Supplementary Material for

Plant Rubisco assembly in *E. coli* with five chloroplast chaperones including BSD2

H. Aigner, R. H. Wilson, A. Bracher, L. Calisse, J. Y. Bhat, F. U. Hartl, M. Hayer-Hartl*

*Corresponding author. Email: M. Hayer-Hartl, mhartl@biochem.mpg.de

This PDF file includes:

Materials and Methods

Plasmids and strains

All plasmids used in this study and the N-terminal sequences of the recombinantly expressed proteins are listed in tables S2 and S3, respectively.

Plasmid p*At*RbcLS contains the *rbcL* (AtCg00490) and *rbcS3B* (At5g38410) genes under the control of an araBAD promoter. To construct this plasmid, a 3400 bp segment of the pBAD33 vector containing an araBAD promoter, an araC gene, a p15A ori, and a chloramphenicol resistance gene, was amplified by PCR. This segment was ligated with an RBS-*At*RbcS3B cassette to result in the plasmid p*At*RbcS. The RBS-*At*RbcS3B cassette was obtained from a previously constructed *At*RbcS3B-pET11a vector by cutting with XbaI and PstI. During the assembly of the construct, the transit peptide of RbcS was removed. The RBS-*At*RbcL cassette of an *At*RbcL-pET11a vector was amplified by PCR and ligated into the pRbcS plasmid. KpnI and NheI sites were used to excise the RBS-*At*RbcL PCR-product, the p*At*RbcS plasmid was cut with KpnI and XbaI.

Plasmid p*At*C60αβ/C20 contains the sequences of the *A. thaliana* chloroplast chaperonins Cpn60α1 (At2g28000), Cpn60β1 (At1g55490), and Cpn20 (At5g20720). As for p*At*RbcLS, each sequence was cloned, without the chloroplast transit peptide (prediction based on peptide homology and using ChloroP) (*43*), into a pET11a vector (plasmids p*At*C60α, p*At*C60β, and p*At*C20). Next, the RBS-*At*C60β1 cassette was excised with XbaI and BamHI from p*At*C60β and ligated into the p*At*C60α plasmid, which was cut by NheI and BamHI. This resulted in plasmid p*At*C60αβ. Plasmid p*At*C60αβ and the RBS-*At*Cpn20 cassette were amplified by PCR to add the restriction sites SalI and KpnI. Ligation of these fragments resulted in plasmid p*At*C60αβ/C20. Deletions of single chaperonin subunit sequences were obtained by site-directed mutagenesis that led to premature stop codons. Replacement of *At*Cpn20 was done by digestion of the plasmid

p*At*C60αβ/C20 with SalI and EcoRI and insertion of either *At*Cpn10 (At2g44650), resulting in

plasmid p*At*C60αβ/C10, or GroES, resulting in plasmid p*At*C60αβ/*Ec*ES. Plasmid

p*At*C60αβ/C20/C10 was generated by insertion of *At*Cpn10 after the *At*Cpn20 sequence by using

the KpnI and EcoRI restriction sites.

Plasmid p*At*R1/R2/X/B2 contains the sequences of *At*Raf1.2 (At3g04550), *At*Raf2

(At5g51110), *At*RbcX2 (At5g19855), and *At*BSD2 (At3g47650). As before, the coding sequences

(without chloroplast transit peptides) were ligated into pET11a vectors. The *At*Raf2, *At*RbcX, and

*At*BSD2 sequences were amplified by PCR so as to contain the T7 promoter, the lac operator, and

the T7 terminator of the pET11a vector. These sequences were cloned in a pTRC99a vector

containing a synthesized multiple cloning site, specifically designed for this purpose. The *At*Raf1

sequence was ligated directly into the first multiple cloning site of a pCDFDuet vector and the

insertion of the Raf2-RbcX-BSD2 cassette into pCDFDuet-*At*Raf1 resulted in plasmid

p*At*R1/R2/X/B2. Single deletions of the assembly chaperone sequences were made by PCR

amplification of the p*At*R1/R2/X/B2 plasmid omitting the sequences to be removed.

The plasmid pCDF-*Te*RbcL(IA)/*At*BSD2 contains the sequence of *Thermosynechococcus*

elongatus BP1 Rubisco (*Te*RbcL) featuring the amino acid substitutions F345I and P415A (IA),

which improve Rubisco yield when expressed in *E. coli* (*33, 34*). The plasmid was constructed

using Gibson cloning by amplifying the genes *[his6]Ubq-TerbcL(IA)* from the vector pHue-

*Te*RbcL(IA), *At*BSD2 from p*At*R1/R2/X/B2, and the intergenic region from pCDF-Duet. The

Gibson fragments were assembled into the target vector pCDF-Duet1 (Novagen) with each gene

partnered with one of the two co-expression promoters. Point mutations to generate the AtBSD2

mutants within p*At*R1/R2/X/B2 were performed using the QuikChange lightning site-directed

mutagenesis kit (Agilent), as per the manufacturer's recommendations.

Plasmid pHue-*At*BSD2 was created by insertion of the *At*BSD2 sequence (signal peptide removed) into the pHUE vector using the restriction sites SacII and SacI. Plasmids p*Nt*Rbc*LS* and p*Nt*R1/*At*R2/X/B2 were created by replacing the corresponding *A. thaliana* sequences with gBlocks Gene Fragments (IDT) in the respective *A. thaliana* plasmids.

E. coli growth

E. coli cells were grown in LB medium, which contained, depending on the plasmids used, the antibiotics ampicillin (200 µg mL^{-1}), chloramphenicol (32 µg mL^{-1}) and streptomycin (100 µg mL^{-1}). For expression of Rubisco, cells with all three plasmids were grown overnight at 37°C. 100 µL of this starter culture were used to inoculate 50 mL LB-medium. Once the cultures reached an OD$_{600}$ of 0.3 they were transferred to 23°C and IPTG (isopropyl β-D-thiogalactoside) was added to a final concentration of 0.3 mM. After 3h the medium was replaced with an equal volume of fresh LB-medium containing the antibiotics and 0.4 % arabinose. The cultures were allowed to grow for 18 h at 23°C, usually reaching a final OD$_{600}$ of ~6.

Antibodies

Rabbit antibodies against *A. thaliana* RbcS, BSD2 and RbcX were produced with standard procedures. Antibodies against *Synechococcus elongatus* PCC7942 Raf1 and *A. thaliana* Raf1.2 were previously described (*22*). Plant RbcL and Cpn60α antibodies (raised against the *N. tabacum* proteins) were kindly supplied by S. Whitney (Australian National University). Antibody against *C. reinhardtii* Raf2 was kindly supplied by S. Merchant (UCLA).

Rubisco expression analysis

Bacterial pellets containing induced protein were re-suspended in Rubisco extraction buffer (50 mM Tris-HCl pH 8.0, 10 mM MgCl$_2$, 1 mM EDTA, 2 mM DTT) containing 0.4 mg mL^{-1} lysozyme, 5 U mL^{-1} benzonase (Sigma), and protease inhibitor cocktail (Roche). Plant tissue was extracted in Rubisco extraction buffer containing plant protease inhibitor cocktail (Sigma) and 1 % w/v polyvinylpolypyrrolidone (Sigma). Bacterial cells were lysed by sonication and plant tissue was lysed using a glass homogenizer. Samples of total cellular protein (lysate) and the soluble protein fractions, prepared by centrifugation for 10 min at 24,000 x g at 4°C, were analyzed by native-PAGE using 4-12 % Tris-glycine gradient gels (Thermo Fisher) and SDS-PAGE using 4-12 % Bis-Tris gradient gels (Thermo Fisher), unless otherwise specified. Immunoblotting against AtRbcS, AtRaf1, AtRbcX, and AtBSD2 was performed using primary antibodies.

Assembled Rubisco was quantified by stoichiometric binding of the essentially irreversible (K$_d$ = 10^{-11} M) active site inhibitor (2-^{14}C)2'-carboxyarabinitol-1,5-bisphosphate (^{14}C-CABP) after pre-incubation of soluble lysates in extract buffer (20 mM Tris-HCl pH 8.0, 10 mM MgCl$_2$, 1 mM EDTA) with 20 mM NaHCO$_3$ and 10 mM MgCl$_2$, as described previously (*33*). Total soluble protein was quantified using the Coomassie dye-binding assay (BioRad) relative to a BSA standard (Pierce).

