Published in final edited form as:
Nature. 2014 November 6; 515(7525): 80–84. doi:10.1038/nature13686.

Architecture of mammalian respiratory complex I

Kutti R. Vinothkumar[#1], **Jiapeng Zhu**[#2], and **Judy Hirst**[2]

[1]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, CB2 0QH, UK

[2]MRC Mitochondrial Biology Unit, Wellcome Trust / MRC Building, Hills Road, Cambridge, CB2 0XY, UK

[#] These authors contributed equally to this work.

Abstract

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 fourteen 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 significantly 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[1] is one of the largest and most complicated enzymes in the cell. Complex I from *Bos taurus* (bovine) heart mitochondria has been characterised extensively as a model for the human enzyme; both enzymes contain 44 different subunits (encoded by both the nuclear and mitochondrial genomes)[2,3] and nine redox cofactors (a flavin mononucleotide and eight iron-sulphur clusters). Fourteen subunits are the 'core' subunits that are conserved in all complexes I; they contain all the mechanistically-critical cofactors and structural elements and are sufficient for catalysis. Crystal structures of intact complex I

Correspondence and requests for materials should be addressed to vkumar@mrc-lmb.cam.ac.uk and jh@mrc-mbu.cam.ac.uk.
Author contributions: KRV carried out electron microscopy experiments and analysis; JZ prepared protein; KRV, JZ and JH modelled and analysed data; JH designed the project; KRV, JZ and JH wrote the paper.

The authors declare no competing financial interests.

Map and model deposition: The EM map of complex I and the associated model have been deposited with accession numbers EMD-2676 and 4UQ8.pdb.

Europe PMC Funders Author Manuscripts

from the thermophilic bacterium *Thermus thermophilus*[4], and of domains of the prokaryotic enzymes from *T. thermophilus* and *Escherichia coli*[5-7] 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 enzyme[2,3] 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.

Due to its size, L-shaped asymmetry, membrane-bound location, and multi-component structure, mammalian complex I has proved difficult to crystallise, 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 modelled[8]. 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 visualised to display their overall L-shaped structures[9-11] although at too low 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 diseases[12].

Imaging and Reconstruction

Complex I was purified from *Bos taurus* (bovine) heart mitochondria in detergent[13], 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 2D 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 RELION[14] and movie frames were used to correct for beam-induced movement[15]. Per-frame reconstruction and b-factor weighting were followed by 3D 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 α-helices (TMHs), are observed. The highest-density peaks in the map reveal the eight iron-sulphur (FeS) clusters that, as in the *T. thermophilus*[4,7] and *Y. lipolytica*[8] enzymes, form a chain through the hydrophilic domain.

Structures of the Core Subunits

The 14 conserved core subunits of complex I[1,4] catalyse the energy transducing reactions: NADH oxidation, ubiquinone reduction and proton translocation (Extended Data Table 1 summarises their nomenclature). The seven nuclear-encoded hydrophilic core subunits harbour a flavin mononucleotide to oxidise 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[4], secondary structure analyses, and sequence alignments, and using structural features and densities from aromatic sidechains (Extended Data Fig. 3). Except for the FeS-cluster ligands they have been modelled as polyalanine chains, with the residue numbering optimised 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 bacterial[4] 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 sixty TMHs of the mammalian core subunits match their *T. thermophilus* counterparts closely (Extended Data Fig. 5) — only the position of ND6-TMH4 is different, and the extra C-terminal TMH particular to *T. thermophilus* ND1 is absent from *B. taurus* (Fig. 2, Extended Data Fig. 5). No significant density is observed in place of the three N-terminal TMHs (present in *T. thermophilus* and *Y. lipolytica*) that have been lost through evolution of mammalian ND2[16], so they have not been substituted structurally by other subunits. Importantly, catalytically-relevant features identified in the antiporter-like subunits of bacterial complex[5,6] are conserved. They include the loops in the six broken TMHs in ND2, ND4 and ND5 (see, for examples, Extended Data Fig. 3) that may constitute part of the proton-translocation mechanism, and the long transverse helix in ND5, a proposed coupling element.

Of the seven hydrophilic core subunits (Fig. 2), the structures of the *B. taurus* 51 kDa (NDUFV1), 49 kDa (NDUFS2), 24 kDa (NDUFV2), PSST (NDUFS7) and TYKY (NDUFS8) subunits, and the small domain of the 75 kDa subunit (NDUFS1) are closely conserved from their *T. thermophilus* homologues[4,7] (Extended Data Fig. 5), with significant 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, due 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*[7] and the 30 kDa subunit (NDUFS3), 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 (NDUFS7) and TYKY (NDUFS8) from oxidative damage. Conversely, the area around the NADH-binding

site, where complex I produces superoxide[17], is bare (Fig. 3), so supernumerary subunits do not shield it from O_2 to minimise superoxide production. The NADH dehydrogenase domain is added at the end of the complex I assembly pathway[18], and the local paucity of supernumerary subunits may facilitate both its integration and its replacement (while retaining the rest of the protein) to mitigate the effects of oxidative damage[19]. Two large supernumerary domains capping ND5 and part of ND4, and 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 play a role similar to that of the stabilising β-hairpin-helix structures observed in the prokaryotic enzyme[5]. 18 supernumerary TMHs are distributed around the core membrane domain (Figs. 3, 4), consistent with the predictions of sequence analyses for 14 to 18 TMHs from these subunits (Extended Data Table 4). In total, therefore, we observe 78 TMHs in the mammalian enzyme. Two TMHs are on the outside of the ND5 transverse helix, appearing to strap it to the core domain, and four more are positioned close to the end of it, appearing as a restraint for its lateral movement (Fig. 3). These observations raise the question of whether large-scale piston-like motions of this helix during catalysis, as postulated from the *T. thermophilus* structure[6], are feasible.

Assignment of 14 Supernumerary Subunits

To identify and assign individual supernumerary subunits to the map for mammalian complex I (Extended Data Table 4 summarises 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, 5). Human B8 (NDUFA2) adopts a thioredoxin fold[20] and its structure (Fig. 5) was located at the tip of the large domain of the 75 kDa subunit (NDUFS1) (Fig. 4), so (contrary to current models[18]) 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[21], and, along with other NADH dehydrogenase domain subunits, it is rapidly exchanged under steady-state conditions[19]. 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[18], the 18 kDa (NDUFS4) and 13 kDa (NDUFS6) subunits (Extended Data Tables 4, 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[22,23].

The 42 kDa subunit (NDUFA10), a member of the nucleoside kinase family[24], was easily located as the density on top of ND2, on the matrix side of the membrane (Figs. 4, 5, Extended Data Tables 4, 5). Its location is neatly confirmed by its absence from the density map of *Y. lipolytica* complex I[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 activity[25], implying both its regulatory role, and a molecular link between PINK1 dysfunction and complex I activity in Parkinson's disease. Further elucidation of this regulatory pathway must now be reconciled with the matrix location of the 42 kDa subunit. The 39 kDa subunit (NDUFA9), a member of the nucleotide-binding short-chain dehydrogenase/reductase family[26] (Extended Data Table 4, 5), was readily located adjacent to PSST (NDUFS7) (Figs. 4, 5), and observed to contain a density consistent with a bound dinucleotide (Extended Data Fig. 3)[27]. Furthermore, it partially encloses the long ND3 loop (resolved only in *T. thermophilus*[4]) that is critical for coupling electron and proton transfer, and which, in a conformational transition now known to also involve the 39 kDa subunit[28], switches the enzyme into a 'deactive' state during ischaemia. Notably, these several proposed regulatory elements are all located close to the junction between the hydrophilic and membrane domains where the energy from the redox reaction is used to initiate proton translocation[1].

