Structure-based validation of recombinant light-harvesting complex II

Soichiro Seki [a,b,*], Tomoko Miyata[c,d], Naoko Norioka[b], Hideaki Tanaka[b], Genji Kurisu [b,d,e], Keiichi Namba [c,d] and Ritsuko Fujii [a,f,g,*]

aGraduate School of Science, Osaka City University, 3-3-138 Sugimoto, Sumiyoshi-ku, Osaka 558-8585, Japan
bInstitute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan
cGraduate School of Frontier Biosciences, Osaka University, Suita, Osaka 565-0871, Japan
dJEOL YOKOGUSHI Research Alliance Laboratories, Osaka University, 1-3 Yamadaoka, Suita, Osaka 565-0871, Japan
eInstitute for Open and Transdisciplinary Research Initiatives, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan
fGraduate School of Science, Osaka Metropolitan University, 3-3-138 Sugimoto, Sumiyoshi-ku, Osaka 558-8585, Japan
gResearch Center for Artificial Photosynthesis, Osaka Metropolitan University, 3-3-138 Sugimoto, Sumiyoshi-ku, Osaka 558-8585, Japan
*To whom correspondence should be addressed: Email: s-seki@protein.osaka-u.ac.jp (S.S.); Email: ritsuko@omu.ac.jp (R.F.)
Edited By Gerhard Hummer

Abstract

Light-harvesting complex II (LHCII) captures sunlight and dissipates excess energy to drive photosynthesis. To elucidate this mechanism, the individual optical properties of pigments in the LHCII protein must be identified. In vitro reconstitution with apoproteins synthesized by *Escherichia coli* and pigment–lipid mixtures from natural sources is an effective approach; however, the local environment surrounding each pigment within reconstituted LHCII (rLHCII) has only been indirectly estimated using spectroscopic and biochemical methods. Here, we used cryo-electron microscopy to determine the 3D structure of the rLHCII trimer and found that rLHCII exhibited a structure that was virtually identical to that of native LHCII, with a few exceptions: some C-terminal amino acids were not visible, likely due to aggregation of the His-tags; a carotenoid at the V1 site was not visible; and at site 614 showed mixed occupancy by both chlorophyll *a* and *b* molecules. Our observations confirmed the applicability of the in vitro reconstitution technique.

Keywords: light-harvesting complex II, recombinant, chlorophyll, carotenoid, photosynthesis

Significance Statement

In plants, the major light-harvesting complex (LHCII) plays a critical role in handling ever-changing sunlight. As LHCII is a multichromophore system, a complete theoretical understanding requires reference to experimental results. In vitro reconstitution is powerful in providing a range of conditions, including differences in pigments and proteins. For decades, spectroscopic studies of in vitro reconstituted LHCII (rLHCII) have been utilized for this purpose without structural proof. Here, we obtained a high-resolution 3D structure of rLHCII using cryo-electron microscopy with the slight protocol modification. We found several differences between rLHCII and its native counterpart, which explain the typical spectroscopic features of rLHCII. The structural validation builds the foundation for structure-based analyses using rLHCII and will open the door to improve research on photosynthesis.

Introduction

Light harvesting is the initial step in photosynthesis, which is a unique biological process that converts sunlight energy into chemical energy (1, 2). Light-harvesting complex II (LHCII) is the most abundant antenna complex in higher plants and green algae and contains the majority of all chlorophyll (Chl) on Earth. LHCII plays a dual role in capturing light and dissipating excess energy to provide an appropriate amount of energy to the photoreaction center (1, 2). LHCII is a trimer (3, 4); each monomer has 3 transmembrane helices and 2 amphipathic helices facing the thylakoid lumen and binds 14 Chl molecules, comprising 8 Chl

a and 6 Chl *b* molecules, 4 carotenoid molecules occupying distinct binding sites (5) (L1, L2, N1, and V1 sites), and 1 phosphatidyl glycerol (PG). Among the lipids that constitute the thylakoid membranes, PG is essential for trimerization (6–8), and monogalactosyldiacylglycerol, although not visible in the 3D structure, contributes to the stabilization of the trimer within the membrane (8, 9). The densely packed pigments in LHCII are allocated within the protein scaffold, allowing directed energy transfer while suppressing spontaneous quenching (1, 3, 10, 11). Spectroscopic techniques have revealed the excited-state dynamics of pigments in LHCII; however, attempting to attribute

the observed transition energies to individual pigments within LHCII is challenging because multiple pigments with the same chemical structure exist within a single protein, each surrounded by a different local environment (10, 11). Therefore, quantum chemical attempts require reference data, specifically, a series of spectroscopic results for LHCII with specific pigments replaced or removed (1, 10, 11). In vitro reconstitution is a powerful tool for this purpose.

In vitro reconstituted LHCII (rLHCII) is formed by self-assembly through the mixing of the apoprotein synthesized by *Escherichia coli* solubilized with a detergent and a pigment–lipid mixture from natural sources, after removing the detergent (12). This technique allows for His-tag affinity purification of rLHCII (13) and modification of the pigment composition (5, 13–15) and apoprotein (16–18). In this regard, in vitro reconstitution is widely used to discern the optical properties of each pigment in LHCII (13, 19). However, each pigment-binding environment in rLHCII has been indirectly evaluated without reference to the structure of rLHCII by combining biochemical assays, pigment composition analyses, and spectroscopic analyses that sensitively reflect the characteristics of the pigments' surroundings (12). Therefore, determining the exact structure of rLHCII is essential to verify whether it is identical to its natural counterpart.

Here, we aimed to capture the high-resolution 3D structures of both the rLHCII and native LHCII trimers (nLHCII) using cryo-electron microscopy (cryo-EM). The overall structure of rLHCII closely resembled that of nLHCII; however, several differences emerged that could explain the frequently observed spectroscopic and biochemical features of rLHCII. This indicates the applicability of the in vitro reconstitution technique.

Results

Strategy to prepare rLHCII for high-resolution structural determination

We targeted the LHCII trimer of *Spinacia oleracea*, which has been extensively studied, and the high-resolution structures were determined using X-ray crystallography (3) and cryo-EM (20). In vitro reconstitution of the C-terminal-tagged Lhcb1 protein (Fig. 1a-ii and b), and the pigment–lipid mixture extracted from purified LHCII (Fig. 1a-i) was performed according to the procedure described by Natali et al. (13) (see Materials and methods section). Trimerization could occur during mixing with the Ni column resin (Fig. 1a-iii). After eliminating other contaminants by affinity purification, the trimer was isolated by sucrose density gradient centrifugation (Fig. 1a-iv).