Protein purification

E. coli containing the three Rubisco expression/assembly vectors (Fig. 1A) were grown, induced, and lysed as described above. AtRbcL$_8$S$_8$ was purified by subjecting the soluble protein to ion-exchange chromatography on a DAEA column (self-packed) equilibrated with buffer A (50 mM Tris-HCl pH 8.0, 10 mM KCl). Proteins were eluted with a linear KCl gradient from 0 to 1 M. Fractions were analyzed for Rubisco by SDS-PAGE. Fractions most enriched in RbcL$_8$S$_8$ were

pooled and desalted using a HiPrep 26/10 column (GE Healthcare). The resulting protein solution was applied onto an ENRichQ 100/10 (Bio-Rad) ion-exchange column and again eluted with a linear salt gradient as above. The Rubisco-containing fractions were applied to a Superdex 200 gel filtration column equilibrated in buffer 50 mM Tris-HCl pH 8.0, 50 mM NaCl. Fractions containing Rubisco were pooled and concentrated using a centrifugal filter (Amicon MWCO 100 kDa). Concentrated Rubisco was then either directly analyzed or dialyzed against storage buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 5 % v/v glycerol), frozen in liquid nitrogen and stored at −80°C.

To purify *At*BSD2, *E. coli* containing the plasmid pHue-*At*BSD2 were grown, induced and lysed as described above. *At*BSD2 was purified using the pHue expression system as described previously (*22*). The pellet of a 3 L overnight culture was dissolved in 80 mL lysis buffer (50 mM Tris-HCl pH 8.0, NaCl 300 mM, 2 mM DTT, 5 U mL^{-1} benzonase (Sigma), protease inhibitor cocktail (Roche)) and lysed by sonication. The lysate was clarified by ultracentrifugation and purified by Ni-affinity chromatography. The His-ubiquitin tag was cleaved by overnight incubation with the deubiquitinating enzyme Usp2 (*44*) and buffer exchanged to 50 mM Tris-HCl pH 8.0, 50 mM NaCl, before removal of the His-ubiquitin tag by Ni-affinity chromatography. The untagged BSD2 eluted in the flow through and was concentrated to ~500 µM (Amicon MWCO 10 kDa), frozen in liquid nitrogen and stored at −80°C.

Purification of *Te*RbcL(IA)$_8$:*At*BSD2$_8$ complexes, *Se*RbcL$_8$, *Se*RbcS, and *At*RbcS was performed as previously described (*12, 22*). All purification steps were performed at 4°C, and protein concentrations were determined spectroscopically at 280 nm.

Rubisco enzyme kinetics

Maximal Rubisco carboxylation rates (V_c^{max}) and affinity for substrate CO_2 in air (K_c^{air}) were measured by $^{14}CO_2$ fixation assays. *E. coli* cell and *A. thaliana* tissue lysates were analyzed in 1.5 mL microcentrifuge tubes using reaction buffer (50 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 1 mM EDTA, 330 µM RuBP (Sigma)) purged with CO_2 free air and containing varying concentrations of $NaH^{14}CO_3$ laced $NaHCO_3$ (~0.5–20 mM $NaHCO_3$ corresponding to ~16–190 µM CO_2). Assays of 0.3 mL total volume were started by the addition of 20 µL soluble protein lysate mixed 1:1 in activation buffer (20 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 1 mM EDTA, 40 mM $NaHCO_3$) and incubated for 1 min at 25°C. Reactions were terminated by the addition of 100 µL of 50 % (v/v) formic acid. Dried reactions were re-dissolved in 500 µL H_2O and mixed with scintillant for counting. Rubisco activity in the soluble protein fraction (k_{cat}^C) was measured similarly with the exception that CO_2 concentration was fixed at ~350 µM (~27 mM $NaHCO_3$) and incubated for 1 or 3 min. Rubisco kinetics were calculated using the specific activity determined by the complete conversion of 33 µM RuBP (Sigma) (9.9 nmol per assay) in reactions incubated with *A. thaliana* tissue extracts for ~30 min and Rubisco active site content determined by [^{14}C]-CABP binding.

Mass spectrometry

Native mass spectrometry (native-MS): Purified *At*BSD2 (11 µM) and *Se*RbcL8 (30 µM) and the binding reactions (15 min at 25°C) of *At*BSD2 to *Se*RbcL8 in buffer (50 mM Tris-HCl pH 8.0, 50 mM NaCl) were buffer-exchanged into 100 mM ammonium acetate, pH 8.0 (Fluka, Sigma) using Micro Bio-Spin 6 chromatography columns (Bio-Rad). Native-MS analyses were performed on a quadrupole IM time-of-flight hybrid mass spectrometer with a Z-spray nano-ESI source (Synapt G2-Si, Waters Corp., USA) in positive ion mode. Gold-plated 10-µm nano-ESI pipettes

(Mascom, Bremen) were used as capillaries. Optimized capillary and sample cone voltages were 1.2-1.7 kV and 130-150 V, respectively. Spectra were acquired and processed in MassLynx 4.1 (Waters Corp., USA). Spectra were calibrated with 30 mg mL^{-1} cesium iodide dissolved in 1:1 acetonitrile/water.

Proteomic analysis: Soluble cell lysate was methanol-precipitated, followed by reduction, alkylation, tryptic digestion and desalting (*45, 46*).Tryptic peptides were dissolved in 5% formic acid and analyzed by nano LC-MS/MS using an EASY-nLC 1000 nano liquid chromatography system (Thermo) coupled to a Q-Exactive mass spectrometer (Thermo). Samples were injected onto a home-made 25 cm silica reversed-phase capillary column (New Objective) packed with 1.9-μm ReproSil-Pur C18-AQ (Dr. Maisch GmbH). Samples were loaded on the column by the nLC autosampler at a flow rate of 0.5 μl min^{-1}, without a trap column. Peptides were separated by a stepwise 120 min gradient of 0–95% acetonitrile (0.2% formic acid in water to 0.2% formic acid in 95% acetonitrile) at a flow rate of 250 nL min^{-1}. MS/MS analysis was performed with standard settings using cycles of 1 high resolution (70,000 full width at half maximum (FWHM) setting) MS scan followed by MS/MS scans of the 10 most intense ions with charge states of 2 or higher at a resolution of 17,500 FWHM. Protein identification was performed with MaxQuant (version 1.3.0.5) using default settings (*47*).

Top-down intact mass determination: Purified Rubisco (10 μg) in buffer was desalted on a home-made porous column followed by direct elution into a Synapt G2-Si mass spectrometer (Waters Corp., USA) through an ESI source. The mass spectra were deconvoluted using MassLynx 4.1 (Waters Corp., USA).

Crystallization

*At*BSD2 was crystallized at 18°C using the sitting-drop vapor diffusion method by equilibrating 100 nL *At*BSD2 at ~12 mg mL^{-1} in buffer 10 mM Tris-HCl pH 8.0, 50 mM NaCl mixed with 100 nL reservoir containing 20 % PEG 10,000, 8 % ethylene glycol and 0.1 M HEPES-NaOH pH 7.5 (Qiagen Classics screen, condition 95) against reservoir. The crystals grew within two weeks after initial precipitation. For cryo-cooling, the crystals were transferred in two steps into 22 % PEG 10,000, 15 % ethylene glycol and 0.1 M HEPES-NaOH pH 7.5 and then dipped into liquid nitrogen.

The *Te*RbcL(IA)$_8$:*At*BSD2$_8$ complex was crystallized at 18°C using the sitting-drop vapor diffusion method by equilibrating 200 nL of complex at 34 mg mL^{-1} in buffer (20 mM Tris, 50 mM NaCl, pH 8.0) mixed with 200 nL reservoir containing 21 % PEG 3350 and 0.12 M DL-malic acid pH 7.0 against reservoir. The crystals grew within 3 days after initial precipitation. For cryo-cooling, the crystals were transferred in two steps into 25 % PEG 3350, 0.12 M DL-malic acid pH 7.0, and 10 % glycerol and then dipped into liquid nitrogen.

