Molecular Plant
Research Article

CellPress
PARTNER JOURNAL

Structure, Diversity, and Evolution of a New Family of Soluble Carotenoid-Binding Proteins in Cyanobacteria

Matthew R. Melnicki[1,2,7], Ryan L. Leverenz[3,7], Markus Sutter[2,3,7], Rocío López-Igual[4,5,8], Adjélé Wilson[4,5], Emily G. Pawlowski[3], François Perreau[6], Diana Kirilovsky[4,5] and Cheryl A. Kerfeld[1,2,3,*]

[1]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA

[2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[3]MSU-DOE Plant Research Laboratory, Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

[4]Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, Université Paris-Saclay, 91198 Gif-sur-Yvette, France

[5]Commissariat à l'Energie Atomique (CEA), Institut de Biologie et Technologies de Saclay (iBiTec-S), 91191 Gif-sur-Yvette, France

[6]INRA, Institut Jean-Pierre Bourgin, UMR 1318, ERL CNRS 3559, Saclay Plant Sciences, RD10, 78026 Versailles, France

[7]These authors contributed equally to this article.

[8]Present address: Unité Plasticité du Génome Bactérien, Département Génomes et Génétique, Institut Pasteur, CNRS, Unité Mixte de Recherche 3525, 75015 Paris, France

***Correspondence: Cheryl A. Kerfeld (ckerfeld@lbl.gov)**

http://dx.doi.org/10.1016/j.molp.2016.06.009

ABSTRACT

Using a phylogenomic approach, we have identified and subclassified a new family of carotenoid-binding proteins. These proteins have sequence homology to the N-terminal domain (NTD) of the Orange Carotenoid Protein (OCP), and are referred to as Helical Carotenoid Proteins (HCPs). These proteins comprise at least nine distinct clades and are found in diverse organisms, frequently as multiple paralogs representing the distinct clades. These seem to be out-paralogs maintained from ancient duplications associated with subfunctionalization. All of the HCPs share conservation of the residues for carotenoid binding, and we confirm that carotenoid binding is a fundamental property of HCPs. We solved two crystal structures of the *Nostoc* sp. PCC 7120 HCP1 protein, each binding a different carotenoid, suggesting that the proteins flexibly bind a range of carotenoids. Based on a comprehensive phylogenetic analysis, we propose that one of the HCP subtypes is likely the evolutionary ancestor of the NTD of the OCP, which arose following a domain fusion event. However, we predict that the majority of HCPs have functions distinct from the NTD of the OCP. Our results demonstrate that the HCPs are a new family of functionally diverse carotenoid-binding proteins found among ecophysiologically diverse cyanobacteria.

Key words: carotenoids, cyanobacteria, phylogenomics, protein structure, photoprotection, molecular evolution

Melnicki M.R., Leverenz R.L., Sutter M., López-Igual R., Wilson A., Pawlowski E.G., Perreau F., Kirilovsky D., and Kerfeld C.A. (2016). Structure, Diversity, and Evolution of a New Family of Soluble Carotenoid-Binding Proteins in Cyanobacteria. Mol. Plant. **9**, 1379–1394.

INTRODUCTION

Carotenoids are produced by all photosynthetic organisms (plants, algae, and cyanobacteria) in which they play two fundamental roles: as light-harvesting pigments and as protectors against oxidative stress. In membranes, they can act as sunscreen or react with dangerous and highly reactive singlet oxygen. In the light-harvesting antennae and other membrane-associated carotenoid–chlorophyll complexes, carotenoids, by interacting with excited chlorophyll molecules, dissipate excess excitation energy as heat to avoid formation of singlet oxygen. They can also directly quench singlet oxygen.

Although carotenoid-containing proteins of photosynthetic organisms typically contain multiple pigments including chlorophyll and are membrane-associated, the Orange Carotenoid

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, SIBS, CAS.

Protein (OCP), unique to cyanobacteria, is water soluble and relies on a single carotenoid to carry out both photosensory and photoprotective functions. It is the only known soluble photosensory protein that uses a carotenoid as the active chromophore. Blue-green light (~500 nm) drives a structural transition from the protein's dark, stable form, OCP^O, to its light-activated form, OCP^R (Wilson et al., 2008). In the OCP^R state, it appears red and can bind to the primary light-harvesting antenna, the phycobilisome (PBS), where it triggers excitation energy dissipation (Wilson et al., 2007; Gwizdala et al., 2011). The OCP is composed of two structural domains (Kerfeld et al., 2003) that have discrete, complementary functions (Leverenz et al., 2014). While the C-terminal domain (CTD) of the OCP confers a regulatory role, the N-terminal domain (NTD) is the effector domain. This domain, when isolated from the OCP by proteolysis, retains the carotenoid and appears red (Holt and Krogmann, 1981; Wu and Krogmann, 1997; Leverenz et al., 2014, 2015) and is an avid quencher of singlet oxygen (Kerfeld et al., 2003). Recently, it was shown that the isolated NTD, also known as the Red Carotenoid Protein (RCP), interacts with the phycobilisome (PBS) antenna complexes to constitutively facilitate energy quenching, independent of the CTD (Leverenz et al., 2014, 2015). The crystal structure of the NTD-binding carotenoid was solved, revealing that upon removal of the CTD, the carotenoid shifts 12 Å deeper into the protein (Leverenz et al., 2015). This intraprotein translocation of the carotenoid defines two carotenoid protein configurations (CPCs), discrete carotenoid protein interactions associated with OCP^R (CPC-R) and OCP^O (CPC-O).

Previously, it was noted that genes encoding for hypothetical proteins with homology to the NTD of the OCP (pfam09150) are present in many cyanobacterial genomes (Kerfeld et al., 2003; Kirilovsky and Kerfeld, 2013). These hypothetical proteins contain only the pfam09150 domain and are present in multiple copies (paralogs). For example, *Nostoc* ("*Anabaena*") sp. PCC 7120 (hereafter *Nostoc* 7120) contains five genes containing the pfam09150 domain, including the OCP and four single-domain homologs (Kerfeld et al., 2003; Kerfeld, 2004a, 2004b). Transcripts for all five genes have been detected in global expression studies (Flaherty et al., 2011). Occasionally these NTD-like proteins are found in genomes that lack OCP or an identifiable OCP-like quenching mechanism (Abasova et al., 2008).

Using a phylogenomic approach, we investigated the diversity of genes encoding single-domain proteins with the pfam09150 domain. We identified nine clades, indicative of ancient duplication events. One of these clades was found to be closely related to the ancestral OCP-NTD, allowing reconstruction of a plausible scenario for the evolutionary history of the OCP. We described the widespread taxonomic distribution of these NTD homologs, showing how multiple paralogs from different clades are frequently encoded within a single genome. We illustrated structural features that distinguish the different clades and propose how they may be related to functional diversification within this family of proteins. Furthermore, we established carotenoid-binding as a fundamental property for the pfam09150 domain and confirmed this by solving two HCP crystal structures from *Nostoc* 7120, showing how the all-helical fold may accommodate different carotenoids within the predicted, conserved binding site. Collectively, this work provides a foundation for determining the physiological role of this new family of pigment-binding proteins and offers insight into the evolution of a modular photoreceptor protein involved in photoprotection in cyanobacteria.

RESULTS

Phylogenetic Analysis and Overview of Homologs to the NTD of the OCP

Homologs to the NTD of the OCP were detected by hits to the Pfam hidden Markov model (HMM) for this domain (pfam09150) in protein sequences compiled in the Integrated Microbial Genomes (IMG) database. All hits belonged to cyanobacterial genomes. Sequences for 196 OCP-NTD-like homologs and 161 full-length OCP-like homologs were gathered and used to construct an alignment block suitable for phylogenetic tree construction using a maximum likelihood method (Figure 1). Among the homologs, several distinct monophyletic clades with long stem lengths were identified, indicating that several subtypes are distinguishable by sequence conservation. Given that the pfam09150 domain is predicted to be all α-helical, and that the homologs all share conservation of the carotenoid-binding residues (in the CPC-R configuration) (see next section), we refer to these as Helical Carotenoid Proteins (HCPs). Accordingly, we have adopted a provisional system of nomenclature to refer to these subgroups (clades) for the remainder of the article.

The HCP and OCP-NTD sequences clearly populate different sides of the unrooted tree, with the majority of OCP homologs (i.e., the NTD) found in a large clade mapped at the opposite end from the HCP9 clade, which contains only HCPs from the basal cyanobacterial genus, *Gloeobacter*. The OCP cluster contains 112 sequences and represents the canonical OCP, including the well-characterized representative from *Synechocystis* sp. PCC 6803 (hereafter Syn6803), the *slr1963* gene product (Figure 1A, indicated with a green square). The amino acid sequence conservation among the OCP-NTDs is quite high: 76 have at least 80% identity with at least one other sequence, and only 46 sequences remain when the redundancy threshold of their subalignment is lowered to 80% identity. For all but seven members of the OCP clade, a gene for the Fluorescence Recovery Protein (FRP), which binds to the CTD of the OCP to regulate its function (Sutter et al., 2013), is also present in the genome.

Of the 154 genomes that contain a gene encoding the pfam09150 domain (OCP or HCP), 87 contain at least one HCP (Figure 1B) and 66 contain two or more HCP paralogs. Among the 255 cyanobacterial genomes examined, 101 did not encode the OCP nor any HCPs (see below). The majority of the HCP-encoding genomes (83) also contain a homolog of the C-terminal domain (CTDH) of the OCP (Figure 1C). For any given genome, the HCP paralogs rarely are members of the same clade but rather are representatives of an assortment of different clades (Figure 1D). For instance, the sequences encoded in the *Nostoc* 7120 genome (indicated by the red squares in Figure 1A) are well distributed throughout the tree. While there is no clear pattern for the maintenance of certain HCP clades in particular genomes, the HCP1–4 clades are the most commonly

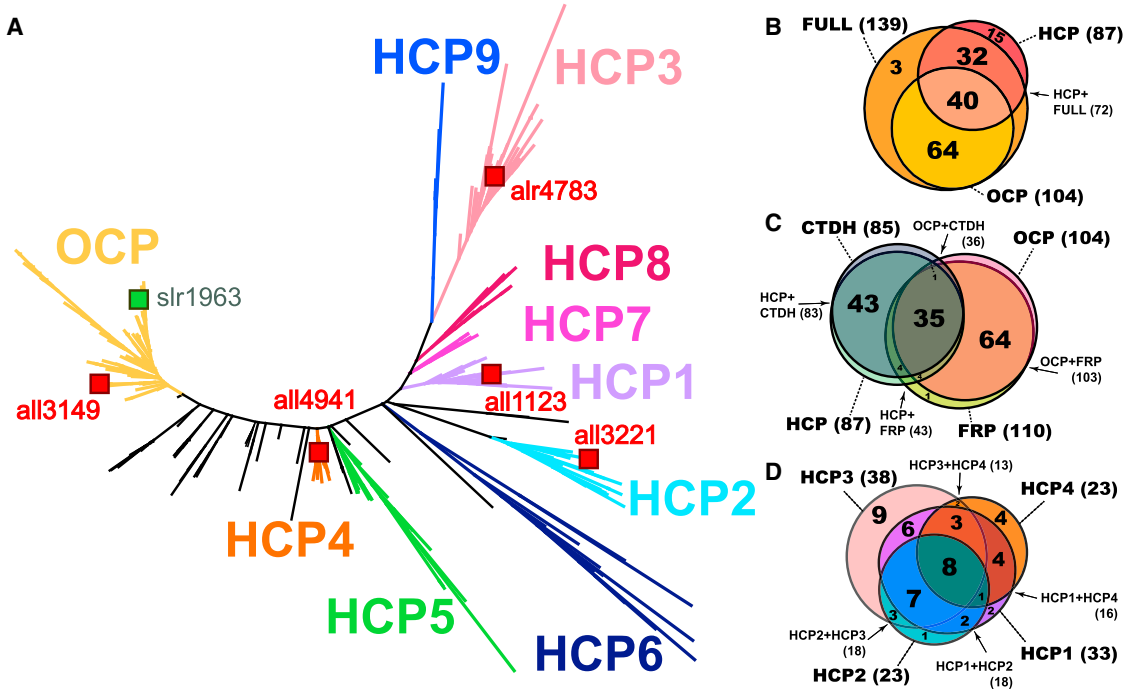


Figure 1. Evolutionary Divergence of HCP Subtypes.

