Structure of mammalian respiratory complex I

Jiapeng Zhu1,*, Kuttii R. Vinothkumar2,3* & Judy Hirst1

Complex I (NADH:ubiquinone oxidoreductase), one of the largest membrane-bound enzymes in the cell, powers ATP synthesis in mammalian mitochondria by using the reducing potential of NADH to drive protons across the inner mitochondrial membrane. Mammalian complex I (ref. 1) contains 45 subunits, comprising 14 core subunits that house the catalytic machinery (and are conserved from bacteria to humans) and a mammalian-specific cohort of 31 supernumerary subunits1,2. Knowledge of the structures and functions of the supernumerary subunits is fragmentary. Here we describe a 4.2-Å resolution single-particle electron cryomicroscopy structure of complex I from Bos taurus. We have located and modelled all 45 subunits, including the 31 supernumerary subunits, to provide the entire structure of the mammalian complex. Computational sorting of the particles identified different structural classes, related by subtle domain movements, which reveal conformationally dynamic regions and match biochemical descriptions of the ‘active-to-de-active’ enzyme transition that occurs during hypoxia3,4. Our structures therefore provide a foundation for understanding complex I assembly5 and the effects of mutations that cause clinically relevant complex I dysfunctions6, give insights into the structural and functional roles of the supernumerary subunits and reveal new information on the mechanism and regulation of catalysis.

Using structures determined for bacterial complex I (refs 7–9) as a starting point, structures of the 14 highly conserved core subunits and their nine cofactors (a flavin mononucleotide (FMN) and eight iron–sulfur (FeS) clusters) have been determined to medium resolution in complex I from both mammals (for Bos taurus)10 and yeast (Yarrowia lipolytica)11. The arrangement and structures of the 31 supernumerary subunits (constituting half the mammalian complex) are, however, far less well defined. The 5-Å resolution electron cryomicroscopy (cryoEM) structure of B. taurus complex I revealed the supernumerary ensemble wrapped around the core, with 14 supernumerary subunits assigned10. Subsequently, eight further assignments were proposed using the crystallographic structure of subcomplex I3 (part of the membrane domain)12. Therefore, nine subunits remain unlocated and models for the supernumerary subunits are fragmentary. The complete structure of mammalian complex I is crucial for elucidating the roles of the supernumerary subunits in complex I function and dysfunction.

Here, we describe a cryoEM map for B. taurus complex I with an overall resolution of 4.16 Å (Fig. 1a and Extended Data Fig. 1), which enabled modelling of all its 45 subunits and 93% of its 8,515 residues (Extended Data Tables 1, 2). Computational sorting of the particles revealed three major classes, with overall resolutions 4.27 Å (class 1), 4.35 Å (class 2) and 5.60 Å (class 3) (Extended Data Fig. 2), for which the quality of the map in several regions was improved substantially. The different classes represent different states of the complex and analysis of each provides new insights into the mechanism of complex I catalysis. Extended Data Figures 3 and 4 present example densities and we use the class 2 map and model to describe the structure, unless indicated otherwise.

Figure 1 presents the structures and locations of all 31 supernumerary subunits in mammalian complex I (see Extended Data Table 3 for subunit–subunit interactions and additional details). The supernumerary subunits are central to the structure, stability and assembly of the complex, and some also have regulatory or independent metabolic roles.

The 18 supernumerary transmembrane helices (TMHs) (Fig. 1b) establish a cage around the core membrane domain. Three TMH-containing subunits, B9 (NDUFA3 in the nomenclature for human complex I), B16.6 (NDUFA13) and MWFE (NDUFA1), interact extensively with PGIV (NDUFA8) on the intermembrane-space (IMS) face, enclosing core subunit ND1. Subunit B14.5b (NDUFC2), bound to ND2, contains two different-length TMHs and attaches KFY1 (NDUFC1) to the complex. Three TMHs that interact with ND4 are assigned to MNLL (NDUFB1), ESSS (NDUFB11) and SGDH (NDUFB8). Four TMHs, assigned to B17 (NDUFB6), AGGG (NDUFB2), B12 (NDUFB3) and ASHI (NDUFB8), are bound to ND5. The TMHs of ASHI and B15 (NDUFB4, on the side of ND4) cross the ND5 transverse helix, and the four TMHs of B14.7 (NDUFA11) appear to support ND5-TMH16 in anchoring it against ND2. Four subunits confined to the IMS (PGIV, the 15 kDa subunit (NDUFS5), PDSW (NDUFB10) and B18 (NDUFB7)) form a helix latticework (together with SGDH and B16.6) on the IMS face (Fig. 2a). PGIV, the 15 kDa subunit and B18 contain CHCH domains (pairs of helices linked by two disulfide bonds)13 and are canonical substrates for the Mia40 oxidative-folding pathway14; PDSW probably contains two further disulfide bonds. These disulfide bonds form during complex I biogenesis and are probably important for enzyme stability. Thus, the supernumerary cage has evolved to become integral to the structure and stability of the membrane domain.

Subunits B14 (NDUFA6) and SDAP–α (NDUFAB1), and B22 (NDUFB9) and SDAP–β (NDUFAB1), constitute matching subdomains on the hydrophilic domain and matrix face of the membrane domain, respectively15–17 (Fig. 1). SDAP-α and SDAP-β are identical to the mitochondrial acyl-carrier protein (ACP) and exhibit densities consistent with the pantheine-4’-phosphate group that covalently attaches an acyl chain to Ser44 (refs 16, 17) (Extended Data Fig. 4e). Their ACP recognition helices interact with arginine- and lysine-rich helices in the Lyr proteins B14 and B22 (Fig. 2b and Extended Data Fig. 4e) as canonical ACPs interact with the enzymes of fatty-acid biosynthesis18. The 42 kDa subunit (NDUFA10, Fig. 2e) contains a central α/β nucleotide kinase fold with a parallel five-strand β-sheet, plus three extensions that dock it to the matrix face of ND2. Although the active site is accessible and the key nucleoside kinase residues are present19, no activity has been reported. The 39 kDa subunit (NDUFA9, Fig. 2d) is attached to core subunits PSST (NDUF57) and the 30 kDa subunit (NDUF33) in the hydrophilic arm. The N-terminal domain of the 39 kDa subunit comprises an α/β short-chain dehydrogenase/reductase fold20 containing an NAD(P)-binding Rossmann fold with a parallel seven-strand β-sheet and density for a bound nucleotide, modelled as NADPH21 (Extended Data Fig. 4). The separate C-terminal
domain interacts with the long matrix loop between TMHs 1 and 2 of ND3.

