PNAS
www.pnas.org

Supplementary Information for

Ycf48 involved in the biogenesis of the oxygen-evolving photosystem II complex is a seven-bladed beta-propeller protein

Jianfeng Yu[a,1,2] Jana Knoppová[b,1], Franck Michoux[a], Wojciech Bialek[a,3], Ernesto Cota[a], Mahendra K. Shukla[b,c], Adéla Strašková[b], Guillem Pascual Aznar[b,c], Roman Sobotka[b,c], Josef Komenda[b,c], James W. Murray[a] and Peter J. Nixon[a,2]

[a]Wolfson Laboratories, Department of Life Sciences, Imperial College London, SW7 2AZ London, United Kingdom

[b]Institute of Microbiology, Center Algatech, Opatovický mlýn, 37981 Třeboň, Czech Republic

[c]Faculty of Science, University of South Bohemia, 37005 České Budějovice, Czech Republic

[1]These authors contributed equally to this work

[2]Correspondence authors

[3]Current address: Department of Biophysics, Faculty of Biotechnology, University of Wroclaw, Wroclaw, Poland

Jianfeng Yu Email: j.yu@imperial.ac.uk
Peter J. Nixon Email: p.nixon@imperial.ac.uk

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1800609115

SI Materials and Methods

Strains and cultivation conditions

The following strains of *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis*) were used in this work: the glucose-tolerant WT strain, GT-P (1); the CP43-less strain, ΔCP43 (2); the CP47 deletion strain, ΔCP47 (3); the PSI-less mutant, ΔPSI (4); and its derivative lacking the *psbC* gene (ΔPSI/ΔCP43) (5). The oligo-histidine tagged strains accumulating unassembled CP43-His, unassembled CP47-His, RC47-His and intact His-tagged PSII core complexes are described in (5, 6).*Thermosynechoccus elongatus* BP-1 (7) was obtained from Prof. James Barber (Imperial College London, UK) and *Cyanidioschyzon merolae* NIES-1332 was from the National Institute for Environmental Studies, Japan.

Two *ycf48* disruption mutants were used in this work. The *ycf48* null strain (ΔYcf48) described in this study was generated in two steps. First, the upstream and downstream flanking sequences of the Ycf48 ORF (*slr2034*) were amplified from WT genomic DNA of *Synechocystis* with primer set Ycf48-1F (AGATTACTAAAGATTTACCCCCAGTTCTCTGCTGTAGT) and Ycf48-2R (AGTAGAGTCTAGTCTAC TATTTTGTGAGCCGCGATATCTGTGAACAGGAATAATCAACGAGTACAGCAAGGGCT), and Ycf48-3F (AGCCCTTGCTGTACTCGTTGATTATTCCTGTTCACAGATATCGCGGCTCACAAAATAGTAGACTAGA CTCTACT) and Ycf48-4R (TTCAATTCCACCGGTTGGCGGTTTGGGTTTGAA). Overlap extension PCR was then performed to fuse the flanking sequences using primer set Ycf48-1F and Ycf48-4R. The resulting PCR fragment starting 445 bp upstream and ending 555 bp downstream of *slr2034*, with the Ycf48 ORF replaced by an EcoRV site, was then cloned into a pGEM-T Easy vector backbone to yield pGEMYcf48. The second step was to introduce a kanamycin-resistance cassette (NCBI accession: X06404) into the EcoRV site of pGEMYcf48 (Fig. S7B). The resulting plasmid, designated pYcf48kan, was then used to transform WT *Synechocystis*, to yield the ΔYcf48 strain (Fig. S7B,C). The other ycf48 disruption mutant used in this study (*ycf48*$^{-}$) refers to a previously constructed strain in which the *ycf48* open reading frame was disrupted via insertion of a chloramphenicol-resistance cassette (8).

To construct the Ycf48-REquad mutant, *slr2034* along with 445 bp of upstream sequence was amplified from WT genomic DNA of *Synechocystis* using primer set Ycf48-1F and Ycf48WT-2R (AAGTAGAGTCTAGTCTACTATTTTGTGAGCCGCGATATCCTAGGGAACCATTGCCACCTCGGTGCT), and 555 bp of downstream sequence was amplified with primer set Ycf48WT-3F (AGCACCGAGGTGGCAATGGTTCCCTAGGATATCGCGGCTCACAAAATAGTAGACTAGACTCTACTT) and Ycf48-4R. The two fragments were then fused by overlap extension PCR using primer set Ycf48-1F and Ycf48-4R and the final PCR product consisting of an EcoRV site behind the STOP codon of *slr2034* was then cloned into pGEM-T Easy vector, to yield pGEMYcf48WT. A gentamycin-resistance cassette (NCBI accession: NC_004998) was inserted into the EcoRV site in the same transcription orientation as *slr2034* to give plasmid pYcf48WTgen, which was used as the parental vector for further mutagenesis work (Fig. S7B).

To generate the Ycf48-REquad mutations, a 351 bp synthetic DNA fragment (Genewiz, USA) carrying the R196E (CGG->GAG), R215E (CGC->GAG), R219E (CGA->GAA) and R220E (CGC->GAG) mutations was used to replace the WT coding sequence between the HindIII and BstEII sites of pYcf48WTgen. The resulting vector pYcf48REgen was used to transform the ΔYcf48 strain to generate the gentamycin-resistant Ycf48-REquad mutant. The Ycf48-RAquad mutant was constructed using the same approach, but with a synthetic DNA fragment carrying R196A (CGG->GCG), R215A (CGC->GCC), R219A (CGA->GCA) and R220A (CGC->GCC) mutations.

To generate the Ycf48-CM strain expressing a chimera of the endogenous Ycf48 from *Synechocystis* and the 19-residue eukaryotic-specific insert found in Ycf48 from *Cyanidioschyzon merolae,* a 408 bp synthetic DNA fragment (Genewiz, USA) in which the sequence GATGCTACCTTAAATCGAAC TATCTCTTCTGGTATCACCGGTGCCAGTTATTTTACC, encoding the sequence DATLNRTISSGITG ASYFT, inserted at its corresponding location in Ycf48 from *Synechocystis* (between the codons encoding V176 and G177), was used to replace the WT coding sequence of Ycf48 between the HindIII and BstEII sites of pYcf48gen. The resulting plasmid, pYcf48CMgen, was used to transform the ΔYcf48 strain to generate the gentamycin-resistant Ycf48-CM mutant.

The transformation vectors used to introduce single Ycf48 R->E mutations were constructed using pYcf48WTgen as a template. Overlap extension PCR reactions were carried out using primer set Ycf48-AvrII-Infu-F (AGCTGATTGACCTAGGGGAAGA) and either Ycf48-R196E-R (GGCTGACACCGCCACA TAA) or Ycf48-R215E-R (GTTATGGGGAGTCCATTCCGT) or Ycf48-R219E-R (GGAAGAATTGC GGTTATGGGG) or Ycf48-R220E-R (TCGGGAAGAATTGCGGTTATG), and primer set Ycf48-R196E-F (TTATGTGGCGGTGTCAGCCGAGGGTAATTTTTATTCCACCTGGG) or Ycf48-R215E-F (ACGGA ATGGACTCCCCATAACGAGAATTCTTCCCGACGCTTACAG) or Ycf48-R219E-F (CCCCATAACCG CAATTCTTCCGAACGCTTACAGACCATGGGCT) or Ycf48-R220E-F (CATAACCGCAATTCTTCCC GAGAGTTACAGACCATGGGCTATGGCA) and Ycf48-BamHI-Infu-R (TGCGCGAACAGGATCC CCATCCT). The resulting two fragments obtained for each construct were then PCR fused with primer set AvrII-Infu-F and Ycf48-BamHI-Infu-R, and integrated into the AvrII and BamHI sites of pYcf48WTgen plasmid through In-Fusion cloning (Takara Bio, USA), thereby replacing the WT sequence in the vector. The resulting plasmids pYcf48R196Egen, pYcf48R215Egen, pYcf48R219Egen and pYcf48R220Egen, were then used to transform the ΔYcf48 strain to yield the single gentamycin-resistant Ycf48-RxxxE mutants, where xxx is the residue mutated (Fig. S7C).

