An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC 6803

Jan Mitschke[a,1], Jens Georg[a,1], Ingeborg Scholz[a], Cynthia M. Sharma[b], Dennis Dienst[c], Jens Bantscheff[a], Björn Voß[a], Claudia Steglich[a], Annegret Wilde[d], Jörg Vogel[b] and Wolfgang R. Hess[a,e,2]

Supplementary Information

1. Methods

2. Supplementary Data Files: please download from http://www.cyanolab.de/Supplementary.html

3. Supplementary Figures

4. Supplementary Tables

5. Supplementary Material References

1. METHODS

Bacterial strains and growth conditions. Liquid cultures of *Synechocystis* 6803 wild-type and mutant strains (Moscow subtype) were grown at 30°C in BG11 medium under continuous illumination with white light of 50 µmol photons m^{-2} s^{-1}. Control cultures that were later studied by transcriptome microarray analysis were grown to an OD of 0.6 at 750 nm and an illumination of 50 µmol photons m^{-2} s^{-1}. Subcultures were either transferred to the dark for 1 h, transferred to high light (500 µmol photons m^{-2} s^{-1}) for 30 min, or depleted for CO_2 by washing once in carbon-free BG11 (w/o $NaCO_2$, pH 7.0) and then cultivated in carbon-free BG11 for 6 h without aeration.

Mutagenesis. The plasmid used for overexpression of SyR1 was constructed by fusing the *petJ* promoter fragment and the *syr1* gene in a two-step PCR using primers PpetJfm, Syr1-PpetJrm, PpetJ-Syr1fm, and Syr1rm (**Table S10**). The resulting 427 bp fragment was ligated into the pDrive vector (Qiagen, Valencia, CA), excised by PstI/SalI digestion and finally inserted into plasmid pVZ-spec (modified version of the self-replicating vector pVZ321 (1)) in which the plasmid's Cm resistance cassette was replaced. As an additional selection marker, this vector contained a spectinomycin/streptomycin resistance cassette, which had been inserted into the plasmid XbaI site. The resulting plasmid, pVZ-syr1, was transferred to WT cells by conjugation (1) and exconjugants were selected on BG11 agar plates containing 40 µg ml^{-1} kanamycin and 20 µg ml^{-1} spectinomycin. Plasmid pVZ-spec was used for conjugation to generate an isogenic wild-type control strain. The *petJ* promoter, which is induced under Cu^{2+} deficiency, was activated for seven days under standard conditions in Cu^{2+}-free medium. Whole cell absorption spectra were recorded with a Shimadzu UV-2401PC spectrophotometer and normalized to the optical density at 750 nm.

Preparation of RNA and pyrosequencing. Total RNA was isolated as previously described (2). Details of the dRNA-seq method were provided elsewhere (3). In bacteria, most primary 5' ends resulting from initiation of transcription of mRNAs as well as ncRNAs carry a characteristic 5' tri-phosphate. In contrast, RNA fragments resulting from processing or degradation carry a 5'- mono-phosphate (5'P). These differences were employed here by synthesizing one cDNA library (-) from the original RNA pool containing both primary and processed transcripts, and another one following enrichment for primary transcripts by selective degradation of RNAs containing a (5'P) by treatment with Terminator^{TM} 5' P-dependent exonuclease (Epicentre), (+) library. For linker ligation RNA was treated with tobacco acid pyrophosphatase to generate 5'-monophosphates. The cDNA libraries were prepared and analyzed on a Roche FLX sequencer as previously described (4). After addition of specific 5'-linkers with unique tags for each library and poly-A-tailing, the RNA was converted into a cDNA library as previously described (5) but omitting size fractionation. A total of 169,360 and 188,723 sequence reads were obtained for the (-) and (+) populations, respectively. From these, 129,346 and 148,767 sequence reads were ≥18 nt in length and from these 106,018 and 131,943 sequence reads matched against the sequences of the genome or one of the four megaplasmids of *Synechocystis* 6803.

Experimental analysis of RNA for verification.

To confirm the pyrosequencing data independently, TSSs for 80 different genes were determined by 5' RACE following published methods (6) with modifications as described (7). Based on the fact that primary transcripts in bacteria carry a 5' tri-phosphate that can be cleaved specifically by tobacco acid pyrophosphatase, the

resulting 5' monophosphates were ligated to the 3' hydroxyl group of an RNA oligonucleotide (5' adaptor), followed by reverse transcription with a gene-specific oligonucleotide that was placed within the first 200 nt downstream of a start codon and PCR amplification with a 5' adaptor and a nested gene-specific primer. Dephosphorylation of RNA (10 µg) prior to ligation (control) was performed using 10 units of calf intestine phosphatase (AP Biotech, Sweden) in a buffer containing 50 mM Tris-HCl, pH 7.9, 100 mM NaCl, 10 mM $MgCl_2$, and 1 mM DTT, at 37 ^{o}C for 30 min, followed by phenol/chloroform extraction and ethanol precipitation. All enzymatic treatments of RNA were performed in the presence of 2 units of Super RNase Inhibitor (Ambion, USA). An amount of cDNA equaling 75 ng of starting RNA was used as a template in 30 µl PCR reactions, of which 15 µl (equal loading throughout the whole set of experiments) were run on 3% Nusieve agarose gels. The amplified fragments were cloned into plasmid PCR 2.1-TOPO (Invitrogen) or pGEMT (Promega). After transformation into *E. coli* XL1-Blue, plasmid inserts were screened by colony PCR. The PCR fragments were then purified on QIAquick spin columns (Qiagen, Germany), and sequenced using an ABI 373 automatic DNA sequencer (Applied Biosystems Inc., Perkin Elmer). Upon cloning and sequencing of the amplification product, the first nucleotide downstream of the 5' adaptor RNA was assigned to the TSS. In total, 244 sequences were generated for 80 different genes of *Synechocystis* 6803. From these, 214 sequences were assigned to a total of 89 different TSS (**Table S7**).

Gene expression microarray

The microarray design, hybridization procedure and data analysis have been described previously (8). The microarray data are available in the GEO database with the accession numbers GSE16162 and GSE14410. Features are stated as

significantly expressed if at least one probe at one condition passed the threshold of $2^{11.12}$ after subtraction of their standard deviation. The threshold was defined by the mean of non-*Synechocystis* control probes (after adding the standard deviation of the control probes).

Oligonucleotides

For each gene to be tested by manual 5'-TAP-RACE, one oligonucleotide was synthesized for reverse transcription (rt-oligo) and a second, nested oligonucleotide was synthesized for the subsequent RACE PCR (p-oligo). Following the identification of a TSS, a third oligonucleotide (t-oligo) was employed for each gene overlapping the TSS, with its 5' end located at least 10 nt upstream the TSS. These served in a second series of RACE reactions as controls to either verify the detected TSS or to detect another, more upstream promoter. A complete list of all oligonucleotides is presented in **Table S10**.

Computational methods

The genome sequence of *Synechocystis* 6803, accession number NC_000911, was downloaded on 25-Jan-2009 from the NCBI ftp-server. The more reads from the cDNA (+) library map to a distinct nucleotide position, the more likely this position is as a functioning TSS. After filtering out all reads mapping to the ribosomal RNA operons, a total of 95,413 sequencing reads for the (+) population ≥18 nt were mapped to the *Synechocystis* 6803 chromosome. All 5' ends located within a window of three consecutive nucleotides were considered a single TSS. As an additional criterion, we derived a position-specific weight matrix (PSWM) for the -10 element as the most common promoter element expected in a distance of 5 to 7 nt positions upstream of these putative TSSs. The relationship between the number

of (+) reads mapped to a certain position, the minimum score for the -10 element, and the resulting number of TSSs for the chromosome of *Synechocystis* 6803 is given in **Table S11**. For example, 1,294 TSSs meet the threshold of at least five reads initiated at one site with a minimum -10 element score of +2.00, and 3,213 with a minimum of 2 reads at a score of +2.00 (**Table S11**). Based on a computation of the possible gain of true positives against the chance to acquire more false positives (**Fig. S4**) and the comparison to published data (**Table S1**) and our own experimental verification **(Table S7)**, we set the TSS minimum threshold at a PSWM score of + 2.00 for the -10 element if followed by at least 2 sequence reads. All 3,213 chromosomally located TSSs possess a suitable -10 element as the most common promoter element (**Fig. S5**). We found only 54 TSSs (21 for tRNAs) with more than 20 sequence reads but lacking a -10 element. The absence of a -10 element might indicate that these promoters are recognized by an RNA polymerase using one of the four type 3 alternative sigma factors (SigF, SigG, SigH and SigI) present in *Synechocystis* 6803. Therefore, these elements were analyzed separately. 27 of those possess a -10 element at a distance of 8 nt to the TSS with a PSWM score from 2.48 to 7.82. Seven gTSS give rise to transcripts that would otherwise not be taken into account due to the lack of a -10 element. From these TSSs, mRNAs originate for the following genes: slr1842, sll1184 (*ho1*), slr1838 (*ccmK3*), sll0927 (*metX*), sll3044, slr1918 and ssr2831 (*psaE*). Some of these genes become induced under conditions of low oxygen (9).

