Architecture of mammalian respiratory complex I

Kutti R. Vinothkumar*1, Jiapeng Zhu*2 & Judy Hirst2

Complex I (NADH:ubiquinone oxidoreductase) is essential for oxidative phosphorylation in mammalian mitochondria. It couples electron transfer from NADH to ubiquinone with proton translocation across the energy-transducing inner membrane, providing electrons for respiration and driving ATP synthesis. Mammalian complex I contains 44 different nuclear- and mitochondrial-encoded subunits, with a combined mass of 1 MDa. The 14 conserved ‘core’ subunits have been structurally defined in the minimal, bacterial complex, but the structures and arrangement of the 30 ‘supernumerary’ subunits are unknown. Here we describe a 5 Å resolution structure of complex I from Bos taurus heart mitochondria, a close relative of the human enzyme, determined by single-particle electron cryo-microscopy. We present the structures of the mammalian core subunits that contain eight iron–sulphur clusters and 60 transmembrane helices, identify 18 supernumerary transmembrane helices, and assign and model 14 supernumerary subunits. Thus, we considerably advance knowledge of the structure of mammalian complex I and the architecture of its supernumerary ensemble around the core domains. Our structure provides insights into the roles of the supernumerary subunits in regulation, assembly and homeostasis, and a basis for understanding the effects of mutations that cause a diverse range of human diseases.

Mammalian complex I (ref. 1) is one of the largest and most complicated enzymes in the cell. Complex I from B. taurus (bovine) heart mitochondria has been characterized extensively as a model for the human enzyme; both enzymes contain 44 different subunits (encoded by both the nuclear and mitochondrial genomes)5–7 and nine redox cofactors (a flavin mononucleotide and eight iron–sulphur clusters). Fourteen subunits are the core subunits that are conserved in all complex I enzymes; they contain all the mechanistically critical cofactors and structural elements and are sufficient for catalysis. Crystal structures of intact complex I from the thermophilic bacterium Thermus thermophilus4, and of domains of the prokaryotic enzymes from T. thermophilus and Escherichia coli11,12 have provided a wealth of information on the structures of these subunits—but they represent only half the mass of the mammalian enzyme. The cohort of 30 supernumerary subunits particular to the mammalian enzyme8,9 has been accumulated through evolution. The supernumerary subunits may have alternative functions or be important for assembly, regulation, stability or protection against oxidative stress—their structures and arrangement around the core subunits are not known.

Owing to its size, L-shaped asymmetry, membrane-bound location, and multi-component structure, mammalian complex I has proved difficult to crystallize, and its high-resolution structure has not yet been determined. Crystallographic information on any eukaryotic complex I is currently limited to a medium-resolution map of the enzyme from the yeast Yarrowia lipolytica, which has been described, but not modelled8. Conversely, the size and shape of complex I make it an attractive target for electron microscopy (EM), and the enzymes from several species have been visualized to display their overall L-shaped structures8,9,10,11, although at too low a resolution to reveal detailed structural information. A high-resolution structure of the mammalian enzyme is essential for understanding how the 30 supernumerary mammalian subunits are arranged around the core domain, how they determine the properties, assembly and activity of the enzyme, and how mutations in both the core and supernumerary subunits cause human diseases12,13.

Imaging and reconstruction

Complex I was purified from B. taurus heart mitochondria in detergent11, and imaged in vitreous ice on holey-carbon grids with a Falcon direct electron detector (see Methods). The enzyme adopts different orientations on the grid, and reference-free two-dimensional class averages clearly show the characteristic L-shape of the minimal prokaryotic form augmented by extra domains from the supernumerary subunits (Extended Data Fig. 1). Refinement was performed in RELION14 and movie frames were used to correct for beam-induced movement15. Per-frame reconstruction and B-factor weighting were followed by three-dimensional classification, resulting in the final map (Fig. 1) obtained from 25,492 particles with an overall resolution of ~5 Å (see Methods and Extended Data Fig. 2). Viewed at a low-density threshold the map is dominated by a disordered detergent–phospholipid belt that encircles the hydrophobic domain and defines the position of the membrane. At intermediate-density threshold, the hydrophilic matrix domain and the extended membrane domain, containing a large number of transmembrane helices (TMHs), are observed. The highest-density peaks in the map reveal the eight iron–sulphur (FeS) clusters that, as in the T. thermophilus4 and Y. lipolytica enzymes, form a chain through the hydrophilic domain.

Structures of the core subunits

The 14 conserved core subunits of complex I (refs 1, 4) catalyse the energy transducing reactions: NADH oxidation, ubiquinone reduction and proton translocation (Extended Data Table 1 summarizes their nomenclature). The seven nuclear-encoded hydrophilic core subunits harbour a flavin mononucleotide to oxidize NADH, FeS clusters for inter-substrate electron transfer, and the ubiquinone-binding site. The seven mitochondrial-encoded membrane core subunits contain four antiporter-like domains for proton translocation. The structures of the mammalian core subunits (Fig. 2) were fitted to the density map (see Methods) using the structure of T. thermophilus complex I (ref. 4), secondary structure analyses and sequence alignments, and using structural features and
Figure 1 | Overall map for complex I from *B. taurus* heart mitochondria determined by single-particle cryo-EM. Three distinct features of the complex are revealed by overlaying maps at different density thresholds. The map at the highest threshold (red) reveals the FeS clusters. The map at medium threshold (grey) reveals the overall architecture of the protein and the 78 TMHs in the membrane domain. The detergent–phospholipid belt observed as a dominant feature at low density threshold (translucent blue) represents the density that remains around the membrane domain after cutting out the final model of the protein, and denotes the position of the complex in the membrane. It is ~30 Å thick, and 3–4 Å thinner at the proximal end of the complex (left) than at the distal end (right).

densities from aromatic side chains (Extended Data Fig. 3). Except for the FeS-cluster ligands they have been modelled as polyalanine chains, with the residue numbering optimized to enable individual residues to be located (Extended Data Table 2). It is not possible to attribute density to any bound ubiquinone species in the present map.

A comparison of the bacterial4 and mammalian core enzymes reveals that the mammalian membrane domain is more strongly curved ‘out’ of the membrane plane (Extended Data Fig. 4). However, within each individual subunit the 60 TMHs of the mammalian core subunits match their *T. thermophilus* counterparts closely (Extended Data Fig. 5)—only the position of TMH4 in subunit ND6 is different, and the extra carboxy-terminal TMH particular to *T. thermophilus* ND1 is absent from *B. taurus* (Fig. 2 and Extended Data Fig. 5). No notable density is observed in place of the three amino-terminal TMHs (present in *T. thermophilus* and *Y. lipolytica*) that have been lost through evolution of mammalian ND2 (ref. 16), so they have not been substituted structurally by other subunits. Importantly, catalytically relevant features identified in the antiporter-like subunits of the bacterial complex5,6 are conserved. They include the loops in the six broken TMHs in ND2, ND4 and ND5 (see Extended Data Fig. 3 for examples) that may constitute part of the proton-translocation mechanism, and the long transverse helix in NDS, a proposed coupling element.