Data collection, structure solution and refinement

Diffraction data for *At*BSD2 crystals were collected at beamline X10SA of the Swiss Light Source (SLS) in Villigen, Switzerland. *At*BSD2 crystals exhibited strong X-ray fluorescence at the Zn L-I absorption edge, indicative of the presence of Zn. Diffraction data were integrated with XDS and further processed with POINTLESS (*48*), AIMLESS (*49*) and CTRUNCATE (*50*) as implemented in the CCP4i graphical user interface (*51*). The structure of *At*BSD2 was solved by Zn-MAD at 1.9 Å resolution using SHELXC/D/E (*52*) as implemented in HKL2MAP (*53*). Three Zn sites were identified. The third zinc forms a crystal contact between *At*BSD2 molecules (employing residues Glu78, Cys80 and Glu115 from one chain and His96 from an adjacent chain), which form spirals along the 4$_1$ screw axes in the crystal lattice. The experimental map

was readily interpretable. The program Buccaneer (*54*) was employed to auto-build a preliminary model, which was completed interactively using Coot (*55*). The structure was refined with REFMAC5 (*56*) using TLS parametrization of the temperature factors. According to MolProbity (*57*), the model exhibits excellent stereochemistry with 100 % of the residues in the favored regions of the Ramachandran plot.

Diffraction data for *Te*RbcL$_8$*At*BSD2$_8$ crystals were collected at beamline ID29 of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. Diffraction data were integrated with XDS and further processed with POINTLESS (*48*), AIMLESS (*49*) and CTRUNCATE (*50*) as implemented in the CCP4i graphical user interface (*51*). The structure was solved by molecular replacement using the program MOLREP (*58*) with the *Te*RbcL$_8$ core structure from pdb file 3ZXW (doi:10.2210/pdb3zxw/pdb) as a search template. The Zn sites of the bound *At*BSD2 were identified in a model-phased anomalous difference map calculated from highly redundant diffraction data collected at the Zn L-I edge (1.2824 Å wavelength, ESRF beamline ID29). *At*BSD2 was placed manually into the difference density calculated after an initial refinement run using REFMAC5 (*56*). The model was completed manually using Coot. REFMAC5 was used for model refinement. The asymmetric unit contains ten copies of the *Te*RbcL$_8$*At*BSD2$_8$ complex. Residues with disordered side chains, which are facing solvent channels, were modeled as alanines. According to MolProbity (*57*), the model exhibits reasonable stereochemistry with 95.8 % of the residues in the favored regions of the Ramachandran plot. Figures were created with PyMol (http://www.pymol.org/) and ESPript (*59*).

The crystallographic models and structure factors have been deposited to wwPDB under accession codes 6EKB (*At*BSD2) and 6EKC (*Te*RbcL$_8$*At*BSD2$_8$).
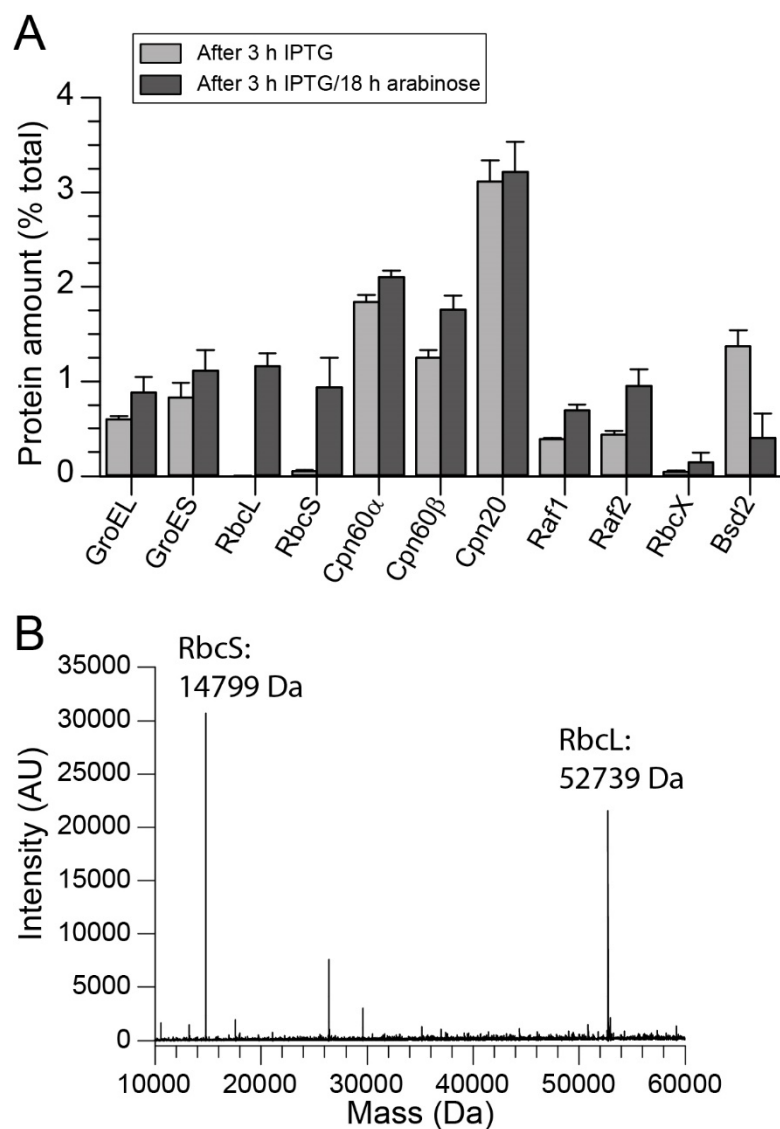
fig. S1

A

B

Fig. S1. **Plant Rubisco folding and assembly in *E. coli* requires co-expression of chloroplast chaperonin and Rubisco auxiliary factors.** (**A**) Total *E. coli* lysates after 3 h of expression of chaperonin and auxiliary factors with IPTG induction (light grey bars), followed by18 h expression of Rubisco with arabinose induction as in Fig. 1B (dark grey bars), were analysed by LC-MS/MS. Protein amounts based on iBAQ values are expressed in % of total lysate protein. Data are averages ± S.D. from three independent experiments. The iBAQ values are the

normalization of the sum of the observed peptide intensities by the number of theoretically observable peptides of a protein (*47*). (**B**) Top-down MS analysis of purified recombinant *At*Rubisco. The mass of RbcL was determined as 52739.00 Da (the theoretical mass of RbcL lacking Met1 and Ser2 with no additional modifications is 52736.63 Da) and that of RbcS at 14799.00 Da (theoretical mass 14797.00 Da).
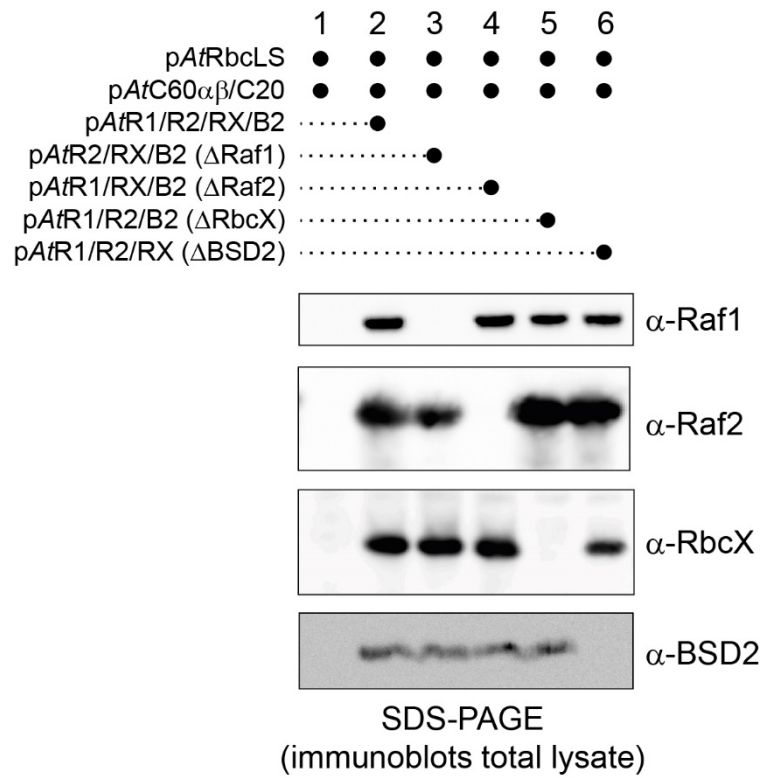
fig. S2

Fig. S2. **Chaperone dependence of *At*Rubisco assembly.** Immunoblot analysis of total lysates

of *E. coli* strains deleted in specific assembly chaperones from p*At*R1/R2/RX/B2 (see Fig. 3).

