ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

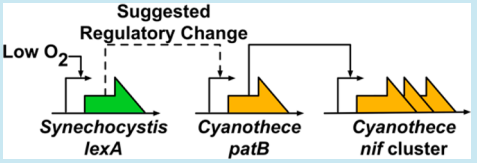
Identifying Regulatory Changes to Facilitate Nitrogen Fixation in the Nondiazotroph *Synechocystis* sp. PCC 6803

Thomas J. Mueller,[†] Eric A. Welsh,[‡] Himadri B. Pakrasi,[§,∥] and Costas D. Maranas*[,†]

[†]Department of Chemical Engineering, Pennsylvania State University, University Park, Pennsylvania 16801, United States
[‡]Cancer Informatics Core, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida 33612, United States
[§]Department of Energy, Environmental, and Chemical Engineering, Washington University, St. Louis, Missouri 63130, United States
[∥]Department of Biology, Washington University, St. Louis, Missouri 63130, United States

Supporting Information

ABSTRACT: The incorporation of biological nitrogen fixation into a nondiazotrophic photosynthetic organism provides a promising solution to the increasing fixed nitrogen demand, but is accompanied by a number of challenges for accommodating two incompatible processes within the same organism. Here we present regulatory influence networks for two cyanobacteria, *Synechocystis* PCC 6803 and *Cyanothece* ATCC 51142, and evaluate them to co-opt native transcription factors that may be used to control the *nif* gene cluster once it is transferred to *Synechocystis*. These networks were further examined to identify candidate transcription factors for other metabolic processes necessary for temporal separation of photosynthesis and nitrogen fixation, glycogen catabolism and cyanophycin synthesis. Two transcription factors native to *Synechocystis*, LexA and Rcp1, were identified as promising candidates for the control of the *nif* gene cluster and other pertinent metabolic processes, respectively. Lessons learned in the incorporation of nitrogen fixation into a nondiazotrophic prokaryote may be leveraged to further progress the incorporation of nitrogen fixation in plants.



KEYWORDS: *cyanobacteria, computational biotechnology, gene regulatory network, modeling, Synechocystis, Cyanothece*

Global use of nitrogen fertilizer has seen a 7-fold increase from 1960 to 1995, and further increases are predicted barring substantial increases in fertilizer efficiency.[1] Increasing levels of biological nitrogen fixation can replace a portion of the required fertilizer, decreasing nitrogen demands.[2,3] Some plants, such as legumes, have formed symbiotic relationships with rhizobium in order to acquire fixed nitrogen,[4] whereas other plants must acquire fixed nitrogen from the environment.[5] The inclusion of nitrogen fixation in a photosynthetic organism requires strict regulation of the oxygen levels within the organism, as nitrogenase is irreversibly inhibited by oxygen.[6] Our goal in this paper is to identify how to provide the regulatory structures to control the inherently incompatible processes of photosynthesis and nitrogen fixation.

Cyanobacteria are photosynthetic prokaryotes that serve as a promising test system in transferring nitrogen-fixing capabilities, as both diazotrophic and nondiazotrophic species exist. Some cyanobacteria, such as *Synechocystis* PCC 6803 (hereafter referred to as *Synechocystis*), require fixed nitrogen from the environment,[7] whereas others, such as *Cyanothece* ATCC 51142 (hereafter referred to as *Cyanothece*) and *Anabaena* PCC 7120, are diazotrophic.[8,9] In order to accommodate both nitrogen fixation and photosynthesis, diazotrophic cyanobacteria separate processes either spatially or temporally.[6] Organisms such as *Anabeana* use specialized cells called heterocysts to physically separate nitrogen fixation from cells performing photosynthesis.[9] Others, such as *Cyanothece*, temporally separate nitrogen fixation and photosynthesis.[8] *Cyanothece* generates glycogen during the day and then consumes it through a burst of respiration in the early dark hours, generating both the energy necessary for nitrogen fixation as well as the requisite anaerobic environment.[10] Therefore, on a standard 12 h/12 h light-dark cycle, the nitrogen fixation genes are highly overexpressed during the dark period under anoxic conditions.[8]

Synechocystis and *Cyanothece* have been extensively studied compared to other cyanobacteria. When both organisms are grown under constant light at 30 °C and non-nitrogen fixing conditions, the doubling times vary from approximately 24 h for *Synechocystis*[11] to 45 h for *Cyanothece*.[12] *Synechocystis* was the first cyanobacterium to have its genome sequenced[13] and has served as a genetically tractable model organism for cyanobacteria.[14] A number of its regulatory genes have been further investigated[15−17] and the structure of segments of the regulatory network has been proposed.[18] While *Synechocystis* is a more common target of research, recent work has compiled both transcriptomic and proteomic data for *Cyanothece*.[8,19] In addition, there have been significant efforts spent elucidating the metabolism of both *Cyanothece* and several closely related strains[11,12,20,21] and the regulatory network of *Cyanothece*.[22] The regulatory network presented here was generated in order

Received: October 15, 2015
Published: December 21, 2015

DOI: 10.1021/acssynbio.5b00202
ACS Synth. Biol. 2016, 5, 250−258

to facilitate comparison between the two organisms. Given the close relationship in *Cyanothece* between nitrogen fixation and its diurnal oscillations, any comparisons between different organisms focusing on nitrogen fixation must trace the light-dark cycle.

Synechocystis will be used as a proof of concept for the incorporation of temporal separation of photosynthesis and nitrogen fixation in a nondiazotrophic organism. *Cyanothece* is a logical source for the nitrogen fixation genes given its close phylogenetic relationship to *Synechocystis* and the fact that it contains the largest contiguous *nif* cluster in cyanobacteria.[10]

In order to identify candidate regulators for the processes necessary to incorporate nitrogen fixation into *Synechocystis,* we generated, examined, and compared regulatory networks for both *Synechocystis* and *Cyanothece*. We identified transcription factors from both *Synechocystis* and *Cyanothece* that could be co-opted to control these processes in *Synechocystis*. Two of these identified transcription factors, LexA and Rcp1, are native to *Synechocystis* and have both expression profiles and environmental responses that make them compelling candidates for controlling the *nif* cluster and other pertinent metabolic processes, respectively. Cyanobacteria have been used as a platform system for investigating CO_2 concentrating mechanisms in plants.[23,24] Therefore, lessons learned in manipulating regulatory structures in cyanobacteria could ultimately inform the incorporation of nitrogen fixation in plants.