Of the seven hydrophilic core subunits (Fig. 2), the structures of the *B. taurus* 51 kDa subunit (human homologue NDUFV1), 49 kDa (NDUF52), 24 kDa (NDUFV2), PSST (NDUF57) and TYKY (NDUF58) subunits, and the small domain of the 75 kDa subunit (NDUF51) are closely conserved from their *T. thermophilus* homologues6,7 (Extended Data Fig. 5), with marked variation only in the length and extent of some of their N and C termini. Consequently, the arrangements of the FeS cluster chains are also very similar (Extended Data Table 3), except that, owing to rotation of the 51 and 24 kDa subunits, the superimposed chains diverge with increasing distance from the membrane (Extended Data Fig. 4). The sequence and structural conservation of the large domain of the 75 kDa subunit, which contains an extra, catalytically redundant cluster in *T. thermophilus*2 and the 30 kDa subunit (NDUF53), are lower (Extended Data Table 2). As neither of them have any known catalytic role, we conclude that the catalytically critical subunits and cofactors are closely conserved in the mitochondrial and bacterial enzymes, supporting their common mechanism of catalysis.

The supernumerary ensemble

Once the core subunits had been modelled, the map revealed that additional densities form an open cage around the core (Fig. 3). These densities are attributed to the supernumerary subunits, and they are arranged predominantly around the membrane domain and lower hydrophilic domain, where they may help to protect FeS-containing PSST and TYKY from oxidative damage19. Two large supernumerary domains, one capping ND5 and part of ND4, the other capping ND2, are observed on the matrix surface of the membrane domain. Facing the intermembrane space, as noted in *Y. lipolytica*8, the supernumerary subunits form a layer of protein that may have a role similar to that of the stabilizing β-hairpin–helix structures observed in the prokaryotic enzyme1. Eighteen supernumerary TMHs are distributed around the core membrane domain...
intermediates lacking the NADH-dehydrogenase module. The models for the core subunits are in light colours (as labelled) in surface representation, and density attributed to the supernumerary subunits, forming a cage around the core subunits, is in dark red. The supernumerary subunits are concentrated on each side of the membrane domain, and around the lower section of the hydrophilic domain. The NADH-binding site in the 51 kDa subunit is indicated, with the predicted positions for the flavin isaloalloxazine (orange spheres) and three conserved phenylalanines at the entry to the site (yellow); the vicinity of this site is devoid of supernumerary subunit density.

Assignment of 14 supernumerary subunits

To identify and assign individual supernumerary subunits to the map for mammalian complex I (Extended Data Table 4 summarizes their nomenclature) we used biochemical, sequence and structural information. Homology models for six of the hydrophilic supernumerary subunits were created using known structures (Extended Data Tables 4 and 5). Human B8 (NDUFAD2) adopts a thioredoxin fold and its structure (Fig. 5) was located at the tip of the large domain of the 75 kDa subunit (Fig. 4), so (contrary to current models) B8 is likely to be assembled into complex I after (or with) the 75 kDa subunit. B8 is extensively degraded in brain mitochondria from patients with Parkinson’s disease, and, along with other NADH dehydrogenase domain subunits, it is rapidly exchanged under steady-state conditions. Therefore, it may help to protect the core enzyme against oxidative damage. Similarly, regions of density consistent with two subunits important for complex I assembly, the 18 kDa (NDUFS4) and 13 kDa (NDUFS6) subunits (Extended Data Tables 4 and 5), were located (Fig. 4). However, they are small proteins with no predicted dominant secondary structure and it cannot be excluded that other supernumerary subunits have similar structures. In the current map, the 18 kDa subunit has been modelled into a density in a cleft between the 75 kDa subunit and the 49 kDa, 30 kDa and TYKY subunits; the density attributed to the 13 kDa subunit suggests that it interacts with the 75 kDa, 49 kDa and TYKY subunits (Fig. 4). These locations may explain why clinically identified mutations in the 18 kDa and 13 kDa subunits lead to accumulation of late-stage interrupted-assembly intermediates lacking the NADH-dehydrogenase module.

The 42 kDa subunit (NDUFA10), a member of the nucleoside kinase family, was easily located as the density on top of ND2, on the matrix side of the membrane (Figs 4, 5 and Extended Data Tables 4, 5). Its location is neatly confirmed by its absence from the density map of Y. lipolytica complex I (ref. 8), which lacks this mammalian-specific subunit. Phosphorylation of a serine in the 42 kDa subunit by a PINK1-dependent mechanism has been proposed to be required for complex I assembly.
units B13 (NDUFA5) and B14 have similar predicted secondary structures, an LYR motif and it is in subcomplex I in loss of catalytic activity. Notably, subunit B22 (NDUFB9) also contains the hydrophilic domain in subcomplex I after fractionation of B. taurus complex I with zwitterionic detergents, and they are missing from Y. lipolytica subcomplex I, which lacks core subunits ND1, 2, 3 and 4L. Therefore, they are assigned to the three TMH densities next to ND1 (Extended Data Table 4 and 5). Sequence analyses predict a single 67-residue helix in subunit B16.6, with the first 20 residues forming a TMH. Correspondingly, one of the three densities is very long and modelled as a single 63-residue helix that interacts with the N terminus of TYKY on the matrix side, spans the membrane, then bends into the intermembrane space and is anchored under the 'hel' (Figs 4 and 5). Therefore, this density is assigned to B16.6, a protein identical to the cell death regulatory gene product GRIM19 (ref. 35). It is currently not possible to confidently deduce assignments for the TMH-containing subunits of subcomplex Iβ.

The PGIV (NDUFA8), 15 kDa (NDUF5) and B18 (NDUF7) subunits contain twin CX4C motifs that form two intramolecular disulphides within 'CHCH' domains, and they have been assigned to the intermembrane surface of complex I (ref. 37). A double L-shaped density, resembling two CHCH domains at right angles, is clearly visible on the hel, clamping B16.6 (Figs 4 and 5) onto the core. PGIV, a subunit present in subcomplex Iα (Extended Data Table 4), contains two CHCH domains, so it is assigned to the L-shaped density feature (Figs 4 and 5), consistent with the position of an antibody label to its homologous subunit in Y. lipolytica. Our structure thus reveals the architecture of the '400 kDa' assembly intermediate of human complex I (refs 18, 33) that contains the core hydrophilic subunits 49 kDa, 30 kDa, PSST and TYKY, core membrane subunit ND1, and the supernumerary subunits PGIV, B9, B16.6 and B13.

