Lawrence Berkeley National Laboratory

Title

Orange carotenoid proteins: structural understanding of evolution and function

Permalink

https://escholarship.org/uc/item/5rf4c5cb

Journal

Trends in Biochemical Sciences, 49(9)

ISSN

0968-0004

Authors

Kerfeld, Cheryl A

Sutter, Markus

Publication Date

2024-09-01

DOI

10.1016/j.tibs.2024.04.010

1   Orange carotenoid proteins: structural understanding of evolution and function

2   Cheryl A. Kerfeld[1,2,3], Markus Sutter[1,2,3]

3

4   [1] MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, MI 48824, USA

5   [2] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory,

6   Berkeley, CA 94720, USA

7   [3] Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National

8   Laboratory, Berkeley, CA 94720, USA

9

10   **Correspondence: ckerfeld@lbl.gov**

11   **Kerfeld orcid: 0000-0002-9977-8482**

12   **Sutter orcid: 0000-0001-6290-4820**

13

16

17   **Abstract**

18   Cyanobacteria uniquely contain a primitive water-soluble carotenoprotein, the Orange

19   Carotenoid Protein (OCP). Nearly all extant cyanobacterial genomes contain genes for the OCP

20   or its homologs, implying an evolutionary constraint for cyanobacteria to conserve its function.

21   Genes encoding the OCP and its two constituent structural domains, the N-terminal domain,

22   Helical Carotenoid Proteins (HCPs), and its C-terminal domain (CCPs), are found in the most basal

23   lineages of extant cyanobacteria. These three carotenoproteins exemplify the importance of the

24   protein for carotenoid properties—including protein dynamics--in response to environmental

25   changes in facilitating a photoresponse and energy quenching. Here we review new structural

26   insights for these carotenoproteins and situate the role of the protein in what is currently

27   understood about their functions.

Carotenoids and the Orange Carotenoid ProteinCarotenoids are considered the oldest, most widespread, and numerous class of pigments and are proposed to be biosignatures for detecting life [1]. They extend the spectral range of light-harvesting antennae of nearly all phototrophic organisms, from bacteria to plants,, where they function as **accessory pigments**, absorbing photons not captured by chlorophylls in photosynthetic complexes. Because carotenoids are fast and effective in deactivating (**quenching**) excited states of nearby **chromophores** through thermal dissipation, they frequently play roles in **photoprotection** mechanisms where they disarm the ROS (reactive oxygen species) inherently produced during oxygenic photosynthesis [2]. It seems likely that the first role of carotenoids in cyanobacteria was as a component of membranes [3], where they provided a structural role and likely influenced permeability to oxygen. The next key event was the complexation of carotenoids to proteins, thereby vastly expanding the potential roles of carotenoids from membrane-structure to light harvesting and its necessary accompaniment, photoprotection.

One of the simplest model systems for the study of how the protein environment tunes carotenoid function is the cyanobacterial Orange Carotenoid Protein (OCP), a 34 kDa protein that noncovalently binds a single carotenoid (Figure 1A) [4, 5]. The OCP regulates photoprotection in cyanobacteria, the organisms that originated oxygenic photosynthesis. The OCP is the only known photoactive protein that uses a carotenoid as its sole chromophore; the absorption of blue-green light causes OCP to convert from a dark stable orange form, OCP^O, to a light-activated red form, OCP^R [6]. The OCP^R directly participates in a photoprotective mechanism by binding to the light-harvesting antenna, the **phycobilisome (PBS)** [7], and dissipating excess captured light energy [6]. Recent structural studies of the OCP and of its primitive antecedents, described below, have provided unexpected new insights into how carotenoid properties are tuned by protein structure, and how protein dynamics—of both large and small amplitudes—regulate carotenoid function. This review will focus on the results of these structural studies and place them in the context of the evolution and dynamics of OCP function. It is known that determinants of quenching properties are inherent in the protein sequence, while the spectroscopic properties play a subsidiary role in in understanding OCP function [8]. Therefore, much of the spectroscopy

57 that has been done on OCP in the past can be reinterpreted in the context of the newly

58 discovered structural and evolutionary aspects described here.

59

The Modularity of OCP Structure and Function

61 The cyanobacterial OCP was discovered by David Krogmann as a contaminant in cytochrome

62 purifications [9], and the crystal structure of the OCP (Figure 1A) was determined in 2003, prior

63 to an understanding of its physiological function; however, its singlet oxygen quenching

64 properties and its ability to reversibly change color from orange to red were described with the

65 structure, and a hypothesis that the OCP may be photoresponsive and play a role in

66 photoprotection was suggested [10-12]. Eventually it was shown to indeed be a photoactive

67 protein responsible for both light sensing and energy dissipation in cyanobacteria [13].

68

69 Structurally, the OCP consists of two domains (Figure 1A): the all-alpha helical N-terminal domain

70 (NTD, formed by residues 1-164, (*Synechocystis* sp. PCC 6803 (hereafter *Syn* 6803) numbering)

71 and the C-terminal domain (CTD, residues 190-314), formed by alpha helices and beta sheet.

72 While the NTD is found only in cyanobacteria, the CTD is a member of the versatile NTF2

73 superfamily, widespread among all kingdoms of life [12, 14]. The two domains are joined by a

74 linker (residues 165-190) that in the orange, inactive state (OCP[O]) is excluded from the interface

75 between the NTD and CTD and wraps around the surface of the CTD (Figure 1B) [12]. A keto-

76 carotenoid—such as canthaxanthin or echinenone— spans the two domains, with a keto group

77 of the CAR (carotenoid) hydrogen bonded to a conserved Trp and Tyr in the CTD. (Figure 1A,

78 inset). In addition to the large interface for interaction across which the CAR spans, the CTD and

79 NTD also interact through a "latch" formed by the N-terminal extension (NTE), residues 1-19, that

80 nestle against the CTD surface opposite of the linker (Figure 1A, B) [12].

81 The structural modularity of the OCP was shown to extend to functional modularity: controlled

82 proteolysis showed that photo activated OCP[R] is sensitive to protease exposure and rapidly

83 converted all OCP into a permanently red and constitutively active form known as the RCP; this

84 proteolytic protein fragment consists of the NTD and the carotenoid [15]. Krogmann had

85 identified similar red proteolytic fragments in purifications from wild cyanobacterial blooms [9],

86   suggesting that, in cells, proteolysis could be used regulate quenching activity Recombinant RCP

87   was readily crystallized and the structure subsequently determined [16]. Combined with X-ray

88   footprinting [17], the structural analysis of the RCP in comparison to the full-length OCP to

89   determine the solvent accessible surface area differences between OCP^O to OCP^R, revealing that

90   the NTD and CTD domains remained structurally intact, but dissociated, with the carotenoid

91   migrating entirely into the NTD during the activation [16].

