Structure, Vol. 11, 55–65, January, 2003, ©2003 Elsevier Science Ltd. All rights reserved. PII S0969-2126(02)00936-X

The Crystal Structure of a Cyanobacterial Water-Soluble Carotenoid Binding Protein

Cheryl A. Kerfeld,[1,*] Michael R. Sawaya,[1] Vishnu Brahmandam,[1] Duilio Cascio,[1] Kwok Ki Ho,[2] Colleen C. Trevithick-Sutton,[3] David W. Krogmann,[2] and Todd O. Yeates[1,3,4]

[1]Molecular Biology Institute
University of California, Los Angeles
Los Angeles, California 90095

[2]Biochemistry Department
Purdue University
West Lafayette, Indiana 47907

[3]UCLA-DOE Center for Genomics and Proteomics
Los Angeles, California 90095

[4]Department of Chemistry and Biochemistry
University of California, Los Angeles
Los Angeles, California 90095

Summary

Carotenoids undergo a wide range of photochemical reactions in animal, plant, and microbial systems. In photosynthetic organisms, in addition to light harvesting, they perform an essential role in protecting against light-induced damage by quenching singlet oxygen, superoxide anion radicals, or triplet-state chlorophyll. We have determined the crystal structure of a water-soluble orange carotenoid protein (OCP) isolated from the cyanobacterium *Arthrospira maxima* at a resolution of 2.1 Å. OCP forms a homodimer with one carotenoid molecule per monomer. The carotenoid binding site is lined by a striking number of methionine residues. The structure reveals several possible ways in which the protein environment influences the spectral properties of the pigment and provides insight into how the OCP carries out its putative functions in photoprotection.

Introduction

Carotenoids are virtually ubiquitous among microorganisms, animals, and plants, where they play a key role in protecting cells from damaging oxygen radicals. The vast majority of these carotenoids are found embedded in membranes, lipid globules, and other hydrophobic environments.

The carotenoid-mediated protective quenching of singlet oxygen is especially important in oxygen-evolving photosynthetic organisms, which generate oxidizing molecules under a broad range of light intensities. Under high illumination, or under other stresses that limit downstream photosynthetic events, excitation energy in the chlorophyll protein complexes exceeds what can be utilized in photosynthesis. Excited states of chlorophyll accumulate and can convert to an excited triplet state, which readily reacts with oxygen to form singlet oxygen. Carotenoids have long been known to interact with triplet chlorophyll, deexciting it in a preemptive reaction before singlet oxygen can be formed (reviewed in [1–3]). For example, transforming cyanobacteria with various carotenogenic enzymes has been shown to confer protection from UV-B radiation [4]. UV-B radiation results in the formation of free radicals that can destroy the D1 and D2 proteins of photosystem II through peroxidation reactions.

In photosynthetic organisms, in addition to playing a role in photoprotection, carotenoids function in light harvesting. Their ability to absorb light in the blue-green region of the spectrum extends the range of usable light beyond what can be absorbed by chlorophyll and related pigments alone. Carotenoids also appear to play a structural role in some proteins, as shown in the structure of the plant light-harvesting complex [5]. In photosynthetic organisms, carotenoids carry out their specialized functions as noncovalently bound components of protein-pigment complexes. These may be integral membrane, membrane associated, or, more rarely, water soluble. There is structural information available on several integral membrane proteins that bind carotenoid in combination with chlorophyll cofactors as part of photosynthetic reaction centers [6, 7], light-harvesting complexes [5, 8, 9], or in photosystems [10]. A water-soluble peridinin-chlorophyll protein carotenoid has also been structurally characterized [11]. In contrast, much less is known structurally about proteins that bind only carotenoid pigments. The structure of a carotenoid binding subunit of a hexadecameric complex isolated from lobster carapace has been determined, but without pigment [12].

A 35 kDa water-soluble orange carotenoid protein (OCP) was first identified in three genera of cyanobacteria in 1981 [13]. Similar proteins were subsequently characterized in several different cyanobacterial species in both natural blooms and laboratory cultures [14–17]. However, these data vary with regard to cellular location, detergent versus water solubility, the number of associated pigment molecules, and molecular mass of the protein. The function of the soluble cyanobacterial carotenoid proteins has not been established firmly, but levels of the protein increase during exposure to intense light [15, 18]. A photoprotective role is also consistent with the absence of chlorophyll or other tetrapyrrole pigments in the OCP, which act as energy acceptors in light-harvesting complexes, but which would be unnecessary in a photoprotective protein. There is some data indicating that photoprotection and light harvesting are carried out by different proteins in higher plants [19].

Here we report the first crystal structure of a protein that binds carotenoids exclusively, the OCP isolated from the cyanobacterium *Arthrospira maxima*. The OCP contains the keto-carotenoid 3-hydroxyechinenone. Levels of 3-hydroxyechinenone are known to increase under stress conditions in cyanobacteria (summarized in [20]).

*Correspondence: kerfeld@mbi.ucla.edu

Key words: carotenoid; methionine; photoprotection; photosynthesis; pigment-protein

Table 1. Data Collection and Refinement Statistics

Data collection	Native	Hg (peak)	Hg (inflection)	Hg (remote)
Space group	C2		C2	
Unit cell dimensions (a, b, c [A], β)	217.3, 40.8, 74.5, 95.8		206.6, 38.6, 75.0, 94.8	
Wavelength (Å)	1.0094	1.00518	1.0094	0.9892
Resolution limit (Å)	2.10	2.50	2.55	2.55
R_{sym} (%)[a]	6.5	6.6	7.8	9.1
R_{sym} (%, last shell)	38.6	23.0	40.0	33.3
I/σ (last shell)	3.5	4.9	2.5	3.9
Total observations	131,739	98,199	55,096	73,565
Unique reflections	37,373	20,903	19,542	19,719
Completeness (%)	94.5	99.5	99.4	99.8
Completeness (%, last shell)	84.3	99.9	98.6	99.9
Phase determination[b]				
R_{Cullis}[c] (%, 20–2.6 Å, acentric/centric, isomorphous)		0.89/0.81		0.93/0.88
R_{Cullis}[d] (%, 20–2.6 Å, anomalous)		0.77	0.93	0.86
Phasing power[e] (20–2.6 Å, acentric/centric)		0.73/0.61		0.50/0.44
Number of sites			6	
Mean overall figure of merit (before/after DM)			0.414/0.709	

Model refinement	Native
R_{work}[f] (20–2.1 Å)	21.5
R_{free}[g] (20–2.1 Å)	27.5
PDB ID code	1M98
Number of residues (protein/water)	630/316
Average B (main chain/side chain)	12.2/18.3
Rmsd bonds (Å)	0.022
Rmsd angles (°)	1.9
B values (Å^2 bonded)	1.6

[a] $R_{sym} = \Sigma|I - <I>|/\Sigma I$
[b] The inflection data set was treated as a reference for phasing.
[c] $R_{Cullis} = \Sigma\epsilon/\Sigma|F_{PH} - F_P|$, where ϵ = lack of closure.
[d] $R_{Cullis} = \Sigma\epsilon/\Sigma|F^+ - F^-|$, where ϵ = lack of closure.
[e] Phasing power = $<F_H/\varepsilon>$.
[f] $R_{work} = \Sigma|F_{obs} - F_{calc}|/\Sigma F_{obs}$
[g] $R_{work} = \Sigma|F_{obs} - F_{calc}|/\Sigma F_{obs}$, where all reflections belong to a test set of 5% randomly selected data.

