Molecular Plant Advance Access published December 3, 2007

Molecular Plant • Pages 1–12, 2007

A Novel Two Domain-Fusion Protein in Cyanobacteria with Similarity to the CAB/ELIP/HLIP Superfamily: Evolutionary Implications and Regulation

Oliver Kilian, Anne Soisig Steunou, Arthur R. Grossman and **Devaki Bhaya**[1]

Department of Plant Biology, The Carnegie Institution, Stanford, CA 94305, USA

ABSTRACT Vascular plants contain abundant, light-harvesting complexes in the thylakoid membrane that are non-covalently associated with chlorophylls and carotenoids. These light-harvesting chlorophyll *a/b* binding (LHC) proteins are members of an extended CAB/ELIP/HLIP superfamily of distantly related polypeptides, which have between one and four transmembrane helices (TMH). This superfamily includes the single TMH, high-light-inducible proteins (Hlips), found in cyanobacteria that are induced by various stress conditions, including high light, and are considered ancestral to the LHC proteins. The roles of, and evolutionary relationships between, these superfamily members are of particular interest, since they function in both light harvesting and photoprotection and may have evolved through tandem gene duplication and fusion events. We have investigated the Hlips (*hli* gene family) in the thermophilic unicellular cyanobacterium *Synechococcus* OS-B'. The five *hli* genes present on the genome of *Synechococcus* OS-B' are relatively similar, but transcript analyses indicate that there are different patterns of transcript accumulation when the cells are exposed to various growth conditions, suggesting that different Hlips may have specific functions. Hlip5 has an additional TMH at the N-terminus as a result of a novel fusion event. This additional TMH is very similar to a conserved hypothetical, single membrane-spanning polypeptide present in most cyanobacteria. The evolutionary significance of these results is discussed.

INTRODUCTION

Cyanobacteria, algae and vascular plants efficiently absorb light energy to drive oxygenic photosynthesis (Grossman et al., 1995). Cyanobacteria and red algae use a large, peripheral antenna called the phycobilisome (PBS) as their major light-harvesting complex (Grossman et al., 1993, 2001), while vascular plants and green algae use a light-harvesting complex comprising integral membrane polypeptides, with three transmembrane α helices (Helix B, Helix C and Helix A or TMH1, TMH2 and TMH3) (Dekker and Boekema, 2005; Jansson, 1994). These light-harvesting chlorophyll *a*/*b* binding (LHC) proteins (also called Chlorophyll a/b binding proteins or CAB) are very abundant in the cell, bind chlorophyll *a*, chlorophyll *b* and carotenoids, associate with photosystem (PS) I and II, and their synthesis is regulated by a number of different environmental parameters, including light and nutrients (Jansson, 1999). LHC-type polypeptides that bind chlorophyll *c* and fucoxanthin are also found in brown algae, diatoms, dinoflagellates, and chrysophytes (Grossman et al., 1995).

LHC polypeptides are well conserved, with a similar three-dimensional structure of internal two-fold symmetry in which TMH1 (Helix B) and TMH3 (Helix A) share sequence similarity. This arrangement led to the hypothesis that LHC polypeptides may have arisen by a gene duplication of an ancestral gene encoding a single TMH polypeptide (Kuhlbrandt et al., 1994). Highly conserved residues in TMH1 and TMH3 enable these domains to associate with each other via the formation of salt bridges and to bind chlorophyll molecules within the thylakoid membrane (Green, 1995; Kuhlbrandt et al., 1994). Early work on LHC proteins focused on structural characterization, gene family identification and regulation, pigment-binding features and integration, and assembly of individual polypeptides into light-harvesting complexes (see recent reviews in Jansson, 1999; Koziol et al., 2007; Lucinski and Jackowski, 2006). However, recent findings have indicated that LHC proteins are part of a large CAB/ELIP/HLIP superfamily,

[1] To whom correspondence should be addressed. E-mail dbhaya@stanford.edu, fax 650 325 6857, tel. 650 325 1521 ×282.

doi: 10.1093/mp/ssm019

with distantly related members having varying numbers of TMHs and different functions (Montane and Kloppstech, 2000).

The identification of the *hli* gene in *Synechococcus* sp. strain PCC7942, encoding a small, single TMH polypeptide, designated Hlip, with homology to TMH1/TMH3 of LHC proteins, provided the first evidence of a putative progenitor of the LHC family in cyanobacteria (Dolganov et al., 1995; Green and Pichersky, 1994). Although the role of Hlips is not entirely understood, based on expression and biochemical analyses, it is likely that they function to protect the photosynthetic apparatus from excess absorbed excitation energy, directly or indirectly (Havaux et al., 2003; He et al., 2001). Hlips may transiently bind chlorophyll (Funk and Vermaas, 1999) and/or carotenoids or regulate tetrapyrrole biosynthesis (Xu et al., 2002). Recently, one-helix proteins (OHPs) with similarity to Hlips and TMH1 of LHC proteins have also been identified in *Arabidopsis*. They are predicted to be targeted to the chloroplasts, although their function has yet to be established experimentally (Andersson et al., 2003; Jansson et al., 2000).

Members of the extended CAB/ELIP/HLIP superfamily in vascular plants contain between one and four TMHs and often play a role in photoprotection. Indeed, it has been argued that photoprotection rather than light harvesting may have been the original role of ancestral members of the CAB/ELIP/HLIP superfamily (Montane and Kloppstech, 2000). For instance, the early light induced proteins (ELIPs) have three TMHs and are transiently expressed during greening of etiolated tissue (Adamska, 1997; Adamska et al., 1992) or when vascular plants are exposed to various stressors (Adamska, 1997; Montane et al., 1997). ELIPs are thought to protect the photosynthetic apparatus from oxidative damage during thylakoid membrane development by acting as pigment-carrier proteins (Adamska, 1997), or by influencing chlorophyll biosynthesis (Tzvetkova-Chevolleau et al., 2007). PSBS—a well conserved PSII-associated membrane protein—has four TMHs, with two that have similarity to TMH1/TMH3 of LHC. PSBS plays a critical role in protecting the photosynthetic apparatus from photo-damage by facilitating nonphotochemical quenching (NPQ) (Green, 1995; Jansson, 1999; Li et al., 2000; Niyogi, 1999). *Arabidopsis thaliana* also contains SEP-1 and SEP-2 (stress-enhanced proteins) and their transcripts are increased under conditions of high light stress. SEPs have two TMHs, of which the first resembles TMH1/TMH3 of the LHCs, while the second TMH has no significant homology to other known proteins (Heddad and Adamska, 2000).

We have examined the Hlip family of polypeptides in two recently sequenced *Synechococcus* isolates from the hot spring microbial mats at Octopus Spring, Yellowstone National Park (Bhaya et al., 2007). The two hot spring isolates, *Synechococcus* OS-A (NC_007775) and *Synechococcus* OS-B′ (NC_007776), are closely related to each other but dominate at different temperatures within the effluent channels, with *Synechococcus* OS-A being more abundant at higher temperatures (55–60°C) (Allewalt et al., 2006; Ramsing et al., 2000; Ward et al., 1998). In the temperature range 50–55°C, *Synechococcus* OS-B′ is abundant, and is present at all vertical depths within the cyanobacterial layer (~1.5 mm) of the mat, even in the top 0.2 mm of the mat, where light irradiances can be >1000 μmol photons m^{-2} s^{-1}. *Synechococcus* OS-B′ grown under defined laboratory conditions is unable to withstand moderately high light for prolonged periods of time, motivating our interest in examining the regulation of the *hli* gene family under light stress conditions (Kilian et al., 2007). We also found that one of the *hli* genes present in both *Synechococcus* OS-A and *Synechococcus* OS-B′ represents a novel fusion between a Hlip domain and an additional TMH which is very similar to a small, conserved hypothetical polypeptide of unknown function which is present in most freshwater cyanobacteria.

MATERIALS AND METHODS

Cultivation and Light-Shift Experiments

Synechococcus OS-B′ (*Synechococcus* sp. JA-2–3B′a (2–13) Genbank NC_007776) was grown in DH-10 medium (i.e. Medium D, containing 10 mM HEPES pH 8.2) supplemented with Va vitamin solution used at 1000 dilution (Davis and Guillard, 1958; Kilian et al., 2007). The cultures were maintained at 52.5°C and 75 μmol photons m^{-2} s^{-1} intensity fluorescent light ('Plant and aquarium' lights from Philips Lighting Company, Somerset, NJ) and bubbled with air enriched with 3% CO_2. For most experiments, cultures were grown to mid-logarithmic phase (OD_{750nm} ~ 0.5) before transferring them to the dark or to different light conditions. Growth at different photon fluences was achieved by positioning the culture flasks at different distances from the light source. Photon fluences were measured using a hand-held quantum photometer (Model Li185B) equipped with a spherical quantum sensor (LI-193), (LiCOR, Lincoln, NJ). For the high-light experiments (Figure 4A), cultures were maintained for 16 h in the dark and then exposed to 200 μmol photons m^{-2} s^{-1} for various times. Aliquots of cells for transcript analysis were collected immediately before the light shift, and then 5, 10, 15, and 60 min following the light shift. To examine changes in the transcript levels in the dark (Figure 4B), *Synechococcus* OS-B′ grown at 200 μmol photons m^{-2} s^{-1} light intensity to mid-logarithmic phase, were transferred to darkness, and cell aliquots removed at various times thereafter. Aliquots, removed at specific time intervals following the transfer (indicated in figures), were rapidly cooled to ~4°C by swirling the flasks in liquid nitrogen. Cells were then collected by centrifugation at 3000 *g* at 4°C for 10 min, frozen in liquid nitrogen, and stored at –80°C until use.