(A) Unrooted Maximum Likelihood phylogenetic tree of pfam09150 amino acid sequences. The OCP cluster and HCP subtypes were identified as monophyletic clades and are designated by numbers and colors. Genes from *Synechocystis* 6803 and *Nostoc* 7120 are indicated by green and red squares, respectively.

(B–D) Euler diagrams showing the number of cyanobacterial genomes with co-occurring sets of **(B)** HCPs and full-length OCP-like homologs, **(C)** HCPs and other OCP-related genes, and **(D)** among the four most common HCP clades (HCP1–4). Unlabeled intersections have zero members and were required for finding an optimal solution. For **(C)**, an intersection between CTDH and FRP (two genomes) could not be drawn.

found and are the most likely to co-occur, as depicted in Supplemental Figure 1, which shows the clade distributions of 12 genomes containing four or more HCP paralogs. Members of the HCP4 and HCP5 clades typically are encoded adjacent to genes for CTDH (36 out 53 sequences) (see Supplemental Figure 2). These clades are found in the middle of the tree and are the HCP clades closest to the OCP clade.

Overall, HCPs do appear to be conserved among certain morphological and taxonomic groups. HCPs are encoded by nearly all heterocystous cyanobacteria (morphological subsections IV/V (Rippka et al., 1979), phylogenetic subclade B1 (Shih et al., 2013)) (Table 1). In contrast, only 21 of 92 Chroococcales genomes encode an HCP, with no clear relationship distinguishing these genomes from the 43 Chroococcales encoding only the OCP. Prochlorales species completely lack HCP (as well as OCP), whereas other alpha cyanobacteria (subclade C1, marine *Synechococcus*, *Cyanobium*) contain OCP without HCP. Among the other strains that lack NTD-like homologs altogether, several (*Moorea* 3L, *Cyanobacterium aponinum* PCC 10605, *Synechococcus* CC9616, *Leptolyngbya* BDU, 20041) contain fragments detectable by translated searches (T-BLASTN) against the nucleotide genome. Fifteen genomes, mostly unicellular, contain only HCPs (without full-length OCP homologs), although these appear in a variety of phylogenetic subclades (see Supplemental Table 1). Interestingly, 12 of these genomes (which include *Thermosynechococcus, Leptolyngbya, Crocosphaera*, and three *Cyanothece* strains) contain an HCP that maps close to the HCP4 and HCP5 clades.

Clade-Specific Features of HCPs

We applied a phylogenomic approach (Sjölander, 2004), parsing the HCP paralogs by their evolutionarily identified clades. To detect clade-specific features of their primary structures, HMMs were built for the sequences within each clade from Figure 1A, and logos were subsequently generated to depict amino acid conservation patterns among the clades (Figure 2).

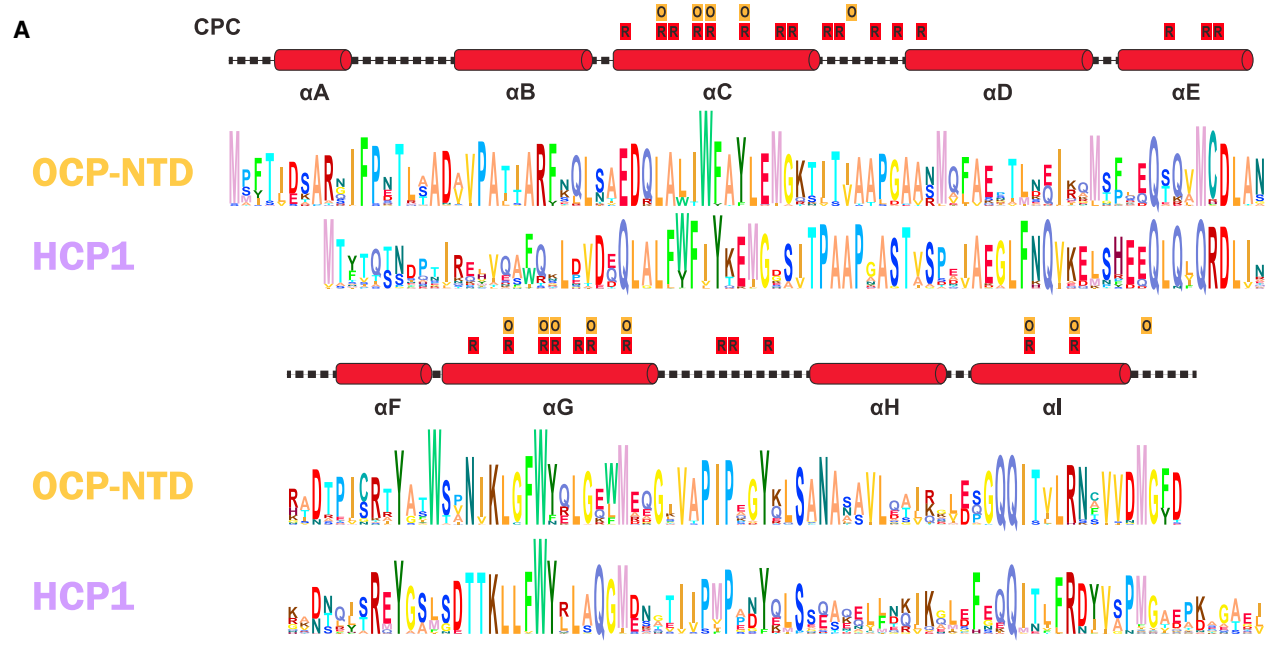
Each clade has a number of specific, conserved features of primary structure, including at positions known to be proximal to the carotenoid in the dark orange OCP (OCP^O) or in the photoactivated red OCP (OCP^R). Together, these positions form the carotenoid-binding "tunnel" (Leverenz et al., 2015) of the OCP (Figure 2; see Supplemental Figure 3 for HMM logos for each clade). There appears to be relatively lower sequence conservation within the HCP6 clade, suggesting that these sequences may have clustered together artificially due to "long-branch attraction" (Philippe et al., 2005) or have experienced a relaxation of evolutionary constraint. To better visualize the position-specific differences between HCP subtypes, consensus sequences from the HMMs were aligned with the primary structure of the NTD of the well-studied Syn6803 OCP (Figure 2B). The most strongly conserved positions in the alignment nearly always correspond with carotenoid-binding

	Genomes analyzed	Both OCP + HCP	HCP only	OCP only	Neither OCP nor HCP	HCP1	HCP2	HCP3	HCP4	HCP5	HCP6	HCP7	HCP8	HCP9	Any HCP	OCP	FULL	CTDH	FRP
Morphological subsection[a]																			
I	151	15	8	43	85	5	5	3	1	10	1	0	5	2	23	51	58	23	56
II	7	5	1	1	0	1	1	1	1	2	1	4	0	0	6	4	6	6	4
III	50	19	5	20	8	3	3	12	0	8	3	3	8	0	24	31	39	22	31
IV	29	21	0	5	3	14	6	16	10	0	5	0	0	0	23	11	28	23	14
V	12	11	1	0	0	10	8	6	11	0	1	0	0	0	12	7	11	12	7
Phylogenetic subclade[b]																			
A	26	6	2	16	2	0	0	5	0	3	0	2	4	0	8	18	22	8	18
B1	47	35	1	5	6	27	16	25	23	0	8	0	0	0	38	18	42	38	21
B2	37	13	6	14	4	1	2	1	0	13	0	4	5	0	19	26	27	19	29
B3	5	4	0	0	1	1	2	2	0	0	0	0	3	0	4	1	4	4	1
C1	66	0	0	10	56	0	0	0	0	0	0	0	0	0	0	9	10	0	10
C2	12	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C3	4	4	0	0	0	0	1	1	0	2	2	0	0	0	4	4	4	3	4
D	2	1	1	0	0	1	0	1	0	0	1	0	0	0	2	1	1	2	1
E	6	0	3	1	2	1	1	0	0	0	0	0	0	0	3	1	1	3	1
F	4	1	0	3	0	0	0	0	0	0	0	0	0	0	1	3	4	1	3
G	5	2	0	0	3	0	0	0	0	0	0	0	0	2	2	0	2	2	0
Taxonomic order																			
Chroococcales	92	13	8	43	28	5	5	3	1	10	1	0	5	0	21	51	56	21	56
Gloeobacterales	2	2	0	0	0	0	0	0	0	0	0	0	0	2	2	0	2	2	0
Melainabacteria	7	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nostocales	30	22	0	5	3	15	7	17	10	0	5	0	0	0	24	12	29	24	15
Oscillatoriales	48	19	5	20	4	3	3	12	0	8	3	3	8	0	24	31	39	22	31
Pleurocapsales	7	5	1	1	0	1	1	1	1	2	1	4	0	0	6	4	6	6	4
Prochlorales	59	0	0	0	59	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Stigonematales	10	10	0	0	0	8	7	5	10	0	1	0	0	0	10	6	10	10	6

Table 1. Taxonomic Distribution of HCP Subtypes and OCP-Related Genes in Sequenced Genomes.

[a]Described in Rippka et al., 1979.

[b]Described in Shih et al., 2013.