Although this carotenoid is not found in higher plants, the introduction of an algal carotenoid biosynthetic gene into tobacco resulted in the accumulation of 3-hydroxyechinenone and other algal pigments [21]. The carotenoid pigment 3-hydroxyechinenone appears yellow (λmax = 450 nm) when isolated in organic solvent, whereas the protein-bound form appears either orange (λmax = 495 nm and 465 nm) or red (λmax = 505 nm). The crystal structure of the *A. maxima* OCP provides structural insights into how the protein environment tunes the spectral properties of the pigment. The structure also provides clues into the OCP's putative function(s) in the vital processes of photoprotection.

Results and Discussion

Structural Overview

The structure of the OCP was determined to 2.1 Å resolution using MAD phasing on a mercury derivative (Table 1). The model is complete relative to the primary structure determined by PCR amplification and sequencing of the gene, with the exception of the N-terminal Met, which is absent in the purified protein. The atomic model is of high quality. All residues in the OCP fall within allowed regions of Ramachandran plots, and all residues are below the 95% confidence limit suggested by ERRAT [22].

The fold of OCP is composed of thirteen α helices (A–N) and six β strands (Figures 1A and 2A) organized into two major domains: an α/β domain and an all-helical domain. The carotenoid spans both domains. Loops connecting the domains (residues 127–129 and 161–185) display higher temperature factors relative to the rest of the model. Searches of the database of known three-dimensional protein structures indicate there are no structural homologs to the overall fold of the OCP.

The all-helical N-terminal domain can be divided into two four-helix bundles. Because these separate bundles are composed from discontinuous segments of the protein sequence and are connected by two crossing regions of the protein, we do not describe them as separate domains here. Bundle 1 is formed by amino acids 19–51 and 131–161 (helices B, C, H, and I in Figure 1; Figure 2A). Bundle 2 is formed from residues 55–90 and 92–122 (helices D–G in Figures 1 and 2A). The two bundles are somewhat similar and can be superimposed with a root-mean-square (rms) Cα deviation of 3.2 Å. While the separate four-helix bundles resemble those seen in many other proteins, the intact domain (containing two bundles) is not similar to any previously reported protein structure. The cleft between the two bundles forms the binding site of the hydroxy end of the 3-hydroxyechinenone molecule.

The C-terminal domain (residues 190–317) of OCP resembles an NTF2 (nuclear transport factor 2) domain

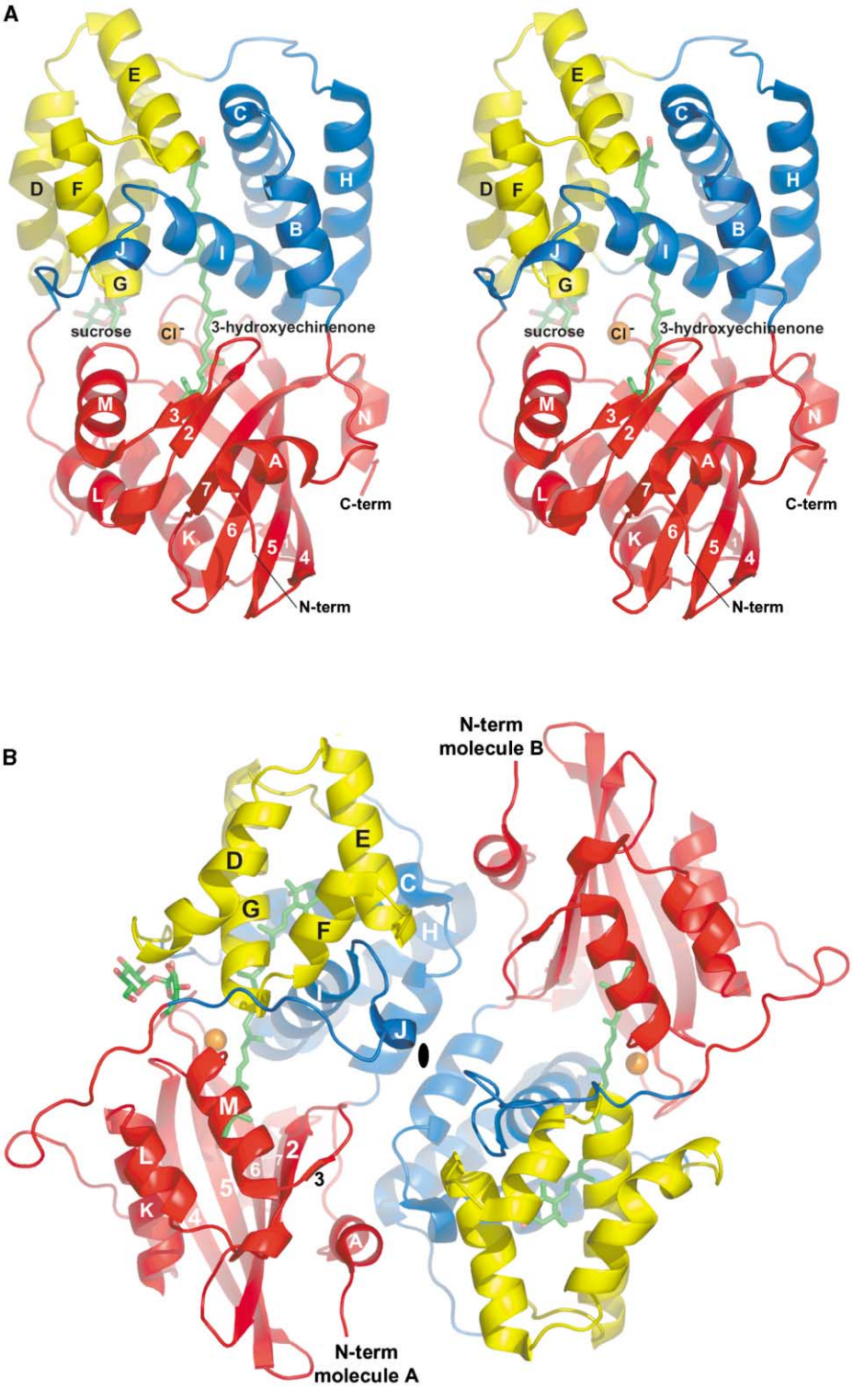
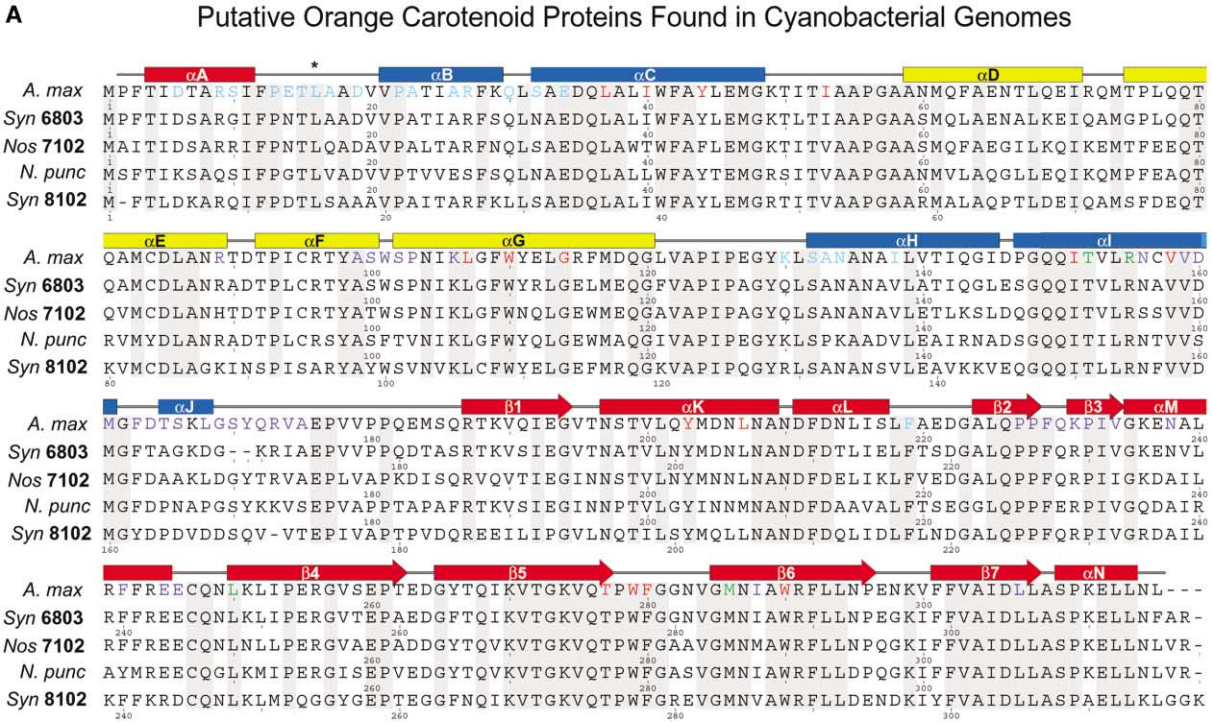


