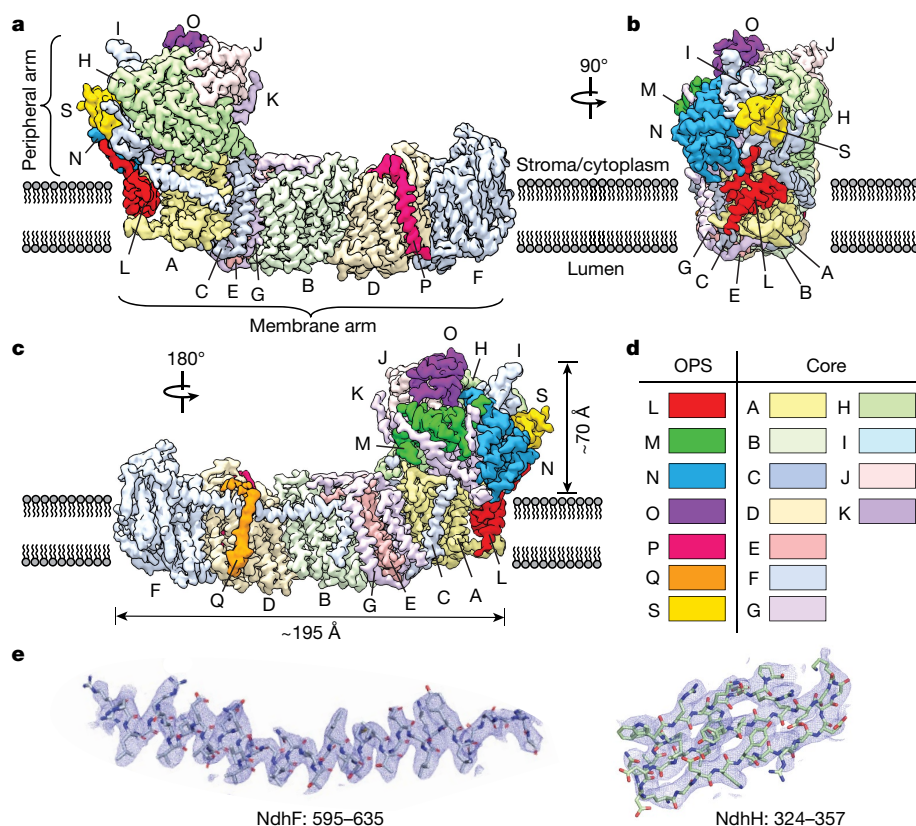# LETTER

Structure of the complex I–like molecule NDH of oxygenic photosynthesis

Thomas G. Laughlin[1,2], Andrew N. Bayne[3,4], Jean-François Trempe[3,4], David F. Savage[1] & Karen M. Davies[1,2]*

Cyclic electron flow around photosystem I (PSI) is a mechanism by which photosynthetic organisms balance the levels of ATP and NADPH necessary for efficient photosynthesis[1,2]. NAD(P)H dehydrogenase-like complex (NDH) is a key component of this pathway in most oxygenic photosynthetic organisms[3,4] and is the last large photosynthetic membrane-protein complex for which the structure remains unknown. Related to the respiratory NADH dehydrogenase complex (complex I), NDH transfers electrons originating from PSI to the plastoquinone pool while pumping protons across the thylakoid membrane, thereby increasing the amount of ATP produced per NADP$^+$ molecule reduced[4,5]. NDH possesses 11 of the 14 core complex I subunits, as well as several oxygenic-photosynthesis-specific (OPS) subunits that are conserved from cyanobacteria to plants[3,6]. However, the three core complex I subunits that are involved in accepting electrons from NAD(P) H are notably absent in NDH[3,5,6], and it is therefore not clear how NDH acquires and transfers electrons to plastoquinone. It is proposed that the OPS subunits—specifically NdhS—enable NDH

to accept electrons from its electron donor, ferredoxin[3–5,7]. Here we report a 3.1 Å structure of the 0.42-MDa NDH complex from the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1, obtained by single-particle cryo-electron microscopy. Our maps reveal the structure and arrangement of the principal OPS subunits in the NDH complex, as well as an unexpected cofactor close to the plastoquinone-binding site in the peripheral arm. The location of the OPS subunits supports a role in electron transfer and defines two potential ferredoxin-binding sites at the apex of the peripheral arm. These results suggest that NDH could possess several electron transfer routes, which would serve to maximize plastoquinone reduction and avoid deleterious off-target chemistry of the semi-plastoquinone radical.

To investigate the role of the OPS subunits in electron transfer through NDH, we performed single-particle cryo-electron microscopy (cryo-EM) on the principal NDH complex of *T. elongatus* (NDH-1L/ NDH-1$_1$)[3,6]. The *T. elongatus* complex represents the fundamental NDH architecture, which is conserved from cyanobacteria to

Fig. 1 | Cryo-EM density of the *T. elongatus* NDH. a–c, Different views of the overall cryo-EM reconstruction of the *T. elongatus* NDH complex segmented by subunit (threshold 0.025), coloured as indicated by the key in **d**, and labelled by one-letter subunit code. Relative orientations are indicated and the membrane is depicted schematically. The map was segmented 3 Å about the model and filtered to a Fourier shell correlation of 0.143 and left unsharpened (*B* = 0). Densities for NdhC, NdhF and NdhQ are from the second dataset map, all others are from the first dataset. The density is representative of reconstruction from the two datasets (*n* = 2). **d**, Colour key defining OPS and homologous core subunits. **e**, Sample fits-to-density (σ = 3) of the final model into the sharpened composite map of an example α-helix (left) and β-sheet (right).
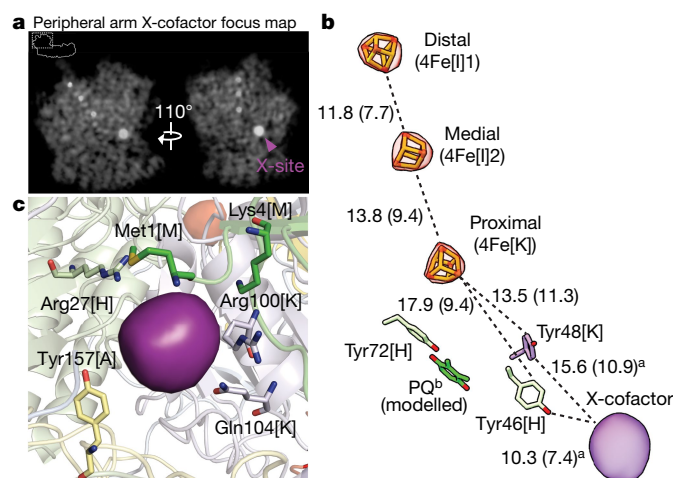
[1]Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA. [2]Molecular Biophysics and Integrative Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [3]Groupe d'étude des protéines membranaires (GEPROM), Montréal, Quebec, Canada. [4]Department of Pharmacology and Therapeutics, McGill University, Montréal, Quebec, Canada. *e-mail: KMDavies@lbl.gov
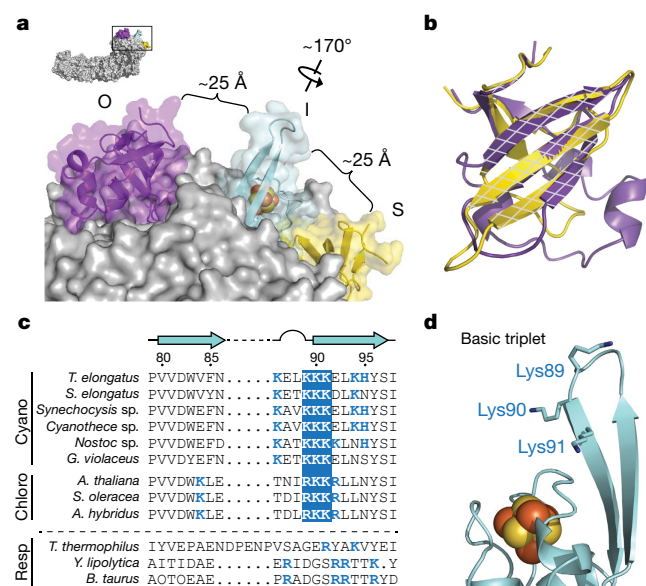
Fig. 2 | Features of OPS subunits NdhO and NdhS and the β-hairpin of core subunit NdhI. a, The main image is a surface model of the boxed region in the inset. Transparent surfaces of NdhI (blue), NdhO (purple) and NdhS (yellow) are shown along with cartoon atomic models. Distances shown are approximate from the nearest edge of NdhO and NdhS to the NdhI β-hairpin. b, Superposition of NdhO and NdhS. Root mean square deviation (r.m.s.d.), 0.92 Å. The hash pattern denotes the surface facing the O- and S-site, respectively. c, Alignment of NdhI homologues from cyanobacterial (cyano), chloroplastic (chloro), and respiratory (resp) complexes, with basic residues in bold blue text and the conserved basic triplet as white text with blue background. The *T. elongatus* secondary structure and numbering is indicated above the alignment. Species and/or strains shown are *Synechococcus elongatus* PCC 7942, *Synechocystis* sp. PCC 6803, *Cyanothece* sp. ATCC 51142, *Nostoc* (*Anabaena*) sp. PCC 7120, *Gloeobacter violaceus* PCC 7421, *Arabidopsis thaliana*, *Spinacia oleracea*, *Amaranthus hybridus*, *Thermus thermophilus*, *Yarrowia lipolytica* and *Bos taurus*. d, NdhI β-hairpin cartoon model with the basic triplet shown as sticks.

