

Specific features and assembly of the plant mitochondrial complex I revealed by cryo-EM

Heddy Soufari¹, Camila Parrot¹, Lauriane Kuhn², Florent Waltz¹⁺ & Yaser Hashem¹⁺

¹ Institut Européen de Chimie et Biologie, U1212 Inserm, Université de Bordeaux, 2 rue R. Escarpit, F-33600 Pessac, France

² Plateforme protéomique Strasbourg Esplanade FRC1589 du CNRS, Université de Strasbourg, Strasbourg, France

+Corresponding authors

Abstract:

Mitochondria are the powerhouses of eukaryotic cells and the site of essential metabolic reactions. Their main purpose is to maintain the high ATP/ADP ratio that is required to fuel the countless biochemical reactions taking place in eukaryotic cells¹. This high ATP/ADP ratio is maintained through oxidative phosphorylation (OXPHOS). Complex I or NADH:ubiquinone oxidoreductase is the main entry site for electrons into the mitochondrial respiratory chain and constitutes the largest of the respiratory complexes². Its structure and composition varies across eukaryotes species. However, high resolution structures are available only for one group of eukaryotes, opisthokonts³⁻⁶. In plants, only biochemical studies were carried out, already hinting the peculiar composition of complex I in the green lineage. Here, we report several cryo-electron microscopy structures of the plant mitochondrial complex I at near-atomic resolution. We describe the structure and composition of the plant complex I including the plant-specific additional domain composed by carbonic anhydrase proteins. We show that the carbonic anhydrase is an heterotrimeric complex with only one conserved active site. This domain is crucial for the overall stability of complex I as well as a peculiar lipid complex composed cardiolipin and phosphatidylinositols. Moreover we also describe the structure of one of the plant-specific complex I assembly intermediate, lacking the whole P_D module, in presence of the maturation factor GLDH. GLDH prevents the binding of the plant specific P1 protein, responsible for the linkage of the P_P to the P_D module. Finally, as the carbonic anhydrase domain is likely to be associated with complex I from numerous other known eukaryotes, we propose that our structure unveils an ancestral-like organization of mitochondrial complex I.

Complex I is the largest multimeric enzyme of the respiratory chain, composed of more than 40 protein subunits. 14 are strictly conserved proteins subunits, vestige of complex I of bacterial origin⁷. The additional subunits, referred to as supernumerary subunits, were acquired during eukaryotes evolution. Part of these additional subunits are conserved among other eukaryotes and play essential roles for the structure, function and the association with the other respiratory chain complexes, eg. to assemble into respirasome in animals⁸. Recently, several high resolution 3D structures of the complete mitochondrial complex I of opisthokonts were determined by cryo-EM in mammalian species⁴⁻⁶ and in the aerobic yeast *Yarrowia lipolytica*^{3,9}, revealing the organization of their additional specie-specific subunits. However, in plants, even though extensive biochemical characterization was conducted^{10,11}, high-resolution structures of mitochondrial complex I are yet to be derived. Early negative staining studies¹² revealed the presence of a large additional membrane attached domain, absent from animal and yeast species, hinting at the peculiar structure and composition of the plant complex I.

In order to obtain a high-resolution structure of the plant mitochondrial complex I and its additional subunits, we purified mitochondria from *Brassica oleracea var. botrytis*, a close relative to the model plant *Arabidopsis* (both belong to the group of Brassicaceae plants), as previously described¹³ (see Methods). We recorded cryo-EM images of membrane complexes purified from sucrose gradient (see Methods), corresponding to mitochondrial complex I. After particle sorting (see Methods) we obtained cryo-EM reconstructions of two main complex I states: the full plant mitochondrial complex I as well as reconstruction of a complex I assembly intermediate, without the PD module, and with the plant-specific assembly factor GLDH¹⁴ (Extended Data Fig. 1). Focused refinement and particle

polishing were performed in RELION3¹⁵ (see Methods), allowing reconstruction of the full complex I at 3.7Å resolution, focused classification on the membrane arm and P_P module yielding 3.4Å. The P_D module being more flexible produced more scant densities reporting lower resolutions (Extended Data Fig. 1 and 2). The cryo-EM maps allowed the determination of many well-resolved features and clear side-chain densities (Extended Data Fig. 3) that enabled modelling of the 45 proteins of the plant mitochondrial complex I (Extended Data Table 1), as well as several characteristic ligands that were directly identified from the density: the eight canonical FeS clusters of the matrix arm, the FMN of the 51kDa subunit, the NADPH molecule in the 39kDa, three cardiolipins two phosphatidylinositol and one phosphatidylethanolamine (Extended Data Fig. 3). Ubiquinone was also clearly visible in the Q module (Extended Data Fig. 4).

General description

The plant mitochondrial complex I has the classical open L-shape formed by the matrix and membrane arms (Fig. 1). In the matrix arm, electrons are transferred from NADH to ubiquinone along the FeS clusters with distances similar to what has been described in bacteria and opisthokonts mitochondrial complex I (Extended Data Fig. 6). The membrane part, where proton pumping takes place, is also conserved. Those highly conserved features are part of the 14 minimal protein subunits conserved in bacteria and other mitochondrial complex I². The additional mitochondria specific proteins enhance the volume and mass of the complex I compared to bacteria. In plant, 25 proteins shared with aerobic yeast or mammals are present. Similarly to the previously obtained complex I structure^{4,6,9,16}, the supernumerary subunits, specific to mitochondrial complex I, form a shell around the core subunits, adding nearly 400kDa of proteins to the conserved subunits (Extended Data Fig. 6). These supernumerary subunits are mainly composed of α -helices and are largely interconnected with the core subunits. In the Q module of the matrix arm, our reconstructions allowed us to clearly visualize ubiquinone in the pocket formed by Nad1, PSST and Nad7 (Extended Data Fig. 4). Our results confirm the recently observed ubiquinone position described for *Y.lipolytica*³.

Carbonic anhydrase: specific component of the plant complex I

The main feature of the plant complex I is the presence of an additional globular matrix-exposed domain, bound to the membrane arm. It holds an overlapping position with the NDUFA10/42-kDa subunit found in mammalian complex I⁴. This domain is formed by a trimeric complex, the carbonic anhydrase (CA). CA are zinc-containing enzymes that catalyze the reversible hydration of CO₂ to HCO₃⁻, whose roles are usually carbon fixation and pH maintenance¹⁷. The carbonic anhydrase activity evolved independently several times during evolution, resulting in several CA enzyme families. The plant mitochondrial CA is part of the gamma-carbonic anhydrase (γ CA) family, which was originally identified in the archaeon *Methanosarcina thermophila*¹⁸. Here, the overall structure of the γ CA is conserved: it is a trimer composed of characteristic left-handed beta-helix monomers, forming triangular prisms (Fig. 2a). However, in contrast to the archaeal enzyme, the γ CA present in plant complex I is an heterotrimer. Indeed, in Arabidopsis five γ CA genes were identified. Three have highly conserved catalytic domains and are termed γ CA1-3, similar to the archaeal enzyme, while the two other show less homology to their prokaryotic homologues especially in the catalytic domain and were thus termed γ CA-like proteins¹⁹. Our reconstruction shows that the γ CA domain is formed by one copy of γ CAL1, representing the most proximal subunit of the membrane part, and two copies of γ CA1 (Fig. 2a and Extended Data Fig. 5). Each subunit have long N and C-terminal extensions interconnecting the 3 subunits and contacting the membrane subunits of complex I (Fig. 2c). These extensions also serve as anchors for the γ CA domain by interacting with the plant specific P2 protein N-terminal part which is anchored in the membrane. The C-terminal part of the canonical Nad6 protein also contacts the CA domain stabilizing it. γ CA is an enzyme, where one histidine of one subunit and two histidine of the adjacent subunit contribute to coordinate a zinc atom. However, as γ CAL proteins lack two of those three histidines, it impairs the formation of the zinc coordination with both γ CA subunits. Hence the only conserved active site is at the interface of the two γ CAs, which we confirmed by visualizing both the zinc and an additional density most likely corresponding to HCO₃⁻ (Fig. 2b and Extended Data Fig. 5). This confirms that the γ CA domain associated with complex I is indeed active. However, the function of complex I-associated γ CA is not well understood. It was suggested to play a role in complex I assembly as γ CA is found in early assembly intermediates of the complex and is essential for complex I formation²⁰. Indeed, embryo development of γ CA mutants is strongly delayed and seed development arrested before maturation. Moreover complemented mutants with inactive γ CA variants is sufficient to complement the mutant phenotype, showing that carbonic anhydrase activity is not essential¹⁹. Recently it was shown that

carbonic anhydrase activity is also present in photosynthetic complex I, but is carried out by a different type of CA enzyme, positioned at the end of the P_D module that would contribute to proton generation²¹. However, in our complex, the only conserved active site is positioned at the most distal position relative to the membrane. Nevertheless, one can't exclude its contribution to proton generation, although its effect would be relatively modest.