Synechocystis PCC6803 was cultivated as described earlier (Bhaya et al., 1999). For high-light treatment, the *Synechocystis* PCC6803 cells were grown for 16 h in the dark before transfer to 850 μmol photons m^{-2} s^{-1}. The high-light source was a reflector lamp installed behind a water-filled glass chamber, which prevented excessive heating of the cultures.

Isolation of RNA, Reverse Transcription, Quantitative PCR (qPCR), and Northern Hybridization

Whole-cell RNA was isolated, purified, and stored at –80°C (Steunou et al., 2006). The RNA used for both reverse transcription PCR (RT-PCR) and quantitative PCR (qPCR) was subjected to DNAse digestion (TurboDNAse, Ambion) and then purified by phenol/chloroform extraction followed by ethanol precipitation. One μl of the RNA sample (between 0.5 and 2 μg RNA) was used as a template in a 50 μl PCR reaction employing specific primers targeted against the *coxA* (CYB_2698) gene (2 min at 94°C, 30 cycles of 1 min at 55, 72, and 94°C; 5 min at 55°C; 5 min at 72°C) to verify that no contaminating genomic DNA remained. For RT-PCR experiments (Figure 5B), further controls were included—a negative control in which reverse transcriptase was excluded from the reaction mix and a positive control in which *Synechococcus* OS-B′ genomic DNA served as the template rather than RNA for the PCR amplification.

Reverse transcriptase (RT) (Superscript III, Invitrogen, Carlsbad, CA) was used to reverse-transcribe 4 μg RNA with 200 ng random primer hexamers (MBI Fermentas, Hanover, MD) in a reaction volume of 20 μl according to the instructions of the manufacturer. When specific reverse primers were used (Table 1 and Figure 5B), we followed the same protocol as previously described (Steunou et al., 2006). A PTC PCR cycler (MJ Research, Inc., Waltham, MA) was used for the reverse-transcription step, under the following conditions: 5 min at room temperature, 5 min at 37°C, 10 min at 45°C, 40 min at 50°C, 20 min at 55°C, followed by 15 min at 75°C for heat inactivation. The MassRuler DNA ladder, low range (Fermentas Cat #SM0383) was used (Figure 5B).

The cDNA was diluted 10-fold in RNAse/DNAse free water and 2.5 μl of the diluted cDNA template were used in a 20 μl qPCR reaction, as described before (Steunou et al., 2006). Melting curves of the PCR products were checked for the presence of multiple/nonspecific products. RNA levels were calculated and normalized utilizing a standard series employing 10-fold dilutions of a specific PCR product as a template. All measurements were carried out in triplicate and standard deviations calculated. The primers used for qPCR on ssl2148 from *Synechocystis* PCC6803 were *ssl2148-F*: TTCCCACCTACTTTGTAGCGGTCT and *ssl2148-R*: TGGAATTGTACCAGGCCACCAAAC.

Northern blot hybridization was carried out with RNA (5 μg per lane) that was denatured for 10 min at 70°C in 2 loading dye solution buffer from Fermentas (Cat #SM1833 (RNA V-buffer V)). RNA was run on a 1.6% (wt/vol) agarose-formaldehyde gel alongside RiboRuler RNA ladder, low range (Fermentas Cat #SM1833). Blots (on Hybond N membrane) were hybridized with the 0.2-kb *hli5* probe obtained by PCR with 16F and 16R primers (Table 1) and labeled with [α-^{32}P] dCTP using Prime—a gene-labeling system kit from Promega (Cat # U1100). After hybridization, the membrane was washed under standard conditions and exposed to a Phosphorimager.

Comparative Analysis of *hli* Genes

The CLUSTALW program (Chenna et al., 2003) was used to generate alignments of the Hlips from *Synechococcus* OS-A, *Synechococcus* OS-B′, and *Synechocystis* PCC6803 (Figure 2A). To generate the comparison between the Hlips shown in Table 2, only the highly conserved regions of the proteins were used, i.e. starting at the sequence GFxxxAERKNGRLAMIGF until the end of the ORF (amino acids 64–108 in Figure 2A), since the N-termini are not conserved between different members of the protein family (even among those that appear to be very closely related, such as HliC and HliD).

Hlip5 (CYB_1999) is annotated as 122 amino acids long compared with its homolog Hlip5 (CYA_0389) in *Synechococcus* OS-A, which is 102 amino acids long. The second in-frame Met codon, which is 20 codons downstream of the first Met in Hlip5, is most likely the correct start site codon. This would make the ORF of CYB_1999 the same size as CYA_0389; moreover, there is no additional Met codon upstream of the predicted reading frame of CYA_0389. Second, the nucleotide identity of the *hli5* genes of *Synechococcus* OS-B′ and *Synechococcus* OS-A is very high from this second Met codon until the end of the ORF. Conversely, there is very little nucleotide

Table 1. Genes in the Hlip Family of *Synechococcus* OS-B′, OS-A and *Synechocystis* sp. Strain PCC6803, Locus IDs and Primer Sequences Used in this Study.

Gene	Size (AA)	Locus ID	Primer	Sequence	Homolog in CYA	Closest homolog in Syn 6803
hli1	73	CYB_1447	13F	TCGGCTGAACAACTACGCCATT	CYA_0142	hliA (scpC)/ hliB (scpD)
			13R	GGTGGTACCAGACAGAGCTTCAAA		ssl2542/ssr2595
hli2	50	CYB_1247	14F	TGTCTAATCCCAATTTGTCTGAGCCC	CYA_0205	hliD (scpB)
			14R	TCAGATCAGGCCCAGCCAGGA		ssr1789
hli3	56	CYB_0506	15F	ATGTCACCCGTTGATGGAGCTA	CYA_2834	hliC (scpE)
			15R	ACATCAGGCCGATTTGAGAGAGGA		ssl1633
hli4	72	CYB_1998	60F	TGCAATCGTCTCCCAGAATCCCTT	CYA_0390	None
			60R	CAATGCTGATCACCAAGCCCAACA		
hli5	102	CYB_1999	16F	TGTTGTGATTGGCTCGATGGCT	CYA_0389	None
			16R	ACCACAACACTGATGGCAAACC		

sequence homology between *hli5* from *Synechococcus* OS-B′ and *hli5* from *Synechococcus* OS-A upstream of this position.

The GeneBee program (www.genebee.msu.su/genebee.html) was used to predict potential stem-loop structures in RNA derived from the intergenic region between *hli4* and *hli5* (Brodsky et al., 1995). HMM-based TMH prediction algorithms such as SOSUI, TMHMM, PHD, and TOP-PRED were used to predict the potential to form an α-helical TMH and membrane topology (www.predictprotein.org/) (Punta et al., 2007; Rost et al., 2004).

RESULTS

All cyanobacteria examined to date have multiple *hli* genes (Bhaya et al., 2002; He et al., 2001; Steglich et al., 2006), but they have also been identified on the chloroplast genomes of the red algae *Porphyra yezoensis* (HYP_537036.1) Hand *Cyanidium caldarium* (Q9TM07) (Glockner et al., 2000), the glaucophyte (*Cyanophora paradoxa*) (NP_043235), and on the nuclear genome of *Arabidopsis* (Andersson et al., 2003; Jansson et al., 2000). We searched the complete genome sequences of the two thermophilic, unicellular cyanobacteria—*Synechococcus* OS-A and *Synechococcus* OS-B′—for the presence of *hli*-like genes (Bhaya et al., 2007). Five *hli* gene sequences were identified (designated *hli1* to *hli5*) on the genomes of both *Synechococcus* OS-A and *Synechococcus* OS-B′.

The *hli1*, *hli2*, and *hli3* genes are not arranged within operons and are not clustered on the genomes, although the genes flanking *hli1*, *hli2*, and *hli3* are syntenic, to some degree, between *Synechococcus* OS-A and *Synechococcus* OS-B′ (data not shown; gene neighborhoods for genomes of choice can be generated at Integrated Microbial Genomes at http://img.jgi.doe.gov/) (Markowitz et al., 2006). In contrast, *hli4* and *hli5* are tandemly arranged on both genomes (Figure 1), and the genes flanking *hli4* and *hli5* have limited synteny between the two organisms. The intergenic distance between *hli4* and *hli5* open reading frames is 110 and 140 bp in *Synechococcus* OS-A and OS-B′, respectively, and there is low nucleotide identity (40%) between these intergenic sequences, consistent with the general finding that non-coding sequences are often quite divergent.

Hlips 1–4 (encoded by *hli1*, *hli2*, *hli3*, and *hli4*, respectively) range in size from 50 to 73 amino acids. They all contain a short N-terminus region followed by a predicted TMH (Figure 2A). In contrast, Hlip5 (CYB_1999 in *Synechococcus* OS-B′ and CYA_0389 in *Synechococcus* OS-A) is 102 amino acids long. It has an N-terminus extension of ~70 amino acids, in addition to the Hlip-like TMH (~35 amino acids) at the C-terminus. This N-terminus extension is highly conserved between *Synechococcus* OS-A and *Synechococcus* OS-B′ and has the potential to form an α-helical TMH (here referred to as TMH1) (Figure 1 and 2A), based on TMH prediction algorithms (Punta et al., 2007; Rost et al., 2004). Thus, Hlip5 is unique in containing *two* potential TMHs, unlike all other Hlips on the cyanobacterial genome. The TMH prediction algorithms suggest that TMH1 and TMH2 are a minimum of 23 amino acids long, but comparison of them to the TMH1 and TMH3 of the LHC crystal structure suggests that they are likely to be ~35 amino acids (Green, 1995; Kuhlbrandt et al., 1994).