Raf1, Raf2, RbcX and BSD2 were detected by immunoblotting with specific antibodies.

fig. S3

A   native-PAGE immunoblot

T
L
RbcL$_8$S$_8$

α-RbcL   α-Raf1   α-Raf2   α-RbcX

B   1   2

pAtC60αβ/C20   ●   ●
pAtR1/R2/RX/B2   ●   ●
pAtRbcL (ΔRbcS)   ●   ●
pAtR1/RX/B2 (ΔRaf2)   ●

native-PAGE
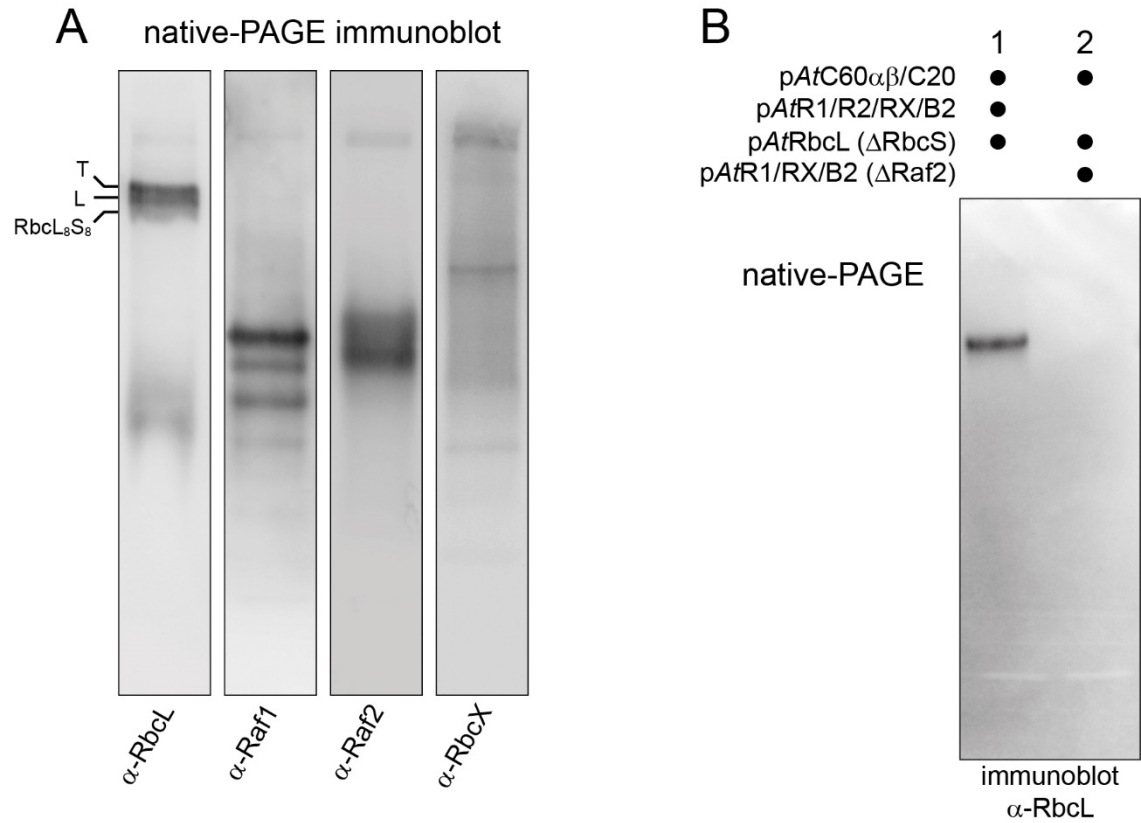
immunoblot
α-RbcL

Fig. S3. **Analysis of RbcL:BSD2 complex.** (**A**) Native-PAGE immunoblot analyses against RbcL, Raf1, Raf2 and RbcX of a Rubisco biosynthesis reaction at RbcS limiting conditions (see Fig. 4A and B, lane 1). (**B**) Analysis by native-PAGE and anti-RbcL immunoblotting of total lysates from *E. coli* cells expressing all *A. thaliana* auxiliary factors and RbcL but no RbcS (lane 1) or expressing the auxiliary factors Raf1, RbcX and BSD2 but no Raf2 and RbcS (lane 2).
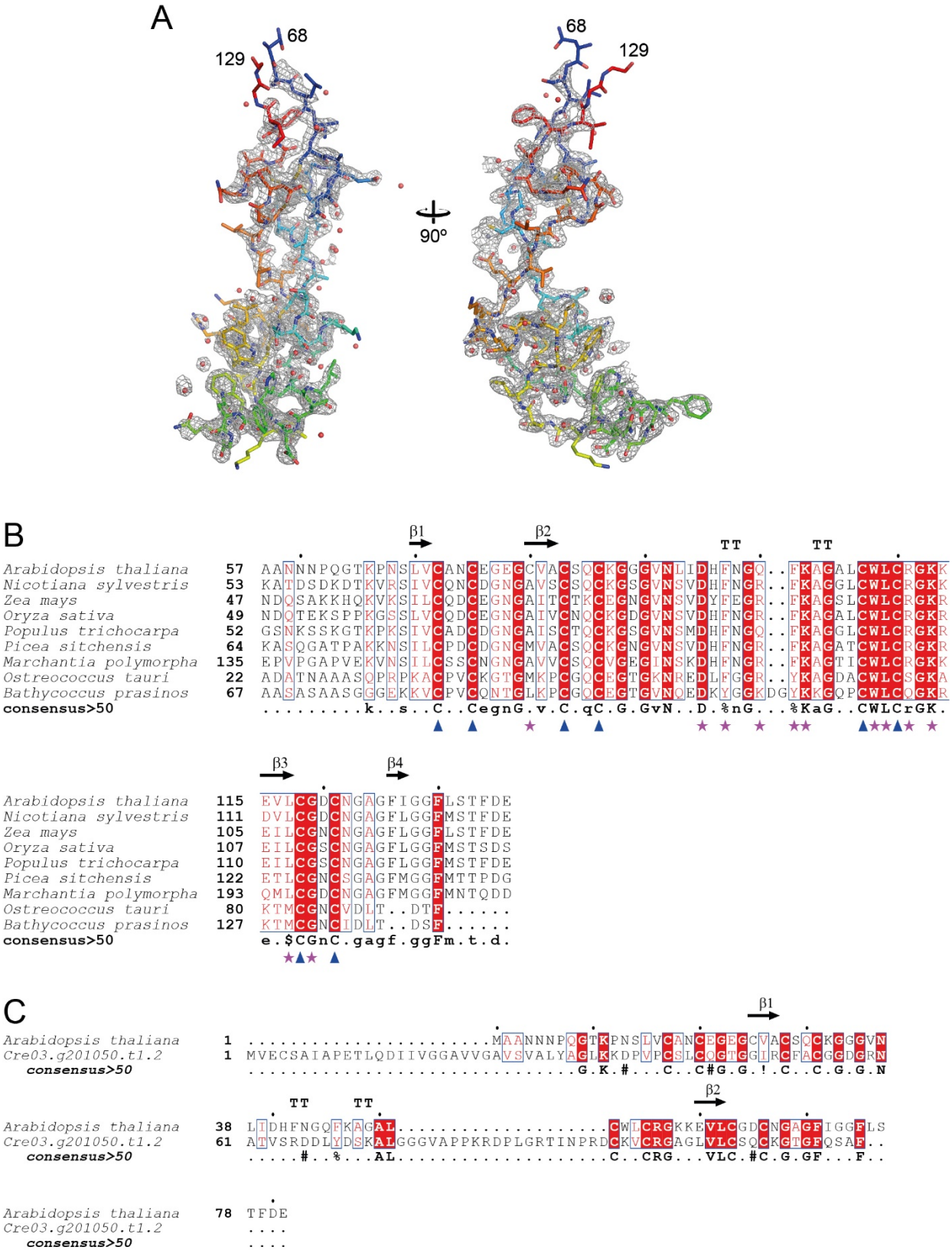
fig. S4