Interestingly, the CA domain is in tight interaction with a peculiar lipid complex positioned below the latter and between the plant specific P2 and proteins B14.5b and Nad2, close to the Nad2-Nad4 junction (Fig. 2d). This lipid complex is composed of one cardiolipin coordinated by two phosphatidylinositol. Each inositol heads sandwich the glycerol linker of the cardiolipin and stabilize it. This lipid block was also observed in the assembly intermediate (see below) and appears to be very stable. Given the essential role of lipids for complex I activity, this lipid block might be crucial for the plant complex I efficiency and stability, which might be why CA mutants are so heavily affected^{3,22}. Altogether, our analysis strongly suggest that the γ CA is essential for the membrane part formation and stability, acting as an essential architectural factor that would remain even in mature complex.

Assembly intermediate of the plant complex I

During 3D classification, two classes of complex I lacking the most distal proton pumping components of the membrane domain – Nad4 and Nad5 and their associated additional protein subunits – were found to naturally accumulate in our sample (Extended Data Fig. 1). One of the two classes presented an additional globular density on the intermembrane space exposed side of complex I. Such density wasn't observed in full complex I classes. This class was thus attributed to an assembly intermediate of complex I (Fig. 3). In plants, such complex was already described to accumulate naturally. Indeed, extensive biochemical analyses identified complex I* as an assembly intermediate lacking the most distal membrane components and presenting an additional factor, GLDH (L-Galactono-1,4-lactone dehydrogenase). GLDH was identified in our mass spectrometry data (Extended Data Table 1), along with the other components of complex I. Likewise γ CA, GLDH is an enzyme catalysing the last enzymatic step of the ascorbate biosynthetic pathway in plants²³. It was never found in the full complex I, however, it was shown to associate with complex I intermediates, complex I* being the largest, as well as with smaller complex intermediates that were not observed here²⁰. Moreover, in *gl dh* knock-out mutant, complex I is undetectable, which hinted its role as a plant specific assembly factor^{14,24}. Using focused classification and refinement, we were able to derive an intermediate-resolution reconstruction on the additional density corresponding to GLDH which was sufficient to attribute the density to the protein. Importantly, the complex I binding domain of GLDH is resolved at near-atomic resolution, thus validating our assignment (Fig. 3b-d). In this assembly intermediate, GLDH interacts with B14.5b - that is slightly displaced compared to the mature complex - and 15kDa. Moreover the whole carbonic anhydrase is slightly shifted toward the missing P_D module (Fig. 3c). We also successfully identified the N-terminal tail of GLDH bound to the maturing complex I, between B14.5b and 15kDa preventing the plant specific protein P1 from docking to the P_P module. P1 appears to functionally replace the NDUFB5 subunit, which is absent in plants. In the GLDH context the P1 path is blocked in the assembly intermediate due to B14.5b movement and N-ter tail of GLDH (Fig. 3). Thus, it appears that GLDH prevents the binding of P1 to the P_P module, until its release, which will then allow association with the P_D module.

In conclusion, our high-resolution cryo-EM structures of the plant mitochondrial complex I, reveal a most likely ancestral features of this respiratory complex (Fig. 4). Indeed, γ CA was identified as a complex I component in all Viridiplantae, including algae of the Chlorophyceae class^{11,12,25-27}, as well as in *Euglena gracilis*, a photosynthetic protozoan related to trypanosomes²⁸ and possibly *Acanthamoeba castellanii*, a protozoan amoeboid²⁹ and in non-photosynthetic organisms (e.g. Tetrahymena and Reclinomonas)²⁹ (Fig. 4). Given the wide presence of the γ CA in eukaryotes – members of the γ CA are found in all eukaryotic lineages but not in opisthokonts – we suggest that γ CA proteins may be an ancestral feature of mitochondrial complex I. Moreover we describe a yet unobserved plant-specific assembly intermediate, shedding light into the respiratory complex I maturation.

METHODS

Methods and any associated references are available in the online version of the paper.

ACKNOWLEDGEMENTS

This work has benefitted from the facilities and expertise of the Biophysical and Structural Chemistry platform (BPCS) at IECB, CNRS UMS3033, Inserm US001, University of Bordeaux. We thank A. Bezault for assistance with the Talos Arctica electron microscope. We thank, J. Chicher and P. Hamman of the Strasbourg Espanade proteomic analysis for the proteomic analysis. This work was supported by a European Research Council Starting Grant (TransTryp ID:759120) and a Agence Nationale de la Recherche (ANR) grant [MITRA, ANR-16-CE11-0024-02]] to YH.

AUTHOR CONTRIBUTIONS

FW, HS and YH designed and coordinated the experiments. FW purified the mitochondria and mitochondrial complex I. HS and CP prepared grids. HS acquired the cryo-EM data and processed the cryo-EM results with FW. HS and FW built the atomic models. LK performed the mass-spectrometry experiments. FW, HS and YH interpreted the structure. FW, HS, CP and YH wrote and edited the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

DATA AVAILABILITY

The cryo-EM maps of plant mitochondrial complex I have been deposited at the Electron Microscopy Data Bank (EMDB): EMDB-XXXXX. The corresponding atomic models been deposited in the Protein Data Bank (PDB) under the accession XXXX. Mass spectrometric data were deposited with the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier XXXXXXXX.

