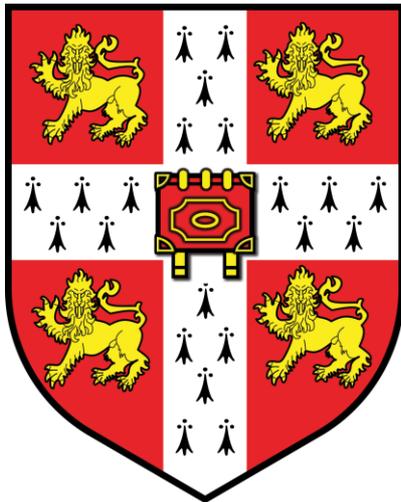# *Cryo-electron microscopy studies on ovine mitochondrial complex I*

**Karol Fiedorczuk**

University of Cambridge

Darwin College

This dissertation is submitted for the degree of Doctor of Philosophy

May 2017

# Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as specified in the text.

I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as specified in the text.

The length of this work does not exceed the prescribed word limit.

Karol Fiedorczuk

May 2017

# Acknowledgements

Firstly, I would like to thank my supervisor Dr Leonid A. Sazanov for his help, support and guidance throughout the course of my PhD. I am very grateful for providing me with this opportunity to participate in this exciting project. Secondly, I would like to thank Dr James A. Letts for his advice in many aspects of the laboratory work and most interesting discussionson various scientific (and non-scientific) topics. Thirdly, I would like to thank Dr Karol Kaszuba, Dr Mark Skehel and Dr Gianluca Degliesposti for all their contributions to this project. I would also like to thank all the previous and current members of the Sazanov group for their help, support and making the laboratory a very enjoyable place to work.

I am deeply grateful to my parents for supporting me through all my life. For the care, advice and love they provided. Without them I would not be where I am today.

This thesis is dedicated to my dearest wife Karolcia. Thank you for your love, patience and faith in me. You are the most wonderful person I know.

# Summary

The main objective of this work is to determine the atomic structure of mammalian respiratory complex I. Mitochondrial complex I (also known as NADH:ubiquinone oxidoreductase) is one of the central enzymes in the oxidative phosphorylation pathway. It couples electron transfer between NADH and ubiquinone to proton translocation across the inner mitochondrial membrane, contributing to cellular energy production. Complex I is the largest and most elaborate protein assembly of the respiratory chain with a total mass of 970 kilodaltons. It consists of 14 conserved 'core subunits' and 31 mitochondria-specific 'supernumerary subunits'. Together they form a giant, L-shaped molecule, with one arm buried in the mitochondrial membrane and another protruding into the mitochondrial matrix.

Here, a novel method for the purification of ovine (Ovis aries) complex I was developed and suitable conditions for cryo-EM imaging established, after extensive screening of detergents and additives. Cryo-EM images were acquired with the recently developed direct electron detector and processed using the latest software. This allowed the solution of the nearly complete atomic model of mitochondrial complex I at 3.9 Å resolution. The membrane part of the complex contains 78 transmembrane helices, mostly contributed by conserved antiporter-like subunits responsible for proton translocation. These helices are stabilized by tightly bound lipids (including cardiolipins). The hydrophilic arm harbours flavin mononucleotide and 8 iron–sulfur clusters involved in electron transfer. Supernumerary subunits build a scaffold around the conserved core, strongly stabilizing the complex. Additionally, subunits containing cofactors (NADPH, zinc ion and phosphopantetheine) may play a regulatory role. Two distinct conformations of the complex are observed, which

may describe the active and deactive states or reflect conformations occurring during the catalytic cycle of the enzyme.

Currently this is the most detailed model of this molecular machine, providing insight into the mechanism, assembly and dysfunction of mitochondrial complex I. It also allows molecular analysis of numerous disease-causing mutations, and so the structure may serve as a stepping-stone for future medical developments.

# Abbreviations

| | |
|---|---|
| Å | Ångstrom |
| ADP | Adenosine diphosphate |
| ATP | Adenosine triphosphate |
| BSA | Bovine serum albumin |
| Ca | Calcium |
| CAC | Citric Acid Cycle |
| CL | Cardiolipin |
| CMOS | Complementary metal oxide semiconductor |
| $CO_2$ | Carbon dioxide |
| CoA | Coenzyme A |
| dALA | δ-Aminolevulinic acid |
| DQ | Decylubiquinone |
| DQE | Detective quantum efficiency |
| DTT | Ditiotreitol |
| e- | Electron |
| EC | Electron counting |
| EDTA | Ethylenediaminetetraacetic acid |
| Em | The standard electric potential |
| EM | Electron Microscopy |
| ER | Endoplasmic reticulum |
| $FADH_2$ | Flavin adenine dinucleotide |
| FEG | Field emission guns |
| FMN | Flavin mononucleotide |
| FSC | Fourier shell correlations |
| GTP | Guanosine tri-phosphate |
| $H^+$ | Proton |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| Hz | Hertz |
| IMM | Internal mitochondrial membrane |

| | |
|---|---|
| IMS | Inter membrane space |
| ISC | Iron-sulphur clusters |
| kbp | Kilobase pair |
| kDa | Kilodalton |
| keV | Kiloelectron volt |
| M | Magnification |
| MAM | Mitochondria-associated ER-membrane |
| MD | Membrane domain |
| MDa | Megadalton |
| mDNA | Mitochondrial DNA |
| MM | Mitochondrial matrix |
| NaCl | Sodium chloride |
| NADPH | Nicotinamide adenine dinucleotide phosphate |
| NADH | Nicotinamide adenine dinucleotide |
| NO | Nitric oxide |
| $O_2$ | Oxygen |
| $O_2$-• | Superoxide radical |
| OMM | External (outer) mitochondrial membrane |
| OXPHOS | Oxidative phosphorylation |
| PA | Peripheral arm |
| PBG | Porphobilinogen |
| PC | Phosphatidylcholine |
| PE | Phosphatidylethanolamine |
| PHG | Porphobilinogen |
| Pi | Phosphate anion |
| PMSF | Phenylmethylsulfonyl fluoride |
| Q | Quinone |
| QH2 | Quinol |
| PNS | Phosphopantetheine |
| ROS | Reactive oxygen species |
| rRNA | Ribosomal RNA |
| S/N | Signal to noise |
| TCA | Tricarboxylic *acid* |
| TM | Trans membrane |
| TMH | Trans membrane helix |

| | |
|---|---|
| tRNA | Transfer RNA |
| UPP | Uroporphyrinogen |
| v/v | volume to volume |
| w/v | weight to volume |
| w/w | weight to weight |
| Δp | Proton gradient |

# Table of content

# Chapter 1. Introduction

## 1.1 Mammalian mitochondria

Mitochondria are organelles present in most eukaryotic organisms (Karnkowska et al., 2016). They are double membrane entities of around 5 µm in size. Although the most prominent function of mitochondria is the production of adenosine triphosphate (ATP), they are involved in many other, very distinct and important processes such as cell signalling, apoptosis or protein maturation (Friedman & Nunnari, 2014).

The number of mitochondria varies depending on the cell type, from none in human erythrocytes (Moritz, Lim, & Wintour, 1997) to a few thousand in skeletal muscle cells, taking up to 20% of the whole cell volume (Alberts, 2015).However on average they are present in numbers of around several hundred. The shape of mitochondria also differs depending on the organism and tissue (Campello & Scorrano, 2010). They are mostly oval but elongated or branched forms are not unusual. The location of mitochondria in the cell is not fixed. Due to cytoplasmic movements or cytoskeleton transporting machinery they can move around or travel in the direction of higher energy demand like myofibrils or synaptic ends of the nerve cells (Anesti & Scorrano, 2006).

### 1.1.1 Evolution

The currently most popular theory states that mitochondria (and similarly plastids) developed during endosymbiosis of bacteria by ancient eukaryotic organisms. The most likely precursor of mitochondria is Rickettsiales family (Dyall, Brown, & Johnson, 2004)(Emelyanov, 2001). The first one to propose the endosymbiotic theory was Ivan Wallin in 1922 (Wallin, 1923). The theory was mostly forgotten for the next few decades until 1967 when it was re-published more completely and with stronger evidence by Lynn Margulis

(Margulis, 1970). It is estimated that the first endosymbiosis events happened almost two billion years ago. During evolution, the ancestral endosymbiotic organism has undergone vast modifications. The host overtook most of its genetic material by moving genes to the nucleus and took control of division regulation, in order to synchronise it with the cell cycle. Nevertheless we can still trace a number of similarities of mitochondria to their bacterial ancestors. They divide by fission; contain circular DNA; contain their own ribosomes – the mito-ribosome – which is more similar to the bacterial 70S ribosome than eukaryotic 80S ribosome; the translational start amino acid residue is N-formylmethionine; and many transport systems and enzymatic pathways of mitochondria strongly resemble bacterial versions (Gray, Burger, & Lang, 1999).

An alternative, less likely, theory states that mitochondria are a product of cell membrane invagination in a process similar to the separation of the nucleus (Gray et al., 1999). Regardless of their origin, development of mitochondria was a tremendous step in evolution enabling cells to perform oxidative respiration on internal membranes and thus more efficiently use available energy sources.

### 1.1.2 Mitochondrial architecture

In mitochondrial morphology we can indicate four different compartments: the internal mitochondrial membrane (IMM), the external (outer) mitochondrial membrane (OMM), the inter membrane space (IMS) and the mitochondrial matrix (MM) (Logan, 2006) (Fig. 1.1.1). External and internal mitochondrial membranes are distinctly different both in function and protein/lipid composition.
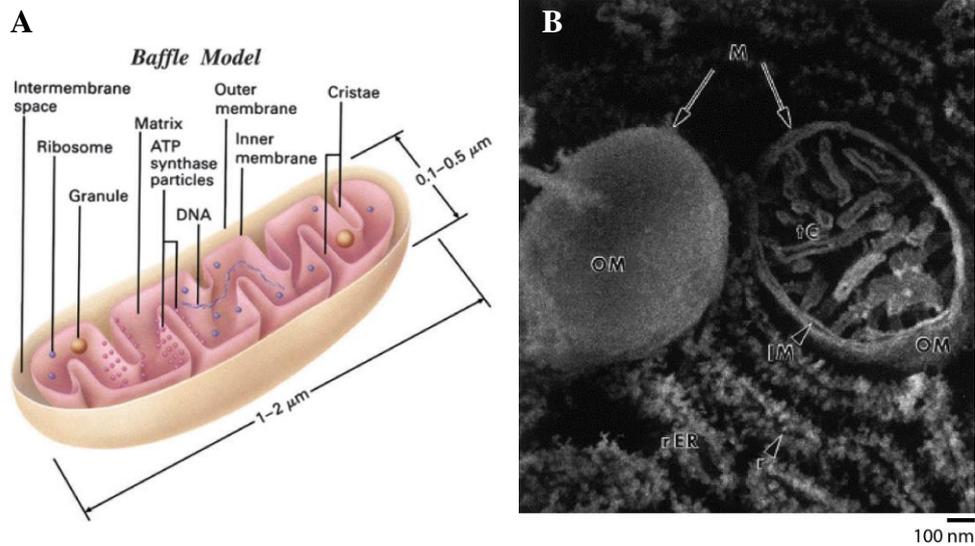
**Figure 1.1** Mitochondrial structure. **A**. The baffle model. As proposed by Palade (Logan, 2006). **B.** Ultra fine structure of mitochondria revealed by high-resolution scanning electron microscopy form rat pancreas exocrine cells. Fracture plane revealing the surface (left) and interior structure (right) of mitochondria (M). The right-hand side mitochondrion appears to have both tubular and lamellar cristae. Labelled features are outer mitochondrial membrane (OM), inner mitochondrial membrane (IM), cristae (C), tubular cristae (tC), rough endoplasmic reticulum (rER) and ribosome (r) (Perkins & Frey, 2000).

The outer mitochondrial membrane separates the organelle from the cytosol. It has a composition characteristic of other mammalian membranes. It contains many porins (2-3 nm in diameter), thus is very permeable for molecules under 5 kDa. Small molecules like water or carbon dioxide freely diffuse through the membrane as well. Bigger proteins are actively transported by translocases (Walther & Rapaport, 2009). The OMM interacts with other cellular organelles like the endoplasmic reticulum (ER) forming the so-called mitochondria-associated ER-membrane (MAM) (Fujimoto & Hayashi, 2011). Enzymatic reactions catalysed on the surface of OMM are mostly involved in secondary metabolism and lipid synthesis (Walther & Rapaport, 2009).

The inter membrane space is the area between the internal and external mitochondrial membranes. Because of the high permeability of the OMM the ion composition there is identical to the cytosol. It is the smallest of all the mitochondrial compartments but plays many very important roles: it oversees the transport and modification of proteins – by the mitochondrial inter

membrane space assembly machinery – and other particles to and from the mitochondria; and it regulates the respiratory chain complexes and coordinates apoptosis (Herrmann & Riemer, 2010). In spite of the general oxidative environment, it contains thioredoxins and thioredoxin reductases that provide the reductive capacity of the compartment important during transport of oxidation sensitive proteins and as a protection from reactive oxygen species (ROS) (Conrad et al., 2004)(Kim et al., 2003). In yeasts it has been shown that Cu,Zn-superoxide dismutase and its metallochaperone (CCS) localize in the IMS and are very important for reactive oxygen physiology (Sturtz, Diekert, Jensen, Lill, & Culotta, 2001).

The internal mitochondrial membrane separates the mitochondrial matrix from the rest of the cell. It is rich in proteins. The weight/weight (w/w) protein-to-lipid ratio is 3:1 (Karp, 2013). Permeability of this membrane, contrary to OMM, is very low (Paradies, Paradies, De Benedictis, Ruggiero, & Petrosillo, 2014). The transport of ions and other small molecules is tightly regulated to maintain their concentrations, enabling preservation of the electrochemical gradient required for ATP production. Twenty percent of all mitochondrial proteins are localised in IMM. Many are involved in electron transfer proton pumping and ATP synthesis (Alberts, 2015). The rest is responsible for the translocation of metabolites, and the synthesis/degradation of proteins. The area of the internal membrane is a few times larger than the outer membrane, most likely to increase the capacity for ATP production. The surface size of IMM varies between types of cells and is bigger in the cells that require more energy (Mannella, 2006).

The mitochondrial matrix (MM) is the inner space of the mitochondria. It contains most of mitochondrial proteins. This is where the β-oxidation and Krebs cycles are localised. Mitochondria just like plastids contain their own DNA (mDNA) stored in the MM (Fig. 1.1.A). It encodes several proteins required for mitochondrial function. To process the mDNA mitochondria posses transcription and translation machinery that is independent from the cellular ones. The mitochondrial genome of humans contains up to 10 copies of circular mDNA of 16569 base pairs (Fig. 1.2) per mitochondrion (Wiesner, Ruegg, & Morano, 1992). They encode 37 genes including 13 proteins, 22 transfer RNAs (tRNA)

and 2 ribosomal RNAs (rRNA). The genetic code of mDNA in higher eukaryotes is similar to that in bacteria and in most cases does not contain introns (Friedman & Nunnari, 2014).
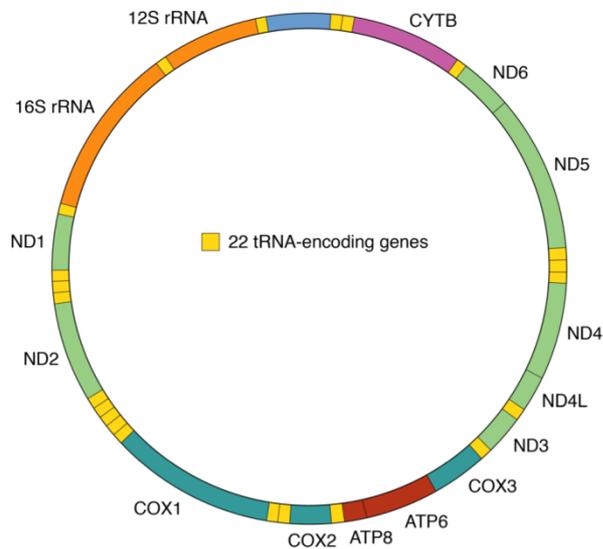
**Figure 1.2** Schematic of human mitochondrial DNA. Human mtDNA encodes 13 proteins, two rRNAs and 22 tRNAs in a 16 kbp circular chromosome. Genes encoding components of complexes I, III, IV and ATP synthase are coloured green, magenta, turquoise and red respectively. Genes encoding tRNAs are coloured gold. Figure adapted from that created by Dr David T. Gallagher.

In mammals, mitochondria are inherited maternally in many copies. The copies are not always identical (heteroplasmic) and can change throughout the life of an individual. During cell division they allocate randomly to the daughter cells (Mishra & Chan, 2014). Damage caused by reactive oxygen species (ROS) – produced mostly by respiratory complex I – can lead to mDNA disease-causing mutations (Murphy, 2009). Such alterations mostly affect tissues of high-energy demand like the brain or skeletal muscles (Gorman et al., 2016). In spite of the mDNA repair pathways (Kazak, Reyes, & Holt, 2012) damage to genetic material is accumulated over time. This phenomenon is the basis for the widely accepted but still unproven mitochondrial free radical theory of ageing, which states that ROS provoke mutations altering mitochondrial function that negatively affect life span in mammals (Sanz & Stefanatos, 2008).

## 1.2 ATP production in mammalian mitochondria

The most important role of mitochondria is the production of ATP. This oxygen dependent process is mostly based on the oxidation of the degradation products from glucose and fatty acids, namely NADH and pyruvate (Berg, Tymoczko, & Stryer, 2002) (Fig 1.3 Stage I). Two distinct mitochondria-located processes are involved in ATP production. First, in the Krebs cycle – located in the mitochondrial matrix (Fig. 1.3 Stage II) – NADH is produced (Akram, 2014). Second, the oxidative phosphorylation (OXPHOS) chain enzymes – located in the IMM (Fig. 1.3 Stage III) – transform adenosine diphosphate (ADP) into ATP (Kuhlbrandt, 2015).
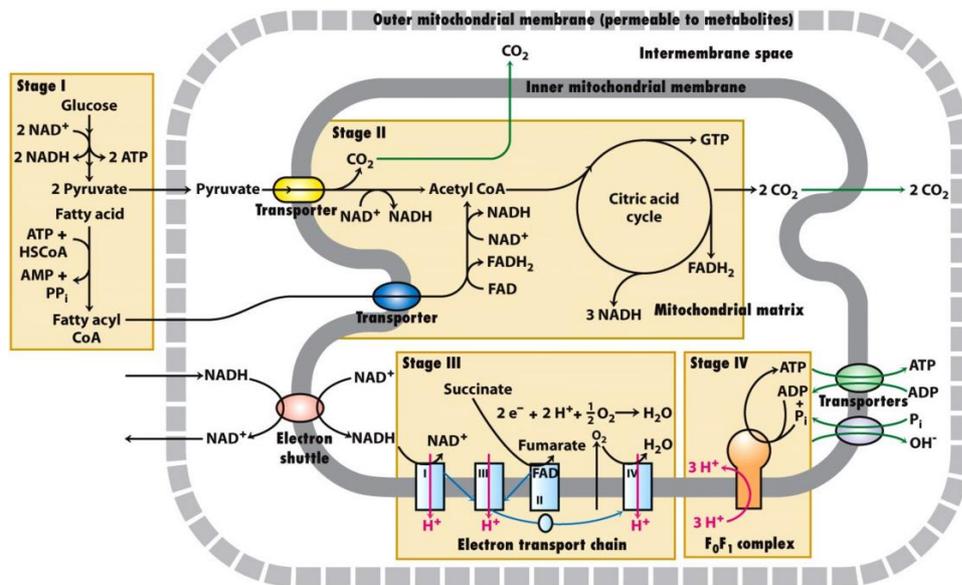
**Figure 1.3** Schematic overview of the metabolic processes that take place within mitochondria. These pathways, such as glycolysis, fatty acid β-oxidation (Stage I) and the TCA cycle (Stage II), break down metabolites thus donating electrons to $NAD^+$ and $FAD^+$. The highly reducing electrons are then donated to the respiratory chain machinery: complexes I-IV (Stage IV). The respiratory chain pumps protons from the matrix to the IMS, thereby generating a proton electrochemical gradient. Energy released by protons moving back along this gradient can be harnessed to make ATP or drive other processes (Stage IV). Note the difference in IMM and OMM permeability and location of metabolites transporters. Figure adapted from (Lodish, 2008).

