Environmental Biotechnology: Concepts and Applications*. (Jördening, H. J., and Winter, J. eds), WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.
4. Hebraud M (2014) Analysis of *Listeria monocytogenes* subproteomes. *Methods Mol Biol* 1157: 109-128.
5. Cabrita P, Trigo MJ, Ferreira RB, Brito L (2014) Is the Exoproteome Important for Bacterial Pathogenesis? Lessons Learned from Interstrain Exoprotein Diversity in *Listeria monocytogenes* Grown at Different Temperatures. *OMICS* 18: 553-569.
6. Siljamaki P, Varmanen P, Kankainen M, Sukura A, Savijoki K, et al. (2014) Comparative exoprotein profiling of different *Staphylococcus epidermidis* strains reveals potential link between nonclassical protein export and virulence. *J Proteome Res* 13: 3249-3261.
7. Ghosh N, Goel AK, Alam SI (2014) Exoproteome analysis of a novel strain of *Bacillus cereus* isolated implicated in disease resembling cutaneous anthrax. *Infect Genet Evol* 22: 1-11.
8. Clair G, Lorphelin A, Armengaud J, Duport C (2013) OhrRA functions as a redox-responsive system controlling toxinogenesis in *Bacillus cereus*. *J Proteomics* 94: 527-539.
9. Vanden Bergh P, Heller M, Braga-Lagache S, Frey J (2013) The *Aeromonas salmonicida* subsp. *salmonicida* exoproteome: global analysis, moonlighting proteins and putative antigens for vaccination against furunculosis. *Proteome Sci* 11: 44.
10. Christie-Oleza JA, Armengaud J (2010) In-depth analysis of exoproteomes from marine bacteria by shotgun liquid chromatography-tandem mass spectrometry: the *Ruegeria pomeroyi* DSS-3 case-study. *Mar Drugs* 8: 2223-2239.
11. Christie-Oleza JA, Pina-Villalonga JM, Bosch R, Nogales B, Armengaud J (2012) Comparative proteogenomics of twelve *Roseobacter* exoproteomes reveals different adaptive strategies among these marine bacteria. *Mol Cell Proteomics* 11: M111 013110.
12. De Marais DJ (2000) Evolution. When did photosynthesis emerge on Earth? *Science* 289: 1703-1705.
13. Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EM, et al. (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 311: 1737-1740.
14. Kasting JF, Siefert JL (2002) Life and the evolution of Earth's atmosphere. *Science* 296: 1066-1068.
15. Goksoyr J (1967) Evolution of eucaryotic cells. *Nature* 214: 1161.
16. Gao L, Shen C, Liao L, Huang X, Liu K, et al. (2014) Functional proteomic discovery of Slr0110 as a central regulator of carbohydrate metabolism in *Synechocystis* species PCC6803. *Mol Cell Proteomics* 13: 204-219.
17. Wang Y, Sun J, Chitnis PR (2000) Proteomic study of the peripheral proteins from thylakoid membranes of the cyanobacterium *Synechocystis* sp. PCC 6803. *Electrophoresis* 21: 1746-1754.
18. Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, et al. (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res* 3: 109-136.
19. Xue Y, Zhang Y, Cheng D, Daddy S, He Q (2014) Genetically engineering *Synechocystis* sp. Pasteur Culture Collection 6803 for the sustainable production of the plant secondary metabolite p-coumaric acid. *Proc Natl Acad Sci U S A* 111: 9449-9454.
20. Liu J, Chen L, Wang J, Qiao J, Zhang W (2012) Proteomic analysis reveals resistance mechanism against biofuel hexane in *Synechocystis* sp. PCC 6803. *Biotechnol Biofuels* 5: 68.
21. Gronenberg LS, Marcheschi RJ, Liao JC (2013) Next generation biofuel engineering in prokaryotes. *Curr Opin Chem Biol* 17: 462-471.
22. Fulda S, Huang F, Nilsson F, Hagemann M, Norling B (2000) Proteomics of *Synechocystis* sp. strain PCC 6803. Identification of periplasmic proteins in cells grown at low and high salt concentrations. *Eur J Biochem* 267: 5900-5907.
23. Wang Y, Xu W, Chitnis PR (2009) Identification and bioinformatic analysis of the membrane proteins of *Synechocystis* sp. PCC 6803. *Proteome Sci* 7: 11.
24. Huang F, Hedman E, Funk C, Kieselbach T, Schroder WP, et al. (2004) Isolation of outer membrane of *Synechocystis* sp. PCC 6803 and its proteomic characterization. *Mol Cell Proteomics* 3: 586-595.
25. Huang F, Parmryd I, Nilsson F, Persson AL, Pakrasi HB, et al. (2002) Proteomics of *Synechocystis* sp. Strain PCC 6803: Identification of Plasma Membrane Proteins. *Mol Cell Proteomics* 1: 956-966.
26. Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using stage tips. *Nat Protoc* 2: 1896-1906.
27. Cox J, and Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26: 1367-1372.
28. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, et al. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* 10: 1794-1805.
29. Cox J, Mann M (2012) 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* 13: S12.
30. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8: 785-786.
31. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, et al. (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* 12: 1652-1662.
32. Bendtsen JD, Kiemer L, Fausboll A, Brunak S (2005) Non-classical protein secretion in bacteria. *BMC Microbiol* 5: 58.
33. Bluthgen N, Brand K, Cajavec B, Swat M, Herzel H, et al. (2005) Biological profiling of gene groups utilizing Gene Ontology. *Genome Inform* 16: 106-115.
34. Sergeenko TV, Los DA (2000) Identification of secreted proteins of the cyanobacterium *Synechocystis* sp. strain PCC 6803. *FEMS Microbiol Lett* 193: 213-216.
35. Trautner C, Vermaas WF (2013) The slr1951 gene encodes the surface layer protein of *Synechocystis* sp. strain PCC 6803. *J Bacteriol* 195: 5370-5380.
36. Awram P, and Smit J (1998) The *Caulobacter crescentus* paracrystalline S-layer protein is secreted by an ABC transporter (type I) secretion apparatus. *J Bacteriol* 180: 3062-3069.
37. Thompson SA, Shedd OL, Ray KC, Beins MH, Jorgensen JP, et al. (1998)