92

93   **The modular OCP Evolved from the Fusion of Two Carotenoproteins**

94   The biochemical and structural studies described above suggest an analogy between the OCP and

95   other modular photoreceptors that employ sensor and effector domain architectures such as

96   BLUF (blue light using flavin) [18] and LOV (light, oxygen voltage) domains [19]. In this context,

97   the NTD could be considered the effector (quenching) domain, activated when the carotenoid

98   was fully enclosed within, and the CTD acting as a regulatory domain [20]. This is consistent with

99   the proposed evolution of the OCP [5, 21, 22]; it evolved from the fusion of two discrete proteins

100   that formed a heterodimer. Already at the time of the first structural description of the OCP,

101   which coincided with the beginning of the genomics era, it was noted that in the few diverse

102   cyanobacterial genomes available, in addition to the OCP, several had genes encoding homologs

103   to both the NTD and the CTD [12]. The NTD homologs are now known collectively as the **Helical**

104   **Carotenoid Proteins** (HCPs) and these have been phylogenetically resolved into at least nine

105   clades [23]. The proteins from all of the clades are proposed to bind carotenoid, and the crystal

106   structures of two of these (*Anabaena* sp. PCC 7120 (hereafter *Anabaena* 7120) HCP1 and HCP2)

107   have been determined [23, 24] (Figure 1C), confirming that they bind a carotenoid as observed

108   in the RCP [15]. Many organisms encode multiple HCPs in addition to the OCP. For example,

109   *Anabaena* 7120 contains four HCPs: HCP1-HCP4 in addition to OCP. Among all HCPs, only HCP4

110   has thus far been verified to quench PBS [25], and among HCP1-4 in *Anabaena* 7120, it has the

111   highest sequence identity (52%) with the NTD of *Anabaena* 7120 OCP. The function of the other

112   HCPs is as-yet unknown (see Box 1), and are likely distinct, given their varying occurrences in

113   expression and proteomics studies (summarized in [23] and, more recently provided in [26-30]).

114

115   The role of HCP4 in both quenching and the evolution of the OCP is also supported by the

116   observation that a gene encoding a CTD homolog is frequently adjacent to an HCP4 locus [21,

117   23]. These proteins, termed CCP (C-terminal domain-like carotenoprotein or CTD homologs

118   (CTDHs)), are now known to bind a single carotenoid as homodimers (Figure 1D), shifting the

119   absorption maximum by 80nm (versus the free carotenoid in solvent) to ~560nm, with a further

120   10nm red shift in a dimer of dimers state at high concentrations [31]. Additionally, some CCPs

121   have been shown to accept carotenoid from HCPs [27, 32-34]. Results from devolving the OCP

122   into its constituent domains and evaluating photoactivation and recovery [21, 35] are consistent

123   with a model that the CCP, the precursor of the CTD, regulated quenching by the HCP by providing

124   a carotenoid environment within the heterodimer that rendered the carotenoid inactive for

125   quenching.

126

127   It also seems likely that this fusion event happened several times in cyanobacterial evolution.

128   This is evident from the multiple discrete families of the OCP that are now known

129   bioinformatically. In 2017 it was shown that there were two clearly resolved families of the OCP,

130   known as OCP1 (the canonical, including the extensively characterized OCP from *Syn* 6803) and

131   OCP2, as well as a third, more mixed group, OCPx [36] which has recently been renamed OCP3

132   and further resolved into 3 subgroups [37]. OCP1-3 all photoactivate from orange to red, and all

133   quench PBS, but with some differences in quenching and regulatory properties [27]. For example,

134   *Tolypothrix* sp. PCC 7601 (*Tolypothrix* hereafter) OCP2 photoactivates faster and recovers faster,

135   and exhibits lower fluorescence quenching than its paralogous OCP1 [36]; it also appears to have

136   less flexibility [8]. Additionally, the OCP^R to OCP^O recovery for OCP1 is accelerated by the binding

137   of the all-helical Fluorescence Recovery Protein (FRP) [38, 39], whereas OCP2 and OCP3 appear

138   to be insensitive to FRP [36, 37]. Interestingly, the evolution of regulation of OCP1 by FRP has

139   been shown to be the result of complementarity between the CTD dimer and the FRP, which was

140   acquired by horizontal gene transfer, likely from an ancestral delta-proteobacterium [40]. These

141   differences in activity and regulation may account for the co-occurrence of OCP paralogs within

142   the same genome—for example, OCP1 and OCP2 in *Tolypothrix* [36] or the two OCP3 paralogs

143   found in *Nostoc flagelliforme* [27] -- to confer more flexibility in the range of photoprotective

144 responses. More specifically, in *Tolypothrix*, OCP1 appears to be constitutively expressed with

145 OCP2 being additionally expressed under high irradiance [36, 41] while the *Nostoc flagelliforme*

146 paralogs likewise exhibited differential expression and differences in quenching properties [27].

147

148 **Large-scale Structural Dynamics Tune OCP Function**

149 While there were differences in the kinetics, as noted above, OCP1, OCP2, and OCP3 all

150 photoactivate and quench PBS. Indeed, the notion that the quenching mechanism is conserved

151 across OCP families was reinforced when Slonimskiy et al demonstrated that OCP3 from

152 *Gloeobacter*, an organism with an atypical, bundle-shaped PBS, could quench the PBS of *Syn* 6803

153 [37]. However, the precise structural details of how the OCP quenches the PBS and the structure

154 of the OCP^R remained elusive. Elements of the activation and quenching process were known:

155 Biophysical studies of the OCP^R had provided some information; in addition to the required

156 carotenoid translocation into the NTD [16], other events such as detachment of the NTE from the

157 CTD and domain separation had been detected [42, 43]. It required capturing the OCP^R in the act

158 of quenching the PBS to learn its structure [44]. Contrary to all previous models, four activated

159 OCP^{R}s, arranged as two dimers, quench the 6.2 MDa PBS with its 396 phycocyanobilin pigments

160 (Figure 2A). Also unexpectedly, quenching of the OCP^R-PBS complex revealed a drastic structural

161 rearrangement relative to the inactive form; the CTD rotated approximately 220 degrees around

162 the NTD and with a net displacement ~60 Å relative to its position in the OCP^O structure (Figure

163 2B) [44]. The structure of the OCP^R bound to the PBS explains why the obvious experiment,

164 illuminate the OCP^O crystals with light and then collect diffraction data, always failed, as such a

165 large structural displacement upon illumination would be prevented by the close crystal packing

166 interactions. Few such large domain motions are known (e.g. a 90 Å translocation as observed

167 for the nuclear localization signal domain of influenza virus polymerase [45]), though the

168 increasing application of cryo-EM and the development of data analysis software (3Dflex [46] and

169 3DVA [46]) promise to reveal more such insights from analysis of conformational flexibility

170 exhibited in single particle datasets.