*1.2.1 Krebs' Cycle (Citric Acid Cycle, CAC; Tricarboxylic acid (TCA) Cycle)*

The final product of glycolysis in the cytosol is pyruvate, which is actively transported to the mitochondrial matrix. It undergoes oxidation and together with coenzyme A (CoA) creates acetyl coenzyme A (acetyl-CoA). As by-products $CO_2$ and NADH are produced. Acetyl-CoA enters the Krebs' cycle in the mitochondrial matrix (Berg et al., 2002). All of the enzymes in the CAC are soluble proteins except for succinate dehydrogenase (complex II) that is bound to the internal mitochondrial membrane. As a result of the process acetyl-CoA is oxidised to $CO_2$ and three types of molecules are produced: 3 molecules of NADH, one of flavin adenine dinucleotide ($FADH_2$)–which will be the source of electrons for the respiratory chain–and a molecule of guanosine tri-phosphate (GTP) – which is eventually converted to ATP by nucleoside - diphosphate kinase (Fig. 1.4) (Berg et al., 2002).
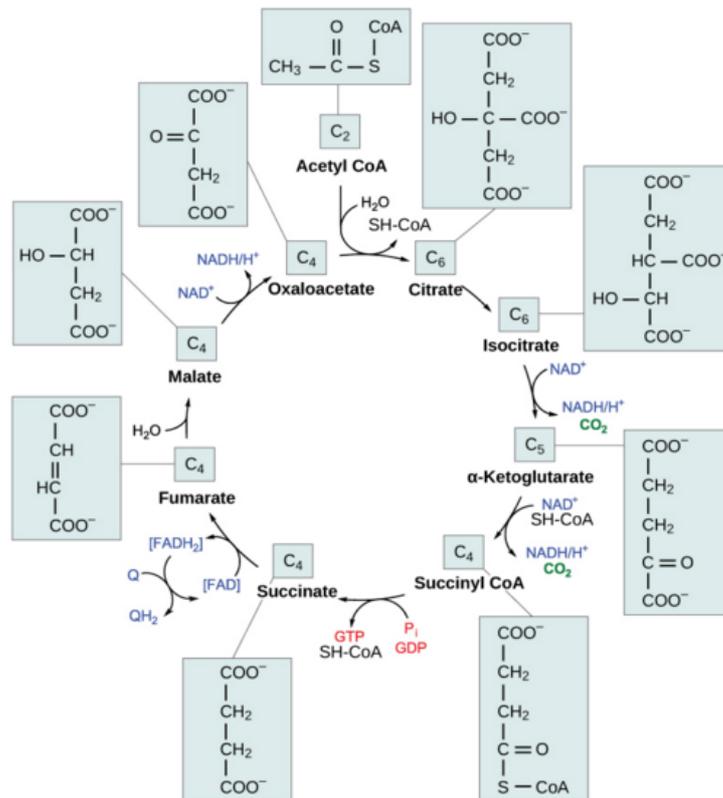
**Figure 1.4** Schematic overview of CAC. The acetyl group from acetyl CoA is attached to a four-carbon oxaloacetate molecule to form a six-carbon citrate molecule. Through a series of steps, citrate is oxidized, releasing two carbon dioxide molecules for each acetyl group fed into the cycle. In the process, three $NAD^+$ molecules are reduced to NADH, one FAD molecule is reduced to $FADH_2$, and one ATP or GTP (depending on the cell type) is produced (by substrate-level phosphorylation). Because the final product of the CAC is also the first reactant, the cycle runs continuously in the presence of sufficient reactants. Figure adapted from work published by Rice University (under a Creative Commons Attribution License 3.0)

### 1.2.2 Respiratory chain overview

NADH and $FADH_2$ are electron sources for the respiratory chain. The molecules are produced mostly by the Krebs' cycle and the ones formed in the cytoplasm are transported into mitochondria via glycerophosphate or malic acid-asparagine carriers (Nicholls & Ferguson, 2002). Harvested electrons are used by NADH dehydrogenase, cytochrome bc1 and cytochrome c oxidase to pump protons ($H^+$) across the mitochondrial membrane and create an electrochemical gradient (Nicholls & Ferguson, 2002). Protons returning to the mitochondrial matrix promote the catalytic movement of ATP synthase, which synthesizes ATP form ADP and an inorganic phosphate anion ($P_i$) (Fig. 1.5) (J. E. Walker, 2013)(Kuhlbrandt, 2015).
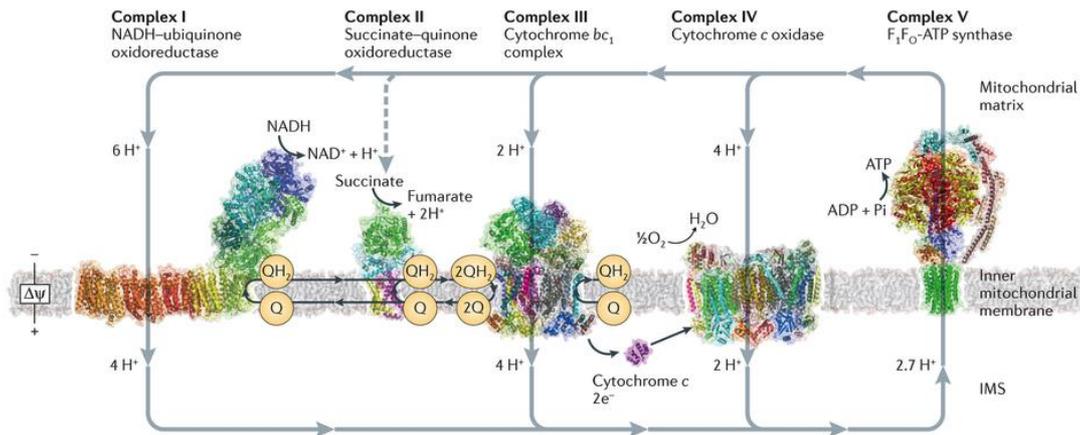
**Figure 1.5** Components of the mitochondrial electron transport chain. Structures of the respiratory complexes are depicted as molecular surfaces overlaid with cartoon representations. Each individual subunit of a protein complex is coloured differently. Complexes I, II and IV pump protons across the inner mitochondrial membrane, thereby generating a proton gradient ($\Delta p$). Protons moving back down this electrochemical gradient provide the driving force for ADP phosphorylation by the $F_1F_O$-ATPase. Complex II does not translocate protons, instead it serves as an entry point for electrons from the CAC intermediate succinate and contributes to the reduction of the quinone (Q) pool. Electrons are transferred from complex III to complex IV by the small soluble protein cytochrome c (coloured purple). Figure adapted from (L. A. Sazanov, 2015).

*1.2.3 Mitochondrial complex I (NADH:ubiquinone oxidoreductase)*

The mitochondrial complex I is the first and largest enzyme of the respiratory chain. In mammals it is composed of 45 subunits making it almost 1 MDa total (J.E. Walker, 1992)(Hirst, 2013).

Complex I harbours eight iron-sulphur (FeS) clusters, of which seven have known catalytic functions. It accepts electrons from NADH produced in the Krebs' Cycle through the two electron acceptor flavin mononucleotide (FMN) non-covalently bound in the 51 kDa subunit. Electrons are transferred through the iron-sulphur clusters to the ubiquinone-10 (Q) molecule. Reduction of the ubiquinone-10 to ubiquinol-10 ($QH_2$) is coupled to proton translocation across the IMM. Proton pumping is performed by the anti-porter like subunits in the membrane part of the protein (Efremov & Sazanov, 2011)(L. A. Sazanov, 2015). From the oxidation of one NADH molecule, complex I reduces one Q molecule and transfers 4 protons across the IMM (Fig. 1.6) (Alexander Galkin, Dröse, & Brandt, 2006).
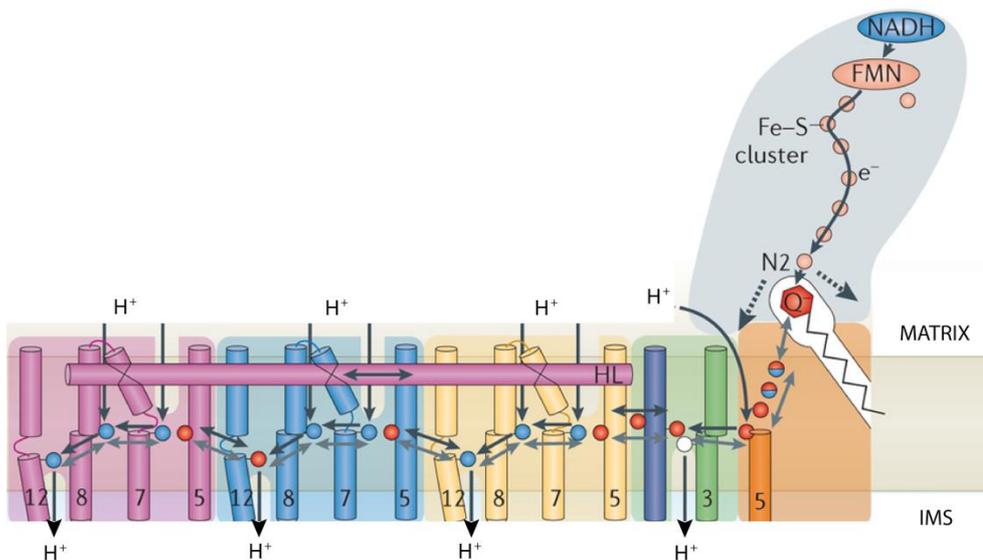
**Figure 1.6** Schematic overview of the catalytic compartments and putative mechanism of respiratory complex I. Upon electron transfer from the last FeS cluster N2, negatively charged quinone (Q) (or charged residues nearby) initiates a cascade of conformational changes, propagating from the E-channel (orange and green blocks) to the antiporters (yellow, blue, magenta blocks) via the central axis (indicated by grey arrows) comprising charged and polar residues that are located around flexible breaks in key transmembrane helices (TMHs). Cluster N2-driven shifts (dashed arrows) may help in the process. Helix HL either helps to coordinate conformational changes by linking discontinuous TMHs between the antiporters or stabilizes the structure. Key charged residues can be protonated from the cytoplasm through several possible pathways, including inter-subunit transfer (indicated by black arrows). Following the reduction of quinone and completion of conformational changes, the antiporters and the E-channel each eject a proton into the periplasm. Crucial TMHs are numbered and key charged residues are indicated by red circles for negative, blue circles for positive, and white circle for uncharged. FMN, flavin mononucleotide. Figure modified from (L. A. Sazanov, 2015).

### 1.2.4 Mitochondrial complex II (succinate dehydrogenase)

The mitochondrial complex II does not contribute to the electrochemical gradient across the mitochondrial membrane and is neutral in respect of $H^+$ balance. Oxidation of succinate produces two $H^+$ which are bound again during the reduction of Q. Complex II via covalently bound flavin adenine dinucleotide (FAD) collects electrons from succinate and via three iron-sulphur centres transfers them to the Q molecule (Fig. 1.7). Overall complex II enlarges the pool of ubiquinol-10 (Crofts, 2004).
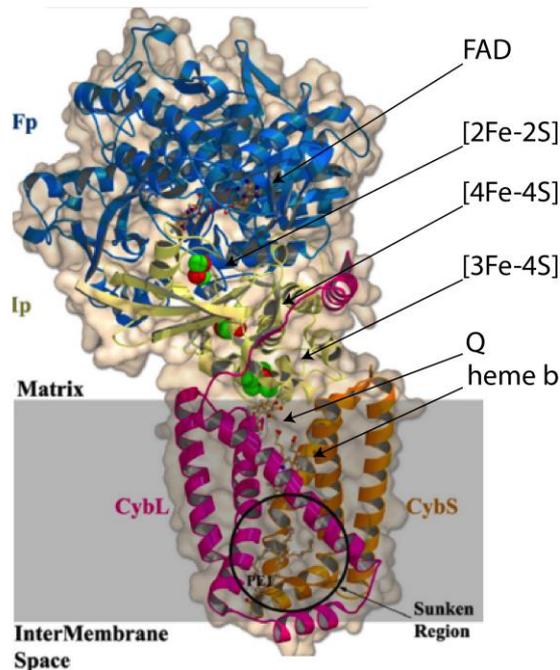
**Figure 1.7** Overall structure of the mitochondrial respiratory Complex II. The ribbon diagram of the complex is superimposed on the semitransparent molecular surface. The FAD binding protein (Fp) is shown in blue; the iron-sulphur protein (Ip) is shown in cream; the two transmembrane proteins CybL and CybS are shown in pink and gold, respectively. The putative membrane region is shaded in grey. FAD, [2Fe-2S], [4Fe-4S], [3Fe-4S], heme, ubiquinone (Q) phosphatidylethanolamine (PE1) and hydrophobic sunken region are labelled. Figure adapted form (Sun et al., 2005).

*1.2.5 Mitochondrial complex III (cytochrome $bc_1$ complex)*

The mitochondrial complex III oxidises the ubiquinol-10 pool in the process known as the Q-cycle (Fig. 1.8.C) (Mitchell, 1975). The first of the electrons harvested from the ubiquinol molecule that binds at the IMS site ($Q_o$) is transferred to the final electron acceptor of complex III – cytochrome *c* (cyt*c*). The second electron is transferred through the hemes $b_L$ and $b_H$ to a ubiquinone molecule at the matrix binding site ($Q_i$) creating ubisemiquinone. Ubisemiquinone is fully reduced to ubiquinol after the second ubiquinol molecule at $Q_o$ gets oxidised, another cyt*c* is reduced and the remaining electron is passed to the ubisemiquinone at the $Q_i$ site forming a fully reduced molecule of ubiquinone. Transfer of electrons from the $Q_o$ site to cytochrome $c_1$ (CYC1) is mediated by movement of the 2Fe-2S cluster of the Rieske subunit between its b-state proximal to $QH_2$ and its c-state proximal to CYC1 (Fig. 1.8A - B). The possible reason for this complicated mechanism is that the ubiquinol molecule is completely reduced when binding to the $Q_o$ site but cyt *c* is only

able to accept one electron at a time. In the Q-cycle process two protons are consumed from the matrix, four protons are released into the inter membrane space and two electrons are passed to two cyt *c* molecules (Moser, Farid, Chobot, & Dutton, 2006).
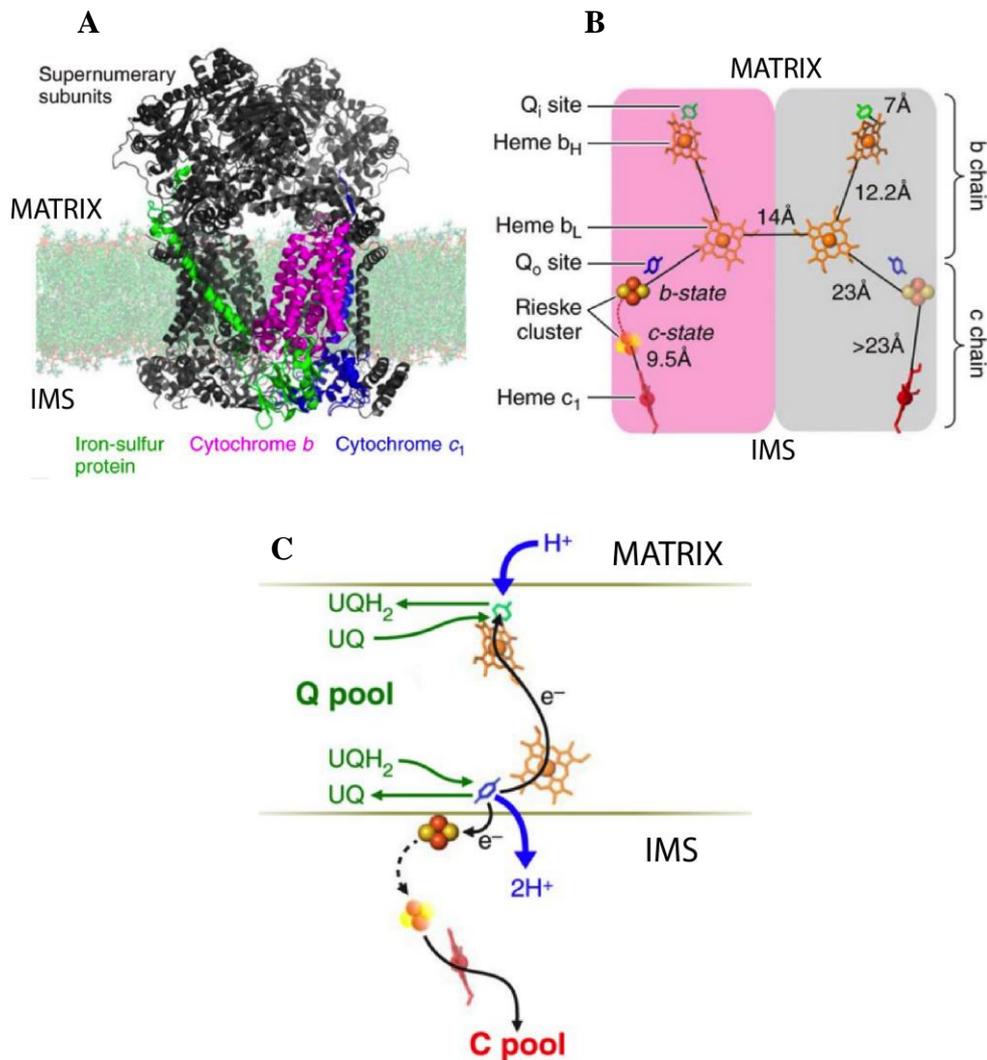


**Figure 1.8** Overview of structural elements and major functional principles of mitochondrial complex III. **A.** Subunit composition of mitochondrial cytochrome bc$_1$ with three catalytic subunits of one monomer coloured. **B.** Spatial arrangement and distances between cofactors and catalytic sites in the dimer. **C.** Scheme of electron and proton transfers within the monomer in the forward mode. Note that dotted lines in **B.** and **C.** denote large-scale movement of the FeS head domain harbouring the Rieske cluster between b- and c-states. Figure adapted from (Sarewicz & Osyczka, 2015).

### 1.2.6 Mitochondrial complex IV (cytochrome c oxidase)

Mitochondrial complex IV is the acceptor of the reduced cyt *c*. Here cyt *c* is oxidised and electrons meet their final acceptor–molecular oxygen (O$_2$). As a result of the oxidoreduction reaction catalysed by complex IV, two molecules of

water are formed using electrons harvested from four cyt *c* molecules and four protons from MM (Fig. 1.9). Additionally four protons are pumped out from the mitochondrial matrix through the complex (Verkhovsky, Jasaitis, Verkhovskaya, Morgan, & Wikstrom, 1999).
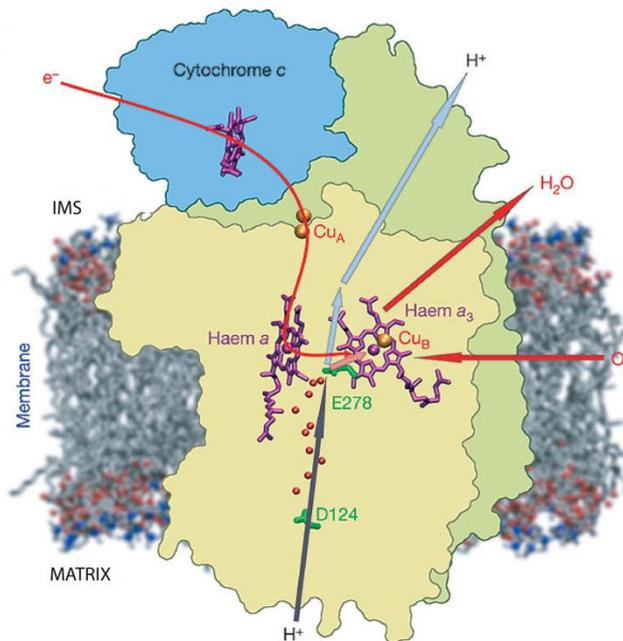


**Figure 1.9.**Overall structure of the mitochondrial respiratory Complex IV. Subunits I (yellow) and II (green) are depicted in the phospholipid membrane, together with docked cyt *c*. Electrons from cyt c are transferred (curved red arrow) via the $Cu_A$ site and haem *a* to the active site (haem a3/$Cu_B$). Proton transfer from the MM side of the membrane via the D-pathway (grey arrow) leads to E278. E278 donates protons, both to the haem a3/$Cu_B$ site to produce water from reduced $O_2$ (light red arrow), and for pumping across the membrane (blue arrows). The red spheres are crystallographically observed water molecules in the D-pathway. Figure adapted form (I. Belevich, Verkhovsky, & Wikstrom, 2006).