Two regions of density corresponding to the structure of the SDAP subunit (NDUFAB1), which is identical to the acyl carrier protein in the mitochondrial matrix[29,30], were identified in the mammalian enzyme (Fig. 4). This result is supported by the presence of SDAP in both subcomplex Iα (which contains the hydrophilic domain) and subcomplex Iβ (the distal portion of the membrane domain) of *B. taurus* complex I[3], and with the presence of two SDAP homologues in *Y. lipolytica* complex I[31]. One SDAP is located at the distal end of the enzyme, above ND5, the other in a peripheral region of the hydrophilic domain, in a subdomain that interacts with the 49 kDa (NDUFS2) and 30 kDa (NDUFS3) core subunits through a three-helix bundle (Figs. 4, 5). From a recent study in *Y. lipolytica*[32] these helices are assigned to subunit B14 (NDUFA6), a protein with an LYR motif that, when deleted in *Y. lipolytica*, results in loss of catalytic activity. Notably, subunit B22 (NDUFB9) also contains an LYR motif and it is in subcomplex Iβ, so it is possible that the distal SDAP molecule interacts with it in a similar fashion. Finally, subunits B13 (NDUFA5) and B14 have similar predicted secondary structures so we ascribe the second three-helix bundle observed on the side of 30 kDa (NDUFS3), adjacent to 42 kDa (NDUFA10), to B13 (Figs. 4, 5).

Continuous density links the four supernumerary TMHs at the end of the transverse helix. Only one subunit, B14.7 (NDUFA11), is predicted to contain more than two TMHs (Extended Data Table 4); secondary structure analyses predict that the first two are unusually long (~30 residues) so they probably correspond to the two highly-tilted TMHs. Therefore, these four TMHs are assigned to subunit B14.7 (Figs. 4, 5, Extended Data Table 5), a protein that is important for the assembly and/or stability of the membrane domain[33]. A second cluster of three TMHs (opposite B14.7) may include two TMHs from B14.5b (NDUFC2), but the connectivity between them is ambiguous and a clear assignment cannot be made. The 11 remaining TMHs are spread around the membrane domain (Fig. 4). Three TMH-containing subunits, B16.6 (NDUFA13), MWFE (NDUFA1) and B9 (NDUFA3), remain associated with the hydrophilic domain in subcomplex Iα, following fractionation of *B. taurus* complex I with zwitterionic detergents[3], and they are missing from *Y. lipolytica* subcomplex Iα, which lacks core subunits ND1, 2, 3 and 4L[34]. Therefore, they are assigned to the three TMH-densities next to ND1 (Extended Data Table 4, 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 (NDUFS8) on the matrix side, spans the membrane, then bends into the intermembrane space and is anchored under the 'heel' (Figs. 4, 5). Therefore, this density is assigned to B16.6, a protein identical to cell death regulatory gene product GRIM19[35]. It is currently not possible to confidently deduce assignments for the TMH-containing subunits of subcomplex Iβ.

The PGIV (NDUFA8), 15 kDa (NDUFS5) and B18 (NDUFB7) subunits contain twin-CX_9C motifs that form two intramolecular disulphides within 'CHCH' domains[36], and they have been assigned to the inter-membrane surface of complex I[37]. A double L-shaped density, resembling two CHCH domains at right-angles, is clearly visible on the heel, clamping B16.6 (NDUFA13, Figs. 4, 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, 5), consistent with the position of an antibody label to its homologous subunit in *Y. lipolytica*[34]. Our structure thus reveals the architecture of the '400 kDa' assembly intermediate of human complex I[18,33] that contains the core hydrophilic subunits 49 kDa (NDUFS2), 30 kDa (NDUFS3), PSST (NDUFS7), and TYKY (NDUFS8), core membrane subunit ND1, and the supernumerary subunits PGIV (NDUFA8), B9 (NDUFA3), B16.6 (NDUFA13), and B13 (NDUFA5).

Conclusions and Perspectives

In summary, 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 complexes I, 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[4]. The 14 supernumerary subunits assigned 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[38-40]. 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 characterise the structural changes that occur during catalysis, a crucial step in defining the mechanism of electron-coupled proton translocation.

METHODS

Protein preparation

Complex I was purified from *B. taurus* heart mitochondrial membranes by solubilisation 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[13].

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[41]. The grids were transferred into cartridges, loaded into an FEI Titan Krios electron microscope, and images were recorded at 2-5 μm underfocus on a Falcon II CMOS (complementary metal oxide semiconductor) direct electron detector at 300 keV at 81,495× magnification (nominally 47,000×), with the specimen temperature at −186 °C using the EPU software (Extended Data Fig. 1). The detector pixel size of 14 μm corresponds to a sampling density of ~1.72 Å $pixel^{-1}$. Each image was exposed for 4 s (total dose ~64 e $Å^{-2}$) and 72 frames were captured. For tilt-pair analysis, the same area was imaged twice, at 0° for 0.8 s and then at 10° for 2.0 s, and analysed with FREALIGN[42].

Image processing and 3D reconstruction

An initial data set for complex I in DDM was obtained by manually picking particles with XIMDISP[43]. 7630 particles, from 366 micrographs, were used to generate initial maps in EMAN2[44]. Particles were boxed in 280 × 280 pixels and contrast transfer function (CTF) parameters estimated internally. Reference-free classification was performed using the default EMAN2 parameters, and classes with distinct orientations selected (see 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. thermophilus* (4HEA.pdb[4]) suggested that it had the wrong hand; this observation was verified using tilt-pair analysis[45] and corrected (Extended Data Fig. 2). All further refinements were performed in RELION[14], 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[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[47]. 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 randomisation 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 3D 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 localised 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[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. Analysis of the local resolution by ResMap[49] showed that the core sections of the molecule, particularly the TMHs, have higher resolutions than the peripheral sections, and that (as expected) the detergent-phospholipid belt is at lower resolution (Extended Data Fig. 2E).

Model building

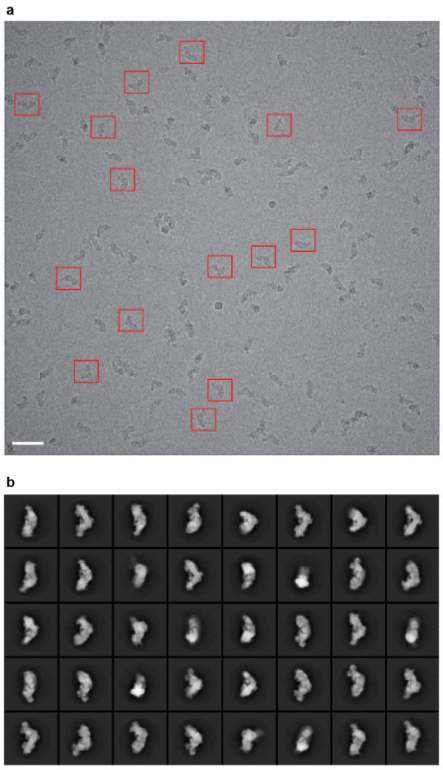
Model building was performed using Coot[50]. All the models described have been built as polyalanine chains, except for the residues that coordinate the FeS clusters. Note that the present model has not been refined, so it inevitably contains some errors and inaccuracies. Examples of the model fitted to the electron density are shown in Extended Data Fig. 3, and figures were created using the PyMOL Molecular Graphics System (Schrödinger, LLC) or UCSF Chimera (http://www.cgl.ucsf.edu/chimera/).