Conclusions and perspectives

We have described a 5 Å resolution cryo-EM density map for mammalian complex I, and used it to produce structural models for the 14 core subunits that are conserved in all complex I enzymes, plus 14 of the supernumerary subunits of the mammalian enzyme. The core subunits comprise the catalytically active centre of the enzyme, and (as expected) they closely resemble their counterparts described by the atomic-resolution structure of bacterial complex I (ref. 4). The 14 supernumerary subunits include two copies of subunit SDAP, bringing the total number of subunits in the mammalian complex up to 45. We have used our structural models for the supernumerary subunits to support and discuss their roles in assembly, homeostasis and regulation. Higher-resolution maps are required for assignment of the remaining 17 supernumerary subunits.

Recent developments in direct electron detectors, microscopy and image processing algorithms have enabled high-resolution structures of biological macromolecules to be determined by single-particle cryo-EM at resolutions that have previously only been routinely possible with X-ray crystallography. Thus, we believe that it will be possible to extend our current study to produce a high-resolution structure for complex I in the near future, to allow us to identify and model all the supernumerary subunits, and to characterize the structural changes that occur during catalysis, a crucial step in defining the mechanism of electron-coupled proton translocation.
1. Hirst, J. Mitochondrial complex I. Annu. Rev. Biochem. 82, 551–575 (2013).
2. Carroll, J., Fearnley, I. M., Shannon, R. J., Hirst, J. & Walker, J. E. Analysis of the subunit composition of complex I from bovine heart mitochondria. Mol. Cell. Proteomics 2, 117–126 (2003).
3. Hirst, J., Carroll, J., Fearnley, I. M., Shannon, R. J. & Walker, J. E. The nuclear encoded subunits of complex I from bovine heart mitochondria. Biochem. Biophys. Acta 1604, 135–150 (2003).
4. Baradaran, R., Bernsford, J. M., Minhas, G. S. & Szakasitz, L. A. Crystal structure of the entire respiratory complex I. Nature 494, 443–448 (2013).
5. Sazanov, L. A. Structure of the membrane domain of respiratory complex I. Nature 476, 414–420 (2011).
6. Sazanov, L. A. & Hinchliffe, P. Structure of the hydrophilic domain of respiratory complex I from Thermus thermophilus. Science 311, 1430–1436 (2006).
7. Hunte, C., Zickermann, V. & Brandt, U. Functional modules and structural basis of conformational coupling in mitochondrial complex I. Science 329, 448–451 (2010).
8. Leonard, K., Haier, H. & Weiss, H. Three-dimensional structure of NADH:ubiquinone reductase (complex I) from Neurospora mitochondria determined by electron microscopy of membrane crystals. J. Mol. Biol. 194, 277–286 (1987).
9. Grigorieff, N. Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 Å in ice. J. Mol. Biol. 277, 1033–1046 (1998).
10. Clason, T. et al. The structure of eukaryotic and prokaryotic complex I. J. Struct. Biol. 169, 81–88 (2010).
11. Fassone, E. & Rahman, S. Complex I deficiency: clinical features, biochemistry and molecular genetics. J. Med. Genet. 49, 578–590 (2012).
12. Sharpley, M. S., Shannon, R. J., Draghi, F. & Hirst, J. Interactions between phospholipids and NADH:ubiquinone oxidoreductase (complex I) from bovine mitochondria. Biochemistry 45, 241–248 (2006).
13. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180, 519–530 (2012).
14. Bai, X.-C., Fernandez, I. S., McMullan, G. & Scheres, S. H. W. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. eLife 2, e00461 (2013).
15. Biellier, J. A. & Hirst, J. Truncation of subunit ND2 disrupts the threefold symmetry of the antipporter-like subunits in complex I from higher metazoans. FEBS Lett. 584, 4247–4252 (2010).
16. Kusmaucl, L. & Hirst, J. The mechanism of superoxide production by NADH:ubiquinone oxidoreductase (complex I) from bovine heart mitochondria. Proc. Natl Acad. Sci. USA 103, 7607–7612 (2006).
17. Mimaki, M., Wang, X., McKenzie, M., Thormburn, D. R. & Ryan, M. T. Understanding mitochondrial complex I assembly in health and disease. Biochim. Biophys. Acta 1817, 851–862 (2012).
18. Dieteren, C. E. J. et al. Subunit-specific incorporation efficiency and kinetics in mitochondrial complex I homeostasis. J. Biol. Chem. 287, 41851–41860 (2012).
19. Brookmann, C. et al. The oxidised subunit BS from human complex I adopts a thioredoxin fold. Structure 12, 1645–1654 (2004).
20. Keeney, M. P., Xie, J., Capaldi, R. A. & Bennett, J. P. Parkinson’s disease brain mitochondrial complex I has oxidatively damaged subunits and is functionally impaired and misassembled. J. Neurosci. 26, 5256–5264 (2006).
21. Leshinsky-Silver, E. et al. NDUF5 mutations cause Leigh syndrome with predominant brainstem involvement. Mol. Genet. Metab. 97, 185–189 (2009).
22. Kirby, D. M. et al. NDUF56 mutations are a novel cause of lethal neonatal mitochondrial complex I deficiency. J. Clin. Invest. 114, 837–845 (2004).
23. Sharpley, M. S. Studies of the Catalytic Activity of NADH:Ubiquinone Oxidoreductase (Complex I) from Bovine Mitochondria. PhD thesis, Cambridge Univ. (2005).
24. Morais, V. A. et al. PINK1 loss of function mutations affect mitochondrial complex I activity via NdufA10 ubiquinone uncoupling. Science 344, 203–207 (2014).
25. Fearnley, I. M. & Walker, J. E. Conservation of sequences of subunits of mitochondrial complex I and their relationships with other proteins. Biochim. Biophys. Acta 1140, 105–134 (1992).
26. Abdurrahmanova, A., Zwick, R., Kerscher, S., Zickermann, V. & Brandt, U. Tight binding of NADPH to the 39-kDa subunit of complex I is not required for catalytic activity but stabilizes the multiprotein complex. Biochim. Biophys. Acta 1757, 1676–1682 (2006).
27. Babot, M. et al. NDS1 and 39 kDa subunits are more exposed in the de-active form of bovine mitochondrial complex I. Biochim. Biophys. Acta 1837, 929–939 (2014).
28. Runswick, M. J., Fearnley, I. M., Skehel, J. M. & Walker, J. E. Presence of an acyl carrier protein in NADH:ubiquinone oxidoreductase from bovine heart mitochondria. FEBS Lett. 286, 121–124 (1991).
29. Onn, J. E., Fearnley, I. M. & Walker, J. E. Mammalian mitochondria contain a soluble acyl carrier protein. FEBS Lett. 579, 4892–4896 (2005).
30. Dobyns, K. N. Characterization of two different acyl carrier proteins in complex I from Yarrowia lipolytica. Biochim. Biophys. Acta 1797, 152–159 (2010).
31. Angerer, H. et al. The LVR protein subunit NB4M/NDUF6/ mitochondrial complex I anchors an acyl carrier protein and is essential for catalytic activity. Proc. Natl Acad. Sci. USA 111, 5207–5212 (2014).
32. Andrews, B., Carroll, J., Ding, S., Fearnley, I. M. & Walker, J. E. Assembly factors for the membrane arm of human complex I. Proc. Natl Acad. Sci. USA 110, 18934–18939 (2013).
33. Angerer, H. et al. A scaffold of accessory subunits links the peripheral arm and the distal proton-pumping module of mitochondrial complex I. Biochem. J. 437, 279–288 (2011).
34. Fearnley, I. M. et al. GRIM-19, a cell death regulatory gene product, is a subunit of bovine mitochondrial NADH:ubiquinone oxidoreductase (complex I). J. Biol. Chem. 276, 38345–38348 (2001).
35. Banci, L. et al. Structural characterization of CHCHD5 and CHCHD7: two atypical human twin CxO4-proteins. J. Struct. Biol. 180, 190–200 (2012).
36. Szklarczyk, R. et al. NDUFB7 and NDUFA8 are located at the intermembrane surface of complex I. FEBS Lett. 585, 737–743 (2011).
37. Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. Nature 504, 107–112 (2013).
38. Amunts, A. et al. Structure of the yeast mitochondrial large ribosomal subunit. Science 343, 1485–1489 (2014).
39. Allegretti, M., Mills, D. J., McMullan, G., Kühnlbrandt, W. & Vonck, J. Atomic model of the F420-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector. eLife 3, e00963 (2014).