This gives the final number of ten net protons translocated from the mitochondrial matrix to inter membrane space by complexes I-IV per one NADH molecule oxidised (Kuhlbrandt, 2015).

### 1.2.7 Mitochondrial complex V ($F_1F_o$ATP synthase)

The ATP synthase is the last element of the OXPHOS system. In general it is built of two components: the $F_o$ domain embedded in the inner mitochondrial membrane and the $F_1$ domain protruding in to the matrix. After the electrochemical gradient has been built up, it is used by the ATP synthase for

ATP production. The $F_o$ domain is built mainly of the so - called c – ring and works as an ion transporter enabling proton's re-flux into the MM. This creates a motive force in the $F_1$ domain facilitating ADP phosphorylation (Fig. 1.10). In this way energy stored originally in the form of NADH and transformed into an electrochemical gradient across IMM is harvested in the form of ATP (J. E. Walker, 2013).
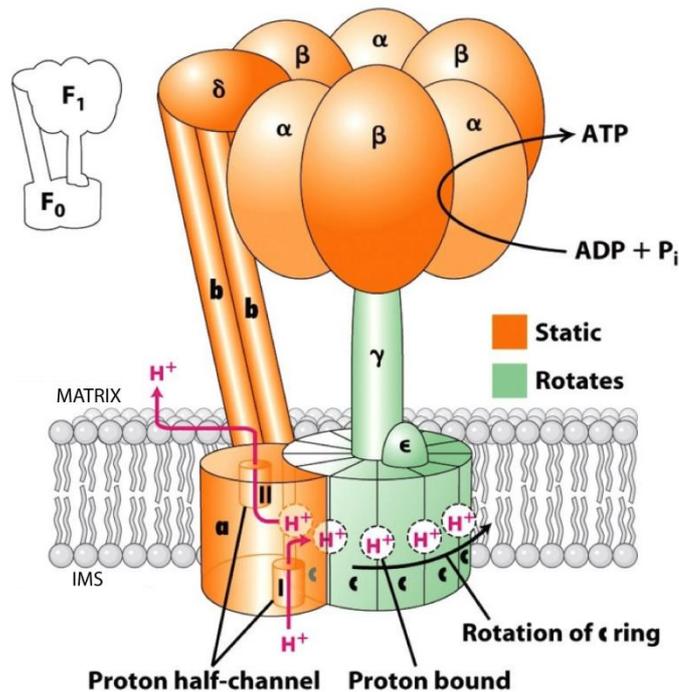


**Figure 1.10.** Schematic overview and the mechanism of ATP synthase. Rotation of the asymmetric γ-stalk drives the interconversion between α and β subunits of the $F_1$ domain facilitating ATP synthesis. Proton translocation through the hydrophobic domain drives c-ring rotation of the $F_o$ domain. Subunit a is proposed to provide a pathway for protons from the IMS to access binding sites on the c subunits as well as a channel for proton release into the matrix. Sequential proton binding and release events drive c-ring rotation, which is transferred to the catalytic domain by the γ-stalk. The peripheral stalk formed by the two b subunits prevents rotation of the $F_1$ α-β domains. Figure adapted from (Lodish, 2008).

## 1.3 Other mitochondrial Functions

### 1.3.1 Reactive Oxygen Species

During the OXPHOS process most of the electrons eventually reach oxygen at cytochrome c oxidase to produce water. However, electron transport through complexes I-IV is not perfect and some of the "leaked" electrons reduce oxygen in an ungoverned manner creating ROS. This contributes to oxidative stress, alters mitochondrial function and leads to cell decay. It is estimated that in isolated mitochondria 0.1 – 2% of OXPHOS electrons reduce oxygen incompletely, mostly at complex I and III producing superoxide radicals ($O_2^{-\bullet}$) (Murphy, 2009).

At high concentrations ROS can wreak havoc to mitochondrial DNA but at moderate concentrations is a desirable cell regulator and a very important component of the cell signalling system. It is produced by tightly regulated enzymes, such as thenitric oxide (NO) synthase and NAD(P)H oxidase isoforms (Thannickal & Fanburg, 2000). It is also produced in a process referred to as the respiratory burst, an important immunological response of phagocytic cells (Robinson, 2008).

### 1.3.2 Apoptosis

In the case of serious damage to the mitochondria, the cell undergoes apoptosis. The Bcl2 protein is a sensor for mitochondrial damage. When activated it heterodimerises with a Bax class protein and together, by interaction with the mitochondrial voltage-dependent anion channel (VDAC), they increase permeability of the mitochondrial inner membrane. As a result the membrane potential is lost and cyt *c* is released. In the cytoplasm cyt*c* interacts with the Apaf-1 protein and together forms the apoptosomes. The outcome is caspase-9 activation and triggering of processes directly leading to death of the cell (C. Wang & Youle, 2009).

### 1.3.3 Heat production

In some cases protons can return to the mitochondrial matrix without involvement of the ATP synthase. This process is called 'proton leaking' and occurs by passive facilitated diffusion through the ion channel thermogenin (UCP1). The energy from the electrochemical gradient is dispersed as heat. This reaction is characteristic for brown adipose tissue (BAT). BAT is present in mammals shortly after birth and is important for those organisms that undergo winter hibernation.

### 1.3.4 Calcium homeostasis

Calcium homeostasis is maintained and regulated through mitochondria. Calcium is a very important ion for cell signalling. Most of the calcium in the cell is stored in the endoplasmic reticulum (ER). However, extensive interactions between ER and mitochondria (Chapter 1.1.2) enable easy exchange of this ion. Calcium diffuses freely through the OMM and can be transported from the IMS to MM by the mitochondrial $Ca^{2+}$ uniporter. Mitochondria can uptake calcium from the cytoplasm and release it rapidly if necessary. That makes them a temporary storage organelle and a very good calcium buffer (Rizzuto, De Stefani, Raffaello, & Mammucari, 2012). Expulsion of the ions causes changes in the membrane potential that invokes certain reactions, for example in a nerve cell, release of a neurotransmitter (Sudhof, 2012).

### 1.3.5 Biosynthesis of proteins and co-factors

Mitochondria are also very important places for the synthesis of different classes of molecules, such as porphyrins (Straka, Rank, & Bloomer, 1990), iron-sulphur-proteins (Ye & Rouault, 2010) and steroids (Ghayee & Auchus, 2007).

Porphyrins are heterocyclic molecules composed of four pyrrole rings connected with methine (-CH=) groups. They mostly provide prosthetic groups of proteins like haemoglobin (Fig 1.11). Since they are often key components of important enzymes their biosynthesis is very well studied (Straka et al., 1990). Porphyrin synthesis is a multistep process taking place in the cytoplasm and IMS. It starts form δ-aminolevulinic acid (dALA) formed from glycine and

succinyl CoA in a reaction known as the Shemin pathway (Straka et al., 1990). Two molecules of dALA are then combined to porphobilinogen (PBG) (a pyrrole ring containing molecule). Four PBGs are fused together to create tetrapyrroleuroporphyrinogen III (UPPIII). UPPIII is decarboxylated to form coproporphyrinogen III, the main porphyrin precursor imported into mitochondria. In the IMS it forms the main product: protoporphyrin IX that can be modified to obtain different compounds and co-factors, for example haem by iron incorporation (Straka et al., 1990).
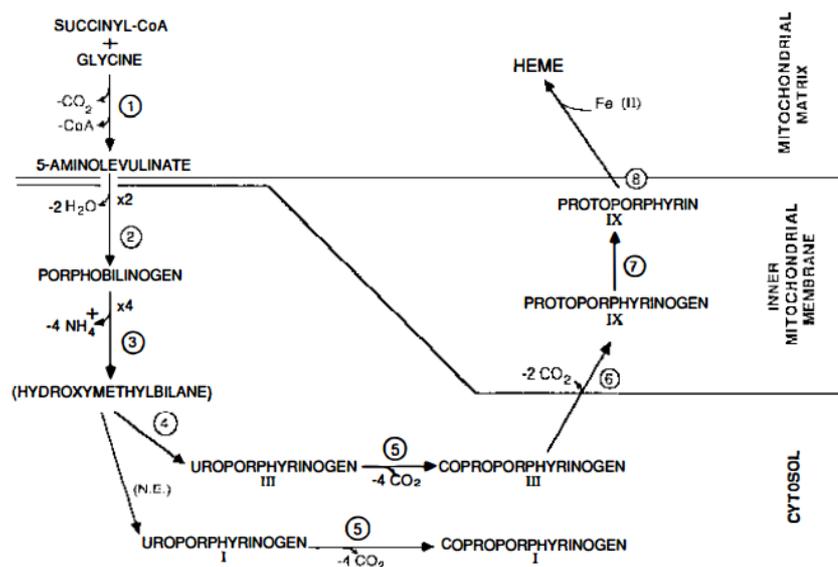


**Figure 1.11** The haem biosynthetic pathway. Enzymes of haem biosynthesis are indicated by numbers. (1) 5-aminolevulinic acid (ALA) synthase. (2) ALA dehydrase. (3) porphobilinogen.(PHG) deaminase. (4) uroporphyrinogen III synthase. (5) uroporphyrinogen decarboxylase. (6) coproporphyrinogen III oxidase. (7) protoporphyrinogen IX oxidase. (8) ferrochelatase. N.E. = nonenzymatic. Figure adapted form (Straka et al., 1990).

Iron-sulphur clusters (ISCs) are ancient co-factors of proteins that are involved in electron transport, catalysis or gene expression (Brzoska, Meczynska, & Kruszewski, 2006). A wide range of functions requires their presence in many different cellular compartments. In non-photosynthetic eukaryotes the machinery of ISC assembly for all cellular ISC-proteins involves mitochondrial components (Ye & Rouault, 2010). The synthesis system is oxidation sensitive, for that reason the reduced environment of the MM provides a perfect location. Mitochondria contain ISC assembly machinery similar to the bacterial ISC

system (Chapter 1.1.1). Pyridoxal phosphate-dependent cysteine desulfurase is the sulphur donor for the mitochondrial ISCs. Sulphur is donated via a cysteinepersulphide intermediate. Iron is most likely delivered by a frataxin protein. Sulphurs of the clusters are initially attached to the scaffold protein Isu1 that works together with a number of other enzymes to form ISC in the MM. Then clusters are exported from mitochondria to the cytoplasm (Rouault, 2015) where they are taken over by the cytosolic iron-sulphur protein assembly machinery (Lill et al., 2015) and incorporated into target proteins. Iron sulphur cluster synthesis is one of the most ancient enzymatic pathways of life. Mutations in any of the involved enzymes cause severe, mostly neurological diseases (Rouault, 2015).

## 1.4 Complex I

Complex I is essential for ATP production. This L-shaped enzyme is an assembly of 45 protein subunits with one arm (membrane domain, MD) buried in the inner mitochondrial membrane and the orthogonal arm (peripheral arm, PA) protruding into the mitochondrial matrix. With 14 central and 30 unique accessory subunits complex I is one of the largest membrane bound assemblies in cells. The function of the enzyme is to couple electron transfer from NADH to quinone with translocation of protons across the inner mitochondrial membrane (IMM) (Chapter 1.2.3).

In mitochondria the PA is composed of 7 core subunits (75 kDa, 51 kDa, 49 kDa, 30 kDa, 24 kDa, PSST and TYKY) binding 8 ISCs (Table S6) and 10 supernumerary subunits (18 kDa, 13k Da, 10 kDa, B17.2, B8, 14.5a, 39 kDa, B13, SDAP-α and B14) (Table S7). The MD is composed of 7 core subunits (ND1, ND2, ND3, ND4, ND5, ND4L and ND6) (Table S6) and 21 supernumerary subunits of which 15 contain transmembrane helices (MWFE, B16.6, B9, B14.7, B17, KFYI, ASHI, B14.5b, AGGG, MNLL, SGDH, B15, ESSS, B12), three are associated with the MD form the matrix site (42 kDa, SDAP-β and B22) and four are associated with the MD from the IMS (15 kDa, PGIV, PDSW and B18) (Table S7). Detailed description of the mitochondrial complex I structure can be found in Chapter 4.2.

*1.4.1 Evolution of mammalian complex I*

Complex I is located in the IMM, however the enzyme itself predates the emergence of mitochondria (Moparthi & Hagerhall, 2011). It can be found in bacteria and archaea in a simpler version.

Most likely complex I evolved form already present smaller protein building blocks. In chloroplast, cyanobacteria, archaea and some eubacterial phyla the enzyme contains only 11 subunits (Moparthi & Hagerhall, 2011). It lacks the NADH oxidizing part comprising FMN and the first ISCs. In some archaea the N-module of the complex I homologue is substituted by F420 dehydrogenase subunit. It is hypothesised that the primordial enzyme, and the many 11 subunit versions present, operate without a designated partner interacting with several electron donors. The 14-subunit complex I is the so-called "minimal catalytically active form" of the "traditional" enzyme and is present in eubacteria and α-proteobacteria (Moparthi & Hagerhall, 2011). Understanding complex I evolution can provide key information on its assembly and mechanism. However, tracing the evolution of complex I is a difficult task. Over the years there have been a number of studies investigating this topic but the process still remains very enigmatic (Moparthi & Hagerhall, 2011) (Gabaldon, Rainey, & Huynen, 2005) (Friedrich & Scheide, 2000) (C. Y. Yip, Harbour, Jayawardena, Fearnley, & Sazanov, 2011). Nevertheless, based on available data, we can sketch the most probable order of events.

**Core subunits**

The first progenitor is likely common to all complex I forms and other hydrogenases. In complex I, it gave rise to the Q module. Most likely it was an ancient form of electron transferring hetero dimer that later formed the PSST/49kDa subunit compartment of complex I and soluble [NiFe] hydrogenases (Friedrich & Scheide, 2000). This theory is also supported by the fact that the quinone binding site of complex I corresponds to the NiFe-active site in hydrogenases (Tocilescu, Zickermann, Zwicker, & Brandt, 2010). The next step was the assembly of the PSST/49kDa module with a progenitor for the TYKY subunit (most likely a ferredoxin - like protein) and an additional molecule that did not play a catalytic role but provided stability and a platform for interactions with electron donors (the 30kDa subunit homologue) (Friedrich & Scheide, 2000). All together these peptides formed the Q module.

The following event was acquiring the membrane binding domain (homologues to ND1 and from one to three antiporter - like subunits) which led to gain of ion transferring capabilities and formation of the common ancestor of complex I and membrane bound hydrogenases. Functionally unrelated to any known ancient proteins the ND1 subunit homologue later, when hydrogenase activity was lost, also allowed for quinone accommodation (Efremov & Sazanov, 2012)(Moparthi & Hagerhall, 2011). It connected the ancestral Q module with a $H^+/Na^+$ antiporter protein – Mrp (L. A. Sazanov, 2015). Subunits MrpA (the ND5 subunit homologue), MrpD (the ND4 and ND2 subunits homologue) and MrpC (the ND4L subunit homologue) gave rise to the proton pumping part of complex I (the P module). The origins of the ND3 and ND6 subunits, which compose most of the proximal proton translocation pathway (ND3+ND4L+ND6 (L. A. Sazanov, 2015)) linking the Q and P modules and thus are very important for coupling mechanism, still remain enigmatic (Moparthi & Hagerhall, 2011). These eleven subunits form the most common form of complex I in all organisms (Moparthi & Hagerhall, 2011). It can be found in archaea, cyanobacteria and bacteria and comprises the core of the chloroplast NDH complex. Such arrangement (without the electron "harvesting" module [N module]) suggests that complex I of some bacteria may be able to use different oxidation substrates and collect electrons from sources like ferredoxin, NADH or NADPH (Friedrich, Steinmuller, & Weiss,

1995)(Mi, Endo, Schreiber, & Asada, 1992). In some archaea the function of the NADH binding module is taken by $F_{420}H_2$ hydrogenase, so the complex functions a as $F_{420}$:methanophenazine oxidoreductase (Friedrich & Scheide, 2000).

The N-module (the NADH dehydrogenase) consists of three core subunits. It evolved as the last supplement to complex I and connects the Krebs' Cycle with the aerobic respiration chain (Friedrich & Scheide, 2000). In mammals it is composed of the 24kDa, 51kDa and 75kDa subunits. However in some microaerophilic organisms like *Campylobacter jejuni* the 51kDa and 24 kDa subunits (FP module) homologues are missing and electrons are donated by flavodoxin (Finel, 1998). That suggests maturation of the N-module in two consecutive steps: first to a 12 subunit complex without a permanently bound electron accepting domain and later to the catalytically complete enzyme (Weerakoon & Olson, 2008).

Complex I in most bacterial organisms contains 14 conserved catalytic core subunits (Sharma, Lu, & Bai, 2009). They are organized in one operon of fourteen genes nuoA - N or nqo1 - 14 (nomenclature depending on the name of electron acceptor, respectively ubiquinone or quinone) (Friedrich & Scheide, 2000). The gene order is also very consistent, with the exception of *Rickettsia prowazekii* whose organisation resembles more the one in the mitochondrial genome (Chapter 1.1.1) (Moparthi & Hagerhall, 2011).
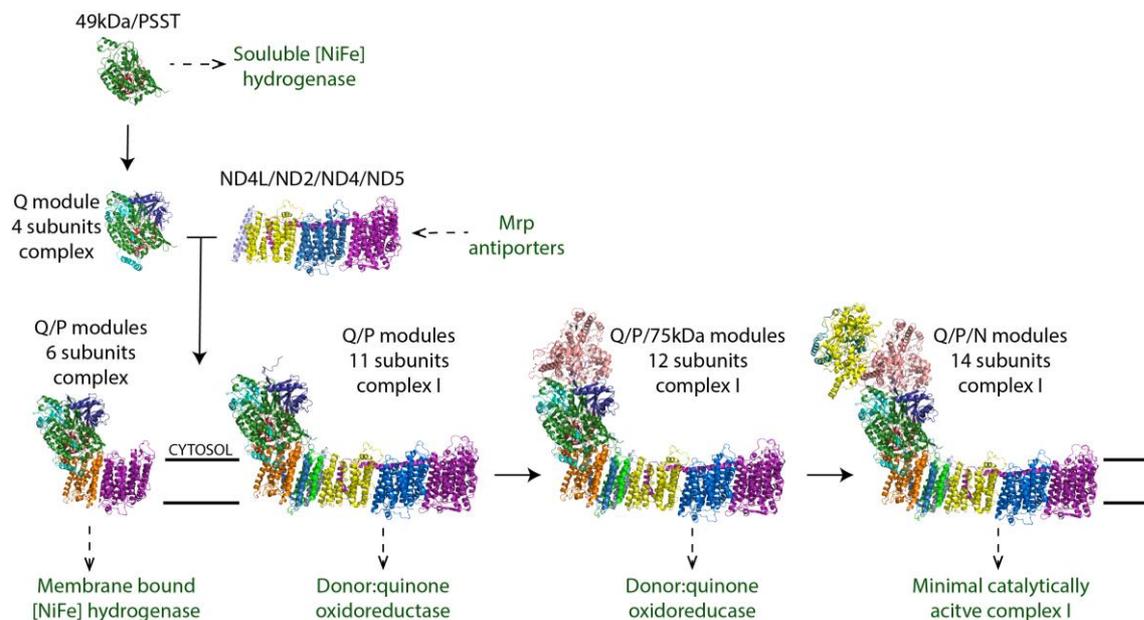
**Figure 1.12.** Hypothetical scheme of modular evolution of the core subunits of complex I. Structures depicted as a cartoon. Approximate lipid bilayer boundaries are indicated. An ancestral hydrogenase made up by the progenitors of the 49kDa (green) and PSST (red) subunits evolved by addition of a ferredoxin (the TYKY subunit, cyan), and a protein of a structural function (the 30kDa subunit, navy).Together with the ND1 homologue protein a progenitor of complex I and the membrane-bound hydrogenases was formed. Addition of a proton transporting protein Mrp (the progenitor of the ND2, ND4 and ND5 subunits, yellow, blue and purple respectively) led to the formation of the common 11 subunits large form of complex I. The catalytically independent complex I emerged by acquisition of the NADH dehydrogenase module in two steps: by first binding the 75kDa subunit (pink) and later the 51kDa/24kDa subunits (FP module, yellow/teal respectively) Detailed description of evolutionary steps can be found in the text. Cartoon models are based on the ovine complex I structure (Fiedorczuk et al., 2016).