We carried out a comparison of amino acid identity within the region conserved in Hlips (amino acids 64–108 in Figure 2A) between *Synechococcus* OS-A and OS-B′ and found that each Hlip is very similar to its putative ortholog, but not nearly as similar to other Hlips (Table 2). Thus, Hlip1 (73 amino acids), Hlip2 (50 amino acids), and Hlip3 (56 amino acids) are 97, 100, and 97% identical between putative orthologs, but share only 41–53% identity with other Hlips encoded on the genome (Table 2). On the other hand, the putative orthologs of Hlip4 and Hlip5 share only 86 and 80% identity, respectively, between *Synechococcus* OS-A and OS-B′. Furthermore, Hlip4 and Hlip5 are more similar to each other (~61–63% identity) than to any other Hlip (~39–47% identity) (Table 2). We also compared the gene products of *hli1-5* of *Synechococcus* OS-A and *Synechococcus* OS-B′ with those from *Synechocystis* PCC6803 (*hliA*, *hliB*, *hliC*, and *hliD*), for which expression data as well as phenotypic information are available (Figure 2A and Table 2). Based on CLUSTALW alignments (Chenna et al., 2003), it appears that Hlip1 from *Synechococcus* OS-A and OS-B′ is

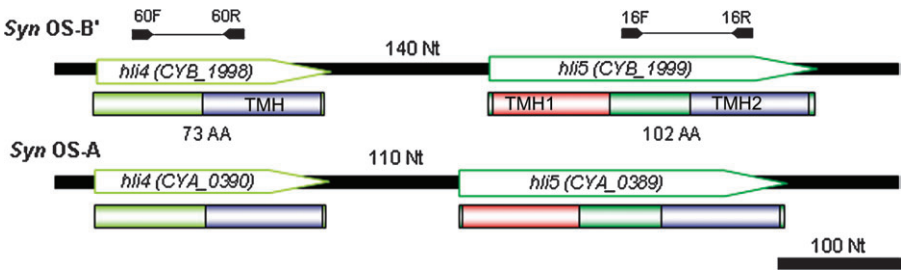

Figure 1. Tandem Arrangement of *hli4* and *hli5* on the Genomes of *Synechococcus* OS-B′ and *Synechococcus* OS-A.

Genes are shown as open arrows with gene designations within brackets. The encoded proteins are shown as filled bars below the genes. The relative positions of the transmembrane helices in Hlip4 (TMH) (purple box) and Hlip5 (TMH1 and TMH2) (red and purple boxes, respectively) are shown. The position of primers used to amplify the genes in *Synechococcus* OS-B′ by PCR is shown in line 1 as filled back arrows. Nt, nucleotides.

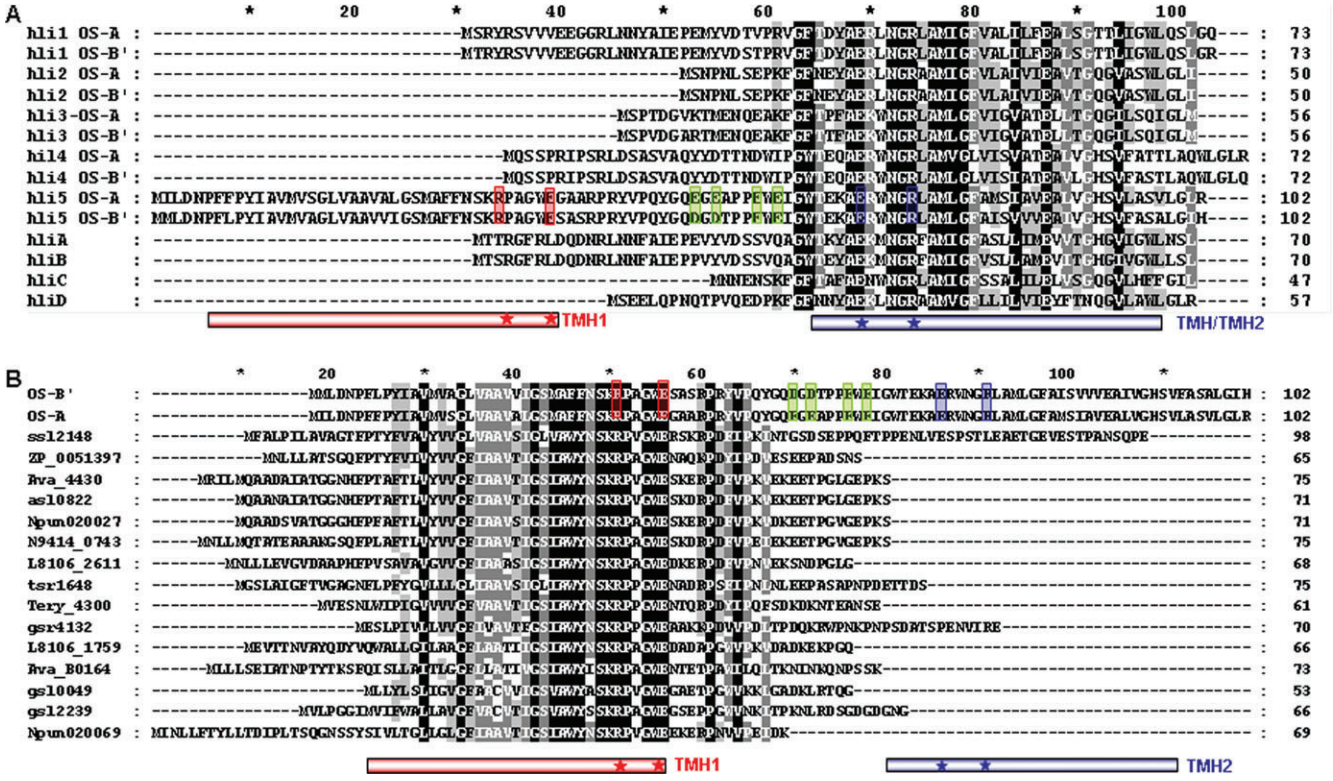

Figure 2. ClustalW alignment of HLIPs.

(A) ClustalW alignment of HLIPs from *Synechococcus* OS-B′ and OS-A and *Synechococcus* PCC6803 are shown. Gene abbreviations used are: *hli1* OS-A to *hli5* OS-A for Hlip1 to 5 of *Synechococcus* OS-A; *hli1* OS-B′ to *hli5* OS- B′ for Hlip1 to 5 of *Synechococcus* OS-B′; hliA, hliB, hliC and hliD for the Hli proteins of *Synechocystis* PCC6803.

(B) ClustalW alignment of HLIPs from*Synechococcus* OS-B′ and OS-A with Coh1 found in other cyanobacteria. See Table 1 and text for Genbank Accession numbers. CY0110_10942 (*Cyanothece* sp. CCY0110) has not been included in this analysis.

Black, dark gray and light gray boxes indicate completely conserved (100%), highly conserved (>80%) or moderately conserved (>60%) residues, respectively. The position of TMH1 (of Hlip5) is shown as a red bar and TMH (of Hlip1–4)/TMH2 (of Hlip5) are shown as a purple bar. The position of the conserved E and R residues in TMH1 and TMH2 of Hlip5 are boxed in red and blue, respectively, in the alignments. They are also marked by stars in the TMH bars below the alignments. The glutamate(E)/aspartate (D) residues in the loop between TMH1 and TMH2 of Hlip5 are boxed in green (residues #53, 55, 59, and 61 in lines 9 and 10 of Figure 2A.

most similar to HliA and HliB; Hlip2 is most similar to HliD, and Hlip3 is most similar to HliC, while Hlip4 and Hlip5 do not appear to have an obvious homolog in *Synechocystis* PCC6803 (Table 2). In general, the identities between the Hlips of *Synechocystis* PCC6803 and the two *Synechococcus* strains are not very high.

Interestingly, comparison of the TMH1 region of Hlip5 with ORFs encoded on cyanobacterial genomes revealed that it has high similarity to a 'hypothetical' or 'conserved hypothetical protein' encoded on cyanobacterial genomes (Figure 2B). This gene, which we will refer to as *coh1* (cyanobacterial one helix), has been identified as a hypothetical protein in most freshwater/brackish water cyanobacteria, but is absent in the genomes of the marine *Synechococcus* sp. or *Prochlorococcus* sp. (of which 18 genomes are either completed or draft versions are available) (www.genomesonline.org/). It is also not present on the genomes of *Synechococcus* OS-A and *Synechococcus* OS-B′. It is found in single copy on the genomes of *Synechocystis* PCC6803 (ssl2148), *Crocosphaera watsoni* (ZP_00051397), *Nodularia spumigena* (N9414_0743), *Trichodesmium erythraeum* (Tery_4300), *Nostoc* sp. PCC7120 (asl0822), *Thermosynechococcus elongatus* (tsr1648), and *Cyanothece* sp. CCY0110 (CY0110_10942); there are two copies on the genomes of *Nostoc punctiforme* (Npun02002710 and Npun02006958), *Anabaena variabilis* (Ava_4430 and Ava_B0164), and *Lyngbya* sp. PCC8106 (L8106_2611 and L8106_1759) and three copies in *Gloeobacter violaceus* PCC7421 (gsr4132, gsl0049, and gsl2239).