B

```
                 1 -MPFTIDSARGIFPN-----TLAADVVPATIARFSQLNAEDQLALIWFAYLEMGKTLTIAAPGAA 59
slr1963
                 1 -MPFTIDSARNIFPN-----TLAADAVPATIARFNQLSAEDQLALIWFAYLEMGKTITIAAPGAA 59
OCP-NTD
                 1 -MTYT----------------QTNDPTIRELVQAFQKLDVDEQLALFWFIYKEMGDSITPAAPGAS 49
   HCP1
                 1 -MTYTTD-----------------ERTKKALEQFQNFDVDTQLALLWYGYLDIKDQLTPANPTSV 47
   HCP2
                 1 -MTASYNKTIP---------QALSQETQNLVEAFNKLDTDAKLAWLYFVYEKMGDSITPAAPAAA 55
   HCP3
                 1 -MTFTTDSASNRFSNAFNYSTQLGDAVASTVAVFKSLSVDDQLAVLWYAYTEMGRSITPAATGAA 64
   HCP4
                 1 -MTYTTTTNNFS---------TQTVDAVPSVTAQFQRLSTDDQLGLLWFIYTNMGRSITPAAPGAA 56
   HCP5
                 1 -MT-----------------STNVNQLEQAVSAFQSLDIDDRLAVLALLYKEIAQAIPADVLNAL 47
   HCP6
                 1 -MTS----------------ANLDQDVQSALSEFKSLSTDDKLALLWYVYTKMGDSITPAAPGAA 48
   HCP7
                 1 -MTFTTKSAYD-----------------QGVEAFKSLSVDDQLALLWFVYTKMGKSITPAAPGAA 47
   HCP8
                 1 VSSEEISDIVDKALDSEAPEAVRSEGLQQAVEQFNALDVDDKLAVLYYVYEAMGESVTPAAPGAA 65
   HCP9

                60 --SMQLAENALKEIQAMGPLQ-QTQAMCD--LANRADTPLCRTYASWSPNIKLGFWYRLGELMEQ 119
slr1963
                60 --NMQFAEPTLNEIKQMSFLE-QSQVMCD--LANRADTPICRTYATWSPNIKLGFWYQLGEWMEQ 119
OCP-NTD
                50 TVSPEIAEGLFNQVKELSHEE-QLQLQRD--LINKADNQISREYGSLSDTTKLLFWYRLAQGMDN 111
   HCP1
                48 ---QDTAAAVFDQIQALSKEE-QLQAQRD--IASGANSDISRAYGALSSSAKLDVWLRLAQGMED 106
   HCP2
                56 --EPELAPMLLGDYLQLSDEE-QLEIMRD--IVNREDTEYSRAYGALKENNQLLVWYAWAEAMGD 115
   HCP3
                65 --RLQLAEGLLNQIKQMSHAE-QLQVMRD--LAAKKNTQISRSYGILSANTKLAFWYELSELMVK 124
   HCP4
                57 --RLQFAEGLLNQVKQMSHDE-QLQFMRD--LVNKVNTPITRAYGVLTNNTKLAFWYQLAEWMRA 116
   HCP5
                48 --FSQAVQNLVAQIQQLSSEE-QLDALRDLELVSGGTTIPTGEYGSLDAESKLAFWYQLAQNLGS 109
   HCP6
                49 --EPEIAEGLFNQVKELSHEE-QLQVQRN--LLEGRDTLISREYGSLSDNTKLLFWYRLAQGMEE 108
   HCP7
                48 --GSEIAQGLFEQVKELPQEE-QLQIMRD--IASKANTEISREYGSLSAETKLAFWYFLAQGMDE 107
   HCP8
                66 --DIELTQAFFGEFDALPQGEAQLEAMRA--LVRCDDNALGREYGKLTENNKLFIWYLLAERMGK 126
   HCP9

slr1963 120 GFVAPIPAGYQLSANANAVLATIQGLESGQQITVLRNAVVDMGFTAGKDGK--RIAEPVVPP 179
OCP-NTD 120 GIVAPIPEGYKLSANASAVLQAIRGLESGQQITVLRNCVVDMGFD------------------ 164
   HCP1 112 GTIIPMPANYQLSSQAQELLNKIKGLEFEQQITLFRDYVSPMGAEPKAGAEI---------- 163
   HCP2 107 GTIIQVPSDYKLPENTKEFVDQIKQLEFEQRVNFMRSAVFEMGAK---------------- 151
   HCP3 116 -TVVDMPSDYEATEAINNLLSQIEGLEFEEQISVLRTIAGQMGYSDVKPIETQAETGKTSSL 176
   HCP4 125 GFVVPMPSNYQLSRDGVQVLEALKQLDFGQQITVLRKVVTDMGVDPLAD------------- 173
   HCP5 117 GDVIPVPADYQLSREANQVLAAIQALDFNQQITVLRNIVVDMGVDPLA---------------- 164
   HCP6 110 -SIVGIPSDYSPSAEATEVLNSLDSLDTEELVSFLKQVL---------------------- 147
   HCP7 109 GTIIPMPEDYKLSGKAQQLLAAIESLDFEQQITFLRKAVIDAGSEPKSGAEI---------- 160
   HCP8 108 GTIIPMPDDYELPQAGKELLGSIESLDFQQQITFLRGIVLNMGAEPAPGAGI---------- 159
   HCP9 127 -DVIGVPADYQLSEDGQQNLVHVKELDFEQQITFLRDVVKPMGRTPIERAV---------- - 176
```

Figure 2. Sequence Conservation among HCP Evolutionary Subtypes.

(A) HMM logos are shown for the OCP-NTD and HCP1 subfamilies of the pfam09150 domain. A full set of logos for all HCP subtypes is available as Supplemental Figure 3. The height of the stack corresponds to conservation at that position, and the height of each individual letter represents the frequency of that amino acid within the distribution. The secondary structure cartoon showing CPC-O/CPC-R positions was adapted from previous work (Leverenz et al., 2015).

(B) Aligned OCP/HCP consensus sequences for HMMs of each clade identified in Figure 1A. Blue coloration indicates conservation strength, according to the BLOSUM62 matrix. Carotenoid-binding residues characteristic of CPC-O (orange highlighted), CPC-R (maroon highlighted), or shared by both CPCs (red highlight) are annotated on the slr1963 OCP-NTD sequence. Red boxes designate residues in a CPC-O/CPC-R-marked column with identity to the OCP-NTD consensus. Insertions unique to HCP6 (24 amino acids) and HCP9 (42 amino acids) were collapsed at the positions of the red arrows.

Species	*Cyanothece* sp. BH68, ATCC 51142		*Nostoc punctiforme* ATCC 29133		
Publication	Bernstein et al., 2015	Toepel et al., 2009	Anderson et al., 2006	Christman et al., 2011	Ow et al., 2009
Study type	Transcript	Transcript	Proteome	Transcript	Proteome
Experimental conditions	N-deprived, CO_2-starved H_2 production	Light/dark cycling	Continuous light, NH_4^+-grown	N-deprivation time course	N_2-fixing vs NH_4^+-grown, filaments vs heterocysts
OCP	Down during N-starvation, rebounds at 15 h, spikes at 21 h	Not reported	1 peptide	Down under N_2 stress	both filament and heterocyst
HCP1	Not present in genome	Not present in genome	3 peptides	Down under N_2 stress	Filament only
HCP2	Not present in genome	Not present in genome	1 peptide	Not reported	Not reported
HCP3	Not present in genome	Not present in genome	Highest abundance (13 peptides)	High in hormogonia	Filament only
HCP4	Not present in genome	Not present in genome	Not present in genome	Not present in genome	Not present in genome
HCP5	High during N-starvation H_2 phase, drops at 21 h	Not reported	Not present in genome	Not present in genome	Not present in genome
HCP6	Not present in genome	Not present in genome	2 peptides	Not reported	Not reported
HCP8	Steady for initial H_2 phase, drops at 15 h, rebounds at 21 h	Up significantly in first dark period, otherwise ND/no change	Not present in genome	Not present in genome	Not present in genome

residues in the NTD of the OCP (Leverenz et al., 2015). Accordingly, we predict that carotenoid binding is a fundamental property of the HCPs and that their tertiary structure resembles the CPC-R form of the OCP-NTD.

The consensus sequence alignments illustrate several additional differences among the primary structures of the HCPs from different clades. Extensions at the N terminus are generally limited to members of the HCP3-5 and HCP8-9 clades, with the longest occurring in the HCP4 and HCP9 clades; conversely, the HCP1, HCP2, HCP6, and HCP7 clades lack this peptide segment. In the OCP, this helix helps stabilize the OCP^O form by interacting with the CTD at the minor interface (Kerfeld et al., 2003; Wilson et al., 2010; Thurotte et al., 2015). A critical PBS-binding arginine, R155 (Syn6803 OCP numbering), is conserved in the vast majority of the OCPs. Only 19 HCP sequences have a different amino acid in this position; six of these are members of the HCP2 clade and have a leucine substitution, and 11 other variants are members of the HCP6 clade, six of which have a lysine in this position, conserving the positive charge (Supplemental Data 1).

We predicted that the HCPs will fold into the two four-helix bundles characteristic of the NTD of the OCP; differences in the primary structures may contribute to clade-specific biochemical characteristics and functions. Bulk biochemical properties for each protein were predicted from the amino acid compositions of the HCPs and were compared with those of the OCP-NTD (summarized in Supplemental Figure 4). Notably, the members of the HCP1 clade are enriched in aromatic residues, particularly in contrast with the NTD of the OCP, which resembles the other major HCP clades in this feature. Also

***Nostoc* sp. PCC 7120**		***Chlorogloeopsis* PCC 9212**	***Nodularia* CCY9414**	***Anabaena variabilis* ATCC 29413**	***Dolichospermum* AWQC131C & AWQC310F**
Ow et al., 2008	Flaherty et al., 2011	Zhang and Bryant, 2015	Voß et al., 2013	Park et al., 2013	D'Agostino et al. 2014
Proteome	Transcript	Transcript	Transcript	Transcript	Proteome
N deprivation, heterocyst vs filaments	N deprivation	Autotrophic vs mixotrophic (acetate supplement)	Standard conditions (N_2 fixing) for start site profiling	Photo vs hetero vs mixotrophy (fructose supplement)	P-deprivation vs N deprivation; toxin production
Lower in NH_4^+ filaments (vs N_2-grown); low counts, slightly above HCP1	Increase after 21 h N deprivation	Lower with acetate supplement	Present (NCBI short read archive SRX23146)	Moderate expression	Lower in 131C (saxitoxin strain) vs 310F; less than HCP4 overall
Up in NH_4^+ (vs N_2-grown) filaments; vegetative and heterocyst equal; low counts	Increase (becomes detectable) after 21 h	Higher with acetate supplement	Not present in genome	Higher in mixotrophic (vs heterotrophic) conditions	Not present in genome
Not reported	Low number of reads	levels unchanged	Not present in genome	Not present in genome	Not present in genome
Not reported	Decrease after 21 h N deprivation	Higher with acetate supplement	Not present in genome	Higher in mixotrophic (vs heterotrophic) conditions	Not present in genome
Not reported	Increase after 12 h N deprivation	lower with acetate supplement	Present (NCBI short read archive SRX23146)	Low expression	Lower in 131C (saxitoxin strain) vs 310F; higher than OCP overall
Not present in genome	Not present in genome	Not present in genome	Not present in genome	Not present in genome	Not present in genome
Not present in genome	Not present in genome	Not present in genome	Not present in genome	Not present in genome	Not present in genome
Not present in genome	Not present in genome	Not present in genome	Not present in genome	Not present in genome	Not present in genome

notable is the observation that members of the HCP4 and HCP5 clades, like the NTD of the OCP, all have a significantly decreased composition of acidic residues compared with the other HCP clades, with an extremely narrow distribution among the HCP4 sequences. HCP4s are also predicted to have a significantly higher isoelectric point (the HCP5 members show this as well but with a much wider distribution). Collectively, these differences likely contribute to distinct functions for the different clades (e.g., altered affinity for a binding partner or interaction with different binding partners).

Global expression studies in cyanobacterial species containing HCPs (previously identified as hypothetical proteins) were re-analyzed (Table 2) in the context of the identified clades. Representatives of all of the clades except HCP7 and HCP9 have been identified in various expression studies, indicating that HCPs maintain some functionality, and thus are not relic sequences. In some cases, in organisms containing multiple paralogs, expression patterns were different among the HCPs and contrasted with that of the OCP. For example, HCP5 transcripts showed inverse dynamics (vs HCP8 and OCP) during N-deprived H_2 production in *Cyanothece* ATCC 51142 (Bernstein et al., 2015). Interestingly, the levels of HCP8 in this strain went up during the dark period of diurnal light–dark cycling experiments (Toepel et al., 2009). HCP3 showed the highest peptide abundances among all HCPs in *Nostoc punctiforme* proteomes (Anderson et al., 2006), and transcripts were upregulated in hormogonia (Christman et al., 2011), whereas OCP and HCP1 transcripts in this strain showed downregulation during N-deprivation. In *Nostoc* 7120, one study showed a decrease of HCP3 but an increase of OCP, HCP1, and HCP4 during the first hours of N deprivation

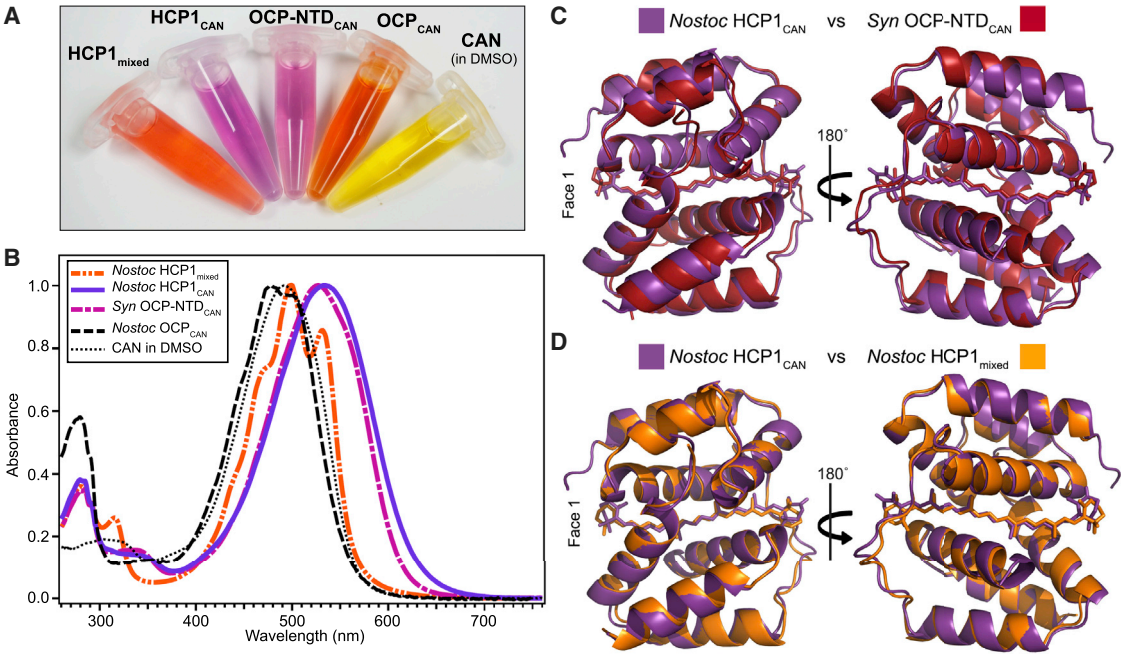


Figure 3. Spectroscopic Properties and Structures of HCPs, Including *Nostoc* HCP1.