The seven core hydrophilic subunits of *B. taurus* complex I were modelled initially using the coordinates of *T. thermophilus* complex I (4HEA.pdb[4]) as a template, trimmed where the densities were ambiguous, and adjusted manually. 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 kDa subunit has dominant secondary structures and, except for the N-terminal peptide, could be completely traced. Similarly, the PSST and TYKY 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 easily using the *T. thermophilus* model, but significant portions

of the large, peripheral domain have poor density, low secondary structure content and low sequence similarity to *T. thermophilus*, 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. thermophilus* subunits. The long transverse helix at the C-terminus of subunit ND5 is well ordered and extends over ND4 and ND2. In better resolved regions of the map, protruding densities that are likely to be side chains of aromatic residues are observed (see Extended Data Fig. 3), and these features, along with secondary structure information and sequence alignments, were used to produce an optimised assignment for the *B. taurus* residue numbers in the modelled subunits (Extended Data Table 2), for use as a guide to the positions of individual residues.

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 supernumerary subunit assignments, the secondary structure of each subunit was predicted using PSIPRED[51] and TMHs were predicted using TMHMM2[52], HMMTOP2[53] and the TOPCONS suite[54] (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[55] and used to build homology models in Modeller[56] and SwissModel[57] (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, jiggle 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 from 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.

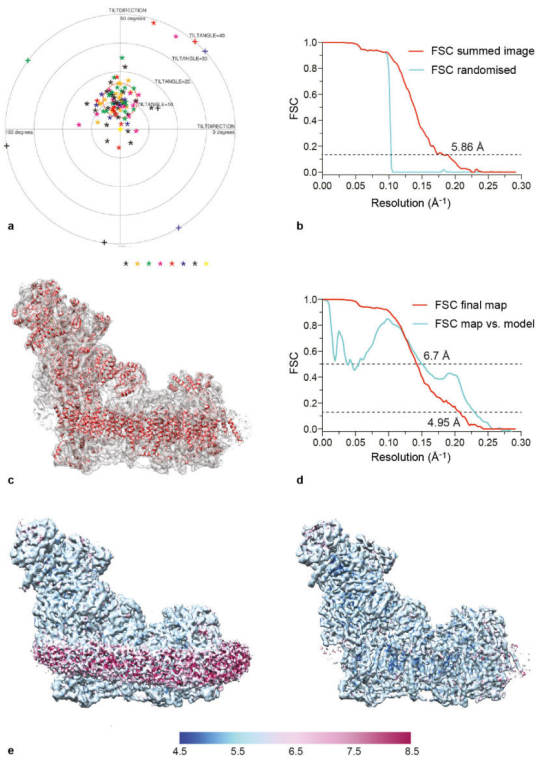
Extended Data



Extended Data Fig. 1. Single particle electron cryo-microscopy analysis of *B. taurus* complex I

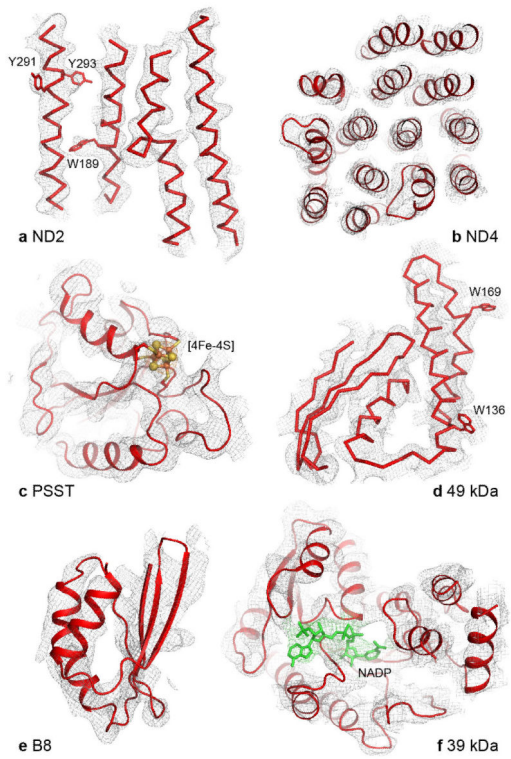
a) Typical micrograph of complex I particles imaged after freezing in vitreous ice on a holey-carbon grid. Some of the selected particles are marked with red boxes. The scale bar represents 50 nm. b) 2D reference classification showing particles lying in different orientations in the ice. The size of each box is 280 pixels and the 2D classification was made in RELION[14].

Europe PMC Funders Author Manuscripts



Extended Data Fig. 2. Validation of the map and resolution

a) Tilt-pair analysis[45] of complex I in cymal-7. 100 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[42]. The outer radius of the plot is 40° and the orange circle centered at the expected tilt angle has a radius of 6°. b) Phase randomisation 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 randomised, 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[14] 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[49]. The left-hand panel (with lower density threshold) shows that the detergent/phospholipid belt is of lower resolution, and the protein regions of the map show resolution distributed from 5 to 6 Å. In the right-hand panel the map is shown at higher density threshold, so the detergent/phospholipid belt is not visualised. Some of the interior parts of the map have resolution of 4.8-5 Å.

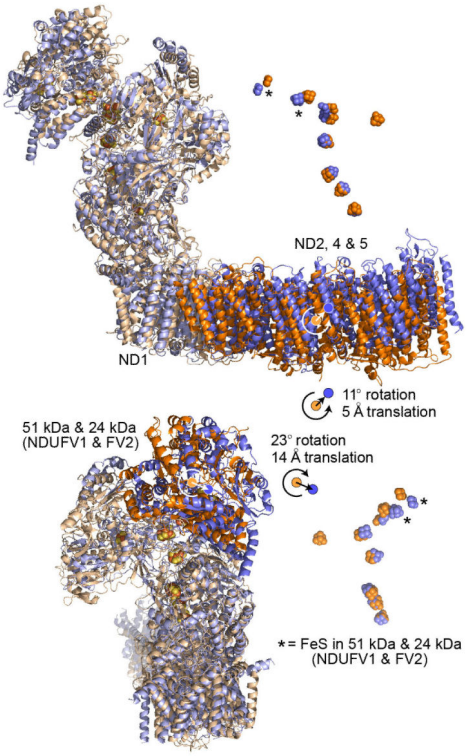


Extended Data Fig. 3. Example regions of the density map with the model fitted to the map

a) ND2 is shown from the membrane plane, high-lighting the densities for three aromatic sidechains and one of the helix-breaking loops. b) Subunit ND4 is 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) In the 39 kDa subunit density consistent with a bound nucleotide is observed, in a similar position to in homologous structures, and as expected from analysis of *Y. lipolytica* complex I[27]. However, the present resolution of the map precludes the inclusion of this nucleotide in the final model.

