DNA Research, 2018, 25(5), 489–497
doi: 10.1093/dnares/dsy019
Advance Access Publication Date: 12 June 2018
Full Paper

OXFORD

Full Paper

The GntR family transcriptional regulator PMM1637 regulates the highly conserved cyanobacterial sRNA Yfr2 in marine picocyanobacteria

S. Joke Lambrecht, J. Mascha L. Wahlig, and Claudia Steglich*

Faculty of Biology, Genetics and Experimental Bioinformatics, University of Freiburg, D-79104 Freiburg, Germany

*To whom correspondence should be addressed. Tel. +49 761 203 6986. Fax. +49 761 203 2745.
Email: claudia.steglich@biologie.uni-freiburg.de

Edited by Dr. Naotake Ogasawara

Received 26 January 2018; Editorial decision 11 May 2018; Accepted 14 May 2018

Abstract

Prochlorococcus is a marine picocyanobacterium with a streamlined genome that is adapted to different ecological niches in the oligotrophic oceans. There are currently >20 regulatory small RNAs (sRNAs) that have been identified in the model strain *Prochlorococcus* MED4. While most of these sRNAs are ecotype-specific, sRNA homologs of Yfr1 and of the Yfr2 family are widely found throughout the cyanobacterial phylum. Although they were identified 13 yrs ago, the functions of Yfr1 and Yfr2 have remained unknown. We observed a strong induction of two Yfr2 sRNA homologs of *Prochlorococcus* MED4 during high light stress and nitrogen starvation. Several *Prochlorococcus* and marine *Synechococcus yfr2* promoter regions contain a conserved motif we named CGRE1 (cyanobacterial GntR family transcriptional regulator responsive element 1). Using the conserved promoter region as bait in a DNA affinity pull-down assay we identified the GntR family transcriptional regulator PMM1637 as a binding partner. Similar to Yfr2, homologs of PMM1637 are universally and exclusively found in cyanobacteria. We suggest that PMM1637 governs the induction of gene expression of Yfr2 homologs containing CGRE1 in their promoters under nitrogen-depleted and high-light stress conditions.

Key words: GntR family transcriptional regulator, nitrogen and high light adaptation, picocyanobacteria, *Prochlorococcus*, regulation of sRNA Yfr2

1.Introduction

Consisting of only 1.65 Mbp the marine cyanobacterium *Prochlorococcus marinus* strain MED4 (*Prochlorococcus* MED4) possesses one of the smallest genomes of any photosynthetic organism.[1] The reduced genome size, correlating with a small cell size, and low GC content of 30.8% reflect adaptions to its extremely nutrient-poor habitat.[2] Members of the genus *Prochlorococcus* can be classified into different ecotypes, primarily driven by their distinct capabilities to cope with high or low light conditions.[3] *Prochlorococcus* MED4 is a laboratory model strain that belongs to the high-light (HL) adapted ecotype.[3] Other environmental factors such as temperature, nutrients, and competitor abundance are driving forces of niche partitioning within *Prochlorococcus* clades.[4,5] Nitrogen acquisition and regulation of nitrogen consumption are one of the greatest challenges that HL adapted *Prochlorococcus* ecotypes have to cope with[6] primarily because the availability of nitrogen dictates the limits of biomass

production.[7] In agreement with the importance of modulating nitrogen consumption is the presence of the nitrogen metabolism core regulatory proteins PII and NtcA in *Prochlorococcus* MED4.[8] In other cyanobacteria such as *Synechococcus elongatus* PCC 7942, the function of both PII and NtcA is modulated by a third protein, PipX. Inactive PipX interacts with PII and stabilizes the active form of NtcA when 2-oxoglutarate levels increase as a consequence of nitrogen starvation.[9] The homologue of *pipX*, encoded by PMM0393, completes the set of nitrogen regulatory proteins in *Prochlorococcus* MED4, though it's functional homology awaits experimental verification. The existence of a complex protein-based interaction network of nitrogen regulation in *Prochlorococcus* is in striking contrast to the general reduction in the number of regulatory proteins in this genus that is thought to have occurred by an intense genome streamlining process. For instance, only four two-component regulatory systems exist in MED4, compared to 44 in the cyanobacterial model *Synechocystis* sp. PCC 6803.[8] However, the reduction in the relative number of regulatory proteins in *Prochlorococcus* MED4 appears to be offset by a relatively high number of sRNAs, which require much less coding capacity and might serve as versatile post-transcriptional regulators of gene expression.[10]

Cyanobacterial functional RNA (Yfr)2 is an abundant sRNA discovered in picocyanobacteria 13 yrs ago.[11] In most cyanobacteria, small gene families with slightly divergent copies exist[12] but their functions and possible targets remained enigmatic. In *Prochlorococcus* MED4, four *yfr2* copies were identified, *yfr2*, *yfr3*, *yfr4*, and *yfr5*, while in the low-light (LL) adapted *Prochlorococcus* strain MIT9313 only two copies exist.[11] In contrast, nine copies were reported for *Synechococcus* CC9311.[12] Though it seems as if *Prochlorococcus* reduced the number of protein regulators relative to the number of sRNAs,[10] there is still the need to regulate the expression of these sRNAs. The genome-wide identification of transcriptional start sites at single-nucleotide resolution[13] enables the systematic study of promoter regions and the search for possible regulatory elements. Here, we identified the conserved motif we call CGRE1 (cyanobacterial GntR family transcriptional regulator responsive element 1). It is found in the promoter regions of several, but not all, *yfr2* homologs in *Prochlorococcus* and marine *Synechococcus*. We used DNA affinity pull-down chromatography to identify the binding of the GntR-like transcriptional regulator PMM1637 to CGRE1 and confirmed this further by *in vitro* and *in vivo* assays. Our results suggest that the induction of the *Prochlorococcus* MED4 homologs Yfr2 and Yfr4 during high light stress and nitrogen starvation is mediated by PMM1637.