(A) Visible appearance of carotenoids and soluble carotenoid proteins described in this work. From left to right: *Nostoc* $HCP1_{mixed}$ (expressed in *Synechocystis*), *Nostoc* $HCP1_{CAN}$ (expressed in CAN-producing *E. coli*), *Synechocystis* $OCP\text{-}NTD_{CAN}$ (Leverenz et al., 2015), *Nostoc* OCP_{CAN} (López-Igual et al., 2016), CAN in DMSO.

(B) UV-visible absorbance spectra of samples in **(A)**.

(C) Structural alignment of *Nostoc* $HCP1_{CAN}$ and *Synechocystis* $OCP\text{-}NTD_{CAN}$ (Leverenz et al., 2015) showing similar folds and CAN bound in a CPC-R-like configuration.

(D) Alignment of *Nostoc* HCP1 binding either homogeneous CAN or a mixture of carotenoids (represented by β-carotene), showing that *Nostoc* HCP1 is observed in a similar fold regardless of the identity of the bound carotenoid.

(Flaherty et al., 2011), while a second study showed an upregulation of OCP and HCP1 under nitrogen starvation (when compared with cultures in the presence of ammonium) (Ow et al., 2008). In *Chlorogloeopsis* PCC 9212, HCP4 and OCP transcript levels decreased during photomixotrophic growth (supplemented with acetate), whereas HCP1 and HCP3 levels went up under these conditions (Zhang and Bryant, 2015). This pattern of upregulation of HCP1 and HCP3 under photomixotrophic conditions was corroborated in studies with *Anabaena variabilis* ATCC 29413 (Park et al., 2013) (which is a closely related sister taxon to *Nostoc* 7120). Proteomic analyses of two strains of *Dolichospermum* under nutrient deprivation both showed higher peptide abundance levels for HCP4 compared with OCP (D'Agostino et al., 2014). HCP2, when detected, showed low levels and did not exhibit any dynamic nor differential expression based on environmental perturbations. Collectively, these results support the hypothesis that the different HCP subclades have distinct functions.

Carotenoid Binding by the HCP1 (All1123) from *Nostoc* 7120

To test the prediction that HCPs bind carotenoid, a *Nostoc* 7120 mutant containing a His-tagged HCP1 protein (All1123) was constructed. Subsequent purification of this HCP1 (hereafter *Nostoc* $HCP1_{Native}$) by immobilized metal ion affinity chromatography (IMAC) yielded a small amount of pigmented protein. The UV-visible spectrum of *Nostoc* $HCP1_{Native}$ (Supplemental Figure 5) was reminiscent of the NTD-OCP from Syn6803 (Leverenz et al., 2014, 2015). An extract analyzed by LC-MS was shown to contain 45%–63% canthaxanthin (CAN), 20%–27% echinenone (ECN), 9%–16% β-carotene, and 7%–10% 3′-hydroxyechinenone. Accordingly, we confirm that *Nostoc* HCP1 binds carotenoids, predominantly C40 carotenoids with 4-keto ring substitutions, with the highest level of binding observed for CAN.

Because the yield of *Nostoc* $HCP1_{Native}$ was insufficient for crystallization, the *all1123* gene was overexpressed with a His-tag in a ΔCrtR Syn6803 mutant (producing both CAN and ECN) and in a background of *Escherichia coli* producing 100% CAN. LC-MS of extracts of the protein purified from the Syn6803 background (hereafter $HCP1_{mixed}$) identified a mixture of deoxymyxoxanthophyll (56%–60%), ECN (15%–20%), CAN (2%–8%), and β-carotene (14%–16%). In contrast, only CAN could be detected in extracts of the protein isolated from the *E. coli* background (hereafter $HCP1_{CAN}$). The visible appearance of $HCP1_{mixed}$ was distinctly different from that of $HCP1_{CAN}$ (Figure 3), with multiple peaks observed in the UV-visible absorbance spectrum of $HCP1_{mixed}$, characteristic of carotenoid vibronic structure (Figure 3B). In contrast, the $HCP1_{CAN}$ spectrum lacked these bands and was significantly red shifted, resembling the synthetically truncated Syn6803 $NTD\text{-}OCP_{CAN}$ (Figure 3B). Hydrophobic interaction chromatography (HIC) was used to

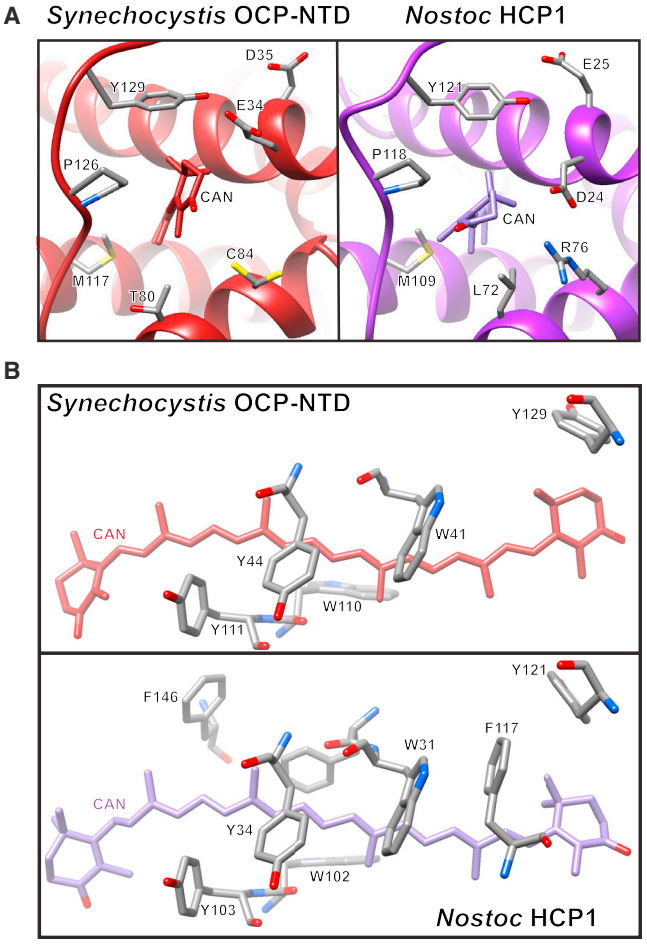


Figure 4. Comparison of Carotenoid-Binding Residues in *Synechocystis* OCP-NTD$_{CAN}$ and *Nostoc* HCP1$_{CAN}$.
(A) CPC-R-like binding is observed in *Nostoc* HCP1$_{CAN}$. Similar 4′-keto ring (B2 ring) binding residues are observed in both structures (with the C84/R76 substitution being a notable difference), as is the *s-trans* configuration of the 4′-keto ring of CAN.
(B) Aromatic residues within 5 Å of the carotenoid constitute a key difference in the binding pocket of the polyene.

separate the apo- and holo forms of the HCP1$_{CAN}$ and produced a homogeneous holoprotein suitable for crystallization and subsequent structural analysis (see Supplemental Figure 6).

Structures of the HCP1 of *Nostoc 7120*

We grew crystals from *Nostoc* HCP1$_{mixed}$ and determined its three-dimensional structure to 2.5 Å resolution (see Methods and Supplemental Table 2 for additional details). The protein crystallized with one molecule per asymmetric unit. Only sparse electron density was observed for the portion of the carotenoid that is exposed at face 1 of the protein (see Supplemental Figure 7). While the electron density of the β-ionyl ring at face 2 is well defined, there is not enough information to conclusively identify the carotenoid, which we modeled as β-carotene because its structural features are shared among the three other carotenoids detected by LC-MS.

Nostoc HCP1$_{CAN}$ was crystallized and the structure was determined to 3.2 Å resolution. An interaction between the two separate protein chains in the asymmetric unit of HCP1$_{CAN}$ crystals occurred across face 1 (see Supplemental Figure 8) indicating proteolysis of a loop connecting the αC and αD helices of the HCP (a region observed to be poorly ordered in OCP-NTD structures). This suggests that the dimerization observed in the crystal is unlikely to be physiological, as the presence of the loops in the native protein would prevent such an association. Moreover, no polar nor salt-bridging interactions were observed at the dimer interface; the interaction appears entirely hydrophobic in nature. The interaction along face 1 includes a coplanar stacking of the 4-keto rings of the CAN molecules, ~ 5 Å in distance. Both *Nostoc* HCP1 structures align well with the recently reported structure of the Syn6803 OCP-NTD$_{CAN}$ (Leverenz et al., 2015), with RMSD of 1.18 and 0.98 Å over 113 and 109 α carbon pairs for HCP$_{CAN}$ and HCP$_{mixed}$, respectively (Figure 3C).

Clear electron density is observed for the noncovalently bound CAN molecule in each chain in the HCP1$_{CAN}$ structure (see Supplemental Figure 7). The CAN is observed to bind in a CPC-R-like binding pocket in the *Nostoc* HCP1$_{CAN}$ structure (Figure 4), with the carotenoid in a similar configuration to that observed in the NTD of the OCP structure (Leverenz et al., 2015). An *s-trans* configuration of the 4′-keto ring of CAN is likewise observed in both structures. Positional sequence conservation across all HCP sequences was mapped onto the HCP1$_{CAN}$ structure (Figure 5), confirming the strong conservation of the amino acids that line the carotenoid-binding pocket and supporting the prediction that carotenoid binding is a conserved structural feature of HCPs. In HCP1, these carotenoid–protein interactions include the presence of two additional aromatic residues (F146 and F117, Figure 4B); F146, in helix αI, is highly conserved in HCP1 sequences (Figure 2; Supplemental Figure 3). While the polyene is completely surrounded by the hydrophobic binding pocket between the two 4-helix bundles, the β-ionyl rings are solvent-exposed in the HCP1 structures.

DISCUSSION

The HCPs Are a Family of Cyanobacterial Carotenoid-Binding Proteins

Here we have shown that at least nine different clades of single-domain homologs to the NTD of the OCP are encoded in cyanobacterial genomes. Due to the consistently strong sequence conservation, particularly among amino acids homologous to the carotenoid-binding tunnel residues of the NTD of the OCP (Figures 2B and 5), we propose that all of these proteins bind carotenoids. Structurally, we predict and demonstrate by crystal structure determination (Figure 3C) that HCPs share the all-helical fold of the pfam09150 domain/the RCP (Leverenz et al., 2015). Literature mining for expression data (Table 2) suggests that these proteins likely have diverse functions. The HCPs constitute a new family of carotenoid-binding proteins.