**Supernumerary subunits**

Interestingly already in some branches of α-proteobacteria, which are proposed to be mitochondrial ancestors, the presence of supernumerary subunits has been discovered. Complex I of *P. denitrificans* was found to consist of 17 subunits. The 3 supernumerary subunits are homologous to mammalian B17.2, 18 kDa, 13 kDa subunits (C. Y. Yip et al., 2011). This indicates that the endosymbiont acquired by eukaryotic cells may have had a complex I already more developed (possibly more efficient) than most eubacteria (and perhaps the host).

It is believed that eukaryotic complex I originates form an α-proteobacterial ancestor, most likely from the Rickettsia genus (Moparthi & Hagerhall, 2011).

The comparison of the DNA and protein sequences from different organisms shows that in addition to the 14 core subunits of complex I, another 6 proteins (SDAP, 39kDa, 13kDa, B17.2, 18 kDa, 42kDa) were adopted from the genome of the proto-mitochondrial ancestor (Gabaldon et al., 2005). It has to be noted that the presence of the genes for these additional 6 subunits in the genome does not necessarily mean that they were part of complex I. For example, the supernumerary SDAP subunit is a ubiquitous protein in bacteria playing an important role in fatty acid maturation (Byers & Gong, 2007). However, in the mitochondrial complex I, it interacts with the core subunits through the LYR-family proteins (the B14 or B22 subunits) (Angerer et al., 2014), whose orthologues have not been identified in α-proteobacteria (Gabaldon et al., 2005). For that reason, we can assume that orthologues for a number of complex I supernumerary subunits are present in bacteria but have little to do with the enzyme. Nevertheless, after origination of mitochondria, those 20 α-proteobacteria proteins (14 core and 6 accessory) were able to combine with the extra 14 host-originated proteins (B14.7, ESSS, B8, PDSW, B13, PGIV, 15kDa, B22, B16.6, B18, B14, B12, MWFE, NUXM) to compose a34 subunit complex I (20 α-proteobacteria and 14 host-originated proteins). This form is most likely the common ancestor of the complex I enzyme in the three eukaryotic kingdoms (Gabaldon et al., 2005). This shows substantial gain in the complexity of complex I, doubling the number of subunits before the emergence of plants.

After separation of the Plantae but still before the kingdoms of Fungi and Animalia have split, homologues of three extra subunits were added to the complex: B15, ASHI and B9 resulting in a 36 subunit enzyme (Gabaldon et al., 2005). After animals diverged, complex I subunit NUXM (found in Fungi) was lost, but another 8 subunits – AGGG, SGDH, B14.5a, B14.5b, B17, 10 kDa, MNLL and KFYI – were gained building the final giant complex of 44 unique subunits in humans (45 total subunits as SDAP is present in two copies) (Gabaldon et al., 2005).
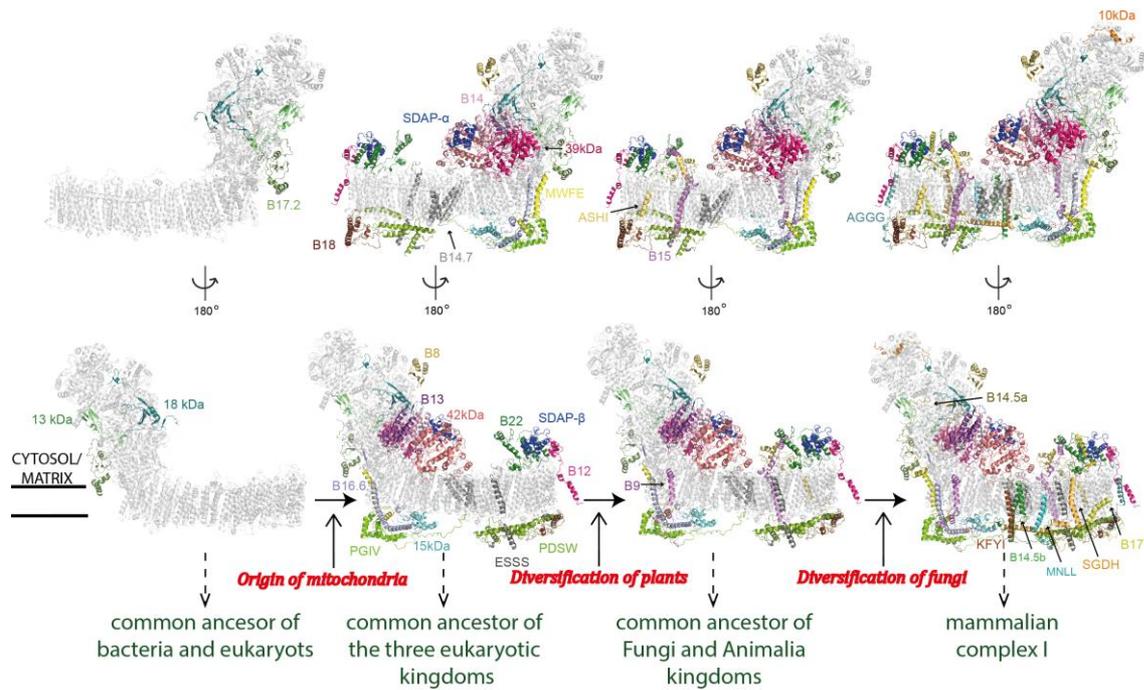
**Figure 1.13** Hypothetical scheme for the evolutionary addition of mitochondrial complex I supernumerary subunits. Structures depicted as a cartoon, with core subunits in grey and transparent and supernumerary subunits coloured and those added at each step labelled. Approximate lipid bilayer boundaries are indicated. Important evolutionary events marked in red. Detailed description of evolutionary steps can be found in the text. Proteins of unknown structure e.g. NUXM not depicted. Cartoon models are based on ovine complex I structure (Fiedorczuk et al., 2016).

This is the most likely order of events that led to development of mammalian mitochondrial complex I. Available data still do not allow us to confidently re-track the order of all the steps and leave many doubts with regard to when exactly each subunit became an integral part of complex I. In most of the cases we can only state when the genes encoding certain protein orthologues appeared. Unfortunately, very little is known about the composition of complex I in non-model species. Most of the data comes from the 10 - 20 most researched organisms. Another elusive aspect of complex I evolution is the development of the complex I assembly factors. Understanding how they evolved can provide an additional key to the map of complex I evolution.

*1.4.2 Assembly of mammalian complex I*

Assembly and maturation of the mammalian complex I is a very complicated and highly regulated process. Although recently progress in understanding of this process has been made, including dynamic complexome profiling performed on human 143B osteosarcoma cells, many steps still remain enigmatic (Guerrero-Castillo et al., 2017; Stroud et al., 2016; Vartak, Deng, Fang, & Bai, 2015).

The assembly of complex I is a stepwise process. Firstly, the functionally independent modules – N, Q and P – are pre-assembled separately and then combined later to form the fully active NADH:Q oxidoreductase (Guerrero-Castillo et al., 2017). Complex I is created from five major assemblies (after Guerrero-Castillo et al.): Q/Pp-a (Quinone/Pumping proximal-a), Pp-b (Pumping proximal-b), Pd-a (Pumping distal-a),Pd-b (Pumping distal-b) and N (NADH oxidising) (Fig. 1.14). Additionally, a cohort of assembly factors transiently associate with specific intermediates to regulate incorporation of the subunits (McKenzie & Ryan, 2010).

**The initial sub-assemblies**

The most primal complex I sub-assemblies are formed before and regardless of the presence of any other subunits (Guerrero-Castillo et al., 2017). That includes:

- Progenitor of the Q module (the 49kDa and 30kDa subunits)
- Progenitor of the Pd' module (the SGDH, PDSW and ESSS subunits)
- N-72kDa module (the 51kDa and 24kDa subunits)
- N-88kDa module (the 75kDa and B8 subunits)

The full Q module is assembled shortly after by adding the 49 kDa, 30 kDa, PSST, TYKY and B13 subunits. At this stage it also contains the assembly factors NDUFAF3 and NDUFAF4 (human nomenclature is used for assembly factors throughout the thesis).

The complete Pd' module is also formed very quickly by addition of the B17 subunit. The presence of Pd' module is very intriguing as all of the subunits building that block belong to complex I accessory proteins. This shows that supernumerary subunits are not necessarily assembled on the core but may create separate sub-assemblies.

At this stage SDAP and B14.7 subunits are also present. A constant presence of the SDAP subunit in mitochondria can be explained by its abundance required for the fatty acids synthesis. Most likely the protein is added to the complex at much later stages. More interesting is the early synthesis of the B14.7 subunit. This trans-membrane (TM) subunit is localized in complex I at the end of amphipathic helix of ND5 locking it to the ND2 subunit (Letts, Fiedorczuk, & Sazanov, 2016)(Fiedorczuk et al., 2016)(J. Zhu, Vinothkumar, & Hirst, 2016). The B14.7 subunits known to be added at the final step of maturation (Guerrero-Castillo et al., 2017), so its presence in the IMM from the very beginning may be a regulatory mechanism (Chapter 4.1.2).

**Formation of the five main assemblies**

Another crucial step during complex I maturation is the formation of the five main assemblies (Fig. 1.14) (Guerrero-Castillo et al., 2017):

- **Pp-b**: formed form the subunits ND2, ND3, ND4L, ND6, KFYI, B14.5b with the assembly factors ACAD9, ECSIT, NDUFAF1 TMEM126B and the putative assembly factors COA1 TMEM186
- **Q/Pp-a**: formed form the Q module, TIMMDC1 assembly factor, and transmembrane subunits: ND1, PGIV, B9 and B16.6
- **N**: formed as a result of joining the N-72kDa and N-88kDa sub-assemblies
- **Pd-a**: formed by adding the MNLL, ND4 subunit and assembly factors TMEM70 and FOXRED1 as well as putative assembly factor ATP5SL to one half of the previously created Pd' module
- **Pd-b**: formed by adding the core ND5 and accessory B18, B12, ASHI, AGGG, SDAP (here called SDAP-β) and B22 subunits to the remaining half of the Pd' module

The Pp-b assembly is formed in two steps: first the ND2, ND4L, KFYI, B14.5b subunits associate with the assembly factors ACAD9, ECSIT, NDUFAF1 and

the putative assembly factor COA1 to create the Pp-b sub-assembly. Shortly thereafter, the remaining proteins ND6, ND3, TMEM 126B, COA1 and TMEM186are added (Guerrero-Castillo et al., 2017). The ND4L subunit is commonly not detected in mass-spectrometry experiments (Letts, Degliesposti, Fiedorczuk, Skehel, & Sazanov, 2016)(Stroud et al., 2016)(Guerrero-Castillo et al., 2017). ND4L is very hydrophobic with no tryptic digestion sites and for that reason it does not give rise to any readily detectable peptides. However, it strongly interacts withND3 and is most likely incorporated into the Pp-b assembly.

Noteworthy is also the fact that the Pd-b assembly migrates in native gels as a 300 kDa assembly (Guerrero-Castillo et al., 2017) although its predicted mass is 154 kDa (Carroll et al., 2006). One of the possible explanations is that it forms a homodimer to stabilize the long amphipathic helix of the ND5 subunit, which otherwise would be just extending into space (Fiedorczuk et al., 2016).

**Joining the main assemblies**

During the next step more global re-arrangements can be observed (Fig. 1.13)(Guerrero-Castillo et al., 2017):

- Pd-a and part of Pd-b assemblies are combined and together with the 42 kDa and B15 subunits form the Pd-680kDa sub-complex
- Q/Pp-a and Pp-b assemblies join and incorporate the B9 and MWFE subunits as well as the NDUFAF2 assembly factor to form the Q/Pp-736kDa sub-complex

**Final maturation of mitochondrial complex I**

The final stage of complex I assembly requires joining of the previously formed modules (Guerrero-Castillo et al., 2017). First combination of the Pd-680kDa, Q/Pp-736 kDa and the remaining pool of the Pd-b assembly form an almost complete complex I termed the Q/P subassembly. Later the N module and the remaining supernumerary subunits (including the B14.7 and SDAP-α) are added. All of the assembly factors (except the COA1 and TMEM186 proteins) leave at this point.
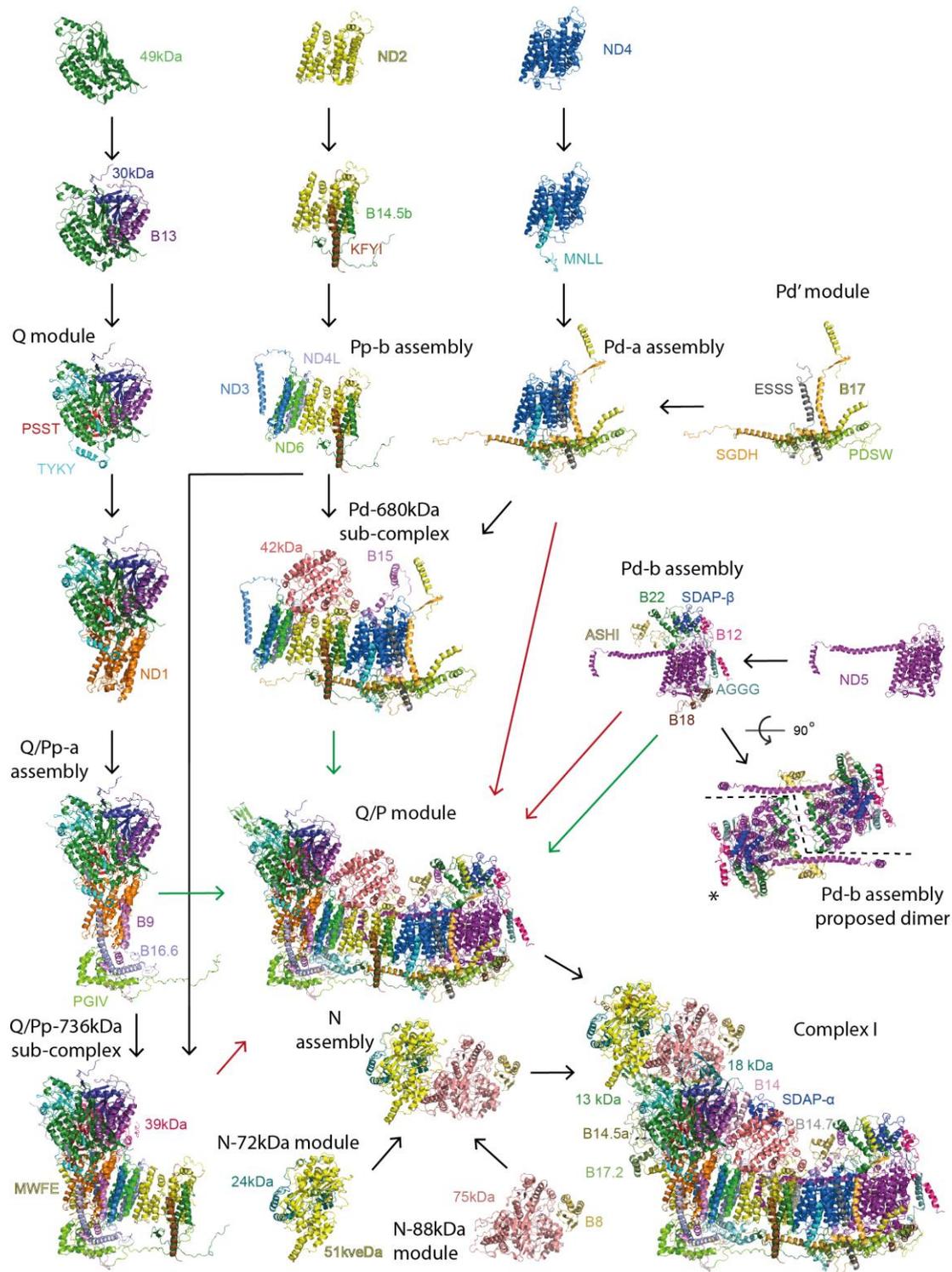
**Figure 1.14** Hypothetical scheme of mitochondrial complex I assembly. Structure depicted as a cartoon, with subunits coloured and labelled. Every important assembly intermediate is labelled and two alternative pathways of Q/P module assembly are marked with red and green arrows. Assembly factors were omitted for clarity. Please note presence of the hypothetical Pd-b assembly dimer; dashed line shows the interaction interface and asterisk marks the added molecule. Cartoon models are based on ovine complex I structure (Fiedorczuk et al., 2016). Figure based on (Guerrero-Castillo et al., 2017)

**Supercomplex formation**

After complex I is assembled, it is additionally stabilised by complexes III and IV, together forming a 1669 kDa supercomplex (the respirasome). In light of recent publications (Guerrero-Castillo et al., 2017; Stroud et al., 2016; Vartak et al., 2015) it is more likely that respirasomes are formed only after the individual complexes are assembled. That rules out the previously suggested theory of supercomplex involvement in the proper complex I maturation (Stroh et al., 2004).

**Complex I recapitulation**

"Ontogeny recapitulates Phylogeny". The theory of recapitulation is a 19th century hypothesis from the field of embryology stating that the development of the embryo of an animal goes through stages resembling sequential phases in the evolution of the animal's ancestors. Currently the theory is largely discredited and treated as a historical side-note (Horder, 2006). Nevertheless, it is interesting to compare complex I evolution and assembly in the context of the "enzymatic recapitulation".

There is no mitochondrial complex I "parallelism" visible. Assembly of individual subunits does not follow the path of evolution with the core assembled first followed by an addition of the supernumerary subunits. However, the modular assembly of the enzyme is preserved from bacteria to mammals. It can be stated that in a way each of the modules evolved separately. Supernumerary subunits are added to the core subunits before the functional modules are joined. This order of events prevents the creation of complex I sub-assemblies of suboptimal activity. Only when all the regulatory and stabilizing supernumerary subunits are present, fragments of the core are assembled and the B14.7 subunit locks the distal pumping modules to the coupling site (Guerrero-Castillo et al., 2017).

*1.4.3 The electron transfer, coupling and proton translocation mechanisms*

**Electron transfer**

Electron transfer in mitochondrial complex I is enabled by 9 co-factors (Chapter 4.2.1) (L. A. Sazanov & Hinchliffe, 2006), a non-covalently bound FMN and 8 ISCs (Table S6). Two of the clusters (N1a and N1b) are binuclear [2Fe-2S] and the remaining six (N2, N3, N4, N5, N6a and N6b) are tetranuclear (Fig. 4.2.B). The FMN molecule is situated at the top of the 51 kDa subunit in a pocket formed by a Rossman fold. ISCs form a path along the PA linking the FMN and the ubiquinone binding chamber. Most of the clusters are co-ordinated by four cysteine residues except for the N5 cluster which is ligated by three cysteines and a histidine (Chapter 4.1.1) (L. A. Sazanov & Hinchliffe, 2006).

