5-19-2009

Transcriptional Analysis of the Unicellular, Diazotrophic Cyanobacterium Cyanothece ATCC 51142 Grown Under Short Day/Night Cycles.

Jörg Toepel
Purdue University

Jason McDermott

Tina C. Summerfield
University of Otago

Louis A. Sherman
Purdue University, lsherman@purdue.edu

Follow this and additional works at: http://docs.lib.purdue.edu/bioscipubs

Recommended Citation
Toepel, Jörg; McDermott, Jason; Summerfield, Tina C.; and Sherman, Louis A., "Transcriptional Analysis of the Unicellular, Diazotrophic Cyanobacterium Cyanothece ATCC 51142 Grown Under Short Day/Night Cycles." (2009). Department of Biological Sciences Faculty Publications. Paper 46.
http://dx.doi.org/10.1111/j.1529-8817.2009.00674.x

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Title: Transcriptional analysis of the unicellular, diazotrophic cyanobacterium *Cyanothece* sp. ATCC 5142 grown under short day/night cycles

Authors: Jörg Toepel, Jason McDermott, Tina C. Summerfield and Louis A. Sherman

Purdue University

Dept. of Biological Sciences

201 S. University Street

47907 West Lafayette, Indiana, USA

Environmental Molecular Science Laboratory

Pacific Northwest National Laboratory (PNNL)

Richland, WA 99352

Univ

Present address:

Department of Biochemistry

University of Otago

PO Box 56, Dunedin, New Zealand

Date of submission: Date of Acceptance:

Corresponding author:

Louis Sherman

Email: lsherman@purdue.edu

Phone: 765-494-8106

Fax: 765-496-1496

Running Title: Short light dark cycle response in *Cyanothece*
Abstract

*Cyanothoece* sp. strain ATCC 51142 is a unicellular, diazotrophic cyanobacterium that demonstrates extensive metabolic periodicities of photosynthesis, respiration and nitrogen fixation when grown under N$_2$-fixing conditions. We have performed a global transcription analysis of this organism using 6 h light/dark cycles in order to determine the response of the cell to these conditions and to differentiate between diurnal and circadian regulated genes. In addition, we used a context-likelihood of relatedness (CLR) analysis with this data and those from two-day light/dark and light-dark plus continuous light experiments to better differentiate between diurnal and circadian regulated genes. *Cyanothoece* sp. adapted in several ways to growth under short light/dark conditions. Nitrogen was fixed in every second dark period and only once in each 24 h period. Nitrogen fixation was strongly correlated to the energy status of the cells and glycogen breakdown and high respiration rates were necessary to provide appropriate energy and anoxic conditions for this process. We conclude that glycogen breakdown is a key regulatory step within these complex processes. Our results demonstrated that the main metabolic genes involved in photosynthesis, respiration, nitrogen fixation and central carbohydrate metabolism have strong (or total) circadian-regulated components. The short light/dark cycles enable us to identify transcriptional differences among the family of *psbA* genes, as well as the differing patterns of the *hup* genes, which follow the same pattern as nitrogenase genes, relative to the *hox* genes which displayed a diurnal, dark-dependent gene expression.

Keywords: circadian and diurnal behavior; cyanobacteria; glycogen metabolism; light-dark cycles; nitrogen fixation; photosystems.

Abbreviations: LP-light period; DP-dark period; LD-light/dark; CLR-context likelihood of relatedness; PPP-pentose phosphate pathway.
Introduction:

Circadian rhythms have been identified and studied in all classes of organisms, including the cyanobacteria. Cyanobacteria are the simplest organisms to have a circadian clock, although the clock components are different than those in higher organisms. One of the more interesting findings in Synechococcus elongatus sp. PCC 7942, the main model cyanobacterium for circadian studies, is that all *S. elongatus* genes apparently are expressed in a circadian rhythmic fashion (Liu et al. 1995a, b, Golden and Canles 2003, Ditty et al. 2003). These findings are based on multiple techniques and it appears evident that actual transcription is globally circadian. However, in practice, it is very likely that only a fraction of the genes will display a perfectly rhythmic pattern. First, mRNAs have different half lives and the gene transcript level at any time period will be the net sum of new transcription plus the accumulation and degradation of older transcripts. Second, the fundamental circadian rhythms will be regulated by many post-transcriptional processes, which will alter the frequency or amplitude of the transcriptional pattern.

Circadian rhythms function to optimize specific physiological mechanisms to the correct time of day (Golden 2007, Golden and Canales 2003). We are particularly interested in the relationship of photosynthesis, which generates O₂, to the O₂-sensitive nitrogen fixation throughout a 24 hr diurnal cycle. Transcript levels may be circadian (light/dark independent) or be directly affected by light or darkness. Such differences will become manifest when cells are grown under photocycles that differ significantly from the usual 12h light/12 h dark diurnal cycle (e.g., continuous light, short light/dark cycles, etc.). Similar considerations are also critical in plants and a comprehensive study in *Arabidopsis thaliana* has investigated the impact of both photocycles and thermocycles on circadian rhythms (Michael et al. 2008). This study led to the conclusion that 89% of *Arabidopsis* transcripts cycled under at least one condition and that the expression of most genes peaked at a specific time of day. Furthermore, this peak can shift depending on the environment (Michael et al. 2008).

This provides the most detailed study to date in a photosynthetic organism on the relationship of the underlying circadian clock to environmental parameters that impact on metabolic rhythms.