Figure 1. Structure of the Orange Carotenoid Protein (OCP)

(A) A stereo view of the OCP monomer. The two four-helix bundles of the N-terminal domains are shown in blue and yellow. The C-terminal α/β (NTF2-like) domain is shown in red. Strands and helices are sequentially labeled (see also Figure 2A). The 3-hydroxyechinenone carotenoid molecule in shown as sticks, with the keto group at the lower end and the hydroxyl group at the top. Also visible are the chloride (Cl^-) ion and the sucrose molecule associated with one of the molecules in the homodimer.

(B) The natural state of the OCP is a dimer. This view down the dyad axis shows the two carotenoid molecules aligned essentially antiparallel to each other. This figure and Figure 3 were prepared with PyMOL [52].



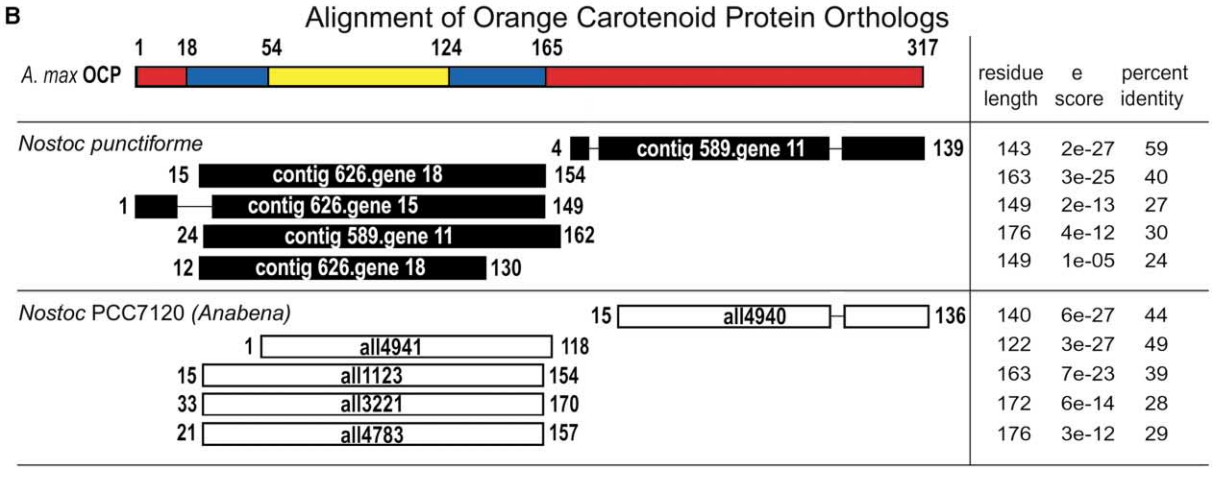


Figure 2. Alignment of the *Nostoc punctiforme* and *Nostoc* PCC7102 OCP Homologs with the *A. maxima* OCP

(A) Sequence alignment of putative orange carotenoid proteins found in cyanobacterial genomes. Abbreviations are as follows: *A. max*, *Arthrospira maxima* (this work); *Syn* 6803, *Synechocystis* PCC6803 (Kazusa Institute, www.kazusa.or.jp/); *Nos* 7102, *Nostoc* PCC7102 (Kazusa Institute, www.kazusa.or.jp/); *N. punc*, *Nostoc punctiforme* (DOE Joint Genome Institute, www.jgi.doe.gov/JGI_microbial/html/). The coloring of the secondary structure for the OCP sequence corresponds to the domains and subdomains as colored in Figure 1. An asterisk marks the beginning of the RCP. Color coding of amino acids is as follows: aqua, residues involved in dimerization; red, residues in the carotenoid binding pocket; purple, residues forming the surface of the large cavity; green, residues forming both the carotenoid binding pocket and the large cavity. Absolutely conserved residues are in gray boxes.

(B) The *A. maxima* OCP primary structure is represented uppermost, colored as in Figure 1. Aligned below the OCP are open reading frames from *N. punctiforme* and *Nostoc* PCC7120 genomes, positioned with respect to their overall alignment with the *A. maxima* OCP. The designation of the open reading frame is given. The numbers flanking each bar correspond to the position within the contig that aligns with the OCP. The alignment score (E score) and identity calculated by BLAST are shown for each on the right. *Nostoc* PCC7120 is DOE project number 63737. *N. punctiforme* was sequenced by the Kazusa Institute. The coloring of the *A. maxima* sequence corresponds to the domains and subdomains as colored in Figure 1.

[23]. The NTF2 fold is an α /β fold with a distinctively bent β sheet that forms a hydrophobic pocket in its interior. In the OCP, this hydrophobic cleft is occupied by the keto end of the 3-hydroxyechinenone molecule (Figure 1). This domain has been observed in several functionally diverse proteins such as the NTF2-GDP-Ran complex that functions in nuclear import [24], the TAP/p15-mRNA export factor [25], and the enzymes ketosteroid isomerase [26] and naphthalene 1,2-dioxygenase [27].

Phylogenetic Distribution of the OCP

The primary structure of the *A. maxima* OCP shows no significant similarity to any proteins of known function in the sequence databases. Initially, N-terminal sequencing of the *A. maxima* OCP suggested that it was related to the product of the open reading frame slr1963 in the *Synechocystis* PCC 6803 genome [28]. Using the recently fully sequenced *A. maxima* OCP gene, a current survey of the six available cyanobacterial genome sequence databases reveals that an open reading frame for the full-length OCP is present in *Synechococcus* WH8102 and *Nostoc* PCC7102 (also known as *Anabaena*). There are two full-length OCP homologs in *Nostoc punctiforme*. Alignment of these sequences shows that the primary structure of the OCP protein is remarkably highly conserved, especially within the C-terminal NTF2-like domain (Figure 2A). Each of the two *Nostoc* genomes also contains five shorter proteins, each of which is homologous to either the N-terminal or the C-terminal domain of the OCP. In both genomes, one of these shorter proteins corresponds to the OCP's C-terminal domain, while there are four paralogs corresponding to the OCP N-terminal domain (Figure 2B).

The OCP is not detected in the genomes of the cyanobacterium *Prochlorococcus marinus* MIT 9313, or in the partial genome of *Prochlorococcus marinus* MED4. The pigment composition of the *Prochlorococcus* genus is unique and lacks the phycobilisomes found in typical cyanobacteria. The *Prochlorococci*, with their relatively small genomes and distinct photosynthetic apparatus, form a separate branch of the cyanobacteria species [29]. Nor is the OCP detectable in the genomes of anoxygenic photosynthetic bacteria (*Chlorobium tepidum*, *Rhodobacter sphaeroides*, *Rhodopseudomonas palustris*, and *Rhodospirillum rubrum*), in the green alga *Chlamydomonas reinhardtii*, or in the higher plant genomes of *Arabidopsis thaliana* and rice. This distribution suggests that the OCP is particularly associated with oxygenic photosynthesis typical of most cyanobacteria, but became unnecessary after compartmentalization of photosynthesis in the chloroplast.

