Supplemental material:

Culturing of *Synechocystis* sp. PCC6803 with N_2/CO_2 in a diel regime shows multi-phase glycogen dynamics and low maintenance costs

S. Andreas Angermayr[a+], Pascal van Alphen[a], Dicle Hasdemir[b], Gertjan Kramer[c]*, Muzamal Iqbal[d], Wilmar van Grondelle[e], Huub C. Hoefsloot[b], Young Hae Choi[d], Klaas J. Hellingwerf[a,e]#

[a]Molecular Microbial Physiology, Swammerdam Institute for Life Sciences, University of Amsterdam, The Netherlands

[b]Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, The Netherlands

[c]Medical Biochemistry, Academic Medical Center, Amsterdam, The Netherlands

[d]Natural Products Laboratory, Institute of Biology, Leiden University, The Netherlands

[e]Photanol B.V., Science Park 408, Amsterdam, The Netherlands

+ present address: Institute of Science and Technology Austria, Klosterneuburg, Austria

* present address: Proteomics Research Group, EMBL Heidelberg, Germany

corresponding author at Swammerdam Institute for Life Sciences, Science Park 904, 1098 XH Amsterdam, The Netherlands, tel: +31-20-525-7055, e-mail: K.J.Hellingwerf@uva.nl

1 Supplemental Materials and Methods:

1.1 Metabolite detection by ^{1}H NMR and statistical evaluation

Rapid sampling from the PBRs for metabolite determination was performed as follows: 50 ml of cell suspension from the PBR vessel was harvested through a sampling port with a syringe into a 50 ml tube. For samples from the dark period the tubes were covered with aluminum foil to avoid unwanted illumination. Cells were pelleted by centrifugation at 6,000 rpm at 4 °C for 5 min (in a pre-chilled centrifuge, Eppendorf centrifuge 5403, 16F6-38 rotor). The supernatant was removed carefully and the remaining wet pellet was snap-frozen in liquid N_2 and stored at -80 °C. After collection of all samples from the whole time period the tubes were collected on liquid N_2, caps were replaced with caps containing holes and samples were freeze-dried. After 24 to 48 hours the tubes were collected on liquid N_2 and stored at -80 °C. This protocol aims for the recovery of a dry cell pellet of about 25 mg for ^{1}H Nuclear Magnetic Resonance (NMR) spectroscopy analysis. The NMR-based metabolite-analysis method followed the guidelines described earlier (1). Briefly, ^{1}H NMR spectra were acquired by the analysis of the extract of the freeze-dried cell material in KH_2PO_4 buffer in D_2O (pH 6). As internal standard for concentration determination 0.54 mM trimethyl-silyl propionic acid solution salt-d_4 (TMSP-d_4) was used with pre-saturating residual HDO signal. Before multivariate analysis of the metabolite profiles, each NMR spectrum was divided into series of bins, generating profiles with approximately 256 peaks using AMIX (ver. 3.6, Bruker, Karlsruhe, Germany), as detailed in Kim *et al.* (1). After baseline correction, partial least square discriminant analysis (PLS-DA) was applied (SIMCA 13.0, Umetrics, Umeå,

Sweden) following Pareto-scaling of the data. PLS-DA allowed us to group the samples according to the time points at which they were taken. The peaks that contributed the most to the separation of samples from different time points were then detected by ranking them based on their variable importance in the projection (VIP) scores (from the SIMCA software package). Manual comparison of the peaks with the highest VIP scores to a database allowed us to identify the metabolites that were the most influential in the separation of samples.

1.2 Proteome determination by mass spectrometry (MS) analysis

Sampling from the PBRs for proteome analysis was performed as follows: 10 ml of cell suspension from the vessel was harvested via a sampling port with a syringe into a 15 ml tube (on ice) that already contained 5 ml of ice-cold 50 mM ammonium bicarbonate (FlukaBioUltra >99.5 %). For samples from the dark period the tubes were thoroughly covered with aluminum foil. Cells were pelleted by centrifugation at 6,000 rpm at 4 °C for 10 min (in a pre-chilled centrifuge, Eppendorf centrifuge 5403, 16F6-38 rotor). The supernatant was discarded completely and cells were re-suspended in 1 ml of ice-cold 50 mM ammonium bicarbonate, snap-frozen in liquid N_2 and stored at -20 °C. Prior to the disruption of the cells by sonication, the cell suspension was supplemented with 0.1 % (w/v) Rapigest SF (Waters Corporation, Manchester, UK). Cell lysis was performed by 1 round of sonication with a micro-tip (duration: 5 min, with output strength of 5, and a 50 % duty cycle, with a probe-type sonicator (Branson) on melting ice). Cell debris was pelleted by centrifugation at 14,000 rpm at 4 °C for 30 min (in a pre-chilled table top-centrifuge). The protein content of the supernatant, containing the soluble fraction of the crude cell-free extract, was immediately determined by a BCA assay, using bovine serum albumin as the standard, according to the manufacturers protocol (Pierce). The protein concentration of the lysates ranges from ~0.5 mg/ml to ~1.0 mg/ml. The concentration of the lysates was brought to the same level (0.5 mg/ml) by diluting with 0.1 % Rapigest SF in 50 mM ammonium bicarbonate. A preliminary investigation of the crude cell-free extract prepared in an identical manner on SDS-PAGE showed no significant protein degradation at room temperature for up to 3 hours (data not shown). Subsequently, samples were reduced with 5 mM dithiothreitol (Sigma-Aldrich) at 60 °C for 30 min and alkylated with 15 mM iodoacetamide (Sigma-Aldrich) at ambient temperature in the dark for 30 min. Proteolytic digestion was initiated by addition of modified trypsin gold grade (Promega, Madison, WI, USA) at a 1:50 protease:protein (w/w) ratio, followed by overnight incubation at 37 °C. Following digestion, breakdown of the acid-labile detergent (Rapigest SF) was induced by the addition of 1 %

trifluoroacetic acid to achieve a pH < 2 and subsequent incubation at 37 °C for 45 min. The peptide solutions were centrifuged at 13,000 rpm for 10 min and supernatant was collected. Prior to analyses a protein digest, MassPREP Protein Digestion Standard (Waters Corporation) was added, at a 1:1 dilution to the sample (50 fmol/µl alcohol dehydrogenase (ADH1) and enolase (ENO) from *Saccharomyces cerevisiae*).