The mutagenesis vectors used to introduce the triple Ycf48 R->E mutations were constructed by overlap extension PCR using pYcf48-REquadgen as the DNA template. The first round of PCR reactions were carried out using primer set Ycf48-AvrII-Infu-F and either Ycf48-R196E-R or Ycf48-R215E-R or Ycf48-R219-R (GGAAGAATTCTCGTTATGGGG) or Ycf48-R220-R (TTCGGAAGAATTCTCGTTATG), and primer set Ycf48-R196-F (TTATGTGGCGGTGTCAGCCCGGGGTAATTTTTATTCCACCTGGG) or Ycf48-R215-F (ACGGAATGGACTCCCCATAACCGCAATTCTTCCGAAGAGTTACAG) or Ycf48-R219-F

(CCCCATAACGAGAATTCTTCCCGAGAGTTACAGACCATGGGCT) or Ycf48-R220-F (CATAACGAGA ATTCTTCCGAACGCTTACAGACCATGGGCTATGGCA) and Ycf48-BamHI-Infu-R. The resulting two fragments for each construct were then PCR fused with primer set AvrII-Infu-F and Ycf48-BamHI-Infu-R, and inserted into the AvrII and BamHI sites of pYcf48WTgen through In-Fusion cloning (Takara Bio, USA), thereby replacing the wild-type sequence. The resulting plasmids pYcf48E196Rtripgen, pYcf48E215Rtripgen, pYcf48E219Rtripgen and pYcf48E220Rtripgen, were then used to transform ΔYcf48 strain to yield the gentamycin-resistant Ycf48-ExxxRtrip mutants, where the number in the name of each vector represents the residue restored from E to R. The genotype of Ycf48 mutants after transformation was verified by PCR amplification from genomic DNA using primer set Ycf48-1F or Ycf48-seg-F (ACTACTTCATCAGCACCGGGG) and Ycf48-seq-R (CTAGTCTACTATTTTGTGAGCCGC). Once segregation was confirmed (Fig. S7C,D), the PCR fragments were gel purified and the correct sequence confirmed.

Strains were grown in liquid BG-11 medium on a rotary shaker under moderate (normal) light intensities (40 µE m^{-2} s^{-1}) at 28 °C either photoautotrophically or mixotrophically in the presence of 5 mM glucose. Strains lacking photosystem I (PSI) were grown at lower light intensities (10 µE m^{-2} s^{-1}). GRO-LUX fluorescent tubes (Sylvania, Germany) were used as the light source, and the light intensity was measured with a portable light meter (Hansatech Instruments, UK).

Cell absorption spectra and determination of chlorophyll content

Absorption spectra of whole cells were measured at room temperature using a UV-3000 spectrophotometer (Shimadzu, Japan). To determine chlorophyll levels, pigments were extracted from cell pellets with 100% methanol and the chlorophyll concentration was determined spectroscopically (9).

Analysis of chlorophyll-derived pigments and precursors by HPLC

For quantitative determination of chlorophyll-related pigments, a 3 ml culture at $OD_{750\ nm}$=0.5-0.6 was spun down and cells resuspended in 20 µl water. Pigments were extracted with 70% (v/v) methanol in water, filtered and immediately injected into an HPLC (Agilent-1200). Separation was carried out on a reverse phase column (ReproSil pur 100, C8, 3 µm particle size, 4x150 mm, Watrex) with 35% (v/v) methanol and 15% (v/v) acetonitrile in 0.25 M pyridine (solvent A) and 50% (v/v) methanol in acetonitrile (solvent B). Pigments were eluted with a gradient of solvent B (40 to 52% in 5 min) followed by 52 to 55% of solvent B in 30 min at a flow rate of 0.8 ml min^{-1} at 40 °C. Eluted pigments were detected by two fluorescence detectors set at several different wavelengths to detect all chlorophyll precursors from coproporphyrin (ogen) III to monovinyl-chlorophyllide; for details see (10).

Preparation of cellular membranes and their trypsinization

Cell cultures were harvested at OD_{750nm}~0.5-0.7. Cells were pelleted, washed, and resuspended in a lysis buffer (25 mM MES/NaOH pH=6.5, 10 mM $CaCl_2$, 10 mM $MgCl_2$, 25% (v/v) glycerol) with additional

supplementation of protease inhibitor (cOmplete, EDTA-free protease inhibitors cocktail tablets, Roche Applied Science, UK). Cells were broken using 0.1 mm diameter zirconia beads in a mini bead beater and the membrane fraction was separated from soluble proteins by centrifugation at high speed (65,000 *x g*, 20 min) and resuspended in the lysis buffer. For the analysis of the cytoplasmic/lumenal location of Ycf48, the membranes were prepared by breaking the cells in a Tris buffer and treated with trypsin as described in (11).

Isolation of YidC-FLAG and PSI-YFP by affinity chromatography

The YidC complex was purified from a *Synechocystis* strain expressing the YidC-3xFLAG protein (YidC-FLAG) under control of the *psbA2* promoter (12). YidC-FLAG was purified from whole-cell extract without any compression of membrane vesicles through centrifugation. Cells from a 4 liter culture were broken in 25 mM K-phosphate buffer, pH=7.5 containing 10% (v/v) glycerol (buffer B), mixed with 1% (w/v) n-dodecyl β-D-maltoside (DDM) and incubated for 30 min at 8 °C. After centrifugation at high speed (65,000 x *g*, 20 min) the supernatant was loaded on an anti-FLAG-M2 agarose column (Sigma-Aldrich, USA). To remove contaminants, the anti-FLAG-resin was washed with 20 column volumes of buffer B containing 0.04% (w/v) DDM. The YidC-FLAG complex was eluted with 2.5 column volumes of buffer B containing 150 mg ml^{-1} 3xFLAG peptide (Sigma-Aldrich, USA) and 0.04% (w/v) DDM. Analysis of the complex by mass spectrometry was performed as described in (12).

The PSI-YFP complex was purified from a *Synechocystis* strain expressing PsaF-YFP using the GFP-Trap as described in (13). Purification of His-tagged PSII complexes was done as described in (5, 6).

Protein electrophoresis, immunoblotting and radiolabelling

For native electrophoresis, solubilized membrane proteins or isolated complexes were separated on 4-12% (w/v) clear native gels (14) or blue native gels (15). Individual components of the protein complexes were resolved by incubating the gel strip from the first dimension in 2% (w/v) sodium dodecyl sulfate (SDS) and 1% (w/v) dithiothreitol for 30 min at room temperature, and proteins were separated in the second dimension by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) in a denaturing 12 to 20% (w/v) polyacrylamide gel containing 7 M urea (16). For standard single dimension SDS-PAGE, membrane proteins from cyanobacteria were solubilized at room temperature for 1 h, with 62.5 mM Tris-HCl pH=6.8, 10% (v/v) glycerol, 2% (w/v) SDS, 0.5% (w/v) bromophenol blue and 5% (v/v) beta-mercaptoethanol, before loading on the gel. Proteins were stained by Coomassie Brilliant Blue (CBB) or by SYPRO Orange; and in the latter case, they were subsequently transferred onto a PVDF membrane. Membranes were incubated with specific primary antibodies and then with a secondary antibody conjugated with horseradish peroxidase (Sigma-Aldrich). The following primary antibodies were used in the study: anti-Ycf48 raised in rabbit against recombinant *Synechocystis* Ycf48 (residues F37-P342 of the predicted protein) expressed and purified as a His-tagged derivative from a 1 litre *Escherichia coli* culture using the modified pRSET-A expression plasmid and methods described in (17), the His-tag was

removed by thrombin digestion prior to immunization (Seqlab, Germany); anti-CP47, anti-CP43, anti-D1, anti-D2 and anti-PsaD (5); anti-FtsH2, anti-FtsH3 and anti-PHB1 (18); anti-Ycf39 (19); and anti-His tag and anti-CyanoP (Agrisera, Sweden).