171

172   The OCP[R]-PBS structure, which has since been extended to a resolution of 1.6-2.1 Å [47], explains

173   how the protein "activates" the carotenoid by providing a snapshot of its environment in the act

174   of quenching. The carotenoid translocation discovered in the RCP structure presented a major

175   advance in understanding of the photoprotective mechanism; however, many details such as the

176   position of the NTE, of the interdomain linker and of the CTD, all of which are not present in the

177   RCP [16], are protein elements that influence the carotenoid properties. The interdomain linker,

178   for example, contributes to the electrostatic environment around the CAR and results in an even

179   more pronounced polarization than in RCP. In the recent high resolution structure [47], modeling

180   the sidechain motions and water molecules permitted quantum mechanical calculations of the

181   transition dipole moment (TDM) conferred by the protein environment. This included the

182   contribution of some PBS residues (in ApcA and ApcB) as well as OCP residues R155, E34 and D35

183   (conserved in OCP1 and OCP2), and showed that it is not distortions in the CAR, but changes in

184   the protein environment, that are the drivers for potentiating quenching [47].

185

186   The OCP[R] dimer structure also suggests a potentially new aspect of the regulation conferred by

187   the CTD. The CTD mediates the dimerization of the OCP[R] through inter-beta sheet interactions

188   (Figure 2C); the dimer forms a buttress-like structure, with one NTD bound to the T cylinder and

189   the second bound to the B cylinder in the PBS core (Figure 2A) [44]. Such an interface seems to

190   be favored by NTF2 domains generally; a very similar interface is observed in the structure of a

191   protein with unknown function (DUF4783 family protein from *Bacteroides uniformis* ATCC 8492,

192   PDB ID 4Q53) that we can structurally superimpose almost perfectly despite only 9% sequence

193   identity (Figure 2C). The predominantly hydrophobic residues forming the CTD dimerization

194   interface are shielded by the NTE in the OCP[O] form [12, 14]. This adds an additional regulatory

195   implication of the oligomeric states of OCP1. A prerequisite to detaching the NTE from the CTD

196   is the dissociation of the OCP1[O] dimer (Figure 1). Based on structural arguments, the dimeric

197   resting OCP1[O] state found in the crystal (Figure 1A) has always been asserted to be biologically

198   relevant by structural biologists because of the surface area excluded [5, 12]. In addition to the

199   obvious occluding of the requisite protein motions for domain separation, the OCP dimer

200   interface is mediated by conserved residues forming the OCP[O] dimer: D6, R9, G10, P13, N14, T15,

201 L16, A17, F227, R229 (Figure 3). Notably, these residues are not well conserved in OCP3,

202 consistent with the observation that it is predominantly monomeric [37]; although there is

203 discrepant earlier data [48]. Recently, the dissociation of the OCP1^O dimer into monomers, the

204 prerequisite for photoactivation from a structural perspective, was shown to be light-driven [49]

205 and important for both photoactivation and recovery of *Syn* 6803 OCP [50]. Recently, protein

206 motion has also been suggested to play a role in regulating the activity of the OCP^R; the dimerized

207 CTDs are the most flexible within the range of motion exhibited by the OCP^R-PBS complex [47].

208 The interdomain linker moves in conjunction with the CTD-CTD dimer which fluctuates as a unit,

209 with linker residues E174-V176 making the closest approach to the CAR through the course of

210 the motion. This potentially modifies the TDM of the carotenoid, expanding the regulatory

211 repertoire to the structural dynamics of the CTD.

212 It was previously demonstrated that OCP1^R is a dimer in solution [36], but the observation of the

213 structure of the dimer bound to the PBS also provided new insight into the regulation of the OCP1

214 by FRP. While precise mechanism of FRP catalysis of dark conversion is not known, several lines

215 of evidence have implicated that it binds the CTD of OCP^R [39, 51]. Solution studies of the

216 interaction of engineered FRP variants with OCP^R in solution led to the suggestion of a 1:1

217 complex of OCPR to FRP [52]. However this model required that FRP, which based on structural

218 arguments was suggested to function as a dimer, was required to monomerize to bind the CTD.

219 Now, with the observation of the quenching-active dimer, the FRP dimer likely interacts with the

220 CTD part of the OCP^R dimer, acting as a sort of chaperone to protect the exposed hydrophobic

221 beta sheet until it can reform with its NTD counterpart. Interestingly, the distance between the

222 two head groups of FRP that are thought to interact with OCP are spaced apart similarly (both

223 ~80 Å) to the distance between the two NTDs in OCP^R dimer associated with the PBS. Each FRP

224 head group binding a CTD would bring the CTDs close to their respective NTDs and facilitate

225 recombination (Figure 2D).

226

227 **Different but Similar—Conserved Features of Quenching by OCP Homologs**

228 FRP binding is not a universal feature of OCP-mediated photoprotective systems; OCP2 and OCP3

229 are insensitive to it [36, 37]. We now consider if other structural features known from the OCP^R

230 structure are relevant to quenching by OCP2, OCP3 and HCP4. While the OCP1 has been

231 extensively characterized structurally and functionally, far less data is available for the OCP2 and

232 OCP3 families. However representatives of both paralogous OCP families, (*Tolypothrix* OCP2 and

233 *Gloeobacter kilaueensis* JS1 OCP3), as well as the *Anabaena* 7120 HCP4 have been shown to

234 quench PBS [25, 36, 37]. Likewise, structural data is now available for the orange form of both

235 OCP2 (PDB IDs 8PYH and 8PZK) [53] and OCP3 (8A0H) [37]; they superimpose on the OCP1^O

236 structure of *Syn* 6803 with less than 1 Å rmsd. A sequence alignment of these with *Syn* 6803 OCP,

237 with residues known from the *Syn* 6803 OCP^R structure to be important for binding and

238 quenching the PBS marked, is shown in Figure 3. The PBS binding residues are largely conserved

239 among the three OCP paralogs. While HCP4 lacks the equivalent of an NTE, all of the other PBS

240 binding residues are conserved relative the OCP NTD (Figure 3), suggesting that the binding sites

241 on the PBS are similar; this is also evident from a cursory survey of their interaction counterparts

242 in these species.

