Supporting Information for

Distinctive and functional pigment arrangements in Lhcp, a prasinophyte-specific light-harvesting complex.

Soichiro Seki[1,2,*], Masato Kubota[3], Nami Yamano[4,5], Eunchul Kim[3,†], Asako Ishii[3], Tomoko Miyata[6,7], Hideaki Tanaka[2], Richard J. Cogdell[8], Jing-Ping Zhang[5], Keiichi Namba[7,9], Genji Kurisu[2,7,9], Jun Minagawa[3], and Ritsuko Fujii[1,10,11,*]

***Corresponding authors.** Soichiro Seki and Ritsuko Fujii.

Email: s-seki@protein.osaka-u.ac.jp (SS), and ritsuko@omu.ac.jp (RF)

This PDF file includes:

Supporting information Text 1 to 3
Figures S1 to S10
Tables S1 to S3
SI References

Supporting Information Text 1. Structural identification of unidentified carotenoid.

The unidentified peak was collected from Lhcp purified with sucrose density gradient centrifugation. Pigments were extracted from lyophilized Lhcp, and purified using high-performance liquid chromatography (column, LiChrosorb Si-60, 10 ID x 250 mm; eluent, 15% (v/v) acetone in *n*-hexane) as described in Seki et al. 2022 [1]. The UV-Vis absorption and CD spectra in *n*-hexane at room temperature were recorded by U-1800 (Shimadzu, Kyoto, Japan) and J-720W (JASCO Corporation, Tokyo, Japan), respectively. ESI-MS was recorded with positive ion mode using AccuTOF LC-plus JMS-T100LP (JEOL, Tokyo, Japan).

ESI-MS represents all the predicted signals listed in the following Table, including Na-adduct ions, confirming the chemical formula of $C_{48}H_{68}O_4$. (See Fig. S4c)

Table. ESI-MS signals

m/z (Obs.)	m/z (Calc.)	Δ [mDa]	Δ [ppm]	Ion
708.51194	708.51176	0.18	0.25	$[M]^{\bullet}$
709.51765	709.51512	2.53	3.57	$[MH]^{+}$
710.52184	710.51847	3.37	4.74	$[MH_2]^{2+}$
731.50170	731.50153	0.17	0.23	$[MNa]^{\bullet}$
732.50428	732.50488	-0.60	-0.82	$[MHNa]^{+}$
733.50660	733.50824	-1.64	-2.24	$[MH_2Na]^{2+}$

Cryo-EM map shows the presence of a 3-acyloxy-5,6-epoxy ionone end group, allowing us to propose a structure of the carotenoid body as antheraxanthin B [(3*S*, 5*S*, 6*R*, 3'*R*)-5,6-epoxy-5,6-dihydro-β-β-carotene-3,3'-diol] [2]. The UV-VIS absorption (421.0, 444.6, and 473.7 nm with %III/II 71% in *n*-hexane) and circular dichroism (CD) profile (211 (+12), 231 (-16), 273 (+23), and 335 (-9) nm) were in good agreement with reported values [2–4]. The cryo-EM map also showed that the acyl group contains at least eight carbons. To fit the MS signal, we tentatively proposed a structure with a double bond at C2"=C3" position based on observed short-chain acyl ester substitutions of carotenoids seen in marine green algae [5]. Overall, the absolute chemical structure of Uid was determined to be (3*S*,5*S*,6*R*,3'*R*)-5,6-epoxy-5,6-dihydro-3'-hydroxy-3-(2-octenoyloxy)-β-β-carotene, and the alternative name for this compound is 3-(2-octenoyl)-antheraxanthin B. The cryo-EM map corroborates the *cis* configuration at C8'-C9' single bond, i.e., 8'-s-*cis* configuration.

(3*S*,5*S*,6*R*,3'*R*)-5,6-epoxy-5,6-dihydro-3'-hydroxy-3-(2-octenoyloxy)-β-β-carotene.
Available amount: ~0.005 mg. UV-VIS in nm: (*n*-hexane) 421, 445, 474, and %III/II = 71. CD in nm (Δε in mdeg): (*n*-hexane) 211 (+12), 231 (-16), 273 (+23), and 335 (-9). ESI-MS, *m/z* = 708.51194 ± 0.00018 [M^+], $C_{48}H_{68}O_4$, *m/z* Calc. = 708.51176.

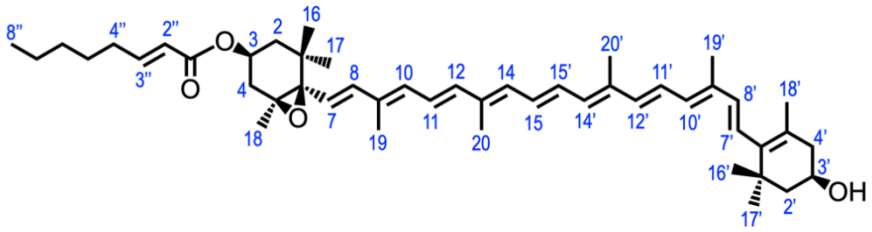


Figure. Chemical structure of 8'-*s*-*cis* form of 3-(2-octenoyl)-antheraxanthin B

Supporting Information Text 2: Excitonic coupling calculations and the constant for Dvp.

To estimate the excitonic coupling between chlorophyll molecules within and between monomers of the trimeric complex based on Amerongen and Grondelle 2001 [6], we implemented a custom Python script. The calculations account for the Mg-Mg distances as well as transition dipole moment vectors between chlorophylls, including Chl *a* (CLA), Chl *b* (CHL), and Dvp. For each pair of chlorophyll molecules, excitonic coupling values were calculated using the formula:

$$V = \frac{\mathrm{C} \cdot \kappa}{R^3}$$

where:

V is the excitonic coupling constant in cm^{-1}

κ is the orientation factor between the transition dipole moments of the two molecules

R is the Mg-Mg distance between the two molecules in nanometers.

C is a constant that equals $(5.04\ f_1^2\ \mu^2)/\varepsilon_r$ in Amerongen and Grondelle 2001 [6]. The values depend on the types of chlorophylls involved (90 $cm^{-1}\ nm^3$ for CLA-CLA, 75 $cm^{-1}\ nm^3$ for CLA-CHL, and 63 $cm^{-1}\ nm^3$ for CHL-CHL). This constant correlates with the dipole strength [6]. Thus, we calculated the ratio of the molar extinction coefficients based on the absorption spectral intensities between Chl *a* and Dvp [7] and determined the constant of CLA-Dvp as 64 $cm^{-1}\ nm^3$, that of CHL-Dvp as 54 $cm^{-1}\ nm^3$.

For each chlorophyll pair, the transition dipole moments were derived from the atomic coordinates of the nitrogen atoms in the porphyrin ring (ND and NB for CLA and CHL; N9 and N21 for DVP). Calculations were performed for all intra-monomer and inter-monomer pairs, averaging the results over the three monomer-monomer interfaces in the trimer.

Supporting Information Text 3. Parameters to calculate the electronic coupling based on dipole-dipole approximation.

Values of angles (θ_{AB}, θ_A, and θ_B) in degrees and a distance (R_{AB}) in Å between two dipoles (μ_A and μ_B) were measured using ChimeraX software based on the definitions given in the right figure.