Spectroscopic characterization of rLHCII

rLHCII function was evaluated using several optical spectroscopy techniques. The absorption spectrum of rLHCII was almost identical to that of nLHCII, with minor differences (Fig. 2a and b), including a decrease in intensity ∼400–500 nm and a blue shift of the peak ∼600–700 nm. The overall circular dichroism (CD) pattern of rLHCII was very similar to that of nLHCII, with small differences in relative intensities (Fig. 2c), such as an increase in the 490 nm negative peak, a decrease in the 470 nm negative peak, a decrease in the 640 nm negative shoulder, and a decrease in the 680 nm negative peak. These differences are typical features of rLHCII (6, 18). The fluorescence spectra of the two samples were identical. The fluorescence excitation spectrum and corresponding absorption spectrum of rLHCII were nearly identical to those of nLHCII (Fig. 2d and e). This indicates that the efficiency of excitation energy transfer in rLHCII is comparable with that in nLHCII.

a *In vitro* reconstitution procedure
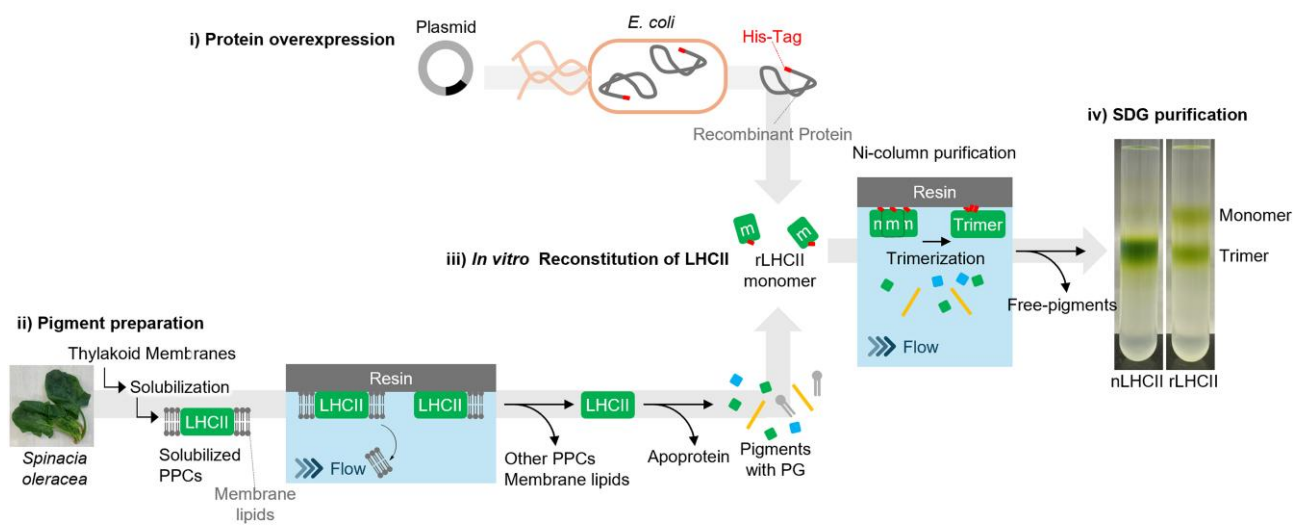
b Designed amino acid sequence

```
              14
MRKTAGKPKTVQSSSPWYGPDRVKYLGPFSGESPSYLTGEFPGDYGWDTAGLSADPETFAKNRELEVIHCRWAMLGALGCVFPELLARN
GVKFGEAVWFKAGSQIFSEGGLDYLGNPSLVHAQSILAIWACQVILMGAVEGYRIAGGPLGEVVDPLYPGGSFDPLGLADDPEAFAELK
VKEIKNGRLAMFSMFGFFVQAIVTGKGPLENLADHLADPVNNNAWNFATNFVPGKHHHHHH
                                              223              His-Tag
```

Fig. 1. Procedure of in vitro reconstitution. The designed amino acid sequence of Lhcb-1 from *S. oleracea* (a) and the flow chart of this method (b). LHCII, light-harvesting complex II; rLHCII, reconstituted light-harvesting complex II; nLHCII, native light-harvesting complex II; SDG, sucrose density gradient; PPCs, pigment–protein complexes.
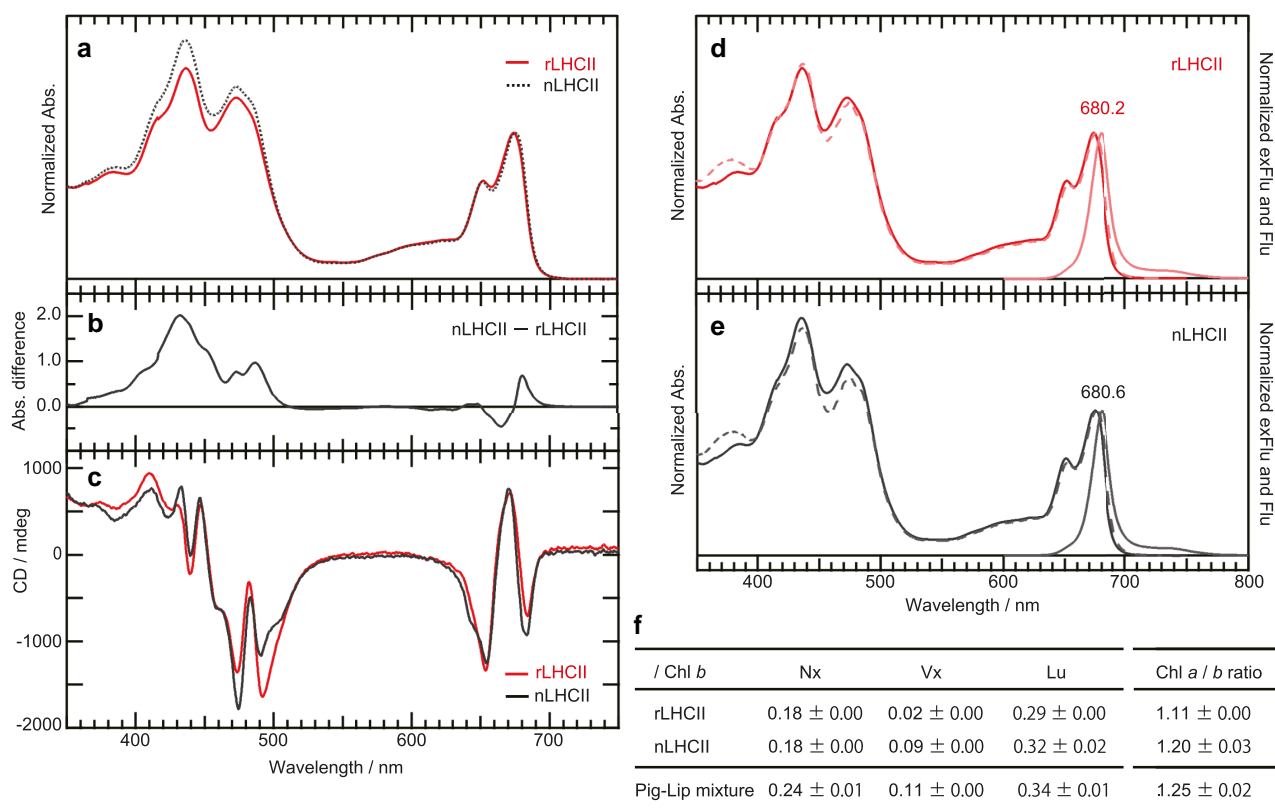
Fig. 2. Spectroscopic characterization of rLHCII. Absorption (a, b), CD (c), Abs, fluorescence, and fluorescence excitation (d, e) spectra, and the pigment compositions per Chl b determined by high-performance liquid chromatography (f). Errors are calculated for three biological replicates. Abs, absorption; CD, circular dichroism; Fluo, fluorescence; FluoEx, fluorescence excitation; Chl, chlorophyll; Nx, 9′-*cis*-neoxanthin; Lu, lutein; Vx, violaxanthin; Pig mixture, pigment–lipid mixture; rLHCII, reconstituted light-harvesting complex II; nLHCII, native light-harvesting complex II.