The final seven supernumerary subunits adorn the hydrophilic domain (Fig. 1c). Thioredoxin-like B8 (NDUFA2) is attached to the 75 kDa subunit (NDUF51), and the three-helix bundle of B13 (NDUF5A5) to the 30 kDa subunit. The remaining five subunits are located at interfaces. The zinc-binding domain of the 13 kDa subunit (NDUFS6) and the four-strand β-sheet and helix of the 18 kDa subunit (NDUFS4) are located where the NADH dehydrogenase domain meets the rest of the complex. All five subunits (the other three are B14.5a (NDUFA7), B17.2 (NDUFA12) and the 10 kDa subunit (NDUFV3)) contain long loops running over the domain surface. A notable example is the extensive loop in B14.5a, which arches up along the TYKY–49 kDa subunit (NDUFS8–NDUFS2) interface, across the 49 kDa subunit, along its interface with the 75 kDa subunit and onto the 30 kDa subunit. The role of the supernumerary subunits in stabilizing interfaces in the hydrophilic domain contrasts sharply with their arrangement into a rigid cage to stabilize the membrane domain.

The structures of the mammalian core subunits (Fig. 3a) closely match those of the bacterial subunits, and contain corresponding mechanistically relevant features. NADH is oxidized by a flavin mononucleotide in the 51 kDa subunit (NDUFV1) (Extended Data Fig. 3c). Electrons then transfer along a chain of FeS clusters to the terminal domain. All five subunits (the other three are B14.5a (NDUFA7), B17.2 (NDUFA12) and the 10 kDa subunit (NDUFV3)) contain long loops running over the domain surface. A notable example is the extensive loop in B14.5a, which arches up along the TYKY–49 kDa subunit (NDUFS8–NDUFS2) interface, across the 49 kDa subunit, along its interface with the 75 kDa subunit and onto the 30 kDa subunit. The role of the supernumerary subunits in stabilizing interfaces in the hydrophilic domain contrasts sharply with their arrangement into a rigid cage to stabilize the membrane domain.

The core subunits are in surface representation and the supernumerary subunits in cartoon; icons show viewpoints and locations in the complex. Subunits in brackets are behind the domain. The assignments of B12 and AGGG may be reversed. See Extended Data Tables 1 and 2 for the subunit nomenclatures in other species.

Four proton-transfer routes (in ND2, ND4, ND5 (ref. 8), and ND1 + ND4L + ND6 (ref. 9)) have been proposed for the four protons that complex I is generally considered to translocate for each NADH molecule oxidized. ND2, ND4 and ND5 each contain two TMHs interrupted by loops in the central membrane plane (TMH4 and TMH9 in ND2; TMH7 and TMH12 in ND4 and ND5, Fig. 3a and Extended Data Fig. 3). The chain of conserved aspartate, glutamate, lysine and histidine residues that runs along the membrane domain (Fig. 3a). This extended conformation explains its susceptibility to proteases, but it is unlikely to be central to the mechanism because it is not conserved in Thermus thermophilus. The chain of conserved aspartate, glutamate, lysine and histidine residues that runs along the membrane domain (Fig. 3a). This extended conformation explains its susceptibility to proteases, but it is unlikely to be central to the mechanism because it is not conserved in Thermus thermophilus, and in Escherichia coli is fused to the C terminus of the 30 kDa subunit. The long matrix loop in ND3, which lies across the front of the hydrophilic domain and is central to the transition between active and de-active states in mammalian complex I (ref. 3), is also resolved (Fig. 3a).
and Lys263 of ND2; Glu123, Lys206, Lys237 and Glu378 of ND4; Glu145, Lys223, His248 and Lys392 of ND5 (Fig. 3a). Distortions of the helical structure are observed in TMH3 of ND6, TMH5 of ND2, TMH8 of ND4 and TMH8 of ND5 (Extended Data Fig. 3). These distortions resemble the π–bulge in bacteriorhodopsin but do not satisfy its technical definition26, perhaps owing to the intermediate resolution of the maps. The distortions are centred on glycine pairs in ND6 (62–3) and ND4 (239–40), on a serine pair in ND5 (249–50), and on Trp167 (flanked by two glycine pairs) in ND2. Notably, TMH3 of ND6 is more distorted in the mammalian structure than in T. thermophilus (which contains only one glycine residue)9, such that Phe67 of ND6 is displaced around the helical axis.

Ubiquinone-10 binds with its redox-active headgroup close to cluster N2, at the top of a cleft between the 49 kDa subunit and PSST (Fig. 3b), while T. thermophilus complex I co-crystallized with decylubiquinone showed it forms hydrogen bonds with His59 and Tyr108 of the 49 kDa subunit. Here, the side chains of Tyr108 and His59 are poorly resolved, and the conformation of the β1–β2 His59-containing loop is different to that in T. thermophilus (Fig. 3d). It therefore appears that the structural elements that form the binding site are flexible, allowing it to organize around substrates and inhibitors (neither of which are present here). The putative ubiquinone-access channel, identified first in T. thermophilus9, connects the cleft to an entrance in ND1 (between TMH1, an amphipathic helix, and TMH6) and can also be detected here (minimum diameter, 2.9 Å). Alternative entrances, between TMH1 and TMH7, and TMH5 and TMH6 of ND1, are also evident but narrower (minimum diameters, 1.9–2.2 Å). However, the planar ubiquinone ring is approximately 6 Å across, so all the channels in the static structure would have to open to allow it to enter. A structure containing ubiquinone-10 (or a long-chain analogue) is therefore required to confirm its access pathway.

In the mammalian complex, further consideration of the most plausible channel (that is, the widest) for ubiquinone reveals a ‘bottleneck’ at the base of the cleft (Fig. 3c). Ubiquinone-10 is highly hydrophobic so most of the channel-lining residues are uncharged and hydrophobic. In contrast, the bottleneck is formed by charged and polar residues including Glu24 and Arg25 (TMH1 of ND1), Arg274 (TMH7 of ND1), and Arg71 and hydroxy-Arg77 (ref. 23) (α2–3 loop of PSST). Nearby, the TMH5–6 loop of ND1, with many acidic residues, contributes more notably to channel formation in T. thermophilus and Y. lipolytica11. This cluster of charged residues suggests the presence of water molecules and appears to be incompatible with a ubiquinone-10 binding channel. However, the PSST loop was modelled incompletely in T. thermophilus and Y. lipolytica, and the ND1 loop is poorly resolved here, indicating their flexibility. It is possible that conformational changes at the bottleneck, linked to ubiquinone binding and dissociation, contribute to coupling of the redox reaction to proton translocation.