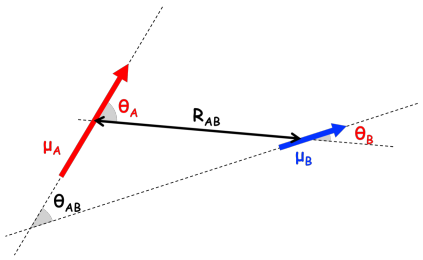


Table. Angles from 3D structure of Lhcp and calculated orientation factor and coupling.

dipole A	dipole B	θ_{AB}	θ_A	θ_B	R_{AB}	κ^{*1}	$V_{DD}{}^{*2}$
Uri: C9→C9'	*b*603: $N_B \rightarrow N_D$	163	91	104	6.5	-0.97	-229
	*a*602: $N_B \rightarrow N_D$	133	35	46	10.1	-2.40	-181
	*b*607: $N_B \rightarrow N_D$	44	148	20	15.1	3.11	58
Pra-1: C7'→C9	*a*612: $N_B \rightarrow N_D$	105	73	32	5.5	-0.99	475
	*a*610: $N_B \rightarrow N_D$	42	148	39	8.9	2.71	298
Mic: C8→C9'	*b*614: $N_B \rightarrow N_D$	116	114	154	8.2	-1.53	-175
	*a*611: $N_B \rightarrow N_D$	124	34	77	15.3	-1.13	-25
Pra-2: C9'→C13	*a*608: $N_B \rightarrow N_D$	62	21	98	14.3	0.85	128

*1 κ was calculated according to the equation, $\kappa = \cos\theta_{AB} - 3\cos\theta_A\cos\theta_B$.

*2 V_{DD} in cm^{-1} was calculated according to the equation, $V_{DD} = \frac{|\mu_A||\mu_B|\kappa}{4\pi\varepsilon_0 R_{AB}^3}$, using the following transition dipole strengths ($|\mu|$): 4.6 D and 3.8 D for the Q_y transitions of Chl *a* and Chl *b* (based on the experimentally determined value in vacuum [8]), and 3.0 D for the S_1-S_0 transition of the carbonyl carotenoids (based on the reported value for the S_1-S_0 transition of peridinin [9]).

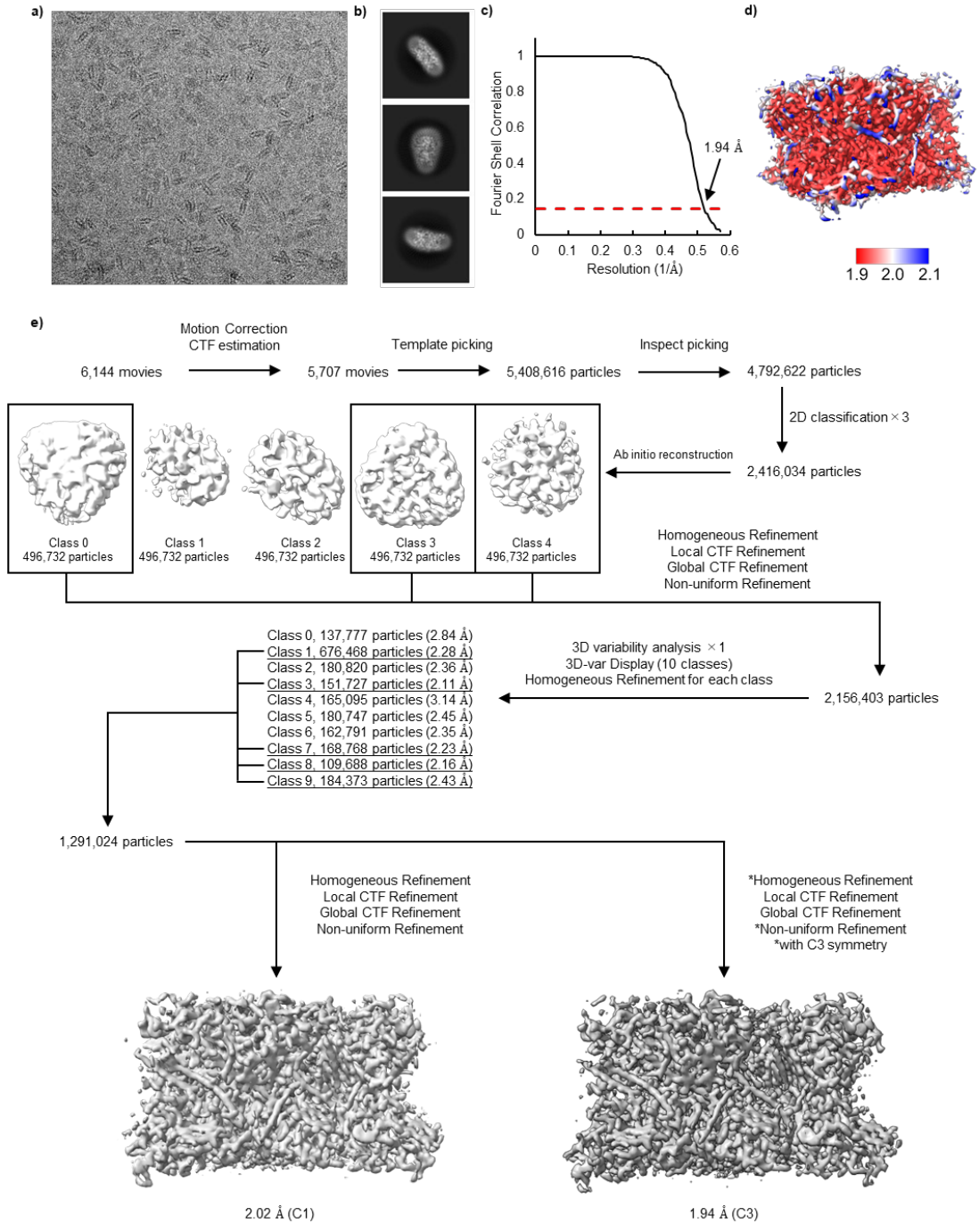
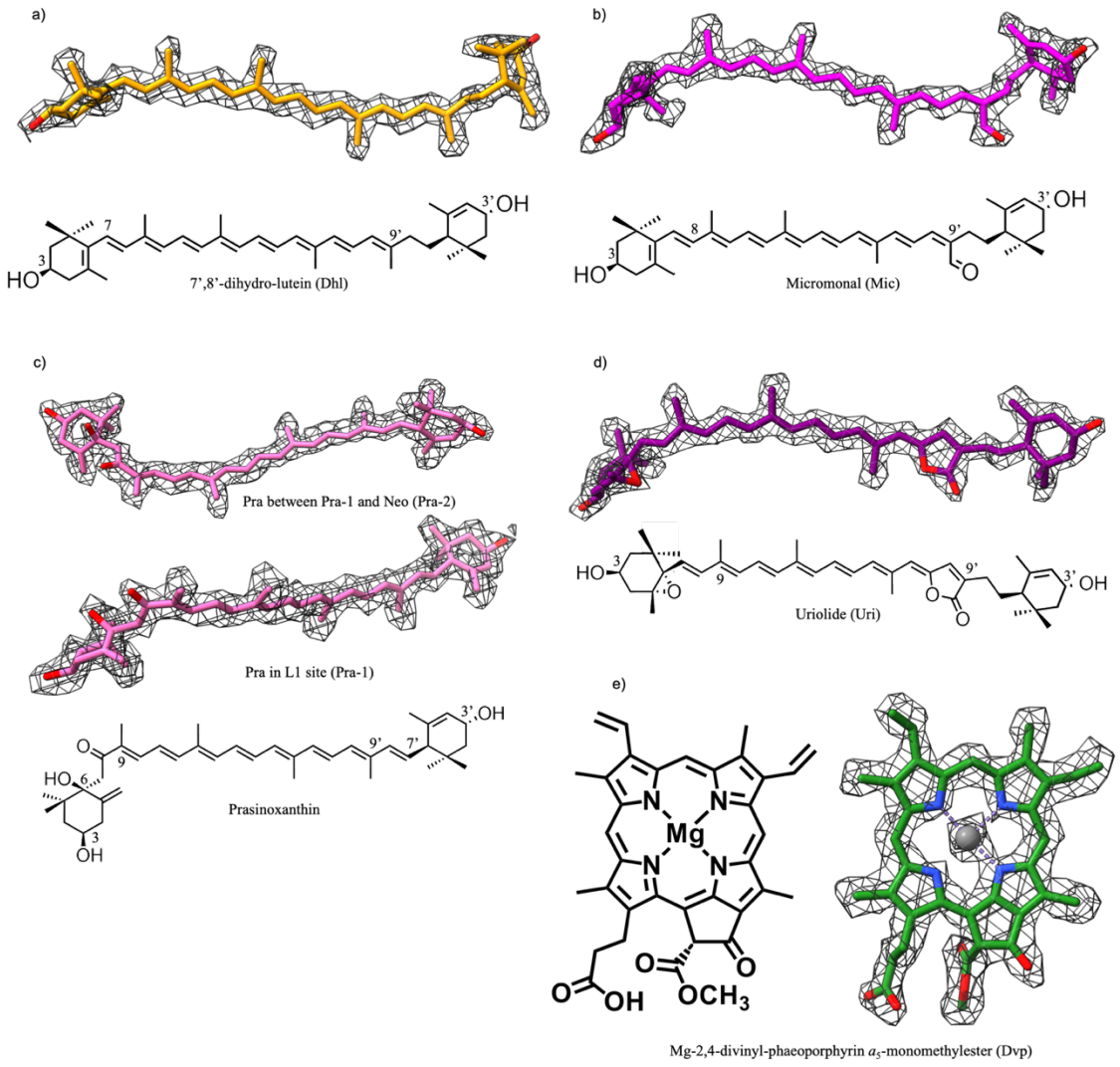