We have been involved with related studies in the unicellular diazotrophic cyanobacterium *Cyanothece* sp. ATCC 51142, hereafter *Cyanothece* sp., (Reddy et al. 1993, Schneegurt et al. 1994) that performs photosynthesis during the light and fixes nitrogen during the dark (Schneegurt et al. 1997a,b, Sherman et al.
Previous studies with *Cyanothece* sp. showed a strong correlation between activity and transcript level for a subset of genes related to photosynthesis (Meunier et al. 1998) and N\textsubscript{2} fixation (Colón-López et al. 1997) during normal 12 h light/dark (LD) cycles. In addition, differential gene expression was reported for the main photosynthetic genes and the nitrogenase genes in the light and dark, respectively (Colón-López et al. 1997, 1998, 1999). Recently, whole genome microarray experiments were carried out to determine the diurnal gene expression under normal LD conditions (12 h L/12 h D) (Stöckel et al. 2008) and under continuous light (Toepel et al. 2008). Both groups found ~30% of the ~5000 genes on the microarray exhibited diurnal oscillations under 12 h LD conditions and Toepel et al. (2008) demonstrated that ~10% of the genes demonstrated circadian behavior during growth in free-running (continuous light) conditions. Toepel et al. (2008) also demonstrated that nitrogenase transcript abundance and nitrogenase activity were correlated in *Cyanothece* sp. under continuous light and that N\textsubscript{2} fixation followed a ~24 h rhythm under these conditions, albeit with reduced rates. These results indicate a LD-independent expression pattern for nitrogenase genes, consistent with the circadian behavior for nitrogenase related genes suggested by Sherman et al. (1998). This is in contrast with results from *Gloeoece* sp., where N\textsubscript{2} fixation appears to be under the control of an undefined endogenous rhythm (Gallon 2001), but one that was not circadian. When grown in continuous light, *Gloeoece* did not display a ~24 h periodicity for N\textsubscript{2} fixation, but reverted to one of 30-40 h (Gallon and Chaplin 1988). In addition, the results of Taniuchi and Ohki (2007) in *Gloeoece* sp. indicated that mRNA abundance and nitrogenase activity were not always connected, and that nitrogenase activity was dependent on the time span between dark periods and occurred 4 h into the dark.

Photosynthetic activity depends upon incident light, although genes encoding photosynthetic proteins display a diurnal or a circadian dependent expression pattern (Stöckel et al. 2008, Michael et al. 2008). In *Cyanothece* sp., maximum photosynthetic rates in a normal LD cycle occurred after 6-8 h light incubation and photosynthetic capacity decreased strongly during the N\textsubscript{2} fixation period (Meunier et al. 1998, Toepel et al. 2008). Toepel et al. (2008) showed that the photosynthetic rates were lower during growth under continuous light and demonstrated no circadian related pattern for photosynthesis genes. In the case of growth under continuous light, the glycogen content stayed at high levels and did not decrease until the cells were again placed in darkness (Toepel et al. 2008). Furthermore, these results indicated that nitrogenase transcription and
activity was metabolically or energetically regulated via glycogen breakdown and suggested that photosynthesis is light activated, but probably regulated by the internal carbohydrate level.

The main circadian control genes found in *Synechococcus elongatus* are the kai-genes that function as internal oscillators (Bell-Pederson et al. 2005) and a cyclic behavior of gene expression during continuous light growth had been demonstrated. Short dark pulses reset the internal clock and lead to a phase shift of cultures grown under continuous light (Mackey and Golden 2007). Therefore, light or dark signals can synchronize the cell’s internal clock with the external environment. Three proteins are described with a receptor function for external signals. The *cikA* (circadian input kinase) (Ivleva et al. 2006) most likely senses the redox state of the plastoquinone (PQ) pool, and *ldpA* (light-dependent period) and *pex* (period extender) determine the period length. The way in which temporal information is converted into global oscillation of transcription remains unsolved, but chromosome compaction (Smith and Williams 2006) and the above signal transduction pathway may well be involved (Takai et al. 2006). However, the CikA kinase does not have a clear homolog in other cyanobacteria, including *Cyanothece* sp. ATCC 51142, and variations in the circadian mechanism should be considered the norm (Golden 2007, Mackey and Golden 2008).

We are interested in relating cellular energy levels to the nature of circadian and diurnal control of gene expression in *Cyanothece* sp. Cellular processes depend on the cellular energy level and enzyme activity is often substrate regulated. Therefore, we asked the following questions: How stable is the circadian or diurnal rhythm of the genome if the ambient LD cycles are permanently different from typical 12 h LD periods? Can *Cyanothece* sp. grow and fix N₂ under 6 h LD, N₂ fixing conditions? What are potential candidates for light/dark perception in *Cyanothece* sp.? Finally, can we use the information from this experiment, in conjunction with the previous transcriptional studies, to better define genes under either circadian or diurnal regulation?

**Material and Methods:**

**Bacterial strains and growth conditions**

*Cyanothece* sp. ATCC 51142 cultures were grown in an airlift bioreactor (BioFlo 3000, 6-L, New Brunswick Scientific, Edison, NJ) in ASP2 media without nitrate at 30°C in 12 h L/12 h D cycles (Toepel et al. 2008). The culture was illuminated by two LED panels using alternating arrays of orange (640 nm) and blue (430 nm).
LED’s, yielding an intensity of ~100 $\mu$mol photons m$^{-2}$ s$^{-1}$ inside the bioreactor (underwater quantum light meter (LI 192, Li-Cor, Lincoln, Nebraska)). Cultures were inoculated in the bioreactor at a cell density of approximately 1 x 10$^6$ cells ml$^{-1}$ and were grown for 5-6 days under LD conditions prior to the experiments. We then applied 6 h light 6 h dark cycles for 24 h and took samples for mRNA isolation and measurements of cell number, chlorophyll (Chl) a concentration, photosynthetic $O_2$ evolution and, respiration and glycogen content (50 ml per time point). Cell number and the Chl a concentration were measured using a Petroff-Hauser cell count chamber and a Perkin-Elmer spectrophotometer (Lambda 40, Shelton, CA), respectively, as described (Schneegurt et al. 1994). Photosynthetic oxygen evolution (using the growth light intensity, ~100 $\mu$mol photons m$^{-2}$ s$^{-1}$ for illumination) and respiration rates (monitored as $O_2$ consumption in the dark) were determined using a Clark type electrode (Hansatech, Norfolk, England), without adding bicarbonate. $N_2$ fixation rates were determined by measuring ethylene on a Hewlett Packard 8460 gas chromatograph (Wilmington, DE) and calculated as described (Colón-López et al. 1997). The rates of photosynthesis, respiration and $N_2$ fixation in Figure 1 are plotted on the right axis as relative activity. Photosynthesis is represented as $O_2$ evolution (positive activity), whereas respiration is represented as $O_2$ uptake (negative activity). This experimental protocol was repeated three times with essentially identical results. 