The pigment composition of rLHCII was similar to that of nLHCII (Fig. 2f). Compared with nLHCII, similar amounts of lutein and 9′-*cis*-neoxanthin were detected in rLHCII; however, the amount of violaxanthin significantly decreased. The Chl *a/b* ratio of rLHCII was 1.11, which was slightly lower than that of nLHCII, implying the presence of Chl *a* to Chl *b* substitution in rLHCII. These differences in pigment composition can explain the differences in the absorption spectra described above; that is, a decrease in 450–500 nm absorption is due to the shortage of violaxanthin, and a decrease ~430 nm and a blue shift ~680 nm may be due to Chl substitution. Therefore, these spectroscopic characterizations confirmed that the rLHCII in this study was a typical rLHCII (6, 12–19, 21–30), and that it has a slight difference in pigment composition when compared with nLHCII.

Comparison of rLHCII and nLHCII cryo-EM maps

The electrostatic potential maps of rLHCII were reconstructed at global resolutions of 2.57 Å without imposing symmetry (C_1) and at 2.37 Å with C_3 symmetry from 5,889 cryo-EM movies of 179,318 particles (Fig. S1). Similarly, those of nLHCII were reconstructed at global resolutions of 2.32 Å (C_1) and 2.17 Å (C_3) from 5,199 movies of 479,808 particles. The electrostatic potential map of rLHCII was nearly identical to that of nLHCII (Fig. 3a–c). In a close comparison (Fig. 3d and inset), the electrostatic potential map of rLHCII slightly lacked density in the C-terminal region when compared with that of nLHCII. Notably, on the electrostatic potential map without imposing symmetry, only rLHCII exhibited one additional density projecting toward the lumen, located asymmetrically with respect to

the C_3 center of symmetry (the red broken circles in Fig. 3h). This area corresponds to the location of the C-terminal residues. Considering that the only difference between the apoproteins of the two LHCII complexes was the additional His-tag in the C-terminal region, the three His-tagged C-termini of each monomer may have aggregated.

Comparison of the structural models between rLHCII and nLHCII

Structural models of both rLHCII and nLHCII were constructed using electrostatic potential maps with C_3 symmetry, based on the native LHCII model determined by X-ray crystallography (3), with the correct absolute configuration of lutein (see below). Note that in previous high-resolution structural models (3, 20), "lutein" was incorrectly registered to be the 3S stereoisomer of naturally occurring lutein (32, 33), (3R,3′R,6′R)-β, ε-carotene-3,3′-diol. The absolute structure of lutein was used for modeling.

The resultant rLHCII model was essentially identical to the nLHCII model (Figs. 4a–f and S2), with a few exceptions: (i) some residues in the C-terminal region were not visible (Fig. 4f), (ii) the carotenoid at the V1 site was not visible (Fig. 4d), and (iii) the Chl molecule at site 614 appeared to be a mixture of Chl *a* and Chl *b* molecules (Fig. 4g and h). In the rLHCII model, the amino acid residues of the Lhcb1 protein were successfully allocated from Ser 14 at the N-terminus to Asp 223; however, the remaining residues until the C-terminal Val 229 were not visible (Figs. 1b, 4f, and S2d). This is likely related to the His-tag introduced in rLHCII because the coordinates of amino acid residues up to Val 229 of nLHCII were consistent between the models determined by
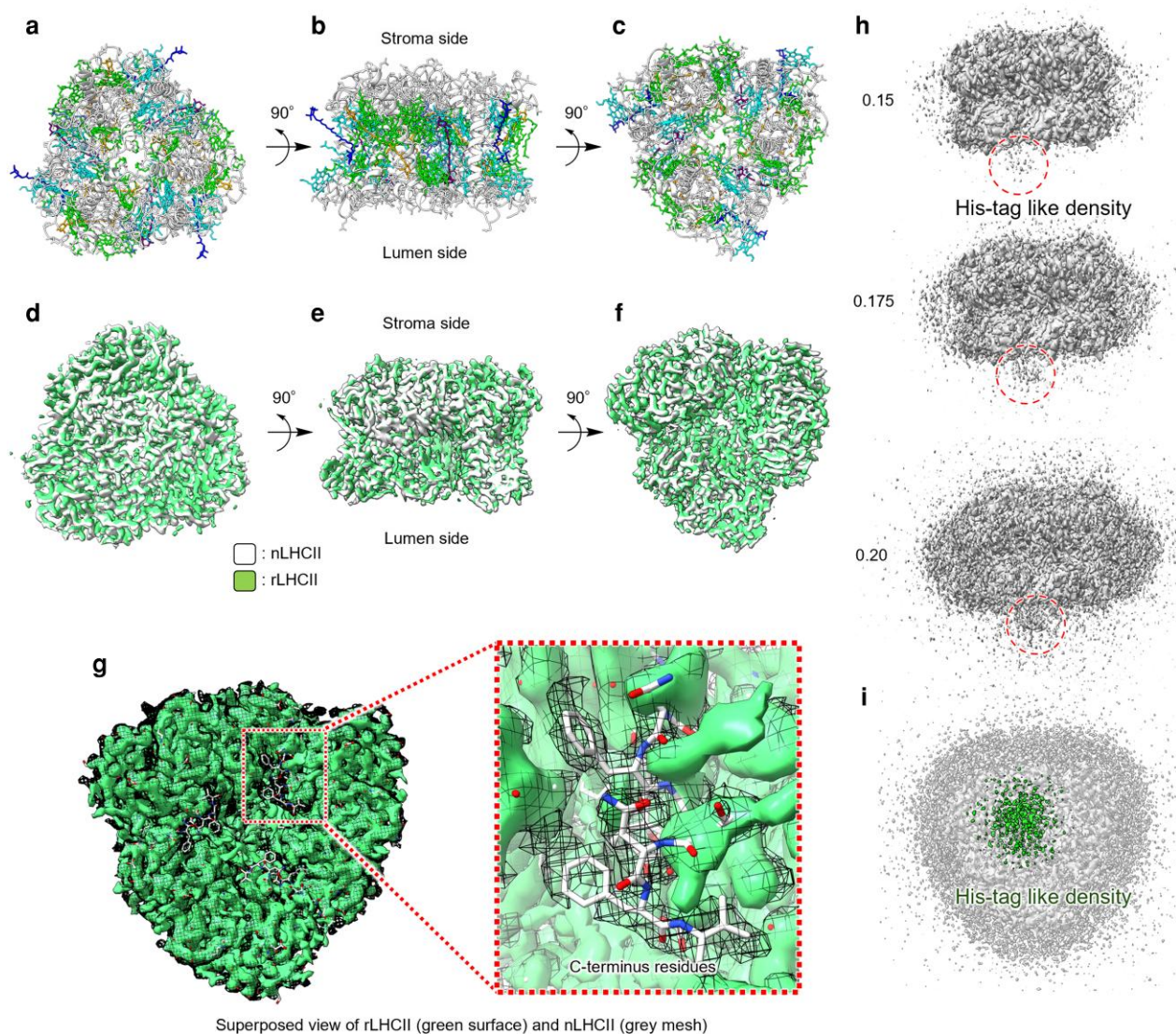
Fig. 3. Cryo-EM map comparison between rLHCII and nLHCII. Superimposed C_1 map of rLHCII (green) to nLHCII (white) (d–f) with coordinates of the same direction (ac). Superimposed C_1 map of rLHCII (green) to nLHCII (mesh) observed from the stromal side with an enlargement (g). Maps for rLHCII without imposing symmetry observed from the side (h). The top view from luminal side was illustrated as highlighted map (i). Cutoff values of ChimeraX are designated beside each map, and the asymmetric density is encircled with a dashed red line. The figures were drawn using ChimeraX (31). rLHCII, reconstituted light-harvesting complex II; nLHCII, native light-harvesting complex II.