Two structures of the *Nostoc* HCP1 were solved from different protein preparations with distinct optical properties and different carotenoid content; nevertheless, the tertiary structure of the protein remained nearly identical (RMSD of 0.47 or 0.46 Å for chain A or B of HCP$_{CAN}$, respectively) over 119 α carbon pairs, Figure 3D). The structures demonstrate that the hydrophobic binding pocket

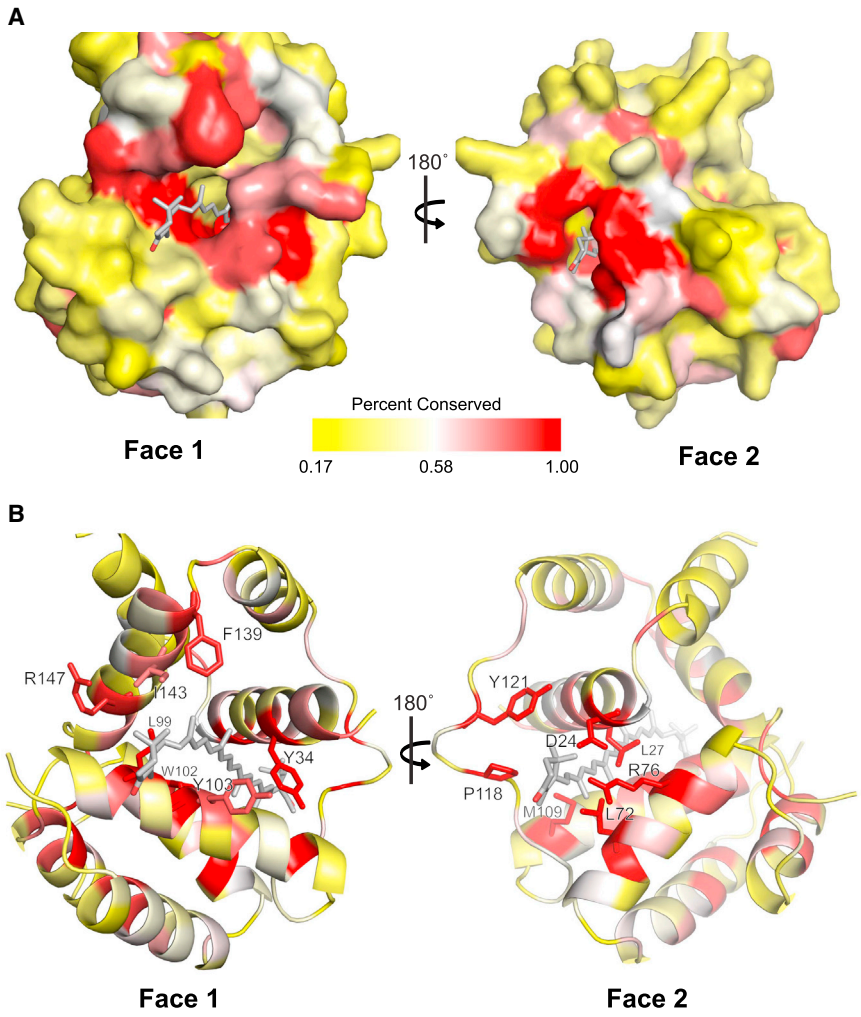


Figure 5. Amino Acid Sequence Conservation of HCPs Mapped on the *Nostoc* HCP1 Structure.

Positions on the *Nostoc* $HCP1_{CAN}$ structure are shown as **(A)** a space-filling model and **(B)** a cartoon model of secondary structure elements, with important residues labeled according to All1123 numbering. Conservation of positions was colored according to percent identity, from least conserved (yellow) to most conserved (red) in accordance with the color bar.

between the two four-helix bundles of HCPs can indeed support binding of a variety of carotenoids in the CPC-R-like configuration observed in the activated red form of the OCP (OCP^R) (Leverenz et al., 2015) (Figure 6). The solvent exposure of the β-ionyl ring region(s) of the carotenoid in the HCP1 structures likely contributes to the apparent preference for the binding of polar carotenoids (e.g., xanthophylls), as confirmed by carotenoid content analyses of HCP1. Moreover, our results suggest that the different HCP subfamilies represent an ancient divergence suggestive of functional diversification, reflected in clade-specific conservation of features of their primary structures (Supplemental Figures 3 and 4, Figure 2), which may influence their relative affinities for different carotenoids.

Our data indicate that HCP1 is competent in binding more than one type of carotenoid, leading to differences in optical properties (Figure 3B). It is possible that (in addition to its specific primary structure), the chemical identity of the carotenoid may also influence the functional properties of an HCP, as it does in the OCP (Punginelli et al., 2009; Wilson et al., 2011; Bourcier de Carbon et al., 2015). We posit that this flexibility in binding different carotenoids may be one way to fine-tune the functional properties of an HCP. For instance, the presence of an intramolecular charge transfer state in 4-keto carotenoids has been suggested to play a role in the quenching function of OCPs (Berera et al., 2012), and this could extend to the HCPs as well. *In vivo*, cells might be able to maintain or control the function of an HCP by regulating the occupancy (i.e., which carotenoid is bound) by availability and/or relative abundance of particular carotenoids. Although speculative, the carotenoid can be thought of as an interchangeable module for regulating the function of an HCP (Figure 6C); this remains to be tested.

Our results also demonstrate that the protein environment influences the properties of the carotenoid. This is apparent when comparing the absorbance spectrum of CAN in organic solvent to CAN bound by the Syn6803 OCP-NTD or by *Nostoc* $HCP1_{CAN}$ (Figure 3B), with the latter two significantly red shifted. Likewise, the absorbance maximum of $HCP1_{mixed}$ (500 nm) is similarly red shifted from that of free myxoxanthophyll (477 nm) (Mohamed and Vermaas, 2004). Thus, a red shift of the carotenoid's ground-state absorbance appears to be a general indicator of CPC-R-like carotenoid binding in these proteins. The protein-induced red shift in CAN's absorbance is also slightly larger in HCP1 compared with the OCP-NTD; the additional aromatic binding residues (Figures 4B and 5) or non-conservative substitutions in the vicinity of the 4′-keto ring (e.g., C84 [OCP-NTD] to R76 [HCP1]) (Figure 4A) could contribute to this difference in the spectra by subtly altering the electrostatic environment and the structure of the CAN molecule. The underlying molecular basis of this red shift awaits elucidation by advanced spectroscopic and computational studies. In much the same way as soluble phycobiliproteins have become valuable scaffolds for studying the effects of a protein environment on a pigment's spectroscopic properties (Wang and Moerner, 2015), the HCPs should be a valuable tool for studying the consequences of a protein matrix on the function and properties of carotenoids, as well as the influences different carotenoids may exert upon a particular protein matrix. The single chromophore per protein chain as well as the solubility of the OCP and the HCPs makes them an ideal system for these studies, in comparison with the majority of carotenoid-binding antenna complexes, which are membrane bound and thus difficult to manipulate.

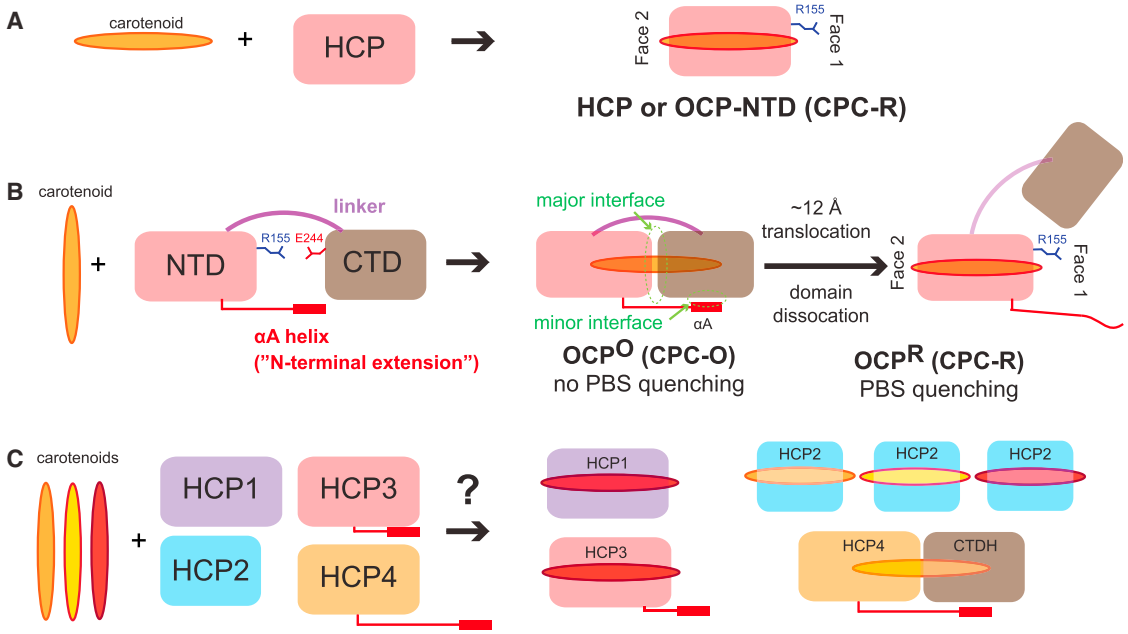


Figure 6. Proposed Modularity of Soluble Carotenoid Proteins.
(A and B) Illustration of Carotenoid Binding and Modular Assembly of HCP **(A)** and OCP **(B)**. The carotenoid is buried entirely within the pfam09150 domain in CPC-R (in HCP, OCP-NTD, and OCP^R), whereas it is simultaneously bound by the NTD and CTD in CPC-O.
(C) The different HCP subtypes are suggested to have different biochemical properties such as carotenoid preference, optical properties, physiological role, and/or interaction partners. As indicated by the question mark, it is hypothesized that some HCP subtypes may bind different carotenoid modules and that a heterocomplex may form between CTDH and an HCP containing the N-terminal extension, resembling extant OCP.

Predicted Functional Diversity Based on Differential Expression and Primary Structure

The pfam09150 domain (as either the NTD of the OCP or as an HCP) is confined to the Cyanobacteria. This domain is widespread throughout the extant taxa of this phylum; HCPs are present in the majority of cyanobacterial species (Table 1) and in all phylogenetic subclades except C1 (alpha cyanobacteria) and C2 (*Synechococcus elongatus*). Multiple HCPs are found even in the *Gloeobacter* species, which are commonly viewed as being the most ancient lineage among modern cyanobacteria, preceding the last common ancestor of the chloroplast (Criscuolo and Gribaldo, 2011; Shih et al., 2013). The majority of genomes that lack both the OCP and HCP are those of atypical cyanobacteria; this includes the Prochlorales, which do not use PBS for their light-harvesting antenna (Partensky and Garczarek, 2010), as well as many symbiotic species such as strains of *Richelia*, *Synechococcus spongiarum*, *Moorea*, *Acaryochloris*, and *Atelocyanobacterium*, which exhibit either genomic streamlining attributed to co-evolution with their host (Thompson et al., 2012; Hilton et al., 2013; Gao et al., 2014) or genome expansion specifically attributed to niche partitioning (Swingley et al., 2008; Engene et al., 2012); in addition, some of these strains have atypical PBS structures. The presence of fragmented OCP-like open reading frames detectable by T-BLASTN in several genomes that lack OCP-like homologs (for instance, *Moorea* 3L, *Cyanobacterium aponinum* PCC 10605, *Synechococcus* CC9616, *Leptolyngbya* BDU 20041) suggests degradation of previously functional genes and supports an ancestral presence of OCP and/or HCP in these strains. The well-studied strain *Synechococcus elongatus* PCC 7942 and its closely related sister taxon PCC 6301 also lack OCP/HCP-homologs; although atypical physiological features are not apparent in these strains, they exist on a long branch (phylogenetic subclade C2) and thus have divergent genomes. Among the 15 genomes containing an HCP but lacking an OCP (Supplemental Table 1), the majority contain an HCP4/HCP5-like sequence with a CTDH gene at a nearby locus (Supplemental Figure 2), suggesting that perhaps some functional aspects of the OCP could be provided by a heterodimeric interaction between a CTDH and an HCP4/HCP5 homolog (Figure 6C; see below). This is consistent with generally accepted models for protein evolution (Enright et al., 1999; Marcotte et al., 1999). On the other hand, in light of the absence of blue-light-inducible phycobilisome quenching in one of these strains, *Thermosynechococcus elongatus* BP-1 (Boulay et al., 2008), the function of a heterodimeric HCP–CTDH complex may differ substantially from the photoprotective mechanism of the OCP, or may be operational only under particular conditions.