243

244 The protein environment of the carotenoid tunes its transition dipole moment to poise it for

245 quenching. Using the OCP1^R -PBS complex as a template and placing homology models for the

246 OCP2^R, OCP3^R and HCP4 in that context, the electrostatics of the carotenoid environment for

247 each was visualized (Figure 4). Given the conservation of the PBS residues at the OCP binding site,

248 the conservation of PBS binding residues among the OCPs, and the experimental observation of

249 PBS quenching by heterologous OCP, such as the *Gloeobacter kilaueensis* JS1 OCP3 quenching

250 the *Syn* 6803 PBS [37], we include the influence of the PBS on the electrostatics (Figure 4). The

251 carotenoid environment electrostatics of the OCP3 and HCP4 are similar to that observed of

252 OCP1^R, while OCP2 is relatively non-polar at the β1 surface, which interacts with the PBS. The

253 recent structure of the OCP^R-PBS complex was of a resolution sufficient for quantum mechanical

254 calculations of the transition dipole moment of the carotenoid [47]; residues that play a

255 significant role in setting the TDM, such as R155 (substituted as Q in all OCP2), E34, D35 are

256 conserved across the homologs. While OCP1 and OCP2 have the linker motif EPV (174-176)

257 suggested to be important in affecting either stability or duration of quenching [47], OCP3 has a

258 substantial difference with an E substituted for Pro175 (Figure 3).

259

260 Another difference between OCP1 and OCP2 that was documented is that OCP1[R] forms dimers,

261 while OCP2[R] was predominantly a monomer [36]; for OCP3 the oligomeric state has only been

262 characterized for OCP[O]. Considering the conservation of residues that stabilize the CTD dimer

263 interface in the OCP[R]-PBS state, and the extent of solvent accessible surface area buried in

264 homology models of the OCP2[R] (962 Å^2) and OCP3[R] (1146 Å^2), suggest that both likely form dimers

265 as in the OCP1 quenching complex. Given the data that OCP2[R] is predominantly a monomer in

266 solution, we propose that association of two monomers of OCP2[R] as they bind the PBS for

267 quenching juxtaposes their CTDs in proximity to facilitate OCP[R] dimer formation.

268

269 **Concluding Remarks**

270 When the ketocarotenoid canthaxanthin becomes associated with protein to form the OCP, its

271 absorption maximum shifts to 495 nm (compared to 490 nm in DMSO). Photoactivation of OCP[O]

272 to OCP[R] further shifts it another 10 nm, the result of the protein environmental change

273 surrounding the CAR with the protein. These visible changes are a result of the sensitive tuning

274 of the carotenoid environment by the protein. The OCP exemplifies how the lack of covalent

275 bonds to CAR, while making their binding sites in proteins difficult to predict, permits CAR

276 function to be tuned by the dynamics of its protein environment [54], including intra-protein and

277 inter-domain motion. Given advances in structural tools such as cryo-EM and its data analysis

278 promise many more surprises in the fundamental understanding of the mechanism of function

279 of carotenoproteins that have implications for engineering of photoprotection; for example,

280 accelerating recovery from photoprotection has been shown to increase primary productivity in

281 plants [55]. Manipulation of photoprotection in cyanobacteria could be especially valuable in

282 production strains and communities used in biotechnology [56]. Likewise, understanding the

283 molecular architecture of pigment interactions in antenna and how energy transfer and

284 dissipation can be regulated is fundamental knowledge forthe construction of artificial

285 photosynthetic systems [57]. Beyond photosynthesis, photoreceptors like OCP have the potential

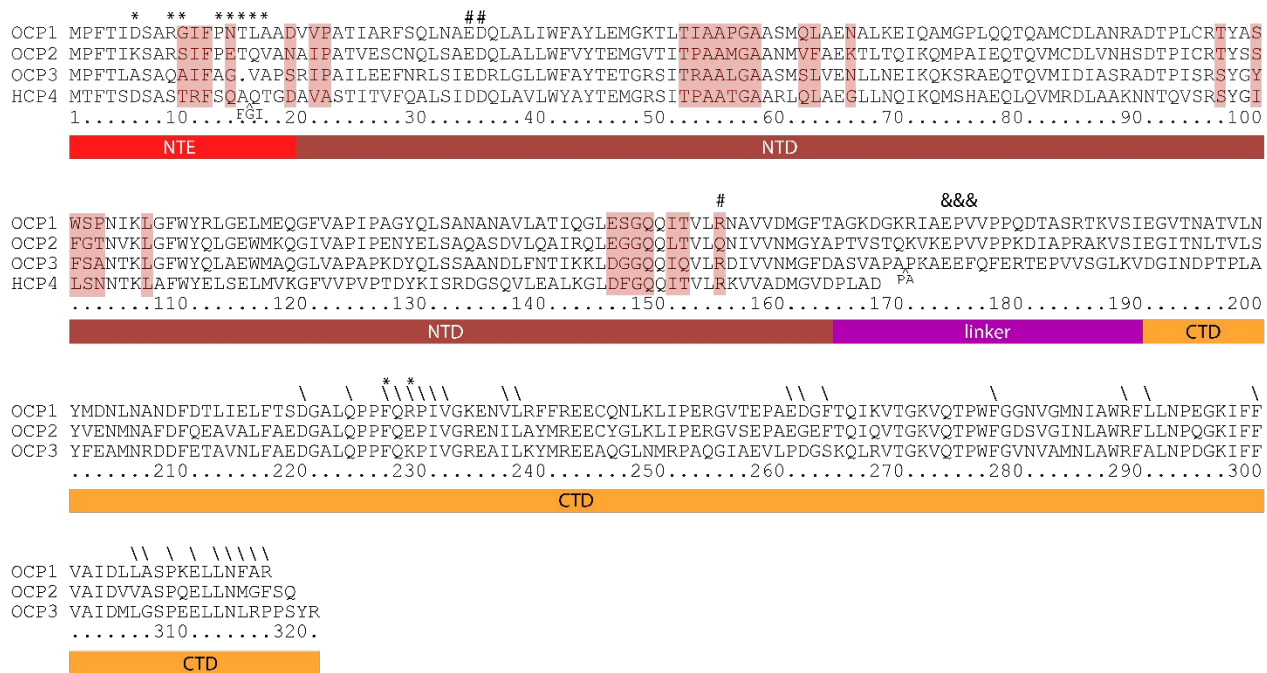286 to become light-activated switches foroptogenetic applications [58, 59].