Acknowledgements We thank R. Henderson, S. H. W. Scheres, G. McMullan, G. Murshudov, P. Emsley and J. E. Walker for helpful advice, the FEI fellows for educating us on use of the Titan Krios, J. Grimmett and T. Darling for computational help, and S. Chen and C. Savva for EM help. This work was supported by the Medical Research Council, grant numbers U105184322 (K.R.V., in R. Henderson’s group) and U105663141 (J.H.).

Author Contributions K.R.V. carried out EM experiments and analysis; J.Z. prepared methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 April; accepted 17 July 2014. Published online 7 September 2014.

©2014 Macmillan Publishers Limited. All rights reserved
METHODS
Protein preparation. Complex I was purified from *B. taurus* heart mitochondrial membranes by solubilization and anion exchange chromatography in n-dodecyl-β-D-maltoside (DDM), and size-exclusion chromatography in DDM or 7-cyclohexyl-1-heptyl-β-D-maltoside (Cymal 7) as described previously.

Cryo-EM specimen preparation and imaging. Aliquots of complex I (3 μl, 3–4.5 mg ml⁻¹) were applied to glow-discharged holey-carbon Quantifoil R 0.6/1 grids, blotted for 15–18 s, then plunge-frozen in liquid ethane using an environmental plunge-freeze apparatus. The grids were transferred into cartridges, loaded into an FEI Titan Krios electron microscope, and images were recorded at 2–5 μm per frame using post-focus CTF parameters estimated internally. Reference-free classification was performed using the default EMAN2 parameters, and classes with distinct orientations selected (Extended Data Fig. 1b for an example) to build initial models. The initial model that best matched the class averages was selected and two cycles of refinement performed in EMAN2. Subsequently, comparison of the model with the structure of complex I from *T. thermophila* (Protein Data Bank (PDB) accession 4HEA) suggested that it had the wrong hand; this observation was verified using tilt-pair analysis and corrected (Extended Data Fig. 2). All further refinements were performed in RELION, starting with maps that were low-pass filtered to 60 Å.

Typical micrographs prepared from complex I in Cymal 7 exhibited, on average, twice as many particles (~40 per micrograph) than those from complex I in DDM, so a larger data set was collected using Cymal 7. The reason for the difference in particle distribution is not clear—it may simply be a product of the grid preparation and freezing protocols. The class averages were used as a reference to pick particles automatically using RELION but many false positives were included, so all the images were inspected manually and particles too close to each other, aggregates and ice contaminants were deleted. The final data set contained 45,618 particles from 1,154 micrographs. The CTF was determined with CTFFIND3 (ref. 46) using the images summed from all 72 frames. Subsequently, refinement was performed using frames 1–32 of each image (the last 40 frames were discarded), to produce a map with resolution of 5.86 Å and orientational accuracy of 1.2°. To check for overfitting, phases were randomized beyond 10 Å on individual images (frames 1–32), followed by refinement as for the normal images. The results clearly show the presence of information beyond 10 Å (Extended Data Fig. 2). Note that RELION divides the data set into two halves at the initial step and calculates the resolution using a gold-standard Fourier shell correlation (FSC), so the phase randomization procedure serves here only as an additional control.

Modelling of the beam-induced movement of the complex I particles (using a running average of 11 movie frames in RELION) provided a modest improvement in resolution to 5.16 Å. The parameters from this analysis were then used to carry out a per-frame reconstruction in RELION (particle-polish), and a B-factor weighting was applied to each frame, resulting in a 4.8 Å map. The B-factor-weighted particles were subjected to 25 iterations of three-dimensional classification into 4 classes; this separated a major class containing 55% of the particles from smaller classes of 24, 14 and 6%. Difference maps revealed minor localized variations in some of the peripheral regions of the molecule, but no large-scale conformational variation was observed. The major class with 25,492 particles was refined and, after post-processing with RELION, a shape-mask, correction for the modulation transfer function (MTF) of the detector and a B-factor of −152 (ref. 48) were applied, and filtered to 4.95 Å resolution (Extended Data Fig. 2C, note that the magnification and CTF values have not been refined). Despite containing a lower number of particles, the maps from this major class and the whole data set were comparable. The 49.5 kDa subunit has dominant secondary structures and, except for the N-terminal peptide, could be computed to 3 Å. Similarly, the PSTT and YYKY subunits, with three FeS clusters, were readily built. The central part of the 30 kDa subunit, containing a mixture of α-helices and β-strands, could be traced, but the path of the long unstructured C terminus is unclear. The 75 kDa subunit is the least conserved hydrophilic core subunit. The small domain containing the three FeS clusters is well resolved and could be traced using the *T. thermophila* model, but considerable portions of the large, peripheral domain have poor density, low secondary structure content and low sequence similarity to *T. thermophila*, and so could not be traced confidently. The seven core subunits in the membrane domain could all be readily traced, except for a few loop regions, and assigned using their similarity to the *T. thermophila* subunits. The long transverse helix at the C terminus of the N-terminal domain is also not resolved.