2.Materials and methods

2.1.Culture growth conditions, RNA isolation, and Northern blot analysis

Prochlorococcus MED4 cultures were grown at 22°C in AMP1 medium[14] under 30 µmol quanta m^{-2} s^{-1} of continuous white cool light to cell densities of 1–3 × 10^8 cells per ml. For HL stress experiments, triplicate 1 l cultures were shifted to 300 µmol quanta m^{-2} s^{-1} of continuous white cool light. For every collection time point (immediately before the shift to HL, T0, and 15 min, 30 min, 1 h, 3 h, 6 h, 12 h, and 24 h after shift to HL) a 100 ml sample was withdrawn and collected by filtration onto 0.45 µm Supor-450 membranes (Pall). Filters were immersed in PGTX,[15] snap frozen in liquid nitrogen and stored at −80°C until further processing. Total RNA was extracted as described.[16] For nitrogen starvation experiments, triplicate 1 l cultures were grown to a cell density of 3 × 10^8 cells ml^{-1} and the T0 time point was collected. Cells were washed twice in nitrogen-free AMP1 medium by centrifugation for 10 min at 15,000*g* at 20°C and the cell pellet was resuspended in the same medium. Cultures were grown in nitrogen-depleted AMP1 medium and samples were collected as described above at 3, 6, 12, 24, 48, and 72 h. Northern blot analyses were performed as described,[17] and primers #82–87 (Supplementary Table S1) were used for preparing probe templates. Cellular transcript abundance was determined as previously described.[18]

2.2.DNA-protein affinity chromatography

Protein pull-down assays using the *yfr2* promoter or Δ*yfr2* promoter (*yfr2* promoter with randomly shuffled CGRE1 motif) were essentially performed as previously described.[19] In brief, 8 l cultures of *Prochlorococcus* MED4 cultures were pelleted by centrifugation at 15,000*g* for 10 min at 20°C and washed twice with nitrogen-free medium. After 24 h of cultivation in nitrogen-free AMP1 medium, cells were harvested by centrifugation at 15,000*g* for 10 min at 20°C and washed with buffer 1 (1.3 M glycine betaine, 10 mM HEPES pH 8). The cell pellet was resuspended in 4 ml BS/THES buffer (22 mM Tris–HCl, 4.4 mM EDTA, 8.9% sucrose (m/v), 62 mM NaCl, 10 mM HEPES, 5 mM $CaCl_2$, 50 mM KCl, 12% glycerol, pH 7.5).[19] BS/THES buffer was supplemented with proteinase inhibitor (cOmplete proteinase inhibitor cocktail, Roche), 0.5 mM $MgCl_2$, 50 µM $ZnSO_4$, 50 µM $FeCl_2$, 50 µM $MnCl_2$, and 50 µM EDTA. Cell lysis was achieved with the One Shot constant cell disruption system (Constant Systems Limited) at 1.4 kbar with one repetition. Cell debris was pelleted by centrifugation at 13,000*g* for 30 min at 4°C. The supernatant was incubated with 200 µl of uncharged Dynabeads M-280 Streptavidin (ThermoFisher Scientific) for 20 min to minimize the non-specific binding of contaminants to the beads before incubating with beads loaded with the bait *yfr2* promoter or Δ*yfr2* promoter for 1 h at room temperature. Before incubating with lysate, Dynabeads were washed and prepared as described by the manufacturer. *Yfr2* and Δ*yfr2* promoters were amplified by PCR using primers #1 and #2 or #3, respectively, with KAPA-HiFi polymerase (Kapa Biosystems). For the biotinylated primer, a limiting amount of 0.2 µM was applied and the PCR products were loaded on Dynabeads in 1× BW Buffer (5 mM Tris–HCl pH 7.5, 0.5 mM EDTA, and 1 M NaCl). After incubating with cell lysate, beads charged with bait promoters and bound proteins were washed with BS/THES buffer and proteins were eluted with elution buffer (25 mM Tris–HCl, 100/250/500/750/2000 mM NaCl). Protein elution fractions were precipitated overnight at 4°C by adding trichloroacetic acid to a final concentration of 15%. Proteins were pelleted by centrifugation at 13,000*g* for 30 min at 4°C. Protein pellets were washed with 50% cold acetone, resuspended in water and supplemented with protein loading buffer. Samples were denatured at 95°C for five min and treated with 10 mM iodoacetamide at room temperature in the dark for 30 min. Elution fractions were analysed on RunBlue SDS Protein Gels 4–20% (Expedeon) in a Tris-Tricine running buffer system. Gels were stained with InstantBlue (Expedeon). Proteins specific for the *yfr2* promoter elution fraction were analysed by mass spectrometry (Core Facility Proteomics, ZBSA, University of Freiburg).

2.3.Generation of expression vectors

The gene PMM1637 was amplified from genomic DNA with primers #13 and #14 (Supplementary Table S1), digested with *Sph*I and *Bam*HI and introduced into the expression vector pQE-70. Primers #20–67 (Supplementary Table S1) were used to generate point mutation variants of PMM1637 using the Q5 Site-Directed Mutagenesis Kit (NEB) or AQUA cloning.[20] The construct pQE-70_PMM1637

encoding the wild-type protein served as a template for codon optimization of the first 59 and last 75 amino acids using Q5 Site-Directed Mutagenesis Kit (NEB) and primers #15–19 (Supplementary Table S1). The PMM1637_R50A mutant was derived from the codon-optimized construct pQE_PMM1637 by PCR using primers #16 and #19 (Supplementary Table S1) and transformed into *E. coli* Top10 F' using AQUA cloning.[20] For recombinant protein expression, the constructs pQE_PMM1637 and PMM1637_R50A were transformed into *E. coli* M15.

The backbone pXG10SF[21] was amplified using primers #72 and #73 (Supplementary Table S1). Wild-type and mutant *yfr2* promoters were generated by annealing phosphorylated complementary primers #68 and #69 or #70 and #71 (Supplementary Table S1) followed by ligation of annealed primers into the pXG10SF backbone. The resulting vectors were templates for PCR amplification of vectors containing *yfr3, yfr4, yfr5*, or PMM1637 promoter regions (for primers see Supplementary Table S1 #74–81). Plasmids containing promoter regions of *yfr3, yfr4, yfr5*, or PMM1637 were generated using the Q5 Site-Directed Mutagenesis Kit (NEB) or by AQUA cloning.[20]

2.4.*E. coli* GFP *in vivo* assay

pXG10SF vectors containing *yfr2*, *yfr3*, *yfr4*, *yfr5*, and *yfr2* mutant or PMM1637 promoter constructs were co-transformed with either pQE_PMM1637 or pQE_PMM1637_R50A into *E. coli* TOP10. Colonies were inoculated into 200 µl of lysogeny broth medium supplemented with chloramphenicol and ampicillin in a 96-well plate and incubated overnight at 37°C with gentle agitation at 150 rpm in an air humidity-saturated environment to prevent evaporation. Cells were diluted 1: 10, fixed with 1% Roti Histofix (Roth) and analysed on an Accuri C6 flow cytometer (Beckman-Coulter) monitoring GFP fluorescence at an excitation wavelength of 488 nm and emission wavelength at 533/30 nm. The mean fluorescence per plasmid combination was calculated from 50,000 events (cells) of six individual clones. Fold changes were calculated by dividing the average of the mean fluorescence from six clones for the respective combinations. For the verification of promoter functionality, 5′ RACE experiments were performed as described.[10] Western blot analysis for the detection of GFP was performed with anti-GFP antibody (Sigma) according to the manufacturer's description. Western blot membranes were stained with Pierce Reversible Protein Stain Kit (Thermo Scientific) according to the manufacturer's description. PMM1637 WT and PMM1637 R50A recombinant proteins were detected with anti-Penta-His antibody conjugated to horseradish peroxidase (Qiagen) according to the manufacturer's description.