NADH donates two electrons simultaneously to FMN, however ISCs can transfer only one electron at a time. For that reason, FMN plays the role of an intermediate electron carrier. It can accept two electrons form NADH, transfer one immediately to the first FeS cluster (N3) and remain relatively stable in the form of a flavosemiquinone until the N3 cluster can accept the second electron. Semi-reduced flavin is very reactive and for that reason the FMN site of complex I is the main source of ROS in mitochondria. It has been proposed that to prevent excessive ROS production the N1a cluster accepts the electron from the flavosemiquinone until the N3 cluster is available for reduction (Leonid A Sazanov, 2007). However, the function of the N1a cluster is still debated and it may play only a structural role (Chapter 4.1.1) (Birrell, Morina, Bridges, Friedrich, & Hirst, 2013). Electrons from the N3 cluster are transported through the chain of the ISCs to ubiquinone. All the clusters are separated by <14 Å, which allows for electron tunnelling on a physiological time-scale (Page, Moser, Chen, & Dutton, 1999). The N2 cluster is the last "stepping stone" for the electron on the way through the PA to ubiquinone (T. Ohnishi, 1998). The force driving transfer of electrons form NADH to ubiquinone is the difference in the midpoint redox potentials ($E_m$; the standard electric potential at which 50% of a redox component is in the reduced state) of the system components. At a pH of 7.0 the $E_m$ of NADH is -320 mV; FMN is −340 mV; the N2 cluster is -100mV; and ubiquinone is+110 mV. The potential of the other (EPR-visible) FeS

clusters oscillates at around -250 mV (Yakovlev, Reda, & Hirst, 2007). More detailed description of the electron transfer path in mammalian complex I can be found in Chapter 4.1.1.

**Proton translocation**

For one NADH molecule oxidized, complex I translocates 4 protons across the IMM (Wikstrom, 1984). This stoichiometry is widely accepted although differences cannot be ruled out between organisms (Chapter 1.4.1). It has also been argued that complex I of some bacteria pump sodium ions instead of protons (Steuber, Schmid, Rufibach, & Dimroth, 2000)(Gemperli, Dimroth, & Steuber, 2002), however, these proposals remain very controversial (Bertsova & Bogachev, 2004). Proton translocation is catalysed by the seven core subunits (Chapter 1.4) that evolved from the $H^+/Na^+$ antiporters (Chapter 1.4.1). On the basis of homology, subunits ND2, ND4 and ND5 were identified as proton translocators. The location of the fourth–most proximal to the Q-site– proton path has been proposed based on the structure of the intact complex I form *T. thermophilus* (R. Baradaran, Berrisford, Minhas, & Sazanov, 2013). More detailed description of the proton transfer paths in mammalian complex I can be found in the Chapter 4.2.1.

**The coupling mechanism**

Whilst the mechanisms by which electron transfer from NADH to Q in the PA and protons are transported through the IMM are relatively well understood (L. A. Sazanov, 2015), coupling of those two processes is still very enigmatic.

In the recent years a number of different models have been proposed on the basis of structural and functional data. Some of them, like that proposed by Ohnishi and co-workers (S. T. Ohnishi, Salerno, & Ohnishi, 2010) which rely on the presence of two bound Q molecules have been ruled out by the recent structural data (R. Baradaran et al., 2013)(Vinothkumar, Zhu, & Hirst, 2014)(Zickermann et al., 2015). The second reduction site for ubiquinone has never been observed. Another proposal in which the lateral helix (HL) ofND5 acts as a mechanical transmission element, providing long distance energy transfer analogous to the coupling rod of a steam engine (Efremov, Baradaran,

& Sazanov, 2010), is at odds with some genetic and biochemical data. HL does not show a clear pattern of conserved residues, mutating potential contact site residues, introducing insertions to reduce helix rigidity or cross links to immobilize it did not decrease complex I proton pumping significantly (S. Zhu & Vik, 2015)(G. Belevich, Knuuti, Verkhovsky, Wikstrom, & Verkhovskaya, 2011). Some substitutions resulted in an unstable enzyme that fell apart during purification suggesting that HL might clamp the MD together (G. Belevich et al., 2011) similar to the β-hairpin elements of the bacterial enzyme (R. Baradaran et al., 2013) or the supernumerary subunits of the IMS in the mammalian enzyme (Fiedorczuk et al., 2016). However, other mutagenesis experiments (Nakamaru-Ogiso et al., 2010) provided more controversial results and at least a partial coupling role for HL cannot be ruled out yet.

Current models of complex I coupling are based on the presence of a single quinone molecule and transmission of conformational changes from the Q site to the proton pumping modules directly via the functional residues of the proton translocation apparatus. These models can be divided into so-called two-stroke, wave-spring and one stroke models.

In the two-stroke model there are two discrete translocation events. First, pumping two protons is driven by formation of anionic $Q^{\bullet-}$. The other two protons are translocated upon full reduction of $Q^{\bullet-}$ to $QH_2$ (U. Brandt, 2011). However, there are some issues with this model. It postulates that only two translocation channels are functional, which is inconsistent with structural and conservation data indicating four channels. It also requires the first proton translocation event to be faster than the delay between first and second quinone reduction events and stands against the experimentally measured redox potentials of the Q and ISCs. The predicted potential of the $Q/Q^{\bullet-}$ is between Em -330 and -240 mV (Moser et al., 2006)(Efremov & Sazanov, 2012). Taking into account that the $E_m$ for the N2 cluster is -100mV,indicates that there is no overall drop of energy for electron transfer until the quinone is fully reduced.

The wave-spring model suggest that a key role is played by the MD axis of charged residues and solvating water molecules that was observed in the high resolution crystal structures of bacterial complex I (Efremov & Sazanov, 2011)(R. Baradaran et al., 2013). This model proposes that in the oxidised state conserved Lys206[ND4] is protonated. Upon reduction of Q (to $Q^{2-}$), the negative charge localized on the ligand's head group repulses nearby amino acid residues that eventually translates to a shift in pKa of Lys105[ND2]. Then Lys105[ND2]collects a proton form matrix and the charge wave continuous along the MD axis driving release of the Lys206[ND4] proton and causing protonation of Lys223[ND5]. The cycle ends with protonation and disassociation of ubiquinol-10 ($QH_2$)that triggers proton release form the ND5 and ND2 modules and proton uptake by the ND4 module (Verkhovskaya & Bloch, 2013). However, to achieve the directionality of the proton transport this model requires exquisite fine-tuning of the pKas of all the key residues, which seems difficult to achieve in a flexible protein environment. It also does not assign any role to the many conserved discontinuous helices observed in the structure in key positions.

According to the one stroke model proposed by Sazanov and colleagues (R. Baradaran et al., 2013) all four protons are translocated as a result of long-range conformational changes in the MD. One proton is released per ND5, 4 and 2 subunit (Efremov et al., 2010) and the fourth proton through the ND1+ND4L+ND6+ND3 subunits translocation pathway (R. Baradaran et al., 2013). The central hydrophilic axis of the MD has a key role in the proton pumping and coupling mechanism. The process of proton translocation is initiated by the Q-chemistry and conformational changes nearby (Fig. 4.3.B), which propagate via the central axis (using all the flexible discontinuous helices) towards the three antiporter-like subunits, resulting in pKa and accessibility changes of the key protonatable residues. HL and IMS supernumerary subunits (βH motif in bacteria) are the most likely structural elements (however the role of HL is still not entirely clear) (Chapter 4.6).

## 1.5 A brief history of structural studies on the respiratory complex I

Complex I was isolated and shown to have DPNH (an obsolete name for NADH) - coenzyme Q oxidoreductase activity for the first time in 1962 (Hatefi, Haavik, & Griffiths, 1962). That same year mitochondrial complexes I-IV got their names and have since been studied as separate entities (Hatefi, Haavik, Fowler, & Griffiths, 1962). In 1973, it was shown that in addition to the oxidoreductase activity, complex I is also a proton pump (Ragan & Racker, 1973). This was one of the milestone findings in the field of mitochondrial bioenergetics. It supported Peter Michel's chemiosmotic theory which six years later won him the Nobel Prize (Mitchell, 1979). Over the next decades developments in the purification methods, mass spectrometry and x-ray crystallography contributed to the gradual understanding of the structural basis of enzymes involved in ATP production (Saraste, 1999)(Kuhlbrandt, 2015). From the mid 90s until 2010 atomic structures of all mitochondrial OXPHOS enzymes, except complex I became available (Sun et al., 2005)(Xia et al., 1997)(Tsukihara et al., 1996)(Abrahams, Leslie, Lutter, & Walker, 1994)(Watt, Montgomery, Runswick, Leslie, & Walker, 2010).

Over forty years of studies on complex I resulted in defining the protein and cofactor composition of the enzyme across many domains of life (Meyer, 2012)(Bridges, Fearnley, & Hirst, 2010)(Abdrakhmanova et al., 2004)(Carroll, Fearnley, Shannon, Hirst, & Walker, 2003)(C.-Y. Yip, Fearnley, Harbour, Jayawardena, & Sazanov, 2010)(L. A. Sazanov, S. Y. Peak-Chew, I. M. Fearnley, & J. E. Walker, 2000)(T. Ohnishi, 1998).

The first breakthrough in structural information was the x-ray structure of the hydrophilic arm of complex I from *T. thermophilus* (L. A. Sazanov & Hinchliffe, 2006), which revealed the electron transfer pathway through the enzyme, the arrangement of seven core subunits, all of the redox cofactors, part of the Q binding cavity and non-covalently bound FMN located in a Rossmann fold (Rossmann, Moras, & Olsen, 1974) of the 51 kDa (Nqo1) subunit. Next came the structure of the membrane arm of *E. coli* complex I (Efremov & Sazanov, 2011), revealing the organisation of proton translocation pathways. The MD

structure contained six core subunits of the complex, composed of 55 TMHs forming three proton-pumping modules. It contained the characteristic long lateral helix of the ND5 (Nuo12) subunit and conserved key discontinuous TMHs in each module. This structure allowed the tracing of the proton translocation paths and identified key charged residues in the middle of the membrane. Although the partial bacterial structures provided a lot of information regarding the NADH oxidation mechanism, the electron transfer path and architecture of the proton pumping modules, organization of the quinone-binding site at the junction of the two arms remained elusive.

The next breakthrough came in 2013 when the first complete atomic structure of complex I from *T. thermophilus* was published (R. Baradaran et al., 2013). It provided insight into the quinone reduction site and provided crucial information about the possible mechanism of coupling between electron transfer and proton translocation in complex I, revealing quinone chemistry as a probable key to this coupling. Due to the high conservation of the core subunits across the species it was also an essential model for studying the eukaryotic enzyme. The *T. thermophilus* structure was followed by low resolution x-ray structures of the eukaryotic enzyme form yeast *Yarrowia lipolytica* (Zickermann et al., 2015) and Iβ subcomplex of *Bos taurus* (J. Zhu et al., 2015). Those structures did not provide much information regarding the mechanism of complex I or function of the supernumerary subunits due to their low resolution, however they gave valuable insight into the enzyme's overall architecture. The architecture of yeast enzyme revealed that the core subunits are very similar to *T. thermophilus*, allowing their modelling on the basis of homology, while supernumerary subunits form a shell around the core.

Meanwhile new developments in cryo-electron microscopy (Chapter 1.6)(Kuhlbrandt, 2014) gave the ability to solve structures of biological complexes in atomic detail. With the new direct electron detectors and better software it became possible to obtain high-resolution maps of difficult to crystalize biological assemblies like complex I. The first 5 Å resolution map of the bovine enzyme was published in 2014 (Vinothkumar et al., 2014). It presented the complete architecture of the complex, however moderate

resolution prevented modelling of the side chains and the information about supernumerary subunits was very limited. Finally last year, almost complete cryo-EM structures of mitochondrial complex I became available (Chapter 1.7)

## 1.6 Electron microscopy

In the early 20th century it was observed that an electron beam in vacuum behaves very similarly to light. Additionally, electric and magnetic fields can bend and focus electrons much like glass lenses shape visible light (Egerton, 2007). This finding inspired student Ernst Ruska and his supervisor Max Knoll to construct the first electron microscope. The apparatus constructed in April 1931 was modestly called a cathode ray oscillograph and had a total magnification (M) of 14.4X (Ruska, 1987). Very interestingly, at that time neither Ruska nor Knoll knew about the wave nature of electrons postulated by de Broglie (although it was put forward already in 1925). Thus they thought that the resolution of the electron microscope was limited by the size of the electron. It was very disappointing for both of them to learn in summer 1931 that, like for light microscopes, the wavelength limits the theoretical resolution of the device (Ruska, 1987). Fortunately, the theoretical resolution limit of the electron microscope is much higher than light microscopes. Already a few years later in 1934, Ruska and Knoll constructed a microscope which obtained a resolution of 100 nm, twice as good as that of any light microscope (Ruska, 1987). Modern electron microscopes allow reaching resolutions higher than 0.05 nm (Haider, Rose, Uhlemann, Kabius, & Urban, 1998).

**Figure 1.15 A.** Sketch of the cathode ray tube for testing the image properties of the non-uniform magnetic field of a short coil (Ruska 1929) **B.** Bacterial culture fixed with formalin and embedded in a supporting film stained with heavy metal salt (Krause 1937). Images adapted from Ruska Nobel Prize Lecture (Ruska, 1987).

Currently electron microscopes are used around the world in many different research fields. They have applications in aeronautics, automotive manufacturing, clothing and apparel, machining, pharmacology, criminology and many more. Because they are used for different purposes special types of EMs have been developed. In structural biology the transmission electron microscope (TEM) is mostly used but there are also scanning electron microscopes (SEM), scanning transmission electron microscopes (STEM, combining principles of TEM and SEM) and focused ion beam microscopes (FIB, similar to SEM but using beam of ions instead of electrons).

Modern TEM comprise an electron optical column, lenses for focusing and deflecting the beam, the high voltage generator for the electron source, a vacuum system, operating console, control panels, electron detector and control software (Egerton, 2007). Electrons are produced by the source, focused and pass through a specimen. Projection of the specimen is displayed on a viewing device on the bottom of the column.

*1.6.1 Single particle cryo - EM*

In EM studies of inorganic materials resolutions better than 1Å have been achieved routinely for the last few decades (Haider et al., 1998). However, this requires high electron doses that cause damage to organic specimens. In the study of the structure of ovine mitochondrial complex I, a technique known as cryo-electron microscopy (cryo-EM) was used (Cheng, Grigorieff, Penczek, & Walz, 2015). This method has been shown to be very powerful for three-dimensional structure determination of biological molecules. Preparation of the microscope sample requires spreading the protein suspension on an EM grid and freezing it rapidly in ethane at near liquid nitrogen temperature (appropriate for vitreous ice formation). This approach, developed in the 1980s by Jacques Dubochet and his colleagues, allows for a great reduction in radiation damage (Adrian, Dubochet, Lepault, & McDowall, 1984). To further minimize the damage, images are recorded at very low electron dose that in turn also decreases signal to noise ratio (S/N) in the micrographs. Low S/N and damage by the electron beam require averaging signal from multiple molecular images to obtain the cryo-EM map. For that reason the first high resolution (R<4Å) structures were obtained by recording images from 2D crystals and averaging over many unit cells (Henderson et al., 1990). The obstacle of the necessity for sample crystallization was overcome by the development of more powerful computer programs. In the first decade of the 21st century the first maps from single particles of icosahedral viruses and helically ordered proteins embedded in ice reached resolutions good enough to visualize side chains. However, averaging of images of millions of asymmetric units was required, similar to what was needed for 2D crystals (Grigorieff & Harrison, 2011).

Efforts for EM structure determination from single particles of asymmetric proteins started in the 1970s (Hoppe, Schramm, Sturm, Hunsmann, & Gassmann, 1976). Statistical methods allowed proper 2D averaging and classification and ten years later introduction of angular reconstruction methods enabled transition from 2D particle averages to 3D maps for objects without symmetry (Radermacher, Wagenknecht, Verschoor, & Frank, 1987)(van Heel et al., 2000)(Vinothkumar &Henderson, 2016). The next few decades brought

further development of data processing software as well as EM hardware (Vinothkumar & Henderson, 2016):

- The first electronic detectors improved the speed of data collection and sample examination, although the detective quantum efficiency [DQE] was still worse compared to film when used with high voltage. High DQE (ideally 1.0) is essential in maximising the S/N ratio.
- Field emission guns (FEGs), which provided a 500-fold increase in coherence comparing to previously used tungsten filaments.
- Improvements in both column and cold stage stability reduced image drift.
- Reduction in ice contamination through better vacuum and cryo-shielding.
- Increased electron accelerating voltage from 100 to 300 keV, improving the microscope resolution, depth-of-focus, aberrations, and beam penetration.

In spite of these developments resolution achieved by single particle cryo-EM of asymmetric particles oscillated around 10Å. The breakthrough was in 2013 when new detectors and computer programs became available (Vinothkumar & Henderson, 2016).

*1.6.2 Recent developments in electron microscopy*

Recently introduced direct electron detectors are built based on complementary metal oxide semiconductor (CMOS) technology (Milazzo et al., 2005). They have a rolling-shutter readout mechanism and can read out images continuously with a frame rate that can range from 1 to 1000 Hz or more. That allows recording information content in movie mode. Data collected in the form of movies is very useful during the processing. Firstly, micrographs can be corrected for beam-induced specimen charging and movement by alignment and weighting of the individual frames. Secondly frames with low S/N, which negatively contribute to the final map, can be discarded. Additionally, the rolling shutter mechanism enables identification of individual electron events and replacement of their signal by a delta function at the centroid of the event

density at sub-pixel resolution, a practice called electron counting (EC). EC is used for recording images at higher resolution than might normally be achievable and subsequently performing binning, which allows higher DQE over broader resolution ranges. The new generation of CMOS detectors are also very thin (by backthinning after wafer fabrication). This reduces the noise in the image by preventing electrons from backscattering and depositing energy in incorrect pixels. Additionally, backthinning improves the detector's lifetime due to lowering the overall energy deposition. All those improvements contributed to EM imaging systems with DQE much higher than ever before (Vinothkumar & Henderson, 2016).

The second great boost in single particle cryo-EM was the development of processing software. Although there are a number of programs (Vinothkumar & Henderson, 2016) for use at different steps of data processing  or providing the whole working pipeline, the most popular currently, also used in this study, is RELION (Scheres, 2012). RELION introduced a major improvement in map reconstruction. To determine the orientations of the individual single particle images in 2D classification and to calculate the 3D structures it uses a regularised maximum likelihood (ML) scoring system, rather than a simple cross-correlation maximisation algorithm. ML-scoring itself was not a real breakthrough, as it was originally implemented in Xmipp (Marabini et al., 1996) but RELION uses also regularisation calculated through the empirical Bayesian approach from the data themselves, which leads to optimal weights on the reconstruction and on the alignments.

Summarizing, single particle cryo-EM can currently be used to analyse protein or nucleic acid structures of sufficient size (which is getting ever smaller). It does not require sample crystallization and is able to deal with sample heterogeneity by 3D classification *in silico*. However, it still has not reached its full potential and during the next few years it may become the leading method in structural biology.

## 1.7 Structural studies on the respiratory complex I using electron microscopy

The first electron microscopy work on complex I dates back to the late 1980s. The first EM structure of complex I from negatively stained 2D crystals (Leonard, Haiker, & Weiss, 1987) provided information about enzyme topology and reported it to have an 'axe' shape. The first report of the L-shape of the molecule with one arm embedded in the lipid membrane comes from the reconstruction generated form 2D crystals of the *N. crassa* enzyme in the early 1990s (Hofhaus, Weiss, & Leonard, 1991). The first single particle study of negatively stained complex at 35 Å resolution (Guenebaut, Vincentelli, Mills, Weiss, & Leonard, 1997) confirmed the overall architecture of the enzyme. The characteristic, conserved L-shape of the molecule was further confirmed in a number of studies on enzymes from *E. coli* (Morgan & Sazanov, 2008), *T. thermophilus* (Rozbeh Baradaran, 2009), *A. aeolicus (G. Peng et al., 2003), P. denitfricans* (C.-Y. Yip et al., 2010), *Y. lipolytica* (Djafarzadeh et al., 2000), *A. thaliana* (Dudkina, Kudryashev, Stahlberg, & Boekema, 2011) and *B. Taurus* (Grigorieff, 1998). Although the resolution of these reconstructions varied mostly between 20 and 40 Å, they provided valuable data on arrangement of subunits and gave the first hints about possible conformational changes during catalytic turnover of the enzyme (Morgan & Sazanov, 2008)(Böttcher, Scheide, Hesterberg, Nagel-Steger, & Friedrich, 2002).