For protein labelling, the cells were incubated with a mixture of [^{35}S]-Met and [^{35}S]-Cys (Translabel; MP Biochemicals) as described previously (16) and two-dimensional protein separation was performed as described above. The 2D gel was stained by CBB, photographed, dried, exposed to a phosphorimager plate (GE Healthcare, Austria) overnight and scanned by Storm 860 (GE Healthcare, Austria).

Vectors used for expression in Escherichia coli

All Ycf48 derivatives used for structural determination were produced from a modified *E. coli* expression vector system pRSET-A (described in (20)). To express *T. elongatus* Ycf48, the coding sequence (Cyanobase designation *tll1695*) was amplified from WT genomic DNA with primer pair Ycf48-Elo-F (TTTTTTGGATCCATTCCTGCTTTAGACTACAACCCT) and Ycf48-Elo-R (TTTTTTAAGCTTTTACGC TGGCGCAGCCGT). The resulting PCR fragment encoding mature Ycf48, without the predicted 38-residue signal peptide at the N-terminus, was then digested with BamHI and HindIII prior to cloning into the corresponding sites of pRSET-A. The resulting expression vector pRSET48TE encodes a fusion protein that consists of an N-terminal 6xHis tag, followed by a thrombin cleavage site and Ycf48 lacking the signal peptide (Fig. S7A).

Plasmid pRSET48RE expressing the Ycf48 quadruple R->E variant was constructed by cloning a synthetic DNA fragment (Genewiz, USA), encoding mature Ycf48 from *T. elongatus* plus the R202E/R221E/R225E/R226E quadruple mutations, into the BamHI and HindIII sites of the pRSET-A vector backbone.

The vector expressing *C. merolae* Ycf48 was constructed by amplifying the coding sequence of Ycf48 from WT genomic DNA of *C. merolae* (GeneID: 16995794/ locus tag: CYME_CMO314C) with primer set Ycf48-Mero-F (TATATACTCGAGGCTTTGCCGCGCAGCGGAGCGCTAGT) and Ycf48-Mero-R (GCGCGCAAGCTTTTATCAGACGTTTTCCGGATGAAAT), and then ligating the resulting PCR fragment into pRSET-A after digestion with XhoI and HindIII. The resulting vector, pRSET48CM, encodes a Ycf48 derivative containing an N-terminal 6xHis tag, followed by a thrombin cleavage site and *C. merolae* Ycf48 lacking the predicted 160-residue N-terminal transit peptide (Fig. S7A).

The vector expressing *Synechocystis* Ycf48 was constructed by PCR amplification of Ycf48 from genomic DNA with primer pair Ycf48-Syn-F (TTTTTTGGATCCTTCAATCCATGGCAAGAAATTGCTT) and Ycf48-Syn-R (TTTTTTCTCGAGCTAGGGAACCATTGCCACCT). The resulting PCR fragment was then digested with BamHI and XhoI and cloned into the corresponding sites of pRSET-A, yielding vector pRSET48SYN (Fig. S7A).

Protein expression, purification and crystallization

Single step (KRX) competent cells (Promega, USA) were used to produce native Ycf48 protein and B834 (DE3) cells (Novagen, USA) were used for incorporation of selenomethionine. For native crystals, 1 liter culture of transformed KRX cells were grown at 37 ^{o}C with orbital shaking at 220 rpm until it reached OD_{600nm}~0.6. The cell culture was then cooled to 18 ^{o}C and induced with 0.1% (w/v) L-rhamnose overnight. For selenomethionine derivatives, transformed B834 (DE3) cells were directly grown in minimal media supplemented with 50 mg L^{-1} selenomethionine (Molecular Dimensions, USA) at 37 °C until the cell density reached OD_{600nm}~0.6-0.8 and induction was performed with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 18 °C overnight. After harvesting, the cell pellet was resuspended in 25 ml lysis buffer (50 mM Tris-HCl pH=7.9, 500 mM NaCl and 1 mM $MgCl_2$) and supplemented with protease inhibitor (cOmplete, EDTA-free protease inhibitors cocktail tablets, Roche Applied Science, UK) prior to disruption. The cell suspension was kept in an ice-water bath and disrupted using a sonicator Q125 (Qsonica, USA) using 5 s on and 5 s off interval, for 20 min. Cell debris and unbroken cells were removed by centrifuging at 18,000 x *g* at 4 ^{o}C for 15 min. The supernatant was then transferred to 5 ml pre-equilibrated Ni-IDA resin and incubated at 4 ^{o}C for 1 h. After binding, the resin was washed with 6 ml wash buffer (20 mM Tris-HCl pH=7.9, 500 mM NaCl and 60 mM imidazole) three times and the target protein was then eluted with elution buffer (20 mM Tris–HCl pH=7.9, 500 mM NaCl and 1 M imidazole). The purified His_6-tag bound Ycf48 derivative was then concentrated using a 3.5 K MWCO concentrator (Sartorius, Germany) and dialysed overnight at 4 °C against 20 mM Tris–HCl pH=7.9 and 200 mM NaCl buffer. The oligo-histidine tag was removed by thrombin (GE Healthcare, UK) cleavage at 1 U/100 µg of tagged protein overnight at 4 ^{o}C and protein lacking the oligo-histidine tag was isolated from the unbound fraction after incubating with Ni-IDA resin for 1 h at 4 ^{o}C and concentrating. Protein samples were subjected to crystallization screens using a Mosquito® robot (TTP, LabTech, UK).

Ycf48 structural determination

Selenomethionine substituted Ycf48 was crystallized in 0.2 M Li_2SO_4, 20% (w/v) PEG 3350, pH=6.0. An anomalous dataset at the Se peak energy was collected at BM14, ESRF, and Se sites located, and phases calculated using the hkl2map pipeline (21). The model was built in COOT (22), and refined against a 1.5 Å resolution dataset from native protein. Later structures were solved with molecular replacement with Phaser (23) using this first model.

The iD1 peptide NAHNFPLDLASAESAPVA (Lifetein, USA) and mD1 NAHNFPLDLA (Activotec, UK) were used without further purification. Crystallization conditions were rescreened with 3-fold molar excess of peptide and 10 mg ml^{-1} protein. These gave two new crystal forms. Ycf48 (HCF136) from *C. merolae* and the *T. elongatus* Ycf48-REquad variant were crystallized using 10 mg ml^{-1} protein according to Table S1, and the structures solved by molecular replacement.

Data for native proteins were collected at Diamond or Soleil synchrotrons, and processed with xia2 (24) or XDS (25). Structures were refined with REFMAC or Phenix. Details of crystallization, data processing and refinement are given in Table S1.

Structural images were made using PyMol, vacuum electrostatics were also generated in PyMol, with the default Blue (positive) to Red (negative) color scheme.

SI Figure legends

Fig. S1 Presence of Ycf48-containing complexes in membranes of *Synechocystis* strains lacking various chlorophyll-binding proteins. Membranes isolated from WT, a *psbC* deletion strain (ΔCP43), a *psbB* deletion strain (ΔCP47), a *psaAB* deletion strain (ΔPSI) and a *psaAB/psbC* deletion strain (ΔPSI/ΔCP43) were analyzed by 2D-Blue Native/SDS-PAGE electrophoresis (2D-BN/SDS-PAGE), the gels were stained by SYPRO Orange, blotted to a PVDF membrane and Ycf48 was detected by a specific antibody (for mutants just segments of the 2D blots with the specific antibody signals are shown, each obtained using membranes from the designated strain). Signals obtained for CP47 are also shown to allow a correct alignment of the major PSII complexes with Ycf48-containing complexes, as indicated by the vertical lines. Designation of complexes in the SYPRO-stained gel of WT sample: PSI(1), monomeric PSI complex; PSI(3), trimeric PSI complex; RCC(2) and RCC(1), dimeric and monomeric PSII core complexes, respectively; U.P., unassembled proteins. In the case of the immunoblots, the migration of PSI and PSII complexes containing Ycf48 are designated by an asterisk. RC47 accumulates in the ΔCP43 mutant; RCII* and RCIIa, which are reaction center complexes lacking CP47 and CP43 but containing Ycf48, accumulate in the ΔCP47 mutant (8, 19). Thylakoid membranes containing 4 µg chlorophyll (1.5 µg chlorophyll for PSI-less strains) were analyzed.