2.5.Recombinant protein expression and purification

The codon-optimized vectors pQE_PMM1637 and pQE_PMM1637_R50A were transformed in *E. coli* M15 for overexpression of recombinant proteins PMM1637 and PMM1637_R50A. Overnight cultures were diluted 1: 100 in fresh terrific broth medium supplemented with ampicillin and kanamycin, and grown to OD_{600} 0.5. Protein expression was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG, 2 mM final concentration) and cells were harvested by centrifugation at 6,000*g* for 10 min at 4°C 5 h after induction. Cell pellets were resuspended in lysis buffer (50 mM NaH_2PO_4 pH 8, 1 M NaCl, 10% glycerol, 0.5% Tween20, 7.5 mM imidazole, complete proteinase inhibitor cocktail, Roche) and lysed using the One Shot constant cell disruption system (Constant Systems Limited, United Kingdom) at 2.4 kbar in two consecutive cycles. Cell debris was pelleted by centrifugation at 13,000*g* for 30 min at 4°C and the lysate was filtered through 0.45 µm Supor-450 filters (Pall). Recombinant proteins were immobilized on a HiTrap Talon crude column (GE Healthcare) and eluted with elution buffer (50 mM NaH_2PO_4 pH 7,400 mM imidazole, 100 mM NaCl). Elution fractions were combined and supplemented with 5 mM dithiothreitol (DTT). Combined immobilized metal ion chromatography elution fractions were loaded on HiTrap SP HP columns equilibrated with buffer A (50 mM NaH_2PO_4 pH 7.6, 25 mM NaCl). The column was washed with 17.5% buffer B (50 mM NaH_2PO_4 pH 7.2, 1 M NaCl) and protein was eluted with 33% buffer B. Elution factions were again supplemented with 5 mM DTT, and the buffer was changed to TBS including 5 mM DTT using Amicon Ultra Centrifugal Filters 10 K (Merck). Protein concentration was calculated by measuring the absorption at 280 nm.

2.6.Electrophoretic mobility shift assay (EMSA)

Cy3 end-labelled probes for EMSA were generated using primers #4–12 (Supplementary Table S1). Reaction conditions for binding protein to DNA were 12 mM HEPES, 15.7 mM Tris, 1 mM EDTA, 1 mM EGTA, 40.16% glycerol, 1 µg LightShift Poly (dIdC) (Thermo Fischer Scientific), 60 mM KCl, 64.2 mM NaCl, and 15.625 nM DNA probe. The concentration of proteins used for shift experiments were 595.4, 426.5, 297.3, 193.4, and 128.4 nM. Samples were run on 1.5% agarose gels in 0.5 × TAE.

3.Results

The degree of conservation and wide distribution of the *yfr2* family within the cyanobacterial phylum prompted us to determine whether the regulation of *yfr2* itself is conserved. Bioinformatic analysis led to the identification of a highly conserved sequence starting 55–57 nucleotides upstream of the *yfr2* transcriptional start site, which is present in *Prochlorococcus* and marine *Synechococcus*. The highly conserved motif was further analysed by multiple sequence alignments of *yfr2* promoters (Supplementary Fig. S1) utilizing primary transcriptome data for the strains MED4 and MIT9313[13] or inferring the first transcribed nucleotide based on sequence comparison.[12] The highly conserved motif is 22 nucleotides long in HL *Prochlorococcus* strains (Fig. 1B); however, in LL-adapted *Prochlorococcus* and marine *Synechococcus* strains, it is one nucleotide shorter (Fig. 1A). The promoter sequence of *yfr2c* in *Prochlorococcus* SS120 contains a completely conserved inverted repeat (Supplementary Fig. S1). In the *yfr2a* promoter of *Prochlorococcus* MIT9313, the start of the conserved motif is not found at the usual −55 to −57 position, but at −20. In *Prochlorococcus* MED4, the highly conserved motif shows the highest conservation in the *yfr2* and *yfr4* promoters, while it is entirely degenerated in the *yfr3* promoter (Supplementary Fig. S1A). One half of the motif is highly conserved in the *yfr5* promoter, while the other half is disrupted by the terminator of the upstream *yfr18* gene. Yfr3 is also the least abundant *yfr2* homolog in *Prochlorococcus* MED4 (Supplementary Fig. S2 and Table S2) indicating that the disrupted motif in *yfr5* might still be sufficient for some transcription.

To identify possible binding proteins of the highly conserved motif, we utilized DNA affinity pull-down.[19] As bait we used a 155 nt *Prochlorococcus* MED4 *yfr2* promoter fragment and a mutated version (Δ*yfr2*) with a randomly shuffled motif (Fig. 2B). Captured proteins were analysed by SDS-PAGE and subsequently by mass spectrometry. A protein band at approximately 45 kDa was present in the 500 mM NaCl elution fraction of the *yfr2* promoter bait that

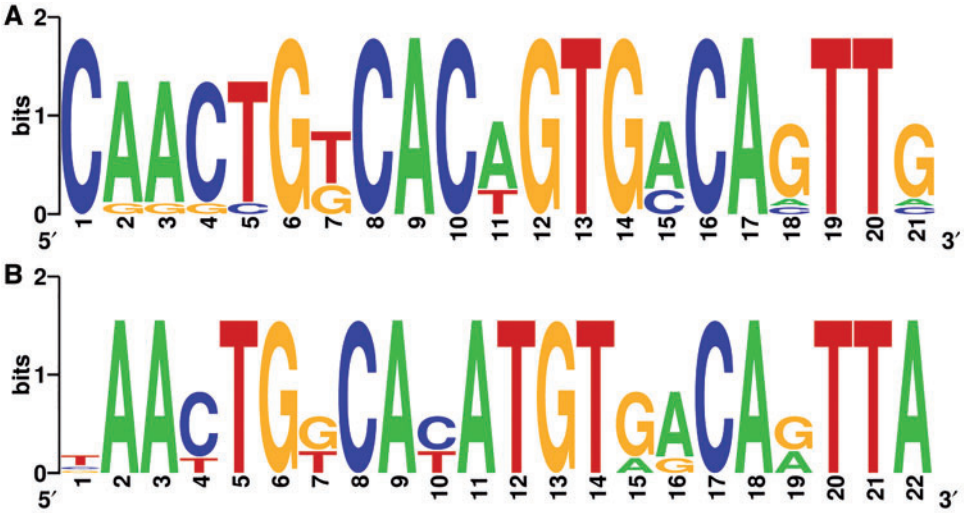

Figure 1. Consensus motif of the highly conserved sequence element. (A) Sequence logo of aligned promoter regions of *yfr2* homologs of the *Prochlorococcus* LL type consisting of LL-adapted *Prochlorococcus* and of marine *Synechococcus* strains. (B) Sequence logo of aligned promoter regions of *yfr2* homologs of the *Prochlorococcus* HL type, exclusively found in HL-adapted *Prochlorococcus* strains. Logos were created based on sequences of the sequence alignment in Supplementary Figure S1 using the sequence logo generator WebLogo.[31]