The OCP Homodimer and Crystal Packing

The *A. maxima* OCP elutes as a dimer in size exclusion chromatography, and the asymmetric unit of the OCP crystal appears to contain the physiological homodimer (Figure 1B). Dimerization buries $\sim$1411 Å^2 of surface area per OCP monomer. In contrast to other proteins containing NTF2-like domains that dimerize across the β sheet, the OCP monomers associate across helices A, B, and H and the loop regions following helix B and preceding helix H. The dimeric interactions are both hydrophobic and polar, with intermolecular salt bridges and hydrogen bonds formed between Asp6 and Gln88, and between Asp19 and both Arg27 and Asn134. Six of the 22 residues involved in the homodimeric interaction are absolutely conserved in the OCP primary structures, while the other 16 residues are conservative substitutions (Figure 2A), suggesting that this dimerization interface may be a common feature of cyanobacterial OCPs. The two molecules of OCP in the homodimer superimpose with an rms deviation on all atoms of 0.24 Å. The largest differences between the two molecules are in the loops defined by residues 124–129, 163–173, and 179–186.

Previously [30], we described an unusual optical property of OCP crystals. When viewed through a single polarizing filter, each crystal was either orange or perfectly colorless, depending upon the orientation of the long axis of the crystal relative to the direction of polarization. The parallel alignment of the two carotenoids in the dimer (Figure 1B) and in the unit cell explains this effect. The pigment molecules are all aligned very nearly parallel to each other and to the c-axis of the crystals.

Protein-Carotenoid Interactions

The pigment in the OCP has been identified as the ketocarotenoid, 3-hydroxyechinenone (Figure 3A). There is one carotenoid bound per protein chain. In the OCP monomer, only 3.4% of the surface of the carotenoid is solvent accessible (35 Å^2 out of a total of 1019 Å^2 surface area for an exposed carotenoid).

As noted above, the carotenoid spans both domains of the protein, with its keto terminus nestled into a hydrophobic pocket within the NTF2 domain and the hydroxyl terminus lodged between the two four-helix bundles (Figures 1 and 3). The residues within 3.7 Å of the carotenoid are shown in red in Figure 2A. Twelve of these 16 residues are absolutely conserved in the OCP primary structures. At one end of the carotenoid, the keto oxygen atom of the 3-hydroxyechinenone molecule is hydrogen bonded to the side chains of the absolutely conserved Trp290 (2.8 Å) and Tyr203 (2.7 Å) in the C-terminal domain.

In contrast, at the other end of the carotenoid, the hydroxyl oxygen atom of the carotenoid does not hydrogen bond to the protein. Instead, it hydrogen bonds to a water molecule that is present in both molecules of the homodimer (Figure 3A; HOH 452 in molecule A and HOH453 in molecule B). This region of the 3-hydroxyechinenone molecule is one of only two sites where the carotenoid is accessible to solvent (Figures 4A and 4B). There are other water molecules buried in the carotenoid binding cleft of each OCP monomer (four in molecule A, two in molecule B), but they are not positionally conserved.

With the exception of the head groups, the 3-hydroxyechinenone molecule is approximately symmetrical. The hydrogen bonding to the protein at the keto end of the pigment is presumably important in binding the carotenoid molecule in the correct orientation within the protein. If the orientation of the pigment were reversed, with the keto end inserted in the all-helical domain and the

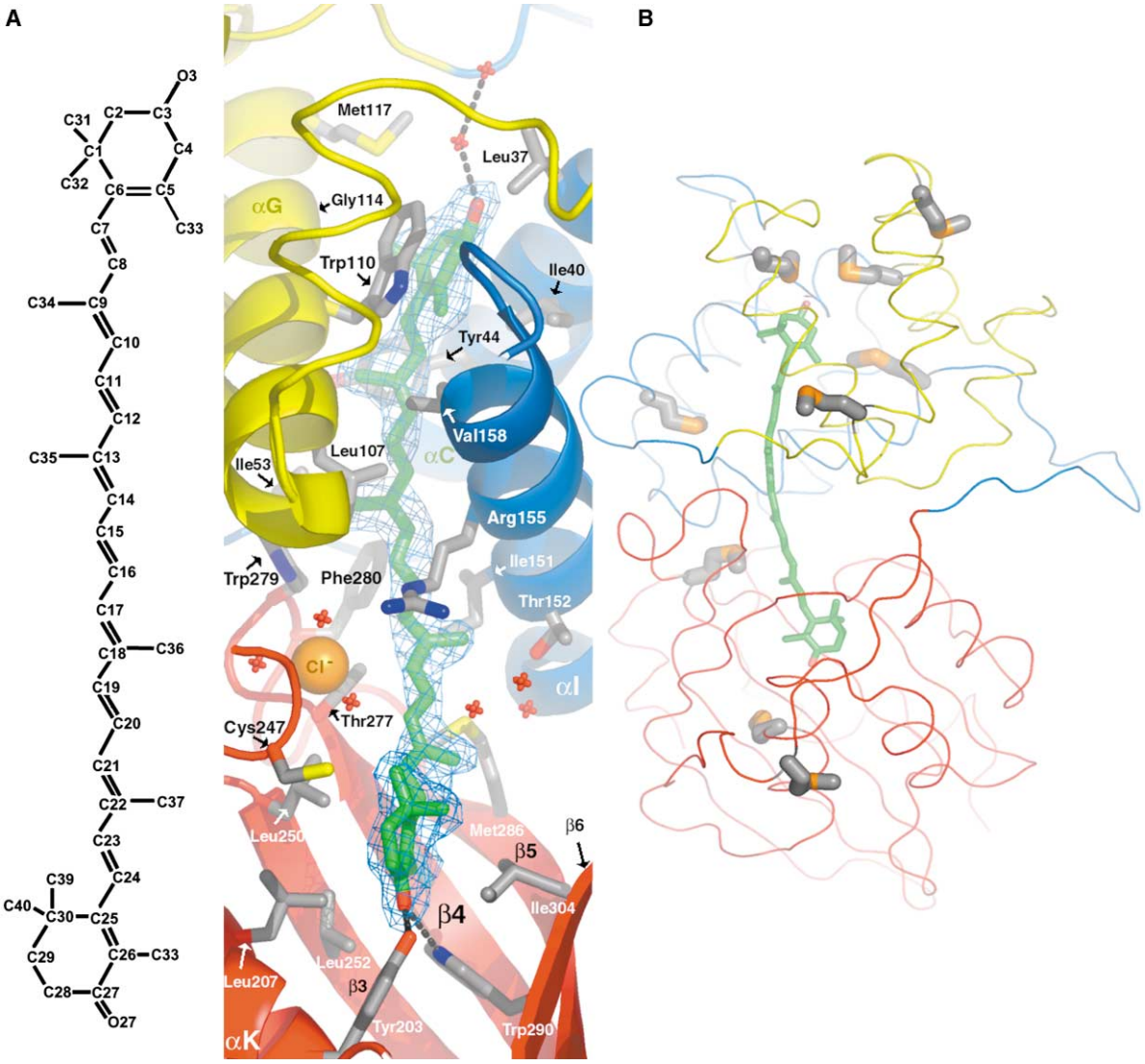


Figure 3. The 3-Hydroxyechinenone Carotenoid Molecule in the OCP

(A) A view of the 3-hydroxyechinenone-protein interactions in OCP. Side chains of residues within 3.7 Å of the pigment are shown. The colors of the domains and the orientation of the center view are the same as in Figure 1. A simulated annealing omit map (Fo − Fc electron density contoured at 3.2 σ) is shown for the carotenoid. The chemical structure of 3-hydroxyechinenone is shown on the left.