It was observed that (as expected) the detergent–phospholipid belt is at lower resolution than the remaining membrane regions, and that (as expected) the detergent–phospholipid belt is at lower resolution than the remaining membrane regions. FeS clusters were located using the highest peak densities in the unsharpened map, and the subunits were built around them. The 24 and 51 kDa subunits were easily built as they are well conserved and the connectivity in the densities is clearly resolved. The 49.5 kDa subunit has dominant secondary structures and, except for the N-terminal peptide, could be computed to 3 Å. Similarly, the PSTT and YYKY subunits, with three FeS clusters, were readily built. The central part of the 30 kDa subunit, containing a mixture of α-helices and β-strands, could be traced, but the path of the long unstructured C terminus is unclear. The 75 kDa subunit is the least conserved hydrophilic core subunit. The small domain containing the three FeS clusters is well resolved and could be traced using the *T. thermophila* model, but considerable portions of the large, peripheral domain have poor density, low secondary structure content and low sequence similarity to *T. thermophila*, and so could not be traced confidently. The seven core subunits in the membrane domain could all be readily traced, except for a few loop regions, and assigned using their similarity to the *T. thermophila* subunits. The long transverse helix at the C terminus of the N-terminal domain is also not resolved.

Once the electron density for the core subunits had been assigned, models for the TMHs of the supernumerary subunits were built. A total of 18 TMHs were modelled, and when the density was clear they were extended. Connectivity was observed between four TMHs adjacent to subunit ND4 so they were combined into a single chain. To aid in superimposing subunit assignments, the secondary structure of each subunit was predicted using SSIPRED and TMs were predicted using TMHMM2 (ref. 52), HMPTOP2 (ref. 53) and the TOPCONS suite (seven methods in total) (Extended Data Table 4). Known structures of soluble proteins with high homology to the complex I supernumerary subunits were identified by HHpred and used to build homology models in Modeller and SwissModel (Extended Data Table 5). Regions of the density map with features corresponding to the predicted structures were located manually. Long loop regions were trimmed, then the models were placed in the density, jiggled fit in COOT was used to find the best fit, and the models were adjusted manually. Finally, several additional tubular densities in the map were built as α-helices. Most of them are located close to TMHs of the supernumerary subunits, but the connectivity to them is not clear; a higher resolution map will be necessary to assign these helices to their respective subunits.

41. Bellare, J. R., Davis, H. T., Scriven, L. E. & Talmon, Y. Controlled environment vitrification system: an improved sample preparation technique. J. Electron Microsc. Tech. 10, 87–111 (1988).
42. Grigorieff, N. Frealign: high-resolution refinement of single particle structures. J. Struct. Biol. 157, 117–125 (2007).
43. Smith, J. M. XMDISP—a visualization tool to aid structure determination from electron microscope images. J. Struct. Biol. 129, 223–228 (1999).
44. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. 157, 38–46 (2007).
45. Henderson, R. et al. Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy. J. Mol. Biol. 413, 1028–1046 (2011).
46. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. 142, 334–347 (2003).
47. Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single-particle electron cryo-microscopy. Ultramicroscopy 135, 24–35 (2013).
48. 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).
49. Kurucsev, L., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. Nature Methods 11, 63–65 (2014).
50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot 0.8. Acta Crystallogr. D 66, 486–491 (2010).
51. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292, 195–202 (1999).
52. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. P. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580 (2001).
53. Tusnady, G. E. & Simon, I. Principles governing amino acid composition of integral membrane proteins: applications to topology prediction. J. Mol. Biol. 283, 489–506 (1998).

PyMOL Molecular Graphics System or UCSC Chimera (http://www.cgl.ucsf.edu/ chimera/).
54. Bernsel, A., Viklund, H., Hennerdal, A. & Elofsson, A. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res.* **37**, W465–W468 (2009).

55. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).

56. Eswar, N. et al. Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinform*. Chapter 5, Unit 5.6 (2006).

57. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006).

58. Efremov, R. G. & Sazanov, L. A. Respiratory complex I: ‘steam engine’ of the cell? *Curr. Opin. Struct. Biol.* **21**, 532–540 (2011).

59. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* **60**, 2256–2268 (2004).

60. Balsa, E. et al. NDUFA4 is a subunit of complex IV of the mammalian electron transport chain. *Cell Metab.* **16**, 378–386 (2012).

61. Johansson, K. et al. Structural basis for substrate specificities of cellular deoxyribonucleoside kinases. *Nature Struct. Biol.* **8**, 616–620 (2001).

62. King, J. D. et al. Predicting protein function from structure - the roles of short-chain dehydrogenase/reductase enzymes in *Bordetella* O-antigen biosynthesis. *J. Mol. Biol.* **374**, 749–763 (2007).

63. Parris, K. D. et al. Crystal structures of substrate binding to *Bacillus subtilis* holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure* **8**, 883–895 (2000).
Extended Data Figure 1 | Single-particle cryo-EM analysis of *B. taurus* complex I.  

**a.** Typical micrograph of complex I particles imaged after freezing in vitreous ice on a holey-carbon grid. Some of the selected particles are marked with red boxes. Scale bar, 50 nm.

**b.** Two-dimensional reference classification showing particles lying in different orientations in the ice. The size of each box is 280 pixels and the two-dimensional classification was made in RELION®.
Extended Data Figure 2 | Validation of the map and resolution. a, Tilt-pair analysis of complex I in Cymal-7. One-hundred complex I particles from eight image pairs, recorded with a relative tilt angle of 10°, were extracted and subjected to tilt-pair analysis with FREALIGN. The outer radius of the plot is 40° and the orange circle centred at the expected tilt angle has a radius of 6°. b, Phase randomization to check for overfitting. Phases that are beyond 10 Å in each of the micrographs used in the final data set (frames 1–32) were randomized, and then refinement was performed as for a normal data set (FSC summed image corresponding to frames 1–32). As expected, the graph shows a drop in the Fourier shell correlation (FSC) curve at 10 Å, validating the presence of information beyond 10 Å in the images. Note that the use of gold-standard refinement procedures in RELION prevents any overfitting, and this test was done only as an additional control. c, An overview of the final map and the model built into it. d, FSC curves of the final map and of the model versus the map. The curve in red is the gold-standard FSC of the final map (after classification) and the resolution at FSC = 0.143 is ~4.95 Å. The curve in cyan is the FSC between the final map and the model, and at FSC = 0.5 the resolution is 6.7 Å. Note that the present model is not complete since it is only a polyalanine model without any side chains, and loop regions in a number of subunits have not been modelled. e, The final map of mammalian complex I was analysed with ResMap. The left-hand panel (with lower density threshold) shows that the detergent–phospholipid belt is of lower resolution, and most of the protein regions of the map show resolution distributed from 5 to 6 Å. In the right-hand panel the map is shown at a higher density threshold, so the detergent–phospholipid belt is not visualized. Some of the interior parts of the map have resolution of 4.8–5 Å.
Extended Data Figure 3 | Example regions of the density map with the model fitted to the map.  