A

129   68   68   129

90°

B

```
                          β1              β2                          TT          TT
Arabidopsis thaliana   57  AANNNPQGTKPNSLVCANCEGEGCVACSQCKGGGVNLIDHFNGQ..FKAGALCWLCRGKK
Nicotiana sylvestris   53  KATDSDKDTKVRSIVCQNCDGNGAVSCSQCKGSGVNSVDHFNGR..FKAGGLCWLCRGKK
Zea mays               47  NDQSAKKHQKVKSILCQDCEGNGAITCTKCEGNGVNSVDYFEGR..FKAGSLCWLCRGKR
Oryza sativa           49  NDQTEKSPPKGSSLVCQDCEGNGAIVCNQCKGDGVNSVDHFNGR..FKAGALCWLCRGKR
Populus trichocarpa    52  GSNKSSKGTKPKSIVCADCDGNGAISCTQCKGSGVNSMDHFNGR..FKAGGLCWLCRGKR
Picea sitchensis       64  KASQGATPAKKNSIILCPDCDGNGMVACSQCKGNGVNSVDHFNGR..FKAGATCWLCRGKR
Marchantia polymorpha 135  EPVPGAPVEKVNSILCSSCNGNGAVVCSQCVGEGINSKDHFNGR..FKAGTICWLCRGKK
Ostreococcus tauri     22  ADATNAAASQPRPKACPVCKGTGMKPCGQCEGTGSNREDLFGGR..YKAGDACWLCSGKA
Bathycoccus prasinos   67  AASAAASGGGEKKVCPVCQNTGLKPCGQCEGTGVNQEDKYGGKDGYKKGQPCWLCQGKR
consensus>50              .........k..s..C..CegnG.v.C.qC.G.GvN..D.%nG...%KaG..CWLCrGK.
                                       ▲     ▲    ★   ▲   ▲           ★ ★   ★  ★★      ▲▲▲▲▲★ ★
```

```
                          β3            β4
Arabidopsis thaliana  115  EVLCGDCNGAGFIGGFLSTFDE
Nicotiana sylvestris  111  DVLCGDCNGAGFLGGFMSTFDE
Zea mays              105  EILCGSCNGAGFLGGFLSTFDE
Oryza sativa          107  EILCGSCNGAGFLGGFMSTSDS
Populus trichocarpa   110  EILCGSCNGAGFLGGFMSTFDE
Picea sitchensis      122  ETLCGNCSGAGFMGGFMTTPDG
Marchantia polymorpha 193  QMLCGDCNGAGFMGGFMNTQDD
Ostreococcus tauri     80  KTMCGNCVDLT..DTF......
Bathycoccus prasinos  127  KTMCGNCIDLT..DSF......
consensus>50              e.$CGnC.gagf.ggFm.t.d.
                             ★ ▲★★   ▲
```

C

```
                                                                                β1
Arabidopsis thaliana    1  ...............................MAANNNPQGTKPNSLVCANCEGEGCVACSQCKGGGVN
Cre03.g201050.t1.2      1  MVECSAIAPETLQDIIVGGAVVGAVSVALYAGLKKDPVPCSLCQGTGGIRCFACGGDGRN
consensus>50               .................................G.K.#...C..C#G.G.!.C..C.G.N
```

```
                              TT          TT                                   β2
Arabidopsis thaliana   38  LIDHFNGQFKAGAL.....................CWLCRGKKEVLCGDCNGAGFIGGFLS
Cre03.g201050.t1.2     61  ATVSRDDLYDSKALGGGVAPPKRDPLGRTINPRDCKVCRGAGLVLCSQCKGTGFQSAF..
consensus>50               ......#..%...AL...................C..CRG...VLC.#C.G.GF...F..
```

```
Arabidopsis thaliana   78  TFDE
Cre03.g201050.t1.2         ....
consensus>50               ....
```

Fig. S4. **Crystal structure of *At*BSD2 and alignment of sequence homologs.** (**A**) Experimental electron density for *At*BSD2. The unbiased experimental density after Zn-MAD phasing and density modification is shown at 1.5 σ as a meshwork together with the final model in stick representation. Carbon atoms in the model are colored in a rainbow color gradient from N- to C-terminus. Red spheres represent ordered water molecules. Two perpendicular views are shown. (**B**) Alignment of BSD2 sequence homologs. Amino acid sequences of a representative set of BSD2 sequence homologs were aligned using the EBI Clustal-Ω server. Secondary structure elements for BSD2 from *A. thaliana* are indicated above the sequences. Similar residues are shown in red and identical residues in white on a red background. Blue frames indicate homologous regions. The consensus sequence is shown at the bottom. The chloroplast transit peptides are not shown. Blue triangles below the sequence indicate residues contacting the structural Zn centers. Asterisks below the sequence indicate mutations in *At*BSD2 analyzed in this study. The Uniprot/genome accession codes for the sequences are: Q9SN73, *Arabidopsis thaliana*; A0A1U7YLB0, *Nicotiana sylvestris*; Q9XF14, *Zea mays*; Q69WH2, *Oryza sativa*; U5FIP7, *Populus trichocarpa*; B8LL53, *Picea sitchensis*; A0A176WCX5, *Marchantia polymorpha*; A0A090MCK2, *Ostreococcus tauri*; K8EHK2, *Bathycoccus prasinos* 1. (**C**) Alignment of the predicted sequences for chloroplastic BSD2 from *A. thaliana* (Q9SN73) and Znf2 from *Chlamydomonas reinhardtii* (A8IXN3).
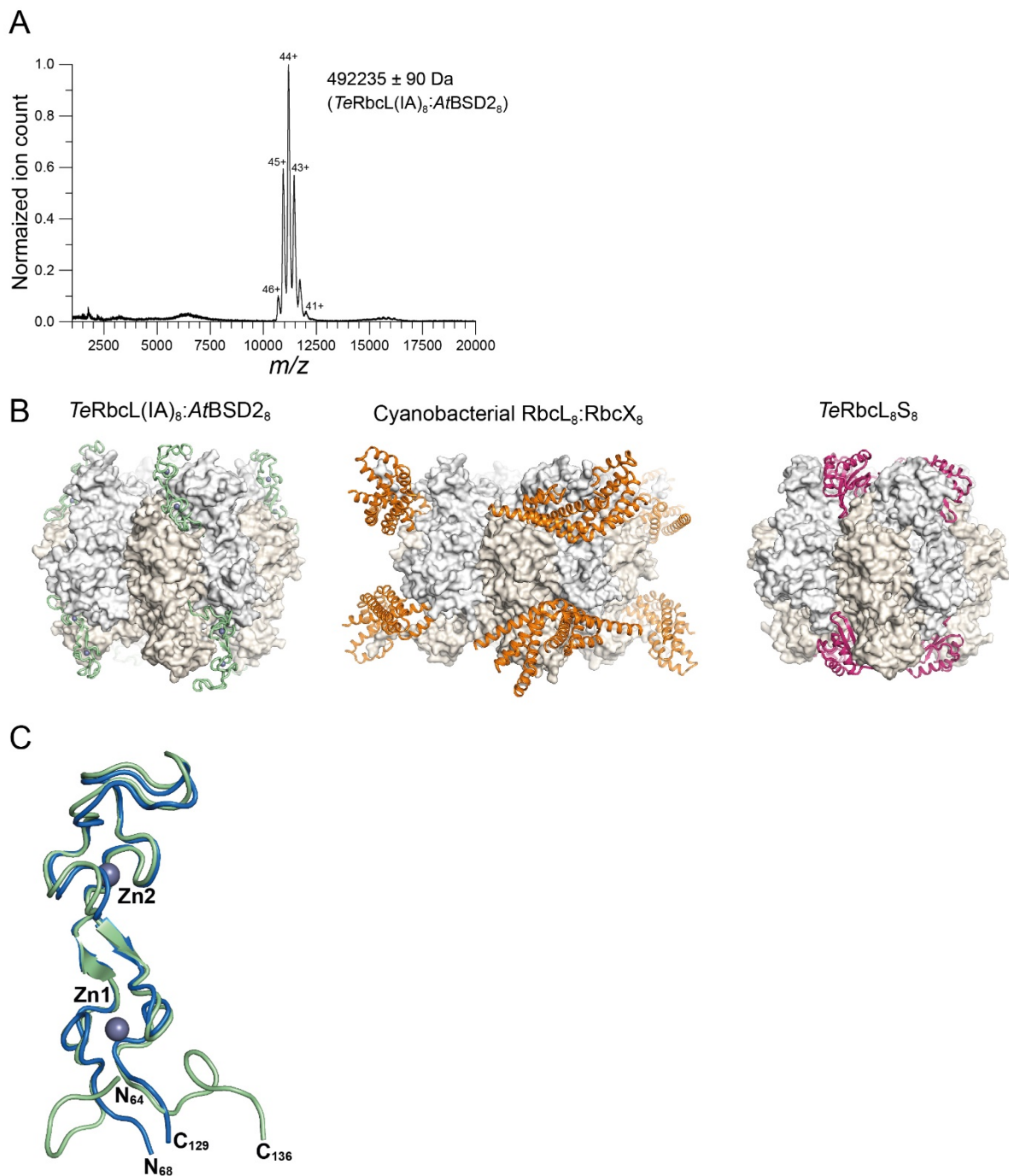
fig. S5