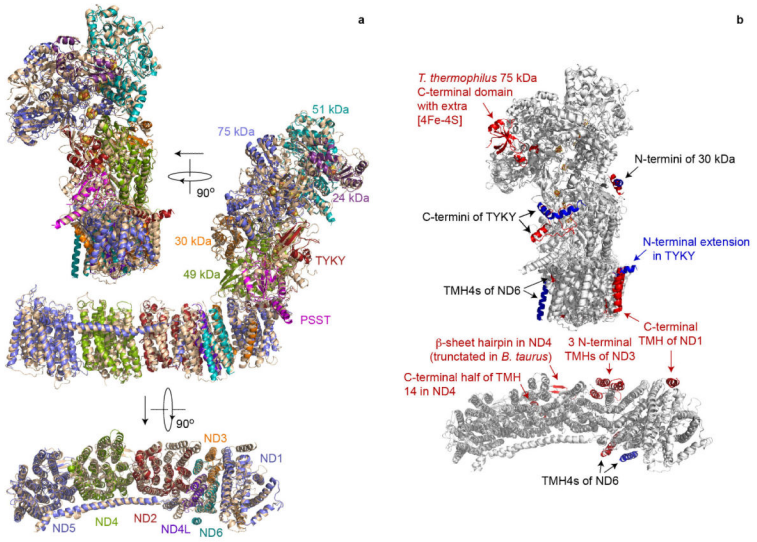
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts



Extended Data Fig. 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* (4HEA.pdb[4]) 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*[58], 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*[58]. 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[59] and the 75 kDa and 30 kDa subunits were not included due to their lower structural conservation.



Extended Data Fig. 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, 4HEA.pdb[4]) 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* (4HEA.pdb[4]); 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*.

TABLE 1
Reference table for the nomenclature of the core subunits of complex I

In the text the names of the subunits from *B. taurus* are used, with the names from the human enzyme presented alongside as appropriate.

Domain	Chain identifier	*Bos taurus*	*Homo sapiens*	*Yarrowia lipolytica*	*Thermus thermophilics*	*Escherichia coli*
Hydrophilic domain	G	75 kDa	NDUFS1	NUAM	Nqo3	NuoG
	F	51 kDa	NDUFV1	NUBM	Nqo1	NuoF
	D	49 kDa	NDUFS2	NUCM	Nqo4	NuoCD
	C	30 kDa	NDUFS3	NUGM	Nqo5	
	E	24 kDa	NDUFV2	NUHM	Nqo2	NuoE
	B	PSST	NDUFS7	NUKM	Nqo6	NuoB
	I	TYKY	NDUFS8	NUIM	Nqo9	Nuol
Membrane domain	H	ND1	ND1	NU1M	Nqo8	NuoH
	N	ND2	ND2	NU2M	Nqo14	NuoN
	A	ND3	ND3	NU3M	Nqo7	NuoA

Domain	Chain identifier	*Bos taurus*	*Homo sapiens*	*Yarrowia lipolytica*	*Thermus thermophilics*	*Escherichia coli*
	M	ND4	ND4	NU4M	Nqo13	NuoM
	L	ND5	ND5	NU5M	Nqo12	NuoL
	J	ND6	ND6	NU6M	Nqo10	NuoJ
	K	ND4L	ND4L	NULM	Nqo11	NuoK

TABLE 2

Summary of the models of the core subunits of *B. taurus* complex I

Subunit	Total residues*	Modelled residues	Poorly resolved / uncertain residue numbering	Unresolved residues	Unresolved elements (>10 residues)	%Modelled	%Identical†	RMSD†
ND1	318	3 - 200 219 - 242 253 - 315		1 - 2 201 - 218 243 - 252 316 - 318	Matrix loop (TMH 5 - 6) IMS loop (TMH 6 - 7)	90% (285/318)	42% (132/318)	1.60 Å
ND2	347	2 - 300 320 - 346	TMH11	1 301 - 319 347	Matrix loop (TMH 10 - 11)	94% (326/347)	25% (86/347)	2.08 Å
ND3	115	2 - 23 52 - 112		1 24 - 51 113 - 115	Matrix loop (TMH 1 - 2)	72% (83/115)	27% (31/115)	2.05 Å
ND4	459	3 - 415 430 - 455	TMH14	1 - 2 416 - 429 456 - 459	Matrix loop (TMH 13 - 14)	96% (439/459)	24% (111/459)	2.20 Å
ND4L	98	1 - 84		85 - 98	Matrix loop (C-terminus)	86% (84/98)	21% (21/98)	2.66 Å
ND5	606	4 - 22 28 - 358 363 - 400 408 - 466 487 - 513 520 - 604	TMH1 TMH13 & TMH14 TMH15 Transverse helix & TMH16	1 - 3 23 - 27 359 - 362 401 - 407 467 - 486 514 - 519 605 - 606	Matrix loop (TMH 1 - 2) Matrix loop (TMH 11 - 12) IMS loop (TMH 12 - 13) IMS loop (TMH 14 - 15) TMH 15 to transverse helix	92% (558/606)	31% (187/606)	2.53 Å
ND6	175	2 - 76 85 - 107 140 - 172	TMH5	1 77 - 84 108 - 139 173 - 175	Matrix loop (TMH 3 - 4) IMS loop (TMH 4 - 5)	75% (131/175)	16% (28/175)	1.83 Å
75 kDa NDUFS1	704	8 - 125 136 - 318 326 - 347 367 - 400 404 - 410 425 - 495 525 - 530 542 - 627	The large domain (222 - 704) is generally poorly resolved. The sequence alignment is weak and the secondary structure content low. Residues 404 - 629 are particularly poorly resolved.	1 - 7 126 - 135 319 - 325 348 - 366 400 - 403 411 -424 496 - 524 531 -541 628 - 704	Probable loop region Probable loop region Probable loop region Probable loop region Probable subdomain	75% (527/704) 1 - 221: 92% 222 - 704: 67%	27% (189/704) 1 - 221 40% 222 - 704: 21%	1.96 Å 1 - 221: 1.57 Å 222 - 704: 2.11 Å
51 kDa NDUFV1	444	31 - 441	Flavin and NADH binding site (63 - 72, 99 - 104, 181 - 189, 300 - 304, 327 - 333)	1 - 30 442 - 444	N-terminal peptide	93% (411/444)	43% (191/444)	1.61 Å
49 kDa NDUFS2	430	47 - 430	3-strand β-sheet (47 - 79)	1 - 46	N-terminal region	89% (384/430)	42% (179/430)	1.41 Å

Subunit	Total residues*	Modelled residues	Poorly resolved / uncertain residue numbering	Unresolved residues	Unresolved elements (>10 residues)	%Modelled	%Identical†	RMSD†
30 kDa NDUFS3	228	15 - 168	Numbering uncertain to 72 Loop / β-strand (73 - 83)	1 - 14 169 - 228	N-terminal peptide C-terminal region	68% (154/228)	24% (54/228)	1.66 Å
24 kDa NDUFV2	217	20 - 178	Loop 126 - 132	1 - 19 179 - 217	N-terminal peptide C-terminal region	73% (159/217)	27% (59/217)	1.57 Å
PSST NDUFS7	179	27 - 169	Loop 68 - 79	1 - 26 170 - 179	N-terminal peptide	80% (143/179)	49% (88/179)	1.44 Å
TYKY NDUFS8	176	15 - 176		1 - 14	N-terminal peptide	92% (162/176)	36% (63/176)	1.89 Å

*For proteins with a mitochondrial-targeting pre-sequence, residue 1 is the first residue of the mature protein[2,3].