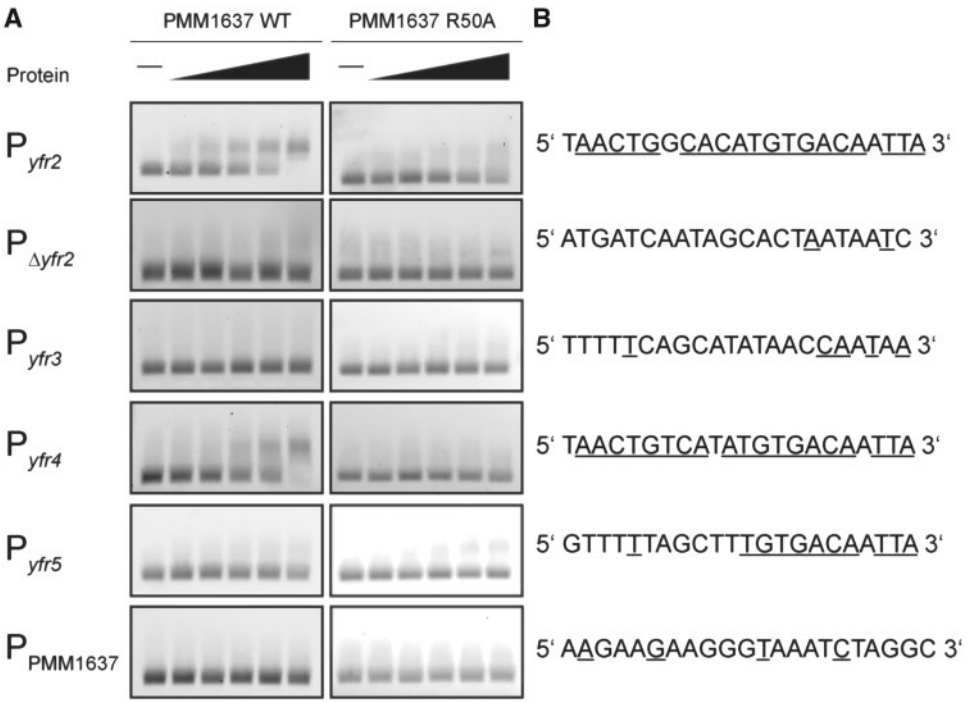

Figure 2. DNA-binding properties of purified recombinant PMM1637 protein. (A) PMM1637 DNA binding was tested by EMSA assays using Cy3-labelled *yfr2*, Δ*yfr2*, *yfr3*, *yfr4*, *yfr5* and PMM1637 promoter fragments. Binding reactions contained 128.4, 193.4, 297.3, 426.5, or 595.4 nM recombinant PMM1637 or PMM1637 R50A protein in the presence of 15.625 nM labelled promoter fragment and 1 μg poly (dI–dC) as a non-specific competitor. (B) Sequences found in the respective promoter regions starting 55–57 bp upstream of the transcriptional start site. Nucleotides that are in agreement with the consensus sequence are underlined.

was not visible in any of the other elution fractions loaded onto the gel (Supplementary Fig. S3). Among the proteins that were identified by mass spectrometry from a gel slice of the specific protein band was PMM1637, which was enriched 28-fold compared to the gel sample of the equivalent area of the Δ*yfr2* promoter 500 mM NaCl elution fraction (Supplementary Table S3). PMM1637 is a 323 amino acid protein that is annotated as a possible transcriptional regulator. Indeed, searches against the pfam, STRING and EggNOG databases showed an N-terminal Winged helix-turn-helix (WHTH) DNA-binding domain, characteristic of the GntR family of transcriptional regulators. Hence, we considered PMM1637 as a promising candidate for a transcription factor governing Yfr2 transcription via

binding to the highly conserved motif. We named the palindromic sequence CGRE1 for cyanobacterial GntR family transcriptional regulator responsive element 1.

Next, we purified the recombinant PMM1637 wild-type protein (PMM1637 WT from here on), overexpressed in *E. coli*, by immobilized metal ion chromatography followed by ion exchange chromatography (Supplementary Fig. S4) and performed electrophoretic mobility shift assay (EMSA). Promoters of *yfr2* and *yfr4* were shifted whereas *yfr3* and *yfr5* promoters were not (Fig. 2A). This result correlates with the presence of CGRE1 in the *yfr2* and *yfr4* promoter regions and the degeneration and partial loss in *yfr3* and *yfr5* promoter regions, respectively. Furthermore, PMM1637 WT did not bind to its own promoter, indicating that it does not autoregulate its own expression. For verification, we mutated an ultraconserved arginine residue at position 50 to alanine in PMM1637 (PMM1637 R50A from here on) (Supplementary Fig. S5A). According to the NCBI conserved domains database, residue R50 is involved in nucleic acid binding in the GntR transcriptional regulator family.[22] R50 is located within the WHTH DNA-binding domain (accession number cd07377) of PMM1637 spanning amino acids 12–77. EMSA analysis showed that the PMM1637 R50A mutant was not able to shift the *yfr2* or *yfr4* promoter (Fig. 2A), hence, confirming that the DNA binding ability of the PMM1637 R50A mutant was greatly reduced.