The conserved *coh1*-encoded proteins range in size from 53 to 98 amino acids, with a short N-terminus region followed by a highly conserved TMH, mostly comprising small hydrophobic residues (Figure 2B). Within the predicted TMH are two highly conserved glycine (G) residues followed by a well conserved eight-residue region (SK**R**P{V/A}GW**E**). Based on membrane prediction programs, it is unclear whether the GWE motif is in or just outside the membrane. This region is followed by a four-residue variable region, and two well conserved proline (P) residues spaced three residues apart. In the case of Hlip5, *both* TMH1 and TMH2 have conserved glutamate (E) and

Table 2. Comparison of Percent Amino Acid Identities of Hlips.

	hli1_B′	hli2_A	Hli2_B′	hli3_A	hli3_B′	hli4_B′	hil4_A	hli5_B′	hli5_A	hliA	hliB	hliC	hliD
hli1_ A	97%	53%	53%	41%	41%	38%	38%	39%	46%	53%	53%	43%	39%
hli1_ B′		53%	53%	41%	41%	38%	38%	39%	46%	53%	53%	43%	39%
hli2_ A			100%	51%	51%	40%	36%	45%	42%	51%	51%	51%	69%
hli2_ B′				51%	51%	40%	36%	45%	42%	51%	51%	51%	69%
hli3_ A					97%	43%	45%	45%	47%	41%	48%	56%	48%
hli3_B′						43%	45%	45%	47%	41%	48%	56%	48%
hli4_B′							86%	63%	63%	36%	40%	36%	27%
hil4_A								63%	61%	34%	40%	34%	29%
hli5_B′									80%	42%	45%	40%	35%
hli5_A										45%	47%	45%	35%
hliA											82%	53%	48%
hliB												46%	46%
hliC													46%

The conserved regions/positions (amino acids 64–108 in Figure 2A) of Hlips of *Synechococcus* OS-B′ (hli1_B′ to hli5_B′), *Synechococcus* OS-A (hli1_A to hli5_A), and *Synechocystis* PCC6803 (hliA, hliB, hliC and hliD) were compared. Black boxes mark Hlips that have the highest identity between *Synechococcus* OS-B′ and *Synechococcus* OS-A. Grey boxes mark the highest identity between the Hlips of *Synechococcus* and those in *Synechocystis* PCC6803.

arginine (R) residues, spaced four residues apart, reminiscent of the arrangement of conserved functional E and R residues in Helix B (or TMH1) and A (or TMH3) of LHC proteins. These residues are marked by red and blue boxes in the alignments (and as stars in the helix bars below the alignments) (lines 9 and 10 of Figure 2A, lines 1 and 2 of Figure 2B). It is worth noting that there are four glutamate (E) residues in the predicted loop between TMH1 and TMH2 of Hlip5 of *Synechococcus* OS-A (in *Synechococcus* OS-B′, two of these residues are replaced by aspartate (D)) (Figure 2, green boxes).

Next, we attempted to align TMH1 from Hlip5 of *Synechococcus* OS-A and OS-B′ with the second TMH (Helix C; TMH2) of members of the CAB/ELIP/Hlip superfamily (Figure 3). However, it has been observed that Helix C is the least conserved of the three helices in the LHCs and, in most cases, this helix has few, if any, conserved residues among the distantly related members of the family (Green, 1995; Heddad and Adamska, 2000). The best alignments were achieved with particular TMH2 sequences of LHCs, so representative sequences from vascular plants and other lineages were manually aligned. We used sequences from the CP24 precursor of *Fragaria ananassa*, LHCII (LHCb1) from *Pisum sativum*, for which the 3.4-Å resolution crystal structure is also available (mmdbId:39082) (Kuhlbrandt et al., 1994), LHCb6 from *Arabidopsis thaliana*, LHC helix 2 (isoform 11) from the haptophyte *Isochrysis galbana*, and a 'chlorophyll *a-b* binding protein' from the green alga *Dunaliella* (Figure 3). We observed that although the sequence similarity is quite low within these TMHs, they predominantly comprise small hydrophobic residues that may assist in tighter packing of helices in the membrane. Shown below the alignment is the position of the predicted Helix C based on the crystal structure of LHCII (LHCb1) from *P. sativum* (Kuhlbrandt et al., 1994). The residues that lie on the face of Helix C, which would be in contact with Helix A and B, are indicated by filled black circles and

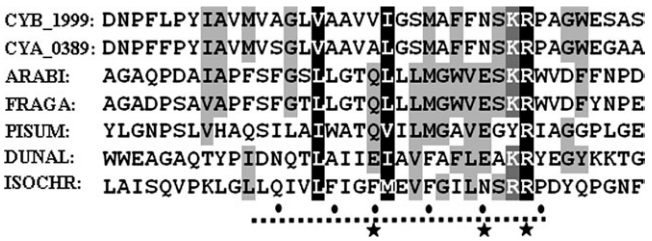

Figure 3. Alignment of TMH1 of *Synechococcus* OS-A and OS-B′Hlip5 with Selected TMH Domains.

TMH1 of Hlip5 of *Synechococcus* OS-A and OS-B′ aligned manually with selected Helix C (TMH2) domains of *Fragaria ananassa* (gi:60549177), *Pisum sativum* (gi:115788), *Arabidopsis thaliana* (gi:4741960), *Isochrysis galbana* (gi:77024113), and *Dunaliella salina* (gi:62638125). Based on the crystal structure of LHCII (Lhcb1) from *Pisum sativum*, the location of TMH2 (Helix C) is indicated by a dotted bar underneath the alignment; conserved residues that are part of *chl a/b* binding fold of pfam PF00504 and are relatively well conserved in all TMHs are shown by small filled black circles; star symbols mark residues involved in chlorophyll binding.

note that several are shared with TMH1 of Hlip5 in the alignments shown here (also seen in the Pfam alignment PF00504). The positions of residues glutamine (Q), glutamate (E), and arginine (R) involved in chlorophyll binding within LHC are indicated by stars; the R residue is conserved in all the sequences shown.

To determine the relationship between *hli* expression and environmental stress conditions, we measured transcript levels of the five *hli* genes of *Synechococcus* OS-B′ (Figure 4A). In initial experiments, cells growing at moderate light levels (75 μmol photons m^{-2} s^{-1}) were kept in the dark for 16 h and subsequently transferred to high light (200 μmol photons m^{-2} s^{-1}) and then harvested at various times following the transfer. Total RNA was isolated and the abundance of specific

transcripts analyzed by qPCR. For this experiment, transcript levels were normalized to the absolute numbers of transcripts, thus reflecting the relative abundance of the different *hli* transcripts to each other at every time point (see 'Materials and Methods' and Figure 4A legend for details). All of the *hli* transcripts increased within 5 min of the transfer and appeared to peak within 10 min (Figure 4A). By 15 min, all transcripts had decreased and, by 1 h, had declined substantially. Transcript levels returned to dark levels after ~24 h (data not shown). It is noticeable that all *hli* genes showed similar kinetics over the time course. Comparing absolute numbers of transcripts, the abundance of *hli3* is much higher than those of the other *hli* transcripts; the *hli1* transcript abundance is about 40% lower and the *hli2*, *hli4*, and *hli5* transcripts are even lower (Figure 4A). In Figure 4B, we show that if cells are grown in high light (200 μmol photons m^{-2} s^{-1}) to mid-logarithmic phase and then transferred to the dark, transcript levels drop almost immediately and reach a minimum after 10 min. By the end of the time course, all the *hli* transcripts had stabilized or increased slightly, but were still low when compared with the time-zero value, suggesting that the *hli* transcripts may be regulated at the level of both synthesis and degradation.