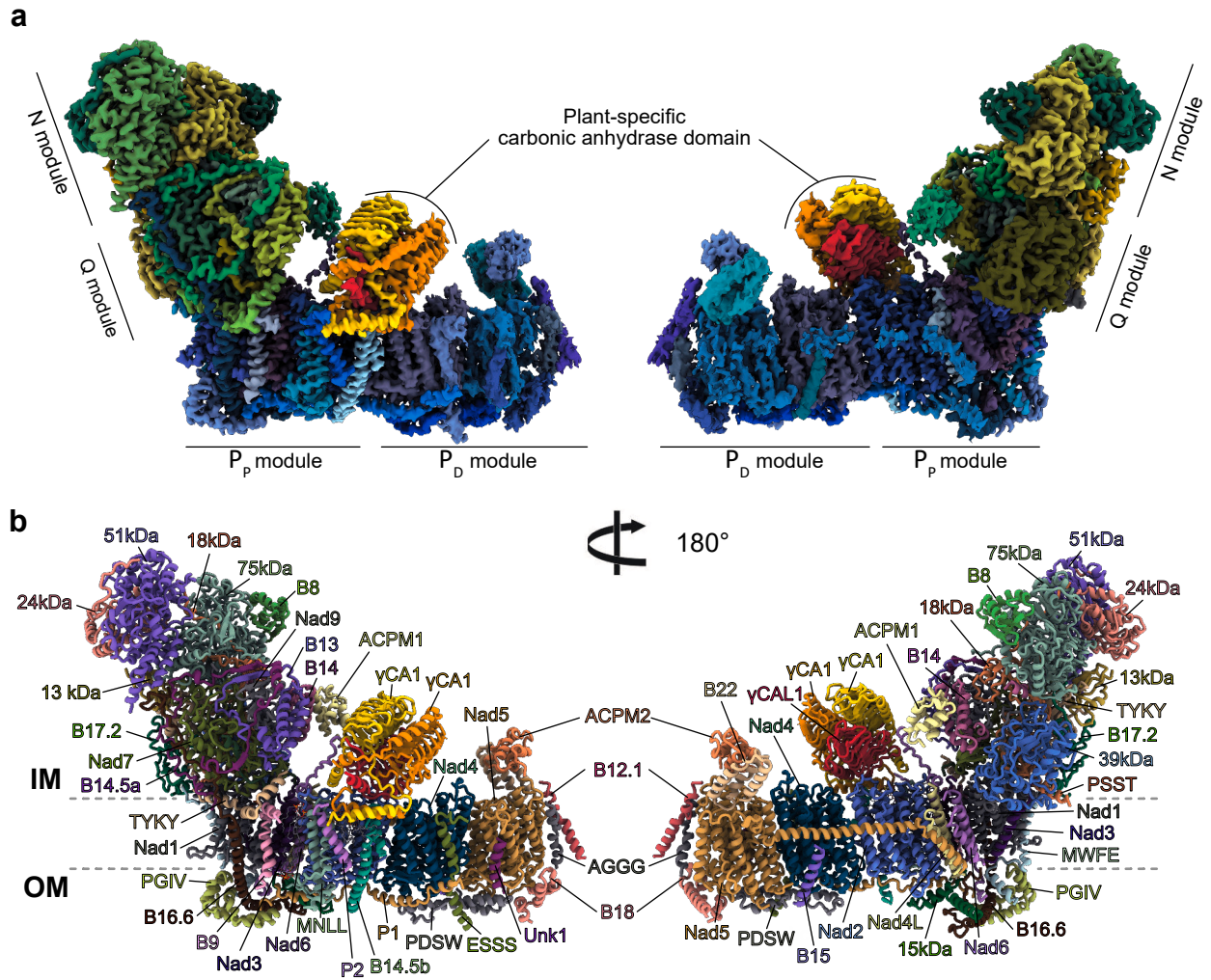


Figure 1. Overall structure of the plant mitochondrial complex I

a Composite cryo-EM map of the complete plant complex I. The membrane part is displayed in blue shades, the matrix arm is displayed in green shades and the specific γ -carbonic anhydrase module is displayed in orange shades. Individual N, Q, P_D and P_P modules are indicated. **b** The resulting atomic model, each individual 45 proteins are annotated and displayed with a different color.

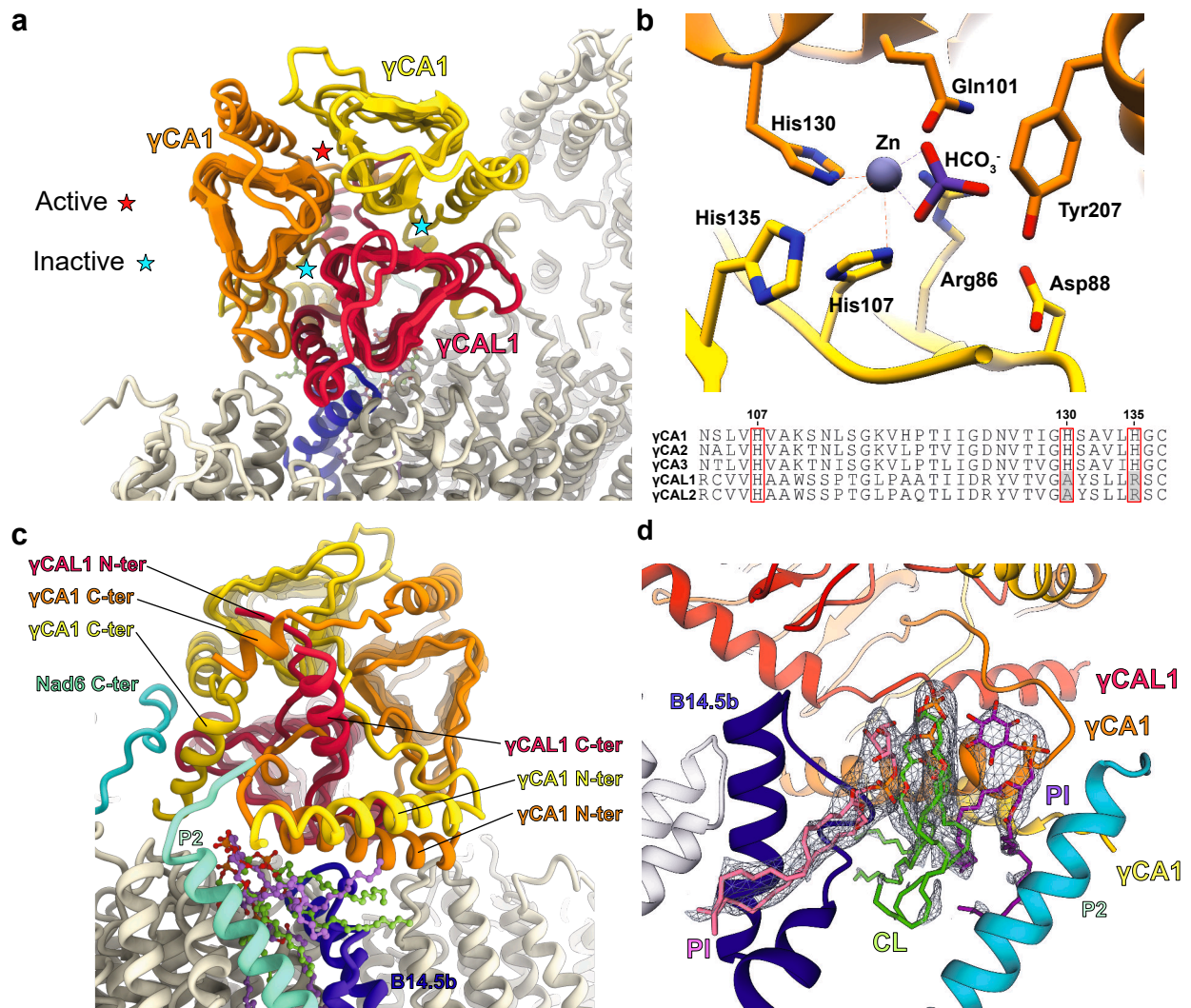


Figure 2. Hetero-trimeric γ -carbonic anhydrase is a specific component of the plant mitochondrial complex I

a-c Front and back view of the carbonic anhydrase showing individual subunits and contacting proteins. Active and inactive sites are shown in **a**. **b** Close up of the γ CA1/ γ CA1 only conserved active site, showing the coordinate zinc by the three histidine residues between the two γ CA1 subunits. Sequence alignment of γ CA and γ CAL highlight the catalytic histidines. **c** Back view of the carbonic anhydrase, highlighting the N and C-ter extensions of the γ CA and γ CAL as well as the proteins allowing to anchor the whole domain to the membrane arm. **d** The lipid block composed of one central cardiolipin and two phosphatidylinositol molecules is displayed in its density. The lipids are sandwiched between the carbonic anhydrase domain, the plant-specific protein P2, Nad2 and B14.5. Both inositol head groups contribute to coordinate the cardiolipin. CL stands for cardiolipin and PI for phosphatidylinositol.