a, ND2 is shown from the membrane plane, highlighting the densities for three aromatic side chains and one of the helix-breaking loops.  
b, Subunit ND4 viewed from the matrix.  
c, The density for a [4Fe–4S] cluster and surrounding protein is shown in the PSST subunit.  
d, A region of the 49 kDa subunit shows a well resolved α-helical stretch and aromatic side chains, and the β-strands are beginning to be resolved.  
e, Subunit B8 is an example of a supernumerary subunit in a peripheral region of the molecule.  
f, Density consistent with a bound nucleotide is observed in the 39 kDa subunit, in a similar position to in homologous structures and as expected from analysis of *Y. lipolytica* complex I (ref. 27). However, the present resolution of the map precludes the inclusion of this nucleotide in the final model.
Extended Data Figure 4 | Global comparison of the core subunit structures of bacterial and mammalian complex I. The core subunits from *B. taurus* are in blue, and from *T. thermophilus* (PDB accession 4HEA) in orange. The structures have been superimposed using ND1 (the heel subunit). Top: the ND2, ND4 and ND5 domain is rotated in *B. taurus* relative to in *T. thermophilus*, increasing the curvature in the *B. taurus* membrane domain. The complex is viewed along the 11° rotation vector (orange) that maps the *T. thermophilus* ND2, ND4 and ND5 domain to the *B. taurus* domain, along with a small 5 Å translation to superimpose the domain centres. Correspondingly, the ND3, ND4L and ND6 domains are superimposed by a 4° rotation and a 1 Å translation. Rotation of ND2, 4 and 5 about the long axis of the domain, as noted for *Y. lipolytica*[^1], is not observed. Bottom: the NADH dehydrogenase domain containing the 51 and 24 kDa subunits is rotated by 23° and translated by 14 Å in *B. taurus*, relative to in *T. thermophilus*, causing the FeS chains to diverge as the distance from ND1 increases. A similar rotation was observed in *Y. lipolytica*. The complex is viewed from behind ND1. Correspondingly, the 49 kDa, PSST and TYKY subunits are superimposed by a 6° rotation and a 2 Å translation. The structures were analysed using Superpose from the CCP4 suite[^2] and the 75 kDa and 30 kDa subunits were not included due to their lower structural conservation.
Extended Data Figure 5 | Comparison of the individual structures of the core subunits of bacterial and mammalian complex I. 

a, The structure of each subunit from *T. thermophilus* (wheat) (PDB accession 4HEA) has been superimposed separately on its corresponding subunit from *B. taurus* (coloured as labelled) with the transverse helix plus TMH16 of ND5 also aligned separately. The complexes are viewed from behind ND1 (top), from the side (middle) and from the matrix (bottom, ND subunits only). 

b, Observed differences in the structures of the core subunits of *B. taurus* and *T. thermophilus* complexes I. Grey, conserved structure from *B. taurus* and *T. thermophilus* (PDB accession 4HEA); red, structural elements present only in *T. thermophilus*; blue, structural elements present only in *B. taurus*. The C-terminal domain of the 75 kDa subunit is not resolved in *B. taurus*, but its structure is clearly different to in *T. thermophilus*. 

©2014 Macmillan Publishers Limited. All rights reserved
## Extended Data Table 1 | Reference table for the nomenclature of the core subunits of complex I

| Domain          | Chain identifier | Bos taurus | Homo sapiens | Yarrowia lipolytica | Thermus thermophilus | Escherichia coli |
|-----------------|------------------|------------|--------------|---------------------|----------------------|------------------|
| Hydrophilic domain |                  |            |              |                     |                      |                  |
| G               | 75 kDa           | NDUF51     | NUAM         | Nqo3                | NuoG                 |                  |
| F               | 51 kDa           | NDUFV1     | NUBM         | Nqo1                | NuoF                 |                  |
| D               | 49 kDa           | NDUF52     | NUCM         | Nqo4                |                      | NuoCD            |
| C               | 30 kDa           | NDUF53     | NUGM         | Nqo5                |                      |                  |
| E               | 24 kDa           | NDUFV2     | NUHM         | Nqo2                |                      | NuoE             |
| B               | PSST             | NDUF57     | NUKM         | Nqo6                |                      | NuoB             |
| I               | TYKY             | NDUF58     | NUIM         | Nqo9                |                      | NuoI             |
| Membrane domain |                  |            |              |                     |                      |                  |
| H               | ND1              | ND1        | NU1M         | Nqo8                | NuoH                 |                  |
| N               | ND2              | ND2        | NU2M         | Nqo14               | NuoN                 |                  |
| A               | ND3              | ND3        | NU3M         | Nqo7                |                      | NuoA             |
| M               | ND4              | ND4        | NU4M         | Nqo13               |                      | NuoM             |
| L               | ND5              | ND5        | NU5M         | Nqo12               |                      | NuoL             |
| J               | ND6              | ND6        | NU6M         | Nqo10               |                      | NuoJ             |
| K               | ND4L             | ND4L       | NULM         | Nqo11               |                      | NuoK             |

In the text the names of the subunits from *B. taurus* are used, with the names from the human enzyme presented alongside as appropriate.
Extended Data Table 2 | Summary of the models of the core subunits of *B. taurus* complex I