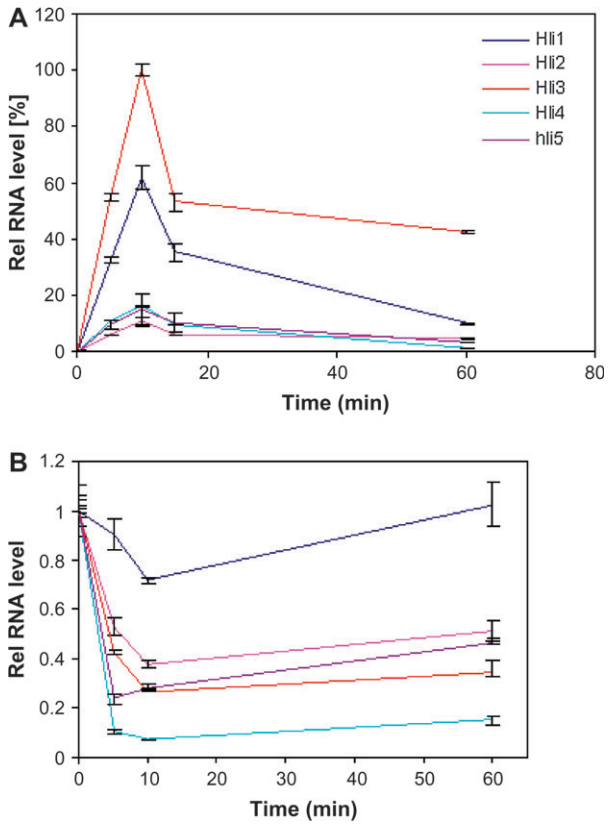

Figure 4. Transcript levels of *hli* genes (*hli1–hli5*).
(A) *hli* transcript levels after transfer to high light. Cells were kept in the dark for 16 h and then transferred to high light (200 μmol photons m^{-2} s^{-1}) and aliquots were taken after 5, 10, 15, and 60 min for RNA extraction. Gene-specific primers were used for qPCR analysis (see 'Material and Methods' for details). Absolute RNA levels were calculated based on a standard series using defined amounts of the respective PCR product for calibration; the highest RNA level which was measured (i.e. *hli3*, at 10 min, based on the qPCR product quantification) was set at 100%.
(B) *hli* transcript levels after transfer to dark. Cells grown in high light (200 μmol photons m^{-2} s^{-1}) to mid-logarithmic phase were transferred to darkness and aliquots were taken for RNA extraction at 0, 5, 10, and 60 min after transfer. The relative RNA levels for each of the *hli* genes was calculated by setting the RNA level at the 0 time point (i.e. immediately before the light shift) arbitrarily to 100% for all transcripts.

The steady-state levels of all five *hli* transcripts were also measured (Figure 5) in cells maintained at different light intensities. The cells were placed in the dark for 16 h or grown to mid-logarithmic phase at photon fluence rates of 25, 75, or 200 μmol photons m^{-2} s^{-1}. These were considered to be low-light (LL), medium-light (ML), or high-light (HL) conditions, respectively, based on our earlier observations with respect to growth rates of *Synechococcus* OS-B' under different photon fluence levels (Kilian et al., 2007). RNA was extracted for measuring transcript abundance, with the levels present in dark-acclimated cultures set arbitrarily to 1. All five *hli* transcripts were most abundant at the highest light intensity (200 μmol photon m^{-2} s^{-1}) used, although the pattern of accumulation as a function of photon fluence level appears to be different for each of the *hli* transcripts. Thus, *hli1* appears to accumulate to about 10-fold of the dark level at all of the three light fluences measured. The *hli2* transcript accumulates at higher levels as a function of light fluence; the transcript levels are 15, 25, and 45-fold relative to the dark level at LL, ML, and HL, respectively. *hli3* and *hli4* transcript levels are low in the dark and in LL and ML, but increased five and 15-fold, respectively, in HL. Thus, it is possible that the Hlip3 and Hlip4 are specifically required for acclimation to high light. The transcript accumulation for *hli5* shows somewhat different steady-state levels; when cells are grown in either LL or ML,

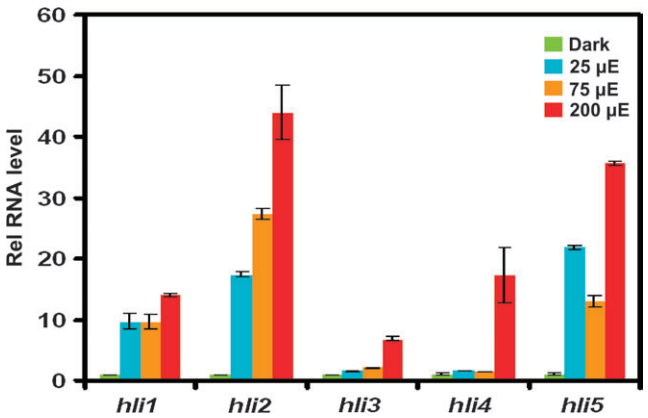

Figure 5. Steady-State Levels of the *hli* Transcripts (*hli1–hli5*). Steady-state levels of the *hli* transcripts (*hli1–hli5*) were measured in cultures kept in the dark for 16 h or grown to mid-logarithmic phase at different light intensities. RNA values are relative to the respective levels of individual *hli* genes in the dark-acclimated cells, set arbitrarily to 1. Specific primers, as described in Table 1, were used for each *hli* gene.

the *hli5* transcript is ~15–20-fold higher than in the dark and increased further in HL (to ~33-fold of the dark level).

The *hli4* and *hli5* genes are tandemly arranged on the genome so a single promoter upstream of *hli4* could drive the synthesis of a transcript spanning both *hli4* and *hli5*. To investigate this possibility, we performed RT-PCR, using RNA from cells maintained in the dark (lanes 2, 5, 7, 10, 12, and 15) or exposed to HL for 15 min (lanes 1, 4, 6, 9, 11, and 14) (Figure 6A). In all cases shown, the *hli5*-specific reverse primer (16R primer, Figure 1 and Table 1) was used to prime the RT-mediated reaction to investigate whether a single long transcript was synthesized for *hli4* and *hli5.* Following the RT reaction, three sets of PCR reactions were carried out. We used *hli4*- (60F and 60R, lanes 1–5) or *hli5*- (16F and 16R, lanes 6–10) specific primers to quantify *hli4* or *hli5* transcript levels. We also used the *hli4* forward and *hli5* reverse primer (60F and 16R, lanes 11–15) to determine whether there was a longer transcript that covered both *hli4* and *hli5*. In all cases, we included a 'no RT added' negative control (lanes 1, 2, 6, 7, 11, and 12) to verify that DNA contamination of the RNA samples was below detectable levels, as well as a positive control using genomic DNA as a template (lanes 3, 8, and 13). As shown in Figure 6A, it appears that for all three primer sets used, transcript abundance was significantly higher in HL-treated samples (lanes 4, 9, and 14) than in the dark samples (lanes 5, 10, and 15), which is consistent with the qPCR data shown in Figure 4. We reproducibly observed a low but detectable level of the *hli5* transcript in the dark (lane 10). This is also consistent with the data presented in Figure 4, which suggests that the *hli5* transcript (unlike the *hli4* transcript) is present in the dark and at very low light intensities. Interestingly, a strong signal was observed using the primer pair (60F and 16R) that partially covers *hli4* and *hli5* and the intergenic region between them, but it was only observed using RNA from cells exposed to high light (lane 14). Thus, there appears to be a transcript that covers *hli4* and *hli5*.

We carried out Northern blot hybridizations to confirm these results (Figure 6B) using RNA isolated from cells kept in the dark for 16 h and subsequently transferred to high light (200 μmol photons m^{-2} s^{-1}) and then harvested at various times, following the transfer (as described for Figure 4A). Using a *hli5*-specific probe, we verified the presence of an abundant, short ~0.3-kb transcript and a longer, less abundant ~0.75-kb transcript. Transcript abundance was highest between 5 and 15 min after transfer to high light, but was much reduced at 60 min, mirroring the data derived from qPCR analysis (Figure 4A). However, we did not detect a transcript in the 'dark' lane, although, using RT-PCR, a faint product was observed (Figure 6A, lane 10). We ascribe this discrepancy to the fact that RT-PCR is a considerably more sensitive assay than Northern hybridization. We estimated that there was ~10-fold more of the short ~0.3-kb *hli5*-specific transcript than of the