**Microarray platform**

The high-density microarray platform consisted of 5,096 ORF’s based upon the rough draft of the *Cyanothecaceae* genome sequence that was obtained by the Washington University Genome Center (Welsh et al. 2008). The ORF’s were identified using Critica and Glimmer and the 60-mers appropriate for each gene were determined by a computer program written by Dr. Rajeev Aurora, St. Louis University, and provided by him to Agilent, Inc (Santa Clara, CA). The microarrays were fabricated by Agilent, provided to MoGene (St. Louis, MO), along with the purified RNA samples for hybridization, scanning and initial data analysis (see Stockel et al. (2008) and Toepel et al. (2008) for further details).

**RNA isolation**

For the microarray experiment, 300 ml samples were taken at each time period (t=2, 4, 8, 10, 14, 16, 20, and 22h) over a 24 h period. The cells were centrifuged at 5,000 x g and re-suspended in STET-Buffer (8%
sucrose, 5% Triton X-100, 50 mM EDTA, 50 mM Tris-HCl, pH8, with DEPC water) and stored at -80°C. The RNA extraction was performed using a slightly modified version of the protocol described in Reddy et al. (1990). The RNA Clean-up Kit-5 columns from Zymo Research Corporation (Orange, CA) were used to remove contamination (e.g., carbohydrates, organic solvents).

Results are for a combination of 3 biological and 2 technical replicates. For each microarray, 1 µg RNA was used (0.5 µg sample + 0.5 µg control). Total RNA was labeled with either cyanine-5 or cyanine-3 using ULS, a RNA fluorescent labeling kit from Kreatech Biotechnology (Amsterdam, Netherlands) according to the manufacturer’s protocol. The labeled material was passed through Zymo RNA Clean-up Kit-5 columns (Zymo Research Corporation, Orange, CA) to remove un-incorporated label and eluted in 15-20 µl of RNase-free water. Concentration of labeled total RNA and label incorporation was determined on a Nanodrop-1000 spectrophotometer (Wilmington, DE). All of the labeling and post labeling procedures were conducted in an ozone-free enclosure to ensure the integrity of the label. Labeled material was hybridized for 17 h in a rotating oven at 65°C in an ozone-free room. Wash conditions used were as outlined in the Agilent processing manual (Santa Clara, CA) and the arrays were scanned by an Agilent scanner. Analysis was performed using Agilent’s Feature Extraction Software version 9.1 and Rosetta Luminator software (Rosetta Biosoftware, Kirkland, WA).

**Semi-Quantitative RT-PCR.**

RNA was treated with DNase I (Invitrogen, Carlsbad, CA) for 1 h at 37°C and successful DNase I treatment was confirmed by PCR on each RNA sample. Reverse transcription was performed using Superscript II (Invitrogen, Carlsbad, CA) and random primers using the manufacturer’s instructions. PCR was carried out at 94°C, 1 min, 30 cycles of 94°C 30 s, 54°C 30 s and 68°C 30 s, to amplify regions of the genes \( nifH, coxCl \) and \( rpnA \). Due to the high transcript level of these genes, amplification using \( psbA \) and \( psbD \) primers was performed using 20 cycles of the PCR conditions described above. The \( rpnA \) transcript abundance was used as a control, since microarray data indicated the transcript level for this gene was unchanged under these growth conditions. The primers and amplified products sizes for the following genes: \( nifH: psbA1, coxCl, rpnA \) (275 bp), and \( psbD \) were as used in Toepel et al. (2008).
**FTIR-spectroscopy**

The macromolecular composition of the cells was determined over the 48 h period with an FTIR-spectrometer (Thermo-Electron, Madison, WI). Samples (2 ml) were taken every 6 h over a 24 h time period, washed twice with distilled water and stored at -80°C. The cells were dried prior to measurements on an IR-slide at 45°C for 4 h. For each sample, 128 spectra x 50 spots were measured and analyzed according to B. Penning (see http://cellwall.genomics.purdue.edu/). The spectra for each sample were baseline corrected, area normalized and averaged.

**Context Likelihood of Relatedness (CLR) Analysis**

Transcriptomic profiles were assembled for each gene in *Cyanothece* from experiments for 12 h LD conditions (Stöckel et al. 2008), continuous light (Toepel et al. 2008), low O₂ conditions for 1, 2 and 6 hours (Toepel, Summerfield and Sherman, unpublished data), and the 6 h LD experiment reported here. These profiles were analyzed using the context likelihood of relatedness (CLR) method (Faith et al. 2007) using the SEBINI framework (Taylor et al. 2006). CLR calculates the mutual information (a sensitive measure of similarity) between transcription profiles from all pairs of genes, and then applies statistical filters to predict relationships among genes. Filtering these relationships to include only those between transcription factors provides a high-confidence regulatory network if enough individual conditions are used (Faith et al. 2007). We did not filter the genes considered, so the relationships predicted are more generic and represent highly significant similarities between the expression profiles among all of the genes. We used a spline degree of 3 and 8 bins for input parameters to CLR. The resulting network (Figure 5) was filtered using a CLR zscore of 5.0 and visualized with Cytoscape (Shannon et al. 2003). Grouping of genes in the network into clusters was accomplished using hierarchical clustering with Cluster 3.0 (Eisen et al. 1998), using average pairwise linkage and the CLR scores as the similarity metric.
Results:

Photosynthesis and N₂ fixation in 6 h LD growth

Growth of Cyanothece sp. in 6 h LD cycles showed a specific cyclic pattern of enzyme activity, as demonstrated by the O₂ evolution capacity, respiration and N₂-fixation during the 24 h time period (Fig. 1a). The net photosynthesis rate increased during the 1st light period (LP) and the capacity for photosynthesis remained high during the following dark period; respiration was low in the 1st LP and in the 1st DP. Net photosynthetic rates decreased steadily in the second half of the 2nd LP and then decreased to 20% of the maximum light rate in the 2nd DP. The rate of respiration increased during the 2nd LP and reached maximum values during the 2nd DP. The gross photosynthetic rates remained constant through the first 12 h of the experiment and decreased slightly during the 2nd LP and decreased further in the 2nd DP (50% decrease, data not shown). The initial N₂ fixation activity rates were very low, but increased in the 2nd DP (Fig.1a). The peak of N₂ fixation was at ~20-22 h, which was a delay of 4-6 h in comparison to the 12 h LD cycle (Schneegurt et al. 1994). The transition to the 6 h LD cycle led to a 24 h rhythm of N₂ fixation (data not shown), which was not influenced by the short LD rhythms (Fig. 1a). Furthermore, the larger decline in the pH during each 2nd DP correlated with higher rates of respiration during that period and the high level of nitrogenase activity (Fig. 1a).