RESULTS AND DISCUSSION

Comparison of Peak Expression Timing. The regulatory networks for both *Synechocystis* PCC 6803 (*Synechocystis*) and *Cyanothece* ATCC 51142 (*Cyanothece*) were generated by minimizing the extent of experimental expression variance not explained for a fixed number of regulatory influences. The number of influences for each gene cluster was chosen as the minimum number of influences that would allow for over 50% of the expression change to be explained. The *Synechocystis* network contains 45 gene clusters, 17 transcription factor (TF) clusters, and an average degree (*i.e.*, TF clusters affecting a gene cluster) of 5. The *Cyanothece* network contains 221 gene clusters, 16 TF clusters, and has an average degree of 2.23 (Table 1). Parts of the networks are shown in Figure 1. The regulatory influences of each TF cluster on each gene cluster can be found in Supplemental File S1. This difference in average degree could be caused in part by the differing methods used to process the original transcriptomic data.

Table 1. Statistics for Both Regulatory Networks

	Synechocystis	*Cyanothece*
Genes	1,423	1,552
Gene clusters	45	221
TFs	76	72
TF clusters	17	16
Average degree	5	2.23
Average % explained expression change	54.25%	60.70%

In order for a nondiazotrophic organism such as *Synechocystis* to temporally separate nitrogen fixation and photosynthesis, a number of genes must have expression profiles similar to those of their *Cyanothece* counterparts. The genes in these two organisms that have differing expression profiles extends past those involved in nitrogen fixation and its associated processes. Performing a bidirectional BLAST search on the set of cycling *Cyanothece* genes identified 306 homologue pairs between the two organisms where both genes were in the set of cycling genes. Only 114 of these pairs have elevated expression for both genes at the same segment of the light-dark cycle. Another 77 peaked at different times during the same period (*e.g.*, a gene peaks in the early light and its homologue peaks in the late light). Figure 2 shows the distribution of the 115 homologue pairs that peak in different periods (*i.e.*, one homologue peaks in the dark period and the other in the light). Of the 14 homologue pairs with differing peaks categorized in photosynthesis and respiration, six were associated with respiratory electron flow and two with phycobilisomes.[25] These unmatched peaks can be explained by the different temporal pattern in glycogen consumption and use of phycobilisomes in the two organisms.[26] Given the number of differences in expression across the genome, we focused only on those genes directly associated with nitrogen fixation or the metabolic processes recruited before or after nitrogen fixation.

Candidates for Controlling *nif* Cluster Expression. Several different avenues are possible for imposing the necessary regulatory action on the genes involved in these processes. By using native transcription factors, the number of genes that would need to be transferred is minimized and the uncertainty in the expression profile of the transcription factor is reduced, as many of its regulatory partners are already present. Candidate native TFs were identified by determining the gene clusters in *Synechocystis* with the appropriate expression profile for this process, and by investigating the TF clusters that were identified as regulatory influences for a number of these gene clusters. Some of these identified TFs have expression profiles similar to that of the final desired expression profile. Controlling gene expression using a transcription factor with the same expression profile is a concept underlying the development in gene circuits.[27] Gene circuits developed using TFs have been designed to respond to environmental cues (*e.g.*, oxygen levels).[28] The strategy outlined here, instead, aims to identify TFs native to the organism that already respond to environmental conditions (*e.g.*, low oxygen, absence of light, *etc.*) that subsequently can be coopted to control the genes of interest. Examining the *Cyanothece* regulatory network and the regulatory influences predicted for the genes associated with the process identified additional candidate regulators. Using these regulators would require the transfer of the regulatory genes to *Synechocystis* but reduce the need to modify promoter regions for the genes transferred from *Cyanothece*.

The expression profile necessary for the transferred *nif* genes is characterized by an increase in expression during the dark and a rapid decrease in expression at the beginning of illumination. The desired profile can be seen in gene clusters 4, 10, 16, 17, 18, 19, and 35 of *Synechocystis*. There are four TF clusters (10, 15, 16, and 17) that are regulatory influences for at least three of these gene clusters. Each TF cluster contains a single gene that has been shown to respond to environmental conditions pertinent to nitrogen fixation. TF cluster 10 contains gene *sll1689*, which codes for the group 2 sigma factor *sigE*. Both the transcript levels of *sigE* and the protein levels of the associated SigE protein increase under nitrogen depletion.[29] Studies have shown that *sigE* expression increases in an NtcA dependent manner.[29] NtcA, a global nitrogen regulator in cyanobacteria,[30] responds to intracellular 2-oxoglutarate levels and regulates a number of genes involved in nitrogen assimilation.[31] In *Cyanothece* ATCC 51142 (*Cyanothece*), it

DOI: 10.1021/acssynbio.5b00202
ACS Synth. Biol. 2016, 5, 250−258

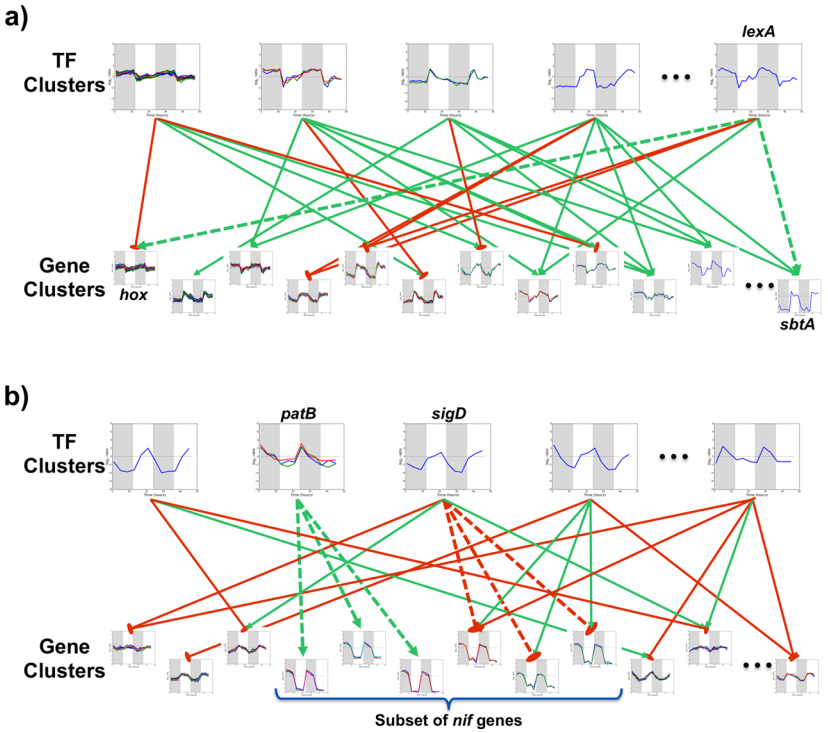


Figure 1. Parts of the regulatory networks for (a) *Synechocystis* and (b) *Cyanothece*. Edges are colored based on whether the predicted interaction is activating (green) or inhibiting (red). A subset of the gene clusters containing *nif* genes is shown above, all regulatory influences on the nif genes are listed in Supplemental File S2. Dashed lines indicate edges with associated literature support discussed in the Results section.

has been shown that *ntcA* expression is out of phase with that of the *nif* genes,[32] and has been suggested that *ntcA* is involved primarily in nitrogen assimilation rather than in nitrogen fixation.[33] Nevertheless, both *ntcA* and *sigE* do not account for the presence of oxygen, a nitrogenase inhibitor, through detection of either light or intracellular oxygen levels. Therefore, additional synthetic regulation is needed to account for such factors.