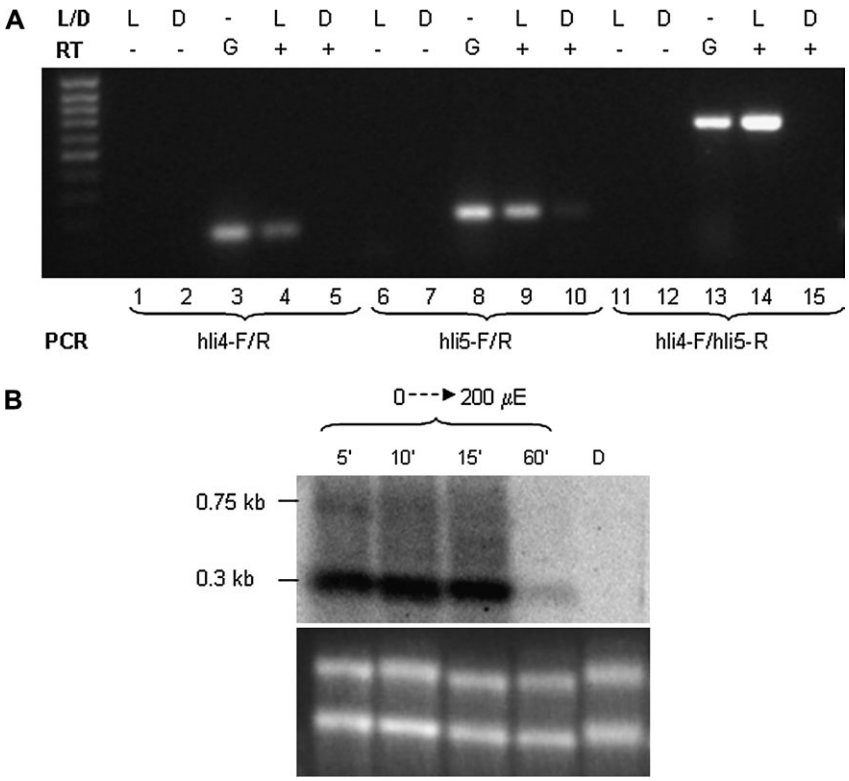

Figure 6. Transcript levels of *hli4* and *hli5*. **(A)** Transcript levels measured using RT-PCR. RT-PCR measurement of transcript levels of *hli4* and *hli5* from *Synechococcus* OS-B' cells grown in the dark (D) or transferred to light (L) (200 μmol photons m^{-2} s^{-1}) for 15 min. In all cases, the reverse primer for *hli5* (16R, Table 1) was used for the RT reaction. Lanes 1–5, PCR with *hli4*-specific primers (expected size 0.15 kb). Lanes 6–10, PCR with *hli5*-specific primers (expected size 0.195 kb). Lanes 11–15, PCR with the forward primer for *hli4* (60F) and the reverse primer for *hli5* (16R) (expected size 0.6 kb). For each set of primers, genomic DNA (G) extracted from *Synechococcus* OS-B' was used as a positive control. Negative control (–) (i.e. no reverse transcriptase added) for RNA extracted from the light (L) and the dark (D) samples (Lanes 1, 2, 6, 7, 11, and 12). Leftmost lane: molecular markers. **(B)** Top: Transcript levels measured using Northern hybridization. RNA was extracted from cells kept in the dark for 16 h (lane 5) and then transferred to high light (200 μmol photons m^{-2} s^{-1}) for 5 min (lane 1), 10 min (lane 2), 15 min (lane 3), or 60 min (lane 4). Bottom: RNA gel (showing the major rRNA bands) which was used for blotting, showing that all lanes were equally loaded.

longer ~0.75-kb transcript that covers *hli4–hli5.* Similar results were obtained using a *hli4*-specific probe (data not shown). From the data acquired using RT-PCR, qPCR, and Northern analysis, it appears that there are abundant short transcripts that specifically cover *hli4* and *hli5*. In addition, there is a longer, less abundant transcript that covers both *hli4* and *hli5*. Unlike the *hli4* transcript, there is a low level of the short *hli5* transcript present in the dark (Figure 6A).

To test whether ssl2148 (the homolog of TMH-1 of Hlip5) in *Synechocystis* PCC6803 was also up-regulated in high light, we monitored the level of the transcript following exposure of cells to high light. Dark acclimated cells (16 h dark) were exposed to high light (850 μmol photons m^{-2} s^{-1}) and collected for RNA extraction at exposure times of 5, 10, 15, and 60 min. qPCR analysis of the extracted RNA revealed a moderate increase (2.5-fold) in the level of the *ssl2148* transcript (data not shown).

DISCUSSION

We have investigated the small *hli* gene family in two closely related, thermophilic cyanobacteria—*Synechococcus* OS-A and *Synechococcus* OS-B′ (Bhaya et al., 2007). Of the five genes that comprise this family, *hli4* and *hli5* are tandemly arranged on the genome of both organisms. The arrangement of homologous genes with a relatively short intergenic region is often taken as evidence of a recent duplication event (Koonin, 2005). TMH2 of Hlip5 and the TMH of Hlip4 are more similar to each other than to Hlip1, Hlip2, or Hlip3, which is also consistent with a recent gene duplication event (Table 2). If *hli4* and *hli5* represent a recent duplication event, then the gene fusion between the *coh1* homolog (represented by TMH1 in Hlip5) and a *hli*-like gene (represented by TMH2 in Hlip5) may also have occurred recently. Thus, if the Coh1 polypeptide is necessary for functionality, Hlip5 represents a novel fusion event between Coh1 and a Hlip domain. It has been suggested that new multi-domain proteins are formed during the course of evolution by the process of gene duplication followed by gene fusion. Such a fusion may also result in tight co-regulation of these genes and serve to expand the functional role of fusion proteins (Bashton and Chothia, 2007; Lerat et al., 2005; Yanai et al., 2000). In fact, among the Hlips examined here, the TMH2 domain of Hlip5 is the least conserved between *Synechococcus* OS-A and *Synechococcus* OS-B′ and might suggest that the TMH2 of Hlip5 is evolving rapidly. It is also worth noting that in most cyanobacteria, ferrochelatase, which catalyzes the insertion of ferrous iron into the protoporphyrin IX ring, yielding protoheme, contains a Hlip domain fused to the C-terminus of the enzyme (Vavilin et al., 2002). However, the ferrochelatase gene (*hemH*) from *Synechococcus* OS-A (CYA_1034) and *Synechococcus* OS-B′ (CYB_2857) do not contain a Hlip domain. The only other example of a ferrochelatase lacking an Hlip domain is *hemH* (gvip102) from *Gloeobacter violaceus* PCC7421—a unicellular cyanobacterium that lacks internal thylakoid membranes and is considered to have diverged very early from other cyanobacteria (Honda et al., 1999; Nakamura et al., 2003).

In *Synechococcus* OS-B′, the *hli* transcripts are rapidly induced within minutes of exposure to HL. The patterns of accumulation of the different *hli* transcripts are not identical when cells are transferred from dark to LL, ML, or HL. Thus, *hli3* and *hli4* are induced primarily by HL, while *hli1*, *hli2*, and *hli5* respond to moderate light levels. The *hli5* transcript is detected under very low light conditions as well as in the dark (Figure 4A and 5B), suggesting that some Hlips may have a functional role under all light conditions. The pattern of transcript abundance in high light was examined for *hli4/hli5* and suggests the presence of a light-inducible promoter upstream of *hli4* and *hli5*, although we were unable to find the HLR motif in the region upstream of the gene. This motif has been found upstream of certain high-light-induced genes (Kappell et al., 2007). Differential cleavage and/or termination can also explain the presence of the multiple transcripts (i.e. *hli4*, *hli5*, and *hli4–5*), which is consistent with the presence of a predicted stem-loop structure in the intergenic region between *hli4* and *hli5* (Brodsky et al., 1995). Based on the data that the short *hli4* and *hli5*-specific transcripts appear more abundant than the long transcript covering *hli4* and *hli5*, it is possible that there are multiple promoters which are responsive to different stress conditions. For instance, using *hli4*- and *hli5*-specific primers for RT-PCR with RNA derived from nitrogen-starved cells, we find that both transcripts are induced, but the kinetics of accumulation are quite different (data not shown). These data are consistent with the finding that in *Synechocystis* PCC6803, which contains four Hlips, accumulation of epitope-tagged Hlips showed a differential response to suboptimal growth conditions such as high light, nitrogen, or sulfur limitation and low temperature (He et al., 2001).

Based on amino acid sequence similarity to the Hlips of *Synechocystis* PCC6803, Hlip1 is most similar to HliA and HliB (which are very similar to each other) and Hlip2 and Hlip3 are most similar to HliD and HliC, respectively. Hlip4 and Hlip5 have no obvious homologs in *Synechocystis* PCC6803 (Table 2). Although transcript accumulation in *Synechococcus* OS-B′ and protein accumulation in *Synechocystis* PCC6803 may not be directly comparable, the relative transcript levels (*hli 3*>*hli1*>*hli2*, *hli4*, *hli5*) in *Synechococcus* OS-B′ as shown here exhibit similar trends as protein accumulation (HliC> HliA/HliB>HliD) in *Synechocystis* sp. strain PCC6803 (He et al., 2001). Thus, there may be conserved functions for the different Hlips across genera. Based on protein similarity algorithms, it was suggested that there may be three basic Hlip types in freshwater species, while marine cyanobacteria appear to form separate clusters (Bhaya et al., 2002), but this has yet to be linked to different functional roles for the Hlips. In the case of the high-light-acclimated ecotype of marine *Prochlorococcus* MED4, there are 22 copies of the *hli* genes, including four tandemly arranged, almost identical copies of *hli* genes (*hli6–hli10*); this entire region is also duplicated (*hli16–hli19*) elsewhere in the genome (Bhaya et al., 2002). Transcripts from all of these genes are significantly up-regulated in high light, although other members of the *hli*

gene family are constitutively expressed (Steglich et al., 2006). Lateral gene transfer of Hlips has been postulated to occur via phage and this may assist in creating and maintaining diversity in marine environments; however, the presence or role of viruses in microbial mats is yet to be determined (Clokie and Mann, 2006; Lindell et al., 2004).

It has been shown that in *Synechococcus* sp. strain PCC7942, a *hli* transcript accumulates to high levels following exposure of cells to high light or UV-A (Dolganov et al., 1995; Salem and van Waasbergen, 2004), while in *Synechocystis* PCC6803, *hli* transcripts are induced under a variety of stress conditions, such as high light, nutrient stress, and low temperature (He et al., 2001; Mikami et al., 2002). A *Synechocystis* PCC6803 mutant in which all four *hli* genes were inactivated was unable to survive under high-light conditions, suggesting that Hlips are crucial for acclimation to high light (Funk and Vermaas, 1999; He et al., 2001). A suppressor of the high-light lethality quadruple mutant was isolated and mapped to a gene (*pfsR*) encoding a regulatory protein. *PfsR* appears to be critical for regulation of iron homeostasis, which, in turn, is important for photosynthesis, controlling oxidative stress and cell survival (Jantaro et al., 2006). ScpE (HliC) and ScpB (HliD) may regulate an early step of tetrapyrrole biosynthesis and, since these proteins contain conserved chlorophyll-binding sites, it was postulated that they act as regulators that function based on chlorophyll availability, such that they activate chlorophyll biosynthesis steps when their pigment binding sites are unoccupied (Xu et al., 2002). However, direct binding of chlorophyll or other pigments to Hlips (Scps) has not been demonstrated. ScpD (HliB) is associated with the periphery of Photosystem II following high-light treatment, suggesting that it may be required for PSII assembly or repair (Promnares et al., 2006; Yao et al., 2007).