We prioritized gTSS over aTSS and iTSS, and all remaining TSS were automatically nTSS. After annotation according to **Fig. 1B,** several additional manual corrections were carried out (**Supplementary data file 2**).

The input data used for the *in silico* studies were 89 TSSs of 80 genes found by a systematic analysis of promoters in *Synechocystis* 6803 (**Table S7**). When

aligning the 89 experimentally determined promoter sequences at their mapped TSS, overrepresented nucleotides only appeared in the -10 region and at the TSS. 85% (185 of 214) of assigned TSS were an adenine or guanine nucleotides, indicating that there is a strong preference for a purine at the first transcribed nucleotide. The other preferences are a thymine at positions -7 and -12 and an adenine at positions -11 and -8. Thus, the presence of these conserved bases indicated typical eubacterial -10 elements.

The -10 region of the final alignment is shown in **Fig. S5** in the form of a sequence logo of all 89 sequences. Adenine (position 2) and thymine (positions 1 and 6) are highly conserved. These base preferences were also apparent from counts of each base at each position shown in **Fig. S5** (upper part), which were calculated to create a scoring matrix for the whole genome analysis.

A set of sites can be used to create a scoring matrix with the nucleotides A, T, C and G as columns and the individual positions as rows. Each entry of the matrix is determined from the logarithm of the ratio of actually observed frequencies and the number of expected nucleotides. The conservation of a position is given by the sum of the scores of one column of the scoring matrix. Thus, if the sum is close to 2 bits, it is completely conserved, and zero stands for no conservation (compare **Fig. S5**). One pseudocount was added to each entry of the table in **Fig. S5** to avoid zeroes which otherwise would occur for small sets of known sites and in order to develop the PSWM. The entries of this PSWM were obtained by taking the logarithm (base 2) of the ratio of observed to expected frequencies, respecting the GC-content of 0.47% within the intergenic regions of the *Synechocystis* 6803 genome (0.235 for C or G and 0.265 for A or T). As an example, if we take the first position, the frequent nucleotide T gets a score of

$$\log 2 \frac{64+1}{0.265*(89+4)} = 1.3991$$

whereas the nucleotide G gets a score of

$$\log 2 \frac{6+1}{0.235*(89+4)} = -1.6425$$

All used software tools are freely available. The multiple alignments were performed with CLUSTALX (10). The sequence logos were created with WebLogo (http://genome.tugraz.at/Logo/) developed by Schneider and Stephens (11). All scripts were written in Python 2.5.2 and Biopython V 1.42 (http://biopython.org/) and are available on request.

In addition to the -10 element, around position -35 of sigma70-specific promoters, a second conserved hexamer (consensus in enteric bacteria: 5'-TTGACA-3') might be expected. A general consensus sequence for the -35 element was not obtained, even if all or only highly expressed RNAs (20 or more reads) were considered. However, in *E. coli*, the first three nucleotides of this hexamer are more conserved than the other (5'-TTG*** -3') (12). Therefore, we selected all hexamers separated from the respective -10 element by a distance of 15-21 nt and containing these three highly conserved nucleotides at the correct position, yielding **Table S3**.

2. Supplementary data files: http://www.cyanolab.de/Supplementary.html

Supplementary data file 1 “Syn6803TSS.gbk“. Genbank file of the *Synechocystis* 6803 chromosome with all annotated TSS.

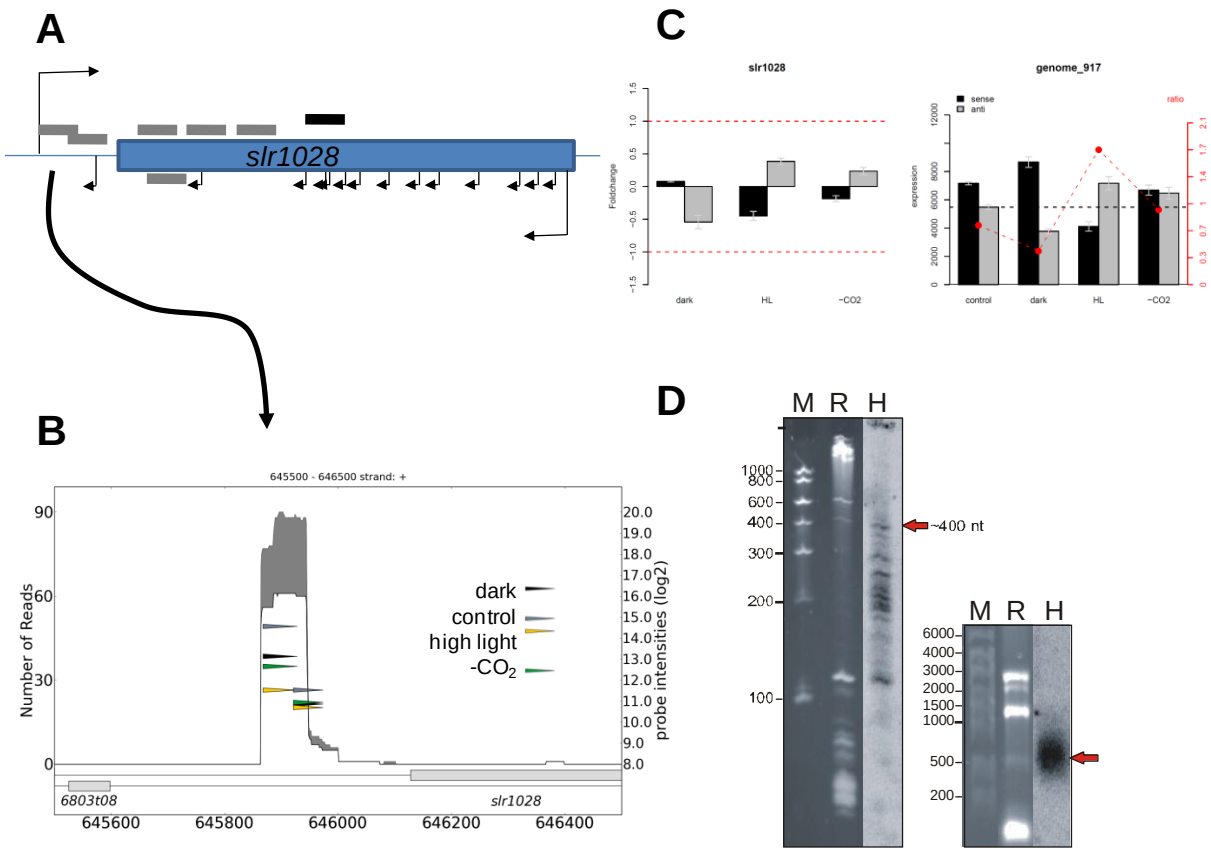
Supplementary data file 2 “Flow chart” illustrates all filter steps and manual changes done for chromosomally located TSSs during the analysis and annotation. All initially found 784 iTSS were screened for possible misannotations (mainly 5’ ends of CDS), lowering the number of iTSS to 752. One iTSS was considered to be an nTSS, another as the gTSS for the downstream located gene slr1470. Another iTSS was corrected to be the gTSS of *phoH* since manual verification showed a long contiguous 5’UTR. One further iTSS was corrected to be the aTSS of as_slr320. The number of 411 initially found nTSS was lowered to 370 nTSS: e.g., 18 were corrected as gTSS upon comparison with the Cyanobase protein database. Another 10 were corrected to be iTSS, 1 indicated a 5' misannotation of a CDS upon a Blastp search, identifying itself as a gTSS of ssl0241, a hypothetical protein. Five nTSS turned out to be aTSS. According to the manual 5’RACE mapping, 5 more nTSS were corrected as gTSS of genes with very long 5'UTRs, one to be the aTSS of asRNA_sll1864. Initially, 1030 aTSS were found. By comparison against Cyanobase, 5 aTSS were corrected as iTSS (1), gTSS (2), nTSS (2). 15 aTSS were corrected to be gTSS according to long secondary reads. Two aTSS were re-annotated as iTSS but then re-re-annotated as gTSS based on conservation of reading frames and corrected 5' misannotations of open reading frames. One aTSS is a gTSS at the +1 position of slr0280, coding for an hypothetical protein. Moreover, 2 aTSS were corrected to gTSS and another one to nTSS (SyR2) based on our manual verification. Thus, 1013 aTSS were left in the dataset. The number of initially 988 gTSS was lowered to 984

by reclassifying three as nTSS and one as an aTSS upon comparison against Cyanobase. Adding the numbers from iTSS, nTSS and aTSS and ten gTSS for 8 novel protein-coding genes yielded the final number of 1098 gTSS. The final numbers for iTSS were 732, and 370 for nTSS, and 1013 for aTSS.