Nanoscale LC separations of tryptic peptides were performed with a NanoAcquity system (Waters Corporation). Samples were loaded onto a Symmetry C18 5 µm, 2 cm x 180 µm trap column (Waters Corporation) at a flow rate of 5 µl/min prior to separation on a Bridged Ethyl HybridC18 1.7 µm, 25 cm x 75 µm analytical reversed phase column (Waters Corporation) by application of a 90 min gradient from 1 % acetonitrile, 0.1 % formic acid to 40 % acetonitrile, 0.1 % formic acid (all v/v) at a column flow rate of 0.300 µl/min. Analysis of eluting tryptic peptides was performed using a Synapt G2 quadrupole time-of-flight mass spectrometer (Waters Corporation, Milford, MA, USA) equipped with a nanolock spray source (Waters Corporation), fitted with a pico-tip emitter (New Objective, Woburn, MA). Operated values were: around 3 kV capillary voltage, cone voltage of 40 V, a source temperature of 90 °C and TOF-voltage set at 7 kV. The collision gas used was argon, maintained at a constant pressure of 2 µbar in the collision cell. The lock mass, [Glu^1]-Fibrinopeptide B, was delivered from the auxiliary pump of the NanoAcquity system at a concentration of 100 fmol/µl and a supply rate of 0.5 µl/min to the reference sprayer of the nanolock spray source, which was sampled every 120 sec. The data were post-acquisition lock-mass corrected using the mono-isotopic mass of the doubly charged precursor of [Glu^1]-Fibrinopeptide B. Accurate mass precursor and fragment ion LC-MS data were collected in data independent $LCMS^E$ mode of acquisition (2, 3). The samples for each of the five time points stem from two biological replicates (the two PBRs) each processed twice with respect to cell lysate preparation and processing and they were additionally injected twice into the LC-MS.

Continuum LC-MS data were processed using ProteinLynxGlobalSERVER version 2.5 (PLGS 2.5, Waters Corporation). Parameter settings: digest reagent: trypsin, allow 1 ‘missed cleavage’, search tolerances: automatic, typically 5 ppm for precursor- and 15 ppm for product ions, fixed modification cysteine: carbamido-methylation, and variable modification: methionine: oxidation. Protein identifications were obtained searching the entries of a *Synechocystis* proteome database (UniProt), with ADH1 and ENO1 of *S. cerevisiae* appended as internal standard to address technical variation and allow for absolute quantitation of

protein amounts. Estimation of false-positive identification rates was done by searching a randomized version of the abovementioned *Synechocystis* protein database generated within PLGS 2.5. HI3 peptide quantitation was used to quantify fmoles (label free absolute quantitation) of each protein measured by use of the known amounts of the added Massprep standards (4). Data were exported as csv-files and fmoles determined by HI3 peptide quantitation were normalized by summed fmoles determined for each LC-MS run to adjust for variability in sample content and instrument performance. Any protein was required to be detected in a minimum of 9 (out of 40) injections, in order to be regarded as reliably identified and quantified, which led to identification of 357 *Synechocystis* proteins with a false discovery rate of < 1 % as determined by the number of identifications from the randomized database (data not shown). As the data acquired needed to reflect the protein profile over the whole time series, to probe the daily cycle of *Synechocystis*, we applied a second round of selection criteria: i.e. proteins needed to be detected over all time points measured, in at least two technical replicates for both PBRs. This provided a list of 176 proteins for which the fmoles (nfmols) are shown in the supplemental material (Additional Material Table A1).

1.3 Transcriptome determination by microarray analysis

Sampling from the two PBRs for transcript analysis was performed as follows: 10 ml of cell suspension from each vessel was harvested via a sampling port with a syringe into a cold 15 ml tube pre-chilled on ice. For samples from the dark period the tubes were shielded from light. Cells were harvested through centrifugation at 6,000 rpm at 4 °C for 10 min (in a pre-chilled centrifuge). The supernatant was discarded and the cell pellet was snap-frozen in liquid N_2 and stored at -80 °C. RNA was extracted following the manufacturers' instructions from the RNeasy mini kit (Qiagen), starting with the respective lysis buffer supplemented with β-mercaptoethanol and opening of cells through bead-beating (4 times at 6,000 rpm in a Precellys®24 bead beater, Bertin technologies) with 4 N HCl-washed and twice autoclaved glass beads (100 µm diameter, Sigma). During RNA isolation on-column DNA digestion was performed using the RNase-Free DNase Set (Qiagen). RNA was eluted with Nuclease-Free Water (not DEPC-Treated) (Life Technologies). The RNA concentrations were measured on the NanoDrop ND-2000 (Thermo Scientific). The integrity of the RNA samples was assessed on a 2200 TapeStation system (Agilent Technologies) using the RNA ScreenTape (Agilent Technologies) yielding RINe values > 9.3. Per sample, 2 µg of total RNA was combined with 1 µg random octamers (Biolegio), denatured at 65 °C for 10 min and placed on ice-water for

10 min. Subsequently, a first strand master mix was added containing first strand buffer (Life Technologies), 0.5 mM dGAC, 0.35 mM dUTP, 0.15 mM dUTP-Cy3 (GE Healthcare) and 200 U SuperScript II (Life Technologies). This mixture was subsequently incubated for 2 min at 25 °C, 120 min at 42 °C and 15 min at 70 °C. Finally, NaOH was added to hydrolyze the remaining RNA by heating at 70 °C for 10 min. The reaction was stopped by adding MOPS buffer and the labeled cDNA was purified with the E.Z.N.A. MicroElute RNA Clean-up Kit (Omega Biotek). Dye incorporation and cDNA yield were measured on the NanoDrop ND-2000 (Thermo Scientific) yielding a frequency of incorporation of > 10 pmol/µg.