cryo-EM (20) and X-ray crystallography (3). All the heteroatoms in rLHCII, such as pigments, PG, and water molecules, were well superimposed on those in nLHCII. The rLHCII model contained 3 carotenoids and 14 Chl molecules (Fig. 4a and b). The electrostatic potential map of rLHCII lacks the density of the V1 carotenoid, resulting in the assignment of only three carotenoids. The two lutein molecules in the L1 and L2 sites and one 9′-*cis*-neoxanthin molecule in the N1 site were confirmed. Careful inspection of the cryo-EM map for carotenoids at both the L1 and L2 sites in rLHCII excluded the possibility of the existence of epoxy carotenoids, such as violaxanthin, at those sites (Fig. S3). Owing to the high-resolution cryo-EM map, seven Chl *a* molecules at sites 602, 603, 604, 610, 611, 612, and 613, and four Chl *b* molecules at sites 605, 606, 607, and 608 were nearly identical between rLHCII and nLHCII (Fig. S4). The remaining two Chl *b* molecules at sites 601 and 609 showed slightly different cryo-EM maps at the C7 position. However, both C7-formyl groups were wrapped up by the cryo-EM maps, and the different appearances could be due to

the different angles of the C7-formyl groups in the Chl *b* molecules of rLHCII and nLHCII (Fig. S4 (34, 35)). In contrast, the cryo-EM map of the Chl molecule at site 614 appeared to have some density beyond the methyl group at the $C7^1$ position (as designated by R in Fig. 4h, see also Fig. S4), indicating that this site was occupied by both Chl *a* and Chl *b* molecules (Fig. 4g and h). In the 3D structural model of rLHCII (Protein Data Bank [PDB] code: 8YEE), Chl *a* was assigned as a representative at site 614; however, in reality, there was mixed occupancy of both Chl *a* and Chl *b* molecules.

Discussion

The His-tag location and its influence on the structure

Our successful affinity purification of rLHCII indicated that the His-tag protruded outside of rLHCII, as designed. This is consistent with our attribution that the additional electrostatic potential
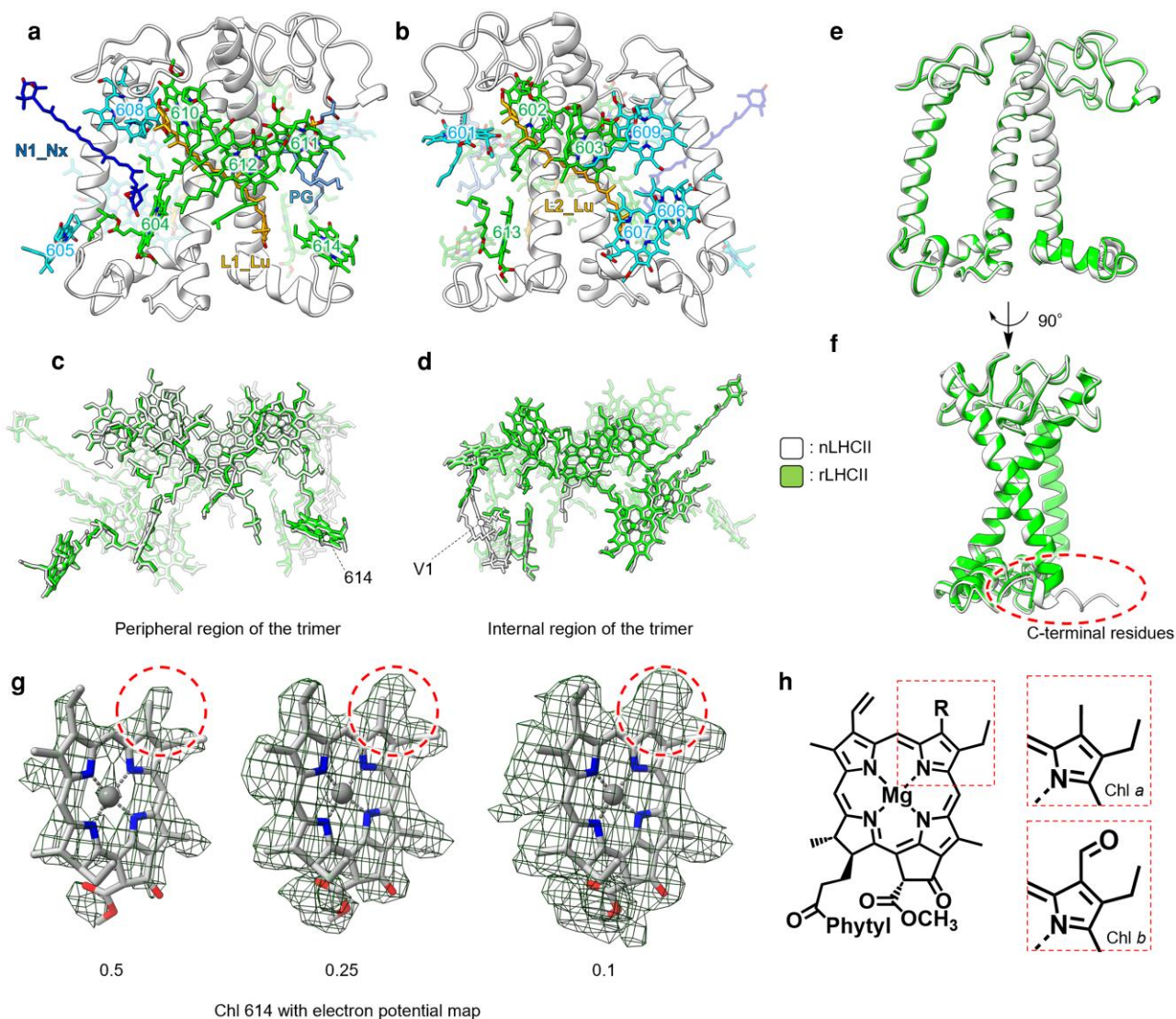
Fig. 4. The 3D model of rLHCII. Monomeric structure of rLHCII (a, b), superposed structure of pigments (c, d), and apoprotein (e, f) between rLHCII (green) and nLHCII (white). The model of Chl 614 with the electrostatic potential map at the series of cutoff values is shown in g. $C7^1$ positions are encircled with the dashed red line. For comparison, the chemical structures of Chl a and Chl b are shown in h. The structures of both LHCII were superposed based on the A, B, and C helices using CCP4MG. LHCII, light-harvesting complex II; rLHCII, reconstituted light-harvesting complex II; nLHCII, native light-harvesting complex II.