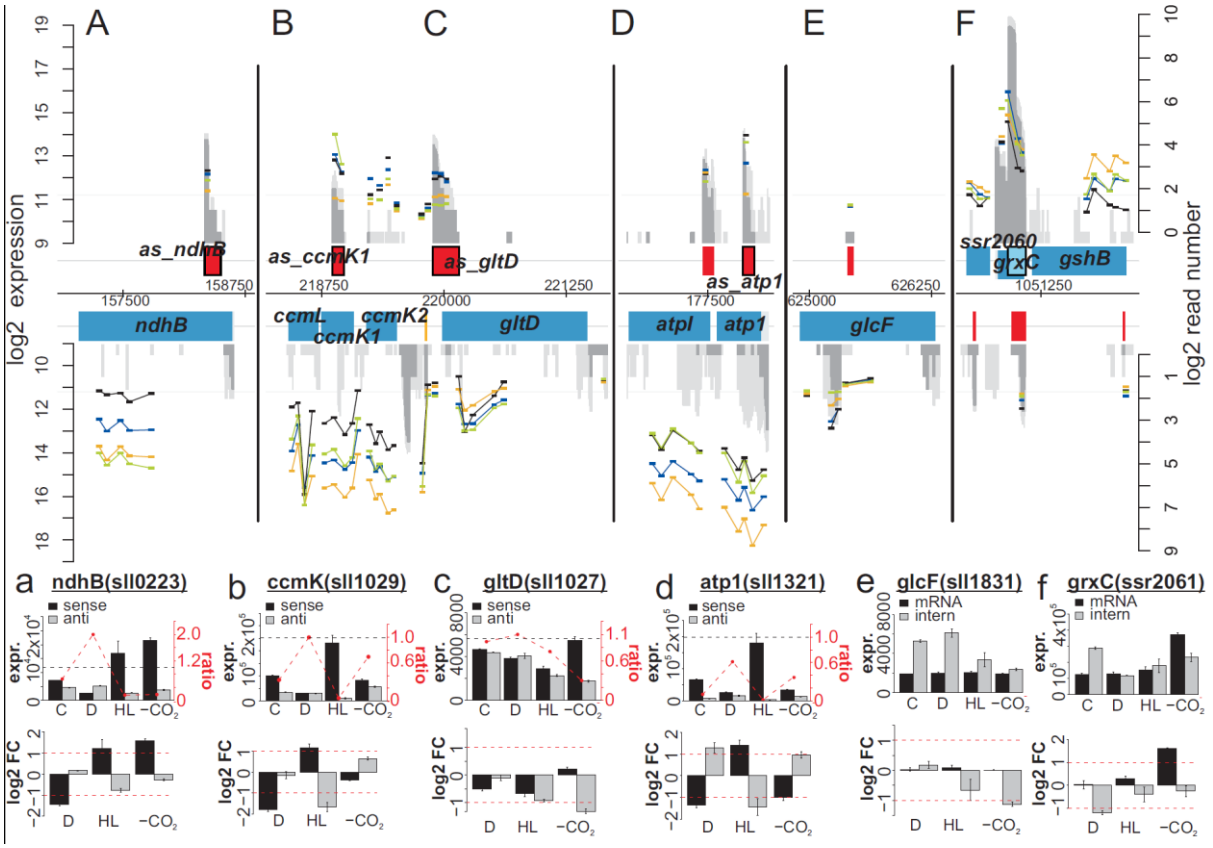
Supplementary data file 3 "genomeplot_conditions_w_lines.pdf". Genome-wide overview combining the number of 454 reads (accumulated read numbers, scale on the right) from the (+) and (-) cDNA populations with the log_2-normalized expression values (left scale) from the microarray analysis of cultures kept under the four different conditions as indicated by the coloured lines.

Supplementary data file 4 “FCratioplots.pdf”. Overview over all asRNA:mRNA pairs measured in the transcriptome microarray. For each pair, the respective fold changes of mRNA and asRNA against the control conditions (log_2) are shown on the left, together with the average of normalized probe set signal intensities from three biological replicates in two technical replicates (right). The ratios of asRNA/mRNA signal intensities are indicated by filled circles in red.