The location of the 49 kDa subunit was proposed by immune-staining experiments (Guenebaut et al., 1997). The ND5 and ND4 subunits were assigned using single particle EM by Sazanov and colleagues (Baranova, Holt, & Sazanov, 2007) and 2D electron crystallography studies by the same group finally resolved the MD of *E. coli* to 8 Å resolution revealing 60 trans-membrane helices with homology to the Mrp family of cation/H$^+$ antiporters and indicated that the coupling of electron transfer to proton pumping may involve long-range conformational changes.
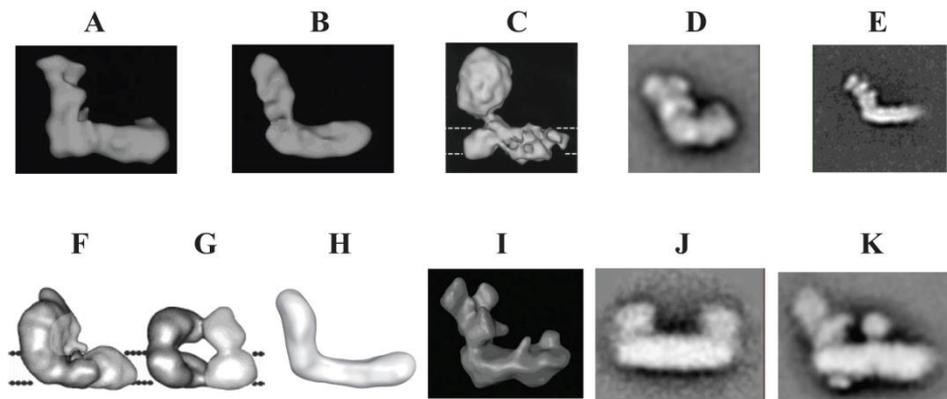
**Figure 1.16** Electron microscopy structures of complex I. Single particle analysis and 3D reconstructions of complex I from: **A.** *N. crassa* (Guenebaut et al., 1997); **B.** *E. coli* (Guenebaut, Schlitt, Weiss, Leonard, & Friedrich, 1998); **C.** Bovine (Grigorieff, 1998); **D.** Bovine (Morgan & Sazanov, 2008); **E.** *A. aeolicus* (L. Peng, Yamamoto, & Shikanai, 2010); **F.** and **G.** *E. coli* inactive and active forms, respectively (Böttcher et al., 2002); **H.** *E. coli* incubated with NAD$^+$ (Morgan & Sazanov, 2008); **I.** *Y. lipolytica* (Radermacher et al., 2006); **J.** *T. elongates* (Arteni et al., 2006); **K.** *A. thaliana* (Dudkina, Heinemeyer, Sunderhaus, Boekema, & Braun, 2006). Figure adapted from that created by Dr Rozbeh Baradaran.

The first high-resolution cryo-EM structure of the bovine complex I was published in 2014 when this project had already started (Vinothkumar et al., 2014). The 5 Å resolution map provided great insights into arrangement and structure of the supernumerary subunits and confirmed the high similarity of the core subunits to bacterial models. However, it was a poly-alanine model that did not provide detailed information about the complex I structure. Finally last year, a partial model (65% atomic for core subunits and 27% atomic for supernumerary subunits) was published for the bovine enzyme at 4.2-4.3 Å resolution (J. Zhu et al., 2016), just before the publication of the first nearly complete (88% atomic) ovine model at 3.9 Å from this work (Fiedorczuk et al., 2016). These publications were followed later by the structure of the porcine enzyme at ~ 3.6 Å (~ 84% atomic)in the supercomplex (Wu, Gu, Guo, Huang, & Yang, 2016).

## 1.8 Aims of this work

This project aimed to generate a complete and accurate atomic model of the mammalian mitochondrial complex I.

First, three easily available model organisms (*B. taurus*, *S. scrofa* and *O. aries*) were evaluated in order to choose the best source for the enzyme, deemed to be ovine due to its high stability and activity. Human complex I for the study would have been more appropriate, however culturing enough immortalized cells is very expensive/time consuming and fresh placentas are difficult to obtain and cause many issues.

Second, a purification protocol was developed to obtain pure and stable sample of the enzyme that can be used for structural experiments.

For many years structural biology has been dominated by X-ray crystallography. Recent developments in cryo-EM have led to a number of near atomic resolution structures for protein complexes difficult to crystallize (Vinothkumar & Henderson, 2016). For that reason, cryo-EM was the first choice technique for this project. Within time constrains of the PhD project (3-4 years) it was possible to optimise cryo-EM grid preparation (one of the most challenging steps in the process), collect and process the data, and obtain the map at a resolution of ~3.9 Å that enabled model building at the atomic level (another challenging part of the project due to the size of the complex).

The first nearly completely atomic model of mitochondrial complex I is an important step forward in our understanding of its function. It provided crucial information about the organisation and function of the supernumerary subunits. Additionally, it shed light on complex I assembly and the molecular basis of disease-causing mutations. The process of coupling electron transfer to proton translocation still remains inscrutable, however, a range of different conformations observed in this work provides possible clues.

# Chapter 2. Materials and Methods

## 2.1 Materials

N-dodecyl-β-maltoside (DDM) and lauryl maltose neopentyl glycol (LMNG) were purchased from Glycon Biochemicals GmbH (Luckenwalde, Brandenburg, Germany); all other detergents were acquired from Anatrace (Maumee, Ohio, USA). NADH was purchased from Santa Cruz Biotechnology Inc. (Heidelberg, Germany). Complete Protease Inhibitor Cocktail tablets (EDTA free) were bought from Roche Diagnostics (Burgess Hill, West Sussex, UK). All other chemicals were purchased from Fisher Scientific UK (Loughborough, Leicestershire, UK), VWR International (Lutterworth, Leicestershire, UK) and Sigma-Aldrich Ltd (Gillingham, Dorset, UK) and were of analytical grade. All solutions, unless otherwise stated, were prepared with ultrapure water filtered by a Milli-Q system (Merck Millipore, Billerica, Massachusetts, USA). Chromatography columns were from GE Healthcare (Amersham, Buckinghamshire, UK). Vivaspin concentrators were purchased from Sartorius Stedim (Epsom, Surrey, UK). Amicon-Ultra centrifugal filter units were supplied by Merck Millipore (Billerica, Massachusetts, USA). Zeba spin desalting columns were acquired from Thermo Fisher Scientific Inc. (Rockford, Illinois, USA). Bio-Beads SM-2 were purchased from BioRad Laboratories Ltd. (Hemel Hempstead, Hertfordshire, UK). Holey carbon-coated grids were supplied by Quantifoil Micro Tools GmbH (Jena, Thuringia, Germany) and Protochips, Inc. (Raleigh, North Carolina, USA). Ovine hearts were purchased from local butcher (Humphreys C & Sons, UK).

## 2.2 Analytical methods

### 2.2.1 Complex I activity assays

All measurements were made using a Shimadzu UV-1601 (Shimadzu UK Ltd., Milton Keynes, Buckinghamshire, UK) spectrophotometer fitted with a peltier temperature controller and automated sample changer (Shimadzu DPS-240A) with magnetic stirring unit (Electronic Stirrer Model 300, Rank Brothers Ltd. Cambridge, UK). The oxidation of NADH was measured by monitoring the change in absorbance at a wavelength of 340 nm. Change in [NADH] was then calculated using the Beer-Lambert law and the appropriate molar extinction coefficient ($\varepsilon 340$ [NADH] = 6220 $M^{-1}$ $cm^{-1}$). Complex I activity is defined as µmol NADH oxidised per min per mg of protein. All reactions were performed at 37 °C.

Complex I dehydrogenase activity was measured using potassium ferricyancide ($K_3[Fe(CN)_6]$; FeCN) as an electron acceptor. Assay buffer containing 50 mM HEPES pH 7.4, 0.1 % (w/v) LMNG and 1 mM FeCN was pre-warmed in 1.5 mL semi-micro disposable cuvettes. Samples containing complex I were added to a final volume of 1 ml. Reactions were then initiated by the addition of 100 µM NADH and the absorbance at 340 nm was monitored for 30 s. If required, the amount of complex I per assay was adjusted to give a linear rate of NADH oxidation over this time period.

Complex I reductase activity was measured using a soluble quinone analogue, decylubiquinone (DQ), as an electron acceptor. Protein was added to reaction buffer containing 20 mM HEPES, pH 7.4, 100 mM NaCl, 10% glycerol, 1.0 mg/ml BSA, 0.1% CHAPS and 0.25 mg/ml bovine heart total lipid extract in 4 ml cuvette. That was followed by addition of 100 µM DQ and incubated in 30 °C for 10 min. Reactions were then initiated by the addition of 100 µM NADH and the absorbance at 340 nm was monitored for 60 s. Sample during incubation and data collection was stirred at ~300 rpm.

*2.2.2 Methods to determine protein concentration*

**BCA assay**

Protein concentration was measured using the Pierce bicinchoninic acid (BCA) assay kit (Thermo Fisher Scientific Inc., Rockford, Illinois, USA) used according to the manufacturer's instructions. A bovine serum albumin (BSA) standard curve (0.1 to 1 mg/ml) was prepared in a buffer matching that of the unknown sample. Colorimetric reactions were performed in Falcon rigid 96 well plates, incubated at 37 °C for 30 min and then read using a SpectraMax M2 plate reader (Molecular Devices, Wokingham, Berkshire, UK). The sample protein concentration was then estimated by comparison with the standard curve.

**Concentration based on theNADH:$K_3$[Fe(CN)$_6$] activity**

Intact *O. aries* complex I was determined to have an NADH:FeCN oxidoreductase activity at ~80 µmol/min/mg (Letts, Degliesposti, et al., 2016). This specific activity value was used to convert the measured activity of a sample (in µmol/min/ml) to protein concentration value in mg/ml for rough estimates of the enzyme concentration during purification.

**A280 absorbance**

Protein absorption at 280 nm was measured to estimate protein concentration of the samples for the cryo-EM grids preparation. This method was used because of its good reproducibility for the protein concentration at the range of 1-5 mg/ml and very small volume of sample required for the measurements (~2 µl). Absorption at 280 nm was measured using a NanoDrop 1000 UV-Vis spectrophotometer (Thermo Fisher Scientific Inc., Rockford, Illinois, USA). The concentration of complex I was calibrated to about 1,65 mg/ml of protein concentration per 1 A280 unit.

*2.2.3 SDS-PAGE*

12- and 15-well pre-cast Novex Tris-glycine 4% to 20 % (w/v) polyacrylamide gradient gels (Life Technologies, Paisley, Renfrewshire, UK) were run according to the manufacturer's instructions. The electrophoresis buffer contained 0.25 M Tris base pH 8.3, 1.92 M glycine and 0.1% (w/v) sodium

46

dodecyl sulphate (SDS). Samples were mixed 4:1 with solubilising buffer comprising 250 mM Tis pH 6.8, 10% (w/v) SDS, 10% (w/v) dithiothreitol (DTT) 50% (v/v) glycerol and 0.05% (w/v) bromophenol blue. 10-15 µl of sample was loaded per well; one lane on each gel was loaded with 6 µl of Precision Plus Protein Standards, unstained (Bio-Rad), gels were then run for 85 min at 150 V. Gels were stained overnight with Instant Blue stain (Expedeon Ltd., Harston, Cambridgeshire, UK) and destained with water the following day.

## 2.3 Sample preparation

### 2.3.1 Preparation of mitochondria
Mitochondrial purification was conducted following previously described protocols (Letts, Degliesposti, et al., 2016). Ovine hearts (36, ~8 kg) are cleaned of fat and connective tissue, then chopped and minced. 1400 ml of isotonic buffer (250mM sucrose, 10mM Tris- HCl pH 7.8, 5 mM 2-mercaptoethanol) is added per 1 kg of minced heart tissue for washing. Buffer is discarded by squeezing the meat through a double-layered muslin and added again with minced tissue to a blender. Material is blended for 30 s. Mitochondria are obtained during two steps of centrifugation. Firstly (H12000 rotor, 4 °C, 2700 rpm, 15 min.), tissue debris, unbroken cells and bigger cellular structures are removed. Secondly, by pelleting (SLA-3000 rotor, 4 °C, 11k rpm 27 min.), mitochondria are separated from soluble proteins and smaller organelles. Mitochondria prepared with this method are portioned (10-20 g), and stored frozen at -80 °C.

### 2.3.2 Preparation of mitochondrial membranes
Mitochondrial membranes are obtained by homogenisation in glass- Teflon homogenizer. 20 g of mitochondria are homogenized in water (100 ml) before adding potassium chloride to 150 mM. Soluble proteins are separated from the membrane fraction, containing complex I, by centrifugation (SS-34 rotor, 20k rpm, 25 min.). After this step membranes can be frozen and stored in –80 °C.

*2.3.3 Complex I preparation*

Protein was purified from *Ovis aries* heart mitochondria following the protocol developed in this work and adapted with some modifications from previously published procedure for bovine enzyme (Letts, Degliesposti, et al., 2016)(L A Sazanov, S Y Peak-Chew, I M Fearnley, & J E Walker, 2000).

**Initial Purification**

Collected membranes were solubilised by incubation in a solubilisation buffer in the cold room (20 mM Tris pH 7.4, 1 mM ethylenediaminetetraacetic acid (EDTA), 10% glycerol, 2 mM dithiothreitol (DTT), 0.002% phenylmethanesulfonyl fluoride (PMSF), 1% n-Dodecyl $\beta$-D-maltoside (DDM). Per 1 g of membrane pellet 8.5 ml of solubilisation buffer was added and stirred for 30 min at 4 °C. Any non-solubilised mitochondrial fraction was removed by centrifugation (SS-34 rotor, 20k rpm, 30 min.).The supernatant that contains complex I is loaded onto a Q-Sepharose HiLoad 26/10 ion exchange column equilibrated in low salt buffer A (20 mM Tis-HCl pH 7.4, 2 mM EDTA, 0.1% DDM, 10% Ethylene Glycol, 1 mM DTT and eluted with 0-800 mM sodium chloride (NaCl) gradient over 500 ml at 1ml/min (from buffer A to buffer A with 1 M NaCl). Complex I elutes at about 220 mM NaCl. Eluted NADH:FeCN active fractions were collected, diluted with buffer A to 100 mM NaCl, then concentrated and the Q-column step was repeated in order to improve purification quality and remove the main contaminant, ATP synthase. The final step of the procedure was size exclusion chromatography using a Superdex 200 HiLoad 16/60 column equilibrated in SizeEx Buffer (20 mM Tris-HCl, 2 mM EDTA, 10% Ethylene Glycol, 2 mM DTT, 100 mM NaCl) with 0.2% DDM, which separates the ~1MDa Complex I from a few remaining proteins. The ÄKTA Explorer chromatography system with 280/460 nm absorbance monitor was used for ion exchange and size exclusion chromatography.

**Improved purification**

Improved purification was performed similarly to the original preparation but it differs in details, which improve the purity and quality of the final sample.

Collected membranes were solubilised by incubation in a solubilisation buffer in the cold room (20 mM HEPES pH 7.4, 1 mM ethylenediaminetetraacetic acid (EDTA),

10% glycerol, 2 mM dithiothreitol (DTT), 0.002% phenylmethanesulfonyl fluoride (PMSF), 1% lauryl maltose neopentyl glycol (LMNG). Per 1 g of membrane pellet 8.5 ml of solubilisation buffer was added and stirred for 30 min at 4 $^o$C. Any non-solubilised mitochondrial fraction was removed by centrifugation (SS-34 rotor, 20k rpm, 30 min.). Protein in a solubilised state was loaded onto a Q-Sepharose HiLoad 26/10 ion exchange column equilibrated in low salt buffer Ai (50 mM HEPES pH 7.4, 2 mM EDTA, 0.1% LMNG, 10% glycerol, 1 mM DTT) and washed with a 30 ml linear gradient of 5-22 mM NaCl followed by150 ml of 22 mM NaCl. Complex I was then eluted with a 300 ml linear gradient of 22-305 mM NaCl. Gradients were obtained by using Ai buffer and Ai buffer supplemented with 1 M NaCl. Complex I eluted at about 220 mM NaCl. Eluted NADH:FeCN active fractions were collected and concentrated. At this point sample contains complex I and two main contaminates: complex IV and 2-oxo-glutarate dehydrogenase.

The final step was size exclusion chromatography, which separated complex I from remaining contaminants, using a Suprose 6 10/30 HR column equilibrated in SizeEx buffer (20 mM HEPES, 2 mM EDTA, 1.5% glycerol, 2 mM DTT, 100 mM NaCl) supplemented with 0.2% of LMNG. The ÄCTA Explorer chromatography system with 280/460 nm absorbance monitor was used for ion exchange and size exclusion chromatography.

**Sucrose gradient purification**

As a gentler, but less efficient, alternative to size exclusion we experimented with using sucrose gradient ultracentrifugation. Solubilized mitochondrial sample after the Q column purification step of the initial purification protocol was loaded on 10-35% sucrose gradient containing buffer A and centrifuged in 4° C for 16.5h with the speed of 38k rpm using SW 41 Ti rotor. This method was proven to work for complex I purification but due to lower purity has not been used in further experiments.

**2.4 Detergent exchange**

*2.4.1 Amphipathic polymers*

Protein sample in detergent (LMNG or DDM) was diluted to ~2 mg/ml and 100 μl volume with SizeEx buffer. Amphipathic polymer (Amphipol A8-35, NV10 or PMALs) in solution was added to protein to achieve a certain w/w ratio (0.3:1, 1:1 or 1:3 reagent to protein) and 250 μl volume and incubated on ice for 10 min. Next, to remove detergent, β-cyclodextrin was added in molar ratio (n/n) 2:1 to detergent and incubated for another 5 min. The mixture was loaded on the Superose 6 HR 10/30 size exclusion column equilibrated with SizeEx buffer without detergent.

*2.4.2 Detergents*

Protein sample purified in LMNG or DDM was diluted to final concentration of ~2 mg/ml and 100 μl volume with SizeEx buffer. Detergent of choice in solution was added to protein to the final ~0.2% concentration. Next, the mixture was loaded on the Suprose 6 HR 10/30 size exclusion column equilibrated with SizeEx buffer supplemented with a detergent of choice at ~0.2% concentration.

**2.5 Cryo-electron microscopy**

*2.5.1 Grids preparation*

2.7 μl of solubilized protein sample concentration of ~1-5 mg/ml was applied to copper grids (Quantifoil mesh 300, R 0.6/1, 1.2/1.3 or 2/2) glow discharged in air at 30 mA for 120 s and blotted for a certain period of time (5-34 s) at 90% humidity in the chamber of FEI Vitrobot IV. Immediately after the sample was snap frozen in liquid ethane.

*2.5.2 Cryo-EM screening conditions*

Initial images were collected at -173 °C on a Tecnai T12 electron microscope (FEI) operating at 120 kV using a Gatan UltraScan 1000 2k × 2k CCD camera, following low-dose procedure. Grids were mounted on a pre-cooled Gatan 626 model single tilt cryo transfer holder and secured using a clip-ring system. Images were recorded at a nominal magnification of 30000 X. Defocus was set for -3 to -7 μm, electron dose 15-25 e$^-$/s/Å$^2$ and exposure time 3-8 s. Micrographs were inspected

using DigitialMicrogprah (Gatan), FIJI and the e2display program from the EMAN2 package.

## 2.5.3 Cryo - EM data collection conditions

Imaging was performed with a 300kV Titan Krios electron microscope equipped with direct electron detector FEI Falcon-II (ETH Zurich, ScopeM center) in automated data collection mode at calibrated magnification of 1.39 Å/pixel (100720 X) and dose of 26 $e^-/s/Å^2$ with total 3s exposure time. The data were collected as seven movie frames fractionated over the first second of exposure and an averaged image over 3 seconds.

## 2.6 Image processing

### 2.6.1 Initial processing

Two datasets of total ~2.6 k micrographs in two datasets were collected. All processing steps were done using RELION (Scheres, 2012) unless stated. For the initial CTF estimations using CTFFIND4 (Rohou & Grigorieff, 2015) and for automated particle picking in RELION averaged images from high dose 3 s exposure were used. MOTIONCORR (Li et al., 2013) was used for whole-image drift correction of movie frames 1 - 7 (1 s) of each micrograph. Contrast transfer function (CTF) parameters of the corrected micrographs were estimated using Gctf and refined locally for each particle (Zhang, 2016). All the consecutive steps of processing were conducted in RELION 1.4 package including focused 3D refinements with the subtraction of signal from the remaining parts of the complex (Bai, Rajendra, Yang, Shi, & Scheres, 2015).