Hybridization mixtures were made using 600 ng cDNA according to the One-Color Microarray-Based Gene Expression Analysis guide version 6.5 (G4140-90040, Agilent Technologies) without the inclusion of the RNA fragmentation mixture. The samples were loaded onto 8x15,000 custom designed microarrays against *Synechocystis* sp. PCC 6803 (5) (Design ID: 047125, Agilent Technologies) and hybridized for 17 hours at 65 °C. Subsequently, the slides were washed and scanned with an Agilent G2565CA scanner as described in the One-Color Microarray-Based Gene Expression Analysis guide version 6.5 (G4140-90040, Agilent Technologies). For data extraction we used the feature extraction software, version 10.7.3.1 protocol GE2_107_Sep09 (Agilent Technologies). For each identifier the gMedianSignal was used. To compensate between-chip differences in our experimental setting, we applied the Least Oscillatory Set (LOS) normalization, following the advice of Lehmann *et al.* (6). Lehmann *et al.* present an extensive study on normalization methods that can be used for oscillatory gene expression data without disturbing the oscillatory behavior of the different genes. The results from their study show that the LOS normalization is the most appropriate method tested and it achieves patterns that agree best with available biological knowledge. We refer the readers to Lehmann *et al.* for more details on the method and its implementation in R. After pre-processing of the dataset, we performed ANOVA to detect the genes which show a time-dependent change.

1.4 Statistical analysis and evaluation of proteome and transcriptome data

To account for a possible PBR effect and biological variation between the two cultures, we performed ‘repeated measurements ANOVA’ with a fixed factor (i.e. time) and a random factor (i. e. PBR). It allowed us to detect time dependent changes in the proteins and genes. When we used ANOVA only as a filtering step prior to clustering, we used a very loose threshold of p-value = 0.1 in order not to exclude many possibly relevant genes from further analysis.

For k-means clustering of both genes and proteins, we used the 'kmeans' function implemented in Matlab (7). The function uses k-means++ algorithm for initializing the cluster centers (8). We repeated the algorithm with 200 different starting points and selected the clustering that gives the smallest distances between the instances and the cluster centroids.

We used the mean of biological or biological and technical replicates for the clustering of the transcriptome and the proteome data, respectively. Prior to clustering, we transformed the data to account for the differences in the absolute values of gene expression and protein abundance. For this purpose, we used z-transformation (see Equation 1).

$$x_Z = \frac{x - \bar{x}}{\sigma_x} \qquad \text{(Equation 1)}$$

where x, x_Z, $\bar{x}$ and σ_x represent the gene expression/protein abundance at a time point, the standardized value of gene expression/protein abundance, the mean of gene expression/ protein abundance across all five time points and the standard deviation of gene expression/ protein abundance across all five time points, respectively. Performing k-means on z-transformed data by using the Euclidean distance metric allowed us to focus only on the relative temporal changes of the genes/proteins.

For the enrichment analysis, we used the functional categories provided in the Cyanobase database (9). We computed the p-values of enrichment by using the cumulative hypergeometric distribution function, 'hygecdf' in Matlab following Equation 2.

$$p - value = 1 - \sum_{i=0}^{x-1} \frac{\binom{K}{i}\binom{M-K}{N-i}}{\binom{M}{N}} \qquad \text{(Equation 2)}$$

where K is the number of the genes in a specific category, N is the total number of genes in the cluster, x is the number of genes that are in the cluster and are associated with the specific category in question and M is the total number of genes in the background set. For the genes, we used a background set of 3,263 genes which is the set of all genes on the microarray. For

the proteins, we used a background set of 176 proteins which is the set of proteins used for further analysis.