When the particles comprising the whole data set were subjected to 3D classification, three major, slightly different classes emerged. Class 3,
The smallest, lowest resolution class, is closer to class 1 than class 2 in structure and is characterized by movement of the ND4–ND5 subdomain (relative to class 1, Extended Data Table 4) and disorder in the ND5 transverse helix and its anchor (TMH16 of ND5). Similar disorder was observed in subcomplex I3 (ref. 12), which comprises the ND4–ND5 subdomain. We therefore suggest that class 3 is a state in which molecules are in the first stages (relative to that in T. thermophilus) were proposed previously as characteristic of the de-active state11, but they vary between our class 2 conformation and that of T. thermophilus (Fig. 3d), and are disordered in class 1. Notably, Y. lipolytica complex I was co-crystallized with a quinazoline inhibitor (Fig. 3d), and cross-linking studies have shown quinazolines interact with sections of the 49 kDa subunit and ND1 that contain the β1–β2 and TMH5–6 loops29. We propose quinazoline binding orders these loops, and the quinazoline-binding site overlaps with (but does not superimpose on) the ubiquinone-binding site. Our interpretation supports biochemical proposals for non-identical but overlapping sites for the myriad inhibitors of ubiquinone reduction29, but does not support an alternative, occluded ubiquinone-binding site in the de-active complex11.

The two states of mammalian complex I described support the idea that dynamic, flexible regions at the hydrophilic–membrane domain interface are important for coupling ubiquinone reduction to proton translocation. The class-1 disordered loops in ND1, ND3 and the 49 kDa subunit all contribute to the ubiquinone-binding site (Fig. 4c). Therefore, we attribute lack of catalytic activity in the de-active state to reversible disruption of this site, which can be recovered when the ubiquinone-binding site in the NADH-reduced enzyme reforms around its substrate. During catalysis, the ND3 loop, which originates in the membrane and interacts extensively with the hydrophilic domain, may restrict conformational changes at the domain interface. Changes in the conformation of the loop of the 49 kDa subunit may trigger proton translocation: molecular simulations were used to outline a mechanism in which the ubiquinol dianion deprotonates Tyr108 and His59, breaking a His59–Asp160 hydrogen bond and displacing Asp160 towards the membrane10. In ND1, TMH2–6 replicate the antipporter-like half-channel motif of ND2, ND4 and ND5 (ref. 9). TMH5 resembles a discontinuous TMH, but with its half helix on the matrix side unstructured and continuous with the TMH5–6 loop at the base of the ubiquinone-binding cleft (Fig. 3d). Like α2–β2 in PSST (Fig. 3c), this loop may change conformation upon ubiquinone binding. Furthermore, the loop carries many conserved acidic residues that may collect protons for Glu143 of ND1 (ref. 9). In turn, Glu143 is connected to the chain of charged residues along the membrane domain by Asp66 of ND2 and the dynamic distortion in TMH3 of ND6 (ref. 9). Thus, a cascade of events originating from the ubiquinone-binding cleft may couple ubiquinone reduction and protonation to proton translocation. Although all such mechanisms for complex I are currently hypothetical, cryoEM now provides a powerful tool to study individual trapped conformations or separate mixed states computationally in order to determine how conformational changes are initiated, coordinated and propagated.

Online Content 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 11 March; accepted 8 July 2016.
Published online 10 August 2016.
1. Hirst, J. Mitochondrial complex I. *Annu. Rev. Biochem.* 82, 551–575 (2013).

2. Hirst, J., Carroll, J., Fearnley, I. M., Shannon, R. J. & Walker, J. E. The nuclear encoded subunits of complex I from bovine heart mitochondria. *Biochim. Biophys. Acta* 1804, 125–150 (2003).

3. Galkin, A. et al. Identification of the mitochondrial ND3 subunit as a structural component involved in the active/deactive enzyme transition of respiratory complex I. *J. Biol. Chem.* 283, 20907–20913 (2008).

4. Galkin, A., Abravam, A. Y., Frakich, N., Duchen, M. R. & Moncada, S. Lack of oxygen deactivates mitochondrial complex I: implications for ischemic injury? *J. Biol. Chem.* 284, 36055–36061 (2009).

5. Sánchez-Caballero, L., Guerrero-Castillo, S. & Nijtmans, L. Unraveling the complexity of mitochondrial complex I assembly: A dynamic process. *Biochim. Biophys. Acta* 1857, 980–990 (2016).

6. Fassone, E. & Rahman, S. Complex I deficiency: clinical features, biochemistry and molecular genetics. *J. Med. Genet.* 49, 578–590 (2012).

7. Sanazan, L. A. & Hinchliffe, P. Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus.* *Science* 311, 1430–1436 (2006).

8. Efremov, R. G. & Sanazan, L. A. Structure of the membrane domain of respiratory complex I. *Nature* 476, 414–420 (2011).

9. Baradaran, R., Bernsford, J. M., Minhas, G. S. & Sanazan, L. A. Crystal structure of the entire respiratory complex I. *Nature* 494, 443–448 (2013).

10. Vinodhukumar, K. R., Zhu, J. & Hirst, J. Architecture of mammalian respiratory complex I. *Nature* 515, 80–84 (2014).

11. Zickermann, V. et al. Structural biology. Mechanistic insight from the crystal structure of mitochondrial complex I. *Science* 347, 44–49 (2015).

12. Zhu, J. et al. Structure of subcomplex II of mammalian respiratory complex I leads to new supernumerary subunit assignments. *Proc. Natl Acad. Sci. USA* 112, 12087–12092 (2015).

13. Szklarczyk, R. et al. NDUF87 and NDUFA8 are located at the intermembrane surface of complex I. *FEBS Lett.* 585, 737–743 (2011).

14. Bani, C. et al. MIA40 is an oxidoreductase that catalyzes oxidative protein folding in mitochondria. *Nat. Struct. Mol. Biol.* 16, 198–206 (2009).

15. Angerer, H. et al. The LYR protein subunit NBM/NDUFA6 of mitochondrial complex I anchors an acyl carrier protein and is essential for catalytic activity. *Proc. Natl Acad. Sci. USA* 111, 5207–5212 (2014).

16. 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).

17. Feng, D., Witkowski, A. & Smith, S. Down-regulation of mitochondrial acyl carrier protein in mammalian cells compromises protein lipoylation and respiratory complex I and results in cell death. *J. Biol. Chem.* 284, 11436–11445 (2009).

18. Chan, D. I. & Vogel, H. J. Current understanding of fatty acid biosynthesis and the acyl carrier protein. *Biochem. J.* 430, 1–19 (2010).

19. Johansson, K. et al. Structural basis for substrate specificities of cellular deoxyribonucleotidases. *Nat. Struct. Mol. Biol.* 8, 616–620 (2001).

20. 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).

21. 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* 1757, 1676–1682 (2006).

22. Kmita, K. et al. Accessory NUMM (NDUF56) subunit harbours a Zn-binding site and is essential for biogenesis of mitochondrial complex I. *Proc. Natl Acad. Sci. USA* 112, 5685–5690 (2015).