### 2.6.2 Separation of different complex I conformations

Improved 3D classifications on the initial "open" and "closed" subsets into ovine classes 1-5 was conducted in RELION 2.0. That includes 3D refinement focused on the peripheral domain of combined classes "open" and "close" with the signal subtraction,3D classification using masked membrane domain without alignments and 3D auto refinements of the separated classes.

## 2.7 Modelling of the ovine mitochondrial complex I

*2.7.1 Model building*

As an initial reference for determination of atomic ovine model of mitochondrial complex I, so far published structures and protein homology recognition software was used. SQWRL4 software (Krivov, Shapovalov, & Dunbrack, 2009) was used for the initial core subunits homology models based on *T. thermophilus* structure (R. Baradaran et al., 2013). Homology models for the supernumerary subunits were generated with Phyre2 (Kelley, Mezulis, Yates, Wass, & Sternberg, 2015) and Swiss-model (Biasini et al., 2014). Secondary structure predictions for all subunits were generated with PredictProtein (Rost, Yachdav, & Liu, 2004), PsiPred (McGuffin, Bryson, & Jones, 2000) and TMHMM (Krogh, Larsson, von Heijne, & Sonnhammer, 2001) servers. The initial models were adjusted to cryo-EM density (in cases when homology models were useful) or built manually in COOT.

Initial models were re-built and refined in Rosetta release version 2016.02.58402 using protocols optimized for cryo-EM maps (DiMaio et al., 2015). For each subunit, 100 different models were produced in Rosetta with optimization of density fit using *elec_dens_fast* function (with *−denswt* = 40, chosen from several trials), selection of the best fitting structure and structure relaxation using *−FastRelax* flag. From the produced structures several best-scoring by density fit and geometry were selected and used in COOT to guide further model building/optimization. This procedure resulted mainly in improvements to backbone geometry, especially in coils, still allowing for the good fit of side-chains into density. All the work involving Rosetta was done by Dr Karol Kaszuba (IST Austria).

*2.7.2 Refinement*

The final model refinement was performed with the Phenix suite (Adams et al., 2010). The Model was refined against "combined" map (Chapter 3.7.2) using the *phenix.real_space_refine* program with following options *run=minimization_global+morphingscattering_table=electron secondary_structure.enabled=True macro_cycles=3 resolution=3.9.* Geometry description files for the co-factors and ligands were calculated using GRADE

server (Vonrhein et al., 2011). After rounds of refinement model was rebuild using COOT and refined again with 5 cycles but without morphing to give the final model.

## 2.8 Cross-linking

All the cross-linking reactions were performed using purified solutions of complex I at a concentration of 1 mg/ml. Following experimental optimisation, ten separate experiments were performed. Experiments varied in relation to the detergent added to the buffer (DDM, LMNG or LDAO/DDM), the cross-linking reagent (targeting lysine or acidic residues) and the protease used to digest the samples (trypsin or endoproteinase Glu-C). Cross-linking experiments were performed by collaborators Dr Gianluca Degliesposti (MRC-LMB Cambridge, UK) and Dr Mark Skehel (MRC-LMB Cambridge, UK). Results in detail were analysed and summarised by Dr James Letts (IST Austria).

## 2.9 Preparation of figures

Figures were prepared using: UCSF Chimera (Pettersen et al., 2004), PyMOL (Shrodinger, Camberley, Surrey, UK), Adobe Illustrator (Adobe Systems Europe Ltd, London, UK).

# Chapter 3. Results

## 3.1 Choice of experimental organism

*Ovis aries* has been chosen as a source for complex I preparation. Recently it has been shown that it may be more suitable for structural studies than the extensively studied bovine enzyme because of its stability (Letts, Degliesposti, et al., 2016).

Sequence similarity analysis revealed that sheep complex I is as similar to human as bovine (the best characterised so far) or porcine enzymes and for that reason it is an equally good model.

| Species | Identity (%) | | |
| --- | --- | --- | --- |
| | Core Nuclear | Core Mitochondrial | Supernumerary |
| Bos taurus | 96.0 | 70.6 | 84.5 |
| Sus scrofa | 95.7 | 69.9 | 83.9 |
| Ovis aries | 96.1 | 70.9 | 83.1 |

**Table 3.1** Comparison of complex I subunits of three organisms easily available for studies (*Sus scrofa, Bos taurus, Ovis areies*) in terms of sequence similarity to human enzyme. Subunits were divided by encoded by nuclear genome, mitochondrial genome, supernumerary and core. Sequences of subunits were obtained from NCBI (Altschul, Gish, Miller, Myers, & Lipman, 1990). Prior to alignment nucleus encoded subunits were truncated to remove mitochondrial targeting sequence based on already published biochemical data or sequence similarities (Letts, Degliesposti, et al., 2016). Detailed result can be found in Supplementary **Table S1**.

## 3.2 Protein purification

### 3.2.1 Initial purification

The initially developed procedure for the purification of ovine complex I (in DDM, Chapter 2.3.3), was based on the original protocol for the bovine enzyme (L A Sazanov et al., 2000). This protocol provided reproducibly pure, homogenous complex I. As a starting material lamb hearts from animals of the same age and sex on the day of slaughter were used. From 36 ovine hearts we obtained ~220 g of mitochondria. The intact enzyme was purified from the internal mitochondrial membrane. During the process pool 36% of the starting amount of complex I is lost on the first Q column step (Fig. 3.1), another 24% during second Q column (Fig. 3.2) and 17% during size exclusion chromatography step (Fig. 3.3) (Table 3.2). Loss of complex I is compensated by increased purity, leaving few minor contaminants after the last step of purification (Fig. 3.4)

| Step of purification | Volume (ml) | Activity (U/ml) | Total Activity (U) |
|---|---|---|---|
| Membrane fraction | 100 | 59.6 | 5964.9 |
| Solubilized membrane | 100 | 63.8 | 6383 |
| Q-Sepharose I | 126 | 32.9 | 4152.9 |
| Q-Sepharose II | 144 | 17.8 | 2570 |
| Superdex 200 | 12 | 123.6 | 1486.7 |

**Table 3.2** Initial purification of complex I from *O. aries*. Starting material was 10g of mitochondria. Final yield of pure complex I was ~15 mg. Activity was measured as (NADH):ferricyanideoxidoreduction.
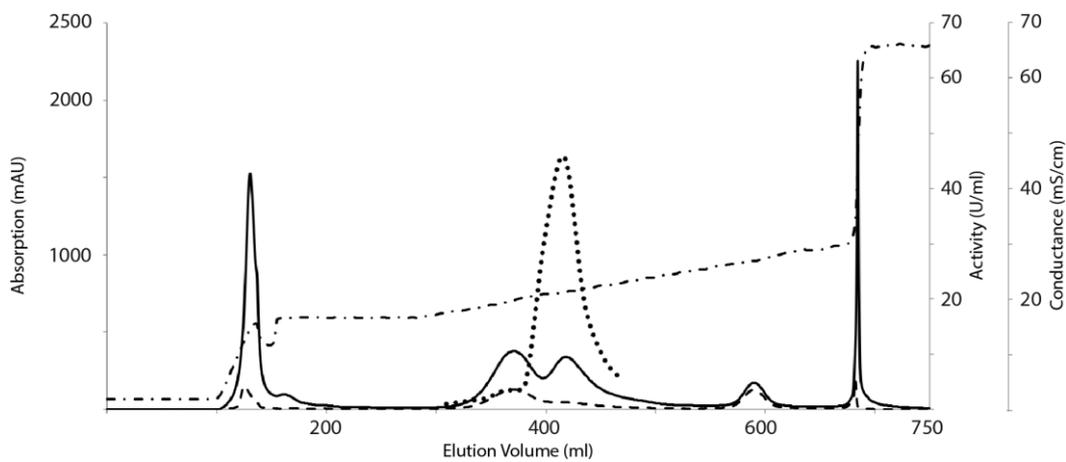
**Figure 3.1** Elution profile of the first Q column of the initial protocol. In solid line absorption at 280nm, in dashed at 420nm, in dots (NADH):ferricyanideoxidoreduction activity and in dashed conductance.
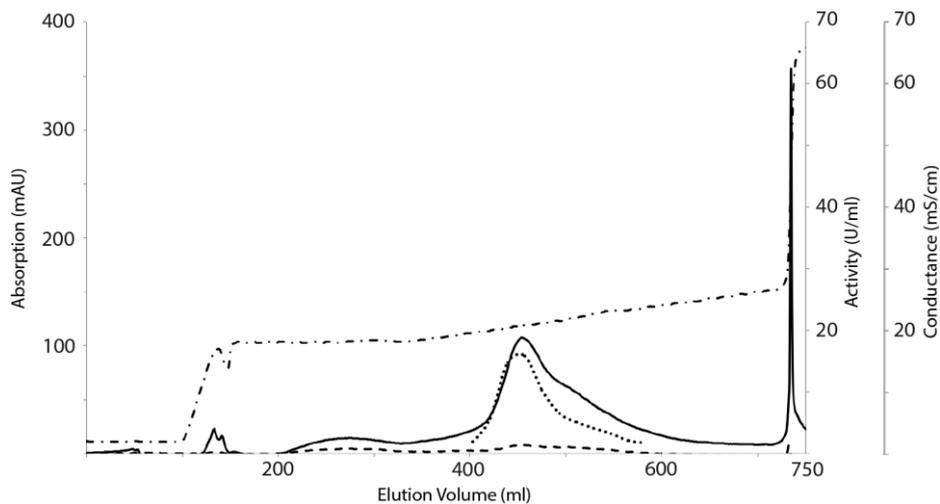


**Figure 3.2** Elution profile of the second Q column of the initial protocol. In solid line absorption at 280nm, in dashed at 420nm, in dots (NADH):ferricyanideoxidoreduction activity, and in dashed conductance.
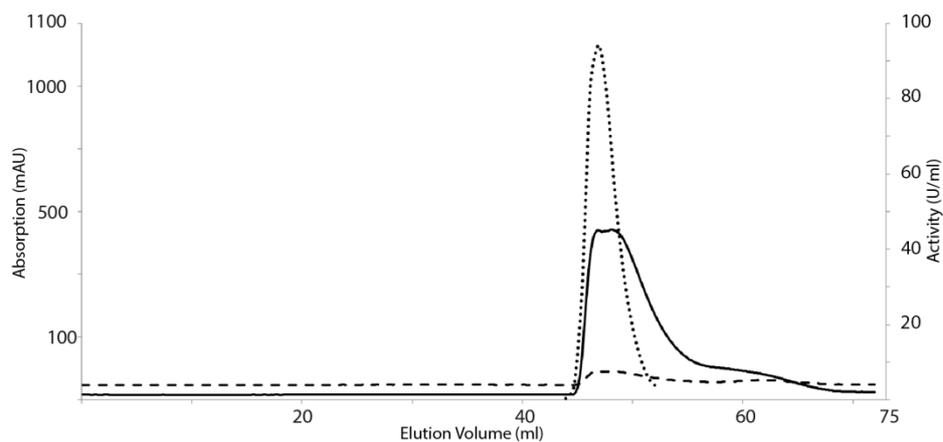
**Figure 3.3** Elution profile of the size exclusion column of the initial protocol. In solid line absorption at 280nm, in dashed at 420nm and in dots (NADH):ferricyanideoxidoreduction activity.
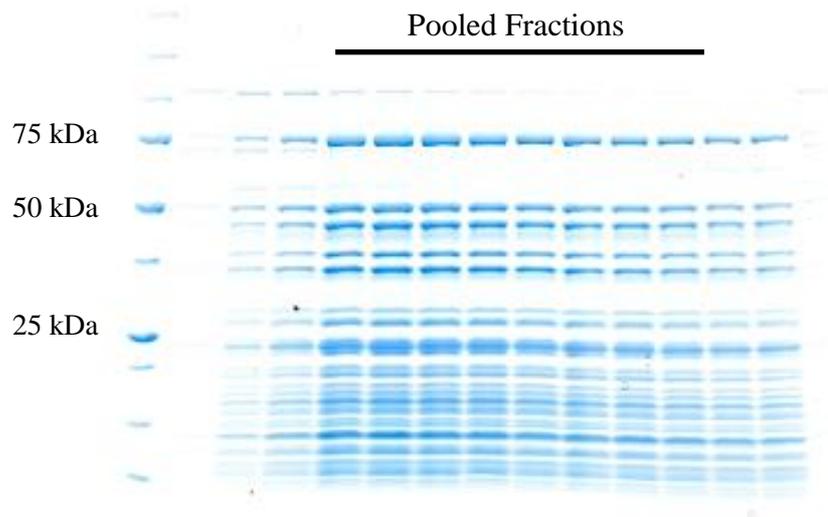


**Figure 3.4** SDS-PAGE gel of every second fraction between 43 and 64 ml (left to right) from the elution profile presented in Fig. 3.3. Indicated are collected fractions of complex I.

Fractions corresponding to complex I (Fig. 3.3) were pooled, concentrated and frozen in liquid nitrogen (LN) with 25% glycerol. Final protein concentration was 36 mg/ml.

Freezing and thawing of the protein sample did not cause any loss of NADH:FeCy activity (Table 3.3), DQ activity (Letts, Degliesposti, et al., 2016) nor protein degradation or aggregation (Fig. 3.5). However it influenced the quality of cryo-EM grids (Chapter 3.5)

| Sample | Activity (U/mg prot.) |
|---|---|
| Complex I before freezing | 101 |
| Complex I after 3 days in LN | 103.8 |

**Table 3.3** (NADH):ferricyanideoxidoreduction activity of complex I before and after freezing.
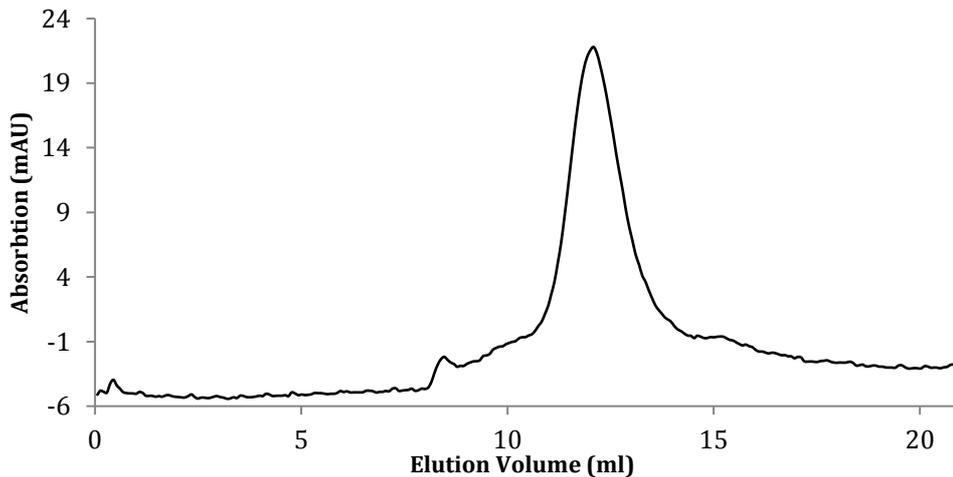


**Figure 3.5** Elution profile of complex I after freezing and storing for 3 days in LN. Sample was run in the buffer used for size exclusion chromatography during purification (SizeEx buffer with 0.2% DDM).In solid line absorption at 280nm.

### 3.2.2 Improved purification

The improved procedure for the purification of ovine complex I (in LMNG, Chapter 2.3.3), was closely based on the already published protocol for the ovine enzyme (Letts, Degliesposti, et al., 2016). This protocol provides reproducibly pure, homogenous complex I. It was found that compared to the initial DDM preparation, this procedure produces more stable, fully active complex and with the easily lost 42 kDa subunit retained (Letts, Degliesposti, et al., 2016). During the process 69% of the initial complex I is lost on the first Q column step (Fig. 3.6) (possibly due to the fact that by solubilizing mitochondrial membranes with LMNG, large percentage of complex I remains in form of supercomplex $I_1III_2$ and elutes at different NaCl concentration [data available in our laboratory, not published]) and 16% during the size exclusion chromatography step (Fig. 3.7) (Table 3.3). Loss of complex I is compensated by increased purity, leaving few minor contaminants after the last step of purification (Table 3.4).

58

| Step of purification | Volume (ml) | Activity/ml | Total Activity |
|---|---|---|---|
| Membrane fraction | 50 | 63.1 | 3157.1 |
| Solubilized membrane | 50 | 59 | 2964 |
| Q-Sepharose | 63 | 17.6 | 1076.2 |
| Superose 6 | 6 | 45.8 | *274.8 |

**Table 3.4.** Improved purification of complex I from *O. aries*. Starting material was 5g of mitochondria. Final yield of pure complex I was ~5 mg. Activity was measured as (NADH):ferricyanideoxidoreduction. Notice that amount of the starting material is twice less than in the initial protocol (Table 3.2)

*Sample was aliquoted and stored in LN after Q-sepharose purification (discussed below). Total activity and volume was estimated based on one sample thawed and run over analytical size Superose 6 column equilibrated in SizeEx buffer supplemented with LMNG. Low total activity may be caused by protein loss during this procedure or total volume underestimation.
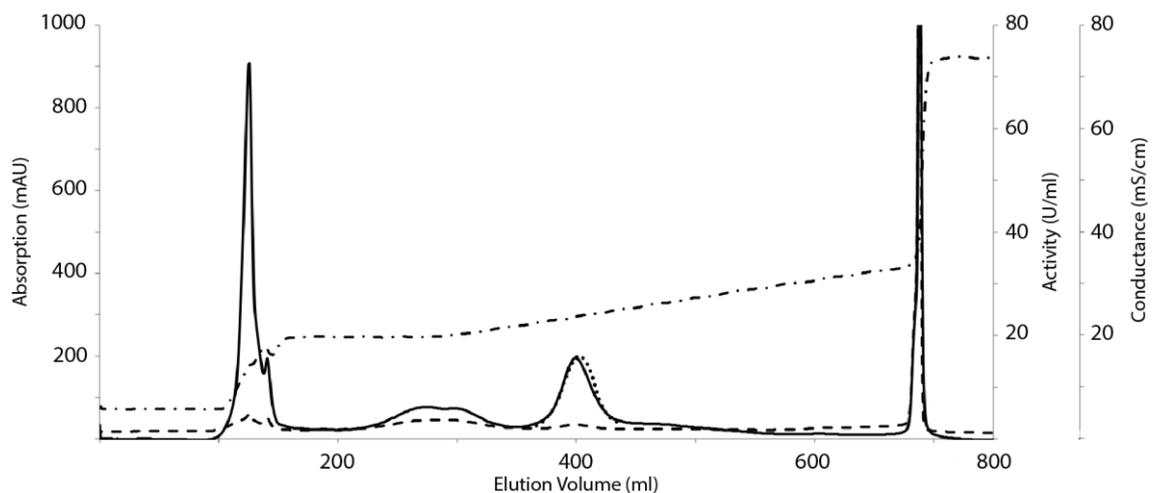


**Figure 3.6.**Elution profile of the Q column. In solid line absorption at 280nm in dashed at 420nm and in dots (NADH):ferricyanideoxidoreduction activity.