There is no genetic system for *Prochlorococcus*. Therefore, we modified the heterologous GFP reporter system that is otherwise used to identify sRNA-mRNA interactions[21] to further analyse the influence of PMM1637 on *yfr2* promoters. We co-expressed the PMM1637 WT or PMM1637 R50A protein encoded on the pQE-70 plasmid with the pXG10-SF plasmid that contains a 5′ UTR conferring high expression of GFP in combination with various promoters (Fig. 3A). The results confirmed that all promoters were recognized in *E. coli* and consequently, GFP was produced (Supplementary Fig. S6A). We can exclude the possibility of erroneous promoter usage, as 5′ RACE experiments showed transcription initiation from the correct start point or in close proximity (Supplementary Figs S7 and S8). Expression from the *yfr2* or *yfr4* promoter was higher in the presence of PM1637 WT in comparison to PMM1637 R50A, while the expression from the Δ*yfr2*, *yfr3*, *yfr5*, and PMM1637 promoters, as well as the native pXG10-SF promoter P_{LtetO} was not altered regardless of the co-expressed protein (Fig. 3B). We can exclude that differences in GFP fluorescence intensities are caused by unequal protein expression levels of PMM1637 WT and PMM1637 R50A (Supplementary Fig. S6B and C). The positive influence of PMM1637 WT on the *yfr2* promoter is also supported by the fact that the transcriptional regulator PMM0637 (Fur), which is not related to Yfr2 expression, showed the same level of expression as the inactive PMM1637 R50A (Fig. 3A–C). All findings are in agreement with the *in silico* analysis and *in vitro* experiments described above. Transcription factors of the GntR family contain the small DNA-binding WHTH domain at their N-terminus and a very diverse regulatory ligand-binding domain within the C-terminal half,[23] which provides the basis for the subfamily classification. Because there is no information about the regulatory domain of PMM1637, we used this GFP system to further elucidate the function of several other highly conserved residues in the C-terminal effector domain (Supplementary Fig. S5B and C). Based on sequence conservation among 20 PMM1637 homologs (Supplementary Fig. S5A), highly conserved amino acids with putative reactive groups such as carboxy-, guanidine-, or amine groups were chosen for point mutations. Among the 24 investigated mutants, five others (D141A, T182A, R200A, R236A, and D276A) showed a comparable average fold change between the *yfr2* and Δ*yfr2* promoter as PMM1637 R50A did (Supplementary Fig. S5C), suggesting that all these six amino acids are important for protein activity. The mutants K217M and S231A showed a slightly higher average fold change than the PMM1637 WT (Supplementary Fig. S5C), hence, these substitutions might possibly have locked the protein into an active conformation.

Previous studies on PMM1637 homologs in *Synechocystis* sp. PCC 6803 and *Synechococcus elongatus* PCC 7942 suggested that this transcription factor is involved in regulation during HL adaptation and nitrogen starvation, respectively.[24–26] We, therefore, monitored *Prochlorococcus* MED4 Yfr2, Yfr3, Yfr4, and Yfr5 expression during these conditions. Nitrogen depletion caused the strongest induction of gene expression for the Yfr2 and Yfr4 homolog, increasing from approximately 100 to 1,000 copies per cell after 48 h and from 20 to 144 copies per cell after 72 h of nitrogen starvation, respectively (Fig. 4A and Supplementary Table S2). The response and accumulation pattern observed here was delayed in comparison to the induction of *ntcA* transcript accumulation that we used as an indicator of nitrogen stress (Fig. 4C). Whereas *ntcA* was already strongly induced after 3 h of nitrogen depletion, Yfr2 reached its maximum only 48 h and Yfr4 only 72 h after nitrogen depletion (Fig. 4A and C). Shifting *Prochlorococcus* MED4 cells from 30 to 300 μE resulted in a rapid increase of Yfr2 and Yfr4 transcript accumulation already visible after 15 min of HL exposure (Fig. 4B and Supplementary Table S2). The expression of Yfr2 began to decline after 30–45 min of HL exposure, while the induced expression of Yfr4 plateaued and only dropped down to standard levels after 24 h of HL treatment (Fig. 4B and Supplementary Table S2). Compared to nitrogen depletion, the response of *ntcA* to HL stress was remarkably weaker. During the course of HL exposure only a slight increase in *ntcA* transcript was visible, which might indicate the first signs of an imbalanced C/N ratio in the cell (Fig. 4D).

4.Discussion

In this study, we identified CGRE1, a conserved motif in the promoter of several genes of the sRNA *yfr2* family that is recognized by the *Prochlorococcus* MED4 GntR family transcriptional regulator PMM1637. CGRE1 motifs observed in HL- and LL-adapted ecotypes differ in length and nucleotide composition. The LL motif is flanked by 'GTTG' and the HL motif by 'GTTA' (Supplementary Fig. S1). The G to A transition is consistent with the low GC content in HL *Prochlorococcus*. The identified CGRE1 motif recognized by PMM1637 contains sequence elements that have been described before for other motifs of GntR family transcriptional regulators. The flanking 'GTTG' of the LL type and the 'GTTA' of the HL type resemble parts of what has been found for CceR from *Rhodobacter sphaeroides*.[27] However, unlike in CGRE1 described here, the sequence and its inverted repeat are separated by only four nucleotides instead of 13–14. In the same study, EMSA assays showed that the T positions are crucial for binding.[27] AkgR from *Rhodobacter sphaeroides* binds to the direct repeat of 'GTGATCAC' separated by 13 nucleotides.[27] We find almost the same sequence in the middle of CGRE1 except that 'AT' in the center is inverted in the HL CGRE1 and one position is deleted in the LL CGRE1. There is also no repetition of 'GTGATCAC' in *yfr2* promoters within a distance of 13 nucleotides.

PMM1637 has a homolog in all cyanobacterial genomes. The first functional characterization of a cyanobacterial GntR family