23. Carroll, J., Ding, S., Fearnley, I. M. & Walker, J. E. Post-translational modifications near the quinone binding site of mammalian complex I. *J. Biol. Chem.* 288, 24799–24808 (2013).

24. Hirst, J. & Roessler, M. Energy conversion, redox catalysis and generation of reactive oxygen species by respiratory complex I. *Biochim. Biophys. Acta* 1857, 872–883 (2016).

25. Muri, M., Mashimo, Y., Hirst, J. & Miyoshi, H. Exploring interactions between the 49 kDa and ND1 subunits in mitochondrial NADH-ubiquinone oxidoreductase (complex I) by photoaffinity labeling. *Biochemistry* 50, 6901–6908 (2011).

26. Cooley, R. B., Arp, D. J. & Karplus, P. A. Evolutionary origin of a secondary structure: α-helices as cryptic but widespread insertion variations of α-helices that enhance protein functionality. *J. Mol. Biol.* 404, 232–246 (2010).

27. Tociescu, M. A., Fendel, U., Zwicker, K., Kerscher, S. & Brandt, U. Exploring the ubiquinone binding cavity of respiratory complex I. *J. Biol. Chem.* 282, 29514–29520 (2007).

28. Babot, M. et al. ND3, ND1 and 39kDa subunits are more exposed in the de-active form of bovine mitochondrial complex I. *Biochim. Biophys. Acta* 1837, 929–939 (2014).

29. Muri, M. & Miyoshi, H. Current topics on inhibitors of respiratory complex I. *Biochim. Biophys. Acta* 1857, 884–891 (2016).

30. Sharma, V. et al. Redox-induced activation of the proton pump in the respiratory complex I. *Proc. Natl Acad. Sci. USA* 112, 11571–11576 (2015).

Acknowledgements We thank R. Henderson and G. Murshudov for discussions, J. Grimmert and T. Darling for computational help, S. Chen and C. Savva for electron microscopy help, and J. N. Blaza for quantifying the ratio of de-active and active enzymes in our preparation. 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 J.Z. prepared protein: K.R.V. carried out electron microscopy data collection and analysis with help from J.Z.; J.Z. built the initial model; J.Z., K.R.V. and J.H. worked together, led by J.H., to model and analyse the data; J.H. designed the project; J.H. wrote the paper with help from J.Z. and K.R.V.

Author Information The electron microscopy maps and models for each class have been deposited in the Electron Microscopy Data Bank (EMDB) with accession numbers EMD-4040 (class 1), EMD-4032 (class 2) and EMD-4041 (class 3), and in the Protein Data Bank with accessions 5LDW (class 1), 5LC5 (class 2) and 5LDC (class 3). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.R.V. (vkumar@mrc-lmb.cam.ac.uk) or J.H. (jh@mrc-mbu.cam.ac.uk).

Reviewer Information Nature thanks R. B. Gennis, M. T. Ryan and the other anonymous reviewer(s) for their contribution to the peer review of this work.
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 7-cyclohexyl-1-heptyl-β-D-maltoside (Cymal 7), as described previously10,31.

CryoEM specimen preparation, imaging and image processing. CryoEM grids were prepared as described previously32. Images were recorded using a 300 keV FEI Titan Krios electron microscope with EPU software, with the specimen temperature maintained at ~186°C. A Falcon II CMOS (complementary metal oxide semiconductor) direct electron detector was used for imaging at 105,263× magnification (nominally 59,000×), corresponding to a sampling density of 1.33 Å pixel⁻¹, at 2.4–4.2 μm under-focus in 0.3 μm steps, with the autofocus routine performed every 8 μm to give a range of defocus. Each image was exposed for 2 s (total dose ~35 e⁻ A⁻²) and an in-house protocol was used to capture 34 movie frames. The frames were aligned using Unblur (without dose filtering)33 and the CTF was determined with CTFIND4 (ref. 33). A total of 139,456 particles were picked manually and extracted using a box of 360 pixels. Further processing was performed with RELION34. Following an initial 2D classification to discard 'bad' particles, 115,966 particles were used for refinement. The 5 Å resolution map described previously10 was low-pass filtered to 60 Å and used as the reference map. The whole data set was subjected to the auto-refine routine in RELION, followed by modelling of the beam-induced movement (using a running average of 7 frames) and B-factor weighting. All the resolutions described here are defined at FSC = 0.143 following application of a shape mask, phase randomization to check for effects of the mask, and correction for the modulation transfer function of the detector. The resolution of the data set containing all the particles after B-factor weighting and refinement was 4.16 Å, with an estimated orientation accuracy of 0.93 degrees.

The B-factor weighted particles were subjected to 3D classification into eight classes using an angular sampling of 0.9° for 25 iterations, with the resolution limited to 8 Å. Three major classes were identified, containing 48,033 (class 1), 33,301 (class 2) and 19,306 (class 3) particles. Each class was refined individually, providing maps with overall resolution of 4.27 Å for class 1, 4.35 Å for class 2 and 5.6 Å for class 3. The maps were sharpened with B-factors of -114 for class 1, -110 for class 2 and -125 for class 3. Local analysis of the resolution was performed using ResMap35 (Extended Data Figs 1, 2). Note that the map used as the reference for refinement is a class 1 map, which we described previously as the major class10, and that the number of particles required to achieve the reported resolutions indicates the need for future improvements in both the biochemical homogeneity and specimen preparation of the samples.

Model building and analysis. Model building was performed using Coot16. The first model was built using the map from the complete data set, with cross-referencing to the maps from classes 1 and 2, using the 5 Å model for *B. taurus* complex I described previously (PDB accession code 4UQ8 (ref. 10)) as the initial template. This unrestrained polyalanine model contains models for the fourteen core subunits that are structural homologues of the subunits of the bacterial enzyme2, partial models for fourteen assigned supernumerary subunits, and a further 21 polypeptide chains from unassigned supernumerary structures. Assignments to some of these chains were subsequently proposed for a further eight supernumerary subunits using the 6.8 Å X-ray crystallographic structure of subcomplex I3 from *B. taurus* complex I (ref. 12). The new maps show clear connectivity within the density features, allowing many of the previously traced chains to be extended considerably, and some of them to be joined together. Furthermore, the helical pitches of most of the TMHs and of many of the helices in the globular subunits are now clear, and the β-strands are well separated. These substantial improvements in the density, together with information from secondary structure analyses and homologous structures, allowed improved and more complete models to be built for all 45 subunits. Note that the former subunit MLQB (as seen in our previous model) has now been assigned to a subunit of complex I29 and that there are two copies of subunit SDAP10. The only substantial un-modelled protein densities are underneath the tip of the membrane domain and are accounted for by the termini of two supernumerary subunits (B18 and ASHI).