(B) A view of the methionine residues in the OCP, six of which have sulfur atoms in proximity (<6.5 Å) of the carotenoid molecule.

hydroxyl end within the C-terminal α/β domain, there would be numerous bad contacts between the protein and pigment, mainly within the C-terminal domain. In the all-helical domain, the side chain of Trp110 would clash with C39 of the carotenoid (as close as 1.7 Å). In this orientation, the hydroxyl oxygen would be able to hydrogen bond with Tyr203, but not with the side chain of Trp290. In the all-helical domain, the keto moiety would be prevented from forming any hydrogen bonds by the side chains of Met117 and Trp41. Both of these residues and Trp110 are absolutely conserved in the primary structure of the OCP, where they may be essential to ensure proper orientation of the pigment.

In the OCP, the carotenoid molecule adopts a *trans* configuration and is approximately 26 Å in length. The radius of curvature is approximately 28 Å for both carotenoids of the dimer. The curvature is achieved by the cumulative effects of small torsions relative to the ideal *trans* configuration. The average deviation from *trans* (180°) is 16° for both carotenoid molecules, although the two carotenoid molecules differ in minor detail. The deviation from planarity could be an important feature in tuning the spectral properties of the bound carotenoid.

The protein environment plays an important role in tuning the spectral properties of the pigment. For example, the absorbance maxima of photosynthetic carotenoids in vivo are "red" (bathochromically) shifted (up to about 30 nm) relative to their absorption in solution. This is likely due to the influence of charged and polarizeable residues on the pigment [31]. One striking feature

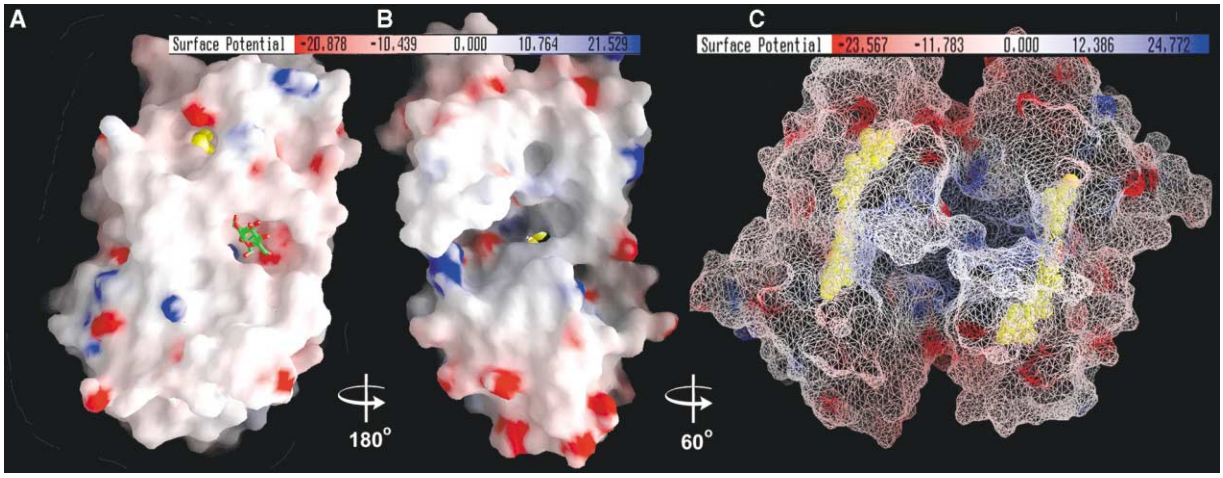


Figure 4. Topography and Charge of the OCP Surface

(A) A view of the "backside" (~180 degree rotation) relative to the view in Figure 1A. The carotenoid is colored yellow. The cavity containing the sucrose molecule found in one of the monomers of the OCP is shown.

(B) A view of the OCP similar to Figure 1A, showing the large surface cavity and the slight accessibility of the carotenoid molecule (in yellow).

(C) A view of the OCP dimer (orientation similar to Figure 1B) showing the continuity of the surface cavities. The side chains of Lys167 from the two monomers appear to contact each other on the surface of the molecule. This figure was prepared with GRASP [53].

of the OCP is the unusual number of methionine residues distributed around the carotenoid. The primary structure of *A. maxima* OCP contains nine methionine residues; eight of these are conserved in the other cyanobacterial open reading frames corresponding to OCP (Figure 2A). In the *A. maxima* OCP structure, six methionine residues (five absolutely conserved among the known primary structures of OCP) are positioned so that their thioether groups extend toward the carotenoid molecule, two in the β domain and four within the N-terminal helical bundle domain (Figure 3B). Specifically, the six sulfur atoms are within an arbitrary distance of 6.5 Å of the carotenoid. A statistical calculation shows that the probability that six of the nine methionine residues should fall so close to the carotenoid is less than one in 1000. A similar abundance of methionine residues is not observed in other known structures of carotenoid binding proteins. Several do contain methionine residues in their carotenoid binding pockets but not in significant numbers.

The polarizeability of the methionine sulfur atom may be an important factor in binding the highly conjugated pigment molecule or in tuning its spectral characteristics. A methionine residue has been shown to be important for tuning the spectral characteristics of bacteriophytochrome [32]. Furthermore, the possible oxidation of the thioether moiety would markedly increase the polarity of the methionine side chain. This, as a response to high concentrations of oxygen radicals, could have significant effects on the spectral characteristics of the protein. A special role for methionines has also been noted for the binding of peptides by calmodulin [33]. Our observation identifies another instance where the unusual properties of methionine may be important in binding and molecular recognition.

Surface Cavities

The surface of the OCP has numerous protuberances and pockets (Figure 4). The pit formed by residues Ala55, Pro56, Gly57, Ala58, Ala59, Asn60, Asn104, Glu176, Gln248, Asn249, Pro278, and Trp279 is filled by a sucrose molecule (used in crystallization) in monomer A (Figures 1 and 4A). All but two (Asn60 and Asn249) of these residues are absolutely conserved in the primary structures of OCP (Figure 2A). Growth of the *A. maxima* OCP crystals required the addition of sucrose, methylpentanediol, or glycerol [30].

There is one unusually large surface cavity that leads to the carotenoid binding cleft (Figure 4B). Using CAST [34], this surface cavity was found to be six times larger than any of the others identified on the OCP surface. This cavity is bounded by regions from both domains, including the loops between strands 1 and 2, and between helices F and G, and I and J. Helices I and M and β strands 2 and 3 also bound part of this pocket. The volume of the surface cavity is 895 $Å^3$ as defined by CAST [34]. The surface cavity is lined with both polar and nonpolar residues (purple in Figure 2A) and is continuous with the portion of the carotenoid binding cleft formed by the C-terminal NTF2-like domain. The cavity and the carotenoid binding cleft also share a chloride ion that reduces the solvent accessibility of the carotenoid (Figure 4B). This chloride ion is positionally conserved in both of the OCP monomers, within 3.9 Å of the pigment. It was assigned as a chloride ion based on the magnitude of the electron density and the crystallization conditions.

The large, polar opening to this cavity has a surface area of 160 $Å^2$ and dimensions of approximately 12.7 Å (Ser166- Lys231) by 19.3 Å (Arg241-Asn156). Other residues forming the boundary of the surface cavity are shown in purple in Figure 2A. Part of the OCP primary structure forming the cavity (residues 162–167) bears some similarity to a loop (GFDPLG/F) implicated to be important in binding carotenoid and in stabilizing the light-harvesting complexes of plants [35, 36] and red algae [37]. Accessibility to the cavity is not occluded by

the dimerization. Instead, the cavities from each monomer are continuous in the dimer and shielded from solvent only by the Lys167 side chains of each monomer, which meet at the 2-fold symmetry axis of the dimer (Figure 4C).