†The %identity and the RMSD (root mean square deviation calculated using PDBeFOLD[59]) are between the sequences and structures of the subunits of *B. taurus and T. thermophilus* (4HEA.pdb) complexes I.

TABLE 3

Distances between the redox cofactors in structural models of complex I

	T. thermophilus				*B. taurus*	
	hydrophilic domain (2FUG.pdb[7])		complex I (4HEA.pdb[4])		complex I (this work)	
Cofactors*	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 Ångstroms, 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 4HEA.pdb.

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	NDUFV3	Iα and Iλ		0

B. taurus* subunit	***H. sapiens* subunit***	**Subcomplex†**	**Sequence information**	**Predicted TMHs‡**
18 kDa	NDUFS4	Iα and Iλ		0
15 kDa	NDUFS5	Iα only	CX_9C motif, intermembrane space[37]	0
13 kDa	NDUFS6	Iα and Iλ	PFAM zinc-finger motif $CX_8HX_{15}CX_2C$	0
MWFE	NDUFA1	Iα only		1
B8	NDUFA2	Iα and Iλ		0
B9	NDUFA3	Iα only		1
B13	NDUFA5	Iα and Iλ		0
B14	NDUFA6	Iα only	LYR motif[32]	0
B14.5a	NDUFA7	Iα and Iλ		0
PGIV	NDUFA8	Iα only	Two CX_9C motifs, PFAM CHCH domain intermembrane space[37]	0
39 kDa	NDUFA9	Iα only	Short-chain dehydrogenase reductase family, NADP binding[26,27]	0
42 kDa	NDUFA10	Iα only (low level)	Similarity to deoxynucleoside kinases[24]	0
B14.7	NDUFA11	Iα (Iλ at low level)		3 or 4
B17.2	NDUFA12	Iα and Iλ		0
B16.6	NDUFA13	Iα and Iλ		1
SDAP	NDUFAB1	both Iα Iβ	Acyl-carrier protein[29,30]	0
MNLL	NDUFB1	Iβ		0 (or 1)
AGGG	NDUFB2	Iβ		1 (or 0)
B12	NDUFB3	Iβ		1
B15	NDUFB4	both Iα and Iβ		1
SGDH	NDUFB5	Iβ		1
B17	NDUFB6	Iβ		1
B18	NDUFB7	Iβ	CX_9C motif, intermembrane space[37]	0
ASHI	NDUFB8	Iβ		1
B22	NDUFB9	Iβ	LYR motif[32]	0
PDSW	NDUFB10	Iβ		0
ESSS	NDUFB11	Iβ		1
KFYI	NDUFC1	none		1
B14.5b	NDUFC2	Iβ (low level)		1 or 2

*The former subunit MLRQ (NDUFA4) is no longer considered a subunit of complex I[60].

†Subcomplex Iλ, which contains the seven hydrophilic core subunits and eight to nine supernumerary subunits, is considered to represent a significant portion of the hydrophilic domain of complex I. Subcomplex Iα, which contains all the subunits of subcomplex Iλ plus core subunit ND6 and nine to ten additional supernumerary subunits, represents the hydrophilic domain of complex I plus associated membrane subunits. Subcomplex Iβ, which contains ND4 and ND5 and 12 to 13 supernumerary subunits, represents part of the membrane domain[3].

‡TMHs were predicted using TMHMM2[52], HMMTOP2[53] and the TOPCONS suite[54] (seven methods in total) and are presented as consensus values with less represented values in brackets and single outliers discarded.

TABLE 5

Summary of the models of the supernumerary subunits of *B. taurus* complex I

Subunit	Chain identifier	Total residues*	PDB model†	Aligned residues	%identical	Modelled residues	%modelled	RMSD‡
42 kDa NDUFA10	O	320	2OCP[61]	21 - 252	21% (49/232)	22 - 54 79 - 167 172 - 210 222 - 241	57% (181/320)	1.91 Å
39 kDa NDUFA9	P	345	2Q1W[62]	19 -325	13% (41/307)	19 - 185 203 - 250 285 - 321	73% (252/345)	2.52 Å
18 kDa§ NDUFS4	Q	133	2JYA	33 - 133	37% (37/101)	33 - 59 76 - 116	52% (69/133)	2.42 Å
13 kDa§ NDUFS6	R	96	2JRR	44 - 96	34% (18/53)	47 - 93	49% (47/96)	1.97 Å
B8 NDUFA2	S	99	1S3A[20]	1 - 99	94% (93/99)	17 - 96	81% (80/99)	2.18 Å
SDAP-α NDUFAB1	T	88	1F80[63]	8 - 84	36% (28/77)	9 - 23 28 - 82	81% (71/88)	1.18 Å
SDAP-β NDUFAB1	U	88	1F80[63]	8 - 84	36% (28/77)	8 - 82	85% (75/88)	1.36 Å
B13§ NDUFA5	V	116				1 - 71¶	61% (71/116)	
B14 NDUFA6	W	128				1 - 72¶	56% (72/128)	
PGIV NDUFA8	X	172	2LQL[36]	35 - 114	23% (18/80)	1 - 80¶	46% (79/172)	2.40 Å
B14.7 NDUFA11	Y	141				1 - 106¶	75% (106/141)	
B16.6 NDUFA13	Z	144				33 - 97	45% (65/144)	
B9§ NDUFA3 or MWFE§ NDUFA1	a b	154				1 - 29¶ 1 - 42¶	46% (71/154)	

*For proteins with a mitochondrial-targeting pre-sequence, residue 1 is the first residue of the mature protein 3.

†Known structures with high homology to the complex I subunits were identified by HHpred[55].

‡RMSD: root mean square deviation calculated using PDBeFOLD[59] between the structures of the subunits of *B. taurus* complex I and the PDB model structures.

§Subunit with less certain assignment.

¶Residue numbers are arbitrary and not assigned to the sequence.

Acknowledgements

We thank Richard Henderson, Sjors Scheres, Greg McMullan, Garib Murshudov, Paul Emsley and John Walker for helpful advice, the FEI fellows for educating us on use of the Titan Krios, Jake Grimmett and Toby Darling for computational help and Shaoxia Chen and Christos Savva for electron microscopy help. This work was supported by The Medical Research Council, grant numbers U105184322 (KRV, in Richard Henderson's group) and U105663141 (JH).