287
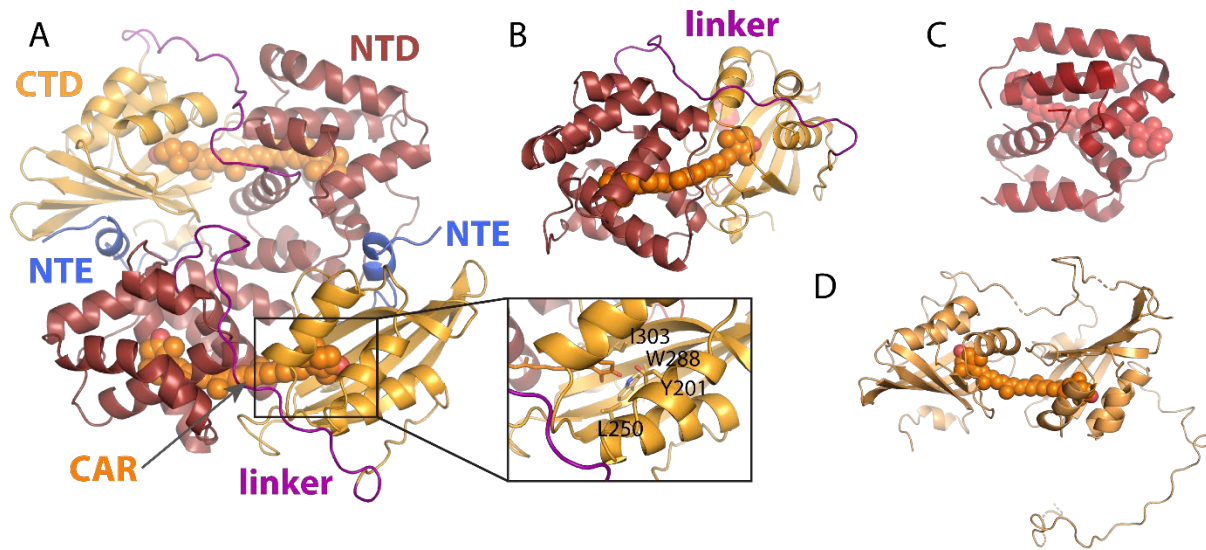
Figure 1. A) OCP^O dimer structure observed in crystallographic studies, highlighting the different domains across which the CA spans (space filling representation) (PDB ID 4XB5). B) OCP^O monomer, highlighting the linker (PDB ID 4XB5) C) HCP2 structure (PDB ID 5FCX), and D) CCP2 dimer model (SASBDB ID SASDHD6). NTD: N-terminal domain, dark red; CTD: C-terminal domain, orange; NTE: N-terminal extension, blue; CAR: carotenoid, orange; linker: magenta.


293
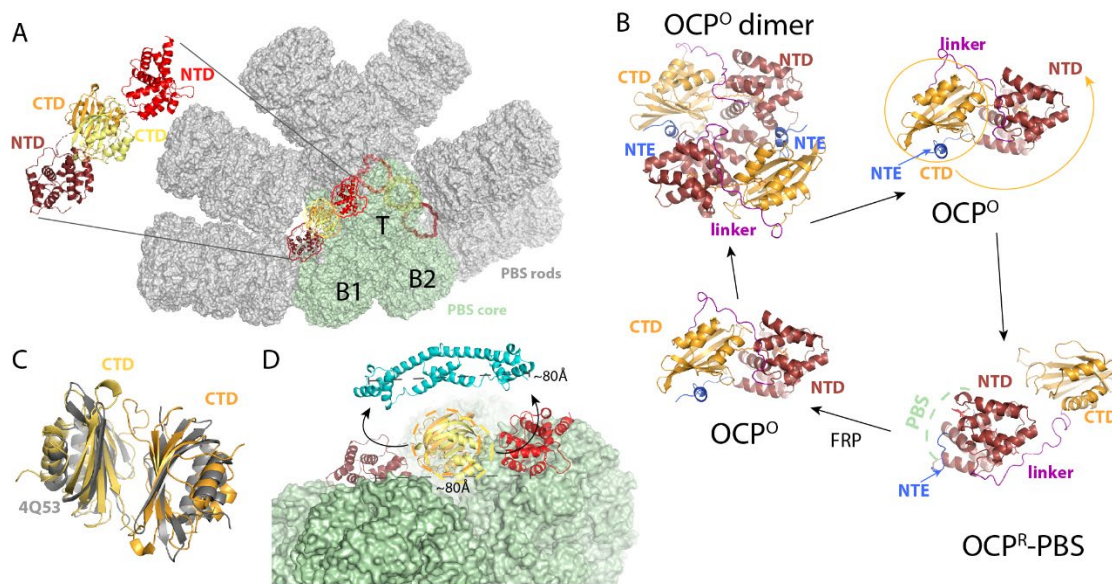
Figure 2. OCP-PBS structure overview. A) Overview of the position of the two OCP dimer pairs, OCP dimer in the back is outlined and dimer in front is shown zoomed in. B) OCP activation and recovery pathway: OCP^O dimer to monomer; structural changes to the OCP^O monomer lead to the OCP^R-PBS form where the NTD and CTD are separated; the FRP returns the OCP^R form back

298   to a monomeric OCP^O which can then dimerize again. C) CTD dimer as observed in the OCP^R-PBS

299   complex and alignment with a structurally homologous NTF2 domain protein (pdb ID 4Q53). D)

300   Distance between the head groups of FRP match the distance between the NTDs in OCP^R-PBS.

301

```
                *   **  *****                          ##
OCP1 MPFTIDSARGIFPNTLAADVVPATIARFSQLNAEDQLALIWFAYLEMGKTLTIAAPGAASMQLAENALKEIQAMGPLQQTQAMCDLANRADTPLCRTYAS
OCP2 MPFTIKSARSIFPETQVANAIPATVESCNQLSAEDQLALLWFVYTEMGVTITPAAMGAANMVFAEKTLTQIKQMPAIEQTQVMCDLVNHSDTPICRTYSS
OCP3 MPFTLASAQAIFAG.VAPSRIPAILEEFNRLSIEDRLGLLWFAYTETGRSITRAALGAASMSLVENLLNEIKQKSRAEQTQVMIDIASRADTPISRSYGY
HCP4 MTFTSDSASTRFSQAQTGDAVASTITVFQALSIDDQLAVLWYAYTEMGRSITPAATGAARLQLAEGLLNQIKQMSHAEQLQVMRDLAAKNNTQVSRSYGI
     1.......10...FGI..20........30........40........50........60........70........80........90.......100
```