- Campylobacter fetus surface layer proteins are transported by a type I secretion system. *J Bacteriol* 180: 6450-6458.
38. Pribyl T, Moche M, Dreisbach A, Bijlsma JJ, Saleh M, et al. (2014) Influence of impaired lipoprotein biogenesis on surface and exoproteome of *Streptococcus pneumoniae*. *J Proteome Res* 13: 650-667.
39. Dixon RA, Chopra I (1986) Leakage of periplasmic proteins from *Escherichia coli* mediated by polymyxin B nonapeptide. *Antimicrob Agents Chemother* 29: 781-788.
40. Rinas U, Hoffmann F (2004) Selective leakage of host-cell proteins during high-cell-density cultivation of recombinant and non-recombinant *Escherichia coli*. *Biotechnol Prog* 20: 679-687.
41. Pacheco LG, Slade SE, Seyffert N, Santos AR, Castro TL, et al. (2011) A combined approach for comparative exoproteome analysis of *Corynebacterium pseudotuberculosis*. *BMC Microbiol* 11: 12.
42. Katoh T (1988) Phycobilisome Stability. *Methods in enzymology* 167.
43. Natale P, Bruser T, and Driessen AJ (2008) Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—distinct translocases and mechanisms. *Biochim Biophys Acta* 1778: 1735-1756.
44. Pugsley AP (1993) The complete general secretory pathway in gram-negative bacteria. *Microbiol Rev* 57: 50-108.
45. Kim YH, Kim JY, Kim SY, Lee JH, Lee JS, et al. (2009) Alteration in the glycan pattern of pilin in a nonmotile mutant of *Synechocystis* sp. PCC 6803. *Proteomics* 9: 1075-1086.
46. Sakiyama T, Ueno H, Homma H, Numata O, Kuwabara T (2006) Purification and characterization of a hemolysin-like protein, SII1951, a nontoxic member of the RTX protein family from the Cyanobacterium *Synechocystis* sp. strain PCC 6803. *J Bacteriol* 188: 3535-3542.
47. Nothaft H, Szymanski CM (2010) Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* 8: 765-778.
48. Mohamed HE, van de Meene AM, Roberson RW, Vermaas WF (2005) Myxoxanthophyll is required for normal cell wall structure and thylakoid organization in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J Bacteriol* 187: 6883-6892.
49. Bhaya D, Bianco NR, Bryant D, Grossman A (2000) Type IV pilus biogenesis and motility in the cyanobacterium *Synechocystis* sp. PCC6803. *Mol Microbiol* 37: 941-951.

## Acknowledgements

This work was supported by the grants 2012CB910900, and 2011CB915402 from the Ministry of Science and Technology of China and the grant 31300677 (to L.G.) from the National Science Foundation of China.