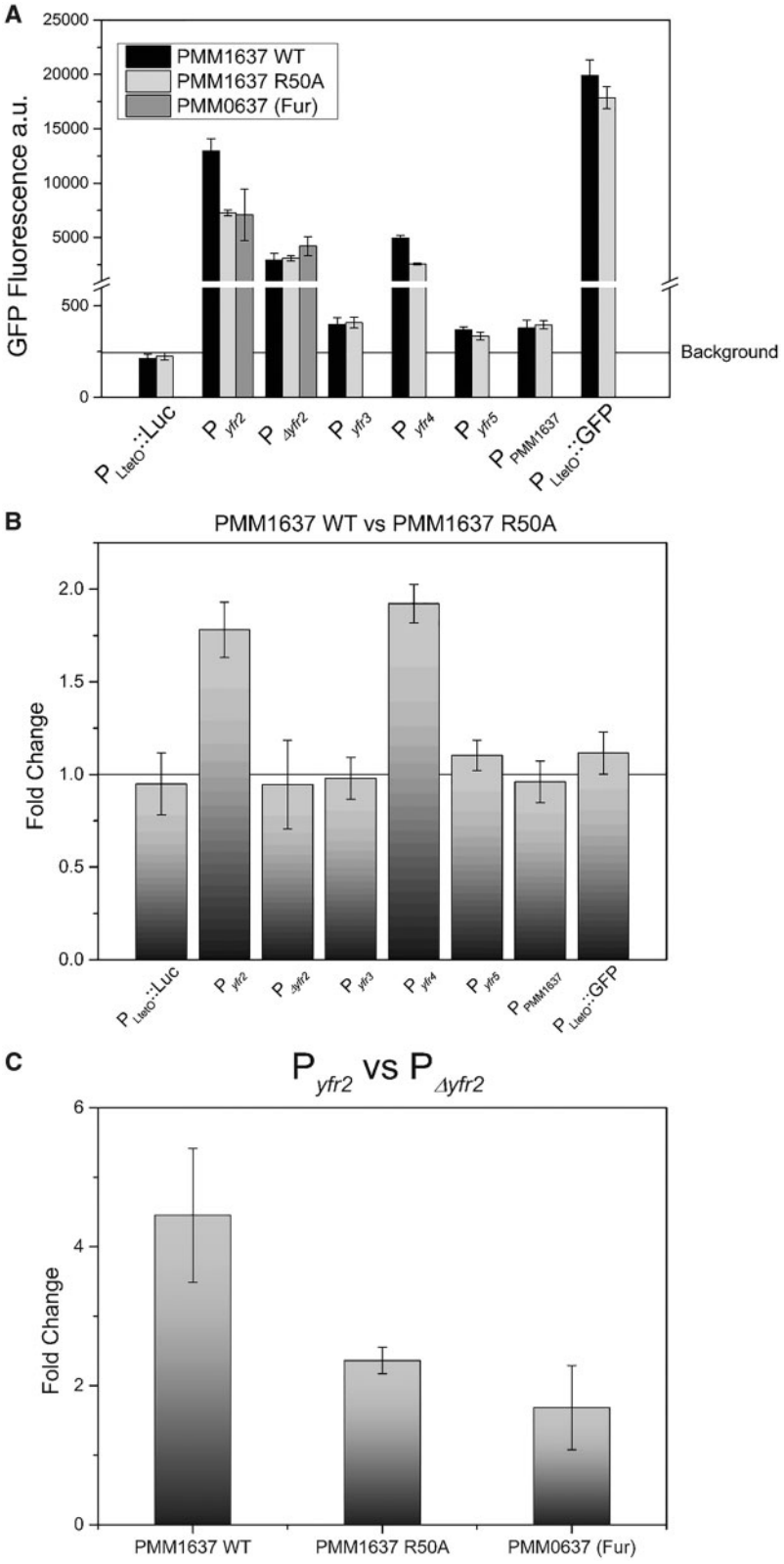

Figure 3. Influence of the GntR family transcriptional regulator PMM1637 on *yfr2* family promoter activity. (A) GFP expression driven by *Prochlorococcus* MED4 *yfr2*, Δ*yfr2*, *yfr3*, *yfr4*, *yfr5*, and PMM1637 promoters, or from the P_{LtetO} promoter. P_{LtetO}:: Luc (pXG-0[32]) codes for luciferase instead of GFP (P_{LtetO}:: GFP) and serves as a control for the background fluorescence of *E. coli* cells. GFP expression was carried out in the presence of PMM1637 WT, PMM1637 R50A, or PMM0637 (Fur) encoded on pQE-70. (B) Fold change differences between PMM1637 WT and PMM1637 R50A for each promoter were calculated from the results shown in A). (C) Fold change differences between *Prochlorococcus* MED4 *yfr2* and Δ*yfr2* promoter in the presence of either PMM1637 WT, PMM1637 R50A, or PMM0637 (Fur) were calculated from results shown in A).