The remaining three TF clusters contain TF genes that respond to these environmental conditions. TF cluster 17 contains the *hik16* (*slr1805*) gene. While *hik16* is most commonly associated with the salt stress response,[34] it also responds to oxidative stress induced by hydrogen peroxide.[35] The decreased expression of *hik16* during the day may be explained by the oxidative stress that can be caused by reactive oxygen species generated by photosynthesis.[36] Given that it responds to additional stimuli, namely salt stress, *hik16* would not be an ideal native TF for use in *nif* cluster regulation. Both TF clusters 15 and 16 contain genes (*lexA* and *rre34*, respectively) that respond to oxygen levels within the environment. In *Synechocystis*, LexA binds to two different sites on the hydrogenase operon and is also known to regulate the *sbtA* gene, which codes for a sodium-bicarbonate symporter.[37] Both of these regulatory influences are predicted in the *Synechocystis* network (Figure 1a). LexA is thought to function as a transcriptional activator[16] and is upregulated after the transition to a low O_2 environment.[38] Similarly, *rre34* (*sll0789*) was shown to be upregulated under low O_2 conditions.[38] Given the irreversible inhibition of nitrogenase by oxygen, a transcriptional activator that is upregulated under low O_2 conditions is a promising candidate for use in regulating the *nif* cluster in *Synechocystis*. When comparing the expression profiles of the two candidates in *Synechocystis* with that of the *nif* cluster in *Cyanothece* (Figure 3) we see that *lexA* peaks at the same time as the *nif* cluster as compared to the later peak in expression for *rre34*. This places *lexA* as the top candidate for regulating the *nif* cluster in a diazotrophic *Synechocystis*.

The *Cyanothece* regulatory network identifies a number of transcription factors involved in the regulation of genes in the *nif* cluster. Thirty-three genes in the *nif* cluster were identified as cycling and were contained within 13 different gene clusters (Supplementary File S2). Many of these gene clusters have different sets of regulatory influences although several TF clusters are prominent (Supplementary File S3). Twenty-seven of the 33 genes have either TF cluster 4 or TF cluster 15 as a regulatory influence. TF cluster 4 contains three genes, one of which is *cce_1898*, annotated as *patB*. In the previous regulatory network of *Cyanothece*, PatB was predicted to influence the expression of the nitrogenase complex.[22] PatB was also shown to be required for diazotrophic growth, and is expressed in the heterocysts of *Anabaena* PCC 7120.[39] TF cluster 15 is composed of only one gene, *sigE*. SigE has been shown to be upregulated in both *Synechocystis* and *Anabeana* PCC 7120 under nitrogen depletion.[9,29] TF cluster 10, another prominent regulator, contains *sigD* and influences 8 genes. The deletion of *sigD* in *Anabeana* impairs the ability of the organism to establish diazotrophic growth[40] (Figure 1b). Given the prediction that PatB controls *nif* cluster expression, the introduction of *patB* into *Synechocystis* and the use of LexA to control its expression is another possible approach to controlling the *nif* cluster expression in *Synechocystis* (Figure 4a). By using the native *nif* cluster regulator, this strategy circumvents the need to modify all promoter regions in the *nif* cluster individually and also retains the native control of the *nif* cluster by PatB.

Candidates for Regulating Other Pertinent Metabolic Processes. In addition to nitrogen fixation two other processes, glycogen catabolism and cyanophycin synthesis, must be active at specific times for the diazotrophic

DOI: 10.1021/acssynbio.5b00202
ACS Synth. Biol. 2016, 5, 250−258

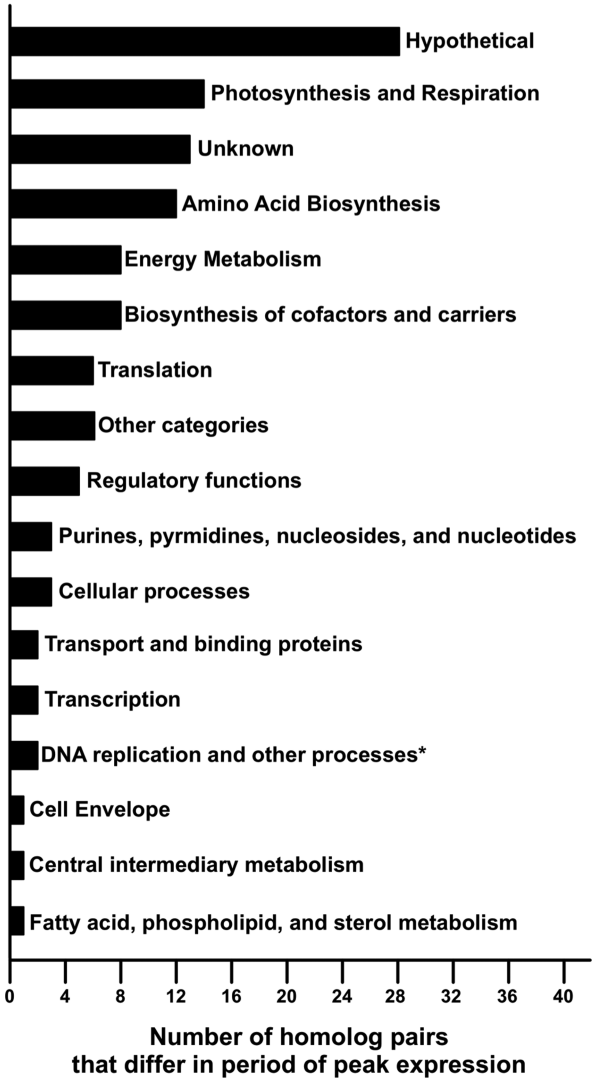


Figure 2. Distribution of homologue pairs in which one peaks in the light period and one peaks in the dark between the different subsystems. Subsystem annotations from Saha *et al.* were used.[45] *"DNA replication and other processes" refers to the subsystem "DNA replication, restriction, modification, recombination, and repair".

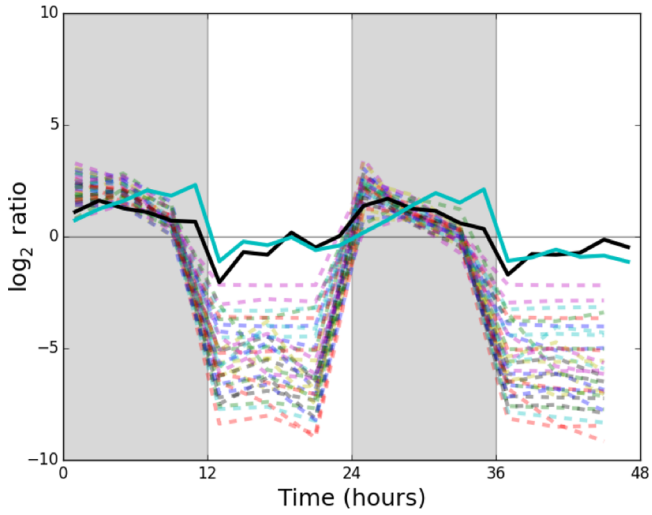


Figure 3. Comparison of candidate native transcription factor expression profiles in *Synechocystis* (*lexA*: black line, *rre34*: cyan line) with those of the *nif* cluster in *Cyanothece* (dashed lines). The dark periods of the light-dark cycle (hours 0−12 and 24−36) are represented with gray shading.