density protruding from rLHCII toward the lumenal side is due to the aggregation of the C-terminal His-tags of the three monomeric units; however, it is unclear when His-tag aggregation occurred. Interestingly, although the His-tag was present in each monomeric protein, the His-tag-like density was only one per trimer, located asymmetrically with respect to the C_3 center of symmetry (Fig. 3e). This can be interpreted as the His-tags of the three monomeric proteins gathering and tending to nestle at a thermodynamically stable position, independent of the center of symmetry of the trimer.

The His-tag-like density was very weak, indicating a disorder. The cryo-EM structures represented the polypeptide chain from Ser 14 to Val 229 in nLHCII and up to Asp 223 in rLHCII (Figs. 3d, 4f, and S2). In other words, six residues in the C-terminal region were not visible in rLHCII. Some of these invisible residues are located near Chl b at site 607 and Chl a at site 613 in nLHCII. Although the coordinates of the Chl molecules in rLHCII are virtually identical to those in nLHCII (Fig. 4d), their binding environments should be somewhat different because the probability of

the presence of nearby C-terminal residues is different. Therefore, when assessing the functions of these two Chl molecules using in vitro reconstitution methods, attention should be paid to their environmental differences from nLHCII.

The influence of His-tags on the native protein structure is usually treated as negligible based on validation using data for proteins consisting of a single polypeptide chain (36, 37). To the best of our knowledge, the length of the His-tag has not yet been optimized. Our results suggest that the optimization of the His-tag may alleviate these effects.

Lacking V1 carotenoids

The cryo-EM structure of rLHCII was trimeric and lacked a carotenoid at the V1 site (Fig. 4b and d). According to pigment composition analyses, the nLHCII trimer contained a small but nonnegligible amount of violaxanthin, which is 50% of neoxanthin (Fig. 2f). The pigment–lipid mixture used in the in vitro reconstitution procedure contained a similar proportion of violaxanthin (~50% of

neoxanthin); however, rLHCII contained very little violaxanthin (~10% of neoxanthin). The decrease in absorption intensity in the 400–500 nm region of rLHCII (Fig. 2a and b) corresponded to a shortage of carotenoids, as described above. These observations were consistent with the lack of V1 carotenoids in the rLHCII structure.

When a pigment–lipid mixture containing all three carotenoids (lutein, neoxanthin, and violaxanthin) in appropriate proportions is used for in vitro reconstitution, very small amounts of violaxanthin are usually detected for rLHCII (5, 14). These studies compared the pigment composition of rLHCII prepared using a pigment–lipid mixture containing a single carotenoid or various proportions of the two carotenoids. In all cases tested, the resulting rLHCII contained fewer than three carotenoids per protein. This suggests that carotenoids bound to the low-affinity V1 site are typically lost during the current reconstitution protocol. Therefore, new approaches are required to overcome this problem. Conversely, rLHCII may be useful for studies focusing on carotenoids bound to the N1, L1, and L2 sites without being influenced by the V1 carotenoid. Other minor light-harvesting complexes (LHCs), such as CP24, CP26, CP29, and Lhca1–4, are monomeric and bind three carotenoids per protein (38). The in vitro reconstitution method can be effective for directly comparing the carotenoid functionality with that of monomeric LHCs.

The absence of V1 carotenoids may affect the excited-state properties of nearby Chl molecules. Chl a 611, Chl b 601, and Chl b 607 are likely to be affected, and Chl a 613 and Chl a 614 may also be affected.

The mixed state of Chl a and Chl b at site 614

As shown in Fig. 2, the optical properties and the Chl a/b ratio indicated that rLHCII contained more Chl b molecules than nLHCII. This suggests that a certain Chl-binding site in rLHCII has a mixed occupancy of both Chl a and Chl b as predicted (12, 18, 39). Cryo-EM maps indicated that site 614 is the most probable candidate for the mixed occupancy (Figs. 4g and S4). However, for the remaining 13 Chl sites, the possibility of low-level occupancy of another Chl type cannot be completely ruled out. Mixed occupancy at site 614 was predicted by Remelli et al. (18) by comparing the spectroscopic results of rLHCII with site-directed mutagenesis; however, they also predicted mixed Chl occupancy at sites 613 and 604. Elias et al. (39) predicted that Chl b preferences at both sites 613 and 614 and mixed occupancy at site 604 using the Molecular Mechanics Poisson–Boltzmann surface area calculations. However, only site 614 appeared to be consistent with our cryo-EM results.

The Chl a/b ratio of the applied pigment–lipid mixture in previous studies was approximately three to reproduce that of nLHCII (12, 13). In the present study, the Chl a/b ratio of the applied mixture was 1.25 (Fig. 2f), corresponding to reconstitution in an environment with low Chl a content. Using a total LHCII extract supplemented with Chl a as the reconstitution mixture would increase the Chl a occupancy at the 614 site of rLHCII, and supplementing with Chl b could increase Chl b occupancy. Among the sites that are experimentally (18) and theoretically (39) predicted to have mixed Chl occupancy, site 614 appears to be controllable by the external Chl a/b ratio. Further investigations are required to confirm this hypothesis.

Peterman et al. (21) revealed an increase at 661 nm and a concomitant decrease at 676 nm in the 77 K absorption of rLHCII compared with that of nLHCII. They also showed that rLHCII decreased in CD intensity at a negative 680-nm signal compared with nLHCII. Based on the mixed occupancy of Chl a and Chl b

at site 614 in this study, these spectroscopic differences can be attributed to the partial replacement of Chl a with Chl b at site 614. This can be interpreted as the transition energy of Chl a at site 614 in nLHCII at 676 nm (77 K). Similarly, by combining structural and spectroscopic information, in vitro reconstitution has great potential for experimentally determining the transition energy of each Chl within a protein with high certainty. This will certainly contribute to the accurate attribution of each transition energy to each pigment within a protein and a deeper understanding of how the protein environment controls the excited state of the pigment.

Applicability of this method

However, this method is currently not applicable to monomeric LHCs. This is because, in single-particle analysis using cryo-EM, improving resolution is challenging as the particle size decreases (40, 41), which hinders obtaining a sufficiently high resolution to distinguish the structures of pigments, particularly between Chl a and Chl b molecules, bound to pigment–protein complexes. The molecular weights of the monomeric LHCs are small (~40 kDa). It is therefore difficult to determine the Euler angle of imaged particles precisely, resulting in inaccurate alignment of the particles. In the future, these measurement and algorithm issues may be resolved, thereby expanding the scope of this method.