REFERENCES

1. Hirst J. Mitochondrial complex I. Annu. Rev. Biochem. 2013; 82:551–575. [PubMed: 23527692]
2. Carroll J, Fearnley IM, Shannon RJ, Hirst J, Walker JE. Analysis of the subunit composition of complex I from bovine heart mitochondria. Mol. Cell. Proteomics. 2003; 2:117–126. [PubMed: 12644575]

3. Hirst J, Carroll J, Fearnley IM, Shannon RJ, Walker JE. The nuclear encoded subunits of complex I from bovine heart mitochondria. Biochim. Biophys. Acta. 2003; 1604:135–150. [PubMed: 12837546]
4. Baradaran R, Berrisford JM, Minhas GS, Sazanov LA. Crystal structure of the entire respiratory complex I. Nature. 2013; 494:443–448. [PubMed: 23417064]
5. Efremov RG, Sazanov LA. Structure of the membrane domain of respiratory complex I. Nature. 2011; 476:414–420. [PubMed: 21822288]
6. Efremov RG, Baradaran R, Sazanov LA. The architecture of respiratory complex I. Nature. 2010; 465:441–445. [PubMed: 20505720]
7. Sazanov LA, Hinchliffe P. Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*. Science. 2006; 311:1430–1436. [PubMed: 16469879]
8. Hunte C, Zickermann V, Brandt U. Functional modules and structural basis of conformational coupling in mitochondrial complex I. Science. 2010; 329:448–451. [PubMed: 20595580]
9. Leonard K, Haiker 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. 1987; 194:277–286. [PubMed: 2956429]
10. Grigorieff N. Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 Å in ice. J. Mol. Biol. 1998; 277:1033–1046. [PubMed: 9571020]
11. Clason T, et al. The structure of eukaryotic and prokaryotic complex I. J. Struct. Biol. 2010; 169:81–88. [PubMed: 19732833]
12. Fassone E, Rahman S. Complex I deficiency: clinical features, biochemistry and molecular genetics. J. Med. Genet. 2012; 49:578–590. [PubMed: 22972949]
13. Sharpley MS, Shannon RJ, Draghi F, Hirst J. Interactions between phospholipids and NADH:ubiquinone oxidoreductase (complex I) from bovine mitochondria. Biochemistry. 2006; 45:241–248. [PubMed: 16388600]
14. Scheres SHW. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 2012; 180:519–530. [PubMed: 23000701]
15. Bai X-C, Fernandez IS, McMullan G, Scheres SH, Ribosome W. structures to near-atomic resolution from thirty thousand cryo-EM particles. eLife. 2013; 2:e00461. [PubMed: 23427024]
16. Birrell JA, Hirst J. Truncation of subunit ND2 disrupts the threefold symmetry of the antiporter-like subunits in complex I from higher metazoans. FEBS Lett. 2010; 584:4247–4252. [PubMed: 20846527]
17. Kussmaul L, Hirst J. The mechanism of superoxide production by NADH:ubiquinone oxidoreductase (complex I) from bovine heart mitochondria. Proc. Natl. Acad. Sci. USA. 2006; 103:7607–7612. [PubMed: 16682634]
18. Mimaki M, Wang X, McKenzie M, Thorburn DR, Ryan MT. Understanding mitochondrial complex I assembly in health and disease. Biochim. Biophys. Acta. 2012; 1817:851–862. [PubMed: 21924235]
19. Dieteren CEJ, et al. Subunit-specific incorporation efficiency and kinetics in mitochondrial complex I homeostasis. J. Biol. Chem. 2012; 287:41851–41860. [PubMed: 23038253]
20. Brockmann C, et al. The oxidised subunit B8 from human complex I adopts a thioredoxin fold. Structure. 2004; 12:1645–1654. [PubMed: 15341729]
21. Keeney PM, Xie J, Capaldi RA, Bennett JP. Parkinson's disease brain mitochondrial complex I has oxidatively damaged subunits and is functionally impaired and misassembled. J. Neurosci. 2006; 26:5256–5264. [PubMed: 16687518]
22. Leshinsky-Silver E, et al. NDUFS4 mutations cause Leigh syndrome with predominant brainstem involvement. Mol. Genet. Metab. 2009; 97:185–189. [PubMed: 19364667]
23. Kirby DM, et al. NDUFS6 mutations are a novel cause of lethal neonatal mitochondrial complex I deficiency. J. Clin. Invest. 2004; 114:837–845. [PubMed: 15372108]
24. Sharpley, MS. Ph.D. Thesis. Cambridge University; 2005.
25. Morais VA, et al. PINK1 loss of function mutations affect mitochondrial complex I activity via NdufA10 ubiquinone uncoupling. Science. 2014; 344:203–207. [PubMed: 24652937]

26. Fearnley IM, Walker JE. Conservation of sequences of subunits of mitochondiral complex I and their relationships with other proteins. Biochim. Biophys. Acta. 1992; 1140:105–134. [PubMed: 1445936]
27. Abdrakhmanova A, Zwicker K, 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. 2006; 1757:1676–1682. [PubMed: 17046710]
28. Babot M, et al. ND3, ND1 and 39 kDa subunits are more exposed in the de-active form of bovine mitochondrial complex I. Biochim. Biophys. Acta. 2014; 1837:929–939. [PubMed: 24560811]
29. Runswick MJ, Fearnley IM, Skehel JM, Walker JE. Presence of an acyl carrier protein in NADH:ubiquinone oxidoreductase from bovine heart mitochondria. FEBS Lett. 1991; 286:121–124. [PubMed: 1907568]
30. Cronan JE, Fearnley IM, Walker JE. Mammalian mitochondria contain a soluble acyl carrier protein. FEBS Lett. 2005; 579:4892–4896. [PubMed: 16109413]
31. Dobrynin K, et al. Characterization of two different acyl carrier proteins in complex I from *Yarrowia lipolytica*. Biochim. Biophys. Acta. 2010; 1797:152–159. [PubMed: 19766092]
32. Angerer H, et al. The LYR protein subunit NB4M/NDUFA6 of mitochondrial complex I anchors an acyl carrier protein and is essential for catalytic activity. Proc. Nat. Acad. Sci. USA. 2014; 111:5207–5212. [PubMed: 24706851]
33. Andrews B, Carroll J, Ding S, Fearnley IM, Walker JE. Assembly factors for the membrane arm of human complex I. Proc Natl Acad Sci USA. 2013; 110:18934–18939. [PubMed: 24191001]
34. 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. 2011; 437:279–288. [PubMed: 21545356]
35. Fearnley IM, et al. GRIM-19, a cell death regulatory gene product, is a subunit of bovine mitochondrial NADH:ubiquinone oxidoreductase (complex I). J. Biol. Chem. 2001; 276:38345–38348. [PubMed: 11522775]
36. Banci L, et al. Structural characterization of CHCHD5 and CHCHD7: two atypical human twin CX_9C proteins. J. Struct. Biol. 2012; 180:190–200. [PubMed: 22842048]
37. Szklarczyk R, et al. NDUFB7 and NDUFA8 are located at the intermembrane surface of complex I. FEBS Lett. 2011; 585:737–743. [PubMed: 21310150]
38. Liao M, Cao E, Julius D, Cheng Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. Nature. 2013; 504:107–112. [PubMed: 24305160]
39. Amunts A, et al. Structure of the yeast mitochondrial large ribosomal subunit. Science. 2014; 343:1485–1489. [PubMed: 24675956]
40. Allegretti M, Mills DJ, McMullan G, Kühlbrandt W, Vonck J. Atomic model of the F_{420}-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector. eLife. 2014; 3:e01963. [PubMed: 24569482]

ADDITIONAL METHODS REFERENCES

41. Bellare JR, Davis HT, Scriven LE, Talmon Y. Controlled environment vitrification system: an improved sample preparation technique. J. Electron Microsc. Tech. 1988; 10:87–111. [PubMed: 3193246]
42. Grigorieff N. FREALIGN: high-resolution refinement of single particle structures. J. Struct. Biol. 2007; 157:117–125. [PubMed: 16828314]
43. Smith JM. XIMDISP – a visualization tool to aid structure determination from electron microscope images. J. Struct. Biol. 1999; 125:223–228. [PubMed: 10222278]
44. Tang G, et al. EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. 2007; 157:38–46. [PubMed: 16859925]
45. Henderson R, et al. Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy. J. Mol. Biol. 2011; 413:1028–1046. [PubMed: 21939668]
46. Mindell JA, Grigorieff N. Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. 2003; 142:334–347.