Synechocystis to create the requisite anaerobic environment and to assimilate and store newly fixed nitrogen. In *Cyanothece* the anaerobic environment and energy necessary for nitrogen fixation are generated through the consumption of glycogen in the early dark period.[8] Glycogen catabolism in *Synechocystis* is a metabolic process that must be placed under different regulatory control. There is a pair of homologues (*cce_1629* and *slr1367*) within the two organisms that are annotated as glycogen phosphorylases and identified as cycling. The current expression profile in *Synechocystis* includes a peak in expression 1 h before the onset of light, and a negative $\log_2$ ratio 1 h into the dark period, when its homologue peaks in *Cyanothece* (Figure 5a). This is in contrast to the glycogen synthase genes, which peak in the beginning of the light period for both organisms, and does not require any modified regulation. Using the steps discussed to identify native transcription factors for the *nif* cluster, two gene clusters, 16 and 34, were found to have the desired profile; low expression during the light and peak expression in the early dark. These two clusters share four regulatory influences, the most significant of these is TF cluster 12, which is composed of the *rcp1* (*slr0474*) gene. The Rcp1 protein is part of a two-component light sensing system with the phytochrome Cph1.[41] Garcia-Dominguez *et al.* showed that *rcp1* is maximally expressed 15 min into the dark period, and its expression levels are almost undetectable 5 min after illumination.[42] Given the role that glycogen catabolism plays in consuming oxygen within the cell, its associated genes should be expressed prior to those for nitrogen fixation, making *rcp1* an ideal candidate for controlling glycogen phosphorylase. The *Cyanothece* network identifies the regulatory influences for glycogen phosphorylase as a hypothetical protein (*cce_4141*) and *sigD*, which is involved in establishing diazotrophic growth in *Anabeana*.[40]

Cyanophycin is a dynamic nitrogen reserve composed of equimolar quantities of arginine and aspartic acid.[26] Cyanophycin accumulation in *Cyanothece* follows the fixation of nitrogen, being mainly generated between two and 6 h after the beginning of the dark period.[26] In contrast, cyanophycin has a minor role in *Synechocystis*, which, instead, degrades phycobilisomes, leading to cell bleaching.[26] Therefore, arginine, aspartic acid, and cyanophycin synthesis in *Synechocystis* should be expressed in a manner similar to that of their counterparts in *Cyanothece*. The genes involved in arginine and aspartic acid synthesis were identified through the gene-protein-reaction relationships of the reactions that carry flux at maximum biomass production for the genome-scale models *i*Syn731 and *i*Cyt773 (Figure 5c).[11] For each reaction, the elevated expression time for each associated gene was compared. Several of the gene pairs shared elevated expression times, although four genes in *Synechocystis* (*sll0573*, *slr0585*, *slr1133*, and *sll0902*) either were not cycling, or had elevated expression during the beginning of the light period. Conversely, their *Cyanothece* counterparts exhibited elevated expression during the beginning of the dark period (Figure 5b). Cyanophycin synthetase catalyzes the elongation reaction of cyanophycin through the incorporation of arginine and aspartic acid.[43] The gene that codes for this enzyme in *Synechocystis*, *slr2002*, is not cycling and therefore will need to be controlled by a different TF. The cyanophycin synthetase gene in *Cyanothece*, *cce_2237*,

DOI: 10.1021/acssynbio.5b00202
ACS Synth. Biol. 2016, 5, 250−258

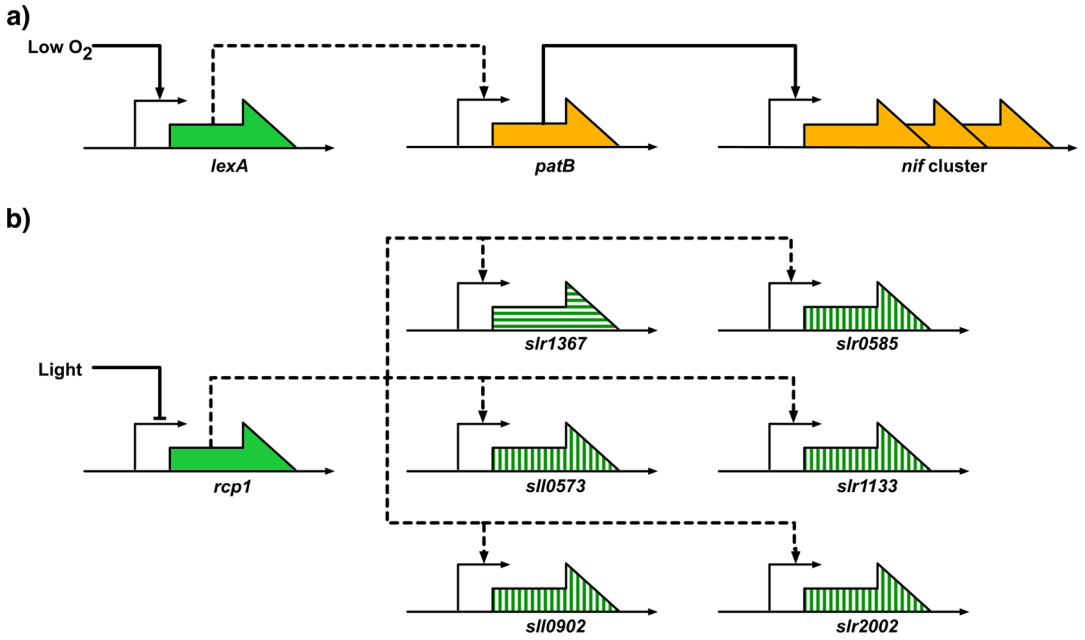


Figure 4. All proposed regulatory changes. Genes native to *Synechocystis* are in green; those from *Cyanothece* are in orange. Proposed regulatory changes are represented as dashed lines. (a) Proposed strategy for control of the *nif* cluster in *Synechocystis* using *lexA* and *patB*. (b) Strategies for control of glycogen catabolism (horizontal stripes) and cyanophycin synthesis (vertical stripes) genes in *Synechocystis* using *rcp1*.