Fig. 3 | Arrangement of redox chain and cofactors within the peripheral arm. a, Solid representation of the X-cofactor focused peripheral arm density map from dataset 1, showing strong peaks for the three homologous core [4Fe-4S] clusters and the X-cofactor at the 'X-site' (indicated by purple arrow). A similar result was obtained with dataset 2 (*n* = 2, Extended Data Fig. 2). b, Putative redox chain residues overlaid with transparent isosurface of X-cofactor focused peripheral arm density map (threshold 0.04) showing only the density for the [4Fe-4S] clusters (red) and X-cofactor (purple). Centre-to-centre and edge-to-edge (in parenthesis) distance indicated in Å between the clusters, possible bridging tyrosine residues and X-cofactor. [a]Distances to X-cofactor are for a pseudo-atom placed at the approximate centre (and edge) of the isosurface shown. The X-cofactor is 6.7 times the volume of the [4Fe-4S] clusters, as measured in Chimera. [b]Plastoquinone (PQ) headgroup position approximated based on coordinating Tyr72 of NdhH. c, Environment of the X-site with the X-cofactor density (purple surface), viewed from arrow in a. Residues within 8 Å of the X-cofactor centre are displayed as sticks.

chloroplasts[3,6,8–10]. The NDH complex was solubilized from thylakoids using detergent and purified by making use of the native 'His-loop' that is distinct to the NdhF(1) subunit of *T. elongatus* NDH-1L/NDH-1₁ (Methods, Extended Data Fig. 1). The identity of the complex was confirmed by mass spectrometry, through which all 11 core subunits (NdhA to NdhK; including NdhF(1)), six of the seven conserved OPS subunits (NdhL to NdhP, NdhS) and cyanobacterial-specific subunit NdhQ were detected (Extended Data Tables 1, 2). Only the conserved OPS subunit NdhV, which is weakly associated with NDH, was absent[11,12]. Single-particle analysis of two independently purified samples resulted in two maps of the overall complex, at 3.1 Å and 3.5 Å resolution (Fig. 1, Extended Data Figs. 2, 3). Both maps are highly consistent and show clear side-chain density for the majority of residues, thus enabling atomic modelling (Extended Data Fig. 3, Extended Data Table 3; see Methods for details). The distal region of the membrane arm is better resolved in the 3.5 Å map, whereas the remainder of the structure is better resolved in the 3.1 Å map (Extended Data Fig. 3c). Density was observed for all subunits except NdhV, which was also absent in the mass spectrometry data (Extended Data Tables 1, 2).

NDH adopts an L-shaped structure that is characteristic of respiratory NADH dehydrogenase complexes[8–10] (Fig. 1). The NDH complex possesses a membrane arm of approximately 195 Å in length, and a peripheral arm that extends only about 70 Å above the membrane, in contrast to the extension of around 135 Å that is observed in respiratory enzymes[8–10]. The shorter peripheral arm in the NDH complex is consistent with the absence of the NADH dehydrogenase domain subunits[13,14]. The 11 core subunits of NDH are positioned similarly to

their counterparts in the respiratory complex (Extended Data Fig. 4a–c, Supplementary Figs. 2–12). NdhB, NdhD and NdhF—homologous to the Mrp Na⁺/H⁺ antiporter-like subunits of complex I—are positioned at the distal end of the membrane arm. NdhA, NdhC, NdhE and NdhG form the proximal membrane-arm region, which links the antiporter-like subunits to the base of the peripheral arm. As is the case in the respiratory enzyme[8,10,15], these subunits form an additional proton-channel and couple plastoquinone reduction to the ion-pumping activities of the complex[4,8,10,15] (Extended Data Fig. 4d, e). The remaining four homologous subunits—NdhH, NdhI, NdhJ and NdhK—form the core of the peripheral arm. The structures of these four subunits are similar to those of their respiratory homologues, with the notable exception of NdhK, which possesses N- and C-terminal extensions that interface with OPS subunits (Extended Data Figs. 4b, 5a).

The seven observed OPS subunits are distributed throughout the NDH complex (Fig. 1). NdhL, NdhP and NdhQ are located in the membrane arm. NdhL consists of a transmembrane helix followed by two short helices—rich in phenylalanine and leucine residues—that are connected to each other by a sharp proline turn within the membrane and an extended C-terminal coil that reaches from NdhA to NdhI (Extended Data Fig. 5b). NdhP and NdhQ are single transmembrane helices and are adjacent to NdhD on opposite sides of the membrane arm. NdhP traps a lipid molecule at the NdhD/NdhF interface (Extended Data Fig. 5c), and NdhQ seems to stabilize the horizontal helix of NdhF (Extended Data Fig. 5d). Thus, the membrane OPS subunits probably stabilize adjacent elements of the complex, as has been observed in previous studies[14,16,17].

The remaining four OPS subunits (NdhM, NdhN, NdhO and NdhS) are located on the peripheral arm (Fig. 1). NdhM and NdhN are α/β proteins located on the side of the peripheral arm and form a series of interactions with the terminal extensions of NdhK, such as inter-subunit antiparallel β-sheets (Extended Data Fig. 5a). NdhS and NdhO are
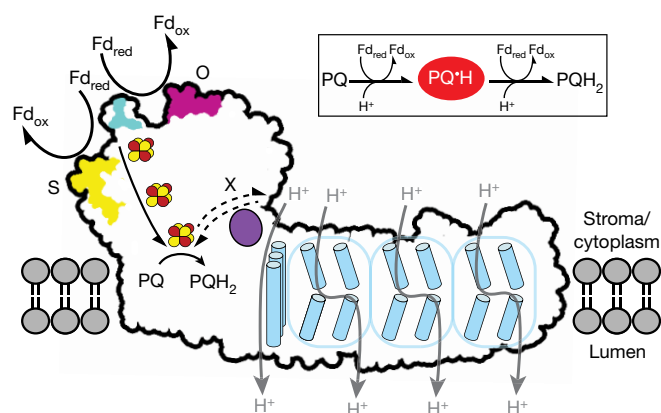
Fig. 4 | Summary of proposed models for the electron-transfer mechanism of NDH. Schematic of the NDH complex, coloured as follows: NdhO, magenta; NdhS, yellow; basic triplet of NdhI, cyan; X-cofactor, purple; and antiporter-like subunits, sky blue. The O-, S- and X-sites are labelled by single letters. Proton pumping in the membrane arm is coupled to reduction of the PQ to PQH_2 at the base of the peripheral arm. Single-barbed arrows indicate potential electron-transfer routes (solid for ferredoxin (Fd) and dashed for X-cofactor routes). The inset shows a schematic of PQ reduction catalysed by NDH using two individual donation events from ferredoxin, highlighting the radical intermediate ($PQ^•H$).

positioned at the apex of the peripheral arm and on opposing sides (related by a rotation of around 170°) of the entry-point iron–sulfur cluster in the core subunit NdhI (Fig. 2a). Their position and distance to the cluster (around 25 Å) suggest a possible role in facilitating redox chemistry.

Previous studies have revealed ferredoxin as the electron donor for the reduction of plastoquinone by NDH[4,5,7,18]. Ferredoxin is a [2Fe-2S] cluster protein with an approximate size of 10 Å × 25 Å, and is a soluble electron-carrier that leverages protein–protein interactions for targeted redox chemistry. Many proteins that interact with ferredoxin contain SH3-like domains[19,20] (for example, PsaE of PSI), and it has been shown that NdhS both adopts an SH3-like fold (PDB: 3C4S, unpublished) and interacts with ferredoxin[5,7,18]. Notably, our data show that NdhO also adopts an SH3-like fold, which is highly similar to that of NdhS (root mean square deviation of 0.92 Å, Fig. 2b); this raises the question of where exactly ferredoxin binds.

The surface of ferredoxin is highly acidic (pI < 4) and its recruitment is often driven by basic patches on the binding partner[21–23]. An electrostatic surface calculated for the peripheral arm (Extended Data Fig. 6a) shows two positively charged putative binding sites that encompass analogous β-strand elements of either NdhS (S-site) or NdhO (O-site), respectively. Previous work in *Arabidopsis* has implicated Arg69 (using *T. elongatus* numbering) of NdhS as forming part of a ferredoxin-binding site[5,18]. However, the S-site is a small, narrow cleft (about 10 Å wide) and Arg69 is largely inaccessible from the surface (Extended Data Fig. 6b) Therefore, if NdhS does bind ferredoxin, it is likely that an alternative surface or conformation would be involved.

The NdhO-adjacent site (O-site) is substantially wider (about 30 Å) and appears to be more suitable for ferredoxin binding (Extended Data Fig. 6c). A notable feature of the O-site is the β-hairpin of NdhI, which extends approximately 15 Å from the surface of the peripheral arm above the entry-point iron–sulfur cluster. This structural element contains a 'basic triplet' (Lys89-Lys90-Lys91), which is conserved across photosynthetic homologues but is absent in respiratory forms (Fig. 2c, d). Along with Lys4-Lys5 of NdhO, this basic triplet imparts the O-site with considerable positive charge, which is characteristic of a ferredoxin-binding site. It is therefore possible that NDH has two ferredoxin-binding sites.

Electron transfer through respiratory complex I involves a series of metal clusters in the peripheral arm[8–10,15]. Solid representation of the

NDH maps reveals four strong densities in the peripheral arm (Fig. 3a). Three of these densities (two in NdhI and one in NdhK) correlate with the final three [4Fe-4S] clusters of the electron-transfer pathway of the respiratory enzyme[8–10] (Extended Data Fig. 4a), which extends from the β-hairpin loop of NdhI to the predicted plastoquinone-coordinating Tyr72 in NdhH (Fig. 3b, Extended Data Fig. 4f). Furthermore, each of these three densities are coordinated by four cysteine residues, which are characteristic of the [4Fe-4S] clusters found in traditional redox pathways in these enzymes.

To our knowledge, the fourth density has never been observed in any homologous structure. This density—which we term the X-cofactor owing to its unknown composition—is at the interface of core subunits NdhA, NdhH, NdhK and OPS subunit NdhM. We term this the 'X-site', and it seems to be solvent-exposed (Fig. 3c, Extended Data Fig. 2). Although the X-cofactor is more than 25 Å from the plastoquinone-proximal cluster of NdhK (Fig. 3b), nearby aromatic residues Tyr46 of NdhH and Tyr48 of NdhK could connect the X-cofactor and the terminal iron–sulfur cluster by acting as an electron-transfer conduit, as has been observed in other redox complexes[24,25]. Focused classification on the X-site in either of the two datasets revealed that only about 30% of particles contained the X-cofactor; however, despite a labile nature, the peak is substantial (Fig. 3a, Extended Data Fig. 2; see Methods for processing details). The local environment around the X-cofactor is composed mainly of hydrophilic residues, including a number of positively charged groups (Fig. 2c). Despite the uncharacteristic local environment, the strong intensity for the X-cofactor is consistent with a metallocofactor, and the location of the two tyrosine residues suggest that it could be involved in the transfer of electrons to plastoquinone.

A critical aspect of the NDH redox mechanism is that ferredoxin shuttles only one electron; however, two donation events are required for the complete reduction of plastoquinone to PQH_2, which proceeds through a semiquinone radical intermediate ($PQ^•H$) (Fig. 4). This is in contrast to complex I, which accepts two electrons from a single NADH molecule to complete its redox mechanism[15]. The plastoquinone site in our NDH structure is similar to the homologous Q-site of complex I, which has been proposed to be capable of stabilizing the semiquinone radical intermediate[9,15]. Therefore, although the mechanism for the stabilization of the radical intermediate may be conserved, our structure reveals how features of the OPS subunits that are distinct to NDH might further mitigate the deleterious effects of a $PQ^•H$ radical. First, NdhS and NdhO may define two binding sites on NDH, enabling more rapid, successive donation events than a single-site mechanism and thus maximizing the complete reduction of plastoquinone to PQH_2. Second, the X-cofactor may participate in redox activity and act as an electron buffer, abstracting an electron from a long-lived $PQ^•H$ and returning the electron when $PQ^•H$ is reformed by a second ferredoxin-donation event, thus completing the reduction to PQH_2. Third, the X-cofactor may provide an alternative point of entry for electrons, thereby enabling the parallel reduction of plastoquinone from two distinct electron-donation pathways. The atomic model presented here provides a framework for future functional studies to discern amongst these putative mechanisms.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-0921-0.