47. Chen S, et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. Ultramicroscopy. 2013; 135:24–35. [PubMed: 23872039]

48. Rosenthal PB, Henderson R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J. Mol. Biol. 2003; 333:721–745. [PubMed: 14568533]

49. Kucukelbir A, Sigworth FJ, Tagare HD. Quantifying the local resolution of cryo-EM density maps. Nat. Methods. 2014; 11:63–65. [PubMed: 24213166]

50. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of *Coot*. Acta Cryst. 2010; D66:486–501.

51. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 1999; 292:195–202. [PubMed: 10493868]

52. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 2001; 305:567–580. [PubMed: 11152613]

53. Tusnády GE, Simon I. Principles governing amino acid composition of integral membrane proteins: applications to topology prediction. J. Mol. Biol. 1998; 283:489–506. [PubMed: 9769220]

54. Bernsel A, Viklund H, Hennerdal A, Elofsson A. TOPCONS: consensus prediction of membrane protein topology. Nucleic Acids Res. 2009; 37:W465–W468. [PubMed: 19429891]

55. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 2005; 33:W244–W248.

56. Eswar N, et al. Comparative protein structure modeling using modeller. Curr. Protoc. Bioinform. 2006; 15:5.6.1–5.6.30.

57. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 2006; 22:195–201. [PubMed: 16301204]

ADDITIONAL EXTENDED DATA REFERENCES

58. Efremov RG, Sazanov LA. Respiratory complex I: ‘steam engine’ of the cell? Curr. Opin. Struct. Biol. 2011; 21:532–540. [PubMed: 21831629]

59. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Cryst. 2004; D60:2256–2268.

60. Balsa E, et al. NDUFA4 is a subunit of complex IV of the mammalian electron transport chain. Cell Metab. 2012; 16:378–386. [PubMed: 22902835]

61. Johansson K, et al. Structural basis for substrate specificities of cellular deoxyribonucleoside kinases. Nat. Struct. Biol. 2001; 8:616–620. [PubMed: 11427893]

62. King JD, et al. Predicting protein function from structure - the roles of short-chain dehydrogenase/reductase enzymes in *Bordetella* O-antigen biosynthesis. J. Mol. Biol. 2007; 374:749–763. [PubMed: 17950751]

63. Parris KD, 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. 2000; 8:883–895. [PubMed: 10997907]

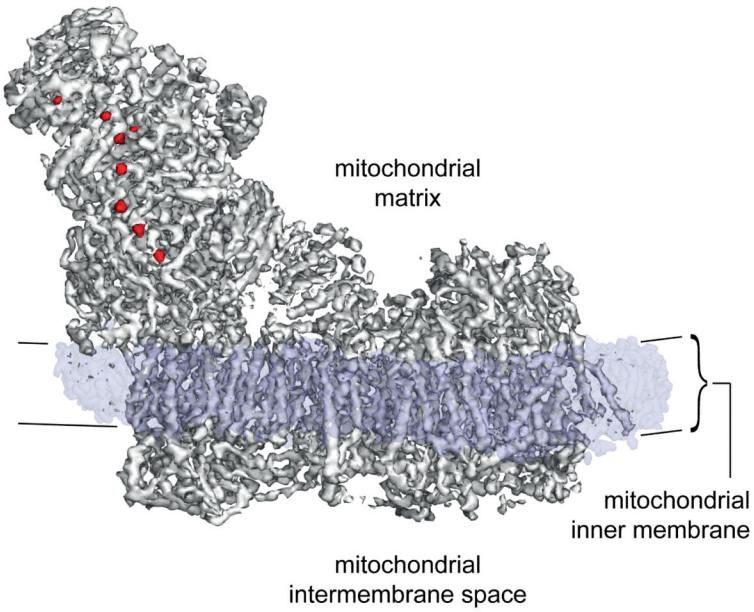


Fig. 1. Overall map for complex I from *B. taurus* heart mitochondria determined by single particle electron cryo-microscopy

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).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

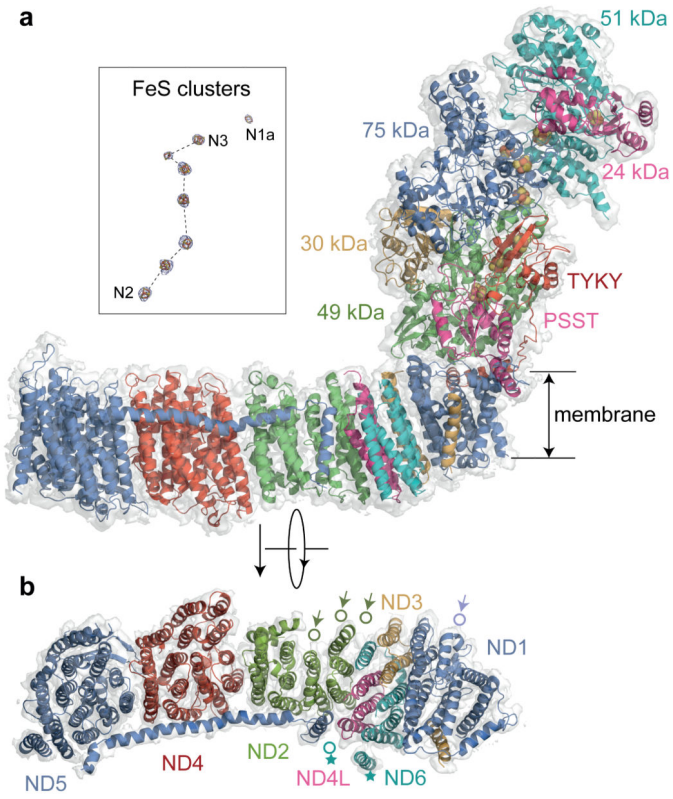


Fig. 2. Structures of the core subunits of mammalian complex I

a) Structural models of the fourteen mammalian core subunits (cartoon representation) and their density (transparent surface); the subunits are coloured individually and labelled with text in the same colours. The chain of FeS clusters is shown modelled to the highest density peaks (blue mesh) in the inset. b) The seven membrane-bound mammalian core subunits, viewed from the matrix. Arrows indicate the positions of the four TMHs in *T. thermophilus* that are not present in *B. taurus*: three N-terminal TMHs in ND2 and one C-terminal TMH in ND1. The position of TMH-4 in ND6 is different in *B. taurus* and *T. thermophilus* (marked with stars). For a detailed comparison of the *B. taurus* and *T. thermophilus* structures see Extended Data Figs. 4 and 5.