Fig. S2 Presence of Ycf48 in various PSII complexes isolated using His-tagged CP43 and CP47 and evidence for lumenal location of Ycf48 in *Synechocystis.* (A) Immobilized Ni-metal affinity purification of detergent-solubilised membrane samples of His-tagged CP47 in a WT background (PSII-His) and a WT control. Samples at each stage of purification were analyzed by SDS-PAGE and the gel was either stained by Coomassie Brilliant Blue (CBB) or blotted onto a PVDF membrane and sequentially probed with antibodies specific for D1 and Ycf48 (segments of the blot with the specific antibody signal are shown). TM: thylakoid membranes, Sol: thylakoid membranes solubilized with DDM, FT: unbound flow-through fraction from the nickel resin, Elu: eluates obtained with 100mM imidazole. 10 µl of Elu and 5 µl of TM, Sol and FT were loaded on the gel. (B) His-tagged PSII sub-complexes CP47-His, RC47-His, CP43-His and more intact core complexes (PSII-His MES) were isolated from the ΔD1/CP47-His, ΔCP43/CP47-His, ΔD1/CP43-His and PSII-His strains, respectively, by a combination of Ni-metal affinity and size exclusion chromatography as described previously (5, 6). The use of a strain lacking D1 blocks assembly of PSII at an early stage thereby allowing accumulation of CP43-His and CP47-His. The complexes were analyzed by SDS-PAGE in an 18% (w/v) gel. WT thylakoid membranes (0.5 µg chlorophyll) and samples of CP47-His, RC47-His, CP43-His and PSII-His (1 µg chlorophyll) were loaded on the gels and proteins in the gel were either stained by CBB or blotted onto a PVDF membrane (Blots) and sequentially probed with the designated antibodies (segments of the blot with the specific antibody signal are shown). (C) Isolated intact membranes of WT *Synechocystis* were treated with

trypsin for 0, 5, 15 and 30 min and analyzed by SDS-PAGE. The gel was stained by SYPRO Orange, and protein then transferred onto a PVDF membrane and sequentially probed with antibodies specific for Ycf48 and the N-terminal part of D1. While Ycf48 remained stable, D1 was cleaved at the Arg8 and Lys238 residues exposed to the cytoplasm resulting in the formation of a slightly shorter D1 and its 20kDa N-terminal fragment (D1-Nfr) (11)

Fig. S3 Site of insertions found in some eukaryotic Ycf48. (A) The top to bottom view of Ycf48 from *C. merolae* (PDB ID: 5OJ3), with the eukaryotic insert marked in purple and the quadruple 'Arg patch' marked in blue found on the surface designated here as top. Insert V160-E163 is relatively conserved in eukaryotic species and is found on the top surface of blade 2 (orange). Insert N322-A324 which is exclusive to *C. merolae* is present on the bottom surface of blade 5, close to the central pore (red). Insert N345-S348, which is underneath the 'Arg patch' of *C. merolae* is replaced with longer inserts in some cyanobacteria (e.g. *Synechococcus elongatus* PCC 7942) and possibly by splicing variants in moss (*Physcomitrella patens*) and several rice species (e.g. *Oryza sativa*) (cyan). Ycf48 in *Auxenochlorella protothecoides* has a 9-residue insert between S176 and G177 on blade 2 (yellow), *Corchorus olitorius* contains an insert between D374 and R375 (green) and Ycf48 in *Chlorella variabillis* has a 3-residue insert between S176 and G177, and an extended C-terminal tail (brown). (B) Sequence alignment of Ycf48 from 8 representative species, the alignment displayed begins from the first universally conserved residue W109 (*C. merolae* numbering), the inserts are color coded as panel A; most inserts are found in the loop regions.

Fig. S4 Characterization of the Ycf48-CM and Ycf48-RAquad mutants. (A) Cell cultures were spotted on BG-11 plates and grown at low irradiance (40-50 µE m^{-2} s^{-1}, top row), semi-high irradiance (70-90 µE m^{-2} s^{-1}, middle row) and high irradiance (90-130 µE m^{-2} s^{-1}, bottom row). Ycf48-WT is a wild-type control strain carrying the gentamycin-resistance cassette downstream of Ycf48; ΔYcf48 is a Ycf48 knock-out mutant; Ycf48-CM is a strain expressing chimeric Ycf48 with a 19-residue eukaryotic loop from *C. merolae*; mutants Ycf48-RAquad and Ycf48-REquad express mutated forms of Ycf48 in which R196, R215, R219 and R220 are replaced by either alanine or glutamate, respectively. (B) Absorption spectra of cells with the same OD_{750nm}=0.1. (C) Membranes isolated from wild-type (WT) and Ycf48-CM were analyzed by 2D-Clear Native/SDS-PAGE electrophoresis (2D-CN/SDS-PAGE), the gels were stained by SYPRO Orange, photographed, blotted to a PVDF membrane and Ycf48 and D1 proteins were consecutively detected by specific antibodies (segments of the blot containing individual antibody signals are shown). Designation of complexes as in Fig. 1.

Fig. S5 View of the 'Arg patch' in Ycf48 from *T. elongatus*. Ribbon diagram of Ycf48 with gray transparent surface, highlighting the side chain positioning (sticks) of the four arginines in the 'Arg patch' found in PDB ID: 2XBG (Fig 4). The side chains of R202 (blue) and R226 (red) point outward from the structure, forming positively charged apices on the surface, whereas, R221 (orange) and R225 (green)

are less prominent. The two conformations of the side chain of R220 modelled into the crystal structure are shown.

Fig. S6 Co-crystallization of Ycf48 and D1 C-terminal peptides. (A) Co-crystal of Ycf48 with a synthetic mD1 peptide consisting of 10 residues (NAHNFPLDLA) corresponding to residues 335-344 of the mature D1 protein. Peptide binding gave a new crystal form that diffracted to atomic resolution, 1.1 Å (PDB ID: 5OF5). The peptide-bound YCF48 was very similar to the unbound form (0.77 Å RMSD) with the peptide bound on the opposite side of the protein to the Arg patch. D1-Leu341 of the peptide (in yellow) fits into a hydrophobic pocket on the side of the torus between blades one and two. The synthetic peptide, illustrated in stick form, adopts a similar extended conformation to the equivalent region of the D1 protein in PSII. The charge distribution on Ycf48 is represented by blue (positive) to red (negative). (B and C) Co-crystal of Ycf48 with a synthetic iD1 peptide consisting of 18 residues (NAHNFPLDLASAESAPVA) corresponding to residues 335-352 of the intermediate precursor D1 protein (26). Four copies of YCF48 were found in the asymmetric unit, which diffracted to 2.0 Å (PDB ID: 5OJR), with two iD1 peptides wrapped around two of the four Ycf48. The peptides adopt similar conformations from residues 335-345, but differ from 346-352 (B, C). The binding site for the iD1 peptide shown in (C) spans along a relatively conserved channel within Ycf48, and overlaps the binding position of mD1 in (A) but the peptides are out of register. Amino acid conservation was scored using ConSurf server, and represented with purple (conserved) to cyan (variable) scheme.