Structural Insights into the Putative Functions of the OCP

In general, the polypeptide chains of pigment proteins serve three functions: to provide a hydrophobic environment for the apolar pigment, to tune the absorption properties of the pigment, and to arrange pigments to facilitate efficient energy transfer between them. In the OCP, a fourth role for the protein can be envisioned: to facilitate controlled quenching of reactive oxygen species. The OCP structure has features consistent with a role in each of these processes.

For example, OCP has been suggested to be a carotenoid transport protein involved in photoprotection [13, 28]. The OCP may shuttle the carotenoid from its site of synthesis in the thylakoid membrane to the cytoplasmic or outer membrane of the organism where the carotenoids are arranged uniformly to form a photoprotective shield [14]. If the carotenoid remains protein bound as part of the photoprotective shield, the parallel packing of the molecules in the OCP crystal demonstrates one way the OCP molecules could pack densely to optimize light absorption. Alternatively, if the OCP functions only as a shuttle, the large cavity on the surface of the protein could be the site for release of the carotenoid at its destination. In this context, one might speculate that the highly conserved sucrose binding site could possibly be an allosteric binding site to control binding and release of the carotenoid. The sucrose is bound at the hinge between the two domains (Figure 1). Furthermore, the presence of an NTF2-like domain, which is found in several transport proteins, may be considered as suggestive of a transport function for OCP.

Another proposed role for the OCP is to quench singlet excited state molecular oxygen generated in photosynthesis. Oxygen could penetrate the protein through the two sites in which the carotenoid is accessible to solvent (Figures 4A and 4B). The carotenoid molecule, 3-hydroxyechinenone, isolated from *A. maxima* OCP, has a near diffusion-controlled rate constant for quenching singlet oxygen of 2.00×10^{10} $M^{-1}s^{-1}$ in chloroform (data not shown), which is typical of the quenching rates for carotenoids [38]. The quenching rate constant of the OCP is lower, 7.56×10^{8} $M^{-1}s^{-1}$ in D_2O (data not shown), presumably because the carotenoid is much less accessible inside the protein. However, the singlet oxygen quenching rate of the OCP is in the same range as other antioxidant proteins such as superoxide dismutase [38].

In order for the OCP to participate in thermal dissipation of excitation energy or in light harvesting, it would be required to interact with chlorophyll-containing proteins. The surface of OCP is rough (Figure 4), containing numerous protuberances and cavities that may be important for interacting with other polypeptides. Interestingly, all proteins known to contain an NTF2 domain form oligomers, utilizing the surface of the NTF2 β sheet for interaction with another monomer or another protein. In the OCP, the β sheet is exposed to solvent.

The dramatic influence of the protein environment on the spectral properties of 3-hydroxyechinenone, and hence its potential utility in light harvesting or in nonradiative dissipation of energy, is underscored by the conversion of OCP into a different spectral form. OCP, by proteolysis, converts into a red carotenoid protein (RCP; $\lambda max = 505$ nm) that is isolated from cells concomitantly with OCP. The biological significance of this derivative form, if any, is unknown. N-terminal sequencing and mass spectrometry indicate that RCP is a product of proteolysis at both the N and C termini of OCP. The N-terminal sequence of RCP begins with residue 16 of OCP. The C-terminal proteolysis site, deduced by comparison of the masses of OCP and RCP, appears to be in the disordered loop connecting the four-helix bundle domain to the NTF2-like domain (Figure 1A). Structurally, the removal of the C-terminal domain would result in the exposure of approximately half of the 3-hydroxyechinenone to solvent, which could explain the observed spectral shift of the pigment. The RCP also has a higher quenching rate constant (4.52×10^{9} $M^{-1}s^{-1}$), most likely due to the greater solvent accessibility of the pigment. In low ionic strength solution, the RCP tends to form an orange aggregate, suggesting that the solvent-exposed carotenoid in RCP becomes shielded in the aggregated protein [39].

Spectral changes that mimic the RCP can also be produced without proteolysis by acidifying the OCP in solution to approximately pH 3.5. This red spectral form of the protein is stable indefinitely at this pH. The spectral shift of the protein may be the result of altered charges in the microenvironment of the carotenoid. Alternatively, acidification may disrupt the numerous polar interactions between the all-helical bundle domain and the C-terminal NTF2-like domain in OCP, resulting in the increased exposure of the pigment to solvent and hence the spectral shift.

The higher plant photoprotective response provides a precedent for a dynamic, pH-mediated, spectral change to light stress. In plants, excess light leads to a large drop in thylakoid lumen pH, which triggers the two-step reversible conversion of violaxanthin into zeaxanthin. This xanthophyll cycle reaction results in a bathochromic shift of the pigment and the direct deexcitation of chlorophyll molecules, dissipating the energy as heat. The lowered pH also results in a protonation-induced conformational change in one or more of the minor light-harvesting complex proteins and their differential aggregation [2]. There is no evidence for a xanthophyll cycle in cyanobacteria. Photoprotective processes are not well characterized in cyanobacteria, but conversion of the OCP between spectral forms could provide a dynamic response in these organisms.

Since its discovery, the OCP has been suggested to have a role in photoprotective processes. The conversion of the OCP into the RCP could be one part of this dynamic process. The function of the shorter OCP paralogs detected in the *N. punctiforme* and *Nostoc* PCCC7120 genomes can be envisioned as important in this context as well. There are three paralogs in each of the *N. punctiforme* and *Nostoc* PCC7120 genomes that approximate the proteolytically formed RCP (Figure 2B). Furthermore, *N. punctiforme* and *Nostoc* PCC7120

each contain one paralog that corresponds to the C-terminal NTF2-like domain of OCP (residues 175–317). Although admittedly speculative, combining different N-terminal domain modules with the C-terminal domain and pigment would allow the organism to form an ensemble of OCP-like proteins, each with different spectral characteristics. Structural characterization of the *A. maxima* RCP is an important next step in testing these hypotheses.

Biological Implications

Carotenoids are virtually ubiquitous among microorganisms, animals, and plants, where they function to protect cells from reactive oxygen species that can cause photooxidative cell damage. In photosynthetic organisms, in addition to a role in light harvesting, they also function in nonphotochemical quenching, the thermal dissipation of chlorophyll excited states in excess of that needed to drive primary photochemistry. In any of these roles, the protein environment surrounding the pigment is crucial for proper function. Understanding the structural basis for how the protein environment regulates the spectral properties of carotenoids may have important applications, such as improving the agronomic and nutritional value of plants or for engineering antioxidative protection systems for animals, who obtain essential carotenoids through their diet.

Experimental Procedures

Protein Preparation and Crystallization

The purification procedure for the *A. maxima* OCP was as described [28] with several important modifications. Cells were harvested from Lake Texcoco near Mexico City. The purification was scaled up as described in Gomez-Lojero and Krogmann [40] by using a 10-fold increase in the amount of cells. Throughout the purification, 5 mM EDTA was present to inhibit the proteases that cleave the OCP to a red form. The preparative isoelectric focusing step previously described was omitted. A final purification on a Sephacryl S-200 column (1 m $\times$ 30 mm) with 0.1 M phosphate buffer (pH 7.0), 5 mM EDTA was used to remove some red carotenoid proteins generated in the preceding DEAE-52 column (likely the result of the ion exchange process separating the OCP from the EDTA).

Purified OCP, concentrated to 2.4 mg/ml in 80 mM Tris (pH 7.9), 2.5 mM EDTA, was used for crystallization. Crystals amenable to cryoprotection conditions were grown by vapor diffusion (4 μl of protein plus 2 μl of a reservoir containing 0.1 M Tris [pH 7.9], 0.2 M NaCl, 25% PEG 3350, 3% sucrose). The OCP crystals indexed in space group C2 (a = 217.2 Å, b = 40.8 Å, c = 75.5 Å, β = 95.8°) with two molecules in the asymmetric unit.