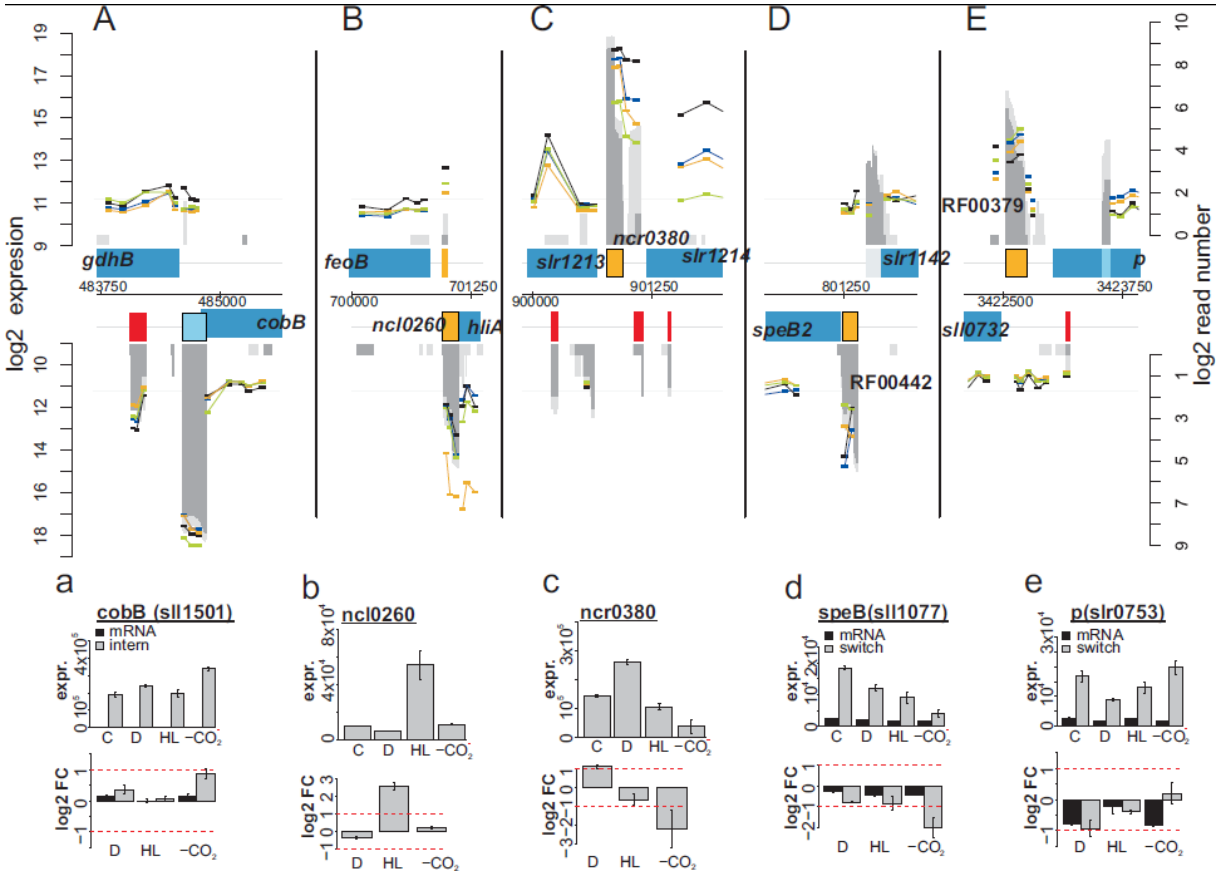
3. Supplementary figures



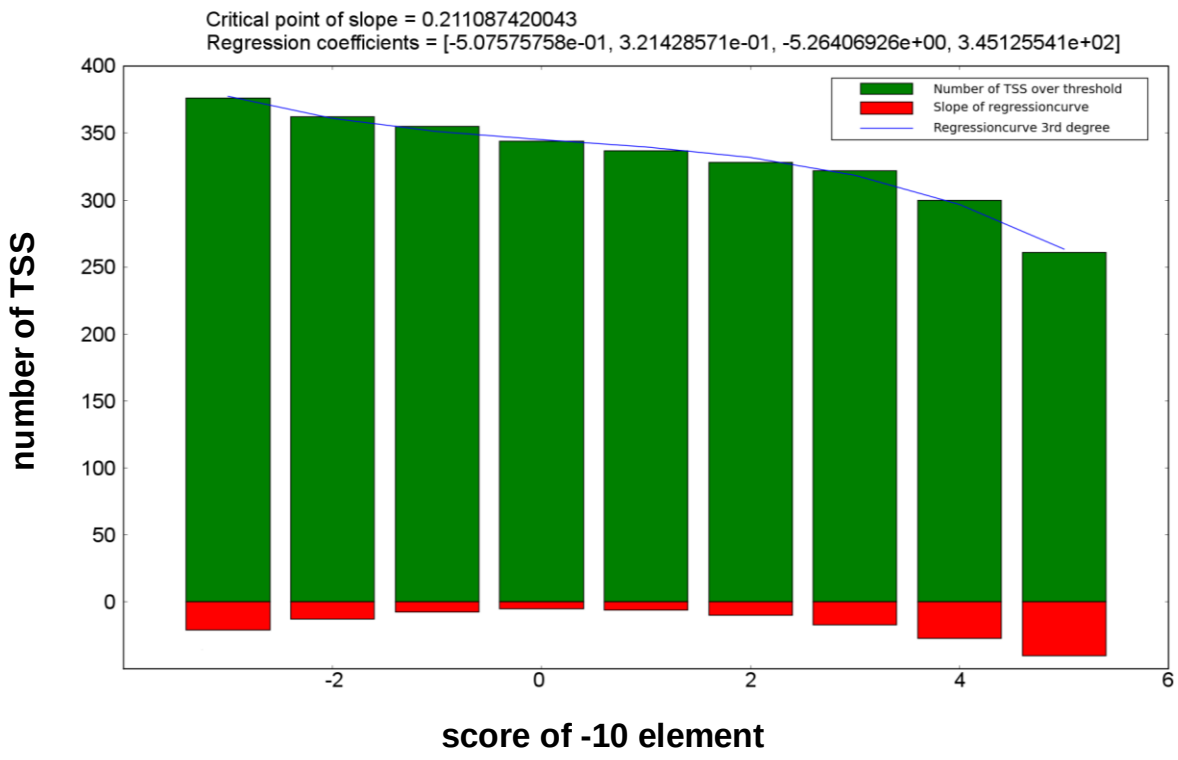
Supplementary Figure S1. Possible riboregulation at gene slr1028 encoding a giant protein. A. TSS within and around slr1028 (black arrows). The protein Slr1028 possesses five integrin alpha domains, one laminin domain, one bulb-type mannose-specific lectin domain and is supposed to be transported to the cell surface, but its precise function is unknown. Several related proteins in *Synechocystis* 6803 include the giant proteins Slr0408, Slr5005, and Sll0723. The location of microarray probes is indicated by grey boxes and of a single stranded RNA probe used in Northern hybridization in part D by a black box. **B.** A particularly strong TSS immediately 5' of slr1028 (number of 454 reads given on x-axis) is downregulated under high light and CO_2 depletion (location of probes indicated by triangles, probe intensities plotted to the right). **C.** Probe intensities of mRNA and asRNAs measured in the microarray and resulting ratio of probe intensities (average of normalized probe set signal intensities from three biological replicates in two technical replicates each). **D.** Northern verification of asRNA accumulation in a high resolution gel (left) and a standard agarose gel (right).



Supplementary Figure S2. Examples of the combined results of the dRNA-seq and microarray experiments in partial-genome plots. (A–F) The normalized $\log_2$ expression values (scale on left) of four different microarray experiments are plotted for each probe as short horizontal bars that span the corresponding hybridization region. All probes of a single RNA feature are connected by lines. The genome plots (A–F) are accompanied by bar plots (a–f), which show the mean expression values and the mean fold changes (FC) of the selected features. C, control; D, dark incubation, HL; high-light stress, $-CO_2$, CO_2 depletion. For asRNAs (A–D and a–d) the data for the corresponding mRNAs and the expression ratios (asRNA/mRNA) are included. For internal RNAs (E–F and e–f) the data for the corresponding genes are added also.



Supplementary Figure S3. Examples for the combined results of the dRNA-seq and microarray experiments in partial genome plots. Examples for (A,B, a,b) intergenic ncRNAs, (D,d) RF00442 riboswitch, (E,e) RF00379 riboswitch.

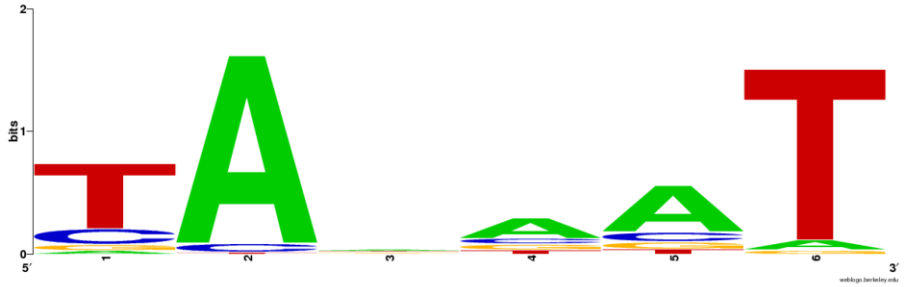


Suppl. Figure S4. Relationship between the -10 element scores (x-axis) and the respective number of TSS (y-axis), based on all 5' ends with ≥20 reads. The slope of the resulting graph is shown in red underneath the x-axis. The turning point of this graph is at 0.2111.

TACAAT TACGGT TATCTT
GAGAAT TAAACT TCATAT
AATAAT TAACGT TATAAT
CAGAAT CACAAT CAAAAT
TACAAT TATGAT TAAACT
TATACT TAGAAT GAACCT
TAGACT CAAAGT TATCAT
TATATT CAGAGT TACAAT
TACCAT TAGCAT TATAAT
TAAATT TAGGAT TAAGAT
GCTAAT TAATAA GATATT
TATAAT TAAAAT TAAACT
TACCCT TATGAT TCGTCA
CAGATA TAACAT TAAAAT
TAAGAT TATTAT TACAAT
TAAGAT TATACT
TAGAAT TAAACT
TAGTGT TAAAAT
CATCAT TAAAAT
TACAAT TAGCAT
TACCCT CAGACT
TAGGAT TATAAT
TATAAT CACACT
AATGAA CAATAG
TAAGGT TAATAT
TAAAAT TAACTA
GAGAAT CAAAAT
TATGGT TAAAAT
TAGTAT CAAGAT
TAGGAT CACAAT
TACGGT GTCTTG
TACAAT ACCAAT
CAAAAT TAAAAT
TAATGT TAGCAT
AAAAAT TAATAT
TATAAT CATAAT
TAAAAT TAACAT

	-12	-11	-10	-9	-8	-7
A	4	84	34	50	60	5
C	15	4	16	14	13	0
G	6	0	18	14	9	2
T	64	1	21	11	7	82

	-12	-11	-10	-9	-8	-7
A	-2.30	1.79	0.51	1.05	1.31	-2.04
C	0.45	-2.13	-0.36	-0.54	-0.64	-4.45
G	-1.64	-4.45	-0.20	-0.54	-1.13	-2.86
T	1.40	-3.62	-0.16	-1.04	-1.62	1.75



Suppl. Figure S5. The -10 element of *Synechocystis* 6803. Alignment of sequences for the -10 elements (positions –12 to –7 with regard to the respective TSS) of 89 (+1 pseudocount) experimentally determined TSS (see **Table S2**) yields the numbers of nucleotides for each position (top), the PSSM, half-adjusting (middle) and the resulting sequence logo (bottom).