2 Supplemental Figures

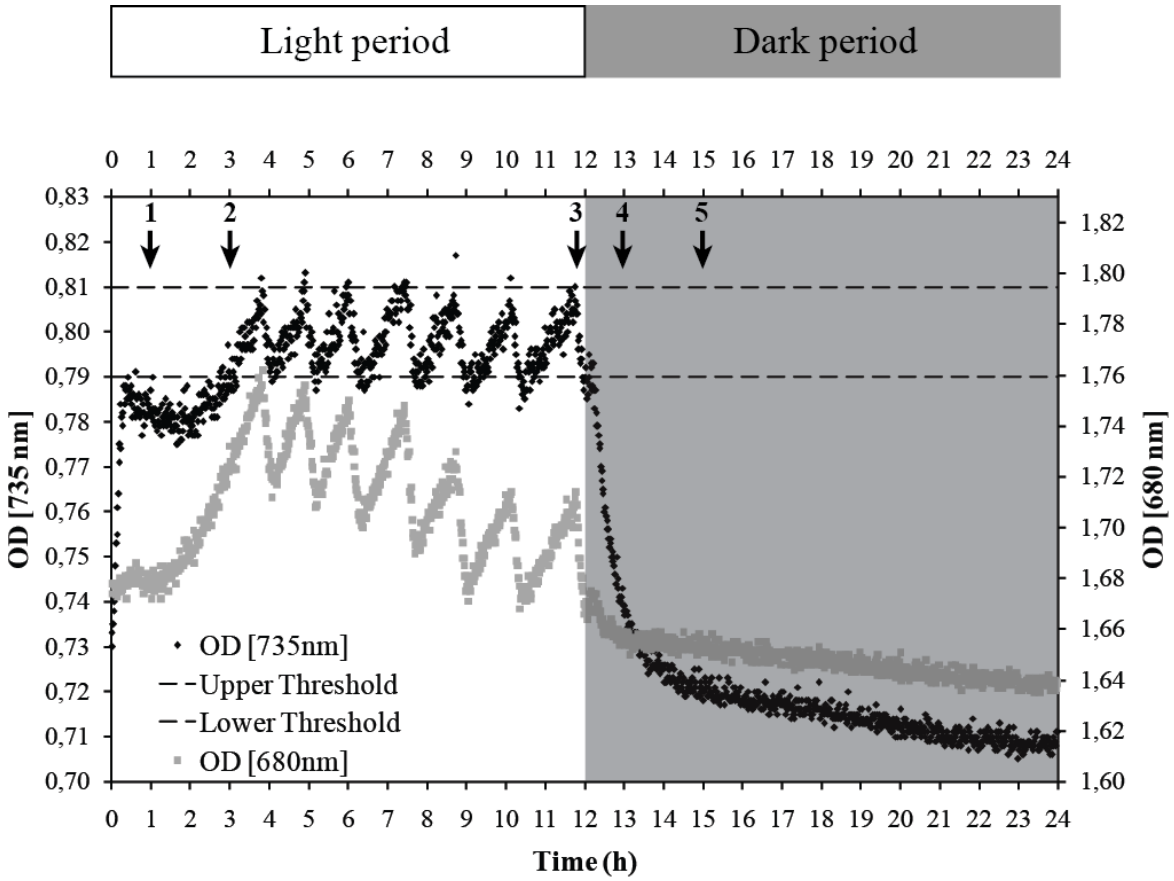

Figure S1: A turbidostat-controlled culture treated with a re-occurring diel light regime of 12 hours light and 12 hours darkness. Optical density (OD) at 735 nm and 680 nm over the period of 24 hours: black and grey dots, respectively. The thresholds for the clamped OD_{735} values are indicated by the dashed lines. The arrows indicate the sampling time points for gDW, metabolites, proteins and transcripts. The rather sharp increase and decrease exactly after the change of the light condition in the OD_{735} reading is considered to be a technical artifact of the detector caused by the temperature change due to the presence and absence of light. The steady decrease of the ODs at 735 and 680 nm during the dark period could originate from cell lysis, changed cellular absorption properties, changed cell size or a combination of those effects.

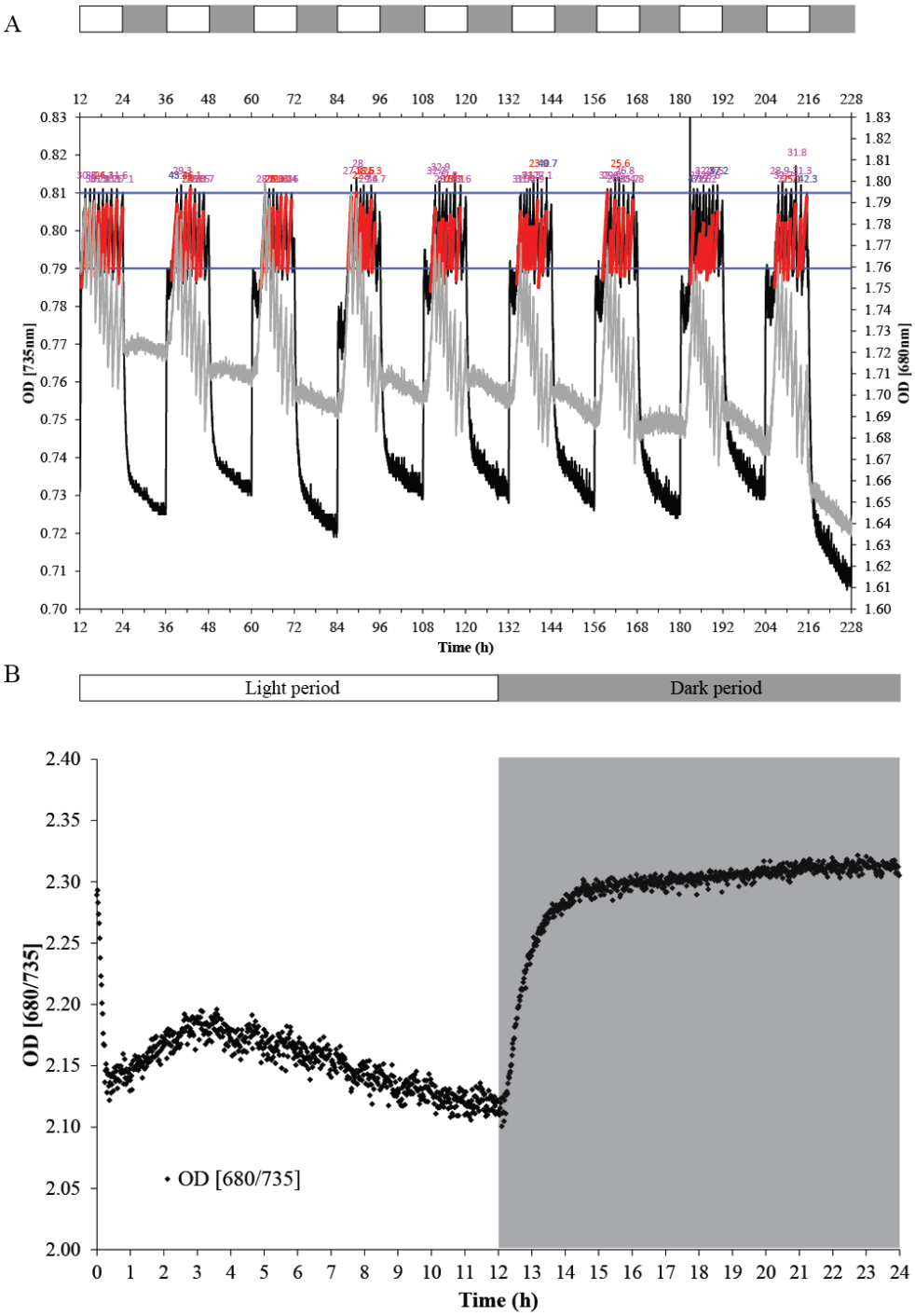

Figure S2: (A) Optical density recordings. OD_{735} (black) and OD_{680} (grey) reading in 1 min intervals over a time period of 1 week, before sampling took place. Absorption values are from the built-in photocells in the PBR. The horizontal blue lines indicate the OD_{735} thresholds for the turbidostat control. The red lines indicate a best fit with an exponential equation for growth rate determination in the time intervals between a dilution by the pump. The numbers on top of the fit indicate the doubling time as calculated from the growth rate. A minor slope for the absolute values of OD_{680} is visible, but the slope of this signal becomes gradually smaller towards the end of the 1 week period, after which sampling took place (on day 8) and the OD_{680}/OD_{720} ratio does not change. **(B)** OD680/OD735 ratio for one day as in