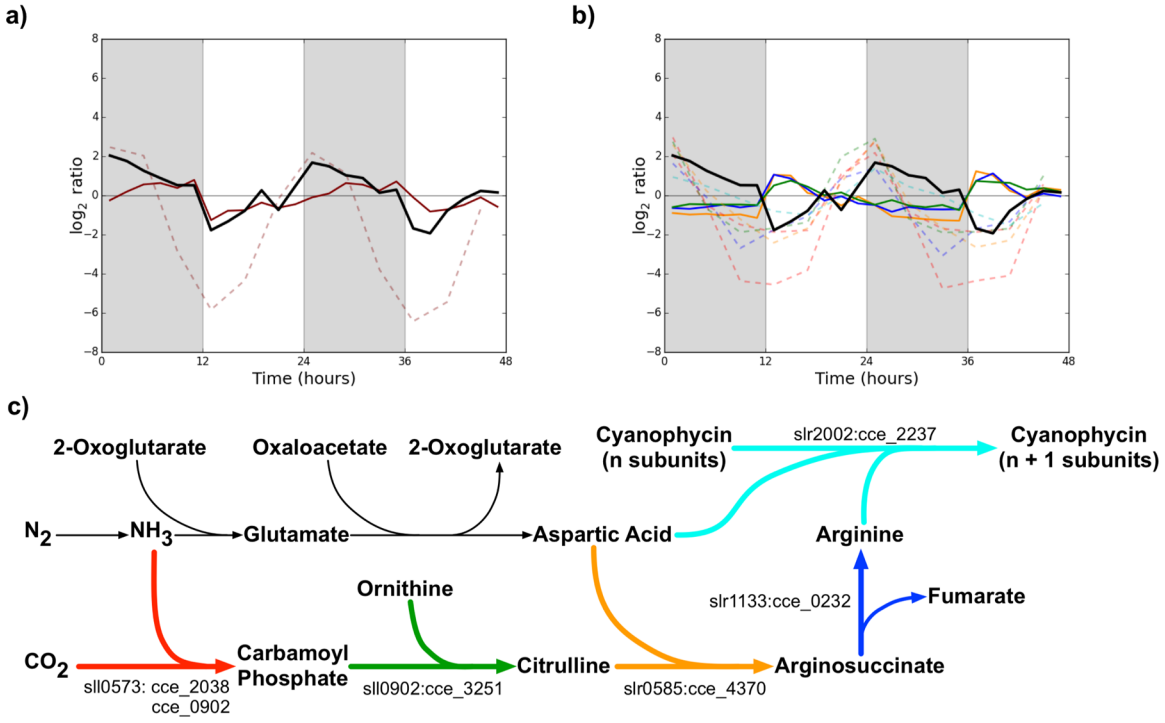


Figure 5. Metabolic processes requiring modified regulation in a diazotrophic *Synechocystis*. (a) Expression profiles of *rcp1* (solid black line) and glycogen phosphorylase (solid brown line) in *Synechocystis,* and glycogen phosphorylase (dashed brown line) in *Cyanothece*. (b) Expression profiles of *rcp1* (solid black line) and cycling cyanophycin synthesis genes (solid lines, colors corresponding to c) in *Synechocystis,* and the expression profiles of the cycling cyanophycin synthesis genes in *Cyanothece* (dashed lines). *Sll0573* and *slr2002* are not shown, as they were not identified as cycling. The dark periods of the light-dark cycle (hours 0−12 and 24−36) are represented with gray shading. (c) Reactions and genes identified by the *i*Syn731 and *i*Cyt773 genome-scale models as being involved in cyanophycin synthesis. Reactions with associated genes whose elevated expression times differ between the two organisms are color-coded.

peaks in expression at the beginning of the dark period, with above average expression continuing through the middle of the dark period. Given the similar timing of necessary expression, the native TF identified as candidates for *nif* cluster and glycogen catabolism regulation, namely, LexA and Rcp1, serve equally well as candidates for control of these genes. Two TF clusters, 3 and 13, regulate the six clusters containing these genes in *Cyanothece*. TF cluster 13 is composed of a FUR family

DOI: 10.1021/acssynbio.5b00202
ACS Synth. Biol. 2016, 5, 250−258

transcription factor, and TF cluster 3 contains *cce_1982*, which is the homologue of *rcp1* in *Synechocystis*, strengthening its candidacy for regulation of cyanophycin synthesis (Figure 4b).

This work focuses on the transcriptional regulation of the *nif* genes and the genes associated with several other key processes. Future work that considers allosteric regulation will be able to provide a more comprehensive set of possible modifications that allow for the incorporation of processes such as nitrogen fixation. The inclusion of additional forms of regulation would capture processes such as the allosteric regulation of NtcA activity through binding with 2-oxoglutarate.[44]

In this paper we elucidated regulatory influence networks for two cyanobacteria, *Synechocystis* PCC 6803 and *Cyanothece* ATCC 51142, generated using transcriptomic data taken during 12 h/12 h light-dark cycles. These networks were inspected to identify what regulation would need to be introduced or modified to incorporate nitrogen fixation into the non-diazotrophic *Synechocystis*. One native transcription factor, LexA, arose as a candidate for regulating *nif* cluster expression, while PatB was identified as a promising candidate to be a regulator of *nif* cluster genes in *Cyanothece* that could be transferred to *Synechocystis*. Using these predictions together, we postulate that transferring *patB* to *Synechocystis* and controlling its expression with LexA would lead to the desired *nif* cluster expression within *Synechocystis*. Other processes with required expression changes dictated by the temporal separation of photosynthesis and nitrogen fixation, specifically glycogen catabolism and cyanophycin anabolism, were analyzed. The native Rcp1 transcription factor arose as an additional candidate for the regulation of these processes within a newly diazotrophic *Synechocystis*. Given the expression profiles and response to environmental conditions, LexA and Rcp1 are transcription factors native to *Synechocystis* that should be considered for use as regulatory control of the *nif* cluster and the associated metabolic processes, respectively.

MATERIALS AND METHODS

Cycling and Homologous Gene Identification. Regulatory influence networks for both organisms were generated using time series transcriptomic data over a standard 12 h/12h light-dark cycle.[8,45] The *Cyanothece* ATCC 51142 (*Cyanothece*) data from Stockel *et al.*[8] was processed using the IRON software tool.[46] The $\log_2$ ratio of expression levels at each time point to the pooled samples was then calculated for each gene and time point. The criteria from Kucho *et al.*[47] were used to identify cycling genes in the *Synechocystis* PCC 6803 (*Synechocystis*) data set[45] and applied to the *Cyanothece* data set as well. The *Synechocystis* data used was not normalized in the range of −1 to +1 in order to be consistent with the *Cyanothece* data from Stockel *et al.*[8] The *nif* genes qualified as cycling using two of the three criteria from Kucho *et al.* (*i.e.*, appropriate period length and statistically significant differences between peak and trough expression each day) but the values of the error function described in Kucho *et al.* were above the specified cutoff. Given the importance of including these genes within the network for identifying regulation on the *nif* cluster, only the period length and statistically significant differences in peak and trough expression were used to identify cycling genes. This resulted in a set of 1552 genes identified as cycling in *Cyanothece* as compared to the 1423 cycling genes in *Synechocystis* (Supplementary File S4).