In well-resolved regions of the map, protruding densities of the side chains of the bulky aromatic residues Phe, Tyr and Trp, along with some side chain densities from Arg and His, are clearly visible (Extended Data Figs 3, 4). For those subunits that had already been assigned, these side chain features were used as landmarks for assigning the sequences. Side chains were added in well-resolved regions, but omitted when their density features are unclear. The assignments of four subunits that were previously assigned in pairs (B9 and MWFE10, and PDSW and B18 (ref. 12)) were also confirmed. For three subunits (B8, SDAP-α and SDAP-β) the models of the human homologues are available in the PDB (PDB accession codes 1S3A (ref. 38) and 2DNW) and were used to assign the residues. For highly conserved regions of the 51 kDa, 24 kDa (NDUFV2) and 75 kDa subunits, residue assignments were supported by the structure of complex I from *T. thermophila*35. In some less well-resolved regions of the map it is not possible to assign the sequence confidently using the current data. In these regions the polyalanine model has been retained (residue names UNK), but the residues for each subunit have been numbered as accurately as possible, to provide a guide to the location of individual residues. The exceptions are subunits B12, 10 kDa and part of B14.5a, for which residue numbers cannot be confidently proposed. In summary, the models described are mixed models in which the residues in some subunits are fully assigned, whereas residues are not assigned at all (Extended Data Tables 1, 2). Next, the hitherto-unknown subunits were assigned. The patterns of the bulky residues observed in the TMH-containing regions of unknown densities were compared with the amino acid sequences of candidate subunits. By combining this information with information from secondary structure analyses, supported with biochemical knowledge, the seven hitherto-unassigned TMHs were assigned and the subunits modelled as described above. Subunits B12 and AGGG were assigned to the two TMHs on the tip of the membrane domain, but lack of clear features in the densities means that our specific assignment of B12 to chain k and AGGG to chain j is less confident, so we cannot exclude the possibility that they have been reversed. The two additional TMH-like densities observed in the structure of subcomplex I3 (ref. 12) are clearly absent from the cryoEM maps, so they are attributed to an artefact produced by crystal contacts with either dissociated subunits or the reorganized transverse helix. Remaining polypeptide chains, located on the outside of the hydrophilic domain and the IMS face of the membrane domain, were assigned by combining secondary structure analyses with biochemical knowledge and by using the densities from bulky residues, and the residues and side chains assigned and modelled where possible. Maps with different B-factor sharpening were used to help with chain tracing and assignment of residues, and the model was checked for consistency with the individual maps from the three different classes. The geometry of the model was improved by cycles of manual adjustment in Coot16, real-space refinement by Phenix19 and refinement by REFMAC40, with secondary structure restraints.

Separate models were subsequently created for classes 1 and 2 by rigid-body fitting of each subunit, manual adjustment to account for substantial local differences identified in the density maps, and cycles of adjustment and refinement as described above. To provide some reassurance of the refinement, the coordinates for classes 1 and 2 were shifted by 0.1 Å and the B-factors re-set to 75 and the coordinates were then refined against one of the half maps. The resulting models were then used to calculate Fourier-shell correlation (FSC) curves for both half maps; as shown in Extended Data Fig. 2 the curves display very little evidence of over-fitting. The class 3 model was created from the class 1 model by rigid-body fitting. It was created purely for comparison with classes 1 and 2 so the individual subunit models were retained unchanged from class 1, the model was not deleted in the regions of poor density reported in Extended Data Fig. 5, and no further refinement was performed. The refinement statistics for classes 1 and 2 are summarized in Extended Data Table 5.

The sequences of all the subunits are numbered starting from residue 1 of the mature proteins2. The naming of the chains has been retained as much as possible from our previous model (PDB accession code 4UQ8). The names are unchanged for chains A–Z and a–n, except that the previous chains d and e have been combined to form new chain d (B14.5b) and chain e has been reallocated to the 15 kDa subunit. Previous chains o–w all represented sections of subunits that have now been combined; new chains n–s have been reallocated.

After modelling the protein, we observed two additional short, elongated densities located at interfaces between core membrane-domain subunits, which may represent phospholipid molecules. They are between TMH10 of ND2 (residues 291–295), TMH5 of ND4 (residues 144–147) and the ND5 transverse helix (residues 564–567), and between TMH11 of ND4 (residues 355–360) and the ND5 residues 356–361. These densities have not been modelled because similar densities are observed elsewhere within the detergent/phospholipid belt but at lower contour levels. As such, we cannot exclude the possibility that they are due to noise in the density map at the current resolution.

Bioinformatics. Secondary structure analyses were carried out using PSIPRED41 and raptorX25. The identification of TMHs in the sequences and the structures of homologous proteins were described previously10. Caviities and channels in the structures were investigated using CAVER35. Figures were created using the PyMOL Molecular Graphics System. The subunit interactions in Extended Data Table 3 were calculated with NCONT in CCP4 (ref. 44), and defined as a distance of less than 5 Å between any two atoms. For some subunits, such as B8 and SDAP-α, the interactions are limited to one subunit, while other subunits with long loops and extended structures, such as SDGH, form multiple interactions. Some of the residues in our current model lack side chains so the number of interactions detected may increase in future models.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
31. 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).

32. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).

33. Rohou, A. & Grigorieff, N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

34. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

35. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).

36. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

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

38. Brockmann, C. et al. The oxidized subunit B8 from human complex I adopts a thioredoxin fold. *Structure* **12**, 1645–1654 (2004).

39. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

40. Brown, A. et al. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71**, 136–153 (2015).

41. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).

42. Källberg, M. et al. Template-based protein structure modeling using the RaptorX web server. *Nat. Protocols* **7**, 1511–1522 (2012).

43. Chovancova, E. et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLOS Comput. Biol.* **8**, e1002708 (2012).

44. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
Extended Data Figure 1 | Resolution estimation and ResMap analysis of the density map for complex I before classification. a, The map, shown at two different contour levels, is coloured according to the local resolution, as determined by ResMap. At the higher contour level (left), the majority of the protein is resolved to 3.9–4.7 Å; only the very peripheral regions (parts of the 51 kDa, 24 kDa and 75 kDa subunits in the matrix arm, and the distal end of the membrane arm) are at lower resolutions of 5–6 Å. At the lower contour level (right), the detergent/lipid belt in the 7–9 Å resolution range dominates the lower-resolution features. b, A slice through the map shows that large portions of the central, core regions are resolved to 4 Å or better. c, The FSC curve defines an estimated overall resolution of 4.16 Å at FSC = 0.143.
Extended Data Figure 2 | Resolution estimation and ResMap analysis of the classes of complex I. a–c, Data on classes 1, 2 and 3, respectively. Classes 1 and 2 display similar distributions in local resolution, with the majority of the protein in the range 4–5 Å. In class 3 the majority of protein displays a resolution of 4.5–5 Å. In all three cases the refined models agree very well with the maps as shown by comparison of the FSC curves (red) and the FSC curves from the half-maps (blue), and the similarity of the resolution values at FSC 0.143 and 0.5. The estimated resolutions, defined where the line at FSC = 0.143 crosses the blue curve, are 4.27 Å for class 1, 4.35 Å for class 2 and 5.6 Å for class 3. d, Cross-validation of the refinement parameters, confirming lack of over-fitting. For classes 1 and 2, one of the half maps was used for refinement then the FSC curves were calculated for each of the two half maps using the same model.
Extended Data Figure 3 | Example regions of the cryoEM density map for the core subunits, and the model fitted to the map. 