1. Kramer, D. M. & Evans, J. R. The importance of energy balance in improving photosynthetic productivity. *Plant Physiol.* **155**, 70–78 (2011).
2. Arnon, D. I. The light reactions of photosynthesis. *Proc. Natl Acad. Sci. USA* **68**, 2883–2892 (1971).
3. Shikanai, T. Chloroplast NDH: A different enzyme with a structure similar to that of respiratory NADH dehydrogenase. *Biochim. Biophys. Acta* **1857**, 1015–1022 (2016).

4. Strand, D. D., Fisher, N. & Kramer, D. M. The higher plant plastid NAD(P)H dehydrogenase-like complex (NDH) is a high efficiency proton pump that increases ATP production by cyclic electron flow. *J. Biol. Chem.* **292**, 11850–11860 (2017).
5. Yamamoto, H., Peng, L., Fukao, Y. & Shikanai, T. An Src homology 3 domain-like fold protein forms a ferredoxin binding site for the chloroplast NADH dehydrogenase-like complex in *Arabidopsis*. *Plant Cell* **23**, 1480–1493 (2011).
6. Peltier, G., Aro, E.-M. & Shikanai, T. NDH-1 and NDH-2 plastoquinone reductases in oxygenic photosynthesis. *Annu. Rev. Plant Biol.* **67**, 55–80 (2016).
7. He, Z. et al. NDH-1L interacts with ferredoxin via the subunit NdhS in *Thermosynechococcus elongatus*. *Photosynth. Res.* **126**, 341–349 (2015).
8. Baradaran, R., Berrisford, J. M., Minhas, G. S. & Sazanov, L. A. Crystal structure of the entire respiratory complex I. *Nature* **494**, 443–448 (2013).
9. Zickermann, V. et al. Mechanistic insight from the crystal structure of mitochondrial complex I. *Science* **347**, 44–49 (2015).
10. Zhu, J., Vinothkumar, K. R. & Hirst, J. Structure of mammalian respiratory complex I. *Nature* **536**, 354–358 (2016).
11. Chen, X., He, Z., Xu, M., Peng, L. & Mi, H. NdhV subunit regulates the activity of type-1 NAD(P)H dehydrogenase under high light conditions in cyanobacterium *Synechocystis* sp. PCC 6803. *Sci. Rep.* **6**, 28361 (2016).
12. Fan, X., Zhang, J., Li, W. & Peng, L. The NdhV subunit is required to stabilize the chloroplast NADH dehydrogenase-like complex in *Arabidopsis*. *Plant J.* **82**, 221–231 (2015).
13. Arteni, A. A. et al. Structural characterization of NDH-1 complexes of *Thermosynechococcus elongatus* by single particle electron microscopy. *Biochim. Biophys. Acta* **1757**, 1469–1475 (2006).
14. Wulfhorst, H., Franken, L. E., Wessinghage, T., Boekema, E. J. & Nowaczyk, M. M. The 5 kDa protein NdhP is essential for stable NDH-1L assembly in *Thermosynechococcus elongatus*. *PLoS ONE* **9**, e103584 (2014).
15. Wirth, C., Brandt, U., Hunte, C. & Zickermann, V. Structure and function of mitochondrial complex I. *Biochim. Biophys. Acta* **1857**, 902–914 (2016).
16. Zhang, J. et al. NdhP is an exclusive subunit of large complex of NADPH dehydrogenase essential to stabilize the complex in *Synechocystis* sp. strain PCC 6803. *J. Biol. Chem.* **289**, 18770–18781 (2014).
17. Shimizu, H. et al. CRR23/NdhL is a subunit of the chloroplast NAD(P)H dehydrogenase complex in *Arabidopsis*. *Plant Cell Physiol.* **49**, 835–842 (2008).
18. Yamamoto, H. & Shikanai, T. *In planta* mutagenesis of Src homology 3 domain-like fold of NdhS, a ferredoxin-binding subunit of the chloroplast NADH dehydrogenase-like complex in *Arabidopsis*: a conserved Arg-193 plays a critical role in ferredoxin binding. *J. Biol. Chem.* **288**, 36328–36337 (2013).
19. Kubota-Kawai, H. et al. X-ray structure of an asymmetrical trimeric ferredoxin–photosystem I complex. *Nat. Plants* **4**, 218–224 (2018).
20. Veit, S. et al. The cyanobacterial cytochrome b_6f subunit PetP adopts an SH3 fold in solution. *Biochim. Biophys. Acta* **1857**, 705–714 (2016).
21. Kurisu, G. et al. Structure of the electron transfer complex between ferredoxin and ferredoxin-NADP+ reductase. *Nat. Struct. Biol.* **8**, 117–121 (2001).
22. Srivastava, A. P. et al. Identification of the ferredoxin-binding site of a ferredoxin-dependent cyanobacterial nitrate reductase. *Biochemistry* **56**, 5582–5592 (2017).
23. Dai, S. et al. Structural snapshots along the reaction pathway of ferredoxin-thioredoxin reductase. *Nature* **448**, 92–96 (2007).
24. Ishikita, H. & Knapp, E.-W. Function of redox-active tyrosine in photosystem II. *Biophys. J.* **90**, 3886–3896 (2006).
25. Ekberg, M., Sahlin, M., Eriksson, M. & Sjöberg, B. M. Two conserved tyrosine residues in protein R1 participate in an intermolecular electron transfer in ribonucleotide reductase. *J. Biol. Chem.* **271**, 20655–20659 (1996).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Purification of NDH from *T. elongatus*. NDH (NDH-1L/NDH-1$_1$) was purified from *T. elongatus* BP-1 by Ni^{2+}-affinity and size-exclusion chromatography using the native histidine-rich loop within the most distal antiporter-like subunit NdhF1 of the NDH-1L/NDH-1$_1$ complex[26]. *T. elongatus* membranes were prepared as previously described[27] and frozen at −80 °C. All subsequent steps were performed on ice or at 4 °C unless otherwise specified. Before extraction, membranes were thawed and washed twice with buffer S (20 mM Bis-Tris, pH 6.0, 100 mM NaCl, 20% v/v glycerol, 1 mM PMSF, one cOmplete, EDTA-free Protease Inhibitor Cocktail per 50 ml) by resuspension and centrifugation at 25,000g for 30 min. Washed membranes were then resuspended at approximately 4 mg protein per ml in buffer S, mixed 1:1 with buffer S containing 2% (w/v) β-dodecyl-maltoside (β-DDM, Anatrace), and gently agitated for 30 min. Insoluble material was removed by centrifugation at 165,000g for 45 min and the supernatant passed through a 0.45 μm filter. Filtered, solubilized material was bound in batch to Ni-Sepharose FF resin (GE Healthcare) for 1 h. The resin was then washed with buffer N (20 mM Bis-Tris, pH 6.0, 100 mM NaCl, 10% v/v glycerol, 0.03% β-DDM) in a gravity-flow column until no protein was detected by a Bradford assay[28]. The bound protein was subsequently eluted with buffer N + 100 mM imidazole. Fractions were pooled and concentrated in a 100-kDa-cutoff spin-concentrator (Vivaspin) to 10–12 mg ml^{-1}. Concentrated sample was then injected onto a Superose 6 10/300 GL column equilibrated in and ran isocratically with buffer G (20 mM Bis-Tris, pH 6.0, 100 mM NaCl, 0.03% β-DDM). Fractions containing NDH were determined by Blue-Native PAGE[29] and negative-stain transmission electron microscopy.

Mass spectrometry of purified NDH complex. NDH proteins (20 μg) obtained as described in the previous section were extracted using the methanol/chloroform precipitation method[30]. The pellet was resuspended in 10 μl of 8 M urea and the pH was adjusted by the addition of 5 μl triethylammonium carbonate 50 mM buffer at pH 8.5. Tris(2-carboxyethyl)phosphine was added to a final concentration of 10 mM and the mixture was incubated for 10 min at 50 °C to reduce disulfide bonds. Iodoacetamide was added to a final concentration of 50 mM followed by incubation at room temperature for 30 min in the dark. The sample was diluted to 80 μl with 50 mM triethylammonium carbonate pH 8.5 (final concentration 1 M urea). The sample was split into two portions and digested overnight at room temperature with 0.5 μg of either trypsin or chymotrypsin (Sigma, proteomics grade). The peptides were then purified using C18 spin columns (Thermo) using the manufacturer's protocol. The peptides were eluted in 2 × 40 μl 70% acetonitrile, dried in a Savant SPD1010 SpeedVac concentrator (Thermo) and resuspended in 24 μl of 4% acetonitrile/0.1% trifluoroacetic acid.

For the in-gel digested samples, two blue native PAGE bands corresponding to the NDH complex were isolated and processed in parallel using a standard in-gel digestion procedure, as previously described[31]. In brief, the gel pieces were excised, reduced with 10 mM DTT, alkylated with 55 mM iodoacetamide, destained and digested overnight with chymotrypsin at 37 °C. Chymotryptic peptides were extracted with acetonitrile, dried in a speed-vac and resuspended in 20 μl of 4% acetonitrile/0.1% trifluoroacetic acid.

Purified peptides (2.5 μg) were captured on a C18 μ-precolumn and eluted onto an Acclaim PepMap100 C18 column (75 μm × 15 cm) with a 1 h 5–40% gradient of acetonitrile in 0.1% formic acid at 300 nl min^{-1}. The eluted peptides were analysed with an Impact II Q-TOF spectrometer equipped with a CaptiveSpray electrospray source with an acetonitrile-enriched NanoBooster gas (Bruker). Data was acquired using data-dependent auto-MS/MS with an *m/z* range of 150–2,200, a fixed cycle time of 3 s, a dynamic exclusion of 1 min, an *m/z*-dependent isolation window of 1.5–5 *m/z* and collision energy in the range of 25–75 eV[32].

The raw data were processed using Andromeda, integrated into MaxQuant[33] (version 1.6.1.0). Specific searches with either trypsin (before K,R) or chymotrypsin (before F,Y,W,L,I,M) were performed against the *T. elongatus* (strain BP-1) proteome (2,451 proteins; ID: UP000000440; NdhP and NdhQ appended). Focused searches were also performed against only Ndh subunits. In both cases, common contaminants provided by MaxQuant were also included. Tryptic and chymotryptic peptides were permitted a maximum of two missed cleavages. All searches were performed with cysteine carbamidomethylation as a fixed modification, as well as N-acetylation and methionine oxidation as variable modifications. Default MaxQuant instrument parameters for Bruker Q-TOF data were used, including a first search peptide mass tolerance of 0.07 Da, main search peptide tolerance of 0.006 Da, and isotope match tolerance of 0.005 Da. The false discovery rate threshold was set to 1%.