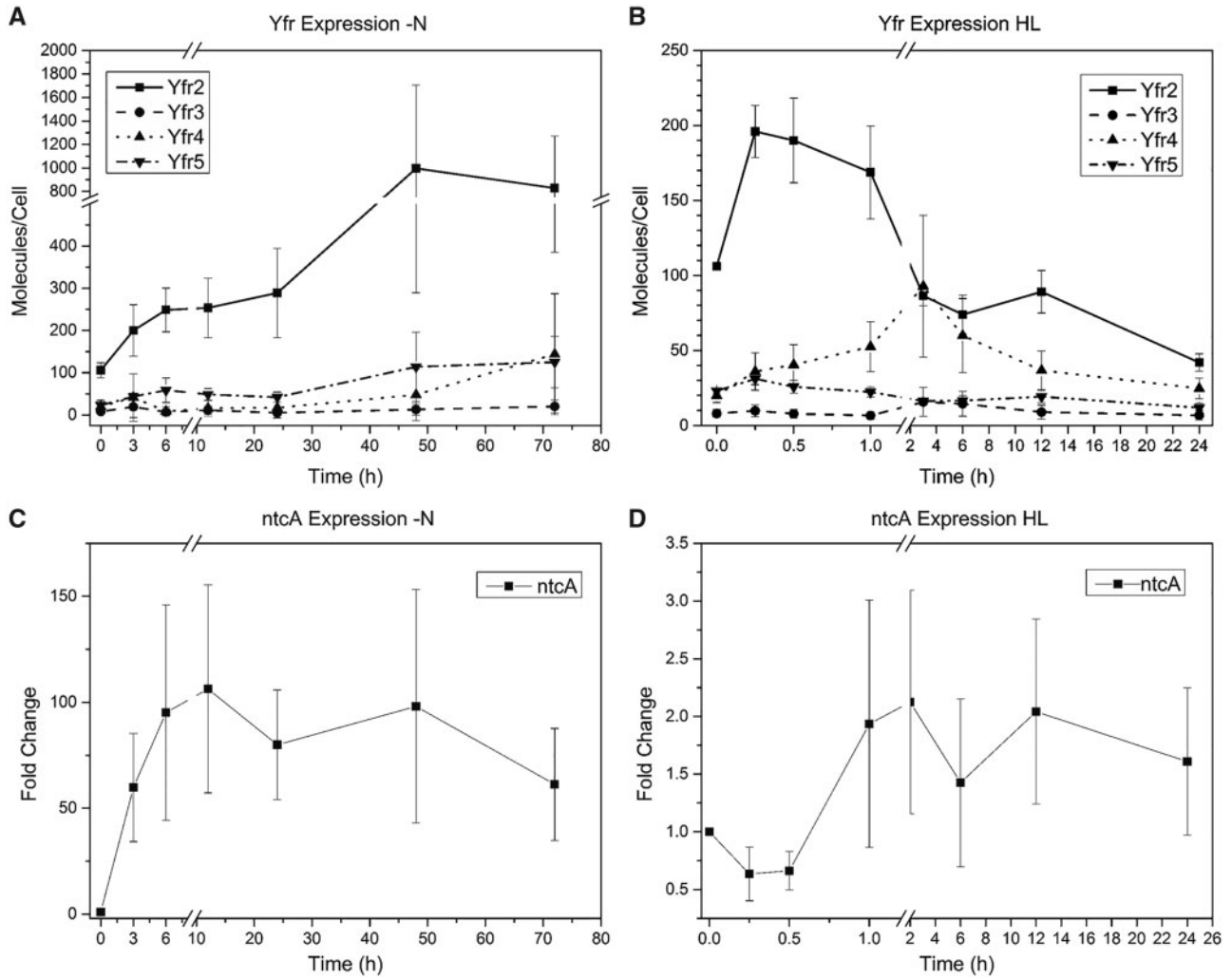

Figure 4. Temporal gene expression kinetics of Yfr2, Yfr3, Yfr4, and Yfr5 during HL acclimation or nitrogen starvation. (A) Yfr2-5 expression of *Prochlorococcus* MED4 samples after 0, 3, 6, 12, 24, 48, and 72 h of nitrogen starvation and of (C) *ntcA* expression. (B) *Prochlorococcus* MED4 cells were shifted from 30 to 300 μE and samples were taken at 0, 0.25, 0.5, 1, 3, 6, 12, and 24 h after stress application and probed against Yfr2-5 or (D) *ntcA*. Transcripts per cell were calculated for Yfr2-5 and relative amounts of transcript for *ntcA*.

transcription regulator was provided by Lee et al., who found evidence that in *Anabaena* sp. PCC 7120 the homolog acts as a plasmid maintenance factor and therefore called it PlmA.[28] The same study defined a new subfamily of GntR family transcriptional regulators based on the divergence in the C-terminal effector domain that includes PMM1637 (formerly annotated as Pmar1604) and PlmA.[28] In a recent yeast three-hybrid screening of genomic libraries from *Synechococcus elongatus* PCC7942, PlmA was found to bind to the complex of PII-PipX.[25] This interaction seems to be facilitated by PII-bound PipX and occurs only during low levels of the nitrogen status reporter molecule 2-oxoglutarate.[25] However, the activation of PMM1637 does not appear to directly depend on 2-oxoglutarate because the accumulation kinetics of its targets Yfr2 and Yfr4 differ from that of *ntcA*, which is activated at high 2-oxoglutarate concentrations.[27] Accumulation of Yfr2 and Yfr4 during early HL adaptation and virtually no changes in *ntcA* gene expression emphasize that 2-oxoglutarate is not involved in PMM1637-dependent regulation. It is therefore likely that PMM1637 is regulated by another cofactor that signals during late nitrogen depletion stress and early HL adaptation, possibly connecting C- and N-regulatory circuits. In *Synechocystis* sp. PCC 6803, the PlmA homolog Sll1961 is involved in phycobilisome degradation during nitrogen starvation.[26] The disruption of *sll1961* leads to a non-bleaching phenotype of the *sll1961* mutant, despite elevated levels of the phycobilisome degradation protein NblA during nitrogen starvation.[26] Another study showed that the *sll1961* mutant strain retained a constant PSII/PSI ratio throughout all tested light conditions while the wild-type ratio increased during acclimation to high light,[24] which could be explained by a weaker decline in the number of PSI reaction center subunits PsaA and PsaB in the mutant compared to the wild-type.[24] The regulation of PsaAB seems to occur at the translational or post-translational level because the decrease in *psaAB* mRNA accumulation was comparable in wild-type and mutant.[24] In *Prochlorococcus* the inverse relationship in the levels of Yfr2 and *psaAB* mRNA under high light[16] and nitrogen starvation[29] suggests that the *psaAB* mRNA might be a target of Yfr2. The latter reports are consistent with our findings of increased Yfr2 levels during nitrogen starvation and HL stress, though the functional orthology between the GntR-like factors of different cyanobacteria certainly requires more detailed investigation.

Pairs of transcription factors and sRNAs are building blocks in the regulatory network, including the response to nitrogen starvation in cyanobacteria, as exemplified by the NtcA-NsiR4 regulatory system in *Synechocystis* sp. PCC 6803.[30] However, the architecture of the regulatory machinery mediating this response was postulated to differ in marine picocyanobacteria. While NtcA is also conserved in *Prochlorococcus*, NsiR4 is not, which exemplarily shows that stress-related sRNAs are mostly restricted to certain groups within cyanobacteria but protein regulators are not, depending on the function. In contrast, the GntR family transcriptional regulator, as well as the *yfr2* family, are globally distributed among cyanobacteria. However, CGRE1 appears to be restricted to marine *Synechococcus* and *Prochlorococcus*. Therefore, regulation of Yfr2 by PMM1637 could be a result of the streamlined regulatory network that is found within the picocyanobacteria. Thus, this finding provides another example how widely conserved genes are differentially used in closely related taxa. Future studies will certainly help to answer the question of whether the homologs of PMM1637 also regulate members of the Yfr2 family in other cyanobacteria or if the targets of Yfr2 in *Prochlorococcus* are elsewhere regulated by the PMM1637 homolog. A study on the global response of gene expression during nitrogen starvation in *Prochlorococcus* MED4 identified the differential gene expression of 16% of the entire genome.[29] However, only a small subset of these genes possesses an NtcA binding site and therefore, many must be controlled by other factors such as PMM1637 and Yfr2. The occurrence of at least two nitrogen stress-related regulons in marine picocyanobacteria might be connected to their nitrogen-poor habitat, leading to the requirement of more regulatory circuits for a fine-tuned acclimation. The results of previous studies and of this work underline the importance of the cyanobacterial GntR family in the regulation of gene expression during nitrogen starvation and hint to an important role of this transcription factor family in maintaining the homeostasis of nitrogen and carbon metabolism. The future identification of targets of the Yfr2-PMM1637 regulon will shed light on their functional role in this complex network and may uncover hidden layers of bacterial gene regulation.

Acknowledgements

We thank Wolfgang R. Hess for helpful discussions and critical reading of the manuscript. The research was supported by the DFG (SPP 1258) to C.S.

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at *DNARES* online.