A

492235 ± 90 Da
(TeRbcL(IA)$_8$:AtBSD2$_8$)

B

TeRbcL(IA)$_8$:AtBSD2$_8$ Cyanobacterial RbcL$_8$:RbcX$_8$ TeRbcL$_8$S$_8$

C

Fig. S5. **Characterisation and crystal structure of the TeRbcL$_8$:AtBSD2$_8$ complex. (A)**

Native-MS analysis of purified, recombinant TeRbcL(IA)$_8$:AtBSD2$_8$ complex (theoretical mass

491528 Da). (**B**) Comparison of binding topologies of BSD2, RbcX and RbcS. Left,

TeRbcL(IA)$_8$:AtBSD2$_8$ complex; middle, cyanobacterial RbcL$_8$:RbcX$_8$ (*10*); right, TeRbcL$_8$S$_8$

17

(doi:10.2210/pdb3zxw/pdb). The RbcL$_8$ core is shown in surface representation, with the subunits of each RbcL$_2$ unit in white and light orange. BSD2 (green), RbcX (orange) and RbcS (magenta) are shown in ribbon representation. RbcS binding sites do not overlap with those of BSD2 or RbcX. (**C**) Overlay of the structures of free BSD2 (teal) and RbcL$_8$ bound BSD2 (green). Structures are shown in ribbon representation.

Table S1. Data collection, phasing and refinement statistics for X-ray crystallographic data.

	*At*BSD2		*Te*RbcL(IA)$_8$:*At*BSD2$_8$ complex
wwPDB	6EKB		6EKC
Data collection			
Space group	$P4_12_12$		$P1$
Cell dimensions			
a, b, c (Å)	55.71, 55.71, 48.13		217.92, 218.59, 310.27
α, β, γ (°)	90, 90, 90		69.40, 81.51, 65.50
	Peak	*Inflection*	
Wavelength	1.2816	1.2826	0.9700
Resolution (Å) *	47.85 – 1.91 (1.95 –1.91)	55.71 – 1.90 (1.94 –1.90)	198.3 – 2.63 (2.73 – 2.63)
R_{merge}	0.138 (4.852)	0.086 (1.732)	0.145 (1.215)
R_{pim}	0.020 (0.699)	0.025 (0.510)	0.090 (0.737)
$CC_{1/2}$	1.000 (0.742)	0.999 (0.760)	0.994 (0.431)
$I / \sigma I$	24.7 (1.1)	19.4 (1.4)	6.2 (1.1)
B_{Wilson} (Å^2)	38	43	42
Completeness (%)	99.7 (95.0)	100 (100)	98.5 (98.2)
Redundancy	48.6 (47.4)	12.2 (12.4)	3.6 (3.7)
Refinement			
Resolution (Å)		30.0 – 1.90	50.0 – 2.63
No. reflections		5503	1356119
R_{work} / R_{free}		0.392 / 0.459	0.362 / 0.372
No. atoms			
Protein		432	321025
Ligand/ion		3	160
Water		29	614
B-factors (Å^2)			
Protein		46	58
Ligand/ion		49	70
Water		46	28
R.m.s deviations			
Bond lengths (Å)		0.02	0.01
Bond angles (°)		1.81	1.18
Ramachandran plot (%) **			
Favored		100	95.8
Outliers		–	0.3

*Values in parentheses are for highest-resolution shell.
**Calculated with MolProbity.

Table S2. Plasmids generated for this study.

plasmid	vector	genes encoded
p*At*RbcLS	pBAD33k	*rbcL* (AtCg00490)
		rbcS3B (At5g38410)
p*At*RbcL	pBAD33k	*rbcL* (AtCg00490)
p*At*RbcS	pBAD33k	*rbcS3B* (At5g38410)
p*Nt*RbcLS	pBAD33k	*rbcL* (NC_001879)
		rbcS (NW_015886701)
pET11a*At*RbcL	pET11a	*rbcL* (AtCg00490)
pET11a*At*RbcS3B	pET11a	*rbcS3B* (At5g38410)
p*At*C60αβ/C20	pET11a	*cpn60α1* (At2g28000)
		cpn60β1 (At1g55490)
		cpn20 (At5g20720)
p*At*C60β/C20	pET11a	*cpn60β1* (At1g55490)
		cpn20 (At5g20720)
p*At*C60α/C20	pET11a	*cpn60α1* (At2g28000)
		cpn20 (At5g20720)
p*At*C60αβ	pET11a	*cpn60α1* (At2g28000)
		cpn60β1 (At1g55490)
p*At*C60αβ/C20/C10	pET11a	*cpn60α1* (At2g28000)
		cpn60β1 (At1g55490)
		cpn20 (At5g20720)
		cpn10 (At2g44650)
p*At*C60αβ/C10	pET11a	*cpn60α1* (At2g28000)
		cpn60β1 (At1g55490)
		cpn10 (At2g44650)
p*At*C60αβ/*Ec*ES	pET11a	*cpn60α1* (At2g28000)
		cpn60β1 (At1g55490)
		groES (NP_418566)
p*Ec*GroEL/ES	pET11a	*groEL* (NP_418567)
		groES (NP_418566)
p*At*C60αβ/C20/RbcS	pET11a	*cpn60α1* (At2g28000)
		cpn60β1 (At1g55490)
		cpn20 (At5g20720)
		rbcS3B (At5g38410)
p*At*C60α	pET11a	*cpn60α1* (At2g28000)
p*At*C60β	pET11a	*cpn60β1* (At1g55490)
p*At*C20	pET11a	*cpn20* (At5g20720)
p*At*R1/R2/X/B2	pCDFDuet	*raf1* (At3g04550)
		raf2 (At5g51110)
		rbcX2 (At5g19855)
		bsd2 (At3g47650)
p*At*R1/R2/X/B2_mut1 (BSD2: D95N/F97S)	pCDFDuet	*raf1* (At3g04550)
		raf2 (At5g51110)
		rbcX2 (At5g19855)
		bsd2-mut1 (At3g47650)
p*At*R1/R2/X/B2_mut2 (BSD2: R111E/K113E)	pCDFDuet	*raf1* (At3g04550)
		raf2 (At5g51110)
		rbcX2 (At5g19855)
		bsd2-mut2 (At3g47650)
p*At*R1/R2/X/B2_mut3 (BSD2: W108A/L109E)	pCDFDuet	*raf1* (At3g04550)
		raf2 (At5g51110)
		rbcX2 (At5g19855)
		bsd2-mut3 (At3g47650)
p*At*R1/R2/X/B2_mut4	pCDFDuet	*raf1* (At3g04550)

plasmid	vector	genes encoded
(BSD2: Q100E/F101A/K102A)		*raf2* (At5g51110)
		rbcX2 (At5g19855)
		bsd2-mut4 (At3g47650)
p*At*R1/R2/X/B2_mut5 (BSD2: L117E/G119T)	pCDFDuet	*raf1* (At3g04550)
		raf2 (At5g51110)
		rbcX2 (At5g19855)
		bsd2-mut5 (At3g47650)
p*At*R2/X/B2	pCDFDuet	*raf2* (At5g51110)
		rbcX2 (At5g19855)
		bsd2 (At3g47650)
p*At*R1/X/B2	pCDFDuet	*raf1* (At3g04550)
		rbcX2 (At5g19855)
		bsd2 (At3g47650)
p*At*R1/R2/B2	pCDFDuet	*raf1* (At3g04550)
		raf2 (At5g51110)
		bsd2 (At3g47650)
p*At*R1/R2/X	pCDFDuet	*raf1* (At3g04550)
		raf2 (At5g51110)
		rbcX2 (At5g19855)
p*Nt*R1/*At*R2/X/B2	pCDFDuet	*Ntraf1* (NW_015927590)
		raf2 (At5g51110)
		rbcX2 (At5g19855)
		bsd2 (At3g47650)
p*At*Raf1	pET11a	*raf1* (At3g04550)
p*At*Raf2	pET11a	*raf2* (At5g51110)
p*At*RbcX	pET11a	*rbcX2* (At5g19855)
p*At*BSD2	pET11a	*bsd2* (At3g47650)
pT*At*R2/X/B2	pTRC99SH	*raf2* (At5g51110)
		rbcX2 (At5g19855)
		bsd2 (At3g47650)
p*Te*RbcL(IA)/*At*BSD2 (RbcL: F345I/P415A)	pCDFDuet	*TerbcL(IA)* (NP_682296.1)
		Atbsd2 (At3g47650)
p*Se*RbcL	pHUE	*SerbcL* (WP_011242444.1)
pH*At*BSD2	pHUE	*bsd2* (At3g47650)

Table S3. N-terminal amino acid sequences of the recombinant proteins used in this study.