NTE   NTD

```
                                                       #                       &&&
OCP1 WSPNIKLGFWYRLGELMEQGFVAPIPAGYQLSANANAVLATIQGLESGQQITVLRNAVVDMGFTAGKDGKRIAEPVVPPQDTASRTKVSIEGVTNATVLN
OCP2 FGTNVKLGFWYQLGEWMKQGIVAPIPENYELSAQASDVLQAIRQLEGGQQLTVLQNIVVNMGYAPTVSTQKVKEPVVPPKDIAPRAKVSIEGITNLTVLS
OCP3 FSANTKLGFWYQLAEWMAQGLVAPAPKDYQLSSAANDLFNTIKKLDGGQQIQVLRDIVVNMGFDASVAPAPKAEEFQFERTEPVVSGLKVDGINDPTPLA
HCP4 LSNNTKLAFWYELSELMVKGFVVPVPTDYKISRDGSQVLEALKGLDFGQQITVLRKVVADMGVDPLAD PA
     .......110.......120.......130.......140.......150.......160.......170.......180.......190.......200
```

NTD   linker   CTD

```
          \     \    *\\*\\\\    \\                   \\ \              \             \ \             \
OCP1 YMDNLNANDFDTLIELFTSDGALQPPFQRPIVGKENVLRFFREECQNLKLIPERGVTEPAEDGFTQIKVTGKVQTPWFGGNVGMNIAWRFLLNPEGKIFF
OCP2 YVENMNAFDFQEAVALFAEDGALQPPFQEPIVGRENILAYMREECYGLKLIPERGVSEPAEGEFTQIQVTGKVQTPWFGDSVGINLAWRFLLNPQGKIFF
OCP3 YFEAMNRDDFETAVNLFAEDGALQPPFQKPIVGREAILKYMREEAQGLNMRPAQGIAEVLPDGSKQLRVTGKVQTPWFGVNVAMNLAWRFALNPDGKIFF
     .......210.......220.......230.......240.......250.......260.......270.......280.......290.......300
```

CTD

```
        \\ \ \ \ \\\\\
OCP1 VAIDLLASPKELLNFAR
OCP2 VAIDVVASPQELLNMGFSQ
OCP3 VAIDMLGSPEELLNLRPPSYR
     .......310.......320.
```

CTD

302

303   Figure 3. Sequence alignment of *Syn* 6803 OCP1, *Tolypothrix* OCP2, *Gloeobacter* OCP3 and HCP4

304   from *Anabaena* 7120. PBS binding residues in the *Syn* 6803 OCP-PBS complex are shaded red,

305   OCP^O dimer residues marked with an asterisk, OCP^R CTD dimer residues with a backslash, TDM

306   residues with a hash and the EPV motif in the linker is denoted with ampersands. Insertion

307   residues for HCP4 (pos. 15) and OCP3 (pos. 170) are shown in small print below the sequence.

308
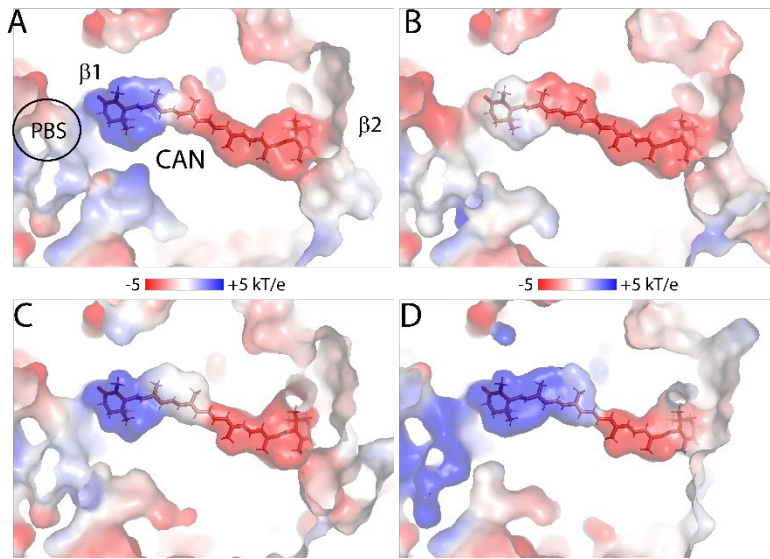
Figure 4. Electrostatic surface representation of the carotenoid environment of (A) the *Synechocystis* sp. PCC 6803 OCP1[R]-PBS complex cryo-EM structure (B-D) homology models based on the same PBS-OCP complex for (B) *Tolypothrix* sp. PCC 7601 OCP2 (C) *Gloeobacter kilaueensis* JS1 OCP3 and (D) *Anabaena* sp. PCC 7120 HCP4. Figure generated with PyMOL using the APBS plugin.

Declaration of Interests

The authors declare no competing interests.