Cryo-electron microscopy grid preparation and image acquisition. To prepare cryo-EM grids for dataset 1, 3 μl of purified NDH (around 35 nM) in buffer G was applied to R1/2 Cu 300 grids (Quantifoil) coated with a thin layer of home-made continuous carbon film and glow-discharged for 15 s immediately before use. The sample was applied to the grid and incubated for 30 s at 100% humidity and 22 °C, before blotting for 2–3 s with blot force 18 using Whatman no. 1 filter paper and plunging into liquid ethane cooled to liquid-nitrogen temperatures[34] using a FEI Vitrobot Mark IV. Dataset 2 grids were prepared in the same way, but at an estimated NDH concentration of around 80 nM and using R2/2 Cu 300 grids (Quantifoil).

Dataset 1 was acquired on a FEI Titan Krios G2 transmission electron microscope operated at 300 kV and equipped with a Gatan K2 Summit direct electron detector camera at the Bay Area Cryo-EM facility at the University of California, Berkeley. Movies were recorded in super-resolution counting mode at an effective pixel size of 0.534 Å, with a cumulative exposure of about 50 e$^-$ A^{-2} distributed uniformly over a 7-s exposure fractionated into 35 frames. Data acquisition was performed using SerialEM[35] with one exposure per hole and focusing for each exposure to a target defocus range of −1.3 μm to −3.0 μm.

Dataset 2 was acquired in a similar manner but with the following modifications. The microscope was operated in EFTEM mode with a Gatan Quantum LS Imaging Filter set to a slit width of 25 eV. Three movies per hole were recorded with a super-resolution pixel size of 0.4516 Å, a cumulative exposure of about 60 e$^-$ A^{-2} distributed uniformly over a 6-s exposure fractionated into 30 frames, and a target defocus range of −0.8 μm to −2.5 μm.

Image processing. For dataset 1, a total of 2,519 movie stacks were collected as described in the previous section. Preprocessing was performed during data collection within Focus[36]. Super-resolution movie stacks were drift-corrected in 5 × 5 patches, dose-weighted, and Fourier-cropped to 1.068 Å per pixel using MotionCor2-v1.0.0[37]. Contrast transfer function parameters were estimated from the non-dose-weighted, aligned micrographs using GCTF-v1.06[38]. The data was manually inspected and micrographs of empty carbon, excessive ice-contamination or poor power spectrum were removed, resulting in 2,192 micrographs that were used for further processing. All classification and reconstruction steps were performed using RELION-v2.1.0[39], unless otherwise specified. A total of 4,285 particles were picked manually from 100 micrographs and extracted in a 360^2 pixel box. These particles were subjected to reference-free 2D-classification to generate 2D references for the auto-picking of all micrographs using Gautomatch-v0.56 (https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch/). A total of 176,394 were extracted using a 360^2 pixel box and subjected to 3D-classification ($K = 5$) using a 60 Å low-pass-filtered ab initio reference generated in cryoSPARC[40]. Selected 3D classes were then subjected to reference-free 2D-classification with selected classes resulting in 119,592 particles. An additional round of 3D-classification ($K = 3$ or $K = 5$) did not yield classes with clearly distinguishable features. Particles were re-extracted and re-centred, and contrast transfer function parameters were refined locally. Unmasked refinement using all particles yielded a reconstruction at 4.1 Å. Local movie-refinement and particle-polishing improved the resolution to 4.0 Å, and inclusion of a soft mask during refinement further improved the resolution to 3.7 Å. All further classifications and focused refinements use soft masks around the regions of interest without image alignment and using the consensus refinement as a starting point.

To further improve the overall map, 3D classification of the entire complex was performed ($K = 3$, $T = 12$) and the most abundant class refined using a soft mask and local-searches to obtain a 3.7 Å overall map from 81,670 particles with improved features compared to the consensus refinement. The peripheral and membrane arms were each subjected to refinement with signal subtraction[41] of the opposite domain. The outputs from refinement were then further classified ($K = 3$, $T = 12$, E-step limit = 8 Å) and the dominant class in each case refined. This resulted in a noticeably improved map for the peripheral arm at 3.6 Å resolution and a slight improvement in the connectivity in areas of the membrane arm at an overall resolution of 3.8 Å. A map with marginally improved connectivity for NdhF was obtained after masked classification ($K = 3$, $T = 12$, E-step limit = 8 Å) and refinement of the entire membrane arm at 3.9 Å resolution. The map quality of regions suspected of partial-occupancy corresponding to NdhS and the X-site were improved by further classification ($K = 3$ for NdhS and $K = 5$ for X-cofactor, $T = 20$, E-step limit = 8 Å) and subsequent refinement of the dominant bound- and apo-states for the entire peripheral arm.

With regards to the processing of the X-site, the X-factor was initially observed in unmasked 3D-classification of the overall complex in dataset 1 at a similar intensity to the [4Fe-4S] cluster. After model-building into the surrounding protein density, characteristic residues for iron–sulfur clusters (for example, cysteine and histidine) were not present at this site. To further improve the density of this region, focused 3D classification on the X-site was performed using a diameter mask of approximately 45 Å centred about the X-site and softened with a cosine-edge of 8 voxels (around 8.5 Å), as depicted in Extended Data Fig. 2. The 3D classification was performed on membrane arm signal-subtracted particles without image alignment and using the peripheral arm map low-pass-filtered to 35 Å as the initial reference, $K = 5$, $T = 20$ and limiting the E-step limit = 8 Å. The X-cofactor class was selected for 3D auto-refinement using the peripheral arm map low-pass-filtered

to 35 Å as a reference, with a mask over the entire peripheral arm softened with a cosine-edge of 8 voxels (around 8.5 Å) and local angular searches.

During preparation of this manuscript, RELION-v3.0.b2[42] was released with a new CtfRefine function, which enabled estimation and correction of beam-tilt. CtfRefine was used for the overall, the X-cofactor-bound and the apo maps, which lead to final resolutions of 3.1 Å, 3.4 Å and 3.3 Å, respectively.

For dataset 2, a total of 1,385 movies were collected as described in the previous section and preprocessed through Focus as described for dataset 1, using the appropriate pixel size and exposure. After manual inspection, 1,364 micrographs were chosen for further processing. The 2D-references from dataset 1 were used for auto-picking using Gautomatch-v0.56 with appropriate scaling. All classification and reconstruction steps were performed using RELION-v3.0.b2. A total of 98,145 particles were extracted using a 448^2 pixel box with a binning factor of 3.5 (Nyquist limit (NQ) of 6.3 Å) and subjected to two rounds of reference-free 2D classification. Classes showing clear secondary structure features were selected, resulting in 90,266 particles. These particles were re-extracted using a 424^2 pixel box with a binning factor of 1.25 (NQ of 2.3 Å). Refinement using a 35 Å low-pass-filtered reference from dataset 1 resulted in a 4.1 Å map with density for an additional transmembrane helix (TMH) adjacent to NdhD and NdhF compared to dataset 1, which was later identified as NdhQ. All further refinement steps were performed using a soft mask and local searches about the consensus refinement. Subsequent refinement of both beam-tilt and per-particle defocus with CtfRefine and Bayesian polishing[43] resulted in improvements of 0.5 Å and 0.2 Å, respectively. Further 3D classification without image alignment ($K = 3$, $T = 12$) and refinement of the dominant class resulted in a 3.3 Å reconstruction from 66,760 particles. To improve the distal region of the membrane arm, focused classification was performed using a soft mask around the distal region of the membrane arm ($K = 3$, $T = 16$, E-step limit = 8). The top class displayed clear density for NdhQ and was selected for subsequent refinement, resulting in a 3.5 Å reconstruction of the overall complex from 34,655 particles. The X-cofactor map for this dataset was obtained as for dataset 1, but using three classes ($K = 3$) for the classification step.

All reported global map resolutions correspond to the 'gold-standard' Fourier shell correlation (FSC) of 0.143 between independently refined half-maps, with phase-randomization to correct for masking artefacts[44] and B-factors calculated as implemented in the post-processing function of RELION[45]. Local resolution maps were generated with RELION using default parameters. Directional FSC curves and map anisotropy were assessed using the 3DFSC[46]. For visualization, maps were corrected for the modulation transfer function of the detector and filtered to the global resolution.

Model building and refinement. In brief, the maps from the two datasets were aligned and resampled onto the same grid using UCSF Chimera-v1.11.2[47]. For the 11 core subunits (NdhA– NdhK), homology models were generated on the basis of the *T. thermophilus* complex I crystal structure[8] (PDB ID: 4HEA) using the I-TASSER webserver[48]. Secondary-structure predictions for all subunits were generated using Phyre2[49], PsiPred[50] and ESPript-3.0[51] webservers. All manual adjustment and building was performed using COOT-v0.8.8[52,53]. The homology models were adjusted and rebuilt into the maps. The remaining density was assigned to the novel OPS subunits. The initial models for the two datasets were then refined in PHENIX-v1.13[54] using *phenix.real_space_refine*[55] with auto-detection of secondary-structure assignments. The models and maps were then inspected to manually adjust secondary-structure assignments. This was followed by iterative rounds of real-space refinement in PHENIX and manual rebuilding in COOT. The 'Composite' map and model was then made from NdhC, NdhF and NdhQ density from dataset 2 and the remainder from dataset 1, using 'vop maximum' in Chimera after normalizing maps to a similar greyscale using 'vop rescale'. The final models were assessed by MolProbity-v4.4[56], EMRinger webserver[57] and map-vs-model FSC.

More specifically, initial modelling relied predominantly on the 3.7 Å overall complex and 3.6 Å peripheral arm maps from dataset 1 obtained using RELION-v2.1.0. Focused classifications and refinements were performed in conjunction with model building, which yielded the improved NdhS region in the 3.8 Å NdhS-focus map of the peripheral arm and marginally improved distal membrane arm region with the 3.9 Å membrane arm map. After placement and rebuilding of homology models for the core subunits, the remaining density was assigned to OPS subunits. The density corresponding to NdhS was initially identified by its all-β-strand secondary structure and a homology model of a crystal structure (PDB: 3C4S) of NdhS from a mesophilic cyanobacterium. NdhL was identified first on the basis of its predicted two TMHs. The density corresponding to a single TMH between NdhD and NdhF was suspected to be either NdhP or NdhQ, and was ultimately identified as NdhP on the basis of the fit of landmark aromatic residues. NdhO was identified on the basis of its predominantly β-strand structure. NdhM and NdhN were identified by testing primary sequence registers after tracing the main chains for these subunits. No density was observed for NdhQ in dataset 1, nor for residues 517–583 of NdhF, which were left unmodelled.