Fig. S1. Cryo-EM data processing procedure. Cryo-EM micrographs (a) and representative merged images (b) of Lhcp trimer were extracted. The resolution of the cryo-EM map was determined using an FSC value equal to 0.143 (c). Local resolution of the electron potential map with imposing C_3 symmetry (d). A detailed procedure for the construction of 3D data (e).



f)

	Chl *b*	Dvp	Neo	Pra	Uri	Mic	Dhl	Uid (Oax)
Pigment ratio (Chl *a* = 7)	6.28 ±0.05	0.75 ±0.01	1.31 ±0.02	2.16 ±0.06	1.02 ±0.05	1.15 ±0.04	1.24 ±0.02	1.00 ± 0.04

Fig. S2. Chemical structure and 3D-structure with electron potential map for unique pigments in Lhcp and the pigment composition in Lhcp. The structures of 7',8'-dihydrolutein (Dhl; a), micromonal (Mic; b), two prasinoxanthin molecules in different binding sites (Pra-1 and Pra-2; c), uriolide (Uri; d), and Mg-2,4-divinyl-phaeoporphyrin a_5-monomethyl ester (Dvp; e) were shown. Pigment composition of Lhcp was shown as the relative number of molecules per 7 Chl *a* molecules, determined using ultra-performance liquid chromatography (f). Each mean value of three biological replicates was shown with standard errors. The structural models with the density were created using ChimeraX[10].