Differential gene expression in 6 h LD cycle

Analysis of the microarrays yielded ~1400 up-regulated genes (>2-fold) during the 24 h experiment out of 5096 annotated genes in Cyanothece sp ATCC 51142. We identified 520 up-regulated genes in the 1st LP, 214 in the 2nd LP, 394 in the 1st DP and 275 in the 2nd DP. The expression patterns for some of the genes encoding energy-related functions are summarized in Table 1. This table defines the expression patterns that we have used to categorize circadian and diurnal regulation. This includes data from the 6h LD experiment described herein, the 12h LD experiment described in Stockel et al (2008) and the LD+LL experiment described in Toepel et al (2008). The darkened sectors indicate where expression increased the most and the shaded sector indicates an increase, but not a peak. The results from the 6h LD experiment show that some genes, such as those encoding phycobiliproteins or the ATP synthase proteins, demonstrated a light-independent gene expression (e.g., circadian), in that they continued to have expression peaks in the 1st LP, although there was a smaller peak in the 2nd LP. Other genes, such as coxABC1, peaked in the 2nd DP, but the
time of peak expression was modified (from \( t=12h \) in 12h LD to \( t=16h \) in 6h LD). A second class of potential circadian genes, with a peak in the 2\(^{nd}\) DP, included the \textit{nif} gene cluster, \textit{hupLS} and \textit{feoA2B2}. At the same time, some genes demonstrated a significant dark regulation and showed increased transcript levels in each DP (e.g., \textit{hoxEFUYH}, \textit{cphAB} and \textit{feoA1B1}). There was a large class of light up-regulated genes, including many of the PSI and PSII genes (not including those encoding the reaction center proteins) and genes encoding nitrate and sulfate transporters. The reaction center genes responded quite differently. The \textit{psaAB} operon followed a circadian pattern, whereas each gene in the \textit{psbA} family demonstrated a unique pattern of transcription. Thus, \textit{psbA1} (and its twin, \textit{psbA5}) peaked in the 1\(^{st}\) LP, but had a smaller peak in the 2\(^{nd}\) LP. The \textit{psbA3} gene had a single peak in the 2\(^{nd}\) LP, but \textit{psbA4} demonstrated a distinct peak in the 1\(^{st}\) DP.

Figure 2 shows the expression pattern of genes involved in several functional processes in \textit{Cyanothece} sp. during the 6 h LD cycles compared to the patterns in the 12 h LD experiment (Toepel et al. 2008). The majority of the genes encoding PS I and PS II proteins were up-regulated at the beginning of both the 1\(^{st}\) LP and the 2\(^{nd}\) LP. Notably, the expression level of all of the photosynthesis-related genes were lower in the 2\(^{nd}\) LP in comparison to the 1\(^{st}\) LP, although the expression was still enhanced \(>2\)-fold in comparison to the DP. The different LD conditions had little impact on the expression of the genes encoding the RUBISCO subunits and PsaAB in the two experiments. The transcripts for \(\text{CO}_2\) fixation (\textit{rbcLSX}) were up-regulated during the 1\(^{st}\) LP in both experiments and again 24 h later. This expression pattern may contribute to the limited photosynthetic capacity during the 2\(^{nd}\) LP in the 6 h LD experiment. The genes encoding the PS I reaction center proteins (PsaAB) demonstrated a maximum expression at \(t=4\) to 8 in both experiments (Fig. 2a). However, the majority of the photosynthetic genes showed light dependent gene expression during both experiments. The respiration related genes (\textit{coxBAC1}) were up-regulated during the 1\(^{st}\) DP and displayed higher expression levels through the rest of the experiment, when the 6 h LD results were compared to the 12 h LD results (Fig. 2c). Since nitrogenase requires both energy and a low level of cellular \(\text{O}_2\), glycogen breakdown and respiration need to be enhanced during this period.

\textit{Nitrogen fixation and nitrogen assimilation}

Genes encoding nitrogenase and related proteins were up-regulated during the 2\(^{nd}\) DP, corresponding to nitrogenase activity, but with a 4 h time shift compared to the normal LD cycle. As shown in previous
experiments (Stöckel et al. 2008, Toepel et al. 2008), several ferredoxins and flavodoxins were up-regulated during the same period. Additionally, the gene for peroxiredoxin, a thiol-based peroxide reductase (Dietz et al. 2006) was 78 fold up-regulated in the 2nd DP, indicating a possible function in detoxification of H$_2$O$_2$ (Table 1). Some genes encoding nitrate assimilation related proteins were also up-regulated at the beginning of the 2nd dark period (t=14 h), similar to the 12 h LD cycle. We found little variation in the expression level of ntcA and the nitrate/nitrate assimilation regulator ntcB was up-regulated at t=14 which represents the normal dark period, the same pattern was detectable during 12 h LD growth.

Light and dark dependent and independent gene expression

The comparison between the 12 h LD cycle experiments and the 6 h LD cycle experiments allowed us to differentiate genes under control of the light/dark signal from genes showing alternative gene expression patterns, e.g. depending on the absolute day time and response to the energy level in the cell. Supplemental Table 1 shows the genes with an LD-dependent gene expression, separated as to the different light and dark period. Analysis of the up-regulated genes during the 1st DP (t=6 to 12 h) yielded several genes encoding energy producing enzymes. The rate-limiting enzymes for the Pentose Phosphate Pathway (zwf and gnd) showed an up-regulation during the 1st DP and remained high during the rest of the experiment (Fig. 3c) (Tian et al. 1998). Several fermentative related genes were up-regulated exclusively during the 1st DP; e.g. the malate dehydrogenase (pyruvate carboxylation) was 8-fold up-regulated and the aldehyde dehydrogenase, probably involved in acetate production, was up-regulated 17-fold. No alcohol dehydrogenase genes were up-regulated during the experiment.