It is possible to speculate that Hlip5, which has two TMH domains, might have increased stability in the membrane and/or additional partner proteins which alter its ability to bind chlorophyll or change its association with complexes of the photosynthetic apparatus (Funk and Vermaas, 1999). One could envisage the formation of a more stable heteromeric structure that is similar in structure to the three TMH-containing LHC proteins, by the interaction of different Hlip domains. For instance, a heterodimer composed of a single membrane Hlip (such as Hlip1, Hlip2, Hlip3, or Hlip4) would be held together to TMH1/TMH2 of Hlip5 by salt bridges created by the conserved charged amino acids, glutamate (E), and arginine (R), thereby forming a typical X-like structure in the membrane (Figure 7) (Kuhlbrandt et al., 1994; Standfuss et al., 2005). A similar salt bridge-connected dimer could also be formed between TMH1 and TMH2 of Hlip5, or by two interacting single TMH proteins, as suggested by Dolganov et al. (1995). Currently, the relative stability and pigment-binding properties of these forms have yet to be experimentally demonstrated.

We have recently shown that *Synechococcus* OS-B' is unable to survive when grown at photon fluence levels of 400 μmol photon m^{-2} s^{-1} (Kilian et al., 2007). Thus, the role of Hlips, and, in particular, the novel two-TMH Hlip5, is of interest.

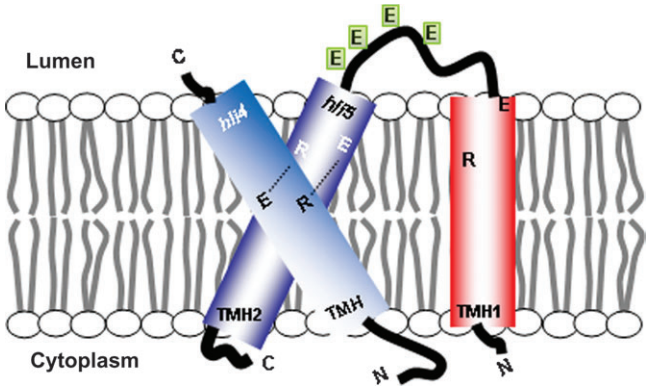

Figure 7. Cartoon of One Possible Configuration of Hlip4 and Hlip5 in the Membrane. Possible dimerization by R–E salt bridges are indicated in the TMH of Hlip4 (light blue) and TMH2 (purple) of Hlip5. The conserved R and E residues in TMH1 (red) of Hlip5 are also shown. The position of the glutamate (E)/aspartate (D) residues (boxed green) is shown in the lumen as predicted by membrane topology.

Membrane topology algorithms predict that the loop between TMH1 and TMH2 is more likely to be located in the lumen; thus, the conserved aspartate/glutamate residues that are positioned between TMH1 and TMH2 would be present in the lumen and could serve a regulatory function. For instance, molecular analysis of nonphotochemical quenching in *A. thaliana* suggested that PSBS could sense hyper-excitation of photosynthetic electron transport via the development of a large ΔpH across the thylakoid membranes which, in turn, led to protonation of conserved, lumenal glutamate residues in PSBS (Li et al., 2004). Furthermore, PSBS may bind carotenoids that can de-excite chlorophyll molecules through the formation of a carotenoid radical (Holt et al., 2005; Standfuss et al., 2005). Whether Hlip5 could play a similar role needs further analysis.

The discovery that the LHC super-family contains members with one to four TMHs has lent support to the model of evolution in which Hlips represent one of the single TMH progenitors of the CAB/ELIP/HLIP superfamily (Green and Pichersky, 1994; Hoffman et al., 1987; Kuhlbrandt et al., 1994). This model postulates that a Hlip sequence fused with another TMH to create a two TMH structure (e.g. the SEPs). A four-helix protein (e.g. PSBS) could represent an event by which a gene for a two-TMH protein experienced a tandem duplication followed by fusion of the duplicated genes. The evolution of a three-TMH LHC or ELIP from a PSBS-type, four-TMH protein would involve loss of the last TMH (Green and Pichersky, 1994; Green et al., 1991; Montane and Kloppstech, 2000). While TMH1 and TMH3 of the LHCs are similar to each other and are also similar to TMHs of the distantly related CAB/ELIP/Hlip polypeptides, the TMH2 of the LHCs is not well conserved among the different LHC family members. This raises questions about both the function and origins of TMH2-like domains (Heddad and Adamska, 2000; Montane and Kloppstech, 2000). Our finding that Hlip5 represents a novel protein containing an Hlip domain and a domain representing a conserved protein in cyanobacteria suggests that closer examination of genomic data from a wide variety of

photosynthetic organisms might reveal the presence of other novel fusion events between ancient *hli* genes and other genes.

ACKNOWLEDGMENTS

We acknowledge the expert help of Fariba Fazeli. This work was funded by the National Science Foundation Frontiers in Integrative Biology Program (FIBR) award EF0328698.

REFERENCES

Adamska, I. (1997). ELIPs: light-induced stress proteins. Physiol. Plant **100,** 794–805.

Adamska, I., Ohad, I., and Kloppstech, K. (1992). Synthesis of the early light-inducible protein is controlled by blue light and related to light stress. Proc. Natl Acad. Sci. U S A **89,** 2610–2613.

Allewalt, J.P., Bateson, M.M., Revsbech, N.P., Slack, K., and Ward, D.M. (2006). Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in the octopus spring microbial mat community of Yellowstone National Park. Appl. Environ. Microbiol. **72,** 544–550.

Andersson, U., Heddad, M., and Adamska, I. (2003). Light stress-induced one-helix protein of the chlorophyll a/b-binding family associated with photosystem I. Plant Physiol. **132,** 811–820.

Bashton, M., and Chothia, C. (2007). The generation of new protein functions by the combination of domains. Structure **15,** 85–99.

Bhaya, D., Grossman, A.R., Steunou, A.-S., Khuri, N., Cohan, F.M., Hamamura, N., Melendrez, M.C., Bateson, M.M., Ward, D.M., and Heidelberg, J.F. (2007). Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. ISME Journal in press.

Bhaya, D., Dufresne, A., Vaulot, D., and Grossman, A. (2002). Analysis of the *hli* gene family in marine and freshwater cyanobacteria. FEMS Microbiol. Lett. **215,** 209–219.

Bhaya, D., Watanabe, N., Ogawa, T., and Grossman, A.R. (1999). The role of an alternative sigma factor in motility and pilus formation in the cyanobacterium *Synechocystis* sp. strain PCC6803. Proc. Natl Acad. Sci. U S A **96,** 3188–3193.

Brodsky, L.I., Ivanov, V.V., Kalai dzidis, Y.L., Leontovich, A.M., Nikolaev, V.K., Feranchuk, S.I., and Drachev, V.A. (1995). GeneBee-NET: Internet-based server for analyzing biopolymers structure. Biochemistry (Mosc.) **60,** 923–928.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. **31,** 3497–3500.

Clokie, M.R., and Mann, N.H. (2006). Marine cyanophages and light. Environ. Microbiol. **8,** 2074–2082.

Davis, H.C., and Guillard, R.R.L. (1958). Relative value of ten genera of micro-organisms as foods for oyster and clam larvae. USFWS Fish Bull. **136,** 293–304.

Dekker, J.P., and Boekema, E.J. (2005). Supramolecular organization of thylakoid membrane proteins in green plants. Biochimica et Biophysica Acta (BBA) – Bioenergetics **1706,** 12.

Dolganov, N.A.M., Bhaya, D., and Grossman, A.R. (1995). Cyanobacterial protein with similarity to the chlorophyll *a/b* -binding proteins of higher plants: evolution and regulation. Proc. Natl Acad. Sci. U S A **92,** 636–640.

Funk, C., and Vermaas, W. (1999). A cyanobacterial gene family coding for single-helix proteins resembling part of the light-harvesting proteins from higher plants. Biochemistry **38,** 9397–9404.

Glockner, G., Rosenthal, A., and Valentin, K. (2000). The structure and gene repertoire of an ancient red algal plastid genome. J. Mol. Evol. **51,** 382–390.

Green, B.R. (1995). Sequence conservation of light-harvesting and stress response proteins in relation to the molecular structure of LHCII. Photosynth. Res. **44,** 139–148.

Green, B.R., and Pichersky, E. (1994). Hypothesis for the evolution of three helix Chla/b and Chla/c light harvesting antenna proteins from two-helix and four helix ancestors. Photosynth. Res. **39,** 149–162.

Green, B.R., Pichersky, E., and Kloppstech, K. (1991). Chlorophyll *a/b* binding proteins: an extended family. Trends Biochem. Sci. **16,** 181–186.

Grossman, A.R., Bhaya, D., Apt, K.E., and Kehoe, D.M. (1995). Light-harvesting complexes in oxygenic photosynthesis: diversity, control, and evolution. Annu. Rev. Genet. **29,** 231–288.

Grossman, A.R., Bhaya, D., and He, Q. (2001). Tracking the light environment by cyanobacteria and the dynamic nature of light harvesting. J. Biol. Chem. **276,** 11449–11452.