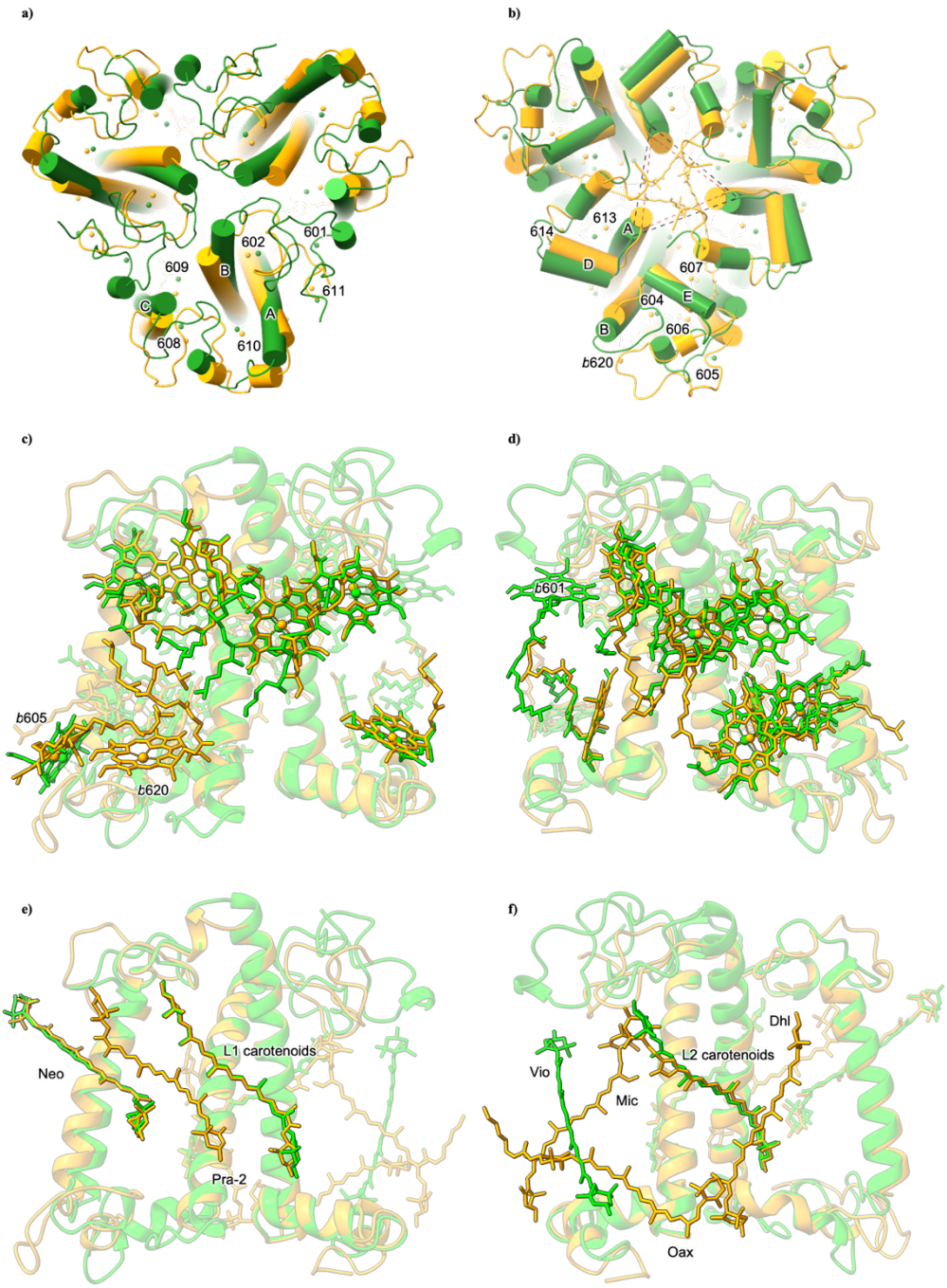


Fig. S3. Superposed structure between spinach LHCII (1RWT, green) and Lhcp (orange) with highlighting helices (a, b), chlorophylls (c, d), and carotenoids (e, f). The views of trimer from stromal (a) and lumenal (c) planes, and views of a monomer from the peripheral side (c, e) and the internal side (d, f) were illustrated. Mg atoms of Chl molecules and Oax are displayed for clarity (a). Dotted line in (b) connects Gly-202 of LHCII (black) or Gln-203 of Lhcp (red) between adjacent monomers, representing an inward shift of helices A by approximately 3 Å. The structural models were created using ChimeraX[10].

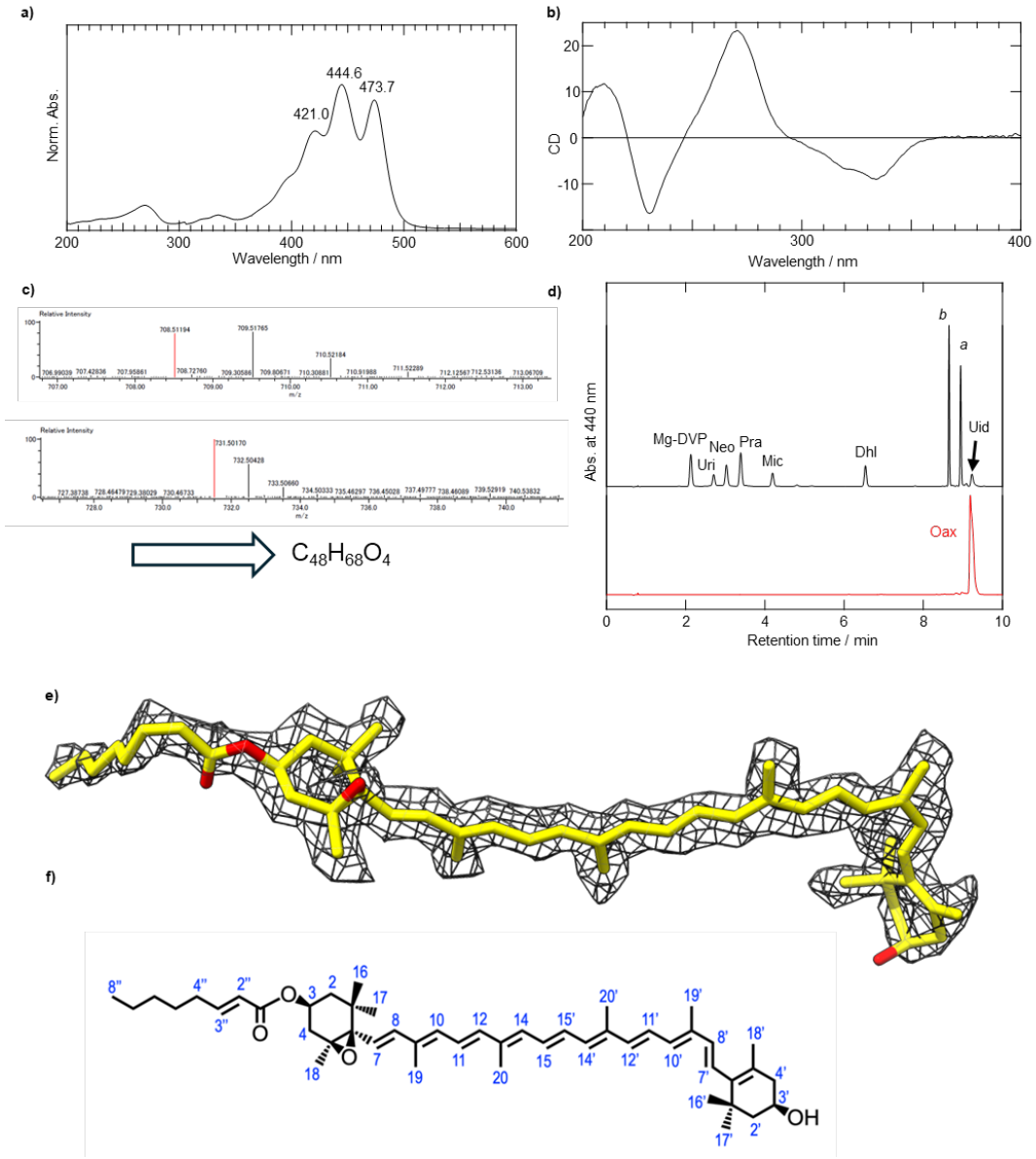


Fig. S4. Structural determination of a novel 3-octenoyloxy-8-s-*cis*-antheraxanthin B in Lhcp. Absorption (a) and circular dichroism (CD; b) spectra of isolated Oax were measured in hexane. The high-resolution mass spectrometry (c) revealed three signals without and with Na^+ ion, indicating the chemical formula of $C_{48}H_{68}O_4$. The isolated Oax was analyzed by HPLC (d) and compared with the extract from Lhcp complex. Based on these results, the absolute structure was determined to be 3-octenoyl antheraxanthin B, although the position of the double bond in the octenoyl group is tentative (Supporting Information Text 1). The unusual 8'-s-*cis* configuration was determined based on the electron potential map.