**a.** A selection of TMHs from the membrane domain: TMH3 from ND1, the distorted TMHs in ND6, ND4 and ND5, and a discontinuous TMH from ND2. The series of TMHs from left to right illustrates the decrease in resolution along the domain.

**b.** The two [4FeS4] clusters in the 75 kDa subunit (density in red, at higher contour level) with the protein ligating one of them.

**c.** The FMN cofactor in the 51 kDa subunit.

**d.** The β-sheet in subunit PSST, showing clear separation between the strands.

**e.** Two helices from the 49 kDa subunit.
Extended Data Figure 4 | Example regions of the cryoEM density map for the supernumerary subunits, and the model fitted to the map. 

a, Subunit MWFE, containing one TMH. b, Subunit B14.5, containing two TMHs. The N- and C-terminal loops are not shown. c, The 15 kDa subunit on the IMS face, containing a CHCH domain with two disulfide bonds. The N- and C-terminal loops are not shown. d, The seven-strand β-sheet in the 39 kDa subunit, showing the separation of the strands, and the bound nucleotide (red density) modelled as NADPH. e, Helix 1, one of the arginine-rich helices, in B22, and SDAP-β, on the matrix side of the tip of the membrane domain. Inset: the weak density attached to Ser44 in SDAP-β attributed to the attached pantetheine-4'-phosphate group (side chain of Ser44 not modelled).
Extended Data Figure 5 | Relationships between classes 1 and 3. a, The structures for classes 1 and 3 have been superimposed using ND1 and ND3. In class 3, relative to class 1, the hydrophilic and distal membrane domains are both rotated and shifted, but the change in the membrane domain dominates. Although the changes appear to make the angle of the L-shaped molecule increase they do not originate from a hinge-like motion at the interface of the hydrophilic and membrane domains. Class 1 is in red, class 3 is in red (ND1 domain), wheat, blue and cyan. Details of the composition and movement of the domains are given in Extended Data Table 4. b, The density for class 3 (white) is presented with the model for well-resolved regions of class 3 in blue (the model is enclosed in the density) and the model for poorly resolved regions in red (the model appears outside the density). The poorly resolved regions include the N terminus of the 49 kDa subunit and the transverse helix in ND5, as well as elements of ND4, ND6, B14.7, ESSS and B15.
Extended Data Table 1 | Summary of the models for the core subunits of *B. taurus* complex I

| Subunit | Other names* | Chain | Total residues | Modelled residues | Assigned residues | Unknown residues | % residues modelled | % residues assigned | % with sidechains | % unknown residues |
|---------|--------------|-------|----------------|-------------------|-------------------|------------------|-------------------|-------------------|------------------|------------------|
| ND1     | Nqo8 NuoH    | H     | 318            | 3-200             | 3-200             | 218-315          | 93.1              | 93.1              | 89.0             | 0                |
| class 2 | Nqo8 NuoH    | H     | 318            | 3-315             | 3-315             | -                | 98.4              | 98.4              | 89.0             | 0                |
| ND2     | Nqo14 NuoN   | N     | 347            | 2-345             | 2-345             | -                | 99.1              | 99.1              | 87.6             | 0                |
| ND3     | Nqo7 NuoA    | A     | 115            | 2-27              | 51-112            | -                | 76.5              | 76.5              | 72.2             | 0                |
| class 1 | NuoA         |       |                |                   |                   |                  |                   |                   |                  |                  |
| ND3     | Nqo7 NuoA    | A     | 115            | 2-112             | 2-112             | -                | 96.5              | 96.5              | 72.2             | 0                |
| class 2 | NuoA         |       |                |                   |                   |                  |                   |                   |                  |                  |
| ND4     | Nqo13 NuoM   | M     | 459            | 3-459             | 3-459             | -                | 99.6              | 99.6              | 98.3             | 0                |
| ND4L    | Nqo11 NuoK   | K     | 98             | 2-96              | 2-96              | -                | 96.9              | 96.9              | 96.9             | 0                |
| ND5     | Nqo12 NuoL   | L     | 606            | 2-605             | 2-605             | -                | 99.7              | 99.7              | 88.1             | 0                |
| ND6     | Nqo10 NuoJ   | J     | 175            | 2-172             | 2-172             | -                | 97.7              | 97.7              | 79.4             | 0                |

75 kDa
NDUF51  Nqo3 NuoG   G  704  8-692  8-209  210-692  97.3  28.7  8.9  68.6
51 kDa
NDUFV1  Nqo1 NuoF   F  444  14-438  14-438  -   95.7  95.7  18.2  0
49 kDa
NDUF52  Nqo4 NuoCD  D  430  5-50  61-430  5-50  61-430  -   96.7  96.7  87.0  0
49 kDa
NDUF52  Nqo4 NuoCD  D  430  5-430  5-430  -   99.1  99.1  89.3  0
30 kDa
NDUF53  Nqo5 NuoCD  C  228  8-213  8-213  -   90.4  90.4  85.5  0
24 kDa
NDUFV2  Nqo2 NuoE   E  217  8-193  8-193  -   85.7  85.7  1.8  0
PSST
NDUF57  Nqo6 NuoB   B  179  27-173  27-173  -   82.1  82.1  82.1  0
TYKY
NDUF58  Nqo9 NuoI   I  176  1-176  1-176  -   100  100  92.0  0