To identify homologue pairs a bidirectional BLAST search between the two organisms was performed using an e value cutoff of 10^{-2}. A result was determined to be a hit if the alignment length was at least 75% of the length of both the query and target genes, and if the raw score was at least 25% of the self-hit score of the query gene. A homologue pair was identified if both genes had only each other as a hit. A gene was defined as having elevated expression if the $\log_2$ ratio at a time point was at least 90% of the maximum $\log_2$ ratio for that gene. To compare peak expression times for homologue pairs, elevated expression was then categorized as occurring in the beginning, middle, or final third of the light or dark period.

Cluster Generation. In order to simplify the system before inferring regulatory influences, hierarchical clustering using Euclidean distances was implemented for both organisms. First the optimal agglomeration method available in R was identified for each organism using the cophenetic correlation.[48] For both organisms the unweighted pair group method with arithmetic mean (UPGMA)[49] method generated the dendrogram whose cophenetic distances most closely correlated with the previously calculated Euclidean distances. The SD validity index[50] was then used to identify the optimal height cutoff for both dendrograms, generating 45 gene clusters for *Synechocystis* and 221 gene clusters for *Cyanothece*.

The transcription factor (TF) clusters were generated in a similar manner. Genes with a regulatory function were first identified using literature sources. Of the 146 regulatory genes identified by Singh *et al.* in *Synechocystis*,[51] 76 are present in the set of cycling genes of Saha *et al.*[45] The cycling *Cyanothece* gene set contained 72 genes that were either annotated as having a regulatory function by Stockel *et al.*[8] or were identified as regulators by McDermott *et al.*[22] Dendrograms for these gene sets were then generated using the same UPGMA agglomeration method. The SD validity index was used to identify the optimal height cutoff for *Cyanothece,* and these 72 genes were clustered into 16 different TF clusters. The identified optimal height cutoff for the *Synechocystis* transcription factors generated only six TF clusters. The subsequently generated network did not adequately capture the expression of the gene clusters, as the percentage of experimental expression change explained by the network (see Network Formulation) ranged from 8.1 to 56.4% at maximum regulatory influences. A lower sampled height cutoff yielded 17 TF clusters, and was chosen to give a comparable number of TF clusters for the two networks. The members of each cluster for both organisms are listed in Supplemental File S5. All clustering and cluster analysis was performed using the R language along with the cluster and clusterCrit packages[52−54]

Network Elucidation Using Optimization Formulation. The network of regulatory influences for each organism was generated by iteratively determining for each gene cluster the maximum amount of experimental expression change that can be accounted for a given number of regulatory influences. J, K, and T denote the total number of gene clusters, TF clusters, and time points, respectively. The experimental expression level of gene cluster j at time point t is represented as X_{jt}, and the experimental expression level of TF cluster k at time point t is denoted as TF_{kt}. C_{kj} is the interaction coefficient that describes the influence of TF cluster k on gene cluster j.

Gene expression was assumed to be a linear additive contribution of the set of influencing TFs. This can be postulated for each gene cluster j and time point t as shown in eq 1.

DOI: 10.1021/acssynbio.5b00202
ACS Synth. Biol. 2016, 5, 250−258

$$\frac{dX_{jt}}{dt} = \sum_{k}^{K} C_{kj}\mathrm{TF}_{kt} \;\forall\, j \in J \,\&\, \forall\, t \in T \quad (1)$$

Backward finite difference was used to approximate the rate of expression term (*i.e.*, left-hand side) of eq 1 in eq 3 in the mixed integer linear program formulation shown below. Solving this formulation identifies the interaction coefficients (C_{kj}) that minimize the amount of change in gene cluster expression not explained by the regulatory influences. The formulation was solved iteratively for every gene cluster and a range of maximum regulatory influences.

$$\left[\begin{array}{ll} \min \sum_{j}^{J}\sum_{t}^{T-1} S_{jt}^{+} + S_{jt}^{-} & (2) \\ \text{s. t.} & \\ X_{jt+1} - X_{jt} - \Delta t \sum_{k}^{K} C_{kj}\mathrm{TF}_{kt+1} & (3) \\ \quad = S_{jt}^{+} - S_{jt}^{-} \;\forall\, t \in \{1, \ldots, T-1\} & \\ -My_{kj} \le C_{kj} \le My_{kj} \;\forall\, k \in K & (4) \\ \sum_{k}^{K} y_{kj} \le \text{max regulatory influences} & (5) \end{array}\right] \forall\, j \in J$$

The slack variables S_{jt}^{+} and S_{jt}^{-} represent the positive and negative deviations from experimental measurements. Eqs 4 and 5 use the binary variables, y_{kj}, to restrict the number of regulatory influences per gene cluster. If a transcription factor in TF cluster k is in gene cluster j then a regulatory interaction will be imposed for that TF cluster gene cluster pair by fixing y_{kj} to one for that pair.

In order to determine the number of regulatory influences per cluster, the percentage of experimental expression change accounted for by the network was calculated for the range from one to 11 regulatory influences for both organisms. For a given number of regulatory influences this percentage of explained expression variance is defined as the difference between the error (*i.e.*, the sum of the slack variables) with zero regulatory influences and the given number of regulatory influences, divided by the error with zero influences. The number of regulatory influences was chosen for each cluster as the minimum number of influences that brought the percentage of explained expression change above 50%. The percentage of explained expression change for each cluster at every number of regulatory influences can be found in Supplemental File S6.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.5b00202.

Regulatory influences for each gene cluster for both networks. (XLSX)

Annotation, gene cluster, and regulatory influences for the cycling genes within the *nif* cluster. (PDF)

Venn diagram depicting those transcription factors that are influences for *nif* genes and have cycling homologues in *Synechocystis*. (PDF)

Transcriptomic $\log_2$ ratios of cycling *Cyanothece* genes and cycling *Synechocystis* genes. (XLSX)

Members of each gene and TF cluster for *Synechocystis* and *Cyanothece*. (XLSX)

Percentage of explained expression change for each gene cluster for 1−10 TF influences. (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: costas@engr.psu.edu.

Author Contributions

TJM created and analyzed regulatory networks. EAW processed *Cyanothece* data. TJM and CDM wrote the paper. All authors read and approved the final manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Fuzhong Zhang and Tae Seok Moon for discussions during the writing of the manuscript. Supported by funding from the National Science Foundation, grant MCB-1331194.

■ REFERENCES

(1) Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., and Polasky, S. (2002) Agricultural sustainability and intensive production practices. *Nature 418*, 671−677.