After initial model building and refinement against the 3.7 Å overall map, the cryo-EM data was reprocessed with estimated beam-tilt correction in the recently released RELION-v3.0.b2 to obtain the 3.1 Å overall map from dataset 1. The 3.7 Å refined model was consistent with the 3.1 Å map, but with improved side-chain density.

Dataset 2 was collected after model-building for dataset 1 using a sample from a different round of purification, and was processed entirely through RELION-v3.0.b2. Dataset 2 yielded a lower global resolution for the overall complex, but with an improved membrane arm (Extended Data Fig. 3c) that revealed density for NdhQ and NdhF residues 517–583.

In addition to the improved distal region of the membrane arm, dataset 2 shows differences in the N termini of NdhA and NdhC. For dataset 1, weak cylindrical density is present adjacent to the N terminus of NdhA. This density extends at a slant through the detergent belt from the lumenal to the stromal side of the complex. The density is consistent with an approximately 24-residue helix that is tentatively assigned as NdhA residues 4–27 in the dataset 1 model (PDB: 6NBQ), but left as UNK owing to the absence of clear side-chain density. A similar density is not observed in the dataset 2 map, therefore it was not included in the composite model (PDB: 6NBY) and may represent a detergent-belt artefact. Adjacent to this density, the NdhC N-terminal TMH (residues 20–43) also differs between the two datasets. The helix in dataset 2 (PDB: 6NBX) most closely resembles the position of its homologue Nqo7 in complex I (PDB: 4HEA), and appears almost upright. In dataset 1, the helix is tilted at an angle of approximately 55° towards the distal region of the membrane arm with respect to the pivot point of residue 20. Connecting density from 43–65 is not observed in either map and is predicted to be a disordered region. It is unclear if this difference in position is related to the NdhA N-terminal density. The position from dataset 2 was used for the composite model owing to its similarity to the homologue in complex I.

Visualization and analysis. Figures of maps and models were prepared with Chimera and PyMOL-v1.8.2[58]. Superposition of structure was performed in PyMOL using the 'super' command. The electrostatic surface representation was generated with the APBS-v2.0[59] plugin in PyMOL using the PARSE force field in PDB2PQR and pH = 8, to approximate the daytime stromal/cytosolic pH, in order to prepare the model for calculations. All alignments were performed using the Clustal Omega webserver[60] with sequences obtained from Uniprot[61].

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Cryo-EM maps for the dataset 1 overall, peripheral arm, with NdhS, without NdhS, with X-cofactor and without X-cofactor and dataset 2 overall have been deposited with the Electron Microscopy Data Bank under accession numbers EMD-0415, EMD-0416, EMD-0417, EMD-0418, EMD-0419, EMD-0420 and EMD-0425, respectively. Atomic coordinates for the overall complex for dataset 1, dataset 2 and composite have been deposited with the PDB under accession codes 6NBQ, 6NBX and 6NBY, respectively. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[62] partner repository with the dataset identifier PXD012206. All other data are available from the corresponding author upon reasonable request.

26. Zhang, P. et al. Isolation, subunit composition and interaction of the NDH-1 complexes from *Thermosynechococcus elongatus* BP-1. *Biochem. J.* **390**, 513–520 (2005).
27. Kern, J. et al. Purification, characterisation and crystallisation of photosystem II from *Thermosynechococcus elongatus* cultivated in a new type of photobioreactor. *Biochim. Biophys. Acta* **1706**, 147–157 (2005).
28. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
29. Wittig, I., Braun, H.-P. & Schägger, H. Blue native PAGE. *Nat. Protoc.* **1**, 418–428 (2006).
30. Wessel, D. & Flügge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138**, 141–143 (1984).
31. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860 (2006).
32. Beck, S. et al. The impact II, a very high-resolution quadrupole time-of-flight instrument for deep shotgun proteomics. *Mol. Cell Proteomics* **14**, 2014–2029 (2015).
33. Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
34. Tivol, W. F., Briegel, A. & Jensen, G. J. An improved cryogen for plunge freezing. *Microsc. Microanal.* **14**, 375–379 (2008).
35. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
36. Biyani, N. et al. Focus: The interface between data collection and data processing in cryo-EM. *J. Struct. Biol.* **198**, 124–133 (2017).

37. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
38. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
39. Kimanius, D., Forsberg, B. O., Scheres, S. H. W. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016).
40. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
41. Bai, X.-C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. W. Sampling the conformational space of the catalytic subunit of human γ-secretase. *eLife* **4**, e11182 (2015).
42. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
43. Zivanov, J., Nakane, T. & Scheres, S. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).
44. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
45. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
46. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
47. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
48. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).
49. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
50. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
51. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
52. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
53. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
54. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
55. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
56. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
57. Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
58. The PyMOL Molecular Graphics System v.1.8.2 (Schrödinger, 2016).
59. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
60. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
61. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).
62. Vizcaíno, J. A. et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, D447–D456 (2016).
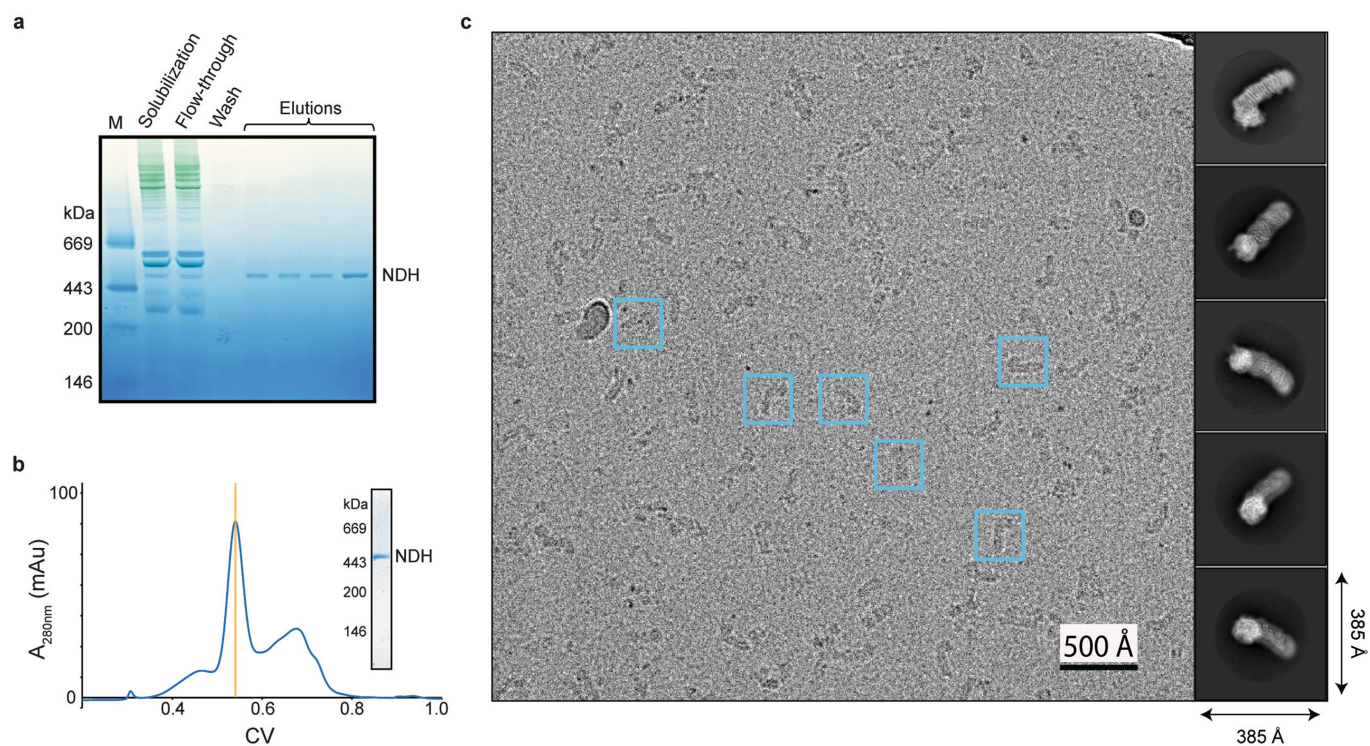
Extended Data Fig. 1 | Purification and cryo-EM of NDH from *T. elongatus*. a, Blue native PAGE gel of NDH solubilized by β-DDM and purified by Ni^{2+}-affinity chromatography with fractions, positions of molecular-weight marker and NDH indicated. **b**, Size-exclusion chromatography trace of concentrated Ni^{2+}-affinity eluates. The gold bar indicates the fraction that shows a single band corresponding to NDH on the blue native PAGE gel (right). See Supplementary Fig. 1 for source gels for **a** and **b**. **c**, Micrograph of frozen hydrated NDH used in this study, with exemplar particles boxed and 2D-class averages generated from dataset 1 showing clear internal features.
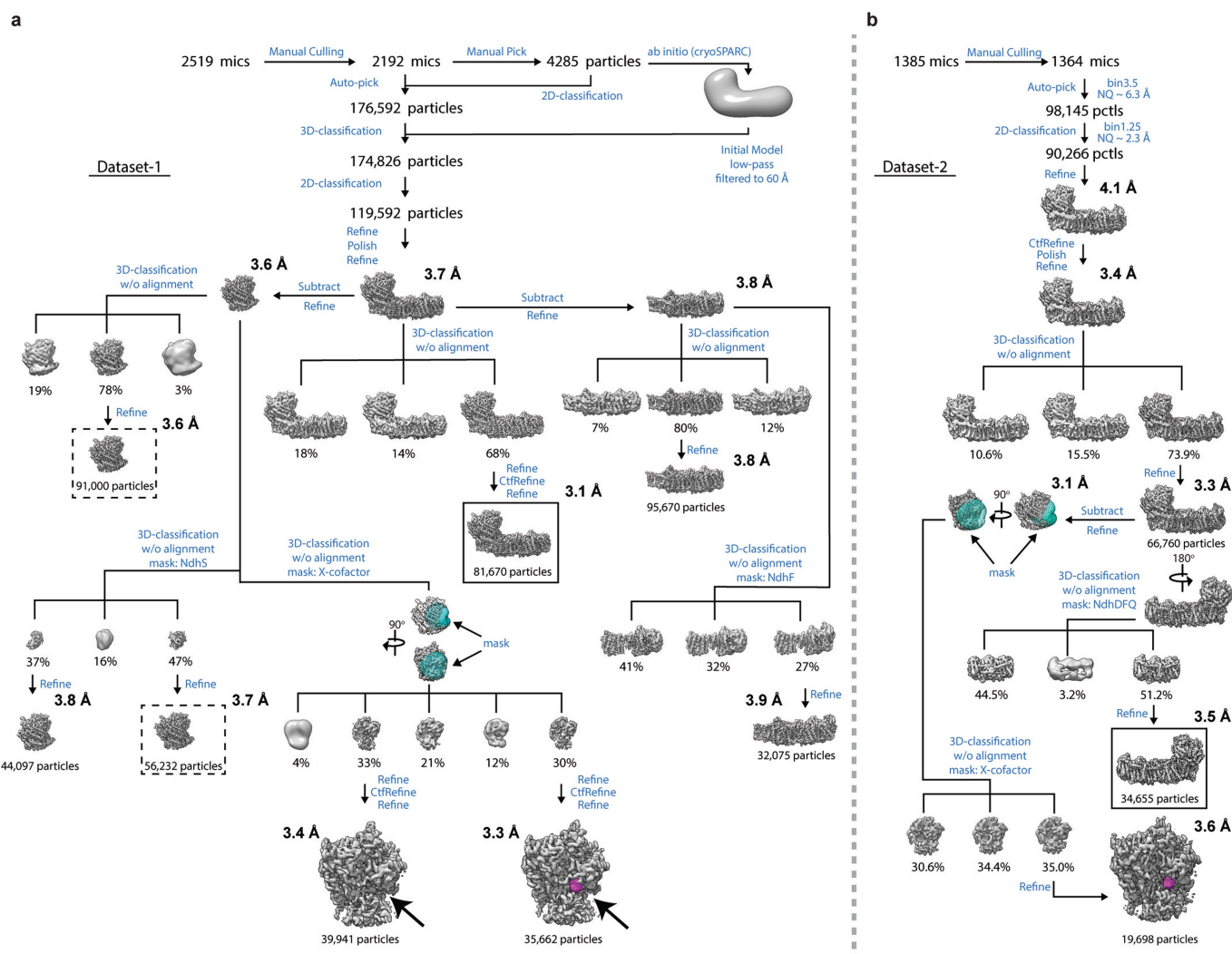
Extended Data Fig. 2 | Cryo-EM data-processing workflow.
a, b, Schematic of pre-processing, classification and refinement procedures used to generate the maps obtained in this study (see Methods and Supplementary Methods for details) for dataset 1 (a) and dataset 2 (b). Boxed maps indicate those used in model building and solid boxes indicate those used for coordinate refinement. The masks used for focused classification of the X-cofactor are indicated in cyan. The apo- and holo X-cofactor focused maps are enlarged to emphasize the X-site (coloured purple). Resolutions are for an FSC of 0.143, as described in Methods.
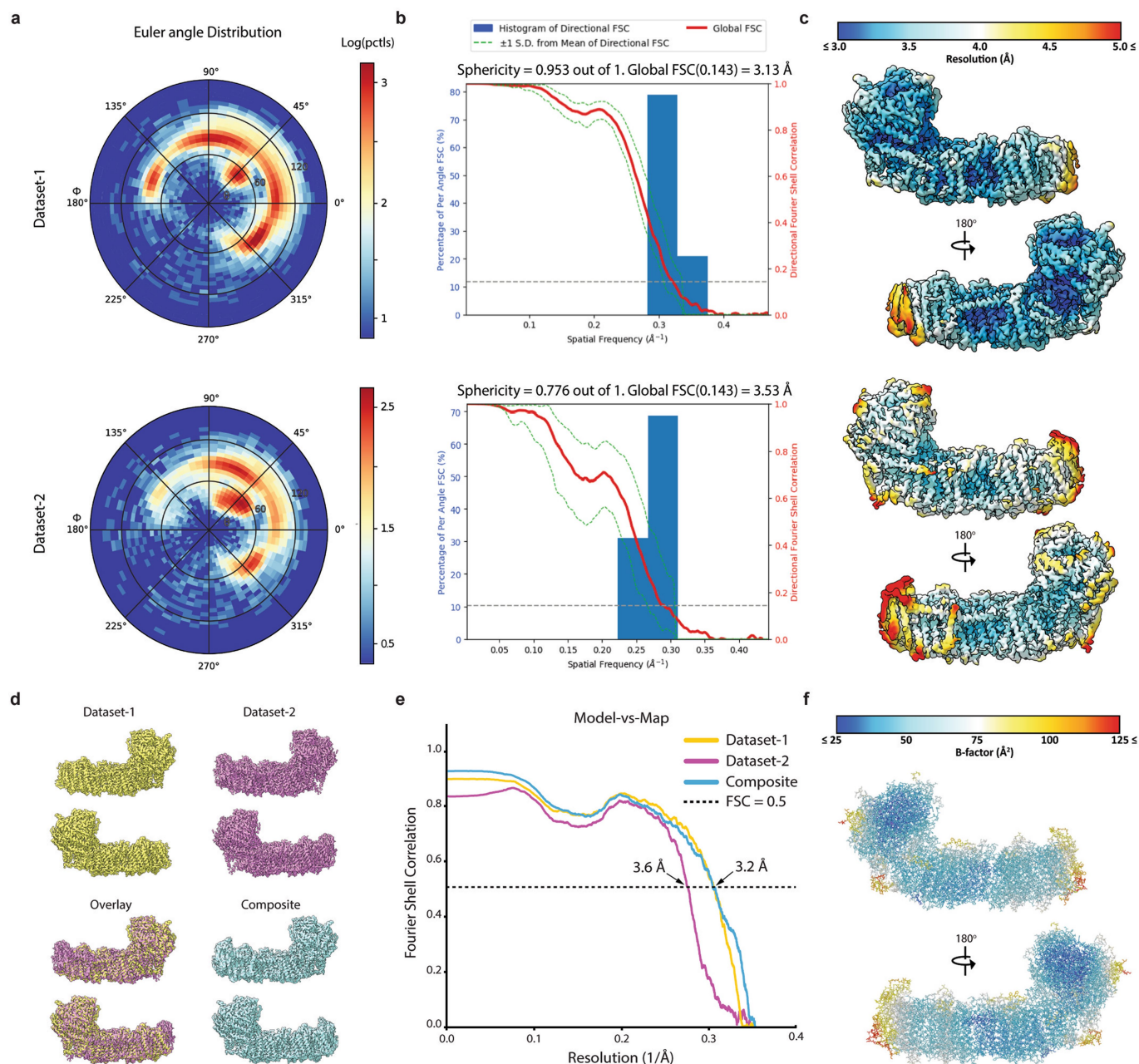