Fig. S7 Vectors and strains used in this study. (A) The N-terminal region of the Ycf48 proteins expressed in *E. coli* for crystallography and Ycf48 antibody production. (B) Schematic representation of the plasmid constructs used to knockout and modify Ycf48 in *Synechocystis*. (C) PCR analysis of genomic DNA extracted from *Synechocystis* mutants confirming segregation; Ycf48 locus was amplified with primer set Ycf48-1F and Ycf48-seg-R (left panel) and Ycf48-seg-F and Ycf48-seg-R (right panel). Ycf48-1F is 445 bp upstream of the start codon of *slr2034*, whereas Ycf48-seg-F is 222 bp upstream, and Ycf48-seg-R is immediately downstream of the STOP codon of *slr2034*. The single Ycf48 R->E mutants are designated as Ycf48-RxxxE, where the number represents the mutated residue number. The triple Ycf48 R->E mutants are designated as Ycf48-ExxxRtrip, where the number represents the Arg residue not been mutated to Glu. ΔYcf48 is the Ycf48 knockout mutant constructed by transforming WT strain with pYcf48kan, and Ycf48-WT is a wild type-like mutant constructed by transforming WT strain with pYcf48WTgen. (D) Thylakoid proteins (equivalent to 1 µg chlorophyll) isolated from the Ycf48 mutants were separated by SDS-PAGE and the gel was either stained by Coomassie Brilliant Blue (CBB) or blotted onto a PVDF membrane and sequentially probed with the antibody specific for the Ycf48 protein (segments of the blot with the specific antibody signal are shown). The four arginines were replaced by alanine in Ycf4-RAquad and glutamate in Ycf48-REquad. Mutant Ycf48-CM expresses a chimeric Ycf48 that has the 19 residue eukaryotic insert from *C. merolae* inserted into the homologous position.

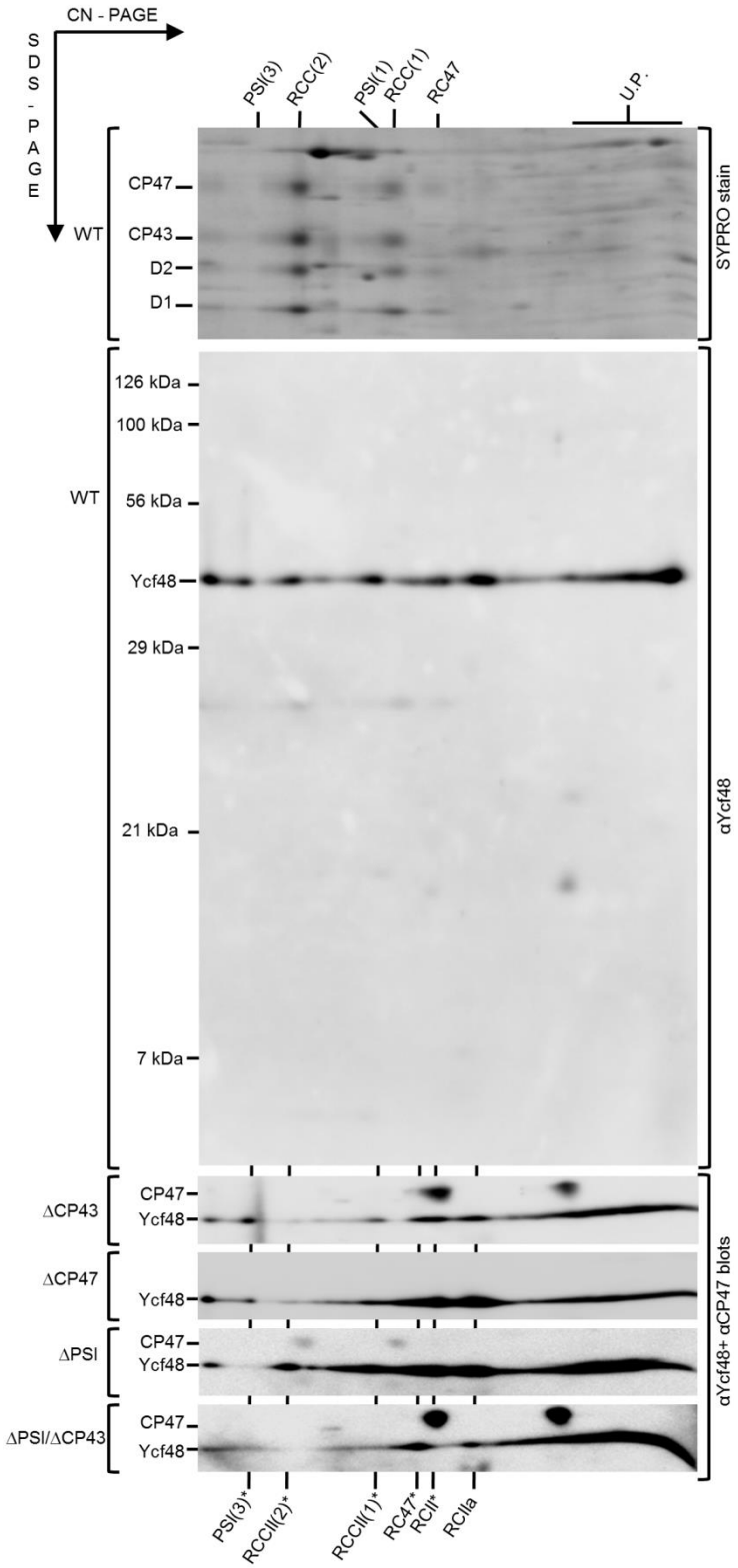

Fig. S1 Presence of Ycf48-containing complexes in membranes of *Synechocystis* strains lacking various chlorophyll-binding proteins.

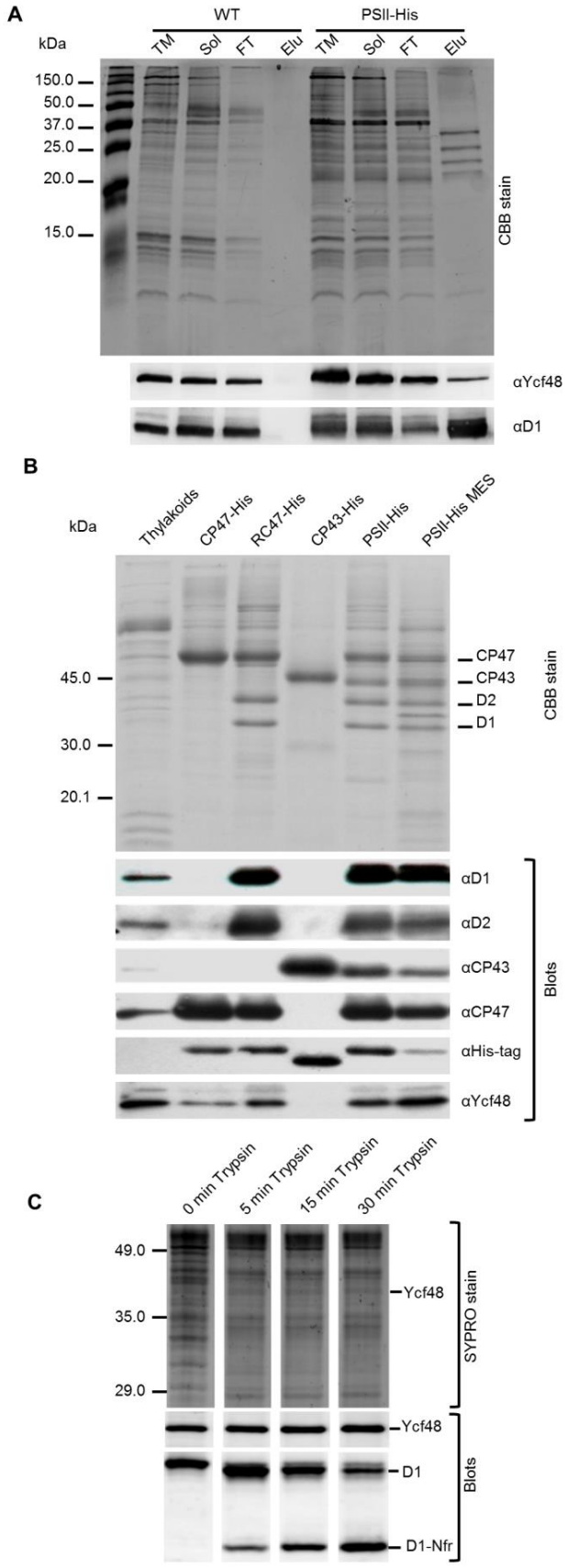

Fig. S2 Presence of Ycf48 in various PSII complexes isolated using His-tagged CP43 and CP47 and evidence for lumenal location of Ycf48 in *Synechocystis.*