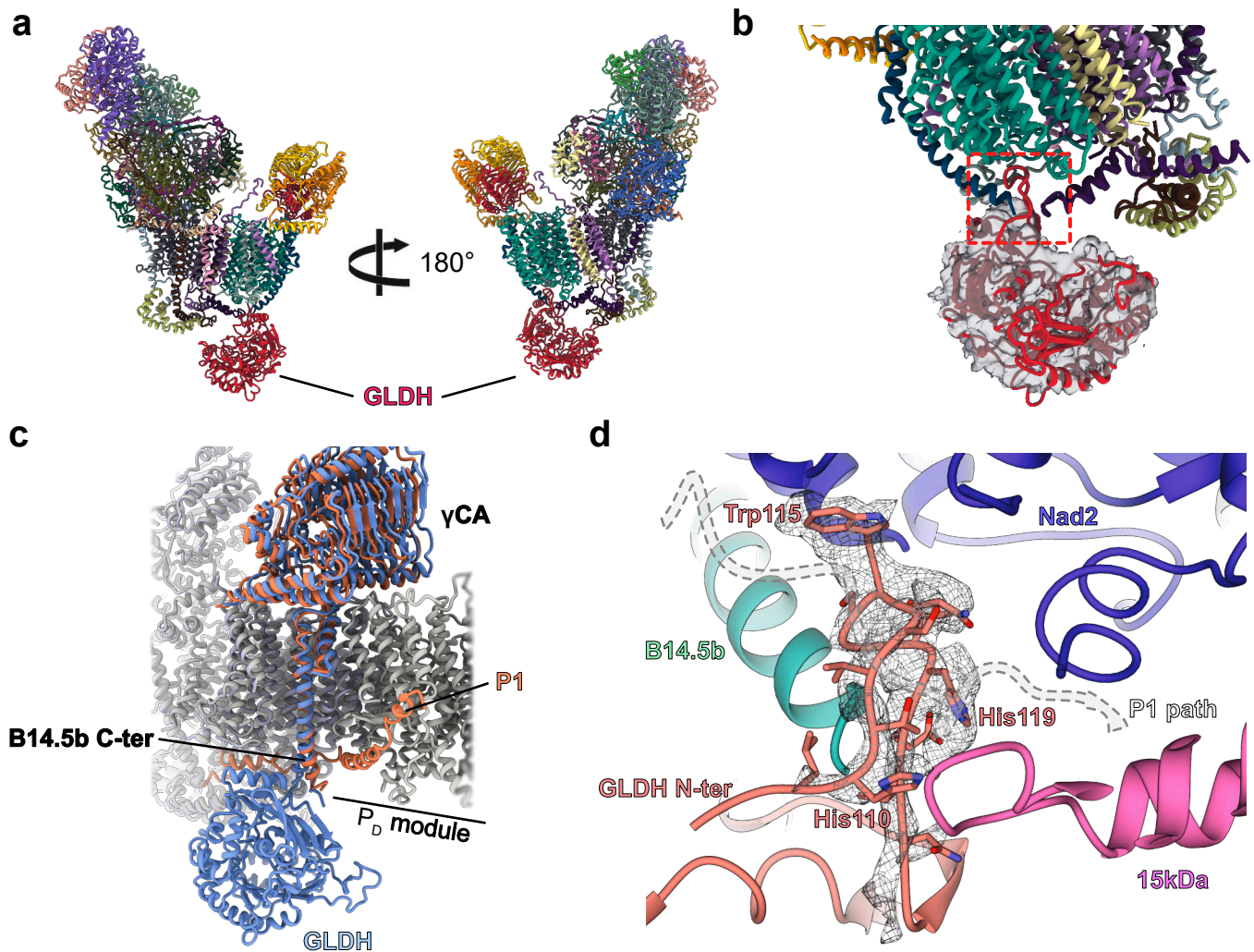


Figure 3. GLDH is a plant-specific assembly factor of mitochondrial complex I

a Atomic model of the assembly intermediate, the whole P_D module is missing and GLDH contacts the P_P module on the intermembrane side. **b** GLDH is shown in cut-out and filtered density. The interaction point is highlighted and a zoomed view is shown in **d**. **c** Overall movement in the assembly intermediate compared to the mature complex, B14.5b C-terminal part is slightly deviated and the whole γ CA domain is shifted, GLDH and absence of the P1 protein are also highlighted. Coral correspond to mature complex I and blue to assembly intermediate. **d** Focused view of the interaction point of GLDH with the complex. The N-terminal part of GLDH was identified thanks to specific aromatic residues and is shown in its density. The N-ter part of GLDH as well as the deviation of B14.5b block P1 path which is shown in dashed lines.

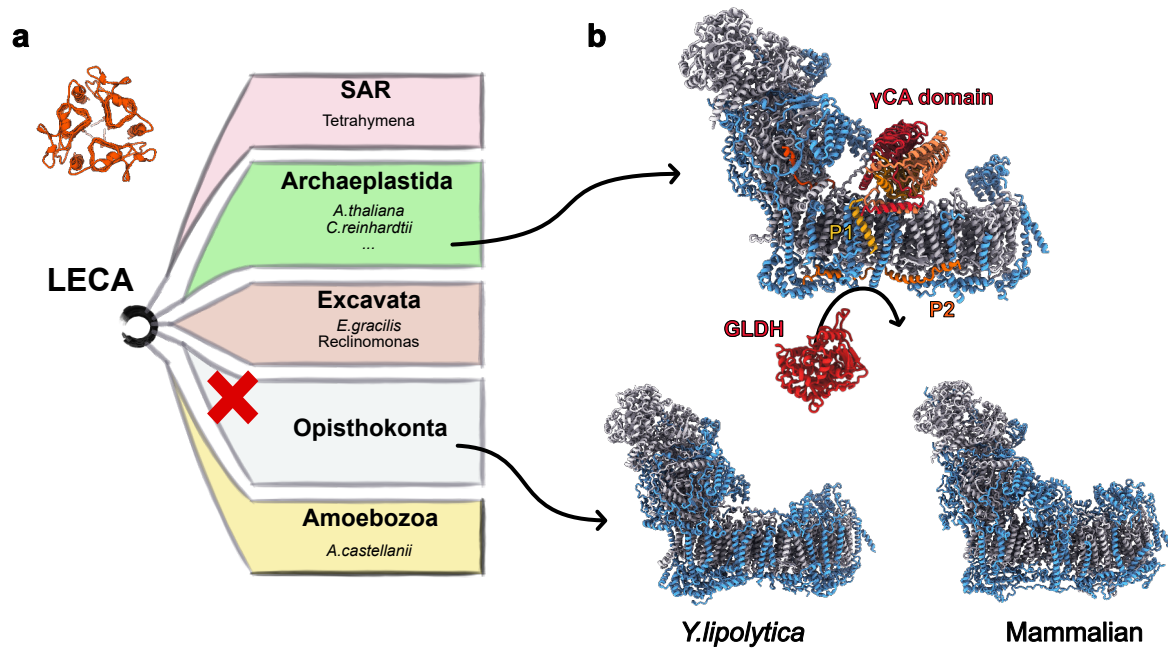
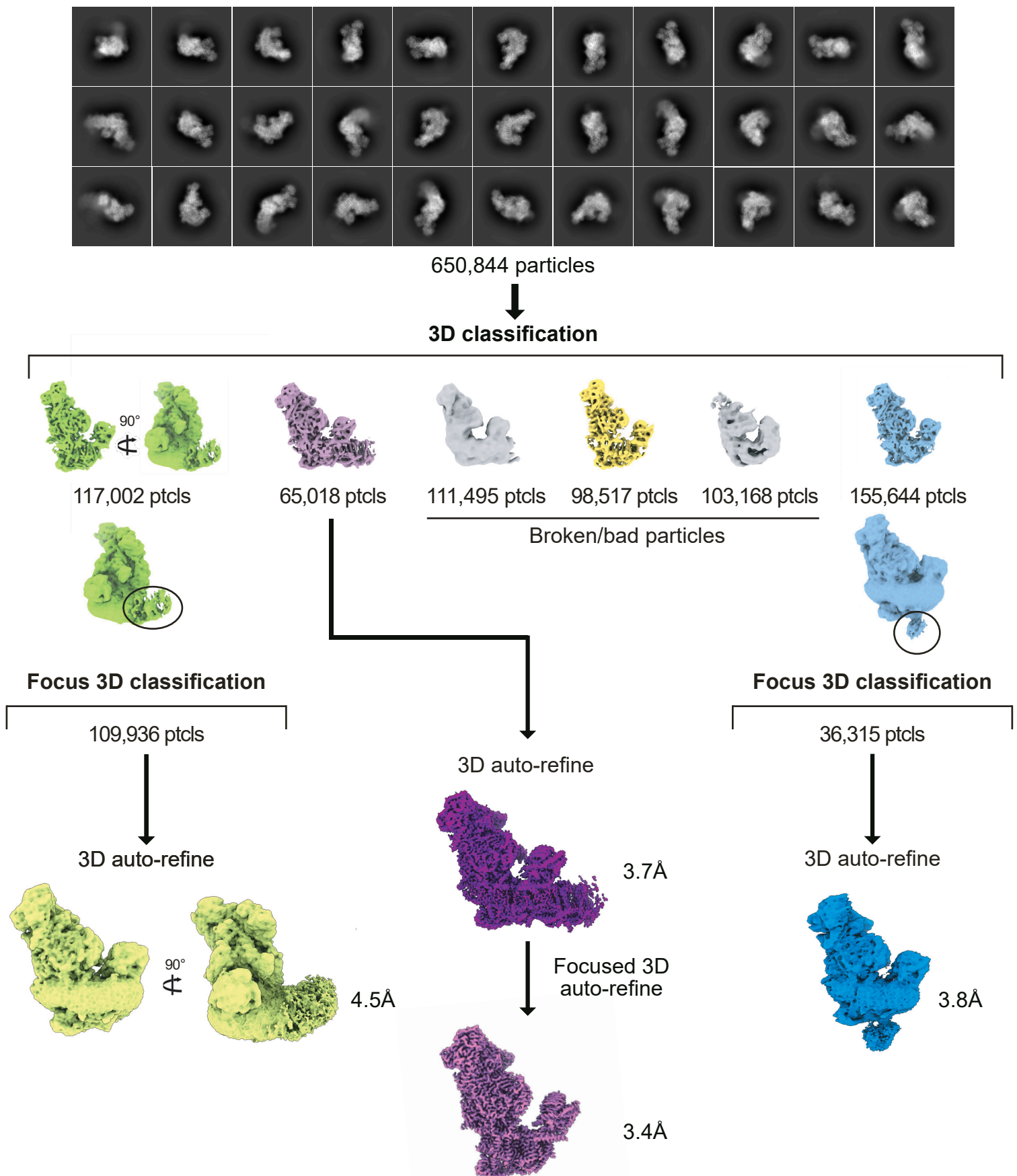