Extended Data Fig. 3 | Resolution assessment of cryo-EM maps and models. a–c, Euler angle distributions (**a**), 3DFSC plots (**b**) and local resolution (RELION, unsharpened) maps (**c**) for the overall cryo-EM maps of dataset 1 (top) and dataset 2 (bottom). **d**, Sharpened maps used for coordinate refinement shown from both sides for dataset 1 (top left, yellow), dataset 2 (top right, pink), and the overlay (bottom left) and composite (bottom right, blue) maps. **e**, Model compared with map FSC for each dataset and composite map with their respective models calculated with PHENIX. **f**, Representation of the refined composite coordinate model, coloured according to *B* factor.
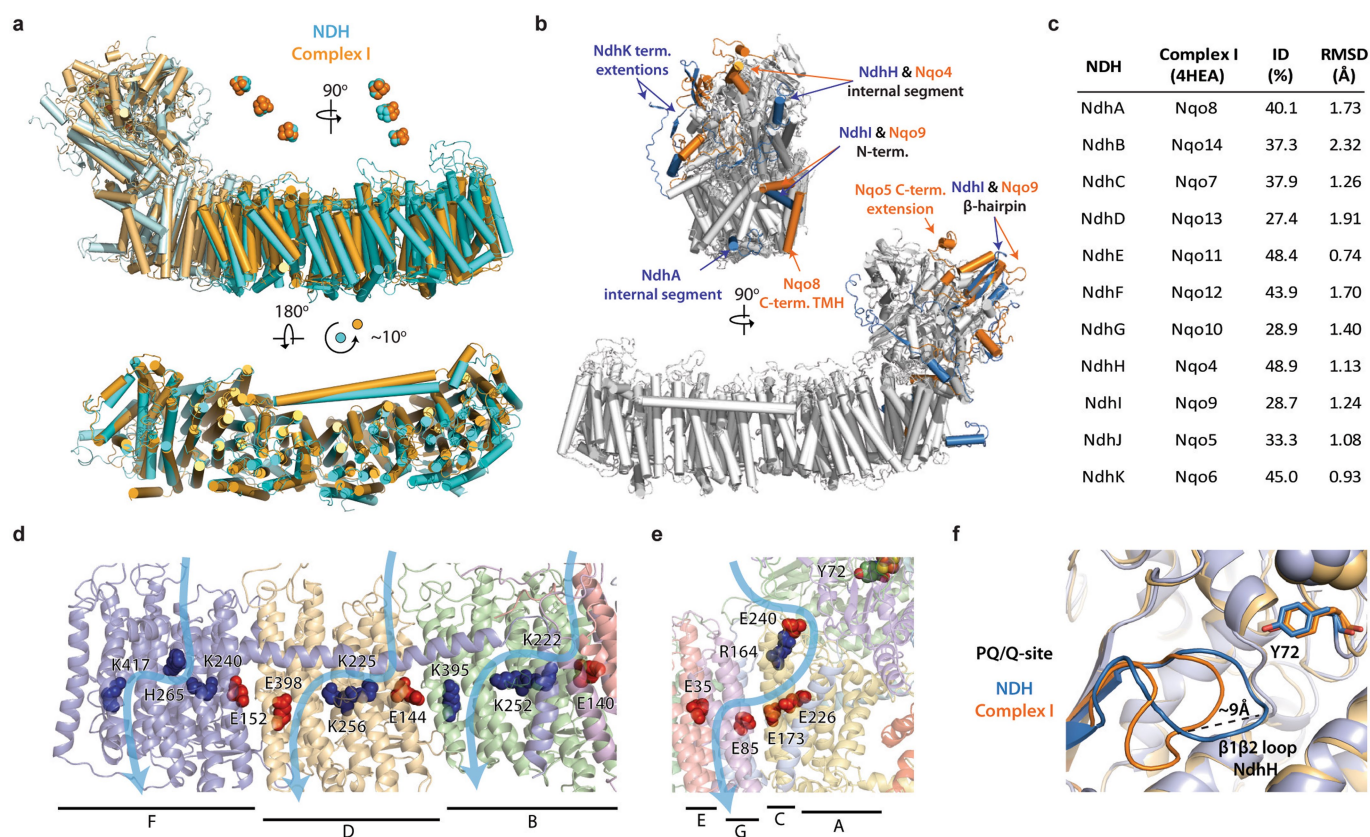
Extended Data Fig. 4 | Comparison of the homologous cores of NDH and complex I. a, Core subunits from *T. elongatus* (cyan) and *T. thermophilus* (orange, PDB: 4HEA) superimposed on NdhA/ND1 (the heel subunit) viewed from side (top) and top-down of the membrane arm (bottom). The antiporter domains are related by a rotation of approximately 10° about their central axis. The iron–sulfur clusters are similarly positioned. **b**, Core subunits and transverse helix aligned individually with observed differences highlighted in colour and labelled (white, conserved structure; blue, distinct to NDH; orange; distinct to complex I). **c**, Table of core subunit sequence homology (ID) and structural similarity (r.m.s.d.). **d**, **e**, Depiction of putative proton-translocation pathways (blue arrows) based on conserved charged residues for the distal (**d**) and PQ/Q-site adjacent (**e**) sites. **f**, Close-up view of the PQ/Q-site based on the location of the coordinating Tyr72 of NdhH (Tyr87 of Nqo4) reveals a difference in the of β1β2 loop of NdhH, which is displaced by approximately 9 Å relative to the complex I homologue Nqo4.
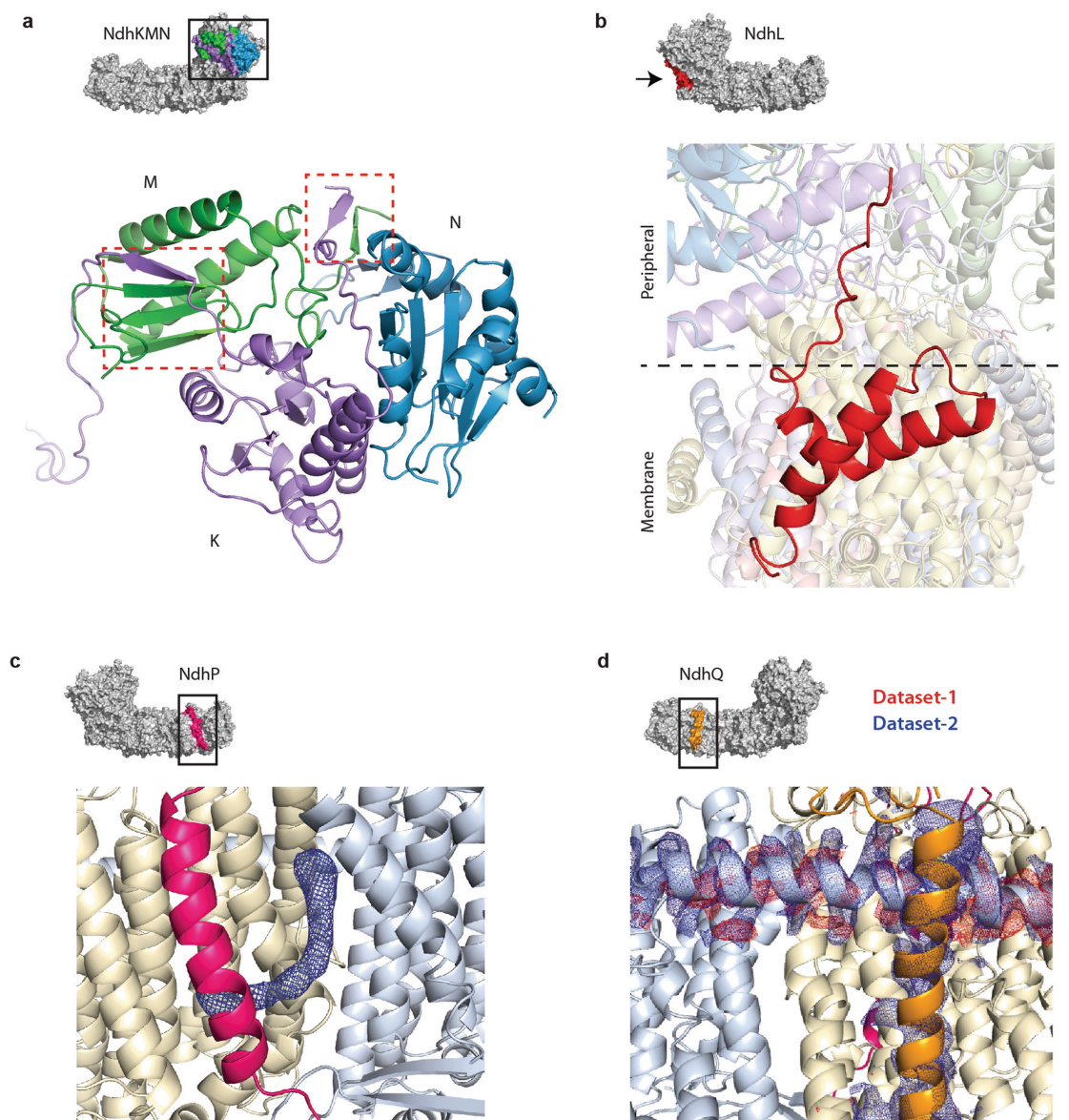
Extended Data Fig. 5 | OPS subunits NdhL, NdhM, NdhN, NdhP and NdhQ. a–d, Cartoon depictions of NdhL, NdhM, NdhN, NdhP and NdhQ, with surface representations indicating their location in the complex. **a**, NdhK (purple), NdhM (green) and NdhN (blue), with complementary β-strands formed between NdhK and NdhM (dashed red boxes). **b**, NdhL (red) connects the peripheral and membrane arms. **c**, NdhP (magenta) trapping a lipid (dark blue mesh) between NdhD (tan) and NdhF (light blue). **d**, NdhQ (orange) with density for dataset 1 (red mesh) and dataset 2 (blue mesh) for NdhQ and the NdhF transverse helix, which is of higher quality in the presence of NdhQ. All density is at $\sigma = 3$.
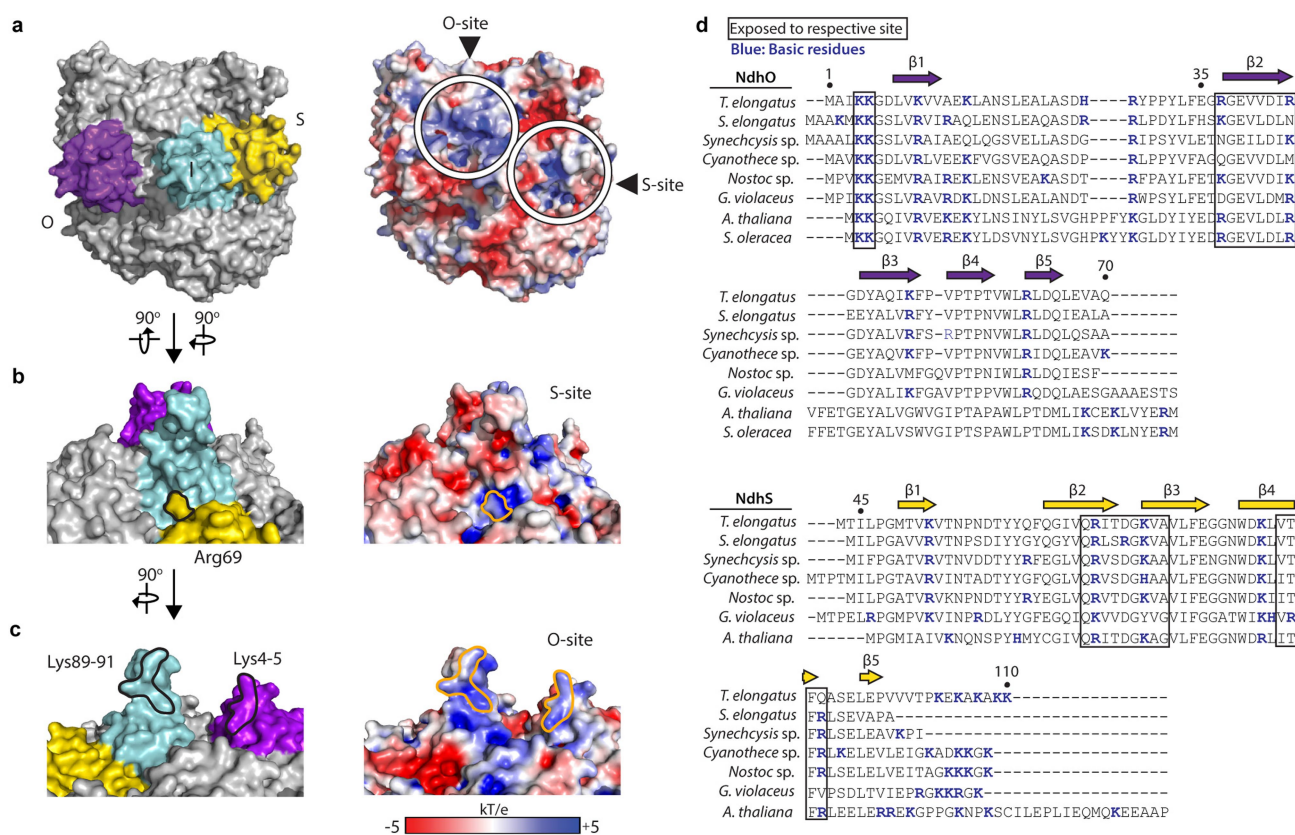
Extended Data Fig. 6 | Electrostatics of the peripheral arm. a–c, Left, surface model views of the peripheral arm with the NdhI β-hairpin, NdhO and NdhS coloured in cyan, purple and yellow, respectively. Right, corresponding electrostatic potential surfaces calculated at pH 8 (see Methods) with the colour key shown at the bottom. **a,** Peripheral arm viewed down the β-hairpin of NdhI. The electrostatic potential of this view shows two positive patches circled in white and labelled as NdhO-adjacent (O-site) and NdhS-adjacent (S-site). **b,** S-site viewed with the indicated rotational relationship to **a** with Arg69 of NdhS outlined in orange. **c,** O-site viewed with the indicated rotational relationship to **b**