The 12 h LD+LL experiments indicated that the expression pattern between the nitrogenase related genes and both hydrogenase systems (hox and hup genes) occurred at the same time. In this experiment, hox and hup genes showed different and distinct expression patterns. The hox genes were up-regulated during the 1st DP (Fig. 3b), whereas the hup genes were mainly up-regulated during the 2nd DP, similar to the nif cluster (Fig. 3a). A similar difference was observed between the two iron-transporter systems feoA1B1 (up-regulated in the 1st and 2nd DP) and feoA2B2 (up-regulated only in the 2nd DP). We also identified genes which showed a specific response to the incident light or dark. One operon (ccel_1684 – cce_1691), with several NADH dehydrogenase subunits, a sulfate-transporter and an Na/$H^+$ antiporter and several unknown genes, was up-
regulated exclusively during both light periods. Another gene cluster that encodes proteins with a 4VR domain (cce_2109 and cce_2112), and which may be involved in signaling, was up-regulated during the beginning of the 1st LP, similar to the 12 h LD experiment (Singh et al. 2004). Finally, the analysis of LD-dependent gene expression patterns yielded several potential regulator genes responding to the varying LD conditions (Figs. 3b and 3c). A potential light-dark receptor, aph1, a phytochrome, was up-regulated at the beginning of each DP.

In addition, we found several histidine kinases up-regulated: during the 1st LP (cce_0754 and cce_0755), during the 2nd LP, (e.g., rpaA/B, Fig. 3c), during the 1st DP (cce_2505, unknown function), during 2nd DP (cce_2366) and during both DPs (cce_0220). The histidine kinase and response regulator cce_0969 and cce_0970 (Fig. 3c) were up-regulated during the normal dark period (time points t=16 to 20h) identical to the previous 12 h LD experiments.

Glycogen production and breakdown

The measurements of relative glycogen content demonstrated that glycogen levels increased during the 1st LP, remained high during the 1st DP and increased in the 2nd LP (Fig. 4). The glycogen content then decreased during the 2nd DP coincident with N2 fixation and the increase in respiration. These results corresponded with the gene expression of glycogen production and breakdown enzymes encoding genes. The glycogen synthase gene (glgA2) was up-regulated at the beginning of the 1st LP and at t=22 h. The transcripts encoding the glycogen debranching enzyme (glgX) was up-regulated during the 1st DP and at t=16 h. However, the glycogen phosphorylase gene (glp3) was up-regulated only during the 2nd DP and it seemed to be the initiating step in glycogen breakdown. The cyanophycin encoding genes cphA and cphB were up-regulated during both dark periods (Table 1).
Identification of Circadian Genes by Context Likelihood of Relatedness (CLR) analysis

A second approach to the identification of genes that displayed a circadian transcriptional pattern used CLR (Faith et al. 2007) analysis of four separate differential transcription experiments: a 2-day 12 h LD experiment (Stöckel et al. 2008); a 12 h LD+LL experiment (Toepel et al. 2008); this 6h LD experiment; and, an experiment that compared an aerobically grown culture before and after a 1 h, 2 h and 6 h transition to anaerobic conditions (Toepel, Summerfield and Sherman, unpublished data). Fig. 5 depicts this analysis as a network of relationships between genes that approximates a 24 h clock. The tightness of a cluster and the large number of contacts indicated groups of genes that are transcribed in a similar fashion. We therefore analyzed the network using a hierarchical clustering approach that identified clusters of genes tightly related by CLR.

From this analysis, we identified 4 large clusters and a number of smaller clusters. We were particularly interested in three clusters that contain many genes displaying circadian behavior and that are circled in Fig. 5: a photosynthesis cluster, a nitrogenase cluster, and a respiration cluster.

The photosynthesis cluster contains genes encoding proteins for all of the major photosynthesis structures, especially PSII, cyt b6f, CO2 fixation, ATPase and the phycobilisomes (Table 2). Interestingly, the only PSI genes in this cluster are psaF/J and all of the other PSI genes, including circadian operon psaAB, are in the predominantly non-circadian cluster around 4-7 o’clock in Fig. 5. Other genes in the photosynthesis cluster include glycogen synthetase and sigma factors sigG and sigH. This cluster is very similar to cluster 3, circadian—light up-regulated in Table 1.

The nitrogenase cluster at 8 o’clock is the tightest and most cohesive cluster as described in previous publications (Stöckel et al. 2008, Toepel et al. 2008). The CLR analysis determined that virtually every gene between cce_0545 and 0580 was in the cluster (Table 2). In addition, hupLS were associated very close to this cluster. In particular, hupL is just outside the nif cluster, but with edges connecting all nitrogenase genes in the cluster. This clearly identifies a functional relationship among these genes. In addition, the rpaA gene, which has been found to be involved in circadian regulation in Synechococcus elongatus (Takai et al. 2006), is at approximately 7 o’clock in the CLR diagram (Fig. 5) and is associated with many nif genes. This might represent a regulatory loop in which rpaA is transcribed, translated and then used in the regulation of the nif cluster.
The cluster at 11 o’clock includes genes involved in respiration and in central metabolism, such as glycolysis, the pentose phosphate pathway and glycogen degradation (Table 2). This is a most interesting cluster because it includes some gene sets that are correlated and others that are anti-correlated. Thus, the cluster includes genes that have peaks at the end of the light period and other genes with peaks toward the end of the dark period. This cluster includes circadian regulated genes that are needed to prepare for the next light or dark phase. Thus, under normal 12 h LD conditions, the cytochrome oxidase genes have peak transcript levels around L10 to D2.