* The names of the human, *T. thermophilus* and *E. coli* subunits (if different to the names in *B. taurus*).
### Extended Data Table 2 | Summary of the models for the supernumerary subunits of *B. taurus* complex I

| Subunit | Human name | Chain | Total residues | Modelled residues | Assigned residues | Unknown residues | % residues modelled | % residues assigned | % with sidechains | % unknown residues |
|---------|------------|-------|----------------|-------------------|-------------------|-----------------|-------------------|-------------------|-----------------|------------------|
| 42 kDa  | NDUFA10    | O     | 320            | 5-318             | 91-318            | 5-90            | 98.1              | 71.3              | 26.9            | 26.9             |
| 39 kDa  | NDUFA9     | P     | 345            | 2-186             | -                 | 2-186           | 82.0              | 0                 | 0               | 82.0             |
| class 1 |            |       |                |                   |                   |                 |                   |                   |                 |                  |
| 39 kDa  | NDUFA9     | P     | 345            | 2-336             | -                 | 2-336           | 97.1              | 0                 | 0               | 97.1             |
| class 2 |            |       |                |                   |                   |                 |                   |                   |                 |                  |
| 18 kDa  | NDUFS4     | Q     | 133            | 16-128            | -                 | 16-128          | 85.0              | 0                 | 0               | 85.0             |
| 15 kDa  | NDUFS5     | e     | 105            | 6-94              | 6-94              | -               | 84.8              | 84.8              | 41.9            | 0                |
| 13 kDa  | NDUFS6     | R     | 96             | 5-93              | 59-93             | 5-58            | 92.7              | 36.5              | 27.1            | 56.3             |
| 10 kDa  | NDUFV3     | s     | 75             | 1-35†             | -                 | 1-35†           | 46.7              | 0                 | 0               | 46.7             |
| AGGG    |            |       |                |                   |                   |                 |                   |                   |                 |                  |
| ASHI    | NDUFB8     | I     | 158            | 5-122             | -                 | 5-122           | 74.7              | 0                 | 0               | 74.7             |
| ESSS    | NDUFB11    | g     | 125            | 25-121            | 25-121            | -               | 77.6              | 77.6              | 40.8            | 0                |
| KFY1    | NDUFC1     | c     | 49             | 1-46              | 1-46              | -               | 93.9              | 93.9              | 55.1            | 0                |
| MNL    | NDUFB1     | f     | 57             | 3-56              | 3-56              | -               | 94.7              | 94.7              | 42.1            | 0                |
| MWFE    | NDUFA1     | a     | 70             | 1-64              | 1-64              | -               | 91.4              | 91.4              | 71.4            | 0                |
| PDSW    | NDUFB10    | p     | 175            | 4-172             | 76-142            | 4-75            | 96.6              | 38.3              | 30.9            | 58.3             |
| PGIV    | NDUFA8     | X     | 171            | 5-168             | 5-113             | 114-168         | 95.9              | 63.7              | 53.8            | 32.2             |
| SDAP-γ  | NDUFA81    | T     | 88             | 8-82              | 8-82              | -               | 85.2              | 85.2              | 0               | 0                |
| SDAP-β  | NDUFA81    | U     | 88             | 4-88              | 4-88              | -               | 96.6              | 96.6              | 0               | 0                |
| SGDH    | NDUFB5     | h     | 143            | 7-140             | 7-45              | 46-140          | 93.7              | 27.3              | 21.0            | 66.4             |
| B22     | NDUFB9     | n     | 178            | 10-175            | 10-136            | 137-175         | 93.3              | 71.3              | 35.4            | 21.9             |
| B18     | NDUFB7     | o     | 138            | 57-114            | 57-114            | -               | 42.6              | 42.6              | 2.9             | 0                |
| B17.2   | NDUFA12    | q     | 146            | 2-139             | 2-139             | -               | 95.2              | 95.2              | 0               | 0                |
| B17     | NDUFB6     | i     | 127            | 6-32             | 40-118            | 83.5            | 21.3              | 14.2              | 62.2            | 0                |
| B16.6   | NDUFA13    | Z     | 143            | 5-142            | 31-99             | 5-30            | 96.5              | 48.3              | 48.3            | 48.3             |
| B15     | NDUFB4     | m     | 128            | 11-128           | 32-128           | 11-31           | 92.2              | 75.8              | 66.4            | 16.4             |
| B14.7   | NDUFA11    | Y     | 140            | 1-138             | 1-138            | -               | 98.6              | 98.6              | 98.6            | 0                |
| B14.5a  | NDUFA7     | r     | 112            | 4-70             | -                 | 4-70           | 77.7              | 0                 | 0               | 77.7             |
| B14.5b  | NDUFC2     | d     | 120            | 3-116            | 29-97            | 3-28            | 95.0              | 57.5              | 57.5            | 37.5             |
| B14     | NDUFA6     | W     | 127            | 16-126            | 16-126         | -               | 87.4              | 87.4              | 57.5            | 0                |
| B13     | NDUFA5     | V     | 115            | 8-113            | 8-113            | -               | 92.2              | 92.2              | 40.0            | 0                |
| B12     | NDUFB3     | k∗†  | 97             | 1-74†             | -                 | 1-74†           | 76.3              | 0                 | 0               | 76.3             |
| B9      | NDUFA3     | b     | 83             | 1-80             | 1-45             | 46-80           | 96.4              | 54.2              | 54.2            | 42.2             |
| B8      | NDUFA2     | S     | 98             | 16-95            | 16-95            | -               | 81.6              | 81.6              | 0               | 0                |

*The assignments of B12 and AGGG may be reversed
†Arbitrary residue numbers

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
### Extended Data Table 3 | Subunit–subunit interactions for the supernumerary subunits