Potential heavy atom derivatives were screened for binding to OCP using the native gel shift approach of Boggon and Shapiro [41]. After identifying $HgCl_2$ as a promising candidate for derivatization, OCP crystals were soaked in 10 mM $HgCl_2$ for 2 hr. For cryoprotection, native and heavy atom crystals were immersed in Paratone-N oil (Paramins) until water no longer diffused out of the crystals (less than a minute). The crystals were then flash-cooled and stored in liquid nitrogen until data collection.

Structure Determination and Refinement

A standard three-wavelength anomalous diffraction data set was collected on the mercury derivative (L-III edge) at Brookhaven's National Synchrotron Light Source beamline X8C. Data were processed using HKL/SCALEPACK (Table 1) [42]. Six mercury sites were identified by SHELXD using anomalous differences from the peak wavelength of the mercury derivative [43]. Initial phases were calculated with MLPHARE and later improved by density modification and 2-fold symmetry averaging with DM [44]. The automatic chain tracing program, MAID, was able to trace about two thirds of the chain [45]. The remainder was built manually using the graphics program O [46]. The model was refined against the peak wavelength data set using conjugate gradient and simulated annealing algorithms as implemented by the program CNS [47]. Strong NCS restraints were used throughout. Due to the lack of isomorphism between the mercury derivative and native data sets, refinement against the native data could proceed only after repositioning the model by molecular replacement, a calculation readily performed using the program Replace [48]. This model was further refined by CNS, then REFMAC [49], in order to introduce TLS parameters in the refinement. The strength of the NCS restraints was reduced in the last round of refinement. Data collection and refinement statistics are given in Table 1.

Primary Structure Determination

A. maxima cells from Lake Texcoco were broken by three cycles of freezing and thawing in 0.09 M phosphate buffer (pH 7.8), 50 mM EDTA. DNA was extracted following a procedure described [50] with modifications. These included (1) adding RNase in the initial lysis buffer and (2) extending the time for protein extraction at 50°C from 2 hr to overnight incubation. The OCP gene was amplified by PCR using the DNA fragments (7–8 kb) obtained from NotI digestion and gel separation. Forward primers, 5′-ATGCCNTTYACNATHGAYWC-3′, were designed on the basis of the known N-terminal amino acid sequence of the OCP from *A. maxima*. Reverse primers, 5′-GCS ARSARRTCDATSGCSAC-3′, were designed from a comparison of the deduced amino acid sequences of the OCP genes from the genome databases of *Synechocystis* sp.PCC6803, *Nostoc* sp.PCC 7120 (Kazusa DNA Research Institute), and *Nostoc punctiforme* and *Synechococcus* WH8102 (DOE Joint Genome Institute). The alignment of the deduced amino acid sequences showed a region of high homology ending at the C terminus.

The OCP gene from *A. maxima* was first amplified using the following PCR program: 1 min denaturation at 95°C followed by 32 cycles for 30 s at 95°C, 30 s at 45°C for primer annealing, 90 s at 72°C for primer extension, and a final cycle at 72°C for 10 min. The amplified products were then subjected to PCR again using the following touchdown program: 1 min denaturation at 94°C, five cycles for 30 s at 94°C, 30 s at 45°C, 90 s at 72°C, 30 cycles for 30 s at 94°C, 30 s at 50°C, 90 s at 72°C, and a final cycle for 10 min at 72°C.

After analysis by agarose gel electrophoresis, the 800–900 bp fragments were cloned into the pGEM-T easy vector (Promega) and amplified by transformation in DH5α *E. coli* cells, and the clone containing the OCP gene was identified by DNA sequencing.

Singlet Oxygen Quenching Assays

Isolated 3-hydroxyechinenone and the OCP of *A. maxima* were assayed for quenching of singlet O_2 in D_2O by the method of Prat et al. [51] following dialysis into 10 mM phosphate buffer (pH 7.36; prepared in D_2O), 2 mM EDTA.

Acknowledgments

The authors thank Dr. Joan Valentine and Dr. Christopher Foote for helpful discussions. C.A.K. acknowledges the U.S. Department of Agriculture for support of this research (USDA1999-01759). M.R.S. and T.O.Y. acknowledge the support of the National Institutes of Health (NIH GM 31299).

Received: August 6, 2002
Revised: October 9, 2002
Accepted: October 10, 2002

References

1. Cogdell, R.J., and Frank, H.A. (1987). How carotenoids function in photosynthetic bacteria. Biochim. Biophys. Acta *895*, 63–79.
2. Yamammoto, H.Y., and Bassi, R. (1996). Carotenoids: localization and function. In Oxygenic Photosynthesis: The Light Reactions, D.R. Ort and C.F. Yocum, eds. (The Netherlands: Kluwer), pp. 539–563.