There does not appear to be any biased co-occurrence between the major HCP subtypes within genomes (Figure 1D and Supplemental Figure 1). Nevertheless, the genomes with the largest inventory of HCP paralogs are filamentous, heterocystous N_2-fixing species (subsections IV and V) (Table 1); however, these strains also tend to have the largest/most expanded genomes among cyanobacteria. Nevertheless, it is a reasonable hypothesis that there exists an association between HCP function and N_2 fixation and/or morphological differentiation. Recently, it was shown that the HCPs in *Nostoc*

7120 are expressed in both vegetative as well as heterocyst cells (López-Igual et al., 2016). HCP1 was downregulated (together with OCP) during an N-deprivation time course in *Nostoc punctiforme* ATCC 29133, whereas HCP3 was shown to be highly upregulated in hormogonia (Christman et al., 2011) (Table 2). Interestingly, an opposite trend was shown for the same HCP subtypes in similar expression studies with *Nostoc* 7120, with increased expression of OCP, HCP1, as well as HCP4 upon the first hours of N deprivation (Flaherty et al., 2011). This discrepancy between strains might derive from genetic changes in *Nostoc* 7120 occurring over decades of lab cultivation (this strain has lost its ability to form akinetes or homogonia; Jack Meeks, personal communication) or from nuances in experimental methods between laboratories. Nevertheless, these differential expression patterns for the different HCPs in each study do support the proposition that the members of the different clades have distinct functions.

In contrast to the highly conserved primary structure of the OCP, the diversity of HCPs reflects a large amount of gene duplication and divergence, reflecting subfunctionalization, which has led to their maintenance in so many cyanobacterial genomes (Figure 1). It is notable that even the evolutionarily basal genus *Gloeobacter* contains both OCP and multiple NTD homologs (Kerfeld, 2004a), although they exhibit relatively greater divergence in primary structure (e.g., HCP9 in Figure 2B; Supplemental Figure 3). Furthermore, HCPs can be found in all four members of phylogenetic subclade F (Table 1), which was also shown to branch earlier than the ancestor of the chloroplast (Shih et al., 2013). Thus, it appears that HCPs are ancient proteins that may have had essential roles because they were largely conserved and they are still present in diverse, extant cyanobacteria species. Thus, it appears that the evolution of HCPs (and possibly the OCP-related photoprotective mechanism) may predate the rise of modern cyanobacteria.

Inspection of published proteomic and transcriptomic datasets shows that at least some of the HCP paralogs (identified as hypothetical proteins) are expressed. Notably, all clades were represented by at least one HCP subtype in global expression studies, with the exception of HCP7 and HCP9, which are not encoded in the genomes of any of the organisms used in these studies (Ow et al., 2009; Christman et al., 2011; Park et al., 2013; Voß et al., 2013; D'Agostino et al., 2014; Bernstein et al., 2015; Zhang and Bryant, 2015) (Table 2). Furthermore, a meta-analysis of these data clearly shows differential patterns of expression in response to various growth conditions. For instance, HCP5 was strongly upregulated (but OCP was down-reguated) during the early phase of H_2 evolution in *Cyanothece* ATCC 51142 (Bernstein et al., 2015), which was induced by abrupt starvation of all external electron acceptors (CO_2, N_2) in NH_4^+-deprived cultures. This early phase of H_2 production was characterized by a transient decline of photosynthetic capacity (Melnicki et al., 2012) associated with elevated ROS production (Natalie Sadler, unpublished results). Similarly, photoheterotrophic supplementation with organic carbon to cultures of *Chlorogloeopsis* PCC 9212 and *Anabaena variabilis* ATCC 29413 led to higher levels of HCP1 and HCP3 transcripts, with *Chlorogloeopsis* also showing decreased levels of both HCP4 and OCP in response to organic carbon supplementation (Park et al., 2013; Zhang and Bryant, 2015). Collectively, these data suggest that HCPs have distinct functions in cells.

The HCP4 Clade and the Evolution of the OCP

Because the *Gloeobacter* HCPs cluster together and are located farthest from the OCP cluster (Figure 1), it can be inferred that the pfam09150 domain was ancestrally derived from an HCP-like protein. The HCP4 clade is found closest to the OCP-like sequences (Figure 1A) and has very short branches, indicating strong conservation of primary structure. Interestingly, nearly all of the genomes with an HCP4 representative also encode an OCP homolog (Table 1), suggesting that a photoprotective role for HCP4 would not be identical with that of OCP. HCP4 representatives are only found in species belonging to phylogenetic subclade B1, which is characterized by the filamentous, heterocystous morphotypes (subsections IV and V); these organisms commonly exhibit complex life cycles such as cell specialization, biofilm formation, anoxic diazotrophy, and chromatic adaptation of the phycobilisome antenna. Thus, perhaps the retention and strong conservation of HCP4 in these strains is associated with particular physiological demands. We can speculate that there may be situations in which indiscriminant or long-duration PBS quenching could be advantageous; for instance, conditions of stress leading to extreme photosensitivity (e.g., desiccation of desert crusts) or upon differentiation of a vegetative cell into specialized cell types no longer performing oxygenic photosynthesis (e.g., heterocysts).

Several additional observations support the contention that the HCP4 clade is ancestral to the NTD of the OCP. For example, 23 of the 26 members of the HCP4 clade are encoded at a locus adjacent to an OCP-CTD homolog (CTDH) (Supplemental Figure 2). Furthermore, HCP4 members tend to have a ~20 amino acid extension with strong sequence similarity to the OCP N-terminal extension α helix (αA in Figure 2A); this extension helps stabilize the closed inactive OCP^O form by interacting with the CTD at the minor interface (Figure 6B). Moreover, the position of the HCP4 clade is furthest from the presumably ancestral HCP9 clade, and it is also the closest to the OCP clade. The high degree of conservation in this clade (short branches in Figure 1A) provides strong evidence for evolutionary constraint, implying a structure–function relationship that cannot be altered without a cost to fitness. Thus, we suggest that at least some of these tandem HCP4–CTDH pairs may interact to form a heterodimer that could possibly constitute a primitive, modular OCP function (as depicted in Figure 6C). The interaction with the CTD might have enabled the acquisition of regulation of the HCP activity by light and the FRP.

Due to the similarity in tertiary structure between the HCPs and the NTD-OCP (Figures 3C and 5) and the conservation of Arg155 (essential for PBS binding) (Figure 2B), it is possible that the HCPs can interact with PBSs. Unless tightly regulated at the level of expression (or by an alternate regulatory mechanism), the HCPs would presumably then be constitutively active in energy quenching. This interaction with the PBSs could lead to irreversible quenching of PBSs even in light-limited conditions; such a function could be desirable for strains

experiencing dehydration while growing on exposed surfaces, such as desert soil crusts (Rajeev et al., 2013).

Accordingly, a plausible scenario for the evolution of OCP-mediated photoprotection involves a series of steps beginning with an HCP and including recruitment of a CTDH, subsequent domain fusion with an ancestor of the HCP4 clade, and finally recruitment of the FRP. An interaction with the CTDH (Figure 6C) and, subsequently, its interaction with the FRP, might have conferred increasingly fine regulatory control on the quenching activity of the OCP-NTD, which alone can be a constitutively active quencher (Leverenz et al., 2014, 2015). Recently it was demonstrated that the HCP4 from *Nostoc* (but not its HCP1, HCP2, or HCP3) indeed acts as a constitutive quencher of PBSs *in vitro* (López-Igual et al., 2016). Accordingly, we suggest that the formation of an HCP4–CTDH complex would represent an intermediate form in the evolution of contemporary OCP. The initial establishment of such an association could have served as a novel way to shut off a constitutively quenching HCP and thereby minimize wasteful energy dissipation from the antenna under favorable environmental conditions. It is well accepted that one widespread mechanism for protein evolution is domain fusion; proteins that functionally interact can become genetically fused into a single polypeptide (Enright et al., 1999; Marcotte et al., 1999). The fusion of an ancestral HCP4 to a CTDH could have thus contributed a regulatory event by providing its portion of CPC-O residues to stabilize the carotenoid in its orange, inactive form, leading to the ability to convert from OCP^O to OCP^R (Figure 6B), effectively leading to "turning on" quenching when necessary as opposed to turning it off. The innovation of OCP photoactivation, which has quite a low quantum yield and thus requires strong (blue) light, perhaps provided some regulation for the activation of the OCP for quenching only when light was in excess. Essentially, we propose that fusion to a CTDH resulted in the gaining of a sensor domain. Finally, recruitment of the FRP and its interaction with the CTD would have afforded a new way to "shut off" (or minimize) quenching under conditions when it would be energetically wasteful. This hypothesis is supported by the observation that nearly all of the full-length OCP sequences occur in genomes that also contain an FRP (Figure 1C), usually located one or two loci away from the OCP gene.

Summary and Outlook

This model for the evolution of the OCP from phylogenomic data explains the structural and functional modularity of this recently identified photoreceptor. Structural and functional modularity is also a hallmark feature of several classes of photoreceptor proteins such as phytochromes and light-oxygen-voltage sensors (Möglich et al., 2010). These proteins consist of a regulatory domain with a sensory function (e.g., the detection of light) and a domain with an effector function (e.g., the selective interaction with a small molecule or another protein that effects a physiological response). It is plausible that one or more of the distinct clades of HCP paralogs could function as an alternate "effector" module in an HCP–CTDH heterodimer. Moreover, our structural analysis confirms that the predicted carotenoid-binding residues are able to support interaction with diverse carotenoids. The factors governing relative affinity for different carotenoids remain to be elucidated; the carotenoid could constitute a third interchangeable module that could influence function. In addition to providing a tractable model system for biophysical and spectroscopic analysis of carotenoid–protein interactions, future detailed studies of HCP expression, the effect of HCP gene knockouts, and *in vitro* activity assays will be critical in establishing the precise functional role(s) of these proteins and their potential uses in optogenetics.

METHODS

Phylogenetic Analysis

Protein sequences pre-assigned to contain the pfam09150 domain were retrieved from IMG on March 30, 2015, and were supplemented by BLAST searches within IMG using all1123 as a seed sequence and 1.0 as the E-value cutoff (Markowitz et al., 2012), resulting in 392 OCP-related sequences (Supplemental Data 1). Twenty-four sequences were removed because they belonged to duplicate genome entries in IMG; rationale for genome de-duplication is documented in Supplemental Data 2. An initial alignment with MUSCLE (Edgar, 2004) was used for manual curation; 11 sequences were removed due to poor alignment or incomplete sequence coverage. The remaining 357 sequences were realigned using MUSCLE, and the NTF2-containing CTD was removed from full-length OCPs using Jalview (Waterhouse et al., 2009) to delete all columns to the right of the MG(F/Y)D motif, which marks the end of the pfam09150 domain and the beginning of the inter-domain linker.