Europe PMC Funders Author Manuscripts

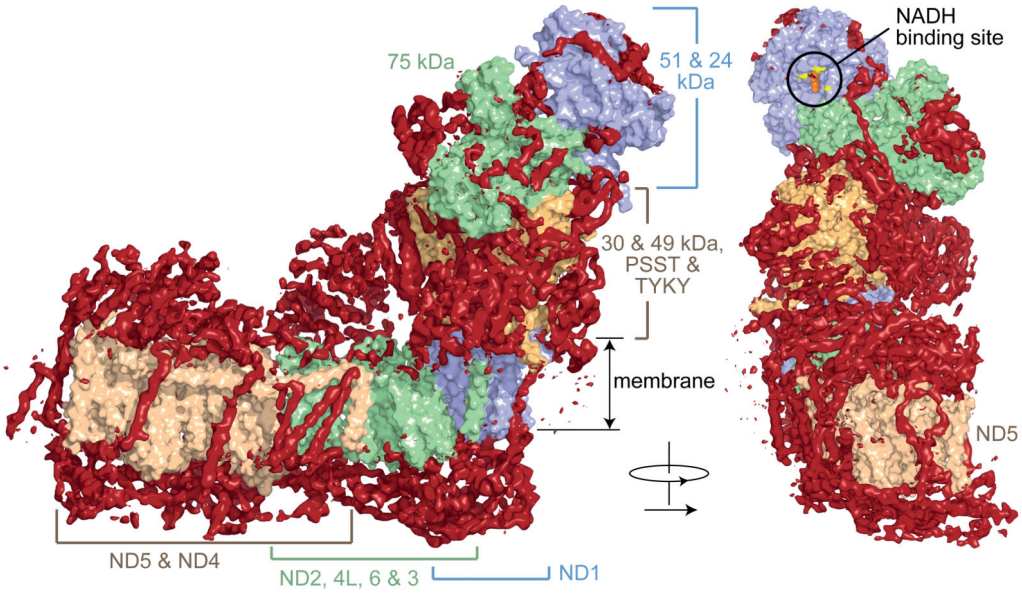


Fig. 3. Architecture of mammalian complex I showing the densities of the supernumerary subunits enclosing the core domain

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 isoalloxazine (orange spheres) and three conserved phenylalanines at the entry to the site (yellow); the vicinity of this site is devoid of supernumerary subunit density.

Europe PMC Funders Author Manuscripts

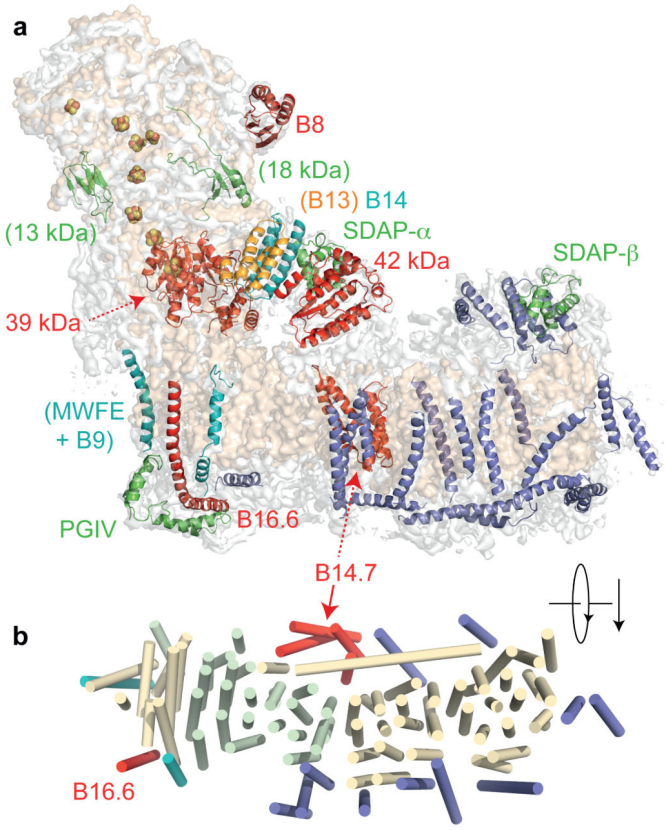


Fig. 4. Structural assignments of supernumerary subunits in mammalian complex I

a) A semi-transparent surface for the density map for mammalian complex I is shown in pale grey, with the surface from the core subunits in wheat. Structural models for the supernumerary subunits are shown in colour and labelled accordingly (dashed lines indicate subunits on the back of the structure). Subunits labelled with brackets are those with less certain assignments, and structural elements, which cannot be assigned confidently in the current map, are in blue. b). Arrangement of TMHs, viewed from the matrix. The core subunits are in light colours (wheat for ND1, ND4 and ND5, green for ND2, ND3, ND4L, and ND6). The supernumerary subunits are coloured as in a).

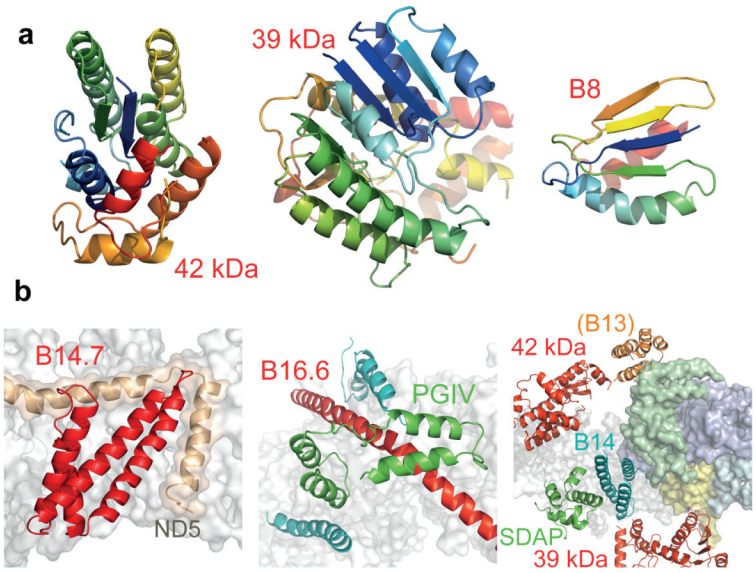


Fig. 5. Structural models for supernumerary subunits in mammalian complex I

a) Models for three supernumerary subunits in cartoon representation, coloured from blue to red (N- to C-termini). b). Structural models and relationships of supernumerary subunits to the core structure. B14.7 is located at the end of the transverse helix, next to ND5-TMH16. The density assigned to PGIV forms an L-shaped ‘clip’ over B16.6, which bends around the heel at ND1. Finally, the supernumerary subunits around the lower section of the hydrophilic domain are viewed in cartoon representation from the matrix. The core membrane subunits (white) and four core hydrophilic subunits, the 49 kDa (blue), 30 kDa (green), PSST (yellow) and TYKY (cyan) subunits, are shown in surface representation.

Europe PMC Funders Author Manuscripts