RT-PCR validation of microarray results

We used RT-PCR to validate the results of the microarray experiments. Figure 6 shows a comparison between the microarray results (top) to the RT-PCR results (bottom) for a select set of genes that demonstrated different types of temporal expression. This included expression of nifH that was dramatically up-regulated in the second dark and two PS II genes (psbA1 and psbD) that were up-regulated in the light, although to a lesser extent in the 2nd LP. We note that the microarray data is plotted on a logarithmic scale because the nifH gene is transcribed at very high levels in the 2nd DP, as described in Toepel et al (2008). We also compared the results for coxC1, which was transcribed at high levels during the dark periods and the second light period. It can be seen that there is a very close correspondence between the microarray data and the RT-PCR data for all 3 types of regulation. These should be compared to the results for rnpA, which were virtually unchanged in both the microarray and the RT-PCR experiments. Based on these results, we concluded that the microarray results were a fair representation of differential gene expression in this organism during the 24 h experiment.
Discussion:

We analyzed growth in short light-dark periods, along with a comprehensive analysis of growth under different light-dark conditions, to develop a fuller understanding of circadian vs. diurnal regulation of gene transcription in *Cyanothecae* sp. The results in Tables 1 and 2 indicate that the main metabolic genes involved in photosynthesis, respiration, nitrogen fixation and central carbohydrate metabolism have strong (or total) circadian-regulated components. The transcription of these gene sets was carefully maintained at the appropriate light or dark period, even during 6 h LD growth, so that the cell could generate sufficient energy and an appropriate anoxic environment, when required. The photosynthate from fixed carbon was placed in glycogen granules and the degradation of this glycogen was used as a substrate for the massive respiratory burst that occurred prior to and concomitant with N2 fixation. Coordinating these metabolic properties in a circadian manner appears to be one of the key regulatory functions of the cell. It is interesting that many PSI and PSII genes are transcribed in a diurnal fashion, whereas the reaction center genes and a few genes encoding assembly-type proteins are transcribed in a circadian manner. These findings might have important ramifications as to the assembly and specific functions of the two photosystems. We have also identified some specific regulatory molecules that may be involved in the intricate timing of gene transcription from light to dark transitions and vice versa (Tables 1 and 2). We will complete a more detailed computational analysis of these data to determine if we can predict specific interrelationships between regulatory and structural proteins that can then be tested via genetics and physiology.

It was evident that *Cyanothecae* undergoes several adaptation processes in order to develop the appropriate environment for N2 fixation. Nitrogenase activity and nif gene expression were in good correspondence. Our results indicated that one 6 h light period is not sufficient for induction of the nitrogenase-related genes, a feature that also was demonstrated in *Gloeobacter* (Taniuchi and Ohki 2007). This is consistent with our hypothesis that N2 fixation is dependent upon the energy status of the cells (Toepel et al. 2008). Our results demonstrated that breakdown of stored glycogen and high respiration rates were connected to the nitrogenase activity and that both processes are necessary to provide sufficient energy equivalents for N2 fixation. Furthermore, our data suggested that regulation of photosynthesis is independent of incident light
during the first 18 h and the decline of photosynthesis is coordinated with the increase in respiration and the
concomitant breakdown of glycogen. Our results support the hypothesis of diurnal control of the major
processes in Cyanothece sp. (Stöckel et al. 2008), but also suggest a strong feedback regulation regarding the
energy status of the cells. Additionally, we demonstrated that respiratory enzyme activity and gene expression
patterns did not always correspond, suggesting post-transcriptional, energy dependent activation of respiration
and a tight co-regulation with N₂ fixation.

Comparison of experiments in which cells were grown in LD + LL (Toepel et al. 2008) and this study
demonstrated that gene expression for glycogen and cyanophycin production/breakdown was LD-dependent. In
both experiments, glycogen breakdown was clearly dark dependent and glycogen phosphorylase appeared to be
the initializing enzyme and one that most likely required a dark signal for expression. Comparison between the
LD+ LL experiments and the 6 h LD experiment, allows us to better differentiate between LD-dependent and
LD-independent expression and identify circadian controlled genes.

Two gene sets, encoding proteins for hydrogenases and photosynthesis, were of particular interest to us.
Toepel et al. (2008) showed that the uptake hydrogenase (hup) and the bidirectional hydrogenase (hox), found
in Cyanothece sp., were co-transcribed with the N₂ fixation genes (nif-genes) and probably consumed the
produced hydrogen. The expression pattern of the uptake hydrogenase-encoding hup genes in Cyanothece sp.
was in good accordance with previous studies, such that the expression of these genes was closely related to
the nif genes. In contrast, the hydrogenase-encoding hox-genes were mostly up-regulated during the first dark
period, and were probably involved in fermentation or acts as an electron valve during photosynthesis (for
review, see Tamagnini et al. 2007). Since the fermentative-related genes for acetate and lactate production and
genes for the PPP cycle (especially gnd) showed elevated expression during the first dark period, it is likely
that Cyanothece sp. could produce reducing power and energy equivalents from fermentation and the PPP
cycle during this dark period.

The photosynthesis genes demonstrated rather important behaviors. The majority of the PSI and PSII
structural genes were diurnally expressed and had peaks in both light periods when grown in 6 h LD periods.
At the same time, the most highly expressed psbA gene, psbA1/psbA5, also demonstrated a diurnal pattern,
although the intensity was lower in the 2nd DP. The expression pattern of the psbA1 gene was in good
agreement with previous results (Colón-López et al. 1998, Stöckel et al. 2008, Toepel et al. 2008). At the same time, *psaAB* demonstrated a circadian behavior and the *psaAB* transcript level peak was always near the end of the light period (see also Colón-López and Sherman 1998). Additionally, two members of the *psbA* family, *psbA3* and *psbA4*, also demonstrated a circadian transcriptional pattern, but with peaks at different times (Table 1). It is possible that the encoded D1 proteins that are produced at specific times specifically to change the properties of PSII, similar to the situation in *Synechococcus elongatus* (Schaefer and Golden 1989a, b). In fact, the transcriptional results for *psbA3* and *psbA4* correlate nicely with the protein results using the *S. elongatus* D1 Form 2 antibody as described in Colón-López and Sherman (1998).