Protein name	N-termimal sequence*	Source
*At*RbcL	MSPQTETKAS…	cDNA
*At*RbcS	MKVWPPIGKK…	cDNA
*At*Raf1	MGMQQLYQPF…	cDNA
*At*Raf2	MMSNLAQDFL…	plasmid (ABRC, U13619)
*At*RbcX	MVSRIANRKS…	cDNA
*At*Bsd2	MMAANNNPQG…	cDNA
*At*Cpn60α	MNVKEIAFDQ…	cDNA
*At*Cpn60ß	MAAKELHFNK…	cDNA
*At*Cpn20	MASVVAPKYT…	cDNA
*At*Cpn10	MSTKWEPTKV…	cDNA
*Nt*RbcL	MSPQTETKAS…	synthesized
*Nt*RbcS	MCMQVWPPIN…	synthesized
*Nt*Raf1	MQQLYQPFRP…	synthesized
*Te*RbcL	MAYTQSKSQK…	genomic DNA
*Se*RbcL	MPKTQSAAGY…	genomic DNA
*Ec*GroEL	MAAKDVKFGN…	genomic DNA
*Ec*GroES	MNIRPLHDRV…	genomic DNA

*Amino acid residues in black are part of the protein sequence, while the amino acid residues in red are added as part of the cloned constructs.

References and Notes

1. I. Andersson, A. Backlund, Structure and function of Rubisco. *Plant Physiol. Biochem.* **46**, 275–291 (2008). doi:10.1016/j.plaphy.2008.01.001 Medline

2. A. Bracher, S. M. Whitney, F. U. Hartl, M. Hayer-Hartl, Biogenesis and metabolic maintenance of Rubisco. *Annu. Rev. Plant Biol.* **68**, 29–60 (2017). doi:10.1146/annurev-arplant-043015-111633 Medline

3. R. H. Wilson, S. M. Whitney, in *Directed Enzyme Evolution: Advances and Applications.*, M. Alcalde, Ed. (Springer International Publishing, 2017), chap. 4, pp. 101–126.

4. R. T. Furbank, W. P. Quick, X. R. R. Sirault, Improving photosynthesis and yield potential in cereal crops by targeted genetic manipulation: Prospects, progress and challenges. *Field Crops Res.* **182**, 19–29 (2015). [doi:10.1016/j.fcr.2015.04.009](doi:10.1016/j.fcr.2015.04.009)

5. T. J. Erb, J. Zarzycki, Biochemical and synthetic biology approaches to improve photosynthetic CO_2-fixation. *Curr. Opin. Chem. Biol.* **34**, 72–79 (2016). doi:10.1016/j.cbpa.2016.06.026 Medline

6. I. Sjuts, J. Soll, B. Bölter, Import of soluble proteins into chloroplasts and potential regulatory mechanisms. *Front. Plant Sci.* **8**, 168 (2017). doi:10.3389/fpls.2017.00168 Medline

7. M. Hayer-Hartl, A. Bracher, F. U. Hartl, The GroEL-GroES chaperonin machine: A nano-cage for protein folding. *Trends Biochem. Sci.* **41**, 62–76 (2016). doi:10.1016/j.tibs.2015.07.009 Medline

8. R. Trösch, T. Mühlhaus, M. Schroda, F. Willmund, ATP-dependent molecular chaperones in plastids—More complex than expected. *Biochim. Biophys. Acta* **1847**, 872–888 (2015). doi:10.1016/j.bbabio.2015.01.002 Medline

9. A. Vitlin Gruber, S. Nisemblat, A. Azem, C. Weiss, The complexity of chloroplast chaperonins. *Trends Plant Sci.* **18**, 688–694 (2013). doi:10.1016/j.tplants.2013.08.001 Medline

10. A. Bracher, A. Starling-Windhof, F. U. Hartl, M. Hayer-Hartl, Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. *Nat. Struct. Mol. Biol.* **18**, 875–880 (2011). doi:10.1038/nsmb.2090 Medline

11. C. Liu, A. L. Young, A. Starling-Windhof, A. Bracher, S. Saschenbrecker, B. V. Rao, K. V. Rao, O. Berninghausen, T. Mielke, F. U. Hartl, R. Beckmann, M. Hayer-Hartl, Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* **463**, 197–202 (2010). doi:10.1038/nature08651 Medline

12. S. Saschenbrecker, A. Bracher, K. V. Rao, B. V. Rao, F. U. Hartl, M. Hayer-Hartl, Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* **129**, 1189–1200 (2007). doi:10.1016/j.cell.2007.04.025 Medline

13. P. Kolesinski, P. Golik, P. Grudnik, J. Piechota, M. Markiewicz, M. Tarnawski, G. Dubin, A. Szczepaniak, Insights into eukaryotic Rubisco assembly—Crystal structures of RbcX chaperones from *Arabidopsis thaliana*. *Biochim. Biophys. Acta* **1830**, 2899–2906 (2013). doi:10.1016/j.bbagen.2012.12.025 Medline

14. D. Emlyn-Jones, F. J. Woodger, G. D. Price, S. M. Whitney, RbcX can function as a rubisco chaperonin, but is non-essential in *Synechococcus* PCC7942. *Plant Cell Physiol.* **47**, 1630–1640 (2006). doi:10.1093/pcp/pcl028 Medline

15. T. Onizuka, S. Endo, H. Akiyama, S. Kanai, M. Hirano, A. Yokota, S. Tanaka, H. Miyasaka, The *rbcX* gene product promotes the production and assembly of ribulose-1,5-bisphosphate carboxylase/oxygenase of *Synechococcus* sp. PCC7002 in *Escherichia coli*. *Plant Cell Physiol.* **45**, 1390–1395 (2004). doi:10.1093/pcp/pch160 Medline

16. S. Tanaka, M. R. Sawaya, C. A. Kerfeld, T. O. Yeates, Structure of the RuBisCO chaperone RbcX from *Synechocystis* sp. PCC6803. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 1109–1112 (2007). doi:10.1107/S090744490704228X Medline

17. M. Tarnawski, B. Gubernator, P. Kolesinski, A. Szczepaniak, Heterologous expression and initial characterization of recombinant RbcX protein from *Thermosynechococcus elongatus* BP-1 and the role of RbcX in RuBisCO assembly. *Acta Biochim. Pol.* **55**, 777–785 (2008). Medline

18. M. Tarnawski, S. Krzywda, W. Bialek, M. Jaskolski, A. Szczepaniak, Structure of the RuBisCO chaperone RbcX from the thermophilic cyanobacterium *Thermosynechococcus elongatus*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **67**, 851–857 (2011). doi:10.1107/S1744309111018860 Medline

19. A. Bracher, T. Hauser, C. Liu, F. U. Hartl, M. Hayer-Hartl, Structural analysis of the Rubisco-assembly chaperone RbcX-II from *Chlamydomonas reinhardtii*. *PLOS ONE* **10**, e0135448 (2015). doi:10.1371/journal.pone.0135448 Medline

20. L. Feiz, R. Williams-Carrier, K. Wostrikoff, S. Belcher, A. Barkan, D. B. Stern, Ribulose-1,5-bis-phosphate carboxylase/oxygenase accumulation factor1 is required for holoenzyme assembly in maize. *Plant Cell* **24**, 3435–3446 (2012). doi:10.1105/tpc.112.102012 Medline

21. P. Kolesinski, I. Belusiak, M. Czarnocki-Cieciura, A. Szczepaniak, Rubisco Accumulation Factor 1 from *Thermosynechococcus elongatus* participates in the final stages of ribulose-1,5-bisphosphate carboxylase/oxygenase assembly in *Escherichia coli* cells and in vitro. *FEBS J.* **281**, 3920–3932 (2014). doi:10.1111/febs.12928 Medline

22. T. Hauser, J. Y. Bhat, G. Miličić, P. Wendler, F. U. Hartl, A. Bracher, M. Hayer-Hartl, Structure and mechanism of the Rubisco-assembly chaperone Raf1. *Nat. Struct. Mol. Biol.* **22**, 720–728 (2015). doi:10.1038/nsmb.3062 Medline