4. Supplementary Tables

Suppl. Table S1. Overview on TSS mapped for plasmids pSYSA, pSYSG, pSYSM and pSYSX.

	pSYSA NC_005230	**pSYSG** NC_005231	**pSYSM** NC_005229	**pSYSX** NC_005232	**Total**
length (nt)	103,307	44,343	119,895	106,004	-
# of genes	106	49	132	110	-
(+) reads	3039	1356	6356	6818	-
(-) reads	11957	930	3699	2777	-
gTSS	19	8	26	14	67
aTSS	30	12	32	25	99
iTSS	27	7	39	16	89
nTSS	20	9	20	10	59
Total:	96	36	117	65	*314*

Suppl. Table S2. Comparison of 454-anchored TSS with previously determined 5' ends. Gene names and references are given for previously determined TSS, followed by the systematic ID and the previously identified TSS position relative to the annotated start codon of an mRNA. Absolute coordinates are indicated for ncRNAs. The position of 454-mapped TSS is given in column 4, followed by the number of associated reads (in brackets). Two or more numbers are given if more than one closely spaced nucleotide was mapped belonging to the same -10 element. In the comment column, "not expressed" indicates that no reads were found by dRNAseq, "not found' indicates the presence of reads in the dRNAseq approach but that previously identified TSS were not confirmed.

If more than one TSS was found, the most distally located was numbered TSS1.

*TSS proposed by authors.

Gene (Reference)	ID	TSS position	Position of dRNAseq-mapped TSS (number of reads)	comment
photosynthesis genes				
psbA2 (13)	slr1311	-49	-49 (13)	TSS confirmed
psbA3 (13)	sll1867	-88	-88 (82), -89 (1), -90 (13)	TSS confirmed
psaAB (14)	slr1834/slr1835	-144	-144 (5),-147 (37)	TSS confirmed
psaC (14)	ssl0563	-63	-63 (105),-54 (2)	TSS confirmed
psaD (14)	slr0737 TSS2	-35,-34,-33*,-32	-32 (23), -31 (1)	TSS P2 (14) confirmed
psaD (14)	slr0737 TSS1	only postulated	-72 (3)	additional TSS
psaE (14)	ssr2831	-90	26	TSS confirmed
psaFJ (14)	ssl0819/sml0008 TSS1	-161*,-162	-161 (9)	TSS P1 (14) confirmed
psaFJ (14)	ssl0819/sml0008 TSS2	-140*,-141	-140 (4)	TSS P2 (14) confirmed
psaK1 (14)	ssr0390	-62,-63	-	not found
psaLI (14)	slr1655/smr0004	-52,-53,-54*,-55	-52 (17)	TSS confirmed
petH (15)	slr1643	-523	-53 (4)	TSS at -53 not found by authors (15), TSS at -523 not found here (possible read-through from *prk* aTSS at -1559?)
isiA (16)	sll0247	-211	N	not expressed
primary metabolism				
adhA (17)	slr1192	-69	-73, -71 (1,10)	TSS confirmed
gap2 (18)	sll1342	-50	-50 (59), -49 (1), -48 (1), -47 (1)	TSS confirmed; gram-positive-like -16 promoter element (18)
prk (15)	sll1525	-219	-219 (182)	TSS confirmed
house-keeping functions				

pilA1 (19)	sll1694	-54	-54 (85)	TSS confirmed; SigF-dependent (19)
sbtA (20)	slr1512	-168	-	not expressed
regulatory genes				
abrB-like (21)	sll0359	-64	-64 (11)	TSS confirmed
pixJ1, pisJ1 , taxD1 (19)	sll0041	-32	-	not found; SigF-dependent (19)
sigE (22)	sll1689	-202	-202 (11)	TSS confirmed (22)
sigF TSS3	slr1564	not detected by authors	-22 (2), -20 (3)	additional TSS
sigF TSS2 (23)	slr1564	-43	-	not found
sigF TSS1 (23)	slr1564	-189	-	not found; SigF-dependent (19)
cpH1/rcp1 (24)	slr0473/slr0474	-149	-149 (10), -148(1)	TSS confirmed
phy, hik35 (25)	slr0473	-149	-148 (1), -149 (10)	TSS confirmed
nitrogen metabolism				
glnA (26)	slr1756	-48,-47	-48(3),-47(11),-46(1)	TSS confirmed
glnN (26)	slr0288	-31	-31(4)	TSS confirmed
gifA (24)	ssl1911	-51	-51 (13)	TSS confirmed
gifB (24)	sll1515	-104	not detected	not expressed
glnB (27)	ssl0707 TSS1	-53, -54	-53 (4)	TSS confirmed
glnB (27)	ssl0707 TSS2	-47	-	not found
glnB (27)	ssl0707 TSS3	-33	-33 (33), -32 (3), -30 (1)	TSS confirmed
nirA (28)	slr0898	-23	-25 (3), -24 (4), -23 (42)	TSS confirmed
nrsBACD (29)	slr0793	-46	-	not expressed
nrsBACD (29)	slr0793	-30	-	not expressed
nrsRS (29)	sll0797	-11	-	not expressed
nrtA (28)	sll1450	-47	-47 (8)	TSS confirmed
ntcA (28)	sll1423	-408,-409	-408 (2)	TSS confirmed plus TSS at -51 with 7 reads
amt1 (30)	sll0108 TSS2	-142	-138 (1), -139 (16), -140 (2)	TSS confirmed
amt1 (30)	sll0108 TSS1	not detected by authors	-148 (2)	additional TSS; -35 element overlaps with NtcA binding site
cpn60 (31)	sll0416	-71	-	not found (not expressed)
stress adaptation				
ggpS (32)	sll1566	-378	-377 (1), -378 (11)	TSS confirmed
deaD, crhR (33)	slr0083	-110	not detected	not found
unkwn (33)	slr0082	-143	-143 (103), -142 (1), -141 (7)	TSS confirmed
groEl-2 (34)	sll0416	-45	-	not expressed
groES (31)	slr2075	-74	-74 (2), -75 (2)	TSS confirmed
hoxE (35)	sll1220	-168	-168 (4)	TSS confirmed
dnaK (36)	sll0170	-1277	-	not found
hsp17 (23, 37)	sll1514	-44	-44 (6)	TSS confirmed
trxA (38)	slr0623, TSS3	-32	-32 (8)	TSS confirmed
trxA (38)	slr0623, TSS2	-38, -36	-38 to -36 (184)	TSS confirmed
trxA	slr0623, TSS1	not detected by authors	-49 (7)	additional TSS
lrtA (39)	sll0947	-312	-312 (6)	TSS confirmed
other protein-coding				
sufBCDS TSS1 (40)	slr0074–slr0077	-126	not detected	not found

sufBCDS TSS2 (40)	slr0074–slr0077	-264	-266 (9)	TSS confirmed; 5'UTR contains highly structured RNA element (41)
unkwn (19)	sll0837	-41	-109 (23)	TSS at -41 not found; SigF-dependent (19)
unkwn (17)	sll1106	-31	-31 (8), -32 (6)	TSS confirmed
non-coding RNAs				
rnpB (42)	RNase P RNA	153166	-3 (1), -1 (1), 153165 (78), +1 (3),	TSS confirmed
ssrA (43)	tmRNA	c3317166	c3317166 (65)	TSS confirmed (43)
isrR (2)	asRNA IsrR	1518034	1518034 (48)	TSS confirmed
Yfr2a (41)	ncRNA	c1558975	-2 (52), -1 (1), c1558975 (323), +1 (1)	TSS confirmed
Yfr2b (41)	ncRNA	2730523	-1 (1), 2730523 (1798), +1 (13), +2 (322)	TSS confirmed
Yfr2c (41)	ncRNA	c3398352	-2 (7), -1 (11), 3398352 (628)	TSS confirmed
SyR1 (41)	ncRNA	1671919	1671919 (58), +2 (1)	TSS confirmed
SyR2 TSS2 (41)	ncRNA	1431853	1431853 (1)	not found
SyR2 TSS1	ncRNA	not detected	-1 (1), 1431912 (9)	additional TSS

Suppl. Table S3. List of the 634 TSS that had a -35 element [5'-TTG*-3']**

See "Suppl. Table 3" within the *Dataset S1*.

Suppl. Table S4. New hypothetical proteins (Norfs) or transcribed pseudogenes. The respective gene name (ID), start and end coordinates, TSS, sequence and score of the -10 element is given, together with strand orientation (s) forward (f) or reverse (r), and the number of reads in the (+) cDNA library.