Figure S1. This ratio functions as a measure for chlorophyll content and shows a pattern of increasing/decreasing chlorophyll which is phase advanced with respect to growth rate (compare also Figure 1).

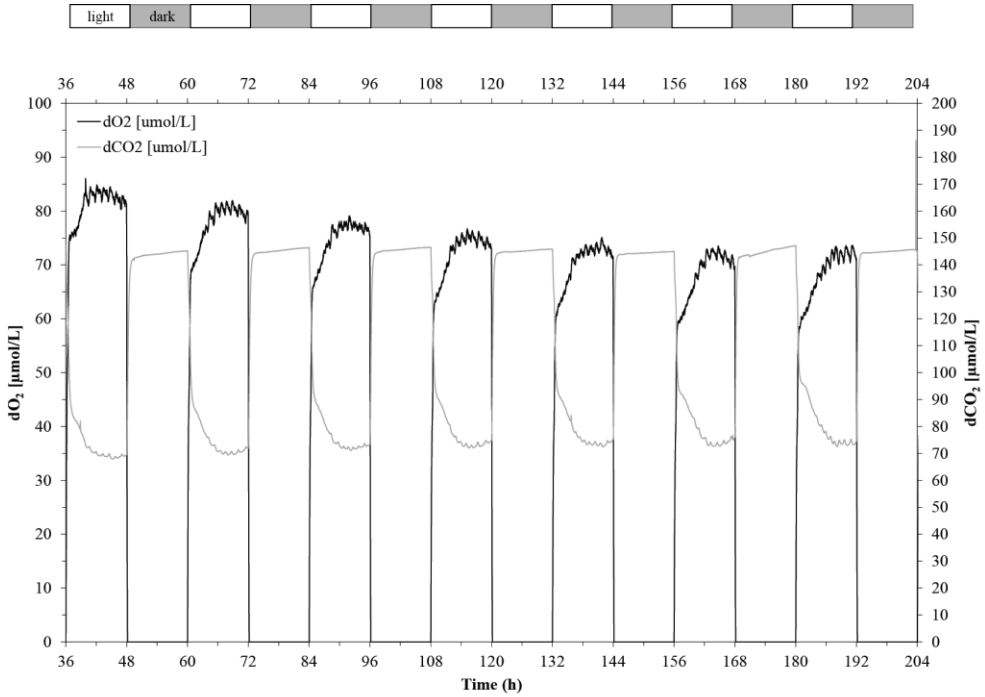

Figure S3: Oxygen and carbon dioxide recordings. dO_2 (black) and dCO_2 (grey) reading in 1 min intervals. A minor slope for the concentration of O_2 is visible, but the slope of this signal becomes gradually smaller towards the end of the 1 week period, after which sampling took place.

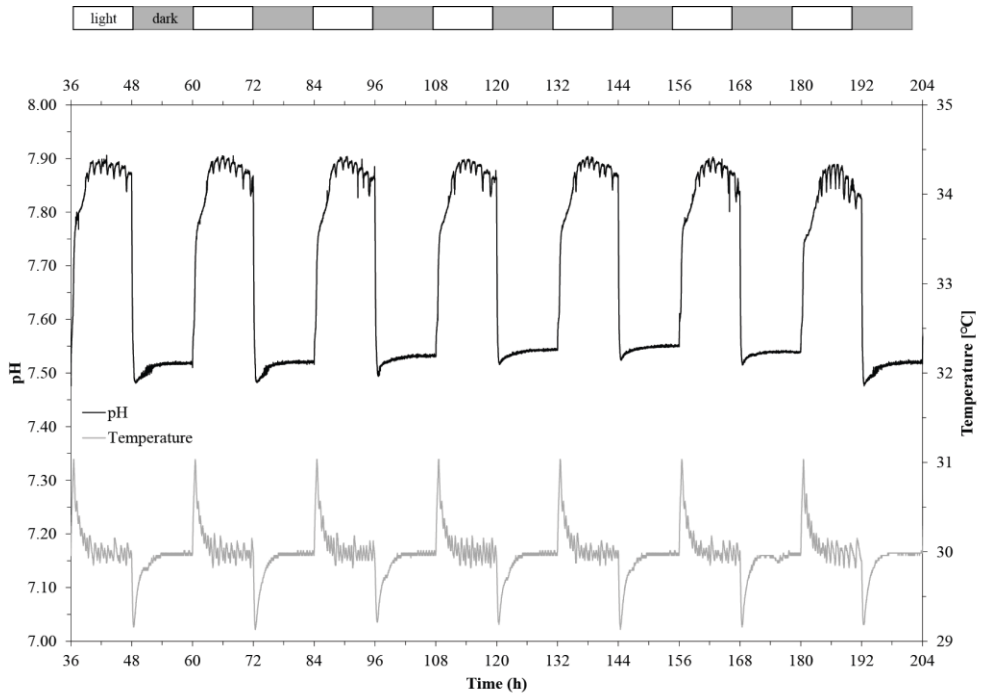

Figure S4: pH and temperature recordings. pH (black) and temperature (grey) reading in 1 min intervals.