Conclusion

We confirmed that the 3D structure of the rLHCII trimer with the correct lutein conformation, which exhibits the typical spectral properties of rLHCII, was virtually identical to that of nLHCII. This ensured the applicability of the in vitro reconstitution technique. However, the following precautions should be noted: (i) the structure of some residues in the C-terminal region differ from monomer to monomer, likely due to the aggregation of C-terminal His-tags of the three monomers; (ii) the V1 carotenoid is almost absent; and (iii) site 614 shows mixed occupancy by both Chl a and Chl b molecules. This structural validation ensured the in vitro reconstitution of LHCII, allowing for further elucidation of the functions of specific pigments.

Materials and methods
Preparation of native LHCII for cryo-EM

The nLHCII for cryo-EM was purified from *S. oleracea* purchased on the market, as described previously (42) with minor modifications: the thylakoid membrane was solubilized using 1% (w/v) n-dodecyl-α-D-maltoside (α-DM) to isolate pure free nLHCII by avoiding the dissociation of supercomplexes. After centrifugation, the supernatant was applied to a sucrose density gradient with a stepwise increase from 0.1 to 0.5 M sucrose and ultracentrifuged (Hitachi, P40ST, 248,500 × g, 16 h, 4 °C). To better separate trimeric LHCII from monomeric LHCs, the range of sucrose densities was narrowed. The nLHCII trimer fraction was further purified using another sucrose density gradient ultracentrifugation (Hitachi, P55ST2, 366,000 × g, 5 h, 4 °C), employing the same stepwise sucrose gradient. Sucrose present in the nLHCII trimer fraction was removed by ultrafiltration using a 50-kDa cutoff filter (Amiconultra-15, Millipore). The purified LHCII trimer was concentrated to 5 mg (total pigment and protein)/mL in 25 mM HEPES–NaOH buffer at pH 7.5 and containing 0.03% (w/v) α-DM, and used for cryo-EM analysis and spectroscopic analyses.

Preparation of pigment–lipid mixture for in vitro reconstitution

The nLHCII, as a source of the pigment–lipid mixture for in vitro reconstitution, was purified through anion exchange chromatography, as described previously (43) with minor modifications. The compositions of buffers were as follows: buffer A, 25 mM HEPES (pH 7.5) containing 0.03% (w/v) β-DM, and buffer B, buffer A with 1 M NaCl. The pigments and PG in the isolated nLHCII were extracted using a prechilled 1/1 (v/v) mixture of acetone and methanol. Precooled dichloromethane and precooled saturated saline were added to the extract in the test tube; after gentle stirring, the dichloromethane layer was taken out and dried with a stream of nitrogen gas to obtain a pigment extract.

DNA cloning and recombinant protein overexpression

The Lhcb1 gene (GenBank: X14341.1) of *S. oleracea* was amplified from the genome of commercially available *S. oleracea* by PCR and cloned into a pET-His expression vector (pET17b). Primers were designed as follows to remove the transit peptide and create recombinant proteins carrying six histidine residues at the C-terminal (Fig. 1b): forward primer, 5′-ATGAGGAAGACTGCAGGAAAG-3′; reverse primer, 5′-TCACTTTCCGGGGACAAAG-3′. Heat shock transformation was used to transform *E. coli* BL21 codon plus RIL (DE3) cells (Takarabio) with the recombinant vectors. His-tagged apoproteins were overexpressed by growing the *E. coli* with isopropyl-β-D-1-thiogalactopyranoside at 37 °C overnight. Finally, inclusion bodies were purified as described (13).

In vitro reconstitution of LHCII trimer and purification

The procedure was conducted following reference (13) with small modifications. The modifications were designed to achieve an in vitro reconstituted LHCII in which all components, such as pigments and lipids, were as similar as possible to nLHCII and to ensure reproducibility. In particular, we used a pigment mixture extracted directly from isolated LHCII rather than a pigment extract from whole cells, aiming to avoid complications such as adding a carotenoid mixture to the whole-cell pigment extract to adjust the pigment composition. In vitro reconstitution was performed without adding extra carotenoids using the pigment–lipid mixture extracted from nLHCII directly. We used 1,600 μg inclusion body and 1,000 μg pigment–lipid mixture for in vitro reconstitution processes. The four batches of reconstituted LHCII were combined and loaded on the Ni column with 1 mL resin (IMAC, BIO-RAD, USA) in an empty column (PD-10, i.d. 14.5 × 50 mm) and mixed by slowly rotating the column for 1 h at 4 °C. After flowing out, ultrafiltration was performed (100 kDa, Amicon Ultra-2 mL). For cryo-EM analysis, the rLHCII trimer was prepared through ultracentrifugation (Hitachi, P55ST2, 366,000 × g, 5 h, 4 °C). For characterization of rLHCII trimer, another ultracentrifugation (Hitachi, P40ST, 248,500 × g, 16 h, 4 °C) was conducted.

Spectroscopic analyses and pigment composition analyses using HPLC

Absorption, fluorescence and fluorescence excitation, and CD spectra were recorded in 25 mM HEPES–NaOH at pH 7.5 and containing 0.03% (w/v) α-DM using a spectrophotometer (UV-1800, Shimadzu, Kyoto, Japan), a fluorometer (FP-8500, JASCO, Tokyo, Japan), and a CD spectroscope (J720W, JASCO), respectively. The pigment composition analyses were performed, as described previously (42).

Cryo-EM data collection

We used the multiple blotting methods (44) with modification to overcome low sample concentration. Plasma treatment of Quantifoil Cu R 1.2/1.3 holey carbon grids was performed on both sides using JEC-3000FC (JEOL, Japan), for 10 s at 10 mA. A 2.5 μL aliquot of two types of LHCII (∼5.0 mg/mL) was applied to both sides of the grid and then manually blotted at room temperature. Quickly moving on to the normal blotting operation, 2.5 μL of the sample was applied on the grid was frozen quickly in liquid ethane using Vitrobot IV system (FEI) at 4 °C, 100% humidity, 3 s blotting time, 5 s waiting time and blotting force –10. Data for rLHCII and nLHCII were collected on cryo-electron microscopes, CRYO ARM 300 (JEM-Z300FSC, JEOL) and CRYO ARM 300 II (JEM-3300, JEOL), respectively, both equipped with a cold field-emission electron gun. The Cryo-EM images were collected using Serial-EM (45). The holes were detected using YoneoLocker (46). Movie frames were recorded using a K3 direct electron detection camera (Gatan, USA) at a nominal magnification of × 60,000 corresponding to pixel sizes of 0.873 Å (CRYO ARM 300) and 0.859 Å (CRYO ARM 300 II) at the specimen level. The data were collected with a total exposure of 3 s fractionated into 40 frames, with a total of ∼80 electrons Å^{-2} in counting mode. In total, 7,735 and 6,145 movies were collected for rLHCII and nLHCII, respectively, with defocus values varying from 0.7 to 2.2 μm. Typical cryo-EM images averaged from motion-corrected movie frames are shown in Fig. S1.