In conclusion, we have used a short day-night cycle to help determine the circadian vs. diurnal regulation of genes encoding important nitrogen fixation and energy-producing functions. We have then used these data, in conjunction with experiments using other light-dark patterns, to develop a comprehensive model of circadian and diurnal associations. Importantly, both approaches resulted in very similar results, thus demonstrating the value of the CLR analysis. Our results clearly indicated that the nitrogenase gene cluster is under very tight circadian regulation, but that other energy-producing functions must be modified to ensure that the cell can provide the appropriate energetic and oxygen environment to permit N₂ fixation under such anoxic conditions. This is a key property of *Cyanothece* sp. ATCC 51142 and one that may provide directions for the synchrony of other important metabolic processes, including photosynthetically-driven hydrogen production.

Acknowledgements:

We would like to thank for Dr. Rajeev Aurora for his efforts to develop the microarray platform and Bryan Penning of the Purdue University for the help with the infrared spectroscopy. This work was supported by the Membrane Biology EMSL Scientific Grand Challenge project at the W. R. Wiley Environmental Molecular Science Laboratory, a national scientific user facility sponsored by the U.S. Department of Energy’s Office of Biological and Environmental Research (BER) program located at Pacific Northwest National Laboratory. PNNL is operated for the Department of Energy by Battelle.
Gene Transcription Classes in *Cyanothece* sp. ATCC 51142 During Growth in 6 h LD Periods.

<table>
<thead>
<tr>
<th>1. <strong>Diurnal—Dark-Up-regulated (1st and 2nd DP)</strong></th>
<th>12h</th>
<th>LL SD&lt;sup&gt;b&lt;/sup&gt;</th>
<th>6h LD</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>hoxEFUYH</em> (cce_2315-19)</td>
<td></td>
<td>L</td>
<td>L L D D</td>
</tr>
<tr>
<td>PPP cycle (<em>gnd</em>, <em>cce_3746</em> and <em>zwf</em> <em>cce_2536</em>)</td>
<td></td>
<td>D</td>
<td>L L</td>
</tr>
<tr>
<td><em>cphAB</em> (ccce_2236 and 2237)</td>
<td></td>
<td></td>
<td>L L</td>
</tr>
<tr>
<td><em>feoA1B1</em> (ccce_0032 and 0033)</td>
<td></td>
<td></td>
<td>D L</td>
</tr>
<tr>
<td><em>nifX</em> like (ccce_1854-56)</td>
<td></td>
<td></td>
<td>D L</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. <strong>Diurnal—Light Up-regulated (1st and 2nd LP)</strong></th>
<th>12h</th>
<th>LL SD&lt;sup&gt;b&lt;/sup&gt;</th>
<th>6h LD</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSI genes (not including reaction center)</td>
<td></td>
<td>L</td>
<td>L L</td>
</tr>
<tr>
<td>PSII genes (not including reaction center)</td>
<td></td>
<td>D</td>
<td>L D</td>
</tr>
<tr>
<td><em>psbA1</em> (A5) (ccce_0636) (lower intensity in 2nd LP)</td>
<td></td>
<td></td>
<td>L L</td>
</tr>
<tr>
<td>nitrate metab. (<em>narB</em>, ccce_1214; <em>nrtP</em>, ccce_1222; <em>nirA</em>, ccce_1223)</td>
<td></td>
<td></td>
<td>D L</td>
</tr>
<tr>
<td>Sulfate/Na&lt;sup&gt;+&lt;/sup&gt;-H&lt;sup&gt;+&lt;/sup&gt; antiporter (ccce_1681-1689)</td>
<td></td>
<td></td>
<td>L L</td>
</tr>
<tr>
<td>NADH dehydrogenase (ccce_1690 to 1693)</td>
<td></td>
<td></td>
<td>D L</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. <strong>Circadian—Light Up-regulated</strong></th>
<th>12h</th>
<th>LL SD&lt;sup&gt;b&lt;/sup&gt;</th>
<th>6h LD</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBS (<em>apcAB</em> ccce_2908-09; <em>cpcABCCDE</em> ccce_2651-2656)</td>
<td></td>
<td>L</td>
<td>L L</td>
</tr>
<tr>
<td>ATP synthase (ccce_2812-2813 and ccce_4482-4489)</td>
<td></td>
<td>D</td>
<td>L D</td>
</tr>
<tr>
<td><em>psaAB</em> (ccce_0989 to 0990)</td>
<td></td>
<td></td>
<td>L L</td>
</tr>
<tr>
<td><em>psbA3</em> (2nd LP) (ccce_0267)</td>
<td></td>
<td></td>
<td>D L</td>
</tr>
<tr>
<td>4VR operon (ccce_2106-12)</td>
<td></td>
<td></td>
<td>L L</td>
</tr>
<tr>
<td>His kinase+response regulator (ccce_0969 and 0970)</td>
<td></td>
<td></td>
<td>D L</td>
</tr>
<tr>
<td>Response regulator (ccce_4542)</td>
<td></td>
<td></td>
<td>L L</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. <strong>Circadian—Dark-up-regulated, Time Modified</strong></th>
<th>12h</th>
<th>LL SD&lt;sup&gt;b&lt;/sup&gt;</th>
<th>6h LD</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>coxABC1</em> (ccce_1975 to 1977)</td>
<td></td>
<td>L</td>
<td>L L</td>
</tr>
<tr>
<td><em>glgP3</em> (ccce_1603)</td>
<td></td>
<td>D</td>
<td>L D</td>
</tr>
<tr>
<td><em>glgX</em> (ccce_3465)</td>
<td></td>
<td></td>
<td>L L</td>
</tr>
<tr>
<td><em>psbA4</em> (1st DP) (ccce_3477)</td>
<td></td>
<td></td>
<td>D L</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5. <strong>Circadian—Dark—Up-regulated (2nd DP)</strong></th>
<th>12h</th>
<th>LL SD&lt;sup&gt;b&lt;/sup&gt;</th>
<th>6h LD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peroxiredoxin (ccce_3126)</td>
<td></td>
<td>L</td>
<td>L L</td>
</tr>
<tr>
<td><em>hupLS</em> (ccce_1063 to 1064)</td>
<td></td>
<td>D</td>
<td>L D</td>
</tr>
<tr>
<td><em>nif</em> cluster (ccce_0545 to 0588)</td>
<td></td>
<td></td>
<td>L L</td>
</tr>
<tr>
<td><em>feoA2B2</em> (ccce_0575 and 0576)</td>
<td></td>
<td></td>
<td>D L</td>
</tr>
<tr>
<td>Response regulator (<em>rpaA</em>, ccce_0298)</td>
<td></td>
<td></td>
<td>D L</td>
</tr>
</tbody>
</table>