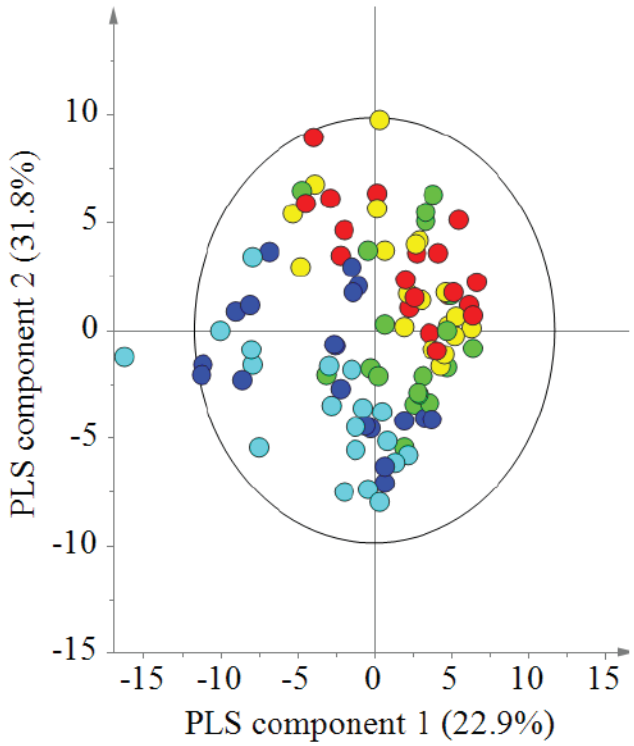

Figure S5: Multivariate data analysis employing partial least square discriminant analysis of the metabolite signals collected by NMR analysis. Significant changes in the metabolite profile over time lead to separation according to the sampling time point. The sampling time points (TP) are color coded as follows: TP1: dark blue, TP2: light blue, TP3: green, TP4: yellow, TP5: red; One dot represents the metabolite profile of one sample time point form one of the two PBRs which were sampled for 9 and 10 consecutive days, respectively.

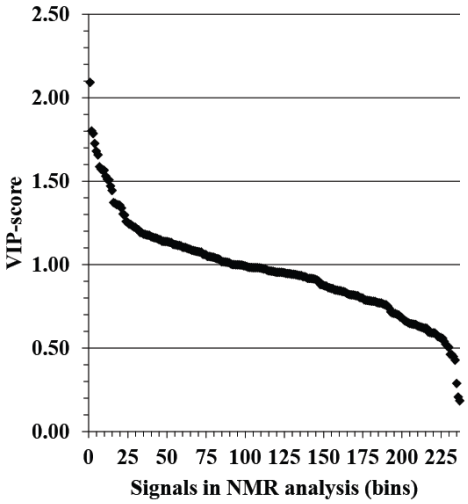

Figure S6: Fraction of dynamic metabolites. About 40 % of the metabolites change over time. A VIP score >1 shows that the respective compound had an influence on the separation in the PLS-DA, whereas a VIP score <1 shows the opposite.