(2) Peoples, M. B., Herridge, D. F., and Ladha, J. K. (1995) Biological Nitrogen-Fixation - an Efficient Source of Nitrogen for Sustainable Agricultural Production. *Plant Soil 174*, 3−28.

(3) Dobereiner, J. (1997) Biological nitrogen fixation in the tropics: Social and economic contributions. *Soil Biol. Biochem. 29*, 771−774.

(4) Gage, D. J. (2004) Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol Mol. Biol. Rev. 68*, 280−300.

(5) Kraiser, T., Gras, D. E., Gutierrez, A. G., Gonzalez, B., and Gutierrez, R. A. (2011) A holistic view of nitrogen acquisition in plants. *J. Exp. Bot. 62*, 1455−1466.

(6) Berman-Frank, I., Lundgren, P., and Falkowski, P. (2003) Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Res. Microbiol. 154*, 157−164.

(7) Herrero, A., Muro-Pastor, A. M., and Flores, E. (2001) Nitrogen control in cyanobacteria. *J. Bacteriol. 183*, 411−425.

(8) Stockel, J., Welsh, E. A., Liberton, M., Kunnvakkam, R., Aurora, R., and Pakrasi, H. B. (2008) Global transcriptomic analysis of Cyanothece 51142 reveals robust diurnal oscillation of central metabolic processes. *Proc. Natl. Acad. Sci. U. S. A. 105*, 6156−6161.

(9) Kumar, K., Mella-Herrera, R. A., and Golden, J. W. (2010) Cyanobacterial Heterocysts. *Cold Spring Harbor Perspect. Biol. 2*, a000315.

(10) Welsh, E. A., Liberton, M., Stoeckel, J., Loh, T., Elvitigala, T., Wang, C., Wollam, A., Fulton, R. S., Clifton, S. W., Jacobs, J. M., Aurora, R., Ghosh, B. K., Sherman, L. A., Smith, R. D., Wilson, R. K., and Pakrasi, H. B. (2008) The genome of Cyanothece 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. *Proc. Natl. Acad. Sci. U. S. A. 105*, 15094−15099.

(11) Saha, R., Verseput, A. T., Berla, B. M., Mueller, T. J., Pakrasi, H. B., and Maranas, C. D. (2012) Reconstruction and Comparison of the Metabolic Potential of Cyanobacteria Cyanothece sp ATCC 51142 and Synechocystis sp PCC 6803. *PLoS One 7*, e48285.

(12) Feng, X. Y., Bandyopadhyay, A., Berla, B., Page, L., Wu, B., Pakrasi, H. B., and Tang, Y. J. (2010) Mixotrophic and photoheterotrophic metabolism in Cyanothece sp ATCC 51142 under continuous light. *Microbiology 156*, 2566−2574.

(13) Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., and Tabata, S. (1996)