<sup>a</sup> Darkened boxes show that the expression of the respective genes peaked at the indicated time in the specific experiment (top). Gray background indicates a significant increase in expression, but not a peak. Experiments are 12 hour light/dark, light/dark followed by continuous light (peak expression in the subjective dark, SD, is shown) and the experiment reported in this paper, 6 hour light/dark. Group headings (1-5) summarize the circadian behavior suggested by the expression patterns shown.
Table 2. Context likelihood of relatedness (CLR) Analysis of *Cyanothece* sp. ATCC 51142 Transcription for four microarray experiments

<table>
<thead>
<tr>
<th>Function</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Photosynthesis Cluster</strong></td>
<td></td>
</tr>
<tr>
<td>PSII</td>
<td>Phycobilisomes (<em>apc &amp; cpc</em>)</td>
</tr>
<tr>
<td></td>
<td>Oxygen-evolving complex (<em>psbO, psb U, psbV</em>)</td>
</tr>
<tr>
<td></td>
<td><em>psbA4</em></td>
</tr>
<tr>
<td></td>
<td><em>psbEFJ, psb28</em></td>
</tr>
<tr>
<td>PSI</td>
<td><em>psa FJ</em></td>
</tr>
<tr>
<td>Cytb₆f</td>
<td><em>petACl, petF3F4</em></td>
</tr>
<tr>
<td>ATPase</td>
<td><em>atp operons</em></td>
</tr>
<tr>
<td>CO₂ Fix</td>
<td><em>rbcLXS</em></td>
</tr>
<tr>
<td></td>
<td><em>ccmMLK1K2</em></td>
</tr>
<tr>
<td>glycogen</td>
<td><em>glgA2</em></td>
</tr>
<tr>
<td>Sigma Factors</td>
<td><em>sigG and sigH</em></td>
</tr>
<tr>
<td><strong>Nitrogen Cluster</strong></td>
<td></td>
</tr>
<tr>
<td>nitrogenase</td>
<td><em>cce_0545 – cce_0580</em> (including <em>nifHDK</em>)</td>
</tr>
<tr>
<td>uptake hydrogenase</td>
<td><em>hupLS</em></td>
</tr>
<tr>
<td>Regulatory</td>
<td><em>cce_0969 + 0970</em></td>
</tr>
<tr>
<td><strong>Respiratory/Carbohydrate Metabolism Cluster</strong></td>
<td></td>
</tr>
<tr>
<td>Regulatory</td>
<td><em>kaiB3</em></td>
</tr>
<tr>
<td>Sigma Factors</td>
<td>hybrid sensor and phytochrome (<em>cce_1982 – 1983</em>)</td>
</tr>
<tr>
<td>Cyt oxidase</td>
<td><em>coxA1B1C1</em></td>
</tr>
<tr>
<td>NADH DH</td>
<td><em>ndh LEID1F1</em></td>
</tr>
<tr>
<td>PPP</td>
<td><em>gnd, zwf, opcA</em></td>
</tr>
<tr>
<td>Carbo. metabolism</td>
<td><em>glpP1, glpX, ppc, talAB gap, pgl</em></td>
</tr>
</tbody>
</table>
References:


Figure Legends

Figure 1: (a) Photosynthetic rates, respiration rates and nitrogenase activity (relative activity) of *Cyanothece* sp. during 6 h LD cycles. Oxygen evolution increases upward, whereas oxygen consumption increases downward. (b) Dissolved oxygen and pH value of *Cyanothece* sp. grown in a bioreactor for 5 days under 12 h LD and 6 h LD cycles.

Figure 2: Comparison of the gene expression pattern of the (a) *rbc*LS, (b) *psa*AB and (c) *cox*BAC1 genes in *Cyanothece* sp. grown under normal 12 h LD (squares) or short LD (6 h) cycles (triangles).

Figure 3: Transcriptional changes in selected genes of *Cyanothece* sp. ATCC 51142 after growth in 6 h LD cycles. (a) *hupS* (cce_1063) (♦) and peroxiredoxin (cce_3126) (■); (b) *cphB* (cce_2236) (▲), *hoxE* (cce_2319) (♦), and response regulator (cce_4542) (■); (c) response regulator, *rpaA*, (cce_0298) (♦), response regulator (cce_0970) (■), and *gnd* (cce_3746) (▲). The results were plotted in separate graphs due to the large differences in transcript levels for the genes in panel a (300,000 counts maximum), panel b (3,000 counts maximum) and panel c (35,000 counts maximum).

Figure 4: Variation of the glycogen content and gene expression of glycogen production (synthase; *glg*A2) and breakdown (debranching; *glg*X and phosphorylase; *glg*P3) genes. The glycogen content is displayed as % of the control (t=0) and the error bars represent standard deviation (n=3).

Figure 5: Inferred network of diurnal and circadian associations. The context likelihood of relatedness (CLR) algorithm was used to infer statistically significant gene to gene associations based on the combined expression profiles from experiments: growth in 12h LD, growth in 12 h LD followed by continuous light, the current experiment, and growth following transition into anaerobic conditions. In the network circles represent genes and edges between them an association with a CLR zscore of 5.0 or higher. The genes are colored according to
the transcriptional levels at 5h Dark (red, high; green, low). Three clusters of interest are labeled and correspond to the groups described in Table 2.

Figure 6: Validation of the microarray results by comparison of the microarray results with those from RT-PCR experiments. The genes chosen were examples of different categories of differential expression, including: $nifH$ (●), highly expressed in the second dark; $coxCI$ (▲), highly expressed during both dark periods and the second light period; $psbA$ (■) and $psbD$ (★), highly expressed in the light periods; and $rnpA$ (▲), little change in expression throughout the experiment. The microarray data is plotted on a logarithmic scale.