REFERENCES

1. Schwieterman, E.W. et al. (2015) Nonphotosynthetic Pigments as Potential Biosignatures. Astrobiology. 15(5): p. 341-361.
2. Hirayama, O. et al. (1994) Singlet oxygen quenching ability of naturally occurring carotenoids. Lipids. 29(2): p. 149-50.
3. Vershinin, A. (1999) Biological functions of carotenoids--diversity and evolution. BioFactors. 10(2-3): p. 99-104.
4. Kirilovsky, D. and C. Kerfeld (2016) Cyanobacterial photoprotection by the orange carotenoid protein. Nature plants. 2(12): p. 16180.
5. Kerfeld, C.A. et al. (2017) Structure, function and evolution of the cyanobacterial orange carotenoid protein and its homologs. New Phytologist. 215(3): p. 937-951.
6. Wilson, A. et al. (2008) A photoactive carotenoid protein acting as light intensity sensor. Proc Natl Acad Sci U S A. 105(33): p. 12075-12080.
7. Adir, N. et al. (2020) The amazing phycobilisome. Biochim Biophys Acta Bioenerg. 1861(4): p. 148047.
8. Kuznetsova, V. et al. (2020) Comparative ultrafast spectroscopy and structural analysis of OCP1 and OCP2 from Tolypothrix. Biochimica Et Biophysica Acta-Bioenergetics. 1861(2).
9. Wu, Y.P. and D.W. Krogmann (1997) The orange carotenoid protein of Synechocystis PCC 6803. Biochimica et Biophysica Acta - Bioenergetics. 1322(1): p. 1-7.
10. Kerfeld, C. (2004) Structure and function of the water-soluble carotenoid-binding proteins of cyanobacteria. Photosynthesis Research. 81(3): p. 215-25.
11. Kerfeld, C. (2004) Water-soluble carotenoid proteins of cyanobacteria. Archives of biochemistry and biophysics. 430(1): p. 2-9.
12. Kerfeld, C. et al. (2003) The crystal structure of a cyanobacterial water-soluble carotenoid binding protein. Structure. 11(1): p. 55-65.
13. Wilson, A. et al. (2006) A soluble carotenoid protein involved in phycobilisome-related energy dissipation in cyanobacteria. The Plant cell. 18(4): p. 992-1007.
14. Wilson, A. et al. (2010) Structural determinants underlying photoprotection in the photoactive orange carotenoid protein of cyanobacteria. The Journal of biological chemistry. 285(24): p. 18364-75.
15. Leverenz, R.L. et al. (2014) Structural and functional modularity of the orange carotenoid protein: distinct roles for the N- and C-terminal domains in cyanobacterial photoprotection. The Plant cell. 26(1): p. 426-37.
16. Leverenz, R.L. et al. (2015) A 12 Å carotenoid translocation in a photoswitch associated with cyanobacterial photoprotection. Science. 348(6242): p. 1463-1466.
17. Gupta, S. et al. (2012) Structure and dynamics of protein waters revealed by radiolysis and mass spectrometry. Proc Natl Acad Sci U S A. 109(37): p. 14882-7.
18. Barends, T.R. et al. (2009) Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. Nature. 459(7249): p. 1015-8.
19. Swartz, T.E. et al. (2007) Blue-light-activated histidine kinases: two-component sensors in bacteria. Science. 317(5841): p. 1090-3.
20. Kirilovsky, D. and C.A. Kerfeld (2013) The Orange Carotenoid Protein: A blue-green light photoactive protein. Photochemistry and Photobiological Sciences 12: p. 1135-1143.
21. Lechno-Yossef, S. et al. (2017) Synthetic OCP heterodimers are photoactive and recapitulate the fusion of two primitive carotenoproteins in the evolution of cyanobacterial photoprotection. Plant Journal. 91(4): p. 646-656.

369 22. Bao, H. et al. (2017) Structure and functions of Orange Carotenoid Protein homologs in
370   cyanobacteria. Current Opinion in Plant Biology. 37: p. 1-9.
371 23. Melnicki, M. et al. (2016) Structure, Diversity, and Evolution of a New Family of Soluble
372   Carotenoid-Binding Proteins in Cyanobacteria. Mol Plant. 9(10): p. 1379-1394.
373 24. Dominguez-Martin, M.A. et al. (2019) Structural and spectroscopic characterization of HCP2.
374   Biochimica Et Biophysica Acta-Bioenergetics. 1860(5): p. 414-424.
375 25. Lopez-Igual, R. et al. (2016) Different Functions of the Paralogs to the N-Terminal Domain of the
376   Orange Carotenoid Protein in the Cyanobacterium *Anabaena* sp. PCC 7120. Plant physiology.
377   171(3): p. 1852-66.
378 26. Ho, M.Y. et al. (2017) Far-red light photoacclimation (FaRLiP) in Synechococcus sp PCC 7335: I.
379   Regulation of FaRLiP gene expression. Photosynthesis Research. 131(2): p. 173-186.
380 27. Yang, Y.W. et al. (2023) Functional specialization of expanded orange carotenoid protein paralogs
381   in subaerial Nostoc species. Plant Physiology. 192(4): p. 2640-2655.
382 28. Yang, Y.W. et al. (2019) Orange and red carotenoid proteins are involved in the adaptation of the
383   terrestrial cyanobacterium Nostoc flagelliforme to desiccation. Photosynthesis Research. 140(1):
384   p. 103-113.
385 29. Llewellyn, C.A. et al. (2020) Synthesis, Regulation and Degradation of Carotenoids Under Low
386   Level UV-B Radiation in the Filamentous Cyanobacterium Chlorogloeopsis fritschii PCC 6912.
387   Frontiers in Microbiology. 11.
388 30. Ho, M.Y. and D.A. Bryant (2019) Global Transcriptional Profiling of the Cyanobacterium
389   Synechocystis sp. PCC 9212 in Far-Red Light: Insights Into the Regulation of Chlorophyll Synthesis.
390   Frontiers in Microbiology. 10.
391 31. Dominguez-Martin, M.A. et al. (2020) Structural analysis of a new carotenoid-binding protein: the
392   C-terminal domain homolog of the OCP. Scientific Reports. 10(1).
393 32. Muzzopappa, F. et al. (2017) Paralogs of the C-Terminal Domain of the Cyanobacterial Orange
394   Carotenoid Protein Are Carotenoid Donors to Helical Carotenoid Proteins. Plant Physiology.
395   175(3): p. 1283-1303.
396 33. Slonimskiy, Y.B. et al. (2019) Light-controlled carotenoid transfer between water-soluble proteins
397   related to cyanobacterial photoprotection. Febs Journal. 286(10): p. 1908-1924.
398 34. Harris, D. et al. (2018) Structural rearrangements in the C-terminal domain homolog of Orange
399   Carotenoid Protein are crucial for carotenoid transfer. Communications Biology. 1.
400 35. Moldenhauer, M. et al. (2017) Assembly of photoactive orange carotenoid protein from its
401   domains unravels a carotenoid shuttle mechanism. Photosynthesis
402   Research(doi:10.1007/s11120-017-0353-3): p. 1-15.
403 36. Bao, H. et al. (2017) Additional families of orange carotenoid proteins in the photoprotective
404   system of cyanobacteria. Nature Plants. 3(8).
405 37. Slonimskiy, Y.B. et al. (2022) A primordial Orange Carotenoid Protein: Structure, photoswitching
406   activity and evolutionary aspects. International Journal of Biological Macromolecules. 222: p. 167-
407   180.
408 38. Boulay, C. et al. (2010) Identification of a protein required for recovery of full antenna capacity in
409   OCP-related photoprotective mechanism in cyanobacteria. Proc Natl Acad Sci USA. 107(25): p.
410   11620-5.
411 39. Sutter, M. et al. (2013) Crystal structure of the FRP and identification of the active site for
412   modulation of OCP-mediated photoprotection in cyanobacteria. Proc Natl Acad Sci USA. 110(24):
413   p. 10022-10027.
414 40. Steube, N. et al. (2023) Fortuitously compatible protein surfaces primed allosteric control in
415   cyanobacterial photoprotection. Nature Ecology & Evolution.