ID	Start	End	s	TSS	-10	Score	Reads (+)	Comments
Norf1	298829	298972	r	298964 (TSS2)	TATGAT	5.54	10	conserved in cyanobacteria
Norf1	298829	298972	r	299115 (TSS1)	GATAAT	4.09	53	conserved in cyanobacteria
Norf2	365077	365280	f	365050	TATTAT	5.04	6	conserved in non-cyanobacteria
Norf3*#	608978	609223	f	608957	TAGAAT	7.09	15	pseudogene ISY120 transposase partial copy; similarity to sll0986, sll1157, slr1903, slr2096
Norf4	1148233	1148325	r	1148394	TACCAT	5.34	3	conserved in cyanobacteria
Norf5	1418691	1418972	r	1418899 (TSS2)	TAGGAT	5.50	4	possible internal start in Norf5 gene
Norf5	1418691	1418972	r	1418990 (TSS1)	TAAAAT	7.80	2	similar to ssr2549, sll1954, ssr0759
Norf6	1842719	1842856	r	1842906	TAAACT	5.85	11	conserved in N_2-fixing cyanobacteria
Norf7	2407236	2407409	f	2407215	TATACT	5.18	4	similar to ssl7074, conserved in cyanobacteria
Norf8#	3067193	3067311	r	3067311	GAGAAT	4.05	15	pseudogene with similarity to IS3/IS911 transposase family domain

*internal stop codon; #partial gene

Suppl. Table S5. Protein-coding genes with a high number of pyrosequencing reads (≥20) in the (+) cDNA population.

See "Suppl. Table 5" within the *Dataset S1*

Suppl. Table S6. List of 58 protein-coding genes with 5'-corrected start codon.

See "Suppl. Table 6" within the *Dataset S1*

Suppl. Table S7. List of primary 5' ends mapped by manual 5'TAP-RACE to infer PSWM. The first transcribed nucleotide is in capital letters and in boldface. The underlined hexanucleotides indicate putative –10 regions. The number of sequenced clones is given for each gene together with the frequency of the respective TSS. In total, 244 sequences were generated, from these 214 were assigned to a TSS. Preference of the first transcribed nt is: A>>G>>C>T (128>>57>17>12). Multiple TSS for one gene are numbered consecutively TSS1, TSS2, etc. *Genes with incorrect annotation of start codons: *ziaA* and *pacL2* (sll0672) begin only at codon 11 (30 nt shorter).

#	Gene & Reference	ID	sequence	TSS position	TAP-RACE clone numbers supporting TSS / total number
1	*sigA*	slr0653	agaccacaattaagccccatttttttgttatcgaggc**tacaat**caatga**A**	-222	13/14
2	*psbD* TSS1	sll0849	cccggcggtgtagtttccaatcgtctcgtcttatta**gagaat**ggagtc**TA**	-81, -80	1/6; 5/6
3	*psbD* TSS2	sll0849	tgagagtgaatatgtcagaatgtaaaatatttgct**aataat**atgta**GA**t**G**	-270, -269, -267	1/6; 4/6; 1/6
4	*psbD* TSS3	sll0849	acccctggcgatcggcgattatgagagtgaatatgt**cagaat**gtaaa**At****A**	-290, -288	2/4; 2/4
5	*psbD2*	slr0927	atccctttcgcactggcagattcccagaatttcctt**tacaat**ggataac**A**	-123	7/7
6	*ureA*	slr1256	tgcctccaaatcattaatattgctcggactaatcgc**tatact**tacggtc**A**	-58	7/7
7	*psbA2*(13)	slr1311	cctgttacaaagctttacaaaactctcattaatcctt**tagact**aagttt**A**	-49	8/9
8	*psbA3*(13)	sll1867	ttttcagcaagctatttacaaattgttacaatcttgt**tatatt**actcat**A**	-88	7/7
9	*petE*	sll0199	gctgtataatctacgacgggctgtcaaacattgtga**taccat**gggcag**AA**	-101, -100	2/7; 5/7
10	*groES*(31)	slr2075	ctcaccggggtgttgcactgggtcaagcaatttagc**taaatt**agcact**CG**	-66, -65	1/6; 5/6
11	*ho1*	sll1184	aaaaaacttaatggtacgactcaacgaatccgtttag**gctaat**aaa**C**t**AG**	-69, -67, -66	3/15; 9/15; 1/15
12	*ndhD2*	slr1291	tctcaaaaaaaattctcaaaagtttacaaagtttaa**tataat**cggagg**AA**	-29, -28	4/9; 2/9
13	*apcE*	slr0335	aagttaacaaagcccctgtagcgttccgaaacgttc**taccct**agatgct**G**	-104	6/7
14	*murF*	slr1351	gtgatgcaagcacccatgggaggtgatctagatca**cagata**aaaattg**CA**	-75, -74	3/8; 3/8
15	*rbcL*	slr0009	agagcattgccataagtaaaggcatcccctgcgtga**taagat**taccttt**CA**	-156, -155	1/12; 6/12
16	*ziaA** TSS1	sll1076	ctttttaggcgacaacagttaagtttgaatctttg**taagat**tggga**AT**c**A**	-39, -38, -37	5/9; 2/9; 2/9
17	*ziaA** TSS2	sll1076	ggcgatcgccattctggtgttatccaccccggccaa**tagaat**gggatc**AT**	-170, -169	2/4,2/4
18	*ftsZ*	sll1633	gctcagattcacattgacaaatttgctcaatatgcc**tagtgt**agggac**GA**	-154, -152	3/5,2/5
19	*sigB*	sll0306	ttcatcttggcccttgtggaatcccttaatgattcgt**catcat**ggtgat**A**	-228	4/8
20	*sigC* TSS1	sll0184	aaaagcctgccatcggccatagatttttttgaaaggttt**tacaat**ggagag**A**	-142	4/4
21	*sigC* TSS2	sll0184	ggcagcgccgacaaaaaaaatagacaattaaagaatgt**taccct**tgactt**A**	-31	6/8
22	*hyp. protein*	sll0815	gccaagtttgagtagatcattgcaaaggggcgttg**taggat**ggaat**G**tc**A**	-273, -270	1/6; 5/6
23	*isiA*(16)	sll0247	Ataattttagttgctataaattctcatttatgcccc**tataat**aat**T**c**GGG**	-216, -214, -213, -212	1/10; 1/10; 2/10; 6/10
24	*pacL1**	sll0672	gtcagtaatttaagcttatgatcgttttcgatt**aaaatgaa**agcttg**T**t**G**	-19, -17	4/7; 2/7
25	*cpcF*	sll1051	aggtatttgtgcaggggttggggtcaaaccaatgt**taaggt**ggttcac**TG**	-31, -30	2/6; 3/6
26	*pacL2*	slr0822	acggtcaattgtagtaggcccccaaacccggccttt**taaaat**aggaaat**AG**	-45, -44	3/8; 3/8
27	*apcA*	slr2067	tgttacgggggcagtgtaatcaggaacgcaatgcct**gagaat**ggttt**GGG**	-124, -123, -122	3/6; 2/6; 1/6
28	*trxA*(38) TSS1	slr0623	cacgatatttttccatacaggggtcaacaattggt**tatggt**agtatt**C**t**A**	-38, -36	6/8; 2/8
29	*trxA*(38) TSS2	slr0623	atatttttccatacaggggtcaacaattggttatgg**tagtatt**c**t**a**a**tc**A**	-32	3/3