The HCPs and truncated OCP-NTDs were again realigned in MUSCLE, as well as with MAFFT (Katoh and Standley, 2013) (run once with the progressive G-INS-1 method and leavegappyregion option, and again with the iterative G-INS-i method), and Probcons (Do et al., 2005). A 50% consensus alignment was obtained from these four MSAs using MergeAlign (Collingridge and Kelly, 2012), maintaining 95 positions for columns that were present in at least two of the four alignments (the 100% consensus preserved only 18 positions). This consensus alignment was used to build a phylogenetic tree using PhyML 3.1 (Guindon et al., 2010) with the LG substitution model, NNI tree topology search settings, and the aBayes algorithm for branch supports. Clades were determined by manual inspection and were colored using Archaeopteryx (http://www.phylosoft.org/archaeopteryx). The resulting phylogenetic tree is available in phyloxml format in Supplemental Data 3 (Han and Zmasek, 2009). HMMs were built with new MUSCLE alignments for each clade using hmmbuild (HMMER 3.0/b, http://hmmer.org), and sequence logos were calculated from these HMMs using Skylign (Wheeler et al., 2014). The consensus sequences for each HMM were aligned in MUSCLE. Biochemical properties of each sequence were predicted using EMBOSS-PEPSTATS (Rice et al., 2000). Area-proportional Euler diagrams for gene co-occurrence within genomes were drawn with VennMaster (Kestler et al., 2008), using Boolean genome incidences, as pivoted from data in Supplemental Data 1.

Cloning of *Nostoc* HCP1 in *Nostoc* sp. PCC 7120

The plasmid for the overexpression of *Nostoc* HCP1 in *Synechocystis* 6803, described previously (Sedoud et al., 2014), was digested using NheI. The resulting 620-bp fragment, containing the *all1123* 3′ sequence with His-tag fusion and Km resistance gene, was cloned in the pRL277 plasmid (Cai and Wolk, 1990), digested with SpeI, and transferred to *Nostoc* 7120 by conjugation using HB101 strains carrying helper and methylation plasmid pRL623 (Elhai et al., 1997). The conjugation was effected by the conjugative plasmid pRL443, carried in *E. coli* ED8654 (provided by Prof. C.P. Wolk) and performed as described previously (Elhai et al., 1997), with selection for Sm/Sp resistance. We obtained a KmSmSp-resistant mutant and confirmed homologous recombination of the His-tag fusion at the native *all1123* locus, as verified by sequencing.

Cloning of *Nostoc* HCP1 in Canthaxanthin-Producing *E. coli*

The *all1123* gene with a sequence encoding for 6xHis in 3′ was amplified by PCR using the restriction site-creating primers 1123NcoI (5′-GAT ATA CCA TGG GCA CTT TTA CAC AAA CAA ACG-3′) and 1123NotI (5′-CAT TAT GCG GCC GCC TAA TGA TGA TGA TGA TGA TG-3′) (and a previously described plasmid containing the *all1123* gene under the control of the *psbA2* promoter as template; Sedoud et al., 2014). The PCR fragment was digested and cloned onto the NcoI and NotI site of pCDFDuet-1 to create the pCDF-HCP-1123Ctag plasmid. The production of holo-protein in a canthaxanthin-producing strain of *E. coli* was described previously (Bourcier de Carbon et al., 2015).

HCP1 Protein Purification

The 6xHis-tagged *Nostoc* $HCP1_{mixed}$ was purified from ΔCrtR *Synechocystis* 6803 by IMAC as described previously (Sedoud et al., 2014), and the same protocol was followed for the 6xHis-tagged $HCP1_{Native}$ in *Nostoc* 7120. The 6xHis-tagged *Nostoc* $HCP1_{CAN}$ was purified from canthaxanthin-producing *E. coli* by IMAC using a protocol similar to that described in Bourcier de Carbon et al. (2015), resulting in a mixture of apo- and holo-HCPs. The apo and holo forms of the protein were separated by HIC as follows: the apo/holo mixture was exchanged into 50 mM Tris-HCl (pH 7.4), 1.5 M $(NH_4)_2SO_4$ using multiple cycles of concentration and buffer exchange in an Amicon 15 centrifugal concentrator (10 kDa MWCO; Millipore). The protein was then loaded on a Tosoh TSK-Gel Phenyl 5PW column and a decreasing linear gradient of 1.5 M $(NH_4)_2SO_4$ to 0 M $(NH_4)_2SO_4$ in 50 mM Tris-HCl (pH 7.4) was used to elute bound protein. Major fractions containing $HCP1_{CAN}$ holoprotein (as judged by the 510 nm:280 nm absorbance ratio in the UV-visible chromatogram; Supplemental Figure 6) were pooled and further purified by SEC on a Superdex 75 pg HiLoad 16/60 column (GE Healthcare) using 50 mM Tris-HCl (pH 7.6), 200 mM NaCl running buffer.

Measurement of Carotenoid Content in HCPs

The carotenoid content of the isolated HCPs was analyzed by high-performance liquid chromatography and mass spectrometry as previously described (Sedoud et al., 2014).

Protein Crystallization and Structure Determination

Crystals of *Nostoc* $HCP1_{CAN}$ were obtained from sitting drop experiments at room temperature by mixing 2 μL of protein solution (3 mg/mL in 10 mM Tris-HCl [pH 7.4]) with 1 μL of reservoir solution containing 400 mM ammonium citrate, 20% PEG 3350. Crystals were stabilized by adding 30% ethylene glycol (in reservoir solution) to the drop, mounted on a nylon loop (CrystalCap ALS HT, Hampton Research), and then frozen in liquid nitrogen.

Crystals of *Nostoc* $HCP1_{mixed}$ were obtained from sitting drop experiments at room temperature by mixing 0.2 μL of protein solution (4 mg/mL in 20 mM NaCl, 20 mM Tris-HCl [pH 7.6]) with 0.2 μL of reservoir solution containing 100 mM HEPES (pH 7.5), 10% 2-propanol, 25% PEG 4000. Crystals were transferred to nylon loops under oil (to prevent drop evaporation) and were frozen directly in liquid nitrogen.

X-ray diffraction was measured at beamlines 5.0.2 and 5.0.3 of the Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory. Diffraction data were integrated with XDS (Kabsch, 2010) and scaled with SCALA (CCP4) (Winn et al., 2011). The HCP structures were solved by molecular replacement with the NTD of Syn6803 OCP (PDB: 3MG1). Automatic building of a model into the density using phenix.autosol (Afonine et al., 2012) was followed by manual rebuilding/refinement cycles using COOT (Emsley and Cowtan, 2004) and phenix.refine (Afonine et al., 2012). Statistics for diffraction data collection, structure determination, and refinement are summarized in Supplemental Table 2.

ACCESSION NUMBERS

Sequence data from this article can be accessed from the IMG database using the gene object identifiers (gene_oid) listed in Supplemental Data 1 (available as an .xls file in the online version of the article).

Structural coordinates have been deposited in the PDB (http://www.rcsb.org/pdb) under the following accession codes: PDB: 5FCX (*Nostoc* $HCP1_{CAN}$) and PDB: 5FCY (*Nostoc* $HCP1_{mixed}$).

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

The work in the Kerfeld Laboratory was supported by the National Science Foundation (IOS 1557324) and by the Office of Science of the U.S. Department of Energy DE-FG02-91ER20021. The research in D.K.'s laboratory was supported by grants from the Agence Nationale de la Recherche (ANR, project CYANOPROTECT), the Centre National de la Recherche Scientifique (CNRS) and the Commissariat à l'Energie Atomique.

AUTHOR CONTRIBUTIONS

Conceptualization, M.R.M., R.L.L., D.K., and C.A.K.; Methodology, M.S., M.R.M., R.L.L., and C.A.K.; Investigation, R.L.L., M.S., E.G.P., R.L.I., A.W., and F.P.; Resources, R.L.I. and A.W.; Formal Analysis, M.S. and M.R.M.; Data Curation, M.R.M; Writing – Original Draft, M.R.M and R.L.L.; Writing – Review & Editing, M.R.M, C.A.K., and D.K.; Funding Acquisition, C.A.K. and D.K.; Supervision, C.A.K.

ACKNOWLEDGMENTS

We thank Prof. C. Peter Wolk (MSU-DOE Plant Research Laboratory and Department of Plant Biology, Michigan State University) for the gift of *E. coli* strain ED8654 and pRL443 plasmid. We are indebted to Dr. Enrique Flores and Dr. Antonia Herrero (IBVF, Seville, Spain) for the *Nostoc* 7120 strain and the pRL277 plasmid. We thank Dr. Fei Cai for comments and assistance in revision of this manuscript. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. No conflict of interest declared.

Received: March 16, 2016
Revised: May 25, 2016
Accepted: June 20, 2016
Published: July 4, 2016

REFERENCES

Abasova, L., Boulay, C., Vass, I., and Kirilovsky, D. (2008). Non-photochemical-quenching mechanisms in the cyanobacterium *Thermosynechococcus elongatus*. In Photosynthesis. Energy from the Sun, J.F. Allen, E. Gantt, J.H. Golbeck, and B. Osmond, eds. (Heidelberg: Springer), pp. 993–996.

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix. refine. Acta Crystallogr. D Biol. Crystallogr. **68**:352–367.

Anderson, D.C., Campbell, E.L., and Meeks, J.C. (2006). A soluble 3D LC/MS/MS proteome of the filamentous cyanobacterium *Nostoc punctiforme*. J. Proteome Res. **5**:3096–3104.

Berera, R., Van Stokkum, I.H.M., Gwizdala, M., Wilson, A., Kirilovsky, D., and Van Grondelle, R. (2012). The photophysics of the orange carotenoid protein, a light-powered molecular switch. J. Phys. Chem. B **116**:2568–2574.

Bernstein, H.C., Charania, M.A., McClure, R.S., Sadler, N.C., Melnicki, M.R., Hill, E.A., Markillie, L.M., Nicora, C.D., Wright, A.T., Romine, M.F., et al. (2015). Multi-omic dynamics associate oxygenic

photosynthesis with nitrogenase-mediated H2 production in *Cyanothece* sp. ATCC 51142. Sci. Rep. **5**:16004.

Boulay, C., Abasova, L., Six, C., Vass, I., and Kirilovsky, D. (2008). Occurrence and function of the orange carotenoid protein in photoprotective mechanisms in various cyanobacteria. Biochim. Biophys. Acta **1777**:1344–1354.

Bourcier de Carbon, C., Thurotte, A., Wilson, A., Perreau, F., and Kirilovsky, D. (2015). Biosynthesis of soluble carotenoid holoproteins in *Escherichia coli*. Sci. Rep. **5**:9085.

Cai, Y.P., and Wolk, C.P. (1990). Use of a conditionally lethal gene in *Anabaena sp.* strain PCC 7120 to select for double recombinants and to entrap insertion sequences. J. Bacteriol. **172**:3138–3145.

Christman, H.D., Campbell, E.L., and Meeks, J.C. (2011). Global transcription profiles of the nitrogen stress response resulting in heterocyst or hormogonium development in *Nostoc punctiforme*. J. Bacteriol. **193**:6874–6886.

Collingridge, P.W., and Kelly, S. (2012). MergeAlign: improving multiple sequence alignment performance by dynamic reconstruction of consensus multiple sequence alignments. BMC Bioinformatics **13**:117.

Criscuolo, A., and Gribaldo, S. (2011). Large-scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria. Mol. Biol. Evol. **28**:3019–3032.

D'Agostino, P.M., Song, X., Neilan, B.A., and Moffitt, M.C. (2014). Comparative proteomics reveals that a saxitoxin-producing and a nontoxic strain of *Anabaena circinalis* are two different ecotypes. J. Proteome Res. **13**:1474–1484.

Do, C.B., Mahabhashyam, M.S.P., Brudno, M., and Batzoglou, S. (2005). ProbCons: probabilistic consistency-based multiple sequence alignment. Genome Res. **15**:330–340.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**:1792–1797.

Elhai, J., Vepritskiy, A., Muro-Pastor, A.M., Flores, E., and Wolk, C.P. (1997). Reduction of conjugal transfer efficiency by three restriction activities of *Anabaena* sp. strain PCC 7120. J. Bacteriol. **179**:1998–2005.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. **60**:2126–2132.

Engene, N., Rottacker, E.C., Kaštovský, J., Byrum, T., Choi, H., Ellisman, M.H., Komárek, J., and Gerwick, W.H. (2012). *Moorea producens* gen. nov., sp. nov. and *Moorea bouillonii* comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. Int. J. Syst. Evol. Microbiol. **62**:1171–1178.