DOI: 10.1021/acssynbio.5b00202
ACS Synth. Biol. 2016, 5, 250−258

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.
(14) Mitschke, J., Georg, J., Scholz, I., Sharma, C. M., Dienst, D., Bantscheff, J., Voss, B., Steglich, C., Wilde, A., Vogel, J., and Hess, W. R. (2011) An experimentally anchored map of transcriptional start sites in the model cyanobacterium Synechocystis sp. PCC6803. *Proc. Natl. Acad. Sci. U. S. A. 108*, 2124−2129.
(15) Garcia-Dominguez, M., Reyes, J. C., and Florencio, F. J. (2000) NtcA represses transcription of gifA and gifB, genes that encode inhibitors of glutamine synthetase type I from Synechocystis sp PCC 6803. *Mol. Microbiol. 35*, 1192−1201.
(16) Gutekunst, K., Phunpruch, S., Schwarz, C., Schuchardt, S., Schulz-Friedrich, R., and Appel, J. (2005) LexA regulates the bidirectional hydrogenase in the cyanobacterium Synechocystis sp PCC 6803 as a transcription activator. *Mol. Microbiol. 58*, 810−823.
(17) Li, H., and Sherman, L. A. (2000) A redox-responsive regulator of photosynthesis gene expression in the cyanobacterium Synechocystis sp strain PCC 6803. *J. Bacteriol. 182*, 4268−4277.
(18) Lemeille, S., Latifi, A., and Geiselmann, J. (2005) Inferring the connectivity of a regulatory network from mRNA quantification in Synechocystis PCC6803. *Nucleic Acids Res. 33*, 3381−3389.
(19) Aryal, U. K., Stockel, J., Krovvidi, R. K., Gritsenko, M. A., Monroe, M. E., Moore, R. J., Koppenaal, D. W., Smith, R. D., Pakrasi, H. B., and Jacobs, J. M. (2011) Dynamic proteomic profiling of a unicellular cyanobacterium Cyanothece ATCC51142 across light-dark diurnal cycles. *BMC Syst. Biol. 5*, 194.
(20) Mueller, T. J., Berla, B. M., Pakrasi, H. B., and Maranas, C. D. (2013) Rapid construction of metabolic models for a family of Cyanobacteria using a multiple source annotation workflow. *BMC Syst. Biol. 7*, 142.
(21) Bandyopadhyay, A., Elvitigala, T., Welsh, E., Stockel, J., Liberton, M., Min, H. T., Sherman, L. A., and Pakrasi, H. B. (2011) Novel Metabolic Attributes of the Genus Cyanothece, Comprising a Group of Unicellular Nitrogen-Fixing Cyanobacteria. *mBio 2*, e00214-11.
(22) McDermott, J. E., Oehmen, C. S., McCue, L. A., Hill, E., Choi, D. M., Stockel, J., Liberton, M., Pakrasi, H. B., and Sherman, L. A. (2011) A model of cyclic transcriptomic behavior in the cyanobacterium Cyanothece sp ATCC 51142. *Mol. BioSyst. 7*, 2407−2418.
(23) Price, G. D., Badger, M. R., Woodger, F. J., and Long, B. M. (2008) Advances in understanding the cyanobacterial CO2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. *J. Exp. Bot. 59*, 1441−1461.
(24) Lieman-Hurwitz, J., Rachmilevitch, S., Mittler, R., Marcus, Y., and Kaplan, A. (2003) Enhanced photosynthesis and growth of transgenic plants that express ictB, a gene involved in HCO3- accumulation in cyanobacteria. *Plant Biotechnol. J. 1*, 43−50.
(25) Vermaas, W. F. J. (2001) *Photosynthesis and Respiration in Cyanobacteria, In eLS*, John Wiley & Sons, Ltd.
(26) Li, H., Sherman, D. M., Bao, S. L., and Sherman, L. A. (2001) Pattern of cyanophycin accumulation in nitrogen-fixing and non-nitrogen-fixing cyanobacteria. *Arch. Microbiol. 176*, 9−18.
(27) Brophy, J. A. N., and Voigt, C. A. (2014) Principles of genetic circuit design. *Nat. Methods 11*, 508−520.
(28) Immethun, C. M., Ng, K. M., DeLorenzo, D. M., Waldron-Feinstein, B., Lee, Y. C., and Moon, T. S. (2015) Oxygen-responsive Genetic Circuits Constructed in Synechocystis sp. PCC 6803. *Biotechnol. Bioeng.*, DOI: 10.1002/bit.25722.
(29) Osanai, T., Ikeuchi, M., and Tanaka, K. (2008) Group 2 sigma factors in cyanobacteria. *Physiol. Plant. 133*, 490−506.
(30) Vega-Palas, M. A., Flores, E., and Herrero, A. (1992) NtcA, a global nitrogen regulator from the cyanobacterium Synechococcus that belongs to the Crp family of bacterial regulators. *Mol. Microbiol. 6*, 1853−1859.
(31) Liu, D., and Yang, C. (2014) The nitrogen-regulated response regulator NrrA controls cyanophycin synthesis and glycogen catabolism in the cyanobacterium Synechocystis sp. PCC 6803. *J. Biol. Chem. 289*, 2055−2071.
(32) Gaudana, S. B., Krishnakumar, S., Alagesan, S., Digmurti, M. G., Viswanathan, G. A., Chetty, M., and Wangikar, P. P. (2013) Rhythmic and sustained oscillations in metabolism and gene expression of Cyanothece sp ATCC 51142 under constant light. *Front. Microbiol.*, DOI: 10.3389/fmicb.2013.00374.
(33) Bradley, R. L., and Reddy, K. J. (1997) Cloning, sequencing, and regulation of the global nitrogen regulator gene ntcA in the unicellular diazotrophic cyanobacterium Cyanothece sp. strain BH68K. *J. Bacteriol. 179*, 4407−4410.
(34) Marin, K., Suzuki, L., Yamaguchi, K., Ribbeck, K., Yamamoto, H., Kanesaki, Y., Hagemann, M., and Murata, N. (2003) Identification of histidine kinases that act as sensors in the perception of salt stress in Synechocystis sp PCC 6803. *Proc. Natl. Acad. Sci. U. S. A. 100*, 9061−9066.
(35) Kanesaki, Y., Yamamoto, H., Paithoonrangsarid, K., Shoumskaya, M., Suzuki, I., Hayashi, H., and Murata, N. (2007) Histidine kinases play important roles in the perception and signal transduction of hydrogen peroxide in the cyanobacterium, Synechocystis sp PCC 6803. *Plant J. 49*, 313−324.
(36) Niyogi, K. K. (1999) Photoprotection revisited: Genetic and molecular approaches. *Annu. Rev. Plant Physiol. Plant Mol. Biol. 50*, 333−359.
(37) Oliveira, P., and Lindblad, P. (2009) Transcriptional regulation of the cyanobacterial bidirectional Hox-hydrogenase. *Dalton T*, 9990−9996.
(38) Summerfield, T. C., Nagarajan, S., and Sherman, L. A. (2011) Gene expression under low-oxygen conditions in the cyanobacterium Synechocystis sp. PCC 6803 demonstrates Hik31-dependent and -independent responses. *Microbiology 157*, 301−312.
(39) Jones, K. M., Buikema, W. J., and Haselkorn, R. (2003) Heterocyst-specific expression of patB, a gene required for nitrogen fixation in Anabaena sp strain PCC 7120. *J. Bacteriol. 185*, 2306−2314.
(40) Khudyakov, I. Y., and Golden, J. W. (2001) Identification and inactivation of three group 2 sigma factor genes in Anabaena sp strain PCC 7120. *J. Bacteriol. 183*, 6667−6675.
(41) Yeh, K. C., Wu, S. H., Murphy, J. T., and Lagarias, J. C. (1997) A cyanobacterial phytochrome two-component light sensory system. *Science 277*, 1505−1508.
(42) Garcia-Dominguez, M., Muro-Pastor, M. I., Reyes, J. C., and Florencio, F. J. (2000) Light-dependent regulation of cyanobacterial phytochrome expression. *J. Bacteriol. 182*, 38−44.
(43) Berg, H., Ziegler, K., Piotukh, K., Baier, K., Lockau, W., and Volkmer-Engert, R. (2000) Biosynthesis of the cyanobacterial reserve polymer multi-L-arginyl-poly-L-aspartic acid (cyanophycin) - Mechanism of the cyanophycin synthetase reaction studied with synthetic primers. *Eur. J. Biochem. 267*, 5561−5570.
(44) Flores, E., and Herrero, A. (2005) Nitrogen assimilation and nitrogen control in cyanobacteria. *Biochem. Soc. Trans. 33*, 164−167.
(45) Saha, R., Liu, D., Hoynes-O'Connor, A., Liberton, M., Yu, J., Bhattacharyya, M., Balassy, A., Zhang, F., Moon, T., Maranas, C. D., and Pakrasi, H. B. (2015) Diurnal Processes in the Cyanobacterium Synechocystis sp. PCC 6803: Insights from Transcriptomic, Fluxomic and Physiological Analyses, in preparation.
(46) Welsh, E. A., Eschrich, S. A., Berglund, A. E., and Fenstermacher, D. A. (2013) Iterative rank-order normalization of gene expression microarray data. *BMC Bioinf. 14*, 153.
(47) Kucho, K., Okamoto, K., Tsuchiya, Y., Nomura, S., Nango, M., Kanehisa, M., and Ishiura, M. (2005) Global analysis of circadian expression in the cyanobacterium Synechocystis sp strain PCC 6803. *J. Bacteriol. 187*, 2190−2199.
(48) Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2001) On clustering validation techniques. *J. Intell Inf Syst 17*, 107−145.
(49) Sneath, P. H. A., and Sokal, R. R. (1973) *Numerical Taxonomy; The Principles and Practice of Numerical Classification*, W. H. Freeman, San Francisco.

DOI: 10.1021/acssynbio.5b00202
ACS Synth. Biol. 2016, 5, 250−258

(50) Halkidi, M., Vazirgiannis, M., and Batistakis, Y. (2000) Quality Scheme Assessment in the Clustering Process, In *Principles of Data Mining and Knowledge Discovery* (Zighed, D., Komorowski, J., and Żytkow, J., Eds.), pp 265−276, Springer, Berlin.
(51) Singh, A. K., Elvitigala, T., Cameron, J. C., Ghosh, B. K., Bhattacharyya-Pakrasi, M., and Pakrasi, H. B. (2010) Integrative analysis of large scale expression profiles reveals core transcriptional response and coordination between multiple cellular processes in a cyanobacterium. *BMC Syst. Biol. 4*, 83.
(52) R Core Team. (2014) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna Austria.
(53) Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2015) *cluster: Cluster Analysis Basics and Extensions*, R package version 2.0.3.
(54) Desgraupes, B. (2014) *clusterCrit: Clustering Indices*.

DOI: 10.1021/acssynbio.5b00202
ACS Synth. Biol. 2016, 5, 250−258