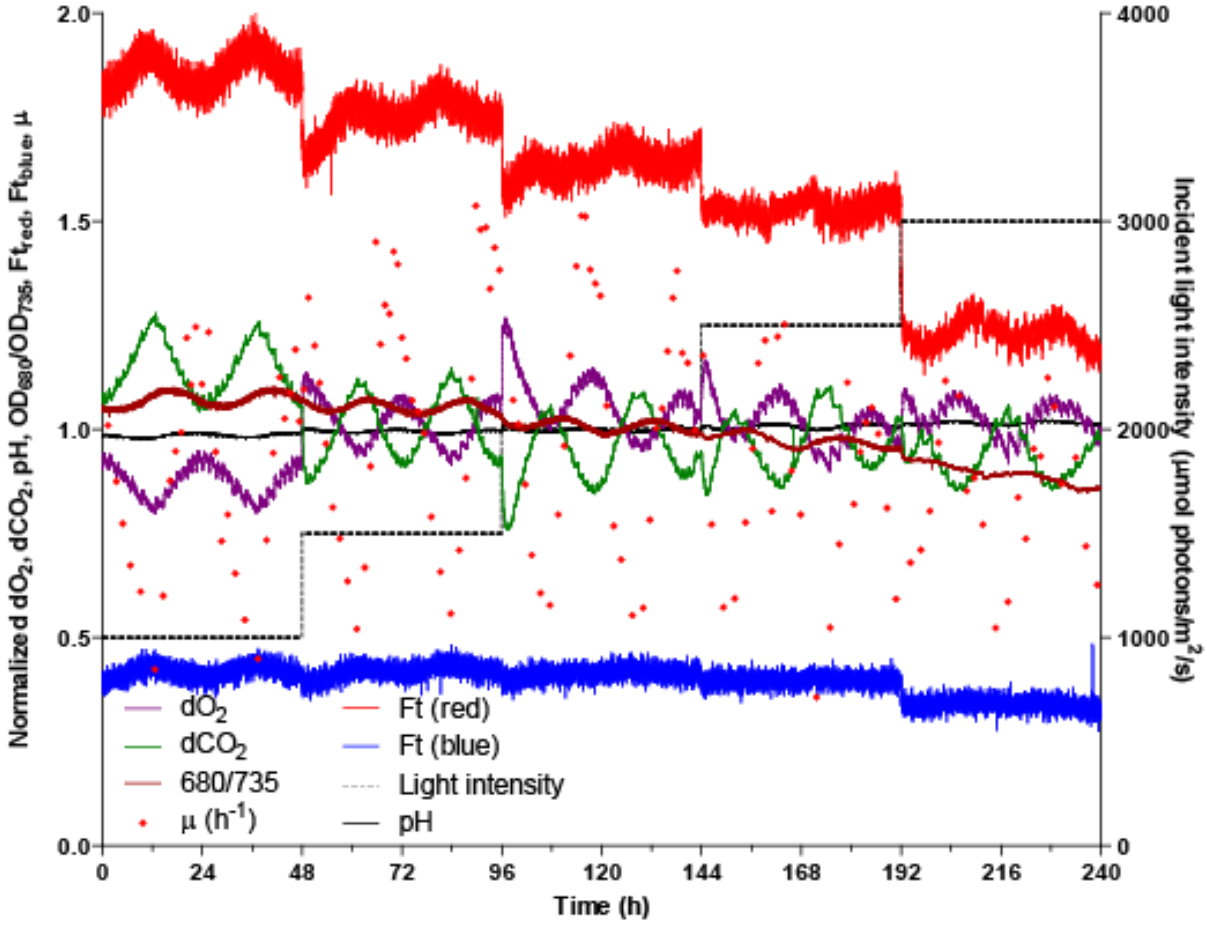

Figure S7: Test for the occurrence of photoinhibition. After concluding sampling under the 12 h / 12 h light/dark (LD) regime, light intensity was increased in steps of continuous illumination to assess whether *Synechocystis* is photoinhibited under the chosen conditions. This would be reflected in a steadily decreasing activity of PSII upon switching to a given light intensity. No evidence for photoinhibition was found up to 3,000 µmol photons/m^2/s of combined red and blue (1:1) light (see Materials and Methods for details). Light intensity (dotted line, right axis) is shown in µmol photons/m^2/s, whereas all other measurements are normalized to their respective mean in order to fit in the same graph. For clarity, Ft (red) and Ft (blue) are normalized to their combined mean. Means are: 7.97 (pH), 74 µM (dO_2), 70 µM (dCO_2), 885 a.u. (Ft), 0.019 h^{-1} (µ), 2.02 (OD_{680}/OD_{735}).

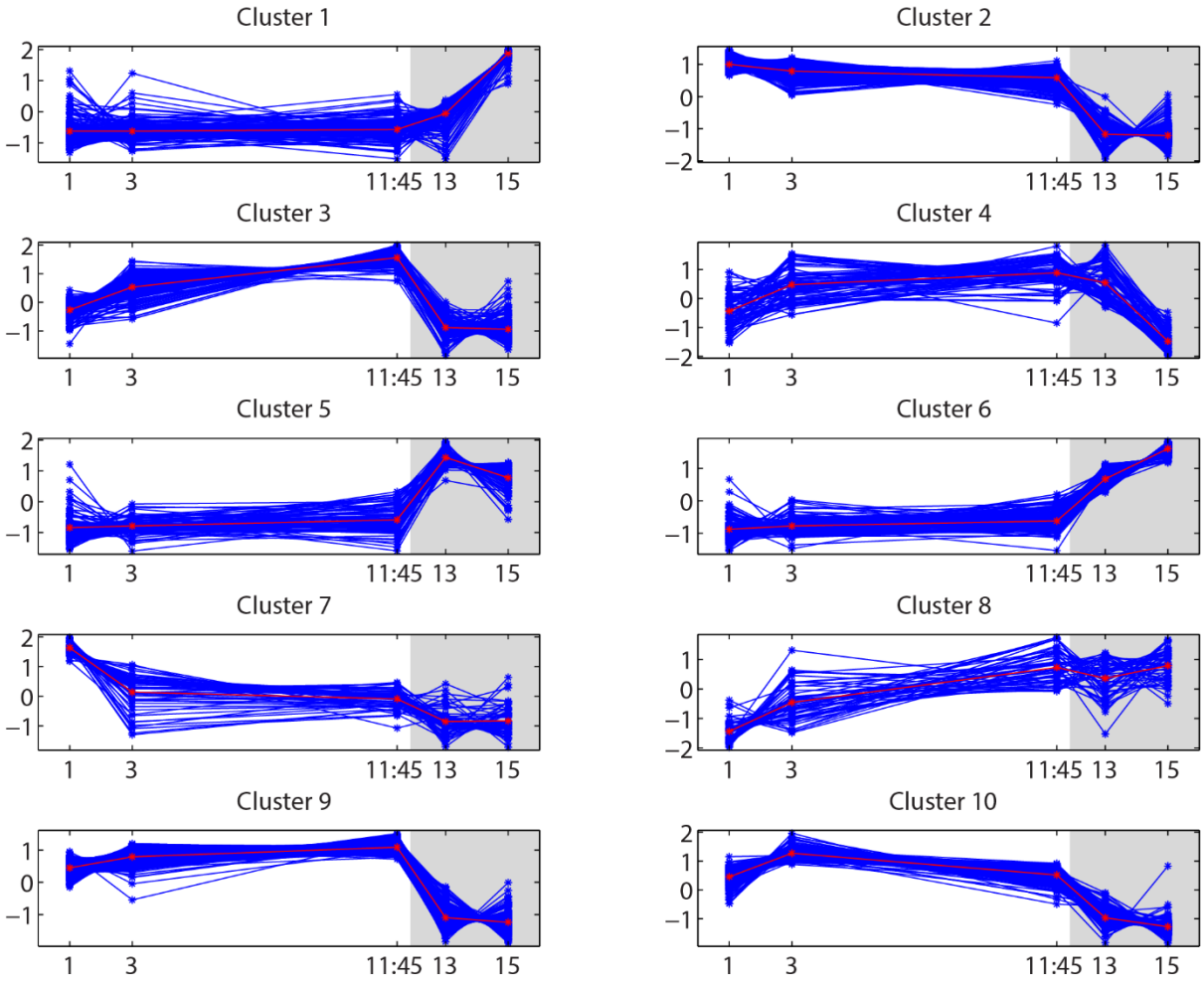

Figure S8: K-means clustering of transcripts which show time-dependent change (i. e. 56 % of all). Red profiles indicate the centroids of each cluster.