After the first Q column, run in LMNG, the purification was finished by gel filtration chromatography in various detergents as needed for cryo-EM experiments. It was noticed that although the protein frozen in Brij-35 does not show signs of decay after size exclusion, it produces cryo-EM specimen with significantly more aggregates than fresh sample (data not presented). For that reason electron microscope grids were always prepared from a sample that was not frozen at any step after membranes solubilisation. To save biological material complex I purifications were performed starting from 5g of mitochondria. Excess protein was frozen in LMNG after the Q-Sepharose purification step and used for initial screening of cryo-EM grids prepared with complex I in other detergents.
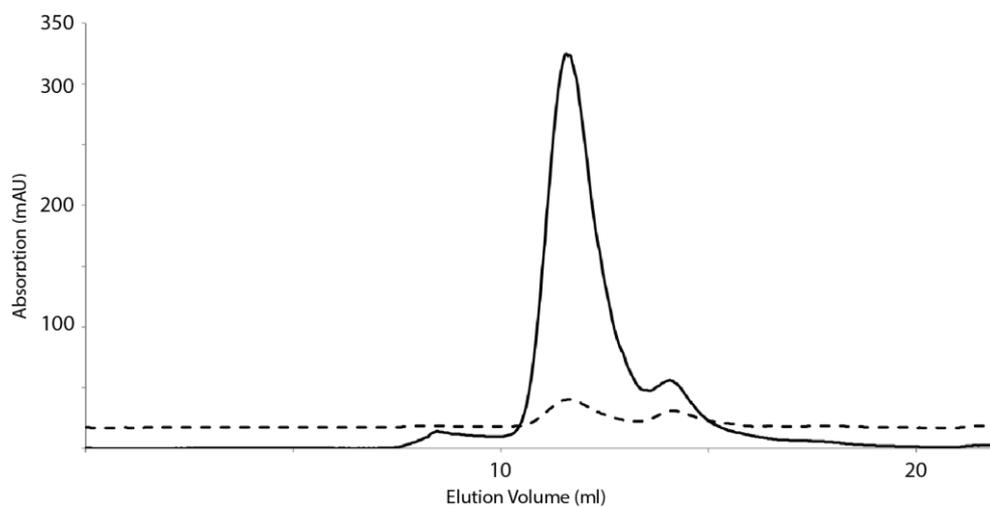
**Figure 3.7.** Elution profile of the analytical size exclusion column in Brij-35, from the sample used for the data collection. In solid line absorption at 280nm and in dashed at 420nm, activity was not measured in order to save sample.
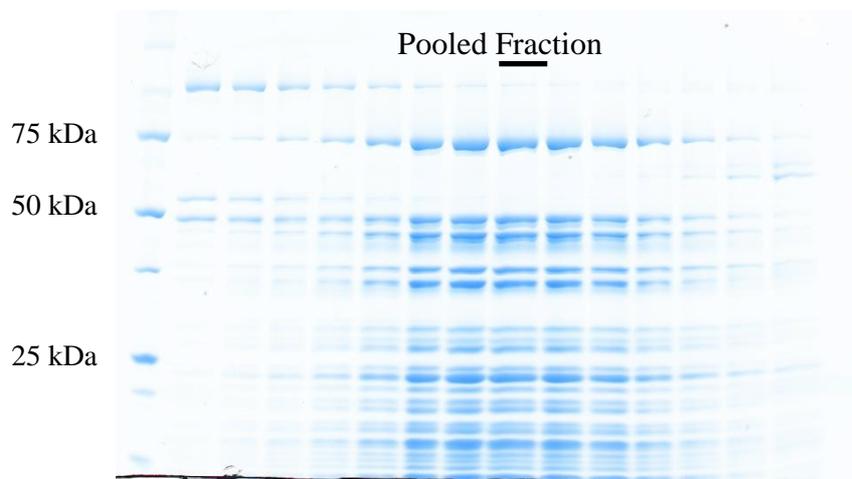


**Figure 3.8** SDS-PAGE gel of every fraction between 10 and 12.6 ml (left to right) from the elution profile presented in Fig. 3.7. Indicated is central peak fraction of complex I (~11.5 ml) that was collected, concentrated and used for grids preparation.

60

## 3.2.3 Sucrose gradient purification

Alternatively, the last step of purification can be substituted by sucrose gradient centrifugation (Fig. 3.9 and 3.10). Sample obtained this way was not used for later experiments due to relatively poor purity (Fig. 3.9).
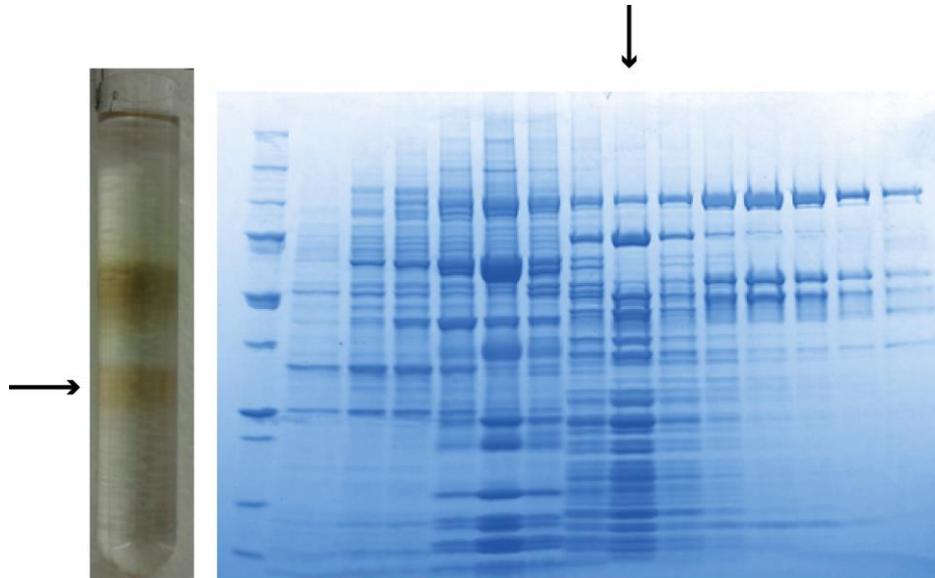


**Figure 3.9A.** Sucrose gradient with complex I present in the lower brown band (indicated by arrow). **B.** SDS-PAGE gel of fractions collected from the sucrose gradient. Arrow indicates fraction containing complex I. Fractions collected from the tube from top and loaded on the gel from the left.
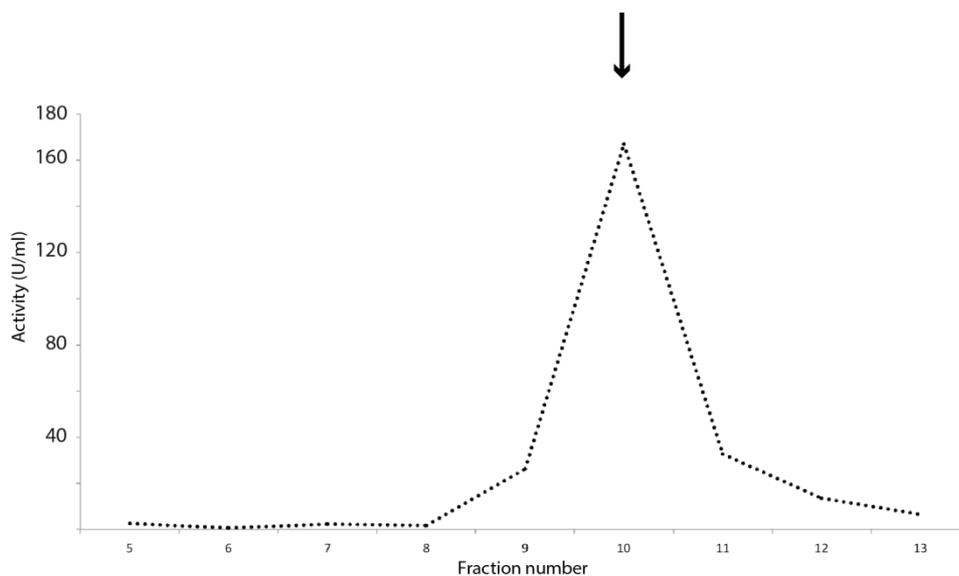


**Figure 3.10** (NADH):ferricyanide oxidation measurements of fractions from the sucrose gradient. In dots (NADH):ferricyanideoxidoreduction activity. Fractions collected from the top, arrow indicates complex I fraction.

*3.2.4 Mass spectroscopy analysis of ovine complex I*

In the sample prepared initially with DDM some subunits were not identified. However, the missing subunits: 10 kDa, B14.7, 15 kDa, MWFE, KFYI, B15, B14.5b, ND3, ND4L, ND6 are very small and/or hydrophobic. Thus their lack in the sample can be rather assigned to sub-optimal conduction of the initial mass spectroscopy experiment and some limitations of the method (Supplementary Table S2, Fig S1).

Mass spectrometry identification of complex I subunits was performed again on complex I purified in LMNG and detergent exchanged to Brij-35 (sample used for structural data collection). It revealed presence of all complex I subunits except ND4L that does not have any triptic digestion sites (Supplementary Table S2). The experiment including SDS-PAGE separation of subunits was performed by Dr. Gianluca Degliesposti and the details including gels can be found in Letts et al. (Letts, Degliesposti, et al., 2016).

## 3.3 Detergent exchange

*3.3.1 Detergents*

In search of the optimal detergent for cryo-EM imaging, stability of the purified complex I was examined in a few common detergents (Table 3.5). Detergent exchange was performed by running the sample after Q-column or size exclusion purification step on the analytical Superose 6 10/30 size exclusion column in SizeEx buffer containing the detergent of choice. The effectiveness of the exchange and stability of the complex was judged by gel filtration chromatography, with intact complex I eluting at 11.5 ml (Fig. 3.11). Disruption of the complex would result in peak broadening and shift to later elution volumes. More extensive experiments were performed by Dr James Letts and not repeated (Letts, Degliesposti, et al., 2016).

| Tested detergent and used concentration | Complex I stability |
|---|---|
| lauryl maltose neopentyl glycol (LMNG) 0.01% | stable |
| Triton X-100 0.05% | stable |
| Brij-35 0.05% | stable |
| Cymal 7 0.05% | stable |

**Table 3.5** Stability of complex I in some common detergents derived from gel filtration chromatography experiments.
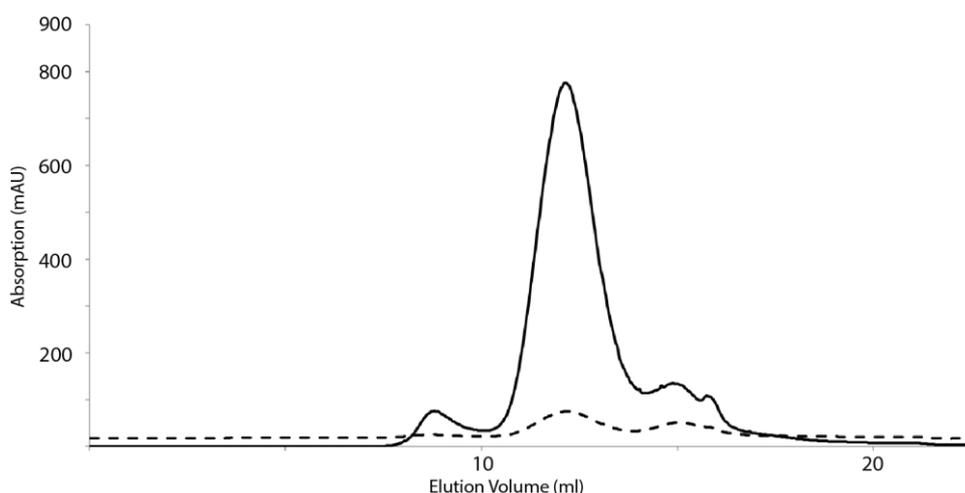


**Figure 3.11** Representative elution profile of detergent exchange run (in this case Brij-35 at 0.05%) yielding stable enzyme. In solid line absorption for 280nm and in dashed for 420nm.

### 3.3.2 Amphipathic polymers

Amphipathic polymers provide an alternative method to solubilise membrane proteins and avoid potential problems of detergent background and "lensing effect" in cryo-EM imaging (Flotenmeyer, Weiss, Tribet, Popot, & Leonard, 2007). The effectiveness of the exchange was judged by gel filtration chromatography, with intact complex I eluting at 11.5 ml (Fig. 3.12).

| Tested amphipathic polymer | Exchange ratio protein/polymer (w/w) | Complex I stability |
|---|---|---|
| Amphipol A8-35 | 1:3 | stable |
| Amphipol A8-35 | 1:1 | stable |
| Amphipol A8-35 | 3:1 | stable |
| PMAL C-8 | 1:1 | not stable |
| PMAL C-12 | 1:1 | not stable |
| PMAL C-16 | 1:1 | not stable |
| NV10 | 1:1 | poor stability |

**Table 3.6** Stability of complex I tested amphipatic polymers. Derived from gel filtration chromatography experiments

Only Amphipol A8-35 was found to be a good solubilisation molecule for complex I. All tested w/w exchange ratios (Table 3.6) gave a mono-disperse peak at the expected elution volume.
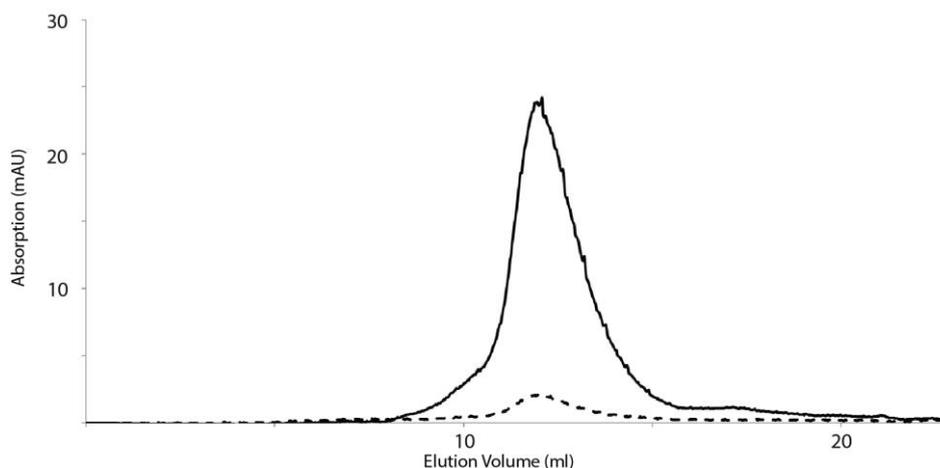


**Figure 3.12** Representative elution profile for complex I sample with DDM exchanged to Amphipol A8-35 (in this case 1:1 protein:amphipol ratio).

PMAL polymers were found to cause immediate aggregation of complex I (Fig. 3.12) that could not have been reversed by adding detergent back to the sample. Protein stability test was conducted by a quick method of measuring the absorbance in the range of 220-350 nm (Porterfield & Zlotnick, 2010). In each of the cases the overall absorbance profile and the A260/280 ratio was examined. For PMAL C16 and PMAL C12 protein aggregation was easily visible by eye due to the milky colour of the protein solution. Complex I with PMAL C8 was transparent and showed A260/A280 ration only slightly higher than expected for proteins ~0.6 (Porterfield & Zlotnick, 2010). However sample injected on the analytical size exclusion column equilibrated in SizeEx buffer with no detergent did not elute, suggesting that protein aggregated and was retained on the filtering system of the purifier.

PMAL C8 A260/280: 0.69



PMAL C12 A260/280: 1.13



PMAL C16 A260/280: 1.27
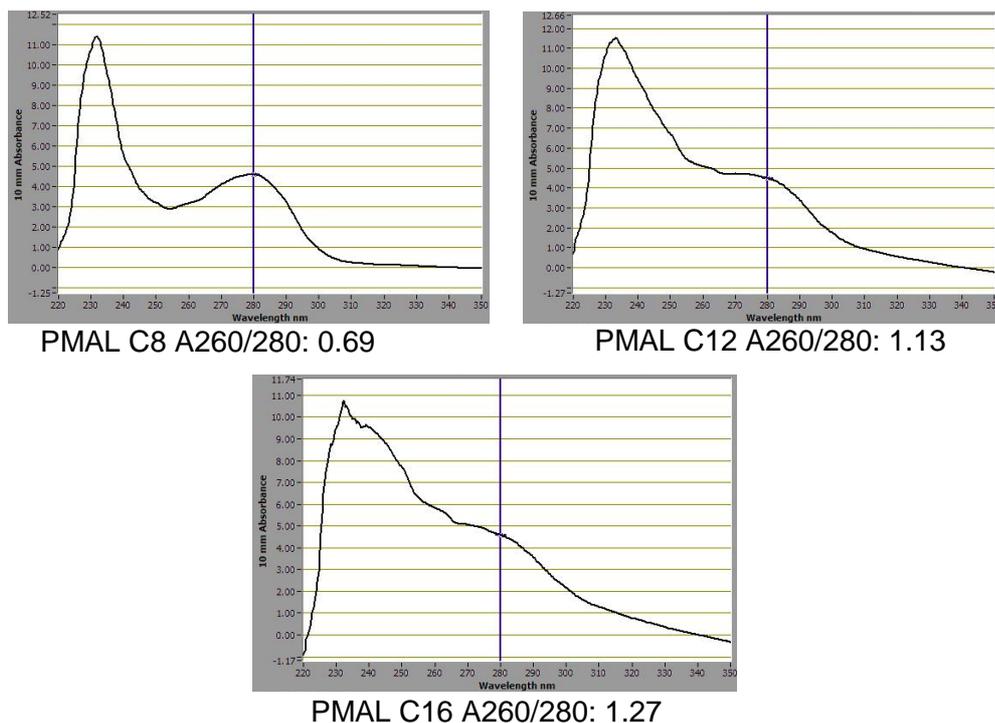
**Figure 3.13** Absorption spectrum for complex I sample with PMALs in 1:1 w/w ratio. The purple vertical line indicates absorption at 280 nm. A260/280 ratios are indicated.

The complex I sample exchanged from detergent to NV10 remained soluble on the size exclusion column in SizeEx buffer without detergent as mono-disperse peak, however the protein eluted 1 ml later than expected (Fig. 3.14). A possible explanation is that the enzyme broke down into the peripheral and membrane domains that co-eluted. Grids prepared from this sample contain aggregates and particles reassembling membrane and matrix arms sub-complexes (Fig. 3.20). Thus NV10 may be a useful reagent for complex I sub-complexes studies.
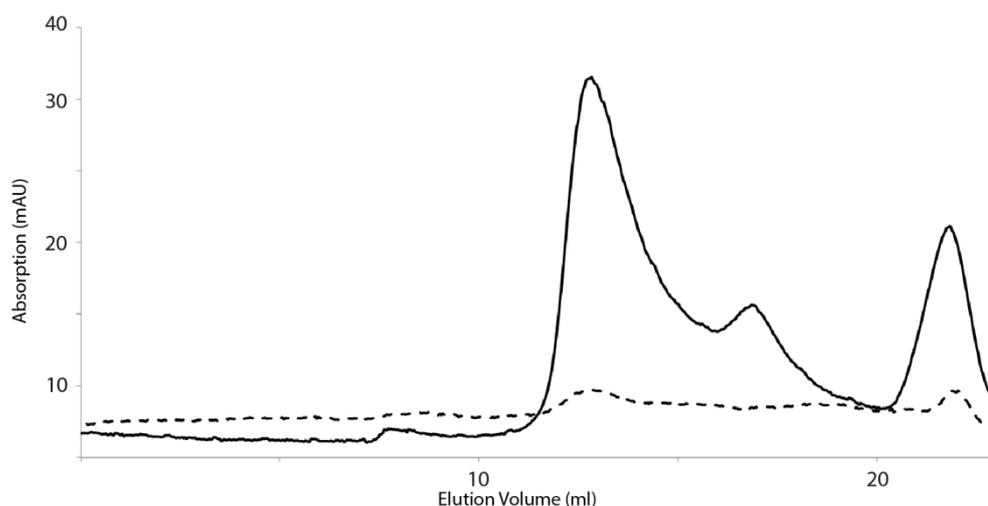
**Figure 3.14.** Elution profile for complex I sample with DDM exchanged to NV10 at 1:1 protein:polymer ratio. The peak at the column volume (22 ml) may contain detached small supernumerary subunits.

## 3.4 Salt concentration screens

In order to examine minimal NaCl concentration required for complex I stability in detergent a number of conditions were screened. Protein purified in DDM detergent was diluted in SizeEx0.2% DDM buffer ranging in the final NaCl concentration from 100 to 2 mM. Results suggest that enzyme aggregates in concentration below 30 mM NaCl however it can be rescued by adding the salt back. The 100 mM NaCl used for cryo-EM grids preparation as well as the 50 mM NaCl used for the Q-column equilibration during purification are sufficient to keep the protein intact.