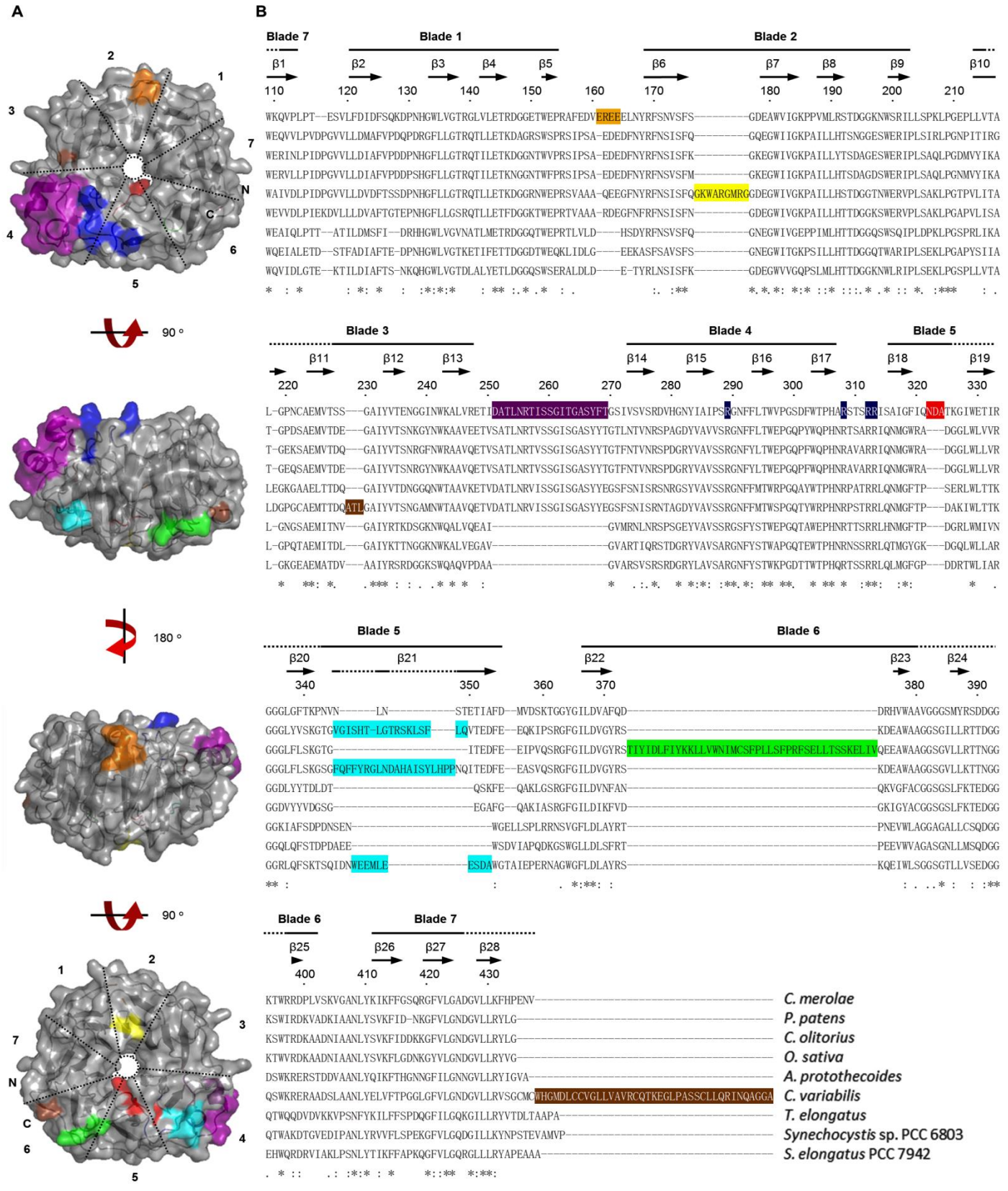

Fig. S3 Site of insertions found in some eukaryotic Ycf48.

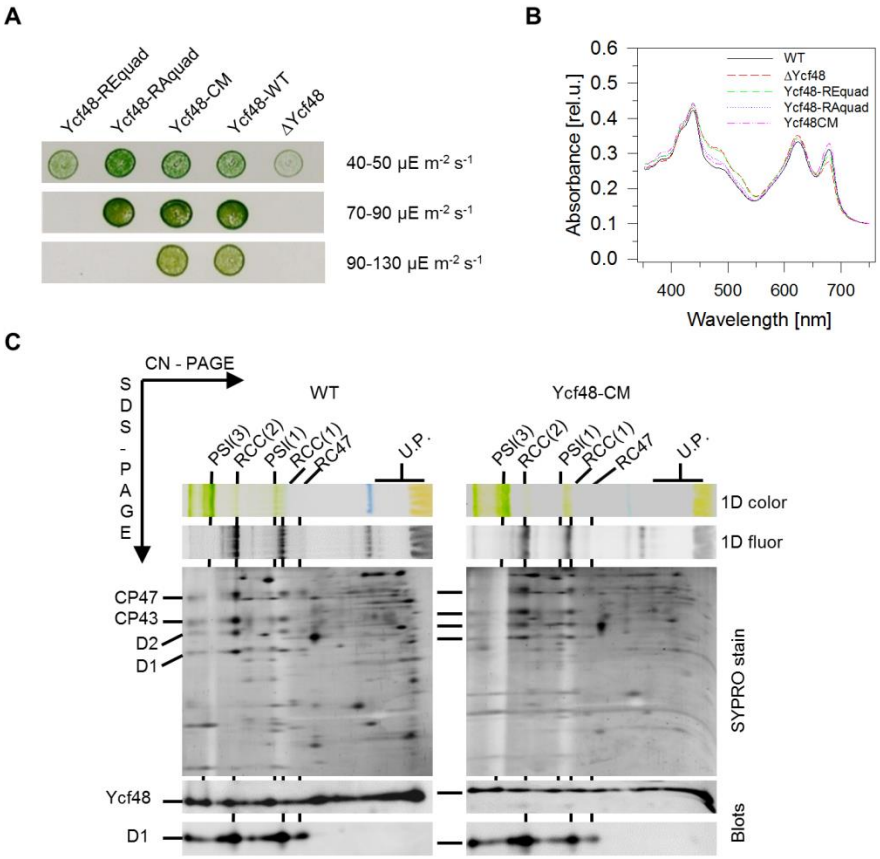

Fig. S4 Characterization of the Ycf48-CM and Ycf48-RAquad mutants.