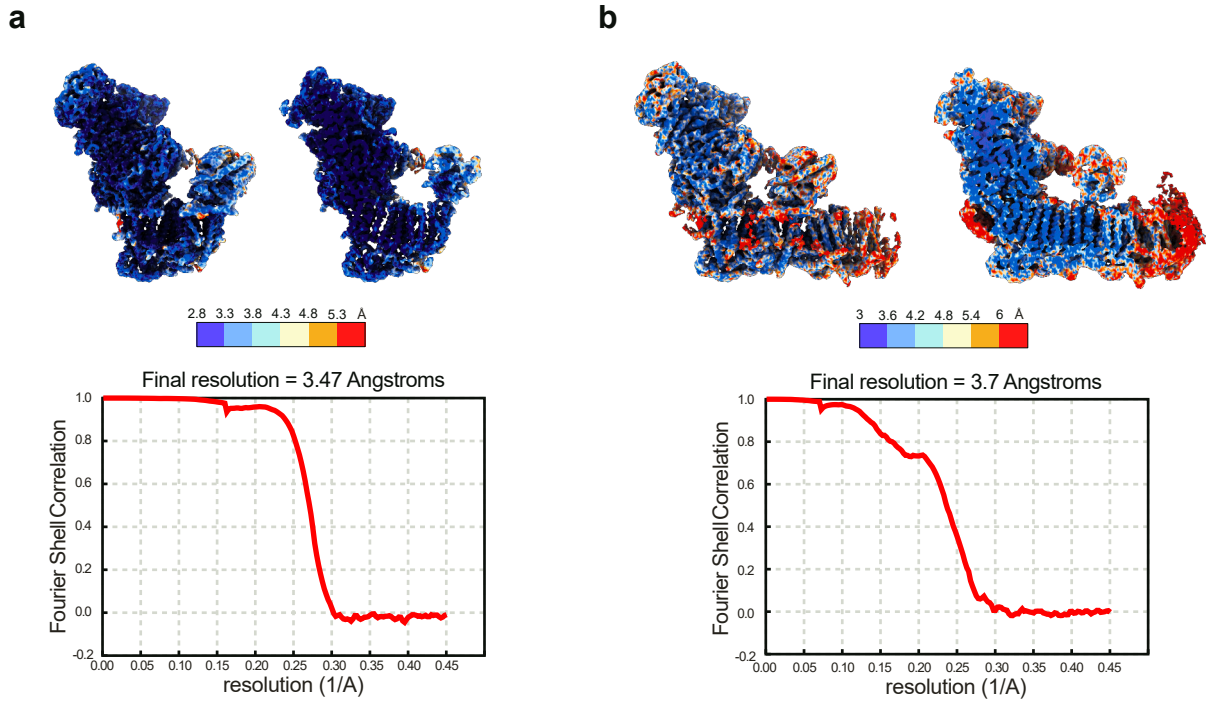
Figure 4. Plant mitochondrial complex I specific subunits

a Schematic representation of the five major eukaryote lineages. Organisms where γ -carbonic anhydrase is hypothesized to be part of mitochondrial complex I are listed next to their respective lineage. The red cross indicate the loss of mitochondrial γ -carbonic anhydrase in opisthokonta. **b** Comparison of mitochondrial complex I structures. Core subunits are shown in light gray and supernumary subunits in light blue. Plant specific proteins are highlighted in red-orange colours. *Y. lipolytica* (PDB:6RFR) and mammalian (PDB:6G2J) complex I don't have the γ -carbonic anhydrase domain.



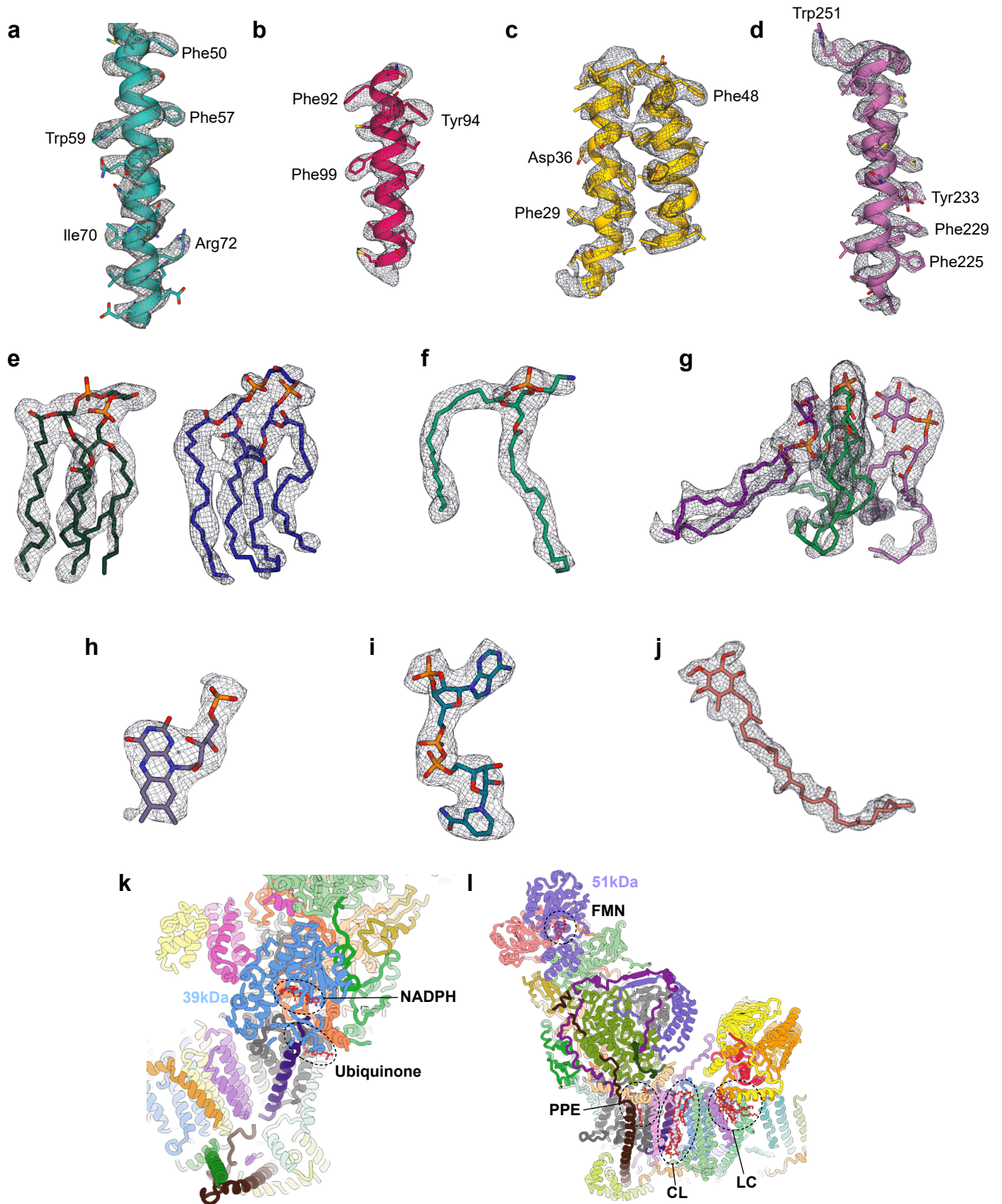
Extended Data Figure 1. Data processing workflow