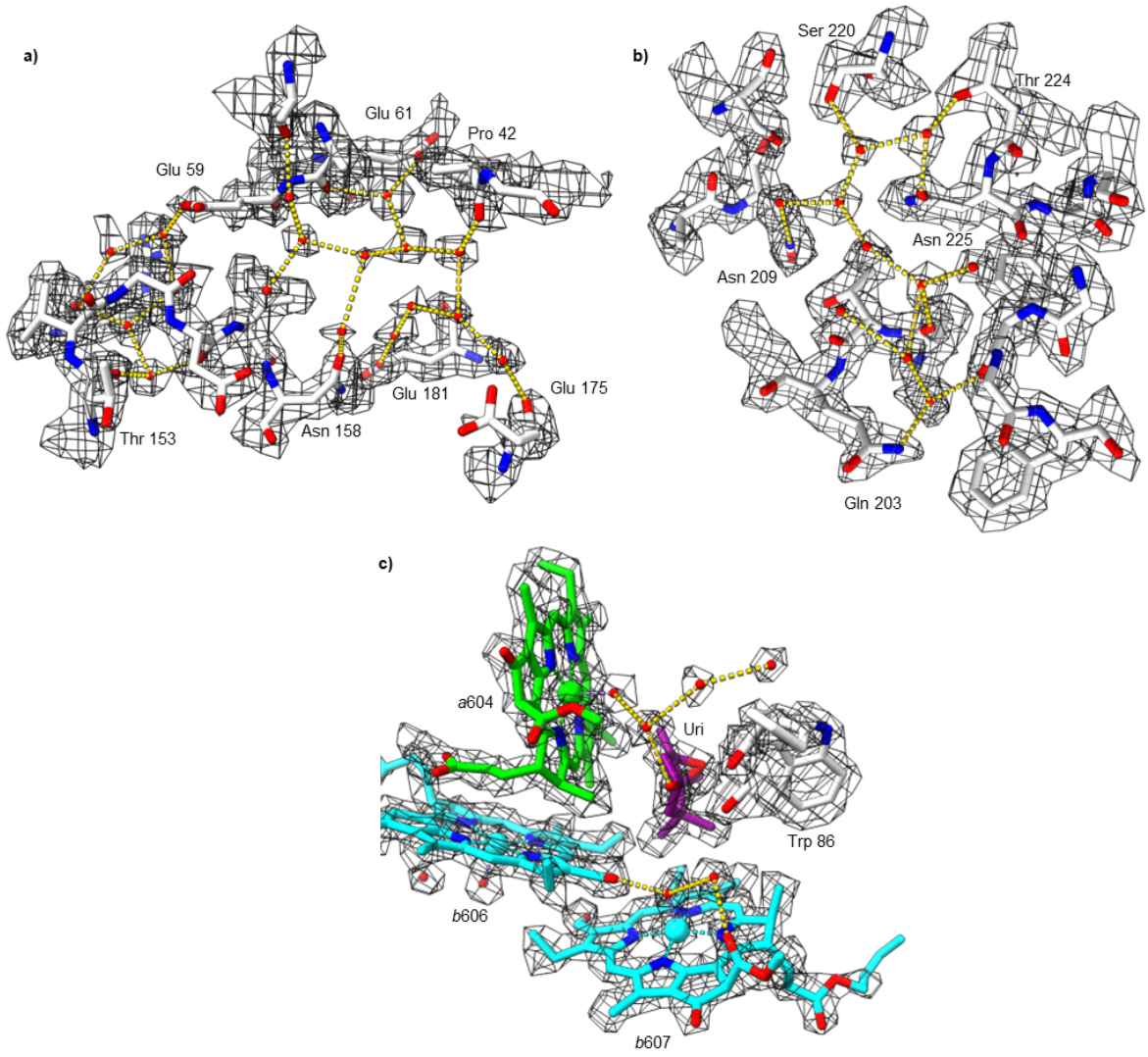


Fig. S5. The hydrogen bonded water clusters on the stromal (a) and lumenal (b and c) sides of Lhcp trimer.

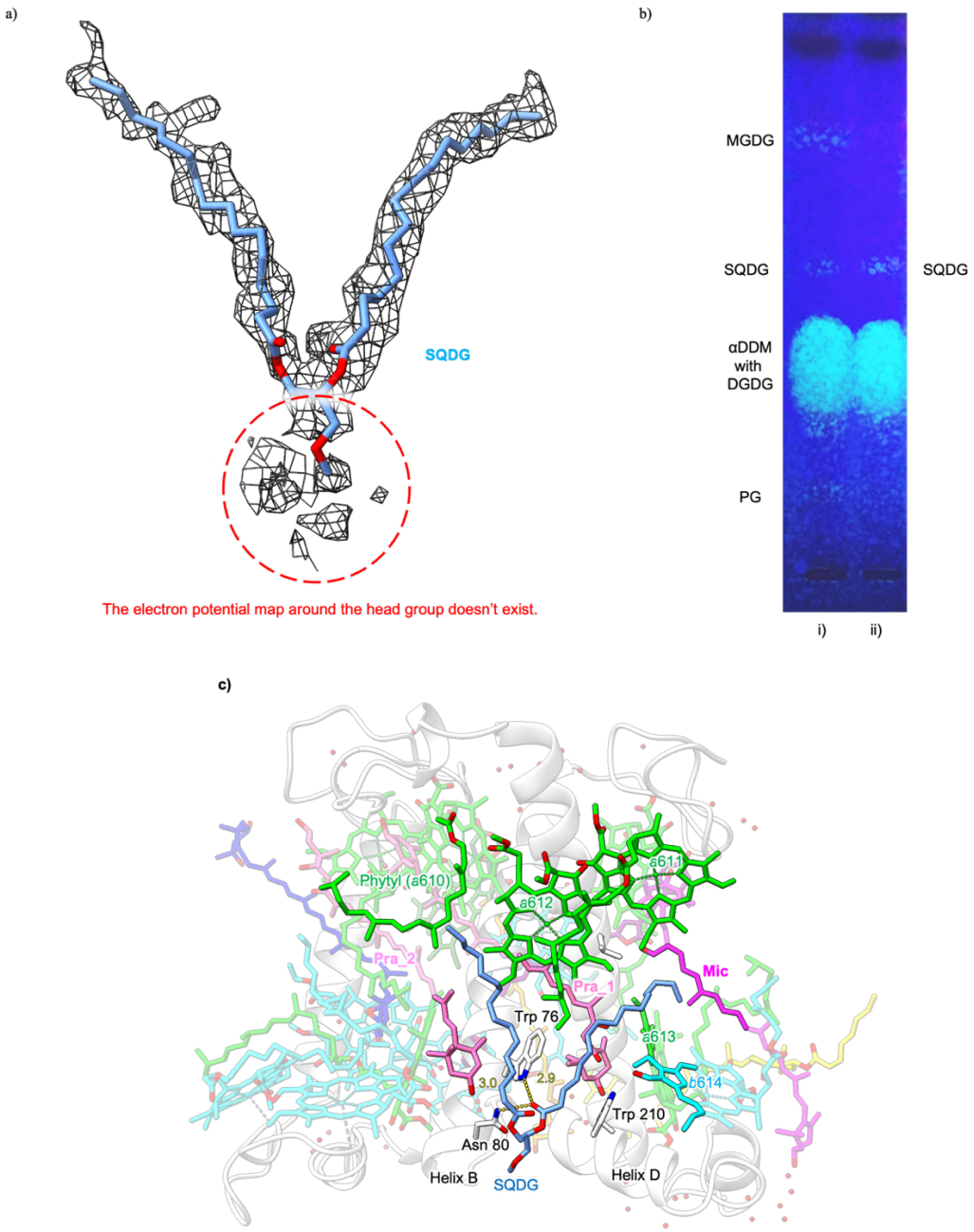


Fig. S6. Determination of SQDG molecule in Lhcp by electron potential map (a) and TLC (b), and stabilization of SQDG in Lhcp (c). Lipid compositions of Lhcp were analyzed using thin-layer chromatography (TLC) to compare the Lhcp trimer isolated using sucrose density gradient (i) with Lhcp trimer subsequently purified using anion exchange chromatography. A clear SQDG band was identified, and at the same time, the MGDG and PG bands disappeared. Two hydrogen bonds with residues Asn-80 and Trp-76 stabilize the acyl carbonyl of SQDG, and hydrophobic environments stabilize the hydrocarbon tails (c). MGDG, Monogalactosyl diacylglycerol; DGDG, Digalactosyl-diacylglycerol; SQDG, sulfoquinovosyl diacylglycerol; PG, phosphatidyl glycerol.