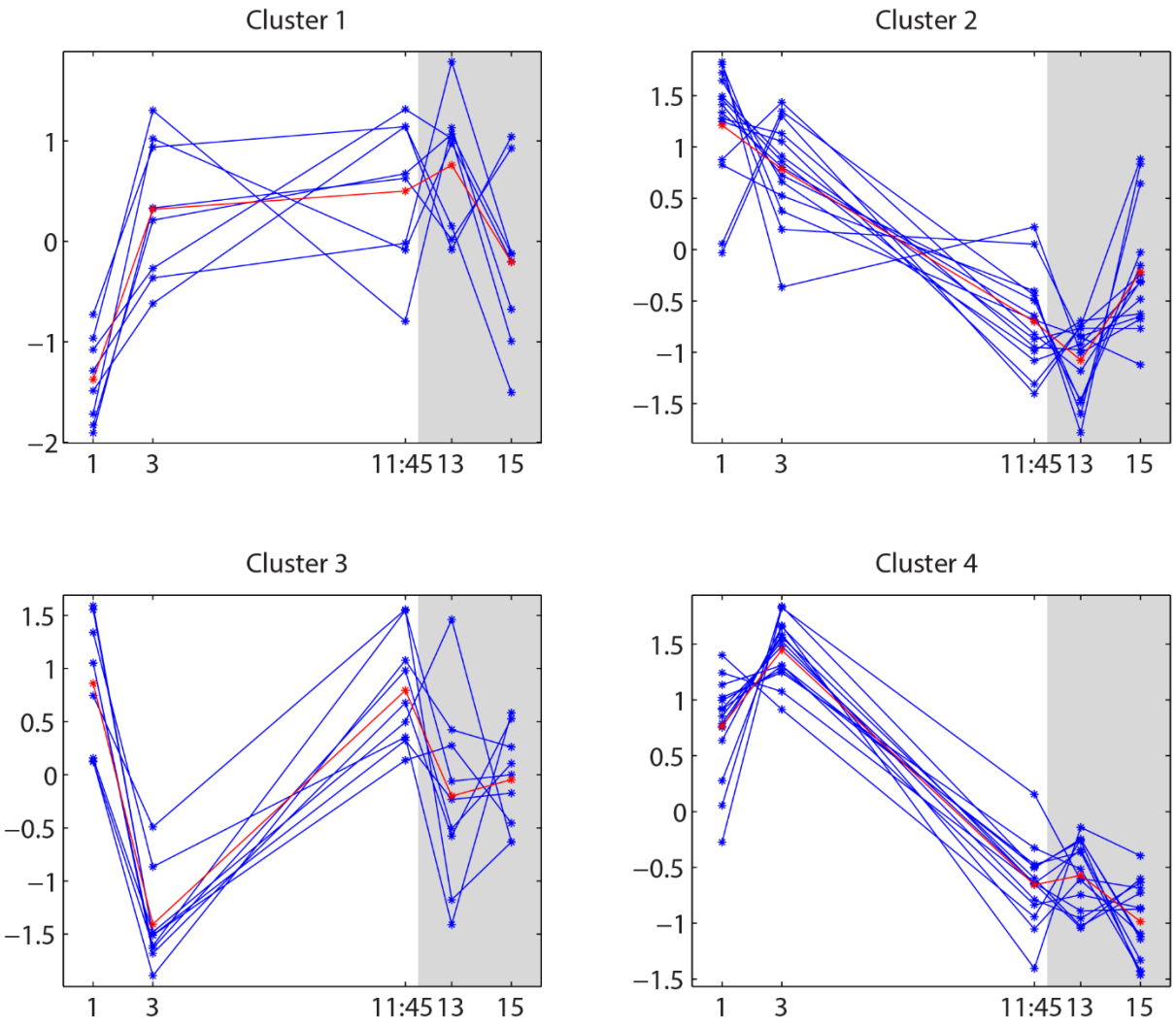

Figure S9: Clustering of dynamic proteins. K-means clustering of the 45 most changing proteins. Red profiles indicate the centroids of each cluster.

3 Supplemental References

1. **Kim HK**, **Choi YH**, **Verpoorte R**. 2010. NMR-based metabolomic analysis of plants. Nat Protoc **5**:536–549.

2. **Geromanos SJ**, **Vissers JPC**, **Silva JC**, **Dorschel CA**, **Li G-Z**, **Gorenstein MV**, **Bateman RH**, **Langridge JI**. 2009. The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. Proteomics **9**:1683–1695.

3. **McLoughlin F**, **Arisz SA**, **Dekker HL**, **Kramer G**, **de Koster CG**, **Haring MA**, **Munnik T**, **Testerink C**. 2013. Identification of novel candidate phosphatidic acid-binding proteins involved in the salt-stress response of Arabidopsis thaliana roots. Biochem J **450**:573–581.

4. **Silva JC**, **Gorenstein MV**, **Li G-Z**, **Vissers JPC**, **Geromanos SJ**. 2006. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics MCP **5**:144–156.

5. **Aguirre von Wobeser E**, **Ibelings BW**, **Bok J**, **Krasikov V**, **Huisman J**, **Matthijs HCP**. 2011. Concerted Changes in Gene Expression and Cell Physiology of the Cyanobacterium Synechocystis sp. Strain PCC 6803 during Transitions between Nitrogen and Light-Limited Growth1[W][OA]. Plant Physiol **155**:1445–1457.

6. **Lehmann R**, **Machné R**, **Georg J**, **Benary M**, **Axmann I**, **Steuer R**. 2013. How cyanobacteria pose new problems to old methods: challenges in microarray time series analysis. BMC Bioinformatics **14**:133.

7. **Lloyd S**. 1982. Least squares quantization in PCM. IEEE Trans Inf Theory **28**:129–137.

8. **Arthur D**, **Vassilvitskii S**. 2007. K-means++: The Advantages of Careful Seeding, p. 1027–1035. *In* Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

9. **Nakao M**, **Okamoto S**, **Kohara M**, **Fujishiro T**, **Fujisawa T**, **Sato S**, **Tabata S**, **Kaneko T**, **Nakamura Y**. 2010. CyanoBase: the cyanobacteria genome database update 2010. Nucleic Acids Res **38**:D379–381.