Graphical summary of the processing workflow described in Methods, with 2D classes presented in **a** and 3D processing and refinement presented in **b**. Density in purple represent the mature and focused reconstructions, density in blue represent the assembly intermediate. The green density maps correspond to a complex where the whole P_D module is shifted, but that complex was not further analyzed.



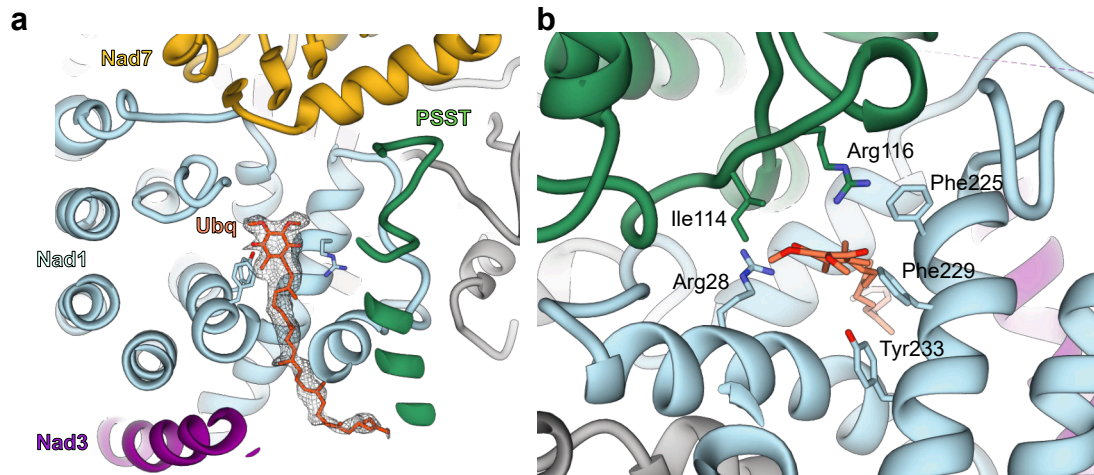
Extended Data Figure 2. Local resolution of the final reconstructions

Local resolutions of both the full **a** and focused **b** complex I are shown. The maps are colored by resolution, generated using ResMap³⁰. Maps are also shown in cut view. For both reconstructions FSC plots are displayed for resolution estimation. The map resolution is calculated at the 0.143 threshold.



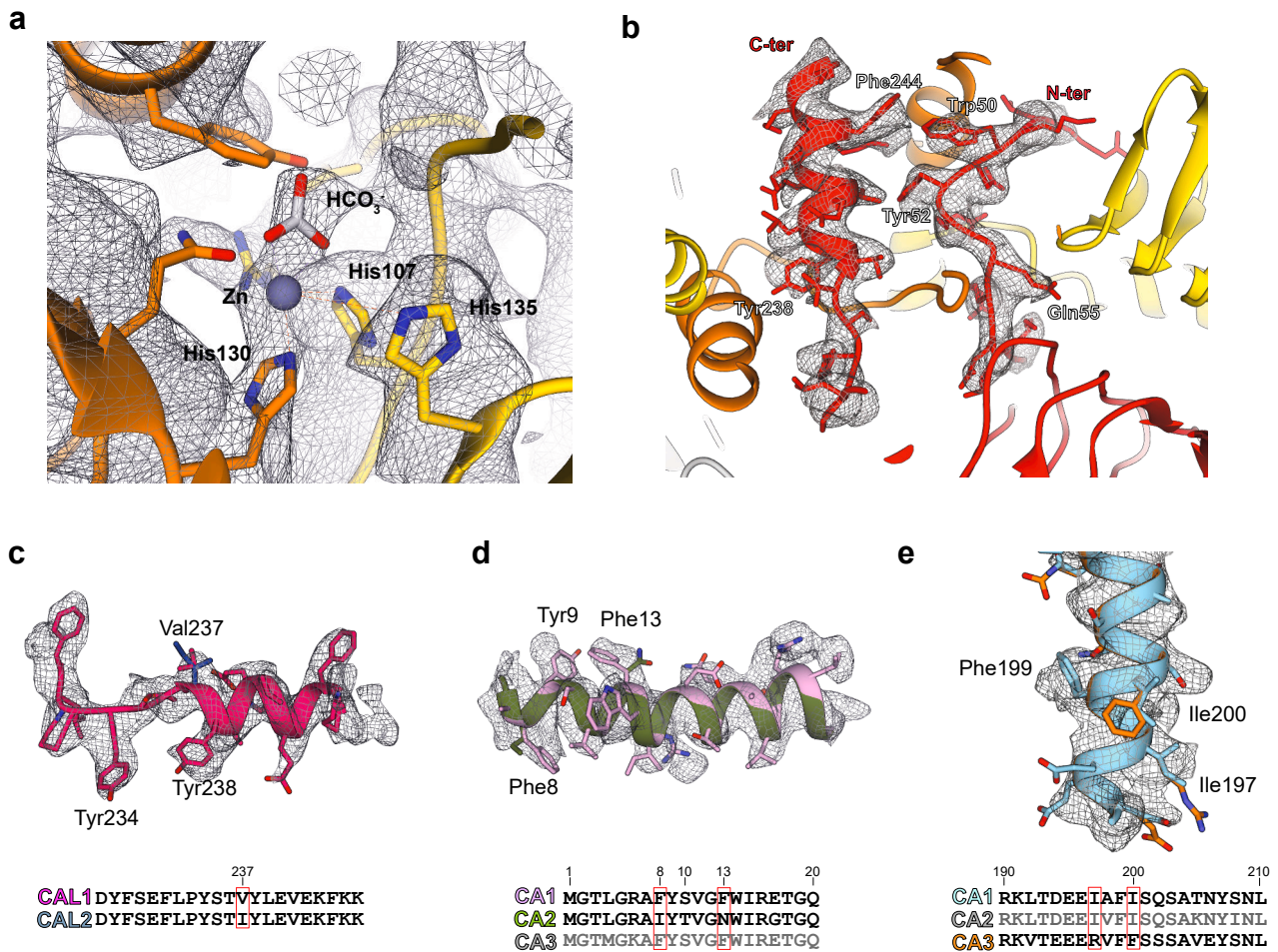
Extended Data Figure 3. Representative cryo-EM densities

a-d Proteins densities with **a** segment of protein B16.6, **b** segment of protein Nad2, **c** segment of protein Nad6 and **d** segment of protein Nad1. **e-f** Lipids identified in their respective densities with **e** cardiolipins, **f** phosphatidylethanolamine and **g** the carbonic anhydrase lipid complex with the cardiolipin in green and phosphatidylinositols in purple and orchid. **h-j** Additional ligands with **h** FMN of 51kDa subunit, **i** NADPH of the 39kDa subunit and **j** the ubiquinone. Positions of the ligands are indicated on the atomic model on **k** and **l** with PPE for phosphatidylethanolamine, CL for cardiolipin and LC for lipid complex.