3. Ort, D.R. (2001). When there is too much light. Plant Physiol. *125*, 29–32.
4. Gotz, T., Windhovel, U., Boger, P., and Sandmann, G. (1999). Protection of photosynthesis against ultraviolet-B radiation by carotenoids in transformants of the cyanobacterium *Synechococcus* PCCC7942. Plant Physiol. *120*, 599–604.
5. Kuhlbrandt, W., Wang, D., and Fujiyoshi, Y. (1994). Atomic model of plant light-harvesting complex by electron crystallography. Nature *367*, 614–621.
6. Deisenhofer, J., Epp, O., Sinning, I., and Michel, H. (1995). Crystallographic refinement at 2.3 Å resolution and refined model of the photosynthetic reaction centre from *Rhodopseudomonas viridis*. J. Mol. Biol. *246*, 429–457.
7. Yeates, T.O., Komiya, H., Chirino, A., Rees, D.C., Allen, J.P., and Feher, G. (1988). Structure of the reaction center from *Rhodobacter sphaeroides* R-26 and 2.4.1: protein-cofactor (bacteriochlorophyll, bacteriopheophytin, and carotenoid) interactions. Proc. Natl. Acad. Sci. USA *85*, 7993–7997.
8. McDermott, G., Prince, S.M., Freer, A.A., Hawthornthwaite-Lawless, A.M., Papiz, M.Z., Cogdell, R.J., and Isaacs, N.W. (1995). Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria. Nature *374*, 517–521.
9. Koepke, J., Hu, X., Muenke, C., Schulten, K., and Michel, H. (1996). The crystal structure of the light-harvesting complex II (B800–850) from *Rhodospirillum molischianum*. Structure *4*, 581–597.
10. Jordan, P., Fromme, P., Witt, H.T., Klukas, O., Saenger, W., and Krauss, N. (2001). Three-dimensional structure of cyanobacterial Photosystem I at 2.5 Å resolution. Nature *411*, 909–917.
11. Hofmann, E., Wrench, P.M., Sharples, F.P., Hiller, R.G., Welte, W., and Diedrichs, K. (1996). Structural basis of light harvesting by carotenoids: peridinin-chlorophyll-protein from *Amphidinium carterae*. Science *272*, 1788–1791.
12. Gordon, E.J., Leonard, G.A., McSweeney, S., and Zagalsky, P.F. (2001). The C_1 subunit of α-Crustacyanin: the de novo phasing of the crystal structure of a 40 kDa homodimeric protein using the anomolous scattering from the S atoms combined with direct methods. Acta Crystallogr. *D57*, 1230–1237.
13. Holt, T.K., and Krogmann, D.W. (1981). A carotenoid protein from cyanobacteria. Biochim. Biophys. Acta *637*, 408–414.
14. Jurgens, U.J., and Weckesser, J. (1985). Carotenoid-containing outer membrane of *Synechocystis* sp. strain PCC6714. J. Bacteriol. *164*, 384–389.
15. Bullerjahn, G.S., and Sherman, L.A. (1986). Identification of a carotenoid-binding protein in the cytoplasmic membrane from the heterotrophic cyanobacterium *Synechocystis* sp. strain PCC6714. J. Bacteriol. *167*, 396–399.
16. Masamoto, K., Riethman, H.C., and Sherman, L.A. (1987). Isolation and characterization of a carotenoid-associated thylakoid protein from the cyanobacterium *Anacystis nidulans* R2. Plant Physiol. *84*, 633–639.
17. Diverse-Pierluissi, M., and Krogmann, D.W. (1988). A zeaxanthin protein from *Anacystis nidulans*. Biochim. Biophys. Acta *933*, 372–377.
18. Walsh, K., Jones, G.J., and Dunstan, R.H. (1997). Effect of irradiance on fatty acid, carotenoid, total protein composition and growth of *Microcystis aeruginosa*. Phytochemistry *44*, 817–824.
19. Li, X.-P., Bjorkman, O., Shih, C., Grossman, A.R., Rosenquist, M., Jannsson, S., and Niyogi, K. (2000). A pigment-binding protein essential for regulation of photosynthetic light harvesting. Nature *430*, 391–395.
20. Hirschberg, J., and Chamovitz, D. (1994). Carotenoids in cyanobacteria. In The Molecular Biology of Cyanobacteria, D.A. Bryant, ed. (The Netherlands: Kluwer Academic), pp. 559–579.
21. Mann, V., Harker, M., Pecker, I., and Hirschberg, J. (2000). Metabolic engineering of astaxanthin production in tobacco flowers. Nat. Biotechnol. *18*, 888–892.
22. Colovos, C., and Yeates, T.O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. *2*, 1511–1519.
23. Bullock, T.L., Clarkson, W.D., Kent, H.M., and Stewart, M. (1996). The 1.6 Å resolution crystal structure of nuclear transport factor 2 (NTF2). J. Mol. Biol. *260*, 422–431.
24. Stewart, M., Kent, H.M., and McCoy, A.J. (1998). Structural basis for molecular recognition between nuclear transport factor 2 (NTF2) and the GDP-bound form of the Ras-family GTPase Ran. J. Mol. Biol. *277*, 635–646.
25. Fribourg, S., Braun, I.C., Izaurralde, E., and Conti, I. (2001). Structural basis for the recognition of a nucleoporin FG repeat by the NTF2-like domain of the TAP/p15 mRNA nuclear export factor. Mol. Cell *8*, 645–656.
26. Cho, H.-S., Choi, G., Choi, K.Y., and Oh, B.-H. (1998). Crystal structure and enzyme mechanism of Δ^5-3-ketosteroid isomerase from *Pseudomonas testoroni*. Biochemistry *37*, 8325–8330.
27. Kauppi, B., Lee, K., Carredano, E., Parales, R.E., Gibson, D.T., Eklund, H., and Ramaswamy, S. (1998). Structure of an aromatic-ring-hydroxylating dioxygenase-naphthalene 1,2-dioxygenase. Structure *6*, 571–586.
28. Wu, Y.P., and Krogmann, D.W. (1997). The orange carotenoid protein of *Synechocystis* PCC 6803. Biochim. Biophys. Acta *1322*, 1–7.
29. Hess, W.R., Rocap, G., Ting, C.S., Larimer, F., Stilwagen, S., Lamerdin, J., and Chisholm, S.W. (2001). The photosynthetic apparatus of *Prochlorococcus*: insights through comparative genomics. Photosynth. Res. *70*, 53–71.
30. Kerfeld, C.A., Wu, Y.P., Chan, C., Krogmann, D.W., and Yeates, T.O. (1997). Crystals of the carotenoid-protein from *Arthrospira maxima* containing uniformly oriented pigment molecules. Acta Crystallogr. *D53*, 720–723.
31. Kakitani, T., Honig, B., and Crofts, A.R. (1982). Theoretical studies of the electrochromic response of carotenoids in photosynthetic membranes. Biophys. J. *39*, 57–63.
32. Davis, S.J., Vener, A.V., and Vierstra, R.D. (1999). Bacteriophytochromes: phytochrome-like receptors from nonphotosynthetic eubacteria. Science *286*, 2517–2520.
33. O'Neil, K.T., Erickson-Viitanen, S., and DeGrado, W.F. (1989). Photolabeling of calmodulin with basic, amphiphilic α-helical peptides containing p-benzoylphenylalanine. J. Biol. Chem. *264*, 14571–14578.
34. Liang, J., Edelsbrunner, H., and Woodward, C. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein Sci. *7*, 1884–1897.
35. Green, B.R., and Kuhlbrandt, W. (1995). Sequence conservation of light-harvesting and stress-response proteins in relation to the three-dimensional molecular structure of LHCII. Photosynth. Res. *44*, 139–148.
36. Heinemann, B., and Paulsen, H. (1999). Random mutations directed to transmembrane and loop domains of the light-harvesting chlorophyll a/b protein: impact on pigment binding. Biochemistry *38*, 14088–14093.
37. Grabowski, B., Cunningham, F.X., and Gantt, E. (2000). Chlorophyll and carotenoid binding in a simple red algal light-harvesting complex crosses phylogenetic lines. Proc. Natl. Acad. Sci. USA *98*, 2911–2916.
38. Matheson, I.B.C., Etheridge, R.D., Kratowich, N.R., and Lee, J. (1975). The quenching of singlet oxygen by amino acids and proteins. Photochem. Photobiol. *21*, 165–171.
39. Knutson, R.K. (1998). The red carotenoid protein from *Arthrospira maxima*. MS thesis, Purdue University, West Lafayette, Indiana.
40. Gomez-Lojero, C., and Krogmann, D.W. (1996). Large-scale preparation of photosynthetic catalysts from cyanobacteria. Photosynth. Res. *47*, 293–299.
41. Boggon, T.J., and Shapiro, L. (2000). Screening for phasing atoms in protein crystallography. Structure *8*, R143–R149.
42. Otwinowsk, Z., and Minor, W. (1997). Processing of the X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276A*, 307–326.
43. Sheldrick, G.M., and Schneider, T.R. (2001). Direct methods for macromolecules. In Methods in Macromolecular Crystallography, D. Turk and L. Johnson, eds. (Amsterdam: IOS Press), pp. 72–81.
44. CCP4 (Collaborative Computational Project 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. *D50*, 760–763.
45. Levitt, D.G. (2001). A new software routine that automates the

fitting of protein X-ray crystallographic electron-density maps. Acta Crystallogr. *D57*, 1013–1019.
46. Jones, T.A., Zou, J.-Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building proteins in electron density maps and the location of errors in these models. Acta Crystallogr. *A47*, 110–119.
47. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. *D54*, 905–921.
48. Tong, L., and Rossmann, M.G. (1997). Rotation function calculations with GLRF program. Methods Enzymol. *276*, 594–611.
49. Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. *D53*, 240–255.
50. Porter, R.D. (1988). DNA transformation. Methods Enzymol. *167*, 703–712.
51. Prat, F., Hou, C.C., and Foote, C.S. (1997). Determination of the quenching rate constants of singlet oxygen by derivatized nucleosides in nonaqueous solution. J. Am. Chem. Soc. *119*, 5051–5052.
52. DeLano, W.L. The PyMOL molecular graphics system (2002). www.pymol.org.
53. Nicholls, A., Bharadwaj, R., and Honig, B. (1993). GRASP: a graphical representation and analysis of surface properties. Biophys. J. *64*, A166.

Accession Numbers

The coordinates for the OCP have been deposited in the Protein Data Bank under ID code 1M98.