nature research

Corresponding author(s): Mei Li, Jun Minagawa, Zhenfeng Liu,

Last updated by author(s): Jun 2, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	SerialEM v3.6/3.6.1, MotionCor2_1.4.0, CTFFIND4/4.1
Data analysis	MotionCor2_1.4.0, CTFFIND4/4.1, RELION-3.0/3.1, ResMap1.1.4, UCSF Chimera 1.9/1.11, Coot-0.8.6.1, PHENIX-v1.15, MolProbity, PyMOL-2.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The atomic coordinates of the CrPSI-LHCI-LHCII supercomplex have been deposited in the Protein Data Bank with accession code 7DZ7 (native supercomplex from pph1;pbcp mutant) and 7DZ8 (supercomplex from ΔLhcbM1 mutant). The cryo-EM maps of the native and ΔLhcbM1 mutant supercomplexes have been deposited in the Electron Microscopy Data Bank with accession code EMD-30925 and EMD-30926, respectively. In addition, the atomic coordinates and locally refined cryo-EM maps of LHCII trimers have been deposited in the Protein Data Bank and the Electron Microscopy Data Bank with accession codes 7E0J and EMD-30934 for LHCII-1 in native supercomplex, 7E0K and EMD-30935 for LHCII-2 in native supercomplex, 7E0H and EMD-30932 for LHCII-1 in supercomplex from ΔLhcbM1 mutant, 7E0I and EMD-30933 for LHCII-2 in supercomplex from ΔLhcbM1 mutant. Source data for Figure 4b, Extended Data Figure 1b, Supplementary Figures 2b and 2c are provided with this paper. All other data generated or analyzed are available from the corresponding authors on reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Statistical methods were not used to predetermine the sample size. The amount of cryo-EM micrographs collected was based on the cryo-EM time allocation and previous knowledge estimating that the size is sufficient to generate a high-resolution density.
Data exclusions	The initial cryo-EM images are screened manually to exclude those with low contrast, thick ice or severe ice contaminations, which is a standard procedure for cryo-EM data processing. No biochemical data have been excluded.
Replication	Most experiments were repeated independently from 2 to 10 times, and were all successfully reproduced.
Randomization	Randomization is not relevant to our study because we tried to solve the structure of the specific protein supercomplex.
Blinding	The investigators were not blinded to allocation during experimental procedures because we are dealing with the specific protein supercomplex.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Rabbit polyclonal antibodies against ATPB (Cat#AS05 085, at 1:10000 dilution) obtained from Agrisera (Sweden) and LhcbM5 raised against the peptide RVNGGPAGEGLD (Eurofins Genomics, at 1:20000 dilution) were used. An anti-rabbit IgG, horseradish-peroxidase-conjugated antibody (Cat#7074, at 1:10000 dilution) obtained from Cell Signaling Technology (Danvers, MA, USA) was used as a secondary antibody.
Validation	Anti-ATPB antibody (Agrisera Cat#AS05 085) was validated by the provider: https://www.agrisera.com/en/artiklar/atpb-beta-subunits-of-atp-synthase-global-antibody. Anti-LhcbM5 antibody was affinity-purified against the peptide RVNGGPAGEGL. The capacity of this polyclonal antibody to recognize LhcbM5 has not been published previously. In the current study, the antibody produced two bands of expected size in western blots. The upper band corresponded to LhcbM5. See Fig.S3b and the relevant source file. Anti-rabbit IgG (Cell Signaling Technology Cat# 7074) was validated by the provider: https://www.cellsignal.com/products/secondary-antibodies/anti-rabbit-igg-hrp-linked-antibody/7074.