30	*chlP*	sll1091	agcggcgatcgcctacccctaagcaaaaagaccgtga**taggat**gcgatc**C**	-190	3/6
31	*hyp. protein*	slr0798	atctgagcatatcttcaggtgtttcaagatttgtgc**tacggt**tcaagg**AG**	-29, -28	4/7; 1/7
32	*yfr1*	-	caccggcaaaaaccctatgccccccgtcccaacctg**tacaat**gaaga**GGG**	ncRNA	1/5; 3/5; 1/5
33	*isrR(2)*	sll0247	tgtccccccaccagatcttctaggctactgatgctgg**caaaat**gggttt**G**	asRNA	10/10
34	*as_uvrA*	slr1844	ccagggtctggcggatgggcacactggttaaattgt**taatgt**tgtactG	asRNA	3/4
35	*as_rpoB*	sll1787	ctaggcgatcgccatcctcctgaaaaactaacggg**aaaaat**ggttggGA	asRNA	2/5; 2/5
36	*as_accA*	sll0728	gtgcatttcaaaccaatcatccgcaatggcttgga**tataat**ccaacgtA	asRNA	4/4
37	*as_kaiBC*	slr0757 / slr0758	gctaaatttagatttttaaagaagaaaaaagttag**taaaat**taaaactA	asRNA	8/10
38	*as_KaiA*	slr0756	tcaaataagggctcagataggctttctcaccatgg**tacggt**gcaagttA	asRNA	6/8
39	*as_hik31*	sll0790	taaatcttcaactaaatcattgagacaacaaagggg**taaact**gggtcGGA	asRNA	2/5; 1/5; 1/5
40	*kaiA*	sll0726	cggaagctatccggccaaggagcactcagattgtgt**taacgt**tcagggG	-122	6/7
41	*as_atpA*	sll1326	attgcagagtggcggggtcgttggcgttggcggccac**cacaat**ggtatA	asRNA	3/3
42	*ndhF*	sll1732	agcaaatggttgttgatctggaattttatttcccc**tatgat**atggtgtA	-241	3/3
43	*slr2006*	slr2006	tggcgatcgccattgctattggtaccaaaacaagga**tagaat**ctatgcA	-112	3/3
44	*slr1804*	slr1804	ccctatgggggggaagtaaaatccattccagtttagg**caaagt**tgagatA	-11	3/3
45	*slr1592*	slr1592	ggctggattttctgcgtaacgctgagtgaactcagc**cagagt**aaaataA	-79	2/3
46	*slr1535*	slr1535	agggaattgtaaccaatgccgaaaccgccccaatgc**tagcat**taacctG	-79	3/3
47	*ncr0380*	slr1214	tcaaaaatcgaccaggacagggttgggaatttttcc**taggat**gggactG	ncRNA	3/3
48	*ycf60 TSS1*	sll1737	tttcggcggtgtaggtgttaaccatggcaaagcgg**taataa**taaaagaA	-226	3/4
49	*ycf60 TSS2*	sll1737	ggggaagaccgaatctgggccatgacgaacccaggt**taaaat**cagggcA	-41	2/3
50	*54f_ncRNA*	-	tactaatctttatattcgcttttggtgggcgaacc**tatgat**ggtcttcA	ncRNA	5/5
51	*yfr2a*	-	ctacaggttcggctatcccgcccttaggcgatcggc**taacat**ttatttG	ncRNA	5/5
52	*yfr2b*	-	tgtcatagtgtccatggggtaggtgggtgaggaagt**tattat**ctagagG	ncRNA	5/5
53	*yfr2c*	-	ccaggtctgaagccttgacaatccccgttggtgatt**tatact**taggacA	ncRNA	2/3
54	*as_sll1586*	sll1586	catcttggagatattgccaggggttatagctaatctt**taaact**ttggaG	asRNA	3/3
55	*as_infB*	slr0744	tgtacttccccttggggtaattgttccaaggagcc**taaaat**agcccccA	asRNA	3/3
56	*as_sll0723*	sll0723	tgaaccagtgcttgtcgccacactgagatcgccag**taaaat**accggtcA	asRNA	3/3
57	*as_sppA*	sll1703	taatatttcttcctgattacccagtaaactctgttg**tagact**gagagtA	asRNA	3/3
58	*as_rfbA*	sll0207	ggttaagaaattaacccccggttgtctttataata**cagact**ctaaattA	asRNA	3/3
59	*as_ycf84*	slr0882	gtcacaaaagtttgtgttttttcatcgaaaaaatag**tataat**gccttcG	asRNA	2/3
60	*as_tktA*	sll1070	attcccattgacggaactgtttaatgtcttcgatggt**cacact**gtcaTA	asRNA	1/3; 2/3
61	*as_sarA*	asRNA	agcggtccataatttgaactgcttgcgagcttgata**caatag**ttgacT	asRNA	3/5
62	*as_sll0503*	sll0503	caaagcaaaaccctaaaagtatcaataactcttg**taatat**cccagtttA	asRNA	1/3
63	*as_sll0503*	sll0503	atcttagggatgattgttttccttctttttggctg**taacta**tggtcaA	asRNA	2/3
64	*as_hat*	slr0143	attgacaaaaacatcgattaatttcctgtcggctgg**caaaat**tctaggG	asRNA	2/3
65	*as_slr0264*	slr0625 / slr0624	accaattgccactggctatctgtgatgtccgtgctg**taaaat**tctcgaA	asRNA	3/3
66	*as_fmu-fmv*	slr0679	gcaatgaaaaaaccatcttggtggtggtgatgggg**caagat**agtcagcA	asRNA	3/3
67	*as_plpA*	sll1124	ggaggagaggccgtaacctggaggggaagattctgc**cacaat**tgacaaA	asRNA	3/3
68	*as_rlpA*	sll0375	cggcccttggtggggaactaaagaaattagttcggt**gtcttg**gaaagcA	asRNA	3/3
69	*as_slr0580*	slr0580	aaagtcagcgagggagccttgacaacttcccctaag**accaat**cacttcT	asRNA	3/3
70	*as_sll0217*	sll0217	aacatggcactccgcacatctccctttgagcggt**taaaat**atacagtC	asRNA	3/3
71	*as_lepA*	slr0604	ttttcccgataaccaattaattgatattccatgc**tagcat**agcccttG	asRNA	3/3
72	*as_ndhH*	slr0261	gggcaacgtcacctagcggtccaccgatcccatga**taatat**cgatactA	asRNA	3/3
73	*as_pknA*	slr1697	gtcaatcccccctgccccaacagagcctccatccgc**catatt**ggaggcA	asRNA	3/3
74	*as_slr0645*	slr0645	cattgttggcgtagatgtaagcaactttctcggcgg**taacat**tgtcatA	asRNA	2/3
75	*as_slr1963*	slr1963	tcacccctcaatggaaactttggtgcggctggcgg**tatctt**ggggagG	asRNA	1/1
76	*as_rpl1*	sll1744	ggccccaattgtttacccaaccgggcaattttgggca**tcatat**cagGgG	asRNA	1/3; 2/3
77	*as_slr1028*	slr1028	agctttgcagggcctgggtaataatctctaagttgc**tataat**attggcC	asRNA	3/3
78	*phoH*	slr2047	gcagaaagggttgaaaacatcaaccgagttttatgg**caaaat**tttgacA	-251	2/2
79	*as_slr1494*	slr1494	atatcagcaatgccattgtcactgcctggataatgg**taaact**aatcctT	*asRNA*	3/3
80	*as_mltA*	sll0016	gtgctccattgaacaaacttctgacacactagctgg**gaacct**agagtcC	*asRNA*	1/4
81	*as_sll1864*	sll1864	tgttaattttgcttggctagtcaaattattatccc**tatcat**tttgtaaA	*ncRNA*	4/7

82	*as_sasA*	sll0750	agcggtccataattttgaactgcttgcgagcttga**tacaat**agttgacT	*asRNA*	3/5
83	*as_dnaB*	slr0833	gactacctttcagctaattctattttattgcttg**tatatt**gaatagcA	*asRNA*	6/6
84	*as_glgC*	slr1176	cgtaatccatgcggtagagatggtcgccggacagaa**taagat**attcatC	*asRNA*	4/7
85	*as_cobN*	slr1211	ttccgttggaaaatttcagtttgagcgcaattaat**gatatt**tttagttA	*asRNA*	3/3
86	*as_cpcE*	slr1878	ctccgtaaatctgtaaatttccaattatttcagag**taaact**atccattA	*asRNA*	1/3
87	*as_moxR*	slr0835	gggctaacacatggggagcaacaaatttcacatcg**tcgtca**atggcatA	*asRNa*	6/7
88	*as_nhaS5*	slr0415	ctcaaaaatgggcggcaggatcagggtcaccaataa**taaaat**agtgaaG	*asRNA*	4/5
89	*ylxR*	ssr1238	tttggcgaattgtccgagtctatccatctcgaactg**tacaat**tagatcA	iTSS	3/3

Suppl. Table S8. TSS of top-scoring ncRNAs according to pyrosequencing (≥20 reads) and their verification by microarray.

See "Suppl. Table 8" within the *Dataset S1*

Suppl. Table S9. List of top-scoring asRNAs (≥10 reads) and asRNAs mentioned in the text. The entries are listed according to the number of aTSS-associated pyrosequencing reads.

See "Suppl. Table 9" within the *Dataset S1*

Suppl. Table 10. List of oligonucleotides. Primer were named after the respective target gene and a letter indicating their purpose (e.g., rt, reverse transcription; p, PCR amplification; t, control primers to search for additional transcription initiation sites).

See "Suppl. Table 10" within the *Dataset S1*

Suppl. Table S11. Relationship between the number of sequence reads, the threshold for the -10 element according to a PSWM and the calculated number of TSS for the chromosome of *Synechocystis* 6803 which meet these criteria.