Grossman, A.R., Schaefer, M.R., Chiang, G.G., and Collier, J.L. (1993). The phycobilisome, a light-harvesting complex responsive to environmental conditions. Microbiol. Rev. **57,** 725–749.

Havaux, M., Guedeney, G., He, Q., and Grossman, A.R. (2003). Elimination of high-light-inducible polypeptides related to eukaryotic chlorophyll a/b-binding proteins results in aberrant photoacclimation in *Synechocystis* PCC6803. Biochim. Biophys. Acta **1557,** 21–33.

He, Q., Dolganov, N., Bjorkman, O., and Grossman, A.R. (2001). The high light-inducible polypeptides in *Synechocystis* PCC6803: expression and function in high light. J. Biol. Chem. **276,** 306–314.

Heddad, M., and Adamska, I. (2000). Light stress-regulated two-helix proteins in Arabidopsis thaliana related to the chlorophyll a/b-binding gene family. Proc. Natl Acad. Sci. U S A **97,** 3741–3746.

Hoffman, N.E., Pichersky, E., Malik, V.S., Castresano, C., Ko, K., Darr, S.C., and Cashmore, A.R. (1987). A cDNA encoding a photosystem I protein with homology to photosystem II chlorophyll a/b-binding proteins. Proc. Natl Acad. Sci. U S A **84,** 8844–8848.

Holt, N.E., Zigmantas, D., Valkunas, L., Li, X.P., Niyogi, K.K., and Fleming, G.R. (2005). Carotenoid cation formation and the regulation of photosynthetic light harvesting. Science **307,** 433–436.

Honda, D., Yokota, A., and Sugiyama, J. (1999). Detection of seven major evolutionary lineages in cyanobacteria based on 16S rRNA gene analysis with new sequences of five marine *Synechococcus* strains. J. Mol. Evol. **48,** 723–739.

Jansson, S. (1999). A guide to the Lhc genes and their relatives in Arabidopsis. Trends Plant Sci. **4,** 236–240.

Jansson, S. (1994). The light-harvesting chlorophyll *a/b* -binding proteins. Biochimica et Biophysica Acta **1184,** 1–19.

Jansson, S., Andersson, J., Jung-Kim, S., and Jackowski, G. (2000). An *Arabidopsis thaliana* protein homologous to cyanobacterial high-light-inducible protiens. Plant Mol. Biol. **42,** 345–351.

Jantaro, S., Ali, Q., Lone, S., and He, Q. (2006). Suppression of the lethality of high light to a quadruple HLI mutant by the inactivation of the regulatory protein PfsR in *Synechocystis* PCC 6803. J. Biol. Chem. **281,** 30865–30874.

Kappell, A.D., and van Waasbergen, L.G. (2007). The response regulator RpaB binds the high light regulatory 1 sequence upstream of the high-light-inducible *hliB* gene from the cyanobacterium *Synechocystis* PCC 6803. Arch. Microbiol. **187,** 337–342.

Kilian, O., Steunou, A.S., Fazeli, F., Bailey, S., Bhaya, D., and Grossman, A.R. (2007). Responses of a thermophilic *Synechococcus* isolate from the microbial mat of Octopus spring to light. Appl. Environ. Microbiol. **73,** 4268–4278.

Koonin, E.V. (2005). Orthologs, paralogs, and evolutionary genomics. Annu. Rev. Genet. **39,** 309–338.

Koziol, A.G., Borza, T., Ishida, K.-I., Keeling, P., Lee, R.W., and Durnford, D.G. (2007). Tracing the evolution of the light-harvesting antennae in chlorophyll a/b-containing organisms. Plant Physiol. **143,** 1802–1816.

Kuhlbrandt, W., Wang, D.N., and Fujiyoshi, Y. (1994). Atomic model of plant light harvesting complex by electron crystallography. Nature (London) **367,** 614–621.

Lerat, E., Daubin, V., Ochman, H., and Moran, N.A. (2005). Evolutionary origins of genomic repertoires in bacteria. PLoS Biol. 3,e130.

Li, X.P., Bjorkman, O., Shih, C., Grossman, A.R., Rosenquist, M., Jansson, S., and Niyogi, K.K. (2000). A pigment-binding protein essential for regulation of photosynthetic light harvesting. Nature **403,** 391–395.

Li, X.P., Gilmore, A.M., Caffarri, S., Bassi, R., Golan, T., Kramer, D., and Niyogi, K.K. (2004). Regulation of photosynthetic light harvesting involves intrathylakoid lumen pH sensing by the PsbS protein. J. Biol. Chem. **279,** 22866–22874.

Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., and Chisholm, S.W. (2004). Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. Proc. Natl Acad. Sci. U S A **101,** 11013–11018.

Lucinski, R., and Jackowski, G. (2006). The structure, functions and degradation of pigment-binding proteins of photosystem II. Acta Biochimica Polonica **53,** 693–708.

Markowitz, V.M., et al. (2006). An experimental metagenome data management and analysis system. Bioinformatics **22,** e359–e367.

Mikami, K., Kanesaki, Y., Suzuki, I., and Murata, N. (2002). The histidine kinase Hik33 perceives osmotic stress and cold stress in *Synechocystis* sp PCC 6803. Mol. Microbiol. **46,** 905–915.

Montane, M.-H., and Kloppstech, K. (2000). The family of light-harvesting-related proteins (LHCs, ELIPs, HLIPs): was the harvesting of light their primary function? Gene **258,** 1.

Montane, M.H., Dreyer, S., Triantaphylides, C., and Kloppstech, K. (1997). Early light-inducible proteins during long-term acclimation of barley to photooxidative stress caused by light and cold: high level of accumulation by posttranscriptional regulation. Planta **202,** 293–302.

Nakamura, Y., et al. (2003). Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. DNA Res. **10,** 137–145.

Niyogi, K.K. (1999). Photoprotection revisited: genetic and molecular approaches. Annu. Rev. Plant Physiol. Plant Mol. Biol. **50,** 333–359.

Promnares, K., Komenda, J., Bumba, L., Nebesarova, J., Vacha, F., and Tichy, M. (2006). Cyanobacterial small chlorophyll-binding protein ScpD (HliB) is located on the periphery of photosystem II in the vicinity of PsbH and CP47 subunits. J. Biol. Chem. **281,** 32705–32713.

Punta, M., Forrest, L.R., Bigelow, H., Kernytsky, A., Liu, J., and Rost, B. (2007). Membrane protein prediction methods. Methods **41,** 460–474.

Ramsing, N.B., Ferris, M.J., and Ward, D.M. (2000). Highly ordered vertical structure of *Synechococcus* populations within the one-millimeter-thick photic zone of a hot spring cyanobacterial mat. Appl. Environ. Microbiol. **66,** 1038–1049.

Rost, B., Yachdav, G., and Liu, J. (2004). PredictProtein: The PredictProtein Server. Nucleic Acids Res. **32 (Web Server issue),** W321–W326.

Salem, K., and van Waasbergen, L.G. (2004). Light control of *hliA* transcription and transcript stability in the cyanobacterium *Synechococcus elongatus* strain PCC 7942. J. Bacteriol. **186,** 1729–1736.

Standfuss, J., Terwisscha van Scheltinga, A.C., Lamborghini, M., and Kuhlbrandt, W. (2005). Mechanisms of photoprotection and nonphotochemical quenching in pea light-harvesting complex at 2.5 A resolution. Embo J **24,** 919–928.

Steglich, C., Futschik, M., Rector, T., Steen, R., and Chisholm, S.W. (2006). Genome-wide analysis of light sensing in *Prochlorococcus*. J. Bacteriol. **188,** 7796–7806.

Steunou, A.S., Bhaya, D., Bateson, M.M., Melendrez, M.C., Ward, D.M., Brecht, E., Peters, J.W., Kuhl, M., and Grossman, A.R. (2006). In situ analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats. Proc. Natl Acad. Sci. U S A **103,** 2398–2403.

Tzvetkova-Chevolleau, T., Franck, F., Alawady, A.E., Dall Osto, L., Carriere, F., Bassi, R., Grimm, B., Nussaume, L., and Havaux, M. (2007). The light stress-induced protein ELIP2 is a regulator of chlorophyll synthesis in *Arabidopsis thaliana*. Plant J **50,** 795–809.

Vavilin, D.V., and Vermaas, W.F. (2002). Regulation of the tetrapyrrole biosynthetic pathway leading to heme and chlorophyll in plants and cyanobacteria. Physiol. Plant. **115,** 9–24.

Ward, D.M., Ferris, M.J., Nold, S.C., and Bateson, M.M. (1998). A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiol. Mol. Biol. Rev. **62,** 1353–1370.

Xu, H., Vavilin, D., Funk, C., and Vermaas, W. (2002). Small Cab-like proteins regulating tetrapyrrole biosynthesis in the cyanobacterium *Synechocystis* sp. PCC 6803. Plant Mol. Biol. **49,** 149–160.

Yanai, I., Camacho, C.J., and DeLisi, C. (2000). Predictions of gene family distributions in microbial genomes: evolution by gene duplication and modification. Phys. Rev. Lett. **85,** 2641–2644.

Yao, D., Kieselbach, T., Komenda, J., Promnares, K., Prieto, M.A., Tichy, M., Vermaas, W., and Funk, C. (2007). Localization of the small CAB-like proteins in photosystem II. J. Biol. Chem. **282,** 267–276.