References

1. Rocap, G., Larimer, F. W., Lamerdin, J., et al. 2003, Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation, *Nature*, **424**, 1042–7.
2. Dufresne, A., Garczarek, L. and Partensky, F. 2005, Accelerated evolution associated with genome reduction in a free-living prokaryote, *Genome Biol.*, **6**, R14.
3. Moore, L. R. and Chisholm, S. W. 1999, Photophysiology of the marine cyanobacterium *Prochlorococcus*: ecotypic differences among cultured isolates, *Limnol. Oceanogr.*, **44**, 628–38.
4. Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. S. and Chisholm, S. W. 2006, Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients, *Science*, **311**, 1737–40.
5. Chandler, J. W., Lin, Y., Gainer, P. J., Post, A. F., Johnson, Z. I. and Zinser, E. R. 2016, Variable but persistent coexistence of *Prochlorococcus* ecotypes along temperature gradients in the ocean's surface mixed layer: temperature impacts *Prochlorococcus* ecotype ratios, *Environ. Microbiol. Rep.*, **8**, 272–84.
6. Moore, L. R., Post, A. F., Rocap, G. and Chisholm, S. W. 2002, Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*, *Limnol. Oceanogr.*, **47**, 989–96.
7. Read, R. W., Berube, P. M., Biller, S. J., et al. 2017, Nitrogen cost minimization is promoted by structural changes in the transcriptome of N-deprived *Prochlorococcus* cells, *ISME J.*, **11**, 2267–78.
8. Scanlan, D. J., Ostrowski, M., Mazard, S., et al. 2009, Ecological genomics of marine picocyanobacteria, *Microbiol. Mol. Biol. Rev.*, **73**, 249–99.
9. Espinosa, J., Rodriguez-Mateos, F., Salinas, P., et al. 2014, PipX, the coactivator of NtcA, is a global regulator in cyanobacteria, *Proc. Natl. Acad. Sci. USA.*, **111**, E2423–30.
10. Steglich, C., Futschik, M. E., Lindell, D., Voss, B., Chisholm, S. W. and Hess, W. R. 2008, The challenge of regulation in a minimal photoautotroph: non-coding RNAs in *Prochlorococcus*, *PLoS Genet.*, **4**, e1000173.
11. Axmann, I. M., Kensche, P., Vogel, J., Kohl, S., Herzel, H. and Hess, W. R. 2005, Identification of cyanobacterial non-coding RNAs by comparative genome analysis, *Genome Biol.*, **6**, R73.
12. Gierga, G., Voss, B. and Hess, W. R. 2009, The Yfr2 ncRNA family, a group of abundant RNA molecules widely conserved in cyanobacteria. *RNA Biol.*, **6**, 222–7.,
13. Voigt, K., Sharma, C. M., Mitschke, J., et al. 2014, Comparative transcriptomics of two environmentally relevant cyanobacteria reveals unexpected transcriptome diversity, *ISME J.*, **8**, 2056–68.
14. Moore, L. R., Coe, A., Zinser, E. R., et al. 2007, Culturing the marine cyanobacterium *Prochlorococcus*: *Prochlorococcus* culturing, *Limnol. Oceanogr. Methods*, **5**, 353–62.
15. Pinto, F. L., Thapper, A., Sontheim, W. and Lindblad, P. 2009, Analysis of current and alternative phenol based RNA extraction methodologies for cyanobacteria, *BMC Mol. Biol.*, **10**, 79.
16. Steglich, C., Futschik, M., Rector, T., Steen, R. and Chisholm, S. W. 2006, Genome-wide analysis of light sensing in *Prochlorococcus*, *J. Bacteriol.*, **188**, 7796–806.
17. Stazic, D., Lindell, D. and Steglich, C. 2011, Antisense RNA protects mRNA from RNase E degradation by RNA-RNA duplex formation during phage infection, *Nucleic Acids Res.*, **39**, 4890–9.
18. Nakamura, T., Naito, K., Yokota, N., Sugita, C. and Sugita, M. 2007, A cyanobacterial non-coding RNA, Yfr1, is required for growth under multiple stress conditions, *Plant Cell Physiol*, **48**, 1309–18.
19. Jutras, B. L., Verma, A. and Stevenson, B. 2012, Identification of novel DNA binding proteins using DNA affinity chromatography-pulldown, *Curr. Protoc. Microbiol.*, **Supplement 24**, Unit1F.1.
20. Beyer, H. M., Gonschorek, P., Samodelov, S. L., Meier, M., Weber, W. and Zurbriggen, M. D. 2015, AQUA cloning: a versatile and simple enzyme-free cloning approach, *PLoS One*, **10**, e0137652.
21. Corcoran, C. P., Podkaminski, D., Papenfort, K., Urban, J. H., Hinton, J. C. D. and Vogel, J. 2012, Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA, *Mol. Microbiol.*, **84**, 428–45.
22. Marchler-Bauer, A., Bo, Y., Han, L., et al. 2017, CDD/SPARCLE: functional classification of proteins via subfamily domain architectures, *Nucleic Acids Res.*, **45**, D200–3.
23. Rigali, S., Derouaux, A., Giannotta, F. and Dusart, J. 2002, Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies, *J. Biol. Chem.*, **277**, 12507–15.
24. Fujimori, T., Higuchi, M., Sato, H., et al. 2005, The mutant of *sll1961*, which encodes a putative transcriptional regulator, has a defect in regulation of photosystem stoichiometry in the Cyanobacterium *Synechocystis* sp. PCC 6803, *Plant Physiol.*, **139**, 408–16.

25. Labella, J. I., Obrebska, A., Espinosa, J., et al. 2016, Expanding the cyanobacterial nitrogen regulatory network: the GntR-like regulator PlmA interacts with the PII-PipX complex, *Front. Microbiol.*, **7**, 1664–302X.
26. Sato, H., Fujimori, T. and Sonoike, K. 2008, sll1961 is a novel regulator of phycobilisome degradation during nitrogen starvation in the cyanobacterium *Synechocystis* sp. PCC 6803, *FEBS Lett.*, **582**, 1093–6.
27. Imam, S., Noguera, D. R. and Donohue, T. J. 2015, CceR and AkgR regulate central carbon and energy metabolism in alphaproteobacteria, *mBio*, **6**, e02461-14.
28. Lee, M. H., Scherer, M., Rigali, S. and Golden, J. W. 2003, PlmA, a new member of the GntR family, has plasmid maintenance functions in *Anabaena* sp. strain PCC 7120, *J. Bacteriol.*, **185**, 4315–25.
29. Tolonen, A. C., Aach, J., Lindell, D., et al. 2006, Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability, *Mol. Syst. Biol.*, **2**, 53.
30. Klähn, S., Schaal, C., Georg, J., et al. 2015, The sRNA NsiR4 is involved in nitrogen assimilation control in cyanobacteria by targeting glutamine synthetase inactivating factor IF7, *Proc. Natl. Acad. Sci. USA.*, **112**, E6243–52.
31. Crooks, G. E., Hon, G., Chandonia, J.-M. and Brenner, S. E. 2004, WebLogo: a sequence logo generator, *Genome Res.*, **14**, 1188–90.
32. Urban, J. H. and Vogel, J. 2007, Translational control and target recognition by *Escherichia coli* small RNAs in vivo, *Nucleic Acids Res.*, **35**, 1018–37.
33. Hall, T. A. 1999, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucleic Acids Symp. Ser.*, **41**, 95–8.
34. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. and Barton, G. J. 2009, Jalview Version 2: a multiple sequence alignment editor and analysis workbench, *Bioinformatics*, **25**, 1189–91.