Enright, A.J., Iliopoulos, I., Kyrpides, N.C., and Ouzounis, C.A. (1999). Protein interaction maps for complete genomes based on gene fusion events. Nature **402**:86–90.

Flaherty, B.L., Van Nieuwerburgh, F., Head, S.R., and Golden, J.W. (2011). Directional RNA deep sequencing sheds new light on the transcriptional response of *Anabaena* sp. strain PCC 7120 to combined-nitrogen deprivation. BMC Genomics **12**:332.

Gao, Z.-M., Wang, Y., Tian, R.-M., Wong, Y.H., Batang, Z.B., Al-Suwailem, A.M., Bajic, V.B., and Qian, P.-Y. (2014). Symbiotic adaptation drives genome streamlining of the cyanobacterial sponge symbiont "Candidatus *Synechococcus spongiarum*". mBio **5**, e00079-14.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. **59**:307–321.

Gwizdala, M., Wilson, A., and Kirilovsky, D. (2011). In vitro reconstitution of the cyanobacterial photoprotective mechanism mediated by the orange carotenoid protein in *Synechocystis* PCC 6803. Plant Cell **23**:2631–2643.

Han, M.V., and Zmasek, C.M. (2009). phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinformatics **10**:356.

Hilton, J.A., Foster, R.A., Tripp, H.J., Carter, B.J., Zehr, J.P., and Villareal, T.A. (2013). Genomic deletions disrupt nitrogen metabolism pathways of a cyanobacterial diatom symbiont. Nat. Commun. **4**:1767.

Holt, T.K., and Krogmann, D.W. (1981). A carotenoid-protein from cyanobacteria. Biochim. Biophys. Acta **637**:408–414.

Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. **66**:125–132.

Katoh, K., and Standley, D.M. (2013). Mafft multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. **30**:772–780.

Kerfeld, C.A. (2004a). Water-soluble carotenoid proteins of cyanobacteria. Arch. Biochem. Biophys. **430**:2–9.

Kerfeld, C.A. (2004b). Structure and function of the water-soluble carotenoid-binding proteins of cyanobacteria. Photosynthesis Research **81**:215–225.

Kerfeld, C.A., Sawaya, M.R., Brahmandam, V., Cascio, D., Ho, K.K., Trevithick-Sutton, C.C., Krogmann, D.W., and Yeates, T.O. (2003). The crystal structure of a cyanobacterial water-soluble carotenoid binding protein. Structure **11**:55–65.

Kestler, H.A., Muller, A., Kraus, J.M., Buchholz, M., Gress, T.M., Liu, H., Kane, D.W., Zeeberg, B.R., and Weinstein, J.N. (2008). VennMaster: area-proportional Euler diagrams for functional GO analysis of microarrays. BMC Bioinformatics **9**:67.

Kirilovsky, D., and Kerfeld, C.A. (2013). The orange carotenoid protein: a blue-green light photoactive protein. Photochem. Photobiol. Sci. **12**:1135–1143.

Leverenz, R.L., Jallet, D., Li, M.D., Mathies, R.A., Kirilovsky, D., and Kerfeld, C.A. (2014). Structural and functional modularity of the orange carotenoid protein: distinct roles for the N- and C-terminal domains in cyanobacterial photoprotection. Plant Cell **26**:426–437.

Leverenz, R.L., Sutter, M., Wilson, A., Gupta, S., Thurotte, A., Bourcier de Carbon, C., Petzold, C.J., Ralston, C., Perreau, F., Kirilovsky, D., et al. (2015). A 12 Å carotenoid translocation in a photoswitch associated with cyanobacterial photoprotection. Science **348**:1463–1466.

López-Igual, R., Wilson, A., Leverenz, R.L., Melnicki, M.R., Bourcier de Carbon, C., Sutter, M., Turmo, A., Perreau, F., Kerfeld, C.A., and Kirilovsky, D. (2016). Different functions of the paralogs to the N-terminal domain of the Orange Carotenoid Protein in the cyanobacterium *Anabaena* sp. PCC 7120. Plant Physiol. http://dx.doi.org/10.1104/pp.16/00502.

Marcotte, E.M., Pellegrini, M., Ng, H.-L., Rice, D.W., Yeates, T.O., and Eisenberg, D. (1999). Detecting protein function and protein-protein interactions from genome sequences. Science **285**:751–753.

Markowitz, V.M., Chen, I.M.A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., et al. (2012). IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. **40**:D115–D122.

Melnicki, M.R., Pinchuk, G.E., Hill, E.A., Kucek, L.A., Fredrickson, J.K., Konopka, A., and Beliaev, A.S. (2012). Sustained H(2) production driven by photosynthetic water splitting in a unicellular cyanobacterium. mBio **3**, e00197–e00112.

Möglich, A., Yang, X., Ayers, R.A., and Moffat, K. (2010). Structure and function of plant photoreceptors. Annu. Rev. Plant Biol. **61**:21–47.

Mohamed, H.E., and Vermaas, W. (2004). Slr1293 in *Synechocystis* sp. strain PCC 6803 is the C-3′,4′ desaturase (CrtD) involved in myxoxanthophyll biosynthesis. J. Bacteriol. **186**:5621.

Ow, S.Y., Cardona, T., Taton, A., Magnuson, A., Lindblad, P., Stensjö, K., and Wright, P.C. (2008). Quantitative shotgun proteomics of enriched heterocysts from *Nostoc* sp. PCC 7120 using 8-plex isobaric peptide tags. J. Proteome Res. **7**:1615–1628.

Ow, S.Y., Nolrel, J., Cardona, T., Taton, A., Lindblad, P., Stensjö, K., and Wright, P.C. (2009). Quantitative overview of N 2 fixation in *Nostoc punctiforme* ATCC 29133 through cellular enrichments and iTRAQ shotgun proteomics. J. Proteome Res. **8**:187–198.

Park, J.-J., Lechno-Yossef, S., Wolk, C.P., and Vieille, C. (2013). Cell-specific gene expression in *Anabaena variabilis* grown phototrophically, mixotrophically, and heterotrophically. BMC Genomics **14**:759.

Partensky, F., and Garczarek, L. (2010). *Prochlorococcus*: advantages and limits of minimalism. Annu. Rev. Mar. Sci. **2**:305–331.

Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., and Delsuc, F. (2005). Heterotachy and long-branch attraction in phylogenetics. BMC Evol. Biol. **5**:50.

Punginelli, C., Wilson, A., Routaboul, J.M., and Kirilovsky, D. (2009). Influence of zeaxanthin and echinenone binding on the activity of the Orange Carotenoid Protein. Biochim. Biophys. Acta **1787**:280–288.

Rajeev, L., Nunes da Rocha, U., Klitgord, N., Luning, E.G., Fortney, J., Axen, S.D., Shih, P.M., Bouskill, N.J., Bowen, B.P., Kerfeld, C.A., et al. (2013). Dynamic cyanobacterial response to hydration and dehydration in a desert biological soil crust. ISME J. **7**:2178–2191.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. Trends Genet. **16**:276–277.

Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., and Stanier, R.Y. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J. Gen. Microbiol. **111**:1–61.

Sedoud, A., López-Igual, R., Ur Rehman, A., Wilson, A., Perreau, F., Boulay, C., Vass, I., Krieger-Liszkay, A., and Kirilovsky, D. (2014). The cyanobacterial photoactive orange carotenoid protein is an excellent singlet oxygen quencher. Plant Cell **26**:1781–1791.

Shih, P.M., Wu, D., Latifi, A., Axen, S.D., Fewer, D.P., Talla, E., Calteau, A., Cai, F., Tandeau de Marsac, N., Rippka, R., et al. (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc. Natl. Acad. Sci. USA **110**:1053–1058.

Sjölander, K. (2004). Phylogenomic inference of protein molecular function: advances and challenges. Bioinformatics **20**:170–179.

Sutter, M., Wilson, A., Leverenz, R.L., López-Igual, R., Thurotte, A., Salmeen, A.E., Kirilovsky, D., and Kerfeld, C.A. (2013). Crystal structure of the FRP and identification of the active site for modulation of OCP-mediated photoprotection in cyanobacteria. Proc. Natl. Acad. Sci. USA **110**:10022–10027.

Swingley, W.D., Chen, M., Cheung, P.C., Conrad, A.L., Dejesa, L.C., Hao, J., Honchak, B.M., Karbach, L.E., Kurdoglu, A., Lahiri, S., et al. (2008). Niche adaptation and genome expansion in the chlorophyll d-producing cyanobacterium *Acaryochloris marina*. Proc. Natl. Acad. Sci. USA **105**:2005–2010.

Thompson, A.W., Foster, R.A., Krupke, A., Carter, B.J., Musat, N., Vaulot, D., Kuypers, M.M.M., and Zehr, J.P. (2012). Unicellular cyanobacterium symbiotic with a single-Celled Eukaryotic Alga. Science **337**:1546–1550.

Thurotte, A., López-Igual, R., Wilson, A., Comolet, L., Bourcier de Carbon, C., Xiao, F., and Kirilovsky, D. (2015). Regulation of orange carotenoid protein activity in cyanobacterial photoprotection. Plant Physiol. **169**:737–747.

Toepel, J., McDermott, J.E., Summerfield, T.C., and Sherman, L.A. (2009). Transcriptional analysis of the unicellular, diazotrophic cyanobacterium *Cyanothece* sp. ATCC 51142 grown under short day/night cycles. J. Phycol. **45**:610–620.

Voß, B., Bolhuis, H., Fewer, D.P., Kopf, M., Möke, F., Haas, F., El-Shehawy, R., Hayes, P., Bergman, B., Sivonen, K., et al. (2013). Insights into the physiology and ecology of the brackish-water-adapted cyanobacterium *Nodularia spumigena* CCY9414 based on a genome-transcriptome analysis. PLoS One **8**:e60224.

Wang, Q., and Moerner, W.E. (2015). Dissecting pigment architecture of individual photosynthetic antenna complexes in solution. Proc. Natl. Acad. Sci. USA **112**:201514027.

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**:1189–1191.

Wheeler, T.J., Clements, J., and Finn, R.D. (2014). Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. BMC Bioinformatics **15**:7.

Wilson, A., Boulay, C., Wilde, A., Kerfeld, C.A., and Kirilovsky, D. (2007). Light-induced energy dissipation in iron-starved cyanobacteria: roles of OCP and IsiA proteins. Plant Cell **19**:656–672.

Wilson, A., Punginelli, C., Gall, A., Bonetti, C., Alexandre, M., Routaboul, J.-m., Kerfeld, C.A., Grondelle, R.V., Robert, B., Kennis, J.T.M., et al. (2008). A photoactive carotenoid protein acting as light intensity sensor. Proc. Natl. Acad. Sci. USA **105**:12075–12080.

Wilson, A., Kinney, J.N., Zwart, P.H., Punginelli, C., D'Haene, S., Perreau, F., Klein, M.G., Kirilovsky, D., and Kerfeld, C.A. (2010). Structural determinants underlying photoprotection in the photoactive orange carotenoid protein of cyanobacteria. J. Biol. Chem. **285**:18364–18375.

Wilson, A., Punginelli, C., Couturier, M., Perreau, F., and Kirilovsky, D. (2011). Essential role of two tyrosines and two tryptophans on the photoprotection activity of the Orange Carotenoid Protein. Biochim. Biophys. Acta **1807**:293–301.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. **67**:235–242.

Wu, Y.P., and Krogmann, D.W. (1997). The orange carotenoid protein of *Synechocystis* PCC 6803. Biochim. Biophys. Acta **1322**:1–7.

Zhang, S., and Bryant, D.A. (2015). Biochemical validation of the glyoxylate cycle in the cyanobacterium *Chlorogloeopsis fritschii* strain PCC 9212. J. Biol. Chem. **290**:14019–14030.