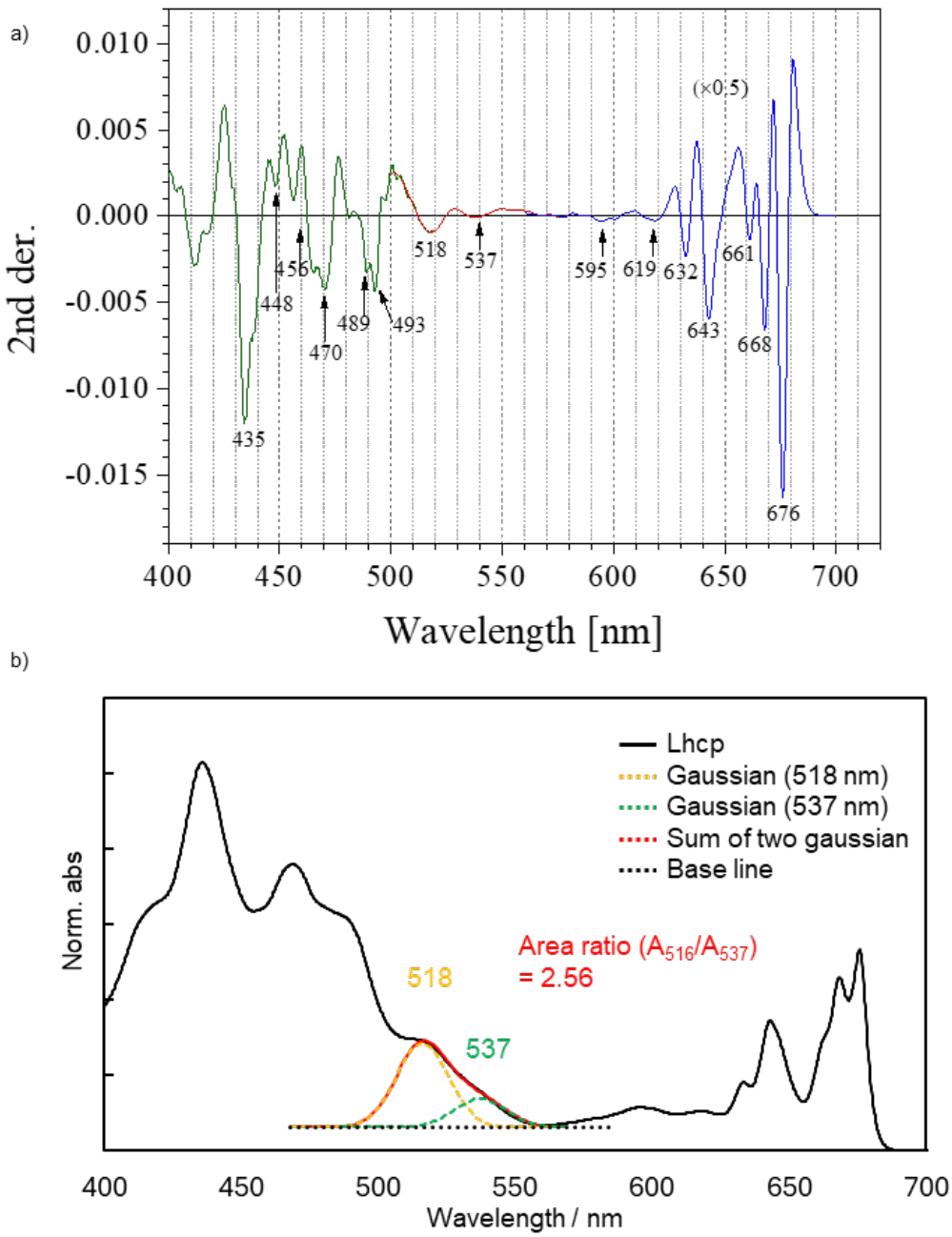


Fig. S7. Second derivative spectrum of the 77 K absorption spectrum (a) and simulation of the green light absorption cross section (b). The 77 K absorption spectrum was divided into three regions, indicated by green, red, and blue, and the second derivative spectrum of each region was calculated using Origin Pro 2023 software. The region indicated by the blue line was multiplied by 0.5. Simulation of the absorption cross section of the green region was performed manually using two Gaussian functions with peaks at 518 and 537 nm.

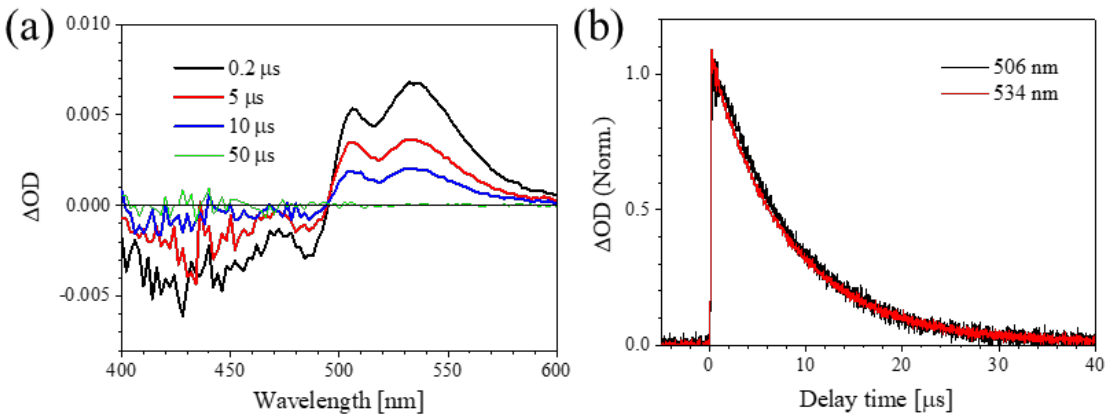


Fig. S8. Nano-second time-resolved absorption analysis of Lhcp trimer. Transient absorption spectra at selected delay time (a) and decay kinetics at 506 nm and 534 nm (b). Lhcp was excited at 672 nm, corresponding to the Q_y band of Chl *a*.