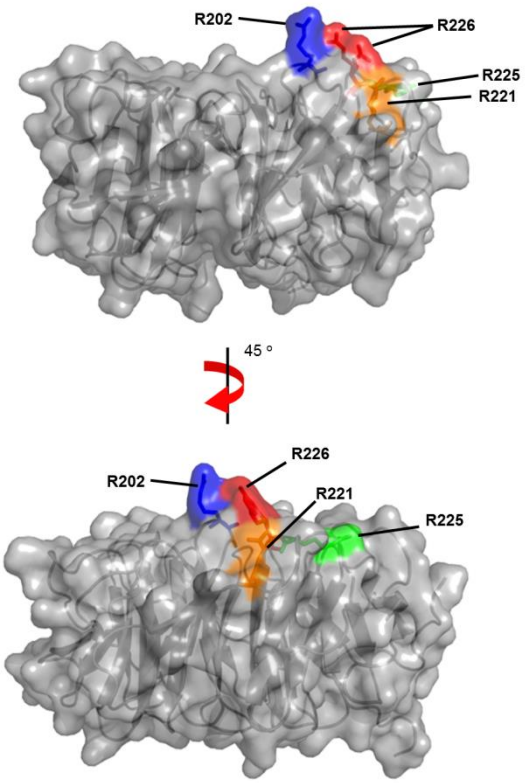

Fig. S5 View of the ‘Arg patch’ in Ycf48 from *T. elongatus*.

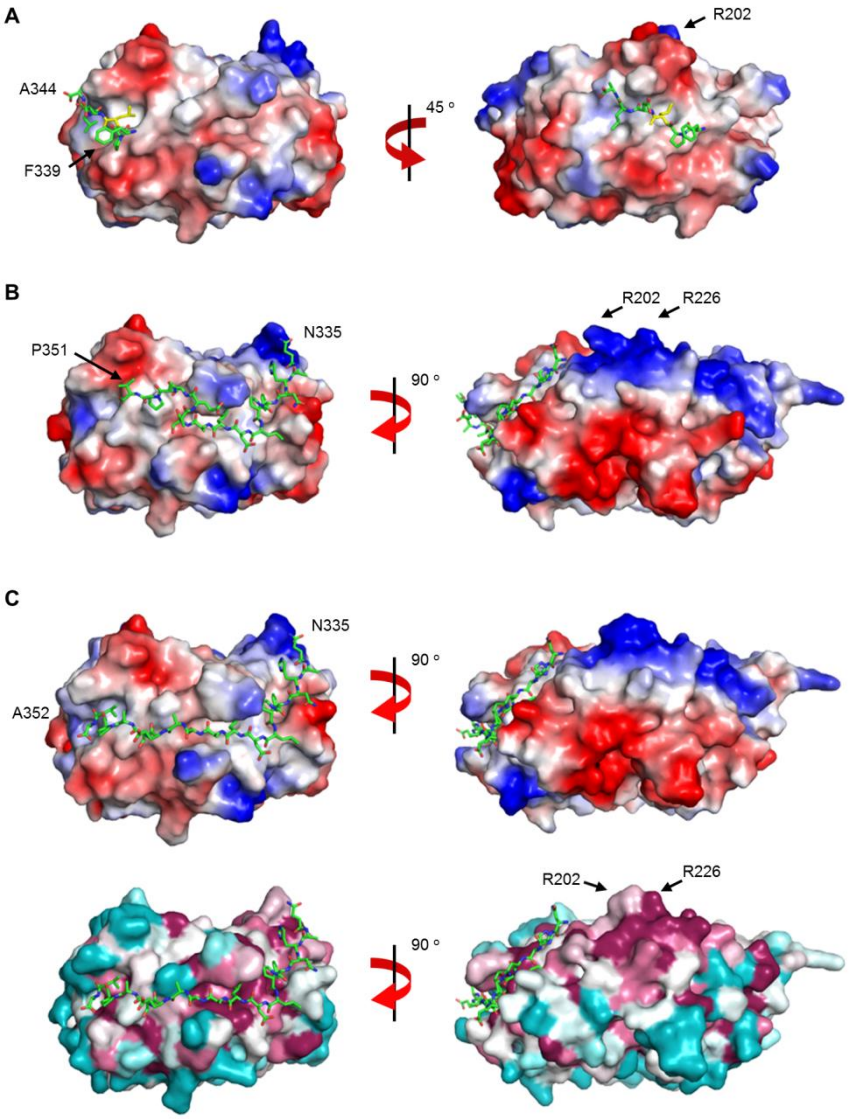

Fig. S6 Co-crystallization of Ycf48 and D1 C-terminal peptides

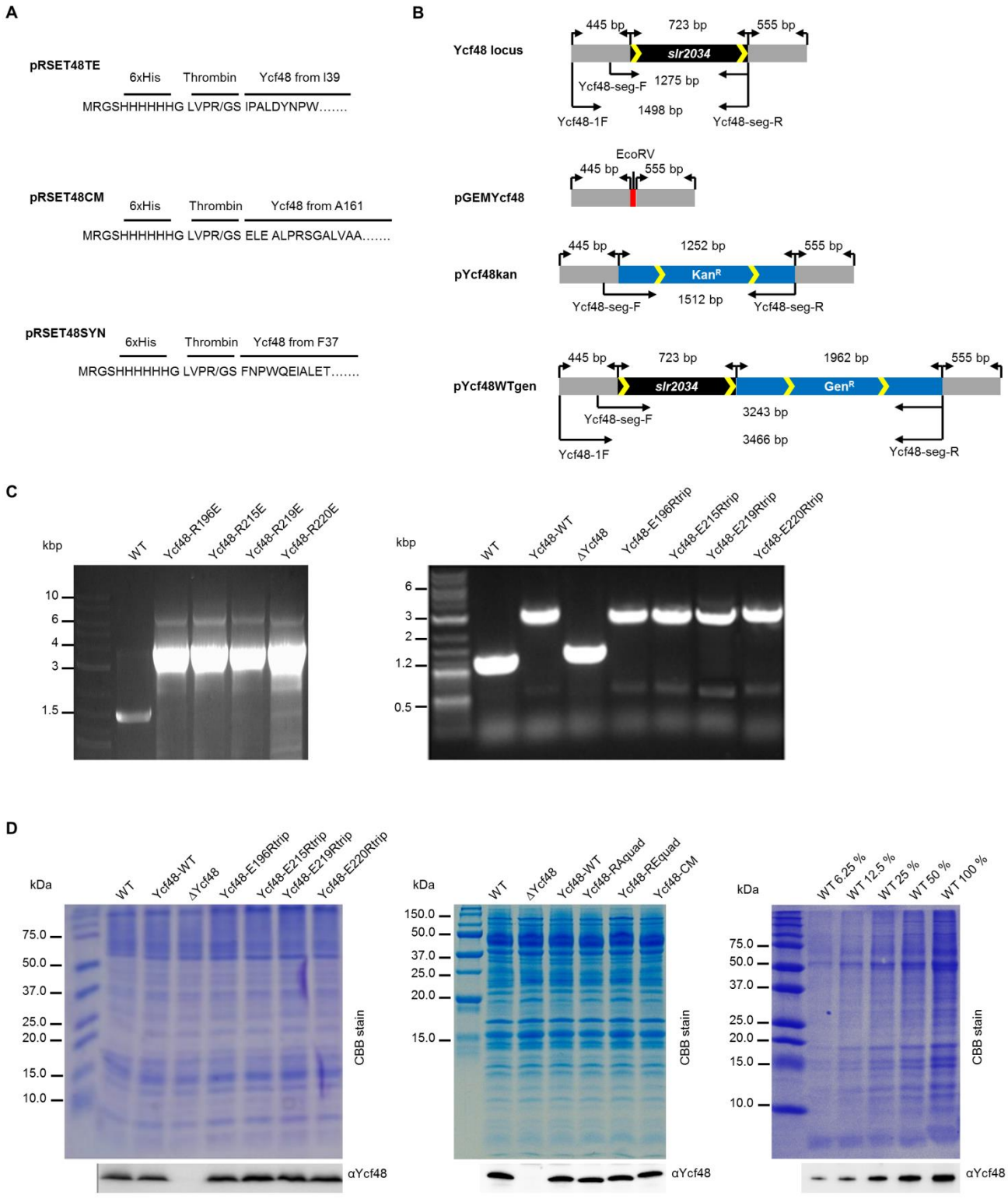

Fig. S7 Vectors and strains used in this study.