	PSWM score at least					
number of reads	**-2**	**-1**	**0**	**1**	**2**	**3**
1	8378	6941	7397	6969	6399	5545
2-4	3862	3696	3573	3427	**3213**	2872
5-9	1464	1427	1387	1343	1294	1204
10-14	728	708	691	672	650	612
15-19	471	459	451	438	428	402
>20	354	343	336	327	321	299

5. Supplementary Material References

1. Zinchenko VV, Piven IV, Melnik VA, Shestakov SV (1999) Vectors for the complementation analysis of cyanobacterial mutants. *Russ J Genet* 35: 228-232.
2. Dühring U, Axmann IM, Hess WR, Wilde A (2006) An internal antisense RNA regulates expression of the photosynthesis gene *isiA*. *Proc Natl Acad Sci USA* 103: 7054-8.
3. Sharma CM, et al. (2010) The primary transcriptome of the major human pathogen, *Helicobacter pylori*. *Nature* 464: 250-255.
4. Sittka A, et al. (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* 4: e1000163.
5. Berezikov E, et al. (2006) Diversity of microRNAs in human and chimpanzee brain. *Nat Genet* 38: 1375-7.
6. Bensing BA, Meyer BJ, Dunny GM (1996) Sensitive detection of bacterial transcription initiation sites and differentiation from RNA processing sites in the pheromone-induced plasmid transfer system of *Enterococcus faecalis*. *Proc Natl Acad Sci USA* 93: 7794-7799.
7. Vogel J, Axmann IM, Herzel H, Hess WR (2003) Experimental and computational analysis of transcriptional start sites in the cyanobacterium *Prochlorococcus* MED4. *Nucleic Acids Res* 31: 2890-2899.
8. Georg J, et al. (2009) Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Mol Syst Biol* 5: 305.
9. Summerfield TC, Toepel J, Sherman LA (2008) Low-oxygen induction of normally cryptic *psbA* genes in cyanobacteria. *Biochemistry* 47: 12939-41.

10. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673-4680.
11. Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* 18: 6097-6100.
12. Hawley DK, McClure WR (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res* 11: 2237-2255.
13. Mohamed A, Eriksson J, Osiewacz HD, Jansson C (1993) Differential expression of the *psbA* genes in the cyanobacterium *Synechocystis* 6803. *Mol Gen Genet* 238: 161-8.
14. Muramatsu M, Hihara Y (2007) Coordinated high-light response of genes encoding subunits of photosystem I is achieved by AT-rich upstream sequences in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J Bacteriol* 189: 2750-8.
15. van Thor JJ, Hellingwerf KJ, Matthijs HC (1998) Characterization and transcriptional regulation of the *Synechocystis* PCC 6803 *petH* gene, encoding ferredoxin-NADP+ oxidoreductase: involvement of a novel type of divergent operator. *Plant Mol Biol* 36: 353-63.
16. Vinnemeier J, Kunert A, Hagemann M (1998) Transcriptional analysis of the *isiAB* operon in salt-stressed cells of the cyanobacterium *Synechocystis* sp. PCC 6803. *FEMS Microbiol Lett* 169: 323-30.
17. Vidal R, Lopez-Maury L, Guerrero MG, Florencio FJ (2009) Characterization of an alcohol dehydrogenase from the cyanobacterium *Synechocystis* sp. PCC 6803 that responds to environmental stress conditions via the Hik34-Rre1 two component system. *J Bacteriol* 191: 4383-91.

18. Figge RM, Cassier-Chauvat C, Chauvat F, Cerff R (2000) The carbon metabolism-controlled *Synechocystis gap2* gene harbours a conserved enhancer element and a Gram-positive-like -16 promoter box retained in some chloroplast genes. *Mol Microbiol* 36: 44-54.
19. Asayama M, Imamura S (2008) Stringent promoter recognition and autoregulation by the group 3 sigma-factor SigF in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Nucleic Acids Res* 36: 5297-305.
20. Lieman-Hurwitz J, et al. (2009) A cyanobacterial AbrB-like protein affects the apparent photosynthetic affinity for CO_2 by modulating low-CO_2-induced gene expression. *Environ Microbiol* 11: 927-36.
21. Oliveira P, Lindblad P (2008) An AbrB-Like protein regulates the expression of the bidirectional hydrogenase in *Synechocystis* sp. strain PCC 6803. *J Bacteriol* 190: 1011-9.
22. Muro-Pastor AM, Herrero A, Flores E (2001) Nitrogen-regulated group 2 sigma factor from *Synechocystis* sp. strain PCC 6803 involved in survival under nitrogen stress. *J Bacteriol* 183: 1090-5.
23. Imamura S, et al. (2003) Antagonistic dark/light-induced SigB/SigD, group 2 sigma factors, expression through redox potential and their roles in cyanobacteria. *FEBS Lett* 554: 357-62.
24. Garcia-Dominguez M, Reyes JC, Florencio FJ (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: 1192-201.
25. Hubschmann T, Borner T, Hartmann E, Lamparter T (2001) Characterization of the Cph1 holo-phytochrome from *Synechocystis* sp. PCC 6803. *Eur J Biochem* 268: 2055-63.

26. Reyes JC, Muro-Pastor MI, Florencio FJ (1997) Transcription of glutamine synthetase genes (*glnA* and *glnN*) from the cyanobacterium *Synechocystis* sp. strain PCC 6803 is differently regulated in response to nitrogen availability. *J Bacteriol* 179: 2678-89.
27. Garcia-Dominguez M, Florencio FJ (1997) Nitrogen availability and electron transport control the expression of *glnB* gene (encoding PII protein) in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Mol Biol* 35: 723-34.
28. Aichi M, Takatani N, Omata T (2001) Role of NtcB in activation of nitrate assimilation genes in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J Bacteriol* 183: 5840-7.
29. Lopez-Maury L, Garcia-Dominguez M, Florencio FJ, Reyes JC (2002) A two-component signal transduction system involved in nickel sensing in the cyanobacterium *Synechocystis* sp. PCC 6803. *Mol Microbiol* 43: 247-56.
30. Montesinos ML, Muro-Pastor AM, Herrero A, Flores E (1998) Ammonium/methylammonium permeases of a Cyanobacterium. Identification and analysis of three nitrogen-regulated *amt* genes in *Synechocystis* sp. PCC 6803. *J Biol Chem* 273: 31463-70.
31. Glatz A, et al. (1997) Chaperonin genes of the *Synechocystis* PCC 6803 are differentially regulated under light-dark transition during heat stress. *Biochem Biophys Res Commun* 239: 291-7.
32. Klähn S, Diederich A, Simon E, Hagemann M (2010) Salt-regulated gene expression in *Synechocystis* sp. PCC 6803: The gene *ssl3076* encodes a DNA-binding protein mediating the salt-induced expression of *ggpS* for the biosynthesis of the compatible solute glucosylglycerol. *J Bacteriol* 192: 4403-12.

33. Vinnemeier J, Hagemann M (1999) Identification of salt-regulated genes in the genome of the cyanobacterium *Synechocystis* sp. strain PCC 6803 by subtractive RNA hybridization. *Arch Microbiol* 172: 377-86.
34. Nakamoto H, Suzuki M, Kojima K (2003) Targeted inactivation of the hrcA repressor gene in cyanobacteria. *FEBS Lett* 549: 57-62.
35. Gutekunst K, et al. (2005) LexA regulates the bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803 as a transcription activator. *Mol Microbiol* 58: 810-23.
36. Aoki S, Kondo T, Ishiura M (1995) Circadian expression of the *dnaK* gene in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J Bacteriol* 177: 5606-11.
37. Fang F, Barnum SR (2004) Expression of the heat shock gene *hsp16.6* and promoter analysis in the cyanobacterium, *Synechocystis* sp. PCC 6803. *Curr Microbiol* 49: 192-8.
38. Navarro F, Martin-Figueroa E, Florencio FJ (2000) Electron transport controls transcription of the thioredoxin gene (*trxA*) in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Mol Biol* 43: 23-32.
39. Imamura S, et al. (2003) Purification, characterization, and gene expression of all sigma factors of RNA polymerase in a cyanobacterium. *J Mol Biol* 325: 857-72.
40. Seki A, et al. (2006) Light-responsive transcriptional regulation of the suf promoters involved in cyanobacterium *Synechocystis* sp. PCC 6803 Fe-S cluster biogenesis. *FEBS Lett* 580: 5044-8.
41. Voss B, et al. (2009) Biocomputational prediction of non-coding RNAs in model cyanobacteria. *BMC Genomics* 10: 123.
42. Vioque A (1992) Analysis of the gene encoding the RNA subunit of ribonuclease P from cyanobacteria. *Nucleic Acids Res* 20: 6331-7.

43. Tous C, Vega-Palas MA, Vioque A (2001) Conditional expression of RNase P in the cyanobacterium *Synechocystis* sp. PCC6803 allows detection of precursor RNAs. Insight in the *in vivo* maturation pathway of transfer and other stable RNAs. *J Biol Chem* 276: 29059-66.