with the basic triplet (Lys89-Lys90-Lys91) of NdhI and Lys4-Lys5 of NdhO outlined in orange. **d,** Alignments of NdhO (top) and NdhS (bottom) homologues with β-strand regions indicated, numbered according to the *T. elongatus* sequence. Boxed regions indicate exposed regions facing the respective sites, and basic residues are coloured blue. Predicted chloroplast localization sequences are removed when applicable. The sequence of *S. oleracea* NdhS is not available. Residues 1–42 for *T. elongatus* NdhS are not conserved and are not observed in our structure, and have been removed for the purposes of alignment.

Extended Data Table 1 | Summary of subunits

	Subunit name			NDH model summary			Chloroplast homolog	Complex I homologues		
	Common	Other(s)	Te[a]	Length (aa)	Range built	Notes	At[b]	Tt[c]	Yl[d]	Bt[e]
Core	NdhA		tlr0667	372	28-362	PQ-site, X-site	AtCg01100	Nqo8	NU1M	ND1
	NdhB		tll0045	515	11-493	antiporter-like	AtCg00890[f]	Nqo14	NU2M	ND2
	NdhC		tlr1429	132	13-45; 65-131		AtCg00440	Nqo7	NU3M	ND3
	NdhD	NdhD1	tll0719	529	2-505	antiporter-like	AtCg01050	Nqo13	NU4M	ND4
	NdhE		tlr0670	101	1-101		AtCg01070	Nqo11	NULM	ND4L
	NdhF	NdhF1	tll0720	656	1-655	antiporter-like, traverse helix, His-loop	AtCg01010	Nqo12	NU5M	ND5
	NdhG		tlr0669	200	4-194		AtCg01080	Nqo10	NU6M	ND6
	NdhH		tlr1288	394	3-394	PQ-site, X-site, O-site	AtCg01110	Nqo4	NUCM	49 kDa
	NdhI		tlr0668	196	6-189	2x 4Fe-4S, O-site, S-site	AtCg01090	Nqo9	NUIM	TYKY
	NdhJ		tlr1430	168	8-168		AtCg00420	Nqo5	NUGM	30 kDa
	NdhK		tlr0705	237	7-218	4Fe-4S, PQ-site, X-site	AtCg00430	Nqo6	NUKM	PSST
								Nqo1	NUBM	51 kDa
								Nqo2	NUHM	24 kDa
								Nqo3	NUAM	75 kDa
OPS	NdhL	CRR23	tsr0706	76	11-76	TMH and partial TMH	At1g70760			
	NdhM		tll0447	111	1-111	X-site	At4g37925			
	NdhN		tlr1130	150	2-149	S-site	At5g58260			
	NdhO		tsl0017	70	2-70	O-site, SH3-like	At1g74880			
	NdhP	PnsB4/NDF6	N/A	44	2-43	TMH, trapped lipid	At1g18730			
	NdhQ		N/A	45	2-44	TMH				
	NdhS	CRR31	tlr0636	110	43-99	S-site, SH3-like	At4g23890			
	NdhV		tlr0472	150	0	Not observed	At2g04039			

[a]*T. elongatus*. [b]*A. thaliana*. [c]*T. thermophilus*. [d]*Y. lipolytica*. [e]*B. torus*. [f]Identical to *A. thaliana* Cg01250.

Extended Data Table 2 | Mass spectrometry data from the purified NDH complex

Subunit	Uniprot ID	Sequence length	MW [kDa]	In-solution					In-gel		
				Peptides			Sequence coverage [%]	MaxQuant Score	Peptides	Sequence coverage [%]	MaxQuant Score
				CHYMO	TRYP	TOTAL					
NdhA	Q8DL32	372	40.55	14 (12)	8 (8)	22 (20)	44.1 (40.6)	260.88 (176.29)	36 (28)	63.7 (49.7)	323.31 (297.92)
NdhB	Q8DMR6	515	55.14	34 (24)	4 (3)	38 (27)	42.5 (34.2)	228.22 (114.06)	42 (32)	45.4 (38.6)	323.31 (323.31)
NdhC	Q8DJ02	132	15.00	4 (2)	0 (0)	4 (2)	28 (12.9)	41.331 (19.486)	6 (6)	28.8 (28.8)	56.185 (52.667)
NdhD/NdhD1	Q8DKY0	514	56.08	30 (22)	6 (6)	36 (28)	49 (40.1)	323.31 (277.07)	34 (29)	44.9 (37.6)	323.31 (323.31)
NdhE	Q8DL29	101	11.13	2 (2)	2 (2)	4 (4)	21.8 (21.8)	7.367 (5.2195)	9 (8)	66.3 (59.4)	282.59 (104.12)
NdhF/NdhF1	Q8DKX9	656	71.97	37 (21)	5 (5)	42 (26)	46.5 (32)	240.12 (114.94)	49 (34)	57.8 (45.6)	323.31 (323.31)
NdhG	Q8DL30	200	21.57	12 (8)	3 (3)	15 (11)	49.5 (40.5)	107.72 (76.66)	12 (10)	48 (33)	69.291 (85.808)
NdhH	Q8DJD9	394	45.22	38 (31)	23 (23)	61 (54)	90.6 (88.8)	323.31 (323.31)	72 (63)	95.4 (91.4)	323.31 (323.31)
NdhI	Q8DL31	196	22.42	18 (16)	13 (12)	31 (28)	92.3 (90.8)	323.31 (323.31)	26 (26)	81.1 (80.6)	323.31 (323.31)
NdhJ	Q8DJ01	168	19.34	19 (16)	17 (17)	36 (33)	99.4 (99.4)	323.31 (323.31)	21 (21)	81.5 (79.2)	323.31 (323.31)
NdhK	Q8DKZ4	237	25.74	20 (17)	11 (11)	31 (28)	84.4 (81.4)	323.31 (323.31)	38 (33)	94.1 (87.8)	323.31 (323.31)
NdhL	Q8DKZ3	76	8.57	0 (0)	0 (0)	0 (0)	0 (0)	N/A	7 (5)	48.7 (36.8)	184.86 (68.945)
NdhM	Q8DLN5	111	12.57	14 (12)	7 (7)	21 (19)	97.3 (95.5)	323.31 (323.31)	11 (11)	73.9 (73.9)	323.31 (219.89)
NdhN	Q8DJU2	150	16.64	14 (10)	9 (8)	23 (18)	88 (85.3)	323.31 (323.31)	23 (17)	87.3 (80.7)	323.31 (323.31)
NdhO	Q8DMU4	70	7.87	7 (7)	9 (8)	16 (15)	98.6 (98.6)	323.31 (323.31)	7 (7)	85.7 (85.7)	226.92 (66.33)
NdhP	N/A	44	4.87	0	1 (1)	1 (1)	36.4 (36.4)	23.472 (22.536)	4 (4)	75 (75)	92.05 (41.566)
NdhQ	N/A	45	4.71	1 (0)	0 (0)	1 (0)	20.5 (0)	5.554 (N/A)	5 (3)	75 (59.1)	143.55 (34.486)
NdhS	Q8DL61	110	12.45	8 (5)	4 (3)	12 (8)	69.1 (53.6)	323.31 (323.31)	6 (6)	60.9 (55.5)	82.066 (44.438)
NdhV	Q8DLL4	150	16.00	0 (0)	0 (0)	0 (0)	0 (0)	N/A	0 (0)	0	N/A

Values from focused (whole proteome) search

Extended Data Table 3 | Cryo-EM data collection, refinement and validation statistics

	Dataset 1 (EMD-0415) (PDB 6NBQ)	Dataset 2 (EMD-0425) (PDB 6NBX)	Composite (PDB 6NBY)
Data collection and processing			
Magnification	53,350	55,555	
Voltage (kV)	300	300	
Electron exposure (e–/Å^2)	~50	~60	
Defocus range (μm)	-1.4 to -3.1	-0.8 to -2.7	
Pixel size (Å)	1.068	0.903	
Symmetry imposed	C1	C1	
Initial particle images (no.)	176,592	98,145	
Final particle images (no.)	81,670	34,655	
Map resolution (Å)	3.13	3.53	
FSC threshold	0.143	0.143	
Map resolution range (Å)	2.8-5.9	3.2-7.3	
Refinement			
Initial model used (PDB code)		4HEA, 3C4S	
Model resolution (Å)	3.23	3.56	3.21
FSC threshold	0.5	0.5	0.5
Model resolution range (Å)	3.23	3.56	3.21
Map sharpening B factor (Å^2)	-30	-20	-30
Model composition			
Non-hydrogen atoms	28296	29401	29401
Protein residues	3755	3854	3854
Ligands	3	3	3
B factors (Å^2)			
Protein	66	43	56
Ligand	42	43	39
R.m.s. deviations			
Bond lengths (Å)	0.007	0.009	0.005
Bond angles (°)	0.967	1.074	0.923
Validation			
MolProbity score	1.85	1.83	1.79
Clashscore	6.21	5.71	5.91
Poor rotamers (%)	0.53	0.71	0.24
Ramachandran plot			
Favored (%)	91.29	91.06	92.43
Allowed (%)	8.63	8.81	7.49
Disallowed (%)	0.08	0.13	0.08

nature research

Corresponding author(s):	Davies K.M.
Last updated by author(s):	Dec 23, 2018

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 Authors & Referees 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
Data analysis	GCTF-v1.06, MotionCor2-v1.0.0, RELION-v2.1.0, RELION-v3.0.b2, cryoSPARC, Andromeda, MaxQuant, Focus, Gautomatch-v0.56, Chimera-v1.11.2, COOT-v0.8.8, PHENIX-v1.8.2, EMRinger, MolProbity-v4.4, Phyre2, PsiPred, SPrit, ClustalOmega and Uniprot.

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/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

Cryo-EM maps for the dataset-1 Overall, Peripheral Arm, with NdhS, without NdhS, with X-cofactor and without X-cofactor and dataset-2 Overall have been deposited with the Electron Microscopy Data Bank under accession numbers EMD-0415, EMD-0416, EMD-0417, EMD-0418, EMD-0419, EMD-0420 and EMD-0425, respectively. Atomic coordinates for the overall complex for dataset-1, dataset-2 and composite have been deposited with the PDB accession codes 6NBQ, 6NBX and 6NBY, respectively. All other data available upon 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	n=2. Two independent protein purifications were performed and imaged separated. The first purification generated dataset 1 and consisted of 2519 micrographs containing 176592 2D projections of protein particles. Dataset 2, contained 1385 micrographs taken of proteins obtained from protein purification experiment 2. The 1385 micrographs contained 98145 2D images of protein particles.
Data exclusions	2D images of protein particles are automatically picked. Images which contain overlapping particles, dirt or have image defects are deselected. During the process of classification and refinement, images of denatured or damaged proteins are removed.
Replication	In this instance, two independent protein structures of NDH-1 were generated by obtaining image data from two independent protein purification. Both structures showed the same overall structure expect dataset 2, contained NdhQ and more lipids.
Randomization	Not applicable with this method. 2D images of protein particles are taken at random in the TEM in areas where the ice appears to be of optimal thickness. Images are taken blind (e.g. no pre-exposure) to avoid radiation damage. Thus it is not know whether protein particles are present in the imaged ice hole until after the image has been acquired.
Blinding	Data is acquired without pre-exposure to avoid radiation damage. Data is classified using a computer based on the % similarity between two images. Prior to data alignment, the images in a dataset are randomly divided into two independent datasets and processed separately. Only being combined during the last iteration to determine how similar the two structures obtained are.

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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