| Subunit | Total residues* | Modeled residues | Poorly resolved / uncertain residue numbering | Unresolved residues (>10 residues) | %Modelled | %Identity† | r.m.s.d.‡ |
|---------|----------------|-----------------|---------------------------------------------|----------------------------------|-----------|------------|-----------|
| ND1     | 318            | 3 - 200         | 219 - 242                                   | 253 - 315                        | 90% (285/318) | 42% (132/318) | 1.60 Å   |
|         |                |                 |                                             | Matrix loop (TMH 5 - 6)           |           |            |          |
|         |                |                 |                                             | IMS loop (TMH 6 - 7)              |           |            |          |
| ND2     | 347            | 2 - 300         | 320 - 346                                   | 1                                | 94% (326/347) | 25% (86/347) | 2.08 Å   |
|         |                |                 |                                             | Matrix loop (TMH 10 - 11)         |           |            |          |
| ND3     | 115            | 2 - 23          | 52 - 112                                    | 1                                | 72% (83/115) | 27% (31/115) | 2.05 Å   |
|         |                |                 |                                             | Matrix loop (TMH 1 - 2)           |           |            |          |
| ND4     | 459            | 3 - 415         | 430 - 455                                   | 1 - 2                            | 96% (439/459) | 24% (111/459) | 2.20 Å   |
|         |                |                 |                                             | Matrix loop (TMH 13 - 14)         |           |            |          |
| ND4L    | 98             | 1 - 84          |                                              | 85 - 98                          | 86% (84/98) | 21% (21/98) | 2.66 Å   |
|         |                |                 |                                             | Matrix loop (C-terminus)          |           |            |          |
| ND5     | 606            | 4 - 22          | 28 - 358                                    | 363 - 400                        | 92% (558/606) | 31% (187/606) | 2.53 Å   |
|         |                |                 |                                             | TMH1                             | 1 - 3      |            |          |
|         |                |                 |                                             | 23 - 27                          |           |            |          |
|         |                |                 |                                             | 359 - 362                        |           |            |          |
|         |                |                 |                                             | 401 - 407                        |           |            |          |
|         |                |                 |                                             | 487 - 488                        |           |            |          |
|         |                |                 |                                             | 514 - 519                        |           |            |          |
|         |                |                 |                                             | 608 - 608                        |           |            |          |
|         |                |                 |                                             | Matrix loop (TMH 1 - 2)           |           |            |          |
|         |                |                 |                                             | Matrix loop (TMH 11 - 12)         |           |            |          |
|         |                |                 |                                             | IMS loop (TMH 12 - 13)            |           |            |          |
|         |                |                 |                                             | IMS loop (TMH 14 - 15)            |           |            |          |
|         |                |                 |                                             | TMH15 to transverse helix         |           |            |          |
| ND6     | 175            | 2 - 76          | 85 - 107                                    | 140 - 172                        | 75% (131/175) | 16% (29/175) | 1.83 Å   |
|         |                |                 |                                             | Matrix loop (TMH 3 - 4)           |           |            |          |
|         |                |                 |                                             | IMS loop (TMH 4 - 5)              |           |            |          |
| 75 kDa  | 704            | 8 - 125         | 136 - 318                                   | 326 - 347                        | 75% (52/704) | 27% (189/704) | 1.96 Å   |
| NDUFS1  |                |                 |                                             | The large domain (222 - 704)      | 1 - 7      |            |          |
|         |                |                 |                                             | is generally poorly resolved.     | 126 - 135  |            |          |
|         |                |                 |                                             | The sequence alignment is weak    | 319 - 325  |            |          |
|         |                |                 |                                             | and the secondary structure       | 348 - 368  |            |          |
|         |                |                 |                                             | content low. Residues 404 - 629   | 400 - 403  |            |          |
|         |                |                 |                                             | are particularly poorly resolved. | 411 - 424  |            |          |
|         |                |                 |                                             | Probable loop region              | 496 - 524  |            |          |
|         |                |                 |                                             | 531 - 541                        |           |            |          |
|         |                |                 |                                             | 628 - 704                        |           |            |          |
| 51 kDa  | 444            | 31 - 441        | Flavin and NADH binding site (63 - 72,     | 1 - 30                            | 93% (41/444) | 43% (191/444) | 1.61 Å   |
| NDUFY1  |                |                 | 99 - 104, 181 - 189, 300 - 304, 327 - 333 | 442 - 444                        |           |            |          |
| 49 kDa  | 430            | 47 - 430        | 3-strand β-sheet (47 - 79)                  | 1 - 46                            | 89% (364/430) | 42% (179/430) | 1.41 Å   |
| NDUFS2  |                |                 |                                             | N-terminal region                 |           |            |          |
| 30 kDa  | 228            | 15 - 168        | Numbering uncertain to 72                   | 1 - 14                            | 68% (154/228) | 24% (54/228) | 1.66 Å   |
| NDUFS3  |                |                 | Loop β-strand (73 - 83)                     | 169 - 228                         |           |            |          |
| 24 kDa  | 217            | 20 - 178        | Loop 126 - 132                              | 1 - 19                            | 73% (159/217) | 27% (59/217) | 1.57 Å   |
| NDUFSV2 |                |                 |                                             | 179 - 217                         |           |            |          |
| PSST    | 179            | 27 - 169        | Loop 68 - 79                                | 1 - 26                            | 80% (143/179) | 49% (89/179) | 1.44 Å   |
| NDUFS7  |                |                 |                                             | 170 - 179                         |           |            |          |
| TYKY    | 176            | 15 - 176        |                                             | 1 - 14                            | 92% (162/176) | 36% (63/176) | 1.89 Å   |
| NDUFS8  |                |                 |                                             | N-terminal peptide                |           |            |          |

* For proteins with a mitochondrial-targeting pre-sequence, residue 1 is the first residue of the mature protein.
† The percentage identity and the root mean squared deviation (r.m.s.d., calculated using PDBeFOLD) are between the sequences and structures of the subunits of *B. taurus* and *T. thermophilus* (PDB accession 4HEA) complex I.
Extended Data Table 3 | Distances between the redox cofactors in structural models of complex I

| Cofactors*                          | T. thermophilus | B. taurus |
|-------------------------------------|-----------------|----------|
|                                     | hydrophilic domain (2FUG.pdb) | complex I (4HEA.pdb) | complex I (this work) |
|                                     | centre† edge‡  | centre† edge‡  | centre† edge‡  |
| N1a - Flavin                        | 15.4 12.3     | 15.9 13.1     | 15.9† 13.1†     |
| Flavin - cluster 1 (N3)             | 12.5 7.6      | 12.2 7.3      | 12.2† 7.2†      |
| N1a - cluster 1 (N3)                | 22.1 19.4     | 22.3 19.7     | 21.1 18.0       |
| Cluster 1 (N3) - cluster 2           | 14.0 11.0     | 13.7 10.7     | 14.0 11.0       |
| Cluster 1 (N3) - cluster 3           | 17.4 13.8     | 17.1 13.4     | 18.4 14.5       |
| Cluster 2 - cluster 3                | 13.5 10.7     | 13.0 9.9      | 12.7 9.7        |
| Cluster 3 - cluster 4                | 12.2 8.5      | 12.4 8.6      | 12.8 8.7        |
| Cluster 4 - cluster 5                | 16.8 14.0     | 16.5 13.6     | 16.8 14.0       |
| Cluster 5 - cluster 6                | 12.1 9.4      | 12.1 9.3      | 12.1 9.3        |
| Cluster 6 - cluster 7 (N2)           | 13.7 10.5     | 13.5 10.2     | 13.6 10.5       |
| Cluster 1 (N3) - cluster 7 (N2)      | 61.1 57.6     | 60.5 57.0     | 61.5 58.1       |