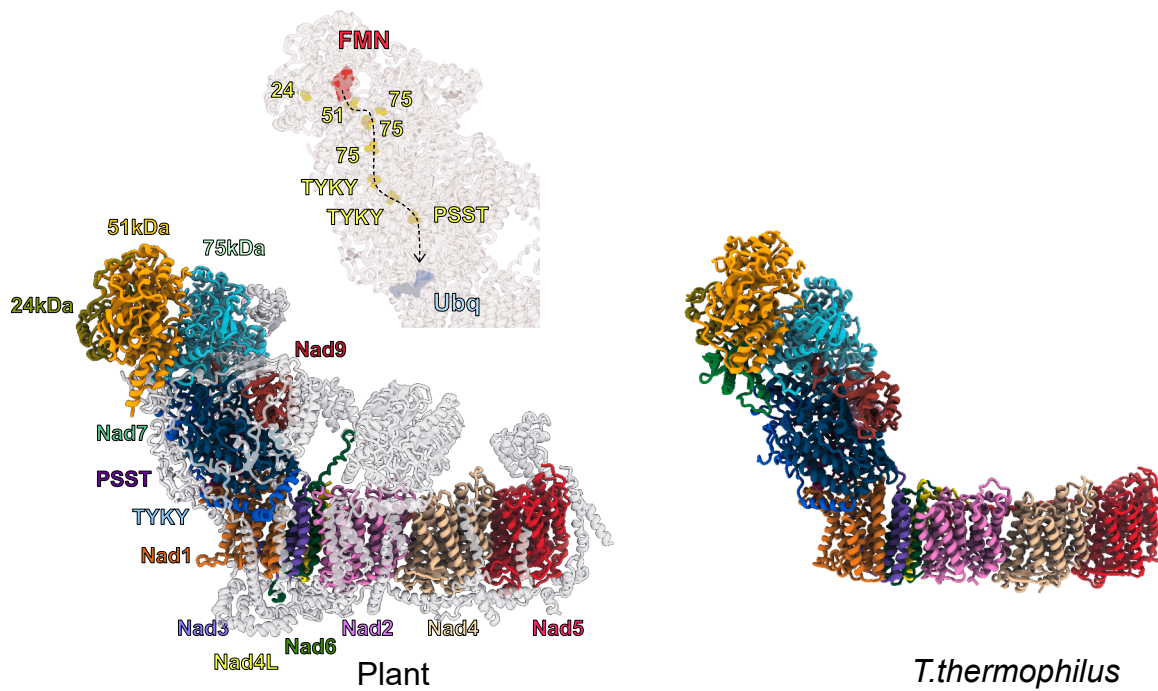
Extended Data Fig4 Ubiquinone binding site

a Overall view of the ubiquinone pocket mainly formed by Nad1 and PPST. The ubiquinone is shown in its density. **b** Focus on the ubiquinone head group and surrounding residues. The position and coordination of ubiquinone are similar to what has been observed in the *Y.lipolytica* mitochondrial complex I³.



Extended Data Figure 5. Identification of carbonic anhydrase subunits

a Density of the zinc and the HCO_3^- are shown, as well as the surrounding residues. The distinction between γCA and γCAL was done based on N and C-ter tails as well as conserved histidine residues, identification of the γCAL is shown in **b**. Models in their respective densities and sequence alignment are presented for the part that contributed to identify each of the subunit. Both possibilities are displayed and alignments are color-coded in accordance to the models. **c** To differentiate between γCAL1 and γCAL2 Val237 was used, which would be an isoleucine in γCAL2 . However, given the high similarity of the two proteins (only 5 different aa excluding the mitochondrial target sequence) and previous genetic screens¹⁹ both could be there. **d** For the γCA subunits, for chain p, residues Phe8 and Phe13 allowed to determine that it was γCA1 or 3 and not γCA2 . **e** Residue Ile200 and Ile197, which would be a Phe and an Arg in γCA3 , were used to distinguish between γCA1 and γCA3 , confirming that the subunit is indeed γCA1 . Still for the second γCA (chain q) the scant density did not allow to clearly distinguish between γCA1 and γCA2 , and was therefore built as γCA1 .



Extended Data Figure 6. Plant mitochondrial complex I comparison with prokaryotic complex I

Plant mitochondrial complex I is shown compared to its bacterial homologue (*T.thermophilus* model PDB:4HEA). The 14 core subunits are colored in different shades and their names are indicated. Mitochondria and plant specific proteins are shown in gray and transparent. Electron path in matrix arm from FMN to ubiquinone is also shown for the plant complex I.

MATERIAL & METHODS

Mitochondrial Complex I purification:

Cauliflower (*Brassica oleracea* var. botrytis) mitochondria were purified as previously described¹³. Quickly, fresh cauliflower inflorescence tissue was blended in extraction buffer containing 0.3 M mannitol, 30 mM sodium pyrophosphate (10.H₂O), 0.5 % BSA, 0.8 % (w/v) polyvinylpyrrolidone-25, 2mM beta-mercaptoethanol, 1 mM EDTA, 20 mM ascorbate and 5 mM cysteine, pH 7.5. Lysate was filtered and clarified by centrifugation at 1.500 g, 10 min at 4°C. Supernatant was kept and centrifuged at 18.000 g, 15 min at 4°C. Organelle pellet was re-suspended in wash buffer (0.3 M mannitol and 10 mM phosphate buffer, 1 mM EDTA, pH 7.5) and the precedent centrifugations were repeated once. The resulting organelle pellet was re-suspended in wash buffer and loaded on a single-step 30% Percoll gradient (in wash buffer without EDTA) and run for 1h30 at 40.000 g. Mitochondria were retrieved, washed two times and flash frozen in liquid nitrogen.

For complex I purification, mitochondria were re-suspended in Lysis buffer (20 mM HEPES-KOH, pH 7.6, 100 mM KCl, 1 mM DTT, 2% n-DDM, 1mM EDTA, supplemented with proteases inhibitors (Complete)) incubated for 15 min in 4°C. Lysate was clarified by centrifugation at 30.000 g, 20 min at 4°C. The supernatant was loaded on a 10-50% sucrose gradient buffer (20 mM HEPES-KOH, pH 7.6, 50 mM KCl, 1 mM DTT, 0.2% n-DDM, 1mM EDTA, supplemented with proteases inhibitors (Complete)) and run for 16 h at 75,000 g. The fraction corresponding to complex I was collected, pelleted and re-suspended in final resuspension buffer (same as sucrose gradient buffer without sucrose and only 0.1% n-DDM).

Grid preparation

4 μ L of the samples at a concentration of 1 μ g/ μ L was applied onto Quantifoil R2/2 300-mesh holey carbon grid, coated with thin home-made continuous carbon film and glow-discharged. The sample was incubated on the grid for 30 sec and then blotted with filter paper for 2.5 sec in a temperature and humidity controlled Vitrobot Mark IV (T = 4°C, humidity 100%, blot force 5) followed by vitrification in liquid ethane.

Single particle cryo-electron microscopy data collection

Data collection was performed on a Talos Arctica instrument (ThermoFisher Company) at 200 kV using the SerialEM software for automated data acquisition. Data were collected at a nominal underfocus of -0.5 to -2.5 μ m at a magnification of 36,000 X yielding a pixel size of 1.2 Å. Micrographs were recorded as movie stack on a K2 direct electron detector (GATAN Company), each movie stack were fractionated into 65 frames for a total exposure of 1 sec corresponding to an electron dose of 45 e⁻/Å².

Electron microscopy image processing

Drift and gain correction and dose weighting were performed using MotionCorr³¹. A dose weighted average image of the whole stack was used to determine the contrast transfer function with the software Gctf³². The following process has been achieved using RELION 3.0¹⁵. Particles were picked using the general model of crYOLO 1.5³³. Output box files from crYOLO were imported in RELION and particles were extracted. After 2D classification, 650,844 particles were extracted with a box size of 360 pixels and binned threefold, resulting in a 120 pixels box size four fold for 3D classification into 6 classes for each subunits (Extended Data Fig. 1). One subclass depicting high-resolution features have been selected for the mature complex I refinement with 65,018 particles. After Bayesian polishing a focused refinement has been performed using mask excluding the P_D module yielding respectively 3.7 and 3.47Å resolution. A second subclass containing the maturation factor GLDH composed of 155,644 particles was selected for further focused 3D classification on the GLDH area. The 36,513 particles of GLDH enriched subclass were extracted for refinement and reached 3.8Å. Determination of the local resolution of the final density map was performed using ResMap³⁰.