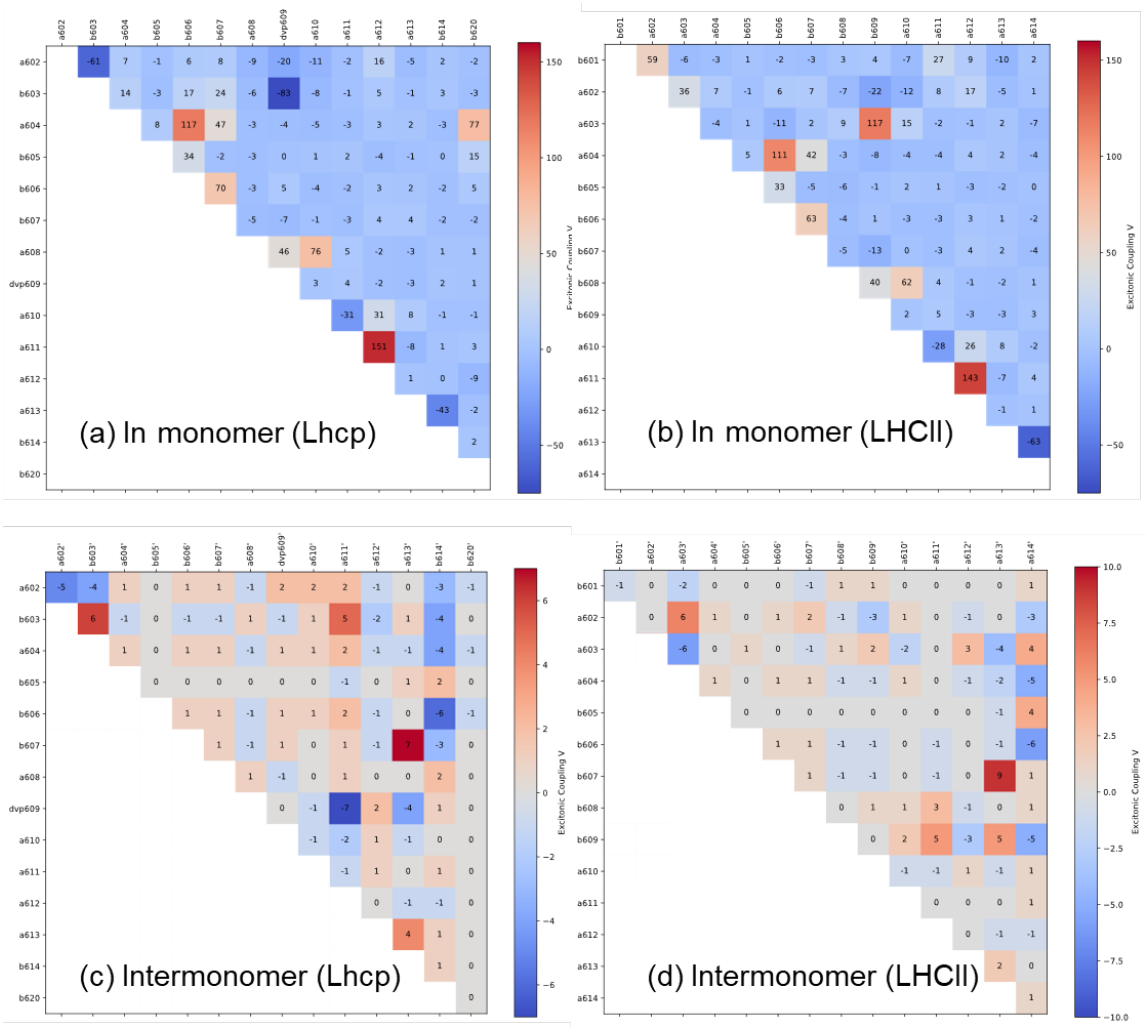


Fig. S9. Excitonic coupling lists of Lhcp and LHCII calculated using the equation described by Amerongen and Grondelle [6]. See Supplementary information 2 for the details.

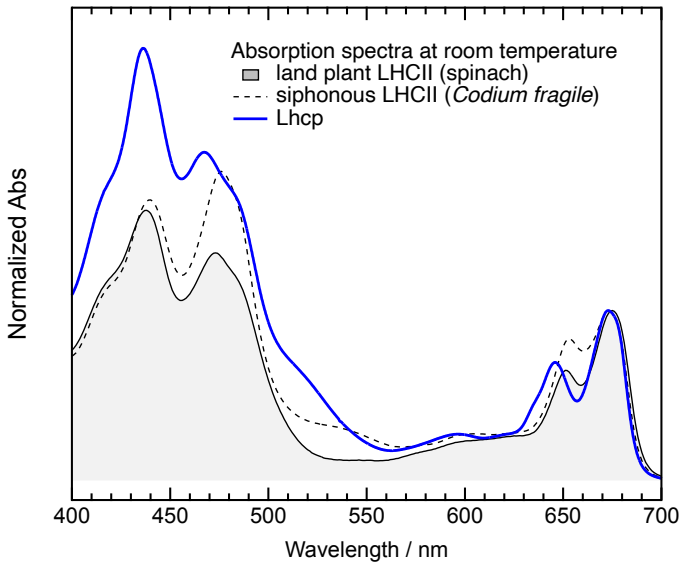


Fig. S10. Comparison of absorption spectra. Land plant LHCII (*Spinacia oleracea*, filled with pale gray, adapted from Seki et al. 2024 [11] with some modifications), siphonous LHCII (*Codium fragile*, black broken line, adapted from Seki et al. 2022 [12] with some modifications), and Lhcp (blue thick solid line) normalized at the Q_y peak maxima were shown.

Table S1. Cryo-EM data collection, refinement and validation statistics.

	Lhcp
PDB	9UC6
EMDB	EMD-64036
Magnification	60,000
Voltage (kV)	300
Electron exposure (e–/Å^2)	80
Defocus range (nm)	700-2,200
Pixel size (Å)	0.859
Symmetry imposed	C3
Initial particle images (no.)	5,408,616
Final particle images (no.)	1,291,024
Map resolution (Å)	1.94
FSC threshold	0.143
Refinement	
Initial model used (PDB code)	8HG5
Model composition	
Non-hydrogen atoms	8379
Protein residues	600
Ligands	69
R.m.s. deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.698
Validation	
MolProbity score	1.24
Clashscore	4.67
Rotamer outlier (%)	0.22
Ramachandran plot	
Favored (%)	99.00
Allowed (%)	1.00
Disallowed (%)	0.00

Table S2. Sequential comparison among several Lhcb or LhcbM proteins, Lhcp1 and Lhcp2. The Lhcb protein from *Spinacia oleracea* (SO-LHCII, PDB 1rwt), LhcbM proteins from *Codium fragile* (CF-Lhcbm11, PDB 7wlm), from *Bryopsis corticulans* (Cry-Lhcb1-3, PDB 8hlv and 8hpd), and Lhcp proteins from *Ostreococcus tauri* (Lhcp1 and Lhcp2, PDB 8hg5, 8hg6) were aligned based on the 3D structure. The amino acid numberings for SO-LHCII and Lhcp, as well as pigment-binding sites are also indicated.