23. S. M. Whitney, R. Birch, C. Kelso, J. L. Beck, M. V. Kapralov, Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3564–3569 (2015). doi:10.1073/pnas.1420536112 Medline

24. L. Feiz, R. Williams-Carrier, S. Belcher, M. Montano, A. Barkan, D. B. Stern, A protein with an inactive pterin-4a-carbinolamine dehydratase domain is required for Rubisco biogenesis in plants. *Plant J.* **80**, 862–869 (2014). doi:10.1111/tpj.12686 Medline

25. N. M. Wheatley, C. D. Sundberg, S. D. Gidaniyan, D. Cascio, T. O. Yeates, Structure and identification of a pterin dehydratase-like protein as a ribulose-bisphosphate

carboxylase/oxygenase (RuBisCO) assembly factor in the α-carboxysome. *J. Biol. Chem.* **289**, 7973–7981 (2014). [doi:10.1074/jbc.M113.531236](doi:10.1074/jbc.M113.531236) Medline

26. T. P. Brutnell, R. J. Sawers, A. Mant, J. A. Langdale, BUNDLE SHEATH DEFECTIVE2, a novel protein required for post-translational regulation of the *rbcL* gene of maize. *Plant Cell* **11**, 849–864 (1999). [doi:10.1105/tpc.11.5.849](doi:10.1105/tpc.11.5.849) Medline

27. L. Doron, N. Segal, H. Gibori, M. Shapira, The BSD2 ortholog in *Chlamydomonas reinhardtii* is a polysome-associated chaperone that co-migrates on sucrose gradients with the *rbcL* transcript encoding the Rubisco large subunit. *Plant J.* **80**, 345–355 (2014). [doi:10.1111/tpj.12638](doi:10.1111/tpj.12638) Medline

28. P. Goloubinoff, A. A. Gatenby, G. H. Lorimer, GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. *Nature* **337**, 44–47 (1989). [doi:10.1038/337044a0](doi:10.1038/337044a0) Medline

29. D. J. Orr, A. Alcântara, M. V. Kapralov, P. J. Andralojc, E. Carmo-Silva, M. A. Parry, Surveying Rubisco diversity and temperature response to improve crop photosynthetic efficiency. *Plant Physiol.* **172**, 707–717 (2016). Medline

30. L. P. Cloney, D. R. Bekkaoui, M. G. Wood, S. M. Hemmingsen, Assessment of plant chaperonin-60 gene function in *Escherichia coli*. *J. Biol. Chem.* **267**, 23333–23336 (1992). Medline

31. F. Baneyx, U. Bertsch, C. E. Kalbach, S. M. van der Vies, J. Soll, A. A. Gatenby, Spinach chloroplast cpn21 co-chaperonin possesses two functional domains fused together in a toroidal structure and exhibits nucleotide-dependent binding to plastid chaperonin 60. *J. Biol. Chem.* **270**, 10695–10702 (1995). [doi:10.1074/jbc.270.18.10695](doi:10.1074/jbc.270.18.10695) Medline

32. Y.-C. C. Tsai, O. Mueller-Cajar, S. Saschenbrecker, F. U. Hartl, M. Hayer-Hartl, Chaperonin cofactors, Cpn10 and Cpn20, of green algae and plants function as hetero-oligomeric ring complexes. *J. Biol. Chem.* **287**, 20471–20481 (2012). [doi:10.1074/jbc.M112.365411](doi:10.1074/jbc.M112.365411) Medline

33. O. Mueller-Cajar, S. M. Whitney, Evolving improved *Synechococcus* Rubisco functional expression in *Escherichia coli*. *Biochem. J.* **414**, 205–214 (2008). [doi:10.1042/BJ20080668](doi:10.1042/BJ20080668) Medline

34. R. H. Wilson, E. Martin-Avila, C. Conlan, S. M. Whitney, An improved *Escherichia coli* screen for Rubisco identifies a protein-protein interface that can enhance CO_2-fixation kinetics. *J. Biol. Chem.* jbc.M117.810861 (2017). [doi:10.1074/jbc.M117.810861](doi:10.1074/jbc.M117.810861) Medline

35. A. P. Duff, T. J. Andrews, P. M. Curmi, The transition between the open and closed states of rubisco is triggered by the inter-phosphate distance of the bound bisphosphate. *J. Mol. Biol.* **298**, 903–916 (2000). doi:10.1006/jmbi.2000.3724 Medline

36. R. E. Sharwood, Engineering chloroplasts to improve Rubisco catalysis: Prospects for translating improvements into food and fiber crops. *New Phytol.* **213**, 494–510 (2017). [doi:10.1111/nph.14351](doi:10.1111/nph.14351) Medline

37. R. E. Sharwood, O. Ghannoum, S. M. Whitney, Prospects for improving CO_2 fixation in C3-crops through understanding C4-Rubisco biogenesis and catalytic diversity. *Curr. Opin. Plant Biol.* **31**, 135–142 (2016). [doi:10.1016/j.pbi.2016.04.002](doi:10.1016/j.pbi.2016.04.002) Medline

38. S. P. Long, A. Marshall-Colon, X. G. Zhu, Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell* **161**, 56–66 (2015). doi:10.1016/j.cell.2015.03.019 Medline

39. J. R. Evans, Improving photosynthesis. *Plant Physiol.* **162**, 1780–1793 (2013). doi:10.1104/pp.113.219006 Medline

40. W. Yamori, K. Hikosaka, D. A. Way, Temperature response of photosynthesis in C3, C4, and CAM plants: Temperature acclimation and temperature adaptation. *Photosynth. Res.* **119**, 101–117 (2014). doi:10.1007/s11120-013-9874-6 Medline

41. K. A. Bishop, A. M. Betzelberger, S. P. Long, E. A. Ainsworth, Is there potential to adapt soybean (*Glycine max Merr.*) to future [CO_2]? An analysis of the yield response of 18 genotypes in free-air CO_2 enrichment. *Plant Cell Environ.* **38**, 1765–1774 (2015). doi:10.1111/pce.12443 Medline

42. K. Wostrikoff, D. Stern, Rubisco large-subunit translation is autoregulated in response to its assembly state in tobacco chloroplasts. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6466–6471 (2007). doi:10.1073/pnas.0610586104 Medline

43. O. Emanuelsson, H. Nielsen, G. von Heijne, P. Chloro, ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* **8**, 978–984 (1999). doi:10.1110/ps.8.5.978 Medline

44. R. T. Baker, A.-M. Catanzariti, Y. Karunasekara, T. A. Soboleva, R. Sharwood, S. Whitney, P. G. Board, Using deubiquitylating enzymes as research tools. *Methods Enzymol.* **398**, 540–554 (2005). doi:10.1016/S0076-6879(05)98044-0 Medline

45. A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996). doi:10.1021/ac950914h Medline

46. J. Rappsilber, Y. Ishihama, M. Mann, Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003). doi:10.1021/ac026117i Medline

47. N. Nagaraj, J. R. Wisniewski, T. Geiger, J. Cox, M. Kircher, J. Kelso, S. Pääbo, M. Mann, Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7**, 548 (2011). doi:10.1038/msb.2011.81 Medline

48. P. Evans, Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006). doi:10.1107/S0907444905036693 Medline

49. P. R. Evans, G. N. Murshudov, How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013). doi:10.1107/S0907444913000061 Medline

50. S. French, K. Wilson, On the treatment of negative intensity observations. *Acta Crystallogr. A Found. Adv.* **34**, 517–525 (1978). doi:10.1107/S0567739478001114

51. E. Potterton, P. Briggs, M. Turkenburg, E. Dodson, A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1131–1137 (2003). doi:10.1107/S0907444903008126 Medline

52. G. M. Sheldrick, Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485 (2010). doi:10.1107/S0907444909038360 Medline

53. T. Pape, T. R. Schneider, HKL2MAP: A graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Cryst.* **37**, 843–844 (2004). doi:10.1107/S0021889804018047

54. K. Cowtan, The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006). doi:10.1107/S0907444906022116 Medline

55. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004). doi:10.1107/S0907444904019158 Medline

56. G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011). doi:10.1107/S0907444911001314 Medline

57. V. B. Chen, W. B. Arendall 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010). doi:10.1107/S0907444909042073 Medline

58. A. Vagin, A. Teplyakov, Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25 (2010). doi:10.1107/S0907444909042589 Medline

59. P. Gouet, X. Robert, E. Courcelle, ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3323 (2003). doi:10.1093/nar/gkg556 Medline