Cryo-EM data processing

A gain reference image was prepared with the relion_estimate_gain command in RELION 4.0 (47) using the first 300 movies. Images were processed using cryoSPARC ver. 4.2.1. (48). All collected movies were imported and motion corrected, and contrast transfer functions (CTFs) were estimated. In total, 5,889 and 5,199 movies from rLHCII and nLHCII were selected with maximum CTF fit resolution of images >5 Å. The particles were automatically picked using a template picker job with a particle diameter at 200 Å using a template made from the electron potential map of marine green algal LHCII trimer (PDB code: 8YYB) for nLHCII and using that of nLHCII for rLHCII. After particle extraction with 2× binning, several rounds of 2D classification into 50 classes were performed to select clear 2D class averages. In the case of rLHCII, three rounds of 2D classification were performed and the particles were extracted again with a box size of 256 × 256 pixel and a pixel size of 0.873 Å to select 311,204 particles. After performing homogeneous, local CTF, global CTF, and nonuniform refinements (48), the resolution reached 2.78 Å. Heterogeneous refinement was performed using the 3D map of 2.78 Å resolution and three maps by ab initio reconstruction. The classes with high resolution were selected, and the 228,507 particles were used for 3D variability analysis. Then, 10 clusters were made by 3D variability display, and for each class, homogeneous refinements were performed on each class. The 5 classes were selected, and a total of 179,318 particles were used for homogeneous refinement, and 2 rounds of local CTF, global CTF, and nonuniform refinements; finally, the resolution reached 2.57 Å with C_1 symmetry. An additional round of these processes was performed with C_3 symmetry for homogeneous and nonuniform refinement, resulting in a 3D map at 2.37 Å. In the case of nLHCII, 2 rounds of 2D classification were performed to select 650,078 particles. After 2D classification, the particles were used for ab initio reconstruction and classified into five classes. Two classes were selected and the 479,808 particles were extracted again with a box size of 256 × 256 pixel and a pixel size of 0.859 Å. After performing

homogeneous refinement, local and global CTF refinement, and nonuniform refinement, the resolution reached 2.42 Å. After using the volume tool for flip hand, an additional round of these refinements was performed and the resolution reached 2.32 Å with C_1 symmetry. The final nonuniform refinement was reperformed with C_3 symmetry, resulting in a final 3D map at 2.17 Å. All resolution calculations were calculated with FSC (Fourier Shell Correlation) = 0.143. The image processing strategy is detailed in Fig. S1.

Model building and validation

The atomic models of nLHCII and rLHCII were constructed using WinCOOT (version 0.9.447) (49), Phenix (version 1.19-415848) (50), and Chimera and ChimeraX (31) software packages. The coordinate of spinach LHCII (PDB: 1RWT) was directly docked as a template into the C_3 symmetry-imposed cryo-EM map of nLHCII and rLHCII using Chimera, so that the transmembrane helices were roughly matched to the electrostatic potential profile of the cryo-EM map. In the case of rLHCII, six C-terminal residues (FATNFV) were deleted. After real-space refinement in WinCOOT, the geometry-optimized model was globally refined in Phenix. The refinement statistics are summarized in Table S1.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Author Contributions

S.S., R.F., and G.K. designed the study; N.N. and R.F. designed the recombinant protein; N.N. prepared the plasmid for the recombinant protein; S.S. prepared the samples and performed the cryo-EM data analyses and the structural analysis; S.S. and T.M. measured EM micrographs, processed the EM data, and reconstructed the final EM map; T.M. and K.N. supervised the EM micrograph measurements and analyses; H.T. and G.K. supervised the structural model refinements; R.F., G.K., and K.N. supervised the project; S.S. and R.F. wrote the draft manuscript; S.S., G.K., K.N., and R.F. contributed for Funding Acquisition; all authors revised the final manuscript and contributed to the interpretation of the results and improvement of the manuscript.

Data Availability

Coordinate data for the structures of rLHCII and nLHCII have been deposited in the PDB under accession codes 8YEE and 8Y15, respectively. Cryo-EM maps for the structures of rLHCII and nLHCII with C_1 and C_3 symmetry have been deposited in the EM Data Bank under accession codes EMD-39305 (rLHCII, C_1), EMD-39192 (rLHCII, C_3), EMD-39304 (nLHCII, C_1), and EMD-38825 (nLHCII, C_3). All protein figures were made using ChimeraX (31).

References

1 Ruban AV. 2022. Chlorophyll a de-excitation pathways in the LHCII antenna. *J Chem Phys.* 156:070902.

2 Croce R, van Amerongen H. 2014. Natural strategies for photosynthetic light harvesting. *Nat Chem Biol.* 10. 492–501.

3 Liu Z, *et al.* 2004. Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature.* 428:287–292.

4 Standfuss J, van Scheltinga ACT, Kamborghini M, Kühlbrandt W. 2005. Mechanisms of photoprotection and nonphotochemical quenching in pea light-harvesting complex at 2.5 Å resolution. *EMBO J.* 24:918–928.

5 Croce R, Weiss S, Bassi R. 1999. Carotenoid-binding sites of the major light-harvesting complex II of higher plants. *J Biol Chem.* 274(42):29613–29623.

6 Hobe S, Prytulla S, Kühlbrandt W, Paulsen H. 1994. Trimerization and crystallization of reconstituted light-harvesting chlorophyll a/b complex. *EMBO J.* 13(15):3423–3429.

7 Dubertet G, Gerard-Hirne C, Trémoliéres A. 2002. Importance of *trans*-Δ^3-hexadecenoic acid containing phosphatidylglycerol in the formation of the trimeric light-harvesting complex in *Chlamydomonas. Plant Physiol Biochem.* 40:829–836.

8 Yoshihara A, Kobayashi K. 2022. Lipids in photosynthetic protein complexes in the thylakoid membrane of plants, algae, and cyanobacteria. *J Exp Bot.* 73(9):2735–2750.

9 Seiwert D, Witt H, Janshoff A, Paulsen H. 2017. The non-bilayer lipid MGDG stabilizes the major light-harvesting complex (LHCII) against unfolding. *Sci Rep.* 7(1):5158.

10 Müh R, Renger T. 2014. Structure-based calculation of pigment-protein and excitonic pigment-pigment coupling in photosynthetic light-harvesting complex. In: Golbeck J, van der Est A, editors. *The biophysics of photosynthesis. Biophysics for the life sciences* 11. New York (NY): Springer. p. 3–45.

11 Krüger TPJ, Novoderezhkin VI, Romero E, van Grondelle R. 2014. Photosynthetic energy transfer and charge separation in higher plants. In: Golbeck J, van der Est A, editors. *The biophysics of photosynthesis. Biophysics for the life sciences.* Vol. 11. New York (NY): Springer. p. 79–81.

12 Belgio E, Ruban AV. 2017. Recombinant light harvesting complexes: views and perspectives. In: Hou H, Najafpour M, Moore G, Allakhverdiev S, editors. *Photosynthesis: structures, mechanisms, and applications.* Cham: Springer. p. 33–49.

13 Natali A, Roy LM, Croce R. 2014. In vitro reconstitution of light-harvesting complexes of plants and green algae. *J Vis Exp.* 92: e51852.