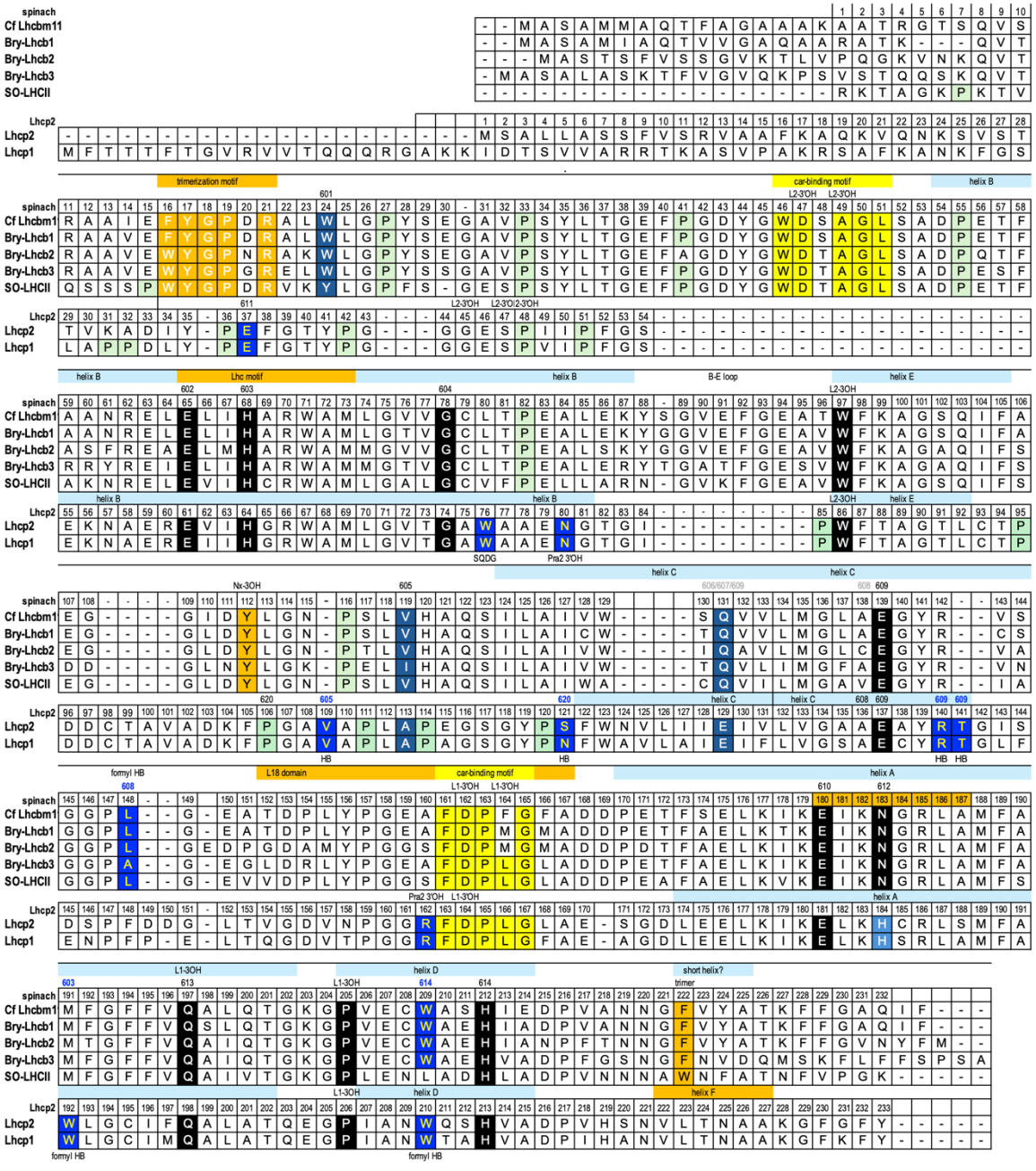


Table S3. Pigment composition differences between LHCII, previous Lhcp and this Lhcp with each axial ligand information. C=O indicates the carbonyl group from main chain.

	LHCII (1RWT)	Axial ligand	Lhcp (8HG4)	Lhcp (9UC6)	Axial ligand
601	*b*	C=O (Tyr 24)			
602	*a*	Glu 65	*a*	*a*	Glu 61
603	*a*	His 68	*a*	*b*	His 64
604	*a*	Water	*a*	*a*	Water
605	*b*	C=O (Val 119)	*b*	*b*	C=O (Leu 112)
606	*b*	Water	*b*	*b*	Water
607	*b*	Water	*b*	*b*	Water
608	*b*	Water	*b*	*a*	Water
609	*b*	Glu 139	Mg-DVP	Mg-DVP	Glu 137
610	*a*	Glu 180	*a*	*a*	Glu 181
611	*a*	PG	*a*	*a*	Glu 37
612	*a*	Asn 183	*a*	*a*	His 184
613	*a*	Gln 197	*a*	*a*	Gln 198
614	*a*	His 212	*a*	*b*	His 213
620			*b*	*b*	C=O (Pro 106)
L1	Lutein		Dihydrolutein	Prasinoxanthin	
L2	Lutein		Dihydrolutein	Uriolide	
N1	9'-*Cis* neoxanthin		9'-*Cis* neoxanthin	9'-*Cis* neoxanthin	
V1	Violaxanthin				
P2			Prasinoxanthin	Prasinoxanthin	
M1			Dihydrolutein	Micromonal	
D1			Dihydrolutein	Dihydrolutein	
O1			Prasinoxanthin	3-Octenoyl-8'-*s-cis* antheraxantin B	

SI References

1. Seki, S., Yamano, Y., Oka, N., Kamei, Y. & Fujii, R. Discovery of a novel siphonaxanthin biosynthetic precursor in Codium fragile that accumulates only by exposure to blue-green light. *FEBS Lett* **596**, 1544–1555 (2022).
2. Britton, G., Liaaen-Jensen, S., Pfander, H. Eds. *Carotenoids Handbook*, (Springer Basel AG, 2004).
3. Buchecker, R. & Noack, K. "Circular Dichroism" in *Carotenoids Vol. 1B: Spectroscopy*, Britton, G., Liaaen-Jensen, S. & Pfander, H. Eds., (Birkhäuser Verlag, 1995), Chapter 3, pp. 63–116.
4. Marki-Fischer, E., Buchecker, R. & Eugster, C. H. Absolute Configuration of Antheraxanthin, '*cis*-Aritheraxanthin' and of the Stereoisomeric Mutatdxanthins. *Helv. Chim. Acta* **65** (7): 2198-2211 (1982).
5. Yoshii, Y., Takaichi, S., Maoka, T., Handa, S. & Inoue, I. Chacterization of two unique carotenoid fatty acid esters from *Pterosperma cristatum* (prasinophyceae, chlorophyta). *J. Pjycol.* **38** (2), 297-303 (2002).
6. Van Amerongen, H. & Van Grondelle, R. Understanding the energy transfer function of LHCII, the major light-harvesting complex of green plants. *J. Phys. Chem. B* **105**, 604–617 (2001).
7. Jeffrey, S. W., Mantoura, R. F. C. & Bjornland, T. "Data for the identification of 47 key phytoplankton pigments" in *Phytoplankton Pigments in Oceanography: Guidelines to Modern Methods*, Jeffrey, S. W., Mantoura, R. F. C. & Wright, S. W. Eds., (UNESCO Publishing, 1997), pp. 488–489.
8. Knox, R. S. Dipole and oscillator strengths of chromophores in solution, *Photochem. Photobiol.* **77**: 492–496 (2003).
9. Krueger, B. P., Lampoura, S. S., van Stokkum, I. H.M., Papagiannakis, E., Salverda, J. M., Gradinaru, C. C., Rutkauskas, D.; Hiller, R. G., van Grondelle, R. "Energy transfer in the peridinin chlorophyll-*a* protein of *Amphidinium carterae* studied by polarized transient absorption and target analysis", *Biophys. J.* **80**:2843-2855 (2001).
10. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Daniel M Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
11. Seki, S., Miyata, T., Norioka, N., Tanaka, H., Kurisu, G., Namba, K., Fujii, R. Structure-based validation of recombinant light-harvesting complex II. *PNAS Nexus* **3**, 405 (2024).
12. Seki, S., Nakaniwa, T., Castro-Hartmann, P., Sader, K., Kawamoto, A., Tanaka, H., Qian, P., Kurisu, G., Fujii, R. Structural insights into blue-green light utilization by marine green algal light harvesting complex II at 2.78 Å. *BBA Advances* **2**, 100064 (2022).