* The [2Fe–2S] cluster in the 24 kDa subunit (known as N1a) is on the other side of the flavin from the main cofactor chain. The [4Fe–4S] cluster in the 51 kDa subunit (known as N3) is the first cluster in the chain and the [4Fe–4S] cluster in subunit PSST (known as N2) is the last (seventh) cluster in the chain.
† The distances are in Å, between the geometric centres of the Fe and S cluster cores or the flavin isoalloxazine ring system (centre), or between the centres of the two closest atoms (edge) as commonly used in calculations of electron transfer rates. Distances are estimated to be accurate to within 1 Å.
‡ The position of the flavin in B. taurus is poorly resolved and has been approximated using its position in PDB accession 4HEA.
Extended Data Table 4 | Knowledge about the supernumerary subunits of *B. taurus* complex I

| *B. taurus* subunit | *H. sapiens* subunit | Subcomplex | Sequence information | Predicted TMHs |
|---------------------|----------------------|------------|----------------------|---------------|
| 10 kDa NDUFA4        | Iα and Iβ            | Subcomplex I | CX8C motif, intermembrane space | 0          |
| 18 kDa NDUFA3        | Iα and Iβ            | Subcomplex I | PFAM zinc-finger motif, CX2H2X12C | 0          |
| 15 kDa NDUFA5        | Iα only              | Subcomplex I | CX8C motif, intermembrane space | 0          |
| 13 kDa NDUFA6        | Iα and Iβ            | Subcomplex I | CX8H2X12C | 0          |
| MWFE NDUFA1          | Iα only              | Subcomplex I | Short-chain dehydrogenase reductase family, NADP | 0          |
| B8 NDUFA2            | Iα and Iβ            | Subcomplex I | Acyl-carrier protein | 0          |
| B9 NDUFA3            | Iα only              | Subcomplex I |             | 0          |
| B13 NDUFA5           | Iα and Iβ            | Subcomplex I | Two CX8C motifs, PFAM CHCH domain | Two CX8C motifs, PFAM CHCH domain | 0          |
| B14 NDUFA8           | Iα only              | Subcomplex I | Lysine-rich motif | 0          |
| B14.5a NDUFA7        | Iα and Iβ            | Subcomplex I |             | 0          |
| 39 kDa NDUFA9        | Iα only              | Subcomplex I |             | 0          |
| 42 kDa NDUFA10       | Iα only (low level)  | Subcomplex I | Similarity to deoxynucleoside kinases | 0          |
| B14.7 NDUFA11        | Iα (Iβ at low level) | Subcomplex I |             | 3 or 4     |
| B17.2 NDUFA12        | Iα and Iβ            | Subcomplex I |             | 0          |
| B16.6 NDUFA13        | Iα and Iβ            | Subcomplex I |             | 1          |
| SDAP NDUFA81         | both Iα and Iβ       | Subcomplex I | Acyl-carrier protein | 0          |
| MNLL NDUFB1          | Iβ                   | Subcomplex I |             | 0 (or 1)   |
| AGGG NDUFB2          | Iβ                   | Subcomplex I |             | 0 (or 0)   |
| B12 NDUFB3           | Iβ                   | Subcomplex I |             | 1          |
| B15 NDUFB4           | both Iα and Iβ       | Subcomplex I |             | 1          |
| SGDH NDUFB5          | Iβ                   | Subcomplex I |             | 1          |
| B17 NDUFB6           | Iβ                   | Subcomplex I |             | 1          |
| B18 NDUFB7           | Iβ                   | Subcomplex I | CX8C motif, intermembrane space | 0          |
| ASHL NDUFB8          | Iβ                   | Subcomplex I |             | 1          |
| B22 NDUFB9           | Iβ                   | Subcomplex I | Lysine-rich motif | 0          |
| PDSW NDUFB10         | Iβ                   | Subcomplex I |             | 0          |
| ESSS NDUFB11         | Iβ                   | Subcomplex I |             | 1          |
| KFYI NDUFC1          | none                 | Subcomplex I |             | 1          |
| B14.5b NDUFC2        | Iβ (low level)       | Subcomplex I |             | 1 or 2     |

*The former subunit MLRQ (NDUFA4) is no longer considered a subunit of complex I (ref. 60). Subcomplex I, which contains the 7 hydrophilic core subunits and 8–9 supernumerary subunits, represents a considerable portion of the hydrophilic domain of complex I. Subcomplex Iα, which contains all the subunits of subcomplex I plus core subunit ND6 and 9–10 additional supernumerary subunits, represents the hydrophilic domain of complex I plus associated membrane subunits. Subcomplex Iβ, which contains ND4 and ND5 and 12–13 supernumerary subunits, represents part of the membrane domain. TMHs were predicted using TMHMM2 (ref. 52), HMMTOP2 (ref. 53) and the TOPCONS suite (seven methods in total) and are presented as consensus values with less represented values in brackets and single outliers discarded.*
## Extended Data Table 5 | Summary of the models of the supernumerary subunits of *B. taurus* complex I

| Subunit | Chain identifier | Total residues | PDB model | Aligned residues | %Identity | Modeled residues | %Modeled | r.m.s.d. |
|---------|------------------|----------------|------------|------------------|-----------|-----------------|----------|---------|
| 42 kDa  | NDUFA10          | O              | 320        | 2OC(F)          | 21 - 262  | 21% (49/232)    | 57% (181/320) | 1.91 Å  |
| 39 kDa  | NDUFA9           | P              | 345        | 2Q1W            | 19 - 325  | 13% (41/307)    | 73% (252/345) | 2.52 Å  |
| 18 kDa(5) | NDUFS4         | Q              | 133        | 2JYA            | 33 - 133  | 37% (37/101)    | 52% (69/133) | 2.42 Å  |
| 13 kDa(5) | NDUFS6         | R              | 96         | 2JRR            | 44 - 96   | 34% (18/53)     | 49% (47/96)  | 1.97 Å  |
| B8      | NDUFA2           | S              | 99         | 1S3A           | 1 - 99    | 94% (93/99)     | 81% (80/99)  | 2.18 Å  |
| SDAP-α  | NDUFA8           | T              | 88         | 1F800          | 8 - 84    | 36% (28/77)     | 81% (71/88)  | 1.18 Å  |
| SDAP-β  | NDUFA8           | U              | 88         | 1F800          | 8 - 84    | 36% (28/77)     | 85% (75/88)  | 1.36 Å  |
| B13(5)  | NDUFA5           | V              | 116        |                 | 1 - 71(1) | 61% (71/116)    |          |        |
| B14     | NDUFA6           | W              | 128        |                 | 1 - 72(1) | 56% (72/128)    |          |        |
| PGIV    | NDUFA8           | X              | 172        | 2LQL           | 35 - 114  | 23% (18/80)     | 46% (79/172) | 2.40 Å  |
| B14.7   | NDUFA11          | Y              | 141        |                 | 1 - 106(1)| 75% (106/141)   |          |        |
| B16.6   | NDUFA13          | Z              | 144        |                 | 33 - 97   | 45% (65/144)    |          |        |
| B9(5)   | NDUFA3           | a              | 154        |                 | 1 - 29(1) | 46% (71/154)    |          |        |

* For proteins with a mitochondrial-targeting pre-sequence, residue 1 is the first residue of the mature protein.  
† Known structures with high homology to the complex I subunits were identified by HHpred.  
‡ The percentage identity and the r.m.s.d. (calculated using PDBeFOLD) are between the sequences and structures of the subunits of *B. taurus* complex I and the PDB models.  
§ Subunit with less certain assignment.  
*Residue numbers are arbitrary and not assigned to the sequence.