| Primary core subunit | Subunit Interactions | Notes |
|----------------------|----------------------|-------|
| 42 kDa               | ND2                  | Nucleoside kinase fold; does not contact hydrophilic core subunits or B14+SDAP-α |
| 39 kDa               | ND2 ND4              | Short-chain dehydrogenase/reductase fold\(^a\) with bound NAD(P)(H)\(^b\) |
| 18 kDa               | ND2 ND3 ND4 ND6 PGIV SGDH B16.6 B14.5b | Four-strand β-sheet with helix located between 75 and 30 kDa |
| 15 kDa               | ND2 ND3 ND4 ND6 PGIV SGDH B16.6 B14.5b | CHCH domain; IMS\(^c\) |
| 13 kDa               | ND5 ND6 B22 B15      | Zinc-binding domain\(^d\) at the interface of TYKY, 75 and 49 kDa |
| 10 kDa               | ND5 ND6 B12          | Present in the three-subunit flavoprotein subcomplex with 51 and 24 kDa |
| AGGG                 | ND5 ND6 B22 B15      | 1 TMH; uncertain assignment of AGGG and B12 between chains j and k |
| ASHI                 | ND5 ND6 B22 B15      | 1 TMH; large globular domain on the matrix face, TMH crosses transverse helix |
| ESSS                 | ND4 ND5 ND6 B22 B15  | 1 TMH; poorly resolved N-terminus on matrix face; possible disulphide 112 to PDSW 154. |
| KFYI                 | ND2 B14.5b           | 1 TMH; attached to complex by B14.5b |
| MNLL                 | ND4 PDSW SGDH       | 1 TMH |
| MWFΔ                 | ND1 B17.2 B16.6 B14.5a | 1 TMH; runs alongside ND1 TMH1 at proposed entrance to Q-binding site |
| PGIV                 | ND1 B16.6 B14.5b     | Two CHCH domains; IMS\(^d\) |
| PDSW                 | ND4 ESSS MNLL SGDH B18 B17 B15 B14.5b | Extensive helix structure on IMS face; two likely internal disulphides (112–124, 76–83); possible disulphide 154 to ESSS-112 |
| SDAP-α               | 30 kDa B14           | ACP on the hydrophilic domain\(^d\) |
| SDAP-β               | ND5 ND6 ASHI B22 B17 B12 | ACP on the membrane domain\(^d\) |
| SGDH                 | ND4 ND5 ND6 PGIV SGDH B17 B16.6 B14.5b | 1 TMH; long helix running along the IMS face of the membrane domain |
| B22                  | ND5 ESSS MNLL PDSW SGDH B18 B17 B15 B14.5a | LVR protein that binds SDAP-β |
| B16                  | ND5 AGGG PDSW B17    | CHCH domain; IMS\(^d\) |
| B17.2                | ND1 B16.6 B14.5a     | Three-strand β-sheet and long loop running across hydrophilic domain |
| B17                  | ND5 PDSW SDAP-β SGDH B22 | 1 TMH; helix on the matrix face, β-strand in IMS augments β-hairpin between NDS TMHs |
| B16.6                | ND1 ND3 ND6 TYKY MWFE PGIV SGDH B14.5a | 1 TMH; 65-residue helix that crosses the membrane then turns along the IMS face |
| B15                  | ND4 ND5 ASHI PDSW B22 B14.7 | 1 TMH; long helix runs across the matrix face of the membrane domain |
| B14.7                | ND7 ND2 ND5 SGDH B15 | 4 TMHs; on the anchor of the transverse helix; likely disulphide 18-75 |
| B14.5a               | ND2 ND4 49 kDa 51 kDa 49 kDa 30 kDa TYKY MWFE B17.2 B16.6 B13 | Long loop structure over hydrophilic domain |
| B14.5b               | ND2 ND4 42 kDa 15 kDa ESSS KFYI PDSW PGIV SGDH B16.6 | 2 TMHs; anchors KFYI to complex |
| B14                  | 30 kDa B17.2 B16.6 B13 | LVR protein that binds SDAP-α\(^d\) |
| B13                  | 30 kDa B14.5a | Three-helix bundle |
| B12                  | ND5 SDAP-β AGGG B22 | 1 TMH; uncertain assignment of AGGG and B12 between chains j and k |
| B9                   | ND1 ND3 ND6 TYKY MWFE PGIV B16.6 | 1 TMH |
| B8                   | 75 kDa              | Thioredoxin-like fold\(^d\); Possible disulphide 23-57 |
Extended Data Table 4 | Allocation of subunits to domains, and the relative movement of domains between classes 1, 2 and 3

| Class 1 vs. class 2 | Subunits | Transformation |
|---------------------|----------|---------------|
| **Heel domain**     | ND1, ND3 | None (reference domain) |
|                     | MWFE, B9, PGIV, B16.6 | |
| **Membrane domain** | ND2, ND4L, ND6, ND4, ND5, N-terminus of 49 kDa subunit* | 3.9° rotation |
|                     | 42 kDa, 15 kDa, KFYI, B14.5b, B14.7, MNLL, AGGG, B12, B15, SGDH, B17, B18, ASHI, B22+SDAP-β, PDSW, ESSS | 0.9 Å shift |
| **Hydrophilic domain** | 75 kDa, 51 kDa, 24 kDa, 30 kDa, 49 kDa (except its N-terminus*), PSST, TYKY | 3.4° rotation |
|                     | B8, B13, B14+SOAP-α, B14.5a, 39 kDa, B17.2, 18 kDa, 13 kDa, 10 kDa | 2.3 Å shift |

| Class 1 vs. class 3 | Subunits | Transformation |
|---------------------|----------|---------------|
| **Heel domain**     | ND1, ND3 | None (reference domain) |
|                     | MWFE, B9, PGIV, B16.6 | |
| **Proximal membrane domain** | ND2, ND4L, ND6 | 0.6° rotation |
|                     | 42 kDa, 15 kDa, KFYI, B14.5b | 0.4 Å shift |
| **Distal membrane domain** | ND4, ND5 | 3.1° rotation |
|                     | B14.7, MNLL, AGGG, B12, B15, SGDH, B17, B18, ASHI, B22+SDAP-β, PDSW, ESSS | 2.9 Å shift |
| **Hydrophilic domain** | 75 kDa, 51 kDa, 24 kDa, 30 kDa, 49 kDa (except its N-terminus), PSST, TYKY | 1.1° rotation |
|                     | B8, B13, B14+SOAP-α, B14.5a, 39 kDa, B17.2, 18 kDa, 13 kDa, 10 kDa | 1.3 Å shift |

*The N terminus of the 49 kDa subunit (residues 5–39) is displaced in class 1 relative to class 2 when considered from the core fold of the subunit because it lies on the surface of the membrane domain and moves with ND2.
Extended Data Table 5 | Data collection, refinement and model statistics for classes 1 and 2

|                             | Class 1 | Class 2 |
|-----------------------------|---------|---------|
| **Data collection**         |         |         |
| Pixel size (Å)              | 1.33    | 1.33    |
| Defocus range (µm)          | 1.8 - 5.5 | 1.8 – 5.5 |
| Voltage (kV)                | 300     | 300     |
| No. of particles            | 48,033  | 33,301  |
| Orientation accuracy (°)    | 0.92    | 0.95    |
| **Model composition**       |         |         |
| Non-hydrogen atoms          | 51,117  | 51,652  |
| Protein residues            | 7,789   | 7,891   |
| % of total                  | 91.5    | 92.7    |
| Core subunit residues       | 4,294   | 4,344   |
| % of total                  | 95.5    | 96.6    |
| Supernumerary subunit residues | 3,495 | 3,547 |
| % of total                  | 87.0    | 88.3    |
| **Refinement**              |         |         |
| Resolution (Å)              | 4.27    | 4.35    |
| Average B-factor (Å²)       | 93.4    | 110.4   |
| **RMS deviations**          |         |         |
| Bonds (Å)                   | 0.008   | 0.008   |
| Angles (°)                  | 1.38    | 1.40    |
| **Validation**              |         |         |
| Molprobity score            | 2.11    | 2.50    |
| Clashscore, all atoms       | 2.91    | 3.35    |
| **Ramachandran plot**       |         |         |
| Favoured (%)                | 86.66   | 86.92   |
| Outliers (%)                | 3.57    | 3.72    |