14 Hobe S, Niemeier H, Bender A, Paulsen H. 2000. Carotenoid binding sites in LHCIIb. *Eur J Biochem.* 267(2):616–624.

15 Elias E, Liguori N, Saga Y, Schäfers J, Croce R. 2021. Harvesting far-red light with plant antenna complexes incorporating chlorophyll d. *Biomacromolecules*. 22(8):3313–3322.

16 Yang C, Kosemund K, Cornet C, Paulsen H. 1999. Exchange of pigment-binding amino acids in light-harvesting chlorophyll ab protein. *Biochemistry*. 38(49):16205–16213.

17 Rogl H, Kühlbrandt W. 1999. Mutant trimers of light-harvesting complex II exhibit altered pigment content and spectroscopic features. *Biochemistry*. 38(49):16214–16222.

18 Remelli R, Varotto C, Sandonà D, Croce R, Bassi R. 1999. Chlorophyll binding to monomeric light-harvesting complex. *J Biol Chem*. 274(47):33510–33521.

19 Kondo T, *et al.* 2017. Single-molecule spectroscopy of LHCSR1 protein dynamics identifies two distinct states responsible for multi-timescale photosynthetic photoprotection. *Nat Chem*. 9(8):772–778.

20 Ruan M, *et al.* 2023. Cryo-EM structures of LHCII in photo-active and photo-protecting states reveal allosteric regulation of light harvesting and excess energy dissipation. *Nat Plants*. 9(9): 1547–1557.

21 Peterman EJ, *et al.* 1996. Low-temperature spectroscopy of monomeric and trimeric forms of reconstituted light-harvesting chlorophyll *ab* complex. *Biochim Biophys Acta*. 1273(2):171–174.

22 Plumley FG, Schmidt GW. 1987. Reconstitution of chlorophyll *a/b* light-harvesting complexes: xanthophyll-dependent assembly and energy transfer. *Proc Natl Acad Sci U S A*. 84(1):146–150.

23 Paulsen H, Rümler U, Rüdiger W. 1990. Reconstitution of pigment-containing complexes from light-harvesting chlorophyll *a/b*-binding protein overexpressed in *Escherichia coli*. *Planta*. 181(2):204–211.

24 Horn R, Paulsen H. 2004. Early steps in the assembly of light-harvesting chlorophyll *a/b* complex. *J Biol Chem*. 279(43): 44400–44406.

25 Horn R, Grundmann G, Paulsen H. 2007. Consecutive binding of chlorophylls *a* and *b* during the assembly in vitro of light-harvesting chlorophyll-a/b protein (LHCIIb). *J Mol Biol*. 366(3): 1045–1054.

26 Dockter C, *et al.* 2009. Refolding of the integral membrane protein light-harvesting complex II monitored by pulse EPR. *Proc Natl Acad Sci U S A*. 106(44):18485–18490.

27 Fehr N, García-Rubio I, Jeschke G, Paulsen H. 2016. Early folding events during light harvesting complex II assembly in vitro monitored by pulsed electron paramagnetic resonance. *Biochim Biophys Acta*. 1857(6):695–704.

28 Hobe S, Foerster R, Klingler J, Paulsen H. 1995. N-proximal sequence motif in light-harvesting chlorophyll a/b-binding protein is essential for the trimerization of light-harvesting chlorophyll a/b complex. *Biochemistry*. 34(32):10224–10228.

29 Van Amerongen H, Van Grondelle R. 2001. Understanding the energy transfer function of LHCII, the major light-harvesting complex of green plants. *J Phys Chem B*. 105(3):604–617.

30 Elias E, *et al.* 2024. Coloring outside the lines: exploiting pigment–protein synergy for far-red absorption in plant light-harvesting complexes. *J Am Chem Soc*. 146:3508–3520.

31 Pettersen EF, *et al.* 2004. UCSF chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 25:1605–1612.

32 Khachik F, Chang AN. 2009. Total synthesis of (3R,3′R,6′R)-lutein and its stereoisomers. *J Org Chem*. 74(10):3875–3885.

33 Britton G, Pfander H, Liaaen-Jensen S. editors. 2004. *Carotenoids handbook*. Basel: Springer Basel AG. p. 133.

34 Gisriel CJ, Wang J, Brudvig GW, Bryant DA. 2020. Opportunities and challenges for assigning cofactors in cryo-EM density maps of chlorophyll-containing proteins. *Commun Biol*. 3(1):408.

35 Gisriel CJ, *et al.* 2021. Quantitative assessment of chlorophyll types in cryo-EM maps of photosystem I acclimated to far-red light. *BBA Adv*. 1:100019.

36 Carson M, Johnson DH, McDonald H, Brouillette C, Delucas LJ. 2007. His-tag impact on structure. *Acta Crystallogr D Biol Crystallogr*. 63:295–301.

37 Ayoub N, Roth P, Ucurum Z, Fotiadis D, Hirschi S. 2023. Structural and biochemical insights into His-tag-induced higher-order oligomerization of membrane proteins by cryo-EM and size exclusion chromatography. *J Struct Biol*. 215(1):107924.

38 Nicol L, Croce R. 2018. Light harvesting in higher plants and green algae. In: van Grondelle R, van Amerongen H, van Stokkum I, editors. *Light harvesting in photosynthesis*. CRC Press. p. 59–67.

39 Elias E, Liguori N, Croce R. 2023. At the origin of the selectivity of the chlorophyll-binding sites in light harvesting complex II (LHCII). *Int J Biol Macromol*. 243:125069.

40 Wu M, Lander GC. 2020. How low can we go? Structure determination of small biological complexes using single-particle cryo-EM. *Curr Opin Struct Biol*. 64:9–16.

41 Nygaard R, Kim J, Mancia F. 2020. Cryo-electron microscopy analysis of small membrane proteins. *Curr Opin Struct Biol*. 64:26–33.

42 Seki S, Yamano Y, Oka N, Kamei Y, Fujii R. 2022. Discovery of a novel siphonaxanthin biosynthetic precursor in codium fragile that accumulates only by exposure to blue-green light. *FEBS Lett*. 596:1544–1555.

43 Seki S, *et al.* 2022. Structural insights into blue-green light utilization by marine green algal light harvesting complex II at 2.78 Å. *BBA Adv*. 2:100064.

44 Snijder J, *et al.* 2017. Vitrification after multiple rounds of sample application and blotting improves particle density on cryo-electron microscopy grids. *J Struct Biol*. 198(1):38–42.

45 Mastronarde DN. 2005. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol*. 152:36–51.

46 Yonekura K, Maki-Yonekura S, Naitow H, Hamaguchi T, Takaba K. 2021. Machine learning-based real-time object locator/evaluator for cryo-EM data collection. *Commun Biol*. 4:1044.

47 Zivanov J, *et al.* 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife*. 7:e42166.

48 Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. 2017. Cryosparc: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Method*. 14:290–296.

49 Emsley P, *et al.* 2010. Features and development of coot. *Acta Crystallogr D Biol Crystallogr*. 66:486–501.

50 Adams PD, *et al.* 2010. Phenix: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 4:43–44.