Structure building and model refinement

The atomic model of the plant mitochondrial complex I was built into the high-resolution maps using Coot, Phenix and Chimera. Atomic models from *Y.lipolytica* (6RFR) and *M.musculus* (6G2J) were used as starting points for protein identification and modelisation. Similarly to Waltz & Soufari 2019¹³, as the genome of cauliflower is not sequenced, and the closest fully sequenced member of the family (*Brassica oleracea* subsp. oleracea) is poorly annotated, to facilitate comprehension and analysis we positioned Arabidopsis proteins in the cauliflower map. The online SWISS-MODEL service was used to generate initial models for bacterial and mitochondria conserved r-proteins. Models were then rigid body fitted to

the density in Chimera³⁴ and all subsequent modeling was done in Coot³⁵. The global atomic model was refined with VMD using the Molecular Dynamic Flexible Fitting (MDFF) then with PHENIX using a combination of real and reciprocal space refinement for proteins.

Proteomic and statistical analyses of mitochondrial ribosome composition

Mass spectrometry analyses of the ribosome fractions were performed at the Strasbourg-Esplanade proteomic platform and performed as previously³⁶. In brief, proteins were trypsin digested, mass spectrometry analyses and quantitative proteomics were carried out by nano LC-ESI-MS/MS analysis on AB Sciex TripleTOF mass spectrometers and quantitative label-free analysis was performed through in-house bioinformatics pipelines. Data were searched against the TAIR *A. thaliana* database with a decoy strategy (release TAIRv10, 27281 forward protein sequences), and home-made cauliflower database extracted from UniProtKB (Swissprot+TrEMBL) including Brassica sub-taxonomy.

Figure preparation

Figures featuring cryo-EM densities as well as atomic models were visualized with UCSF ChimeraX³⁷ and Chimera³⁴.

References

1. Gray, M. W. Mitochondrial Evolution. *Cold Spring Harb. Perspect. Biol.* **4**, a011403–a011403 (2012).
2. Sazanov, L. A. A giant molecular proton pump: structure and mechanism of respiratory complex I. *Nat. Rev. Mol. Cell Biol.* **16**, 375–388 (2015).
3. Parey, K. *et al.* High-resolution cryo-EM structures of respiratory complex I: Mechanism, assembly, and disease. *Sci. Adv.* **5**, eaax9484 (2019).
4. Agip, A.-N. A. *et al.* Cryo-EM structures of complex I from mouse heart mitochondria in two biochemically defined states. *Nat. Struct. Mol. Biol.* **25**, 548–556 (2018).
5. Fiedorczuk, K. *et al.* Atomic structure of the entire mammalian mitochondrial complex I. *Nature* **538**, 406–410 (2016).
6. Zhu, J., Vinothkumar, K. R. & Hirst, J. Structure of mammalian respiratory complex I. *Nature* **536**, 354–358 (2016).
7. Baradaran, R., Berrisford, J. M., Minhas, G. S. & Sazanov, L. A. Crystal structure of the entire respiratory complex I. *Nature* **494**, 443–448 (2013).
8. Gu, J. *et al.* The architecture of the mammalian respirasome. *Nature* **537**, 639–643 (2016).
9. Parey, K. *et al.* Cryo-EM structure of respiratory complex I at work. *Elife* **7**, (2018).
10. Eubel, H., Heinemeyer, J. & Braun, H.-P. Identification and characterization of respirasomes in potato mitochondria. *Plant Physiol.* **134**, 1450–9 (2004).
11. Dudkina, N. V., Eubel, H., Keegstra, W., Boekema, E. J. & Braun, H.-P. Structure of a mitochondrial supercomplex formed by respiratory-chain complexes I and III. *Proc. Natl. Acad. Sci.* **102**, 3225–3229 (2005).
12. Sunderhaus, S. *et al.* Carbonic Anhydrase Subunits Form a Matrix-exposed Domain Attached to the Membrane Arm of Mitochondrial Complex I in Plants. *J. Biol. Chem.* **281**, 6482–6488 (2006).
13. Waltz, F., Soufari, H., Bochler, A., Giegé, P. & Hashem, Y. Cryo-EM structure of the RNA-rich plant mitochondrial ribosome. *bioRxiv* 777342 (2019). doi:10.1101/777342
14. Schimmeyer, J., Bock, R. & Meyer, E. H. I-Galactono-1,4-lactone dehydrogenase is an assembly factor of the membrane arm of mitochondrial complex I in Arabidopsis. *Plant Mol. Biol.* **90**, 117–126 (2016).
15. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, (2018).
16. Letts, J. A., Fiedorczuk, K. & Sazanov, L. A. The architecture of respiratory supercomplexes. *Nature* **537**, 644–648 (2016).
17. Momayyezi, M., McKown, A. D., Bell, S. C. S. & Guy, R. D. Emerging roles for carbonic anhydrase in mesophyll conductance and photosynthesis. *Plant J.* tpj.14638 (2020). doi:10.1111/tpj.14638
18. Parisi, G. *et al.* Gamma carbonic anhydrases in plant mitochondria. *Plant Mol. Biol.* **55**, 193–207 (2004).
19. Fromm, S., Senkler, J., Zabaleta, E., Peterhänsel, C. & Braun, H.-P. The carbonic anhydrase domain of plant mitochondrial complex I. *Physiol. Plant.* **157**, 289–296 (2016).
20. Ligas, J., Pineau, E., Bock, R., Huynen, M. A. & Meyer, E. H. The assembly pathway of complex I in Arabidopsis thaliana. *Plant J.* **97**, 447–459 (2019).
21. Schuller, J. M. *et al.* Redox-coupled proton pumping drives carbon concentration in the photosynthetic complex I. *Nat. Commun.* **11**, 494 (2020).
22. Jussupow, A., Di Luca, A. & Kaila, V. R. I. How cardiolipin modulates the dynamics of respiratory complex I. *Sci. Adv.* **5**, eaav1850 (2019).
23. Millar, A. H. *et al.* Control of ascorbate synthesis by respiration and its implications for stress responses. *Plant Physiol.* **133**, 443–7 (2003).
24. Pineau, B., Layoune, O., Danon, A. & De Paepe, R. I-Galactono-1,4-lactone Dehydrogenase Is Required for the Accumulation of Plant Respiratory Complex I. *J. Biol. Chem.* **283**, 32500–32505 (2008).
25. Davies, K. M., Blum, T. B. & Kühlbrandt, W. Conserved in situ arrangement of complex I and III₂ in mitochondrial respiratory chain supercomplexes of mammals, yeast, and plants. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 3024–3029 (2018).
26. Subrahmanian, N., Remacle, C. & Hamel, P. P. Plant mitochondrial Complex I composition and assembly: A review. *Biochim. Biophys. Acta - Bioenerg.* **1857**, 1001–1014 (2016).
27. Klodmann, J., Sunderhaus, S., Nitz, M., Jansch, L. & Braun, H. P. Internal Architecture of Mitochondrial Complex I from Arabidopsis thaliana. *Plant Cell* **22**, 797–810 (2010).
28. Miranda-Astudillo, H. V. *et al.* The atypical subunit composition of respiratory complexes I and

- IV is associated with original extra structural domains in *Euglena gracilis*. *Sci. Rep.* **8**, 9698 (2018).
29. Gawryluk, R. M. R. & Gray, M. W. Evidence for an early evolutionary emergence of gamma-type carbonic anhydrases as components of mitochondrial respiratory complex I. *BMC Evol. Biol.* **10**, 176 (2010).
 30. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
 31. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
 32. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
 33. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).
 34. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–12 (2004).
 35. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
 36. Waltz, F. *et al.* Small is big in *Arabidopsis* mitochondrial ribosome. *Nat. Plants* **5**, 106–117 (2019).
 37. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).