416 41. Petrescu, D.I. et al. (2021) Environmental Tuning of Homologs of the Orange Carotenoid Protein-
417   Encoding Gene in the Cyanobacterium. Frontiers in Microbiology. 12.
418 42. Bandara, S. et al. (2017) Photoactivation mechanism of a carotenoid-based photoreceptor. Proc
419   Natl Acad Sci USA. 114(24): p. 6286-6291.
420 43. Gupta, S. et al. (2015) Local and global structural drivers for the photoactivation of the orange
421   carotenoid protein. Proc Natl Acad Sci USA. 112(41): p. E5567-74.
422 44. Domínguez-Martín, M.A. et al. (2022) Structures of a phycobilisome in light-harvesting and
423   photoprotected states. Nature. 609(7928): p. 835-845.
424 45. Thierry, E. et al. (2016) Influenza Polymerase Can Adopt an Alternative Configuration Involving a
425   Radical Repacking of PB2 Domains. Molecular Cell. 61(1): p. 125-137.
426 46. Punjani, A. and D.J. Fleet (2021) 3D variability analysis: Resolving continuous flexibility and
427   discrete heterogeneity from single particle cryo-EM. Journal of Structural Biology. 213(2).
428 47. Sauer, P.V. et al. (2024) Structural and quantum chemical basis for OCP-mediated quenching of
429   phycobilisomes. Sci Adv.
430 48. Muzzopappa, F. et al. (2019) Interdomain interactions reveal the molecular evolution of the
431   orange carotenoid protein. Nature Plants. 5(10): p. 1076-1086.
432 49. Rose, J.B. et al. (2023) Photoactivation of the orange carotenoid protein requires two light-driven
433   reactions mediated by a metastable monomeric intermediate. Physical Chemistry Chemical
434   Physics. 25(48): p. 33000-33012.
435 50. Andreeva, E.A. et al. (2022) Oligomerization processes limit photoactivation and recovery of the
436   orange carotenoid protein. Biophysical Journal. 121(15): p. 2849-2872.
437 51. Tsoraev, G.V. et al. (2023) Stages of OCP-FRP Interactions in the Regulation of Photoprotection in
438   Cyanobacteria, Part 1: Time-Resolved Spectroscopy. Journal of Physical Chemistry B.
439 52. Sluchanko, N.N. et al. (2018) OCP-FRP protein complex topologies suggest a mechanism for
440   controlling high light tolerance in cyanobacteria. Nature Communications. 9.
441 53. Sluchanko, N.N. et al. (2024) Structural framework for the understanding spectroscopic and
442   functional signatures of the cyanobacterial Orange Carotenoid Protein families. International
443   Journal of Biological Macromolecules. 254.
444 54. Arcidiacono, A. et al. (2023) How orange carotenoid protein controls the excited state dynamics
445   of canthaxanthin. Chemical Science. 14(40): p. 11158-11169.
446 55. De Souza, A.P. et al. (2022) Soybean photosynthesis and crop yield are improved by accelerating
447   recovery from photoprotection. Science. 377(6608): p. 851-+.
448 56. Abed, R.M.M. et al. (2009) Applications of cyanobacteria in biotechnology. Journal of Applied
449   Microbiology. 106(1): p. 1-12.
450 57. Kloz, M. et al. (2011) Carotenoid Photoprotection in Artificial Photosynthetic Antennas. Journal of
451   the American Chemical Society. 133(18): p. 7007-7015.
452 58. Piccinini, L. et al. (2022) A synthetic switch based on orange carotenoid protein to control blue-
453   green light responses in chloroplasts. Plant Physiology. 189(2): p. 1153-1168.
454 59. Dominguez-Martin, M.A. and C.A. Kerfeld (2019) Engineering the orange carotenoid protein for
455   applications in synthetic biology. Current Opinion in Structural Biology. 57: p. 110-117.
456 60. Chothia, C. and J. Gough (2009) Genomic and structural aspects of protein evolution. Biochemical
457   Journal. 419: p. 15-28.
458 61. Kondrashov, F.A. et al. (2002) Selection in the evolution of gene duplications. Genome Biology.
459   3(2).
460 62. Galhardo, R.S. et al. (2007) Mutation as a stress response and the regulation of evolvability.
461   Critical Reviews in Biochemistry and Molecular Biology. 42(5): p. 399-435.

462

TEXT BOX—Puzzling Pieces: The HCPs

Protein domains are evolutionary modules/units of structure and function. Once internally duplicated within a genome, a domain is under less selective pressure to remain the same. Phylogenomic analysis suggests that HCP paralogs and the NTD of the OCP arose from gene duplication and subfunctionalization, with PBS quenching just one of the possible functions [23]. Despite structural redundancy, including the binding or carotenoids, the relatively low sequence homology between HCP paralogs within a genome supports the idea that each has a distinct function. For example, *Anabaena* 7120 which contains HCP1-HCP4, with the exception of the HCP1:HCP4 homology (49% identity) all other pairwise alignments are between 23-36% identical. Change in function is typically associated with sequence identity less than 40% [60]. In protein evolution the acquisition of new function is the strongest driver of protein sequence change with paralogs evolving more rapidy than their corresponding orthologs [61]. Furthermore, cellular stress has been shown to correlate with high mutation rates [62]; and the inherent dangers of light harvesting as well as the advent of oxygenic photosynthesis would suggest primitive carotenoproteins like the HCPs strong candidates for mutation and divergence. We suggest that the HCP families evolved discrete functions, with only some, such primitive HCP4, able to bind and quench the PBS. Accordingly, the extant function of, for example, *Anabaena* 7120 HCP1-3 likely reflects their more ancient roles as cyanobacterial carotenoproteins.

Glossary

Carotenoids

Yellow, orange or red pigments that are produced by a variety of organisms and are frequently involved in photosynthesis. The basic building blocks are isoprene units that are linked into chains of conjugated double bonds.

Accessory pigments

Accessory pigments function to extend the range of light that photosynthetic organisms can utilize but are not directly used for the conversion of light energy into cellular energy.

Chromophore

A molecule that can absorb light and emit in the visible spectrum, hence giving it a color. A carotenoid, for example, achieves this property by multiple conjugated double bonds.

Photoprotection

Mechanisms of photosynthetic organisms to prevent damage to the photosystem caused by excess light. In Cyanobacteria, the OCP is the primary means of non-photochemical quenching and dissipates excess energy absorbed by phycobilisomes and prevents damage to the Photosystem II.

Phycobilisome (PBS)

Light harvesting antenna complex of cyanobacteria and certain algae to facilitate light absorption. They are large protein complexes consisting of several hundreds of protein chains that form cylindrical cores and rods. There are two types of proteins, phycobiliproteins that bind chromophores and linker proteins to connect them.

Helical Carotenoid Proteins

Homologs to the N-terminal domain of OCP. They contain a carotenoid in the same position as an activated OCP^R.