Table S1 Data collection and refinement statistics for Ycf48 structures. Values in brackets refer to the high resolution shell.

PDB	2XBG	5OJ3	5OF5	5OJR	5OJP
Structure	YCF48 *T. elongatus*	YCF48 *C. merolae*	YCF48 mD1 peptide	YCF48 iD1 peptide	Ycf48-REquad
Crystallization condition	0.1 M Magnesium sulfate, 10% w/v PEG 3350	10% w/v PEG 1000, 10% w/v PEG 8000	0.2 M Ammonium tartrate dibasic, PEG 3350 20% w/v pH=6.6	0.1 M Sodium citrate tribasic pH=5.6	0.2 M Sodium formate, 20% w/v PEG 3350, pH=7.2
Beamline	Diamond-I03	Soleil-Proxima 1	Soleil-Proxima 1	Diamond-I04	Diamond-I03
Wavelength (Å)	0.9184	1.12713	0.91840	0.9795	1.21484
Spacegroup	$C222_1$	$P2_12_12_1$	C2	$P2_1$	P1
Unit cell dimensions a,b,c (Å) α, β, γ (º)	55.726 75.676 148.549 90.000 90.000 90.000	60.780 97.060 110.560 90.000 90.000 90.000	98.730 68.650 56.190 90.000 105.900 90.000	97.320 60.460 103.460 90.000 96.780 90.000	54.441 60.286 75.398 68.435 74.828 85.428
Resolution	44.87-1.50 (1.58-1.50)	37.69-2.98 (3.03-2.98)	47.48-1.08 (1.10-1.08)	74.95-1.96 (1.99-1.96)	56.06-1.86 (1.91-1.86)
Total number of reflections	213013 (26735)	54330 (2768)	545015 (24280)	321472 (14125)	227562 (11690)
Number of unique reflections	50546 (7318)	13864 (688)	149739 (7286)	85525 (4040)	68907 (4141)
Completeness (%)	99.8 (99.9)	99.7 (99.7)	97.2 (95.0)	99.3 (99.5)	95.3 (77.5)
Multiplicity	4.2 (3.7)	3.9 (4.0)	3.6 (3.3)	3.8 (3.5)	3.3 (2.8)
<I/sigmaI>	11.9 (3.7)	10.6 (1.6)	10.7 (1.8)	8.1 (1.2)	8.2 (1.4)
Rmerge	0.068 (0.292)	0.128 (0.951)	0.071 (0.756)	0.110 (0.900)	0.056 (0.810)
Refinement					
Programme	REFMAC 5.5	PHENIX	PHENIX	PHENIX	PHENIX
Rcryst (%)	16.36	22.30	14.38	19.29	23.59
Rfree (%)	21.64	28.01	16.22	23.81	25.25
RMS					
Bonds (Å)	0.012	0.004	0.014	0.010	0.014
Angles (º)	1.1	0.942	1.32	1.109	1.371
Ramachandran plot (molprobity)					
Most favoured (%)	98.4	96.0	99.4	98.39	99.22
Outliers (%)	0.00	0.00	0.00	0.00	0.00

SI References

1. Tichý M, et al. (2016) Strain of *Synechocystis* PCC 6803 with aberrant assembly of photosystem II contains tandem duplication of a large chromosomal region. *Front Plant Sci* 7:648.
2. Vermaas WFJ, Ikeuchi M, Inoue Y (1988) Protein composition of the photosystem II core complex in genetically engineered mutants of the cyanobacterium *Synechocystis* sp. PCC 6803. *Photosynth Res* 113:97–113.
3. Eaton-Rye JJ, Vermaas WFJ (1991) Oligonucleotide-directed mutagenesis of *psbB*, the gene encoding CP47, employing a deletion mutant strain of the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Mol Biol* 17(6):1165–1177.
4. Shen G, Boussiba S, Vermaas WF (1993) *Synechocystis* sp PCC 6803 strains lacking photosystem I and phycobilisome function. *Plant Cell* 5(12):1853–1863.
5. Komenda J, et al. (2012) The Psb27 assembly factor binds to the CP43 complex of photosystem II in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol* 158(1):476–486.
6. Boehm M, et al. (2011) Investigating the early stages of photosystem II assembly in *Synechocystis* sp. PCC 6803: isolation of CP47 and CP43 complexes. *J Biol Chem* 286(17):14812–14819.
7. Nakamura Y, et al. (2002) Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. *DNA Res* 9(4):123–130.
8. Komenda J, et al. (2008) The cyanobacterial homologue of HCF136/YCF48 is a component of an early photosystem II assembly complex and is important for both the efficient assembly and repair of photosystem II in *Synechocystis* sp. PCC 6803. *J Biol Chem* 283(33):22390–22399.
9. Porra RJ, Thompson WA, Kriedemann PE (1989) Determination of accurate extinction coefficients and simultaneous-equations for assaying chlorophyll a and chlorophyll b extracted with 4 different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochim Biophys Acta* 975(3):384–394.
10. Pilný J, Kopečná J, Noda J, Sobotka R (2015) Detection and quantification of heme and chlorophyll precursors using a high performance liquid chromatography (HPLC) system equipped with two fluorescence detectors. *Bio-Protocol* 5(3):e1390.
11. Komenda J (2000) Role of two forms of the D1 protein in the recovery from photoinhibition of photosystem II in the cyanobacterium *Synechococcus* PCC 7942. *Biochim Biophys Acta - Bioenerg* 1457(3):243–252.
12. Chidgey JW, et al. (2014) A cyanobacterial chlorophyll synthase-HliD complex associates

with the Ycf39 protein and the YidC/Alb3 insertase. *Plant Cell* 26(3):1267–79.

13. Strašková A, Knoppová J, Komenda J (2018) Isolation of the cyanobacterial YFP-tagged photosystem I using *GFP-Trap*®. *Photosynthetica* 56:300–305.
14. Wittig I, Karas M, Schägger H (2007) High resolution clear native electrophoresis for in-gel functional assays and fluorescence studies of membrane protein complexes. *Mol Cell Proteomics* 6(7):1215–1225.
15. Schägger H, von Jagow G (1991) Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal Biochem* 199(2):223–231.
16. Dobáková M, Sobotka R, Tichý M, Komenda J (2008) Psb28 protein is involved in the biogenesis of the photosystem II inner antenna CP47 (PsbB) in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol* 149(2):1076–1086.
17. Michoux F, Takasaka K, Boehm M, Nixon PJ, Murray JW (2010) Structure of CyanoP at 2.8 Å: implications for the evolution and function of the PsbP subunit of photosystem II. *Biochemistry* 49(35):7411–7413.
18. Boehm M, et al. (2012) Subunit organization of a *Synechocystis* hetero-oligomeric thylakoid FtsH complex involved in photosystem II repair. *Plant Cell* 24(9):3669–3683.
19. Knoppová J, et al. (2014) Discovery of a chlorophyll binding protein complex involved in the early steps of photosystem II assembly in *Synechocystis*. *Plant Cell* 26(3):1200–1212.
20. Michoux F, et al. (2012) Crystal structure of the Psb27 assembly factor at 1.6 Å: implications for binding to photosystem II. *Photosynth Res* 110(3):169–175.
21. Pape T, Schneider TR (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J Appl Crystallogr* 37(5):843–844.
22. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr Sect D Biol Crystallogr* 66(4):486–501.
23. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(4):658–674.
24. Winter G (2010) *xia2*: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* 43(1):186–190.
25. Kabsch W (2010) *XDS*. *Acta Crystallogr Sect D Biol Crystallogr* 66(2):125–132.
26. Komenda J, et al. (2007) Cleavage after residue Ala352 in the C-terminal extension is an early step in the maturation of the D1 subunit of photosystem II in *Synechocystis* PCC 6803. *Biochim Biophys Acta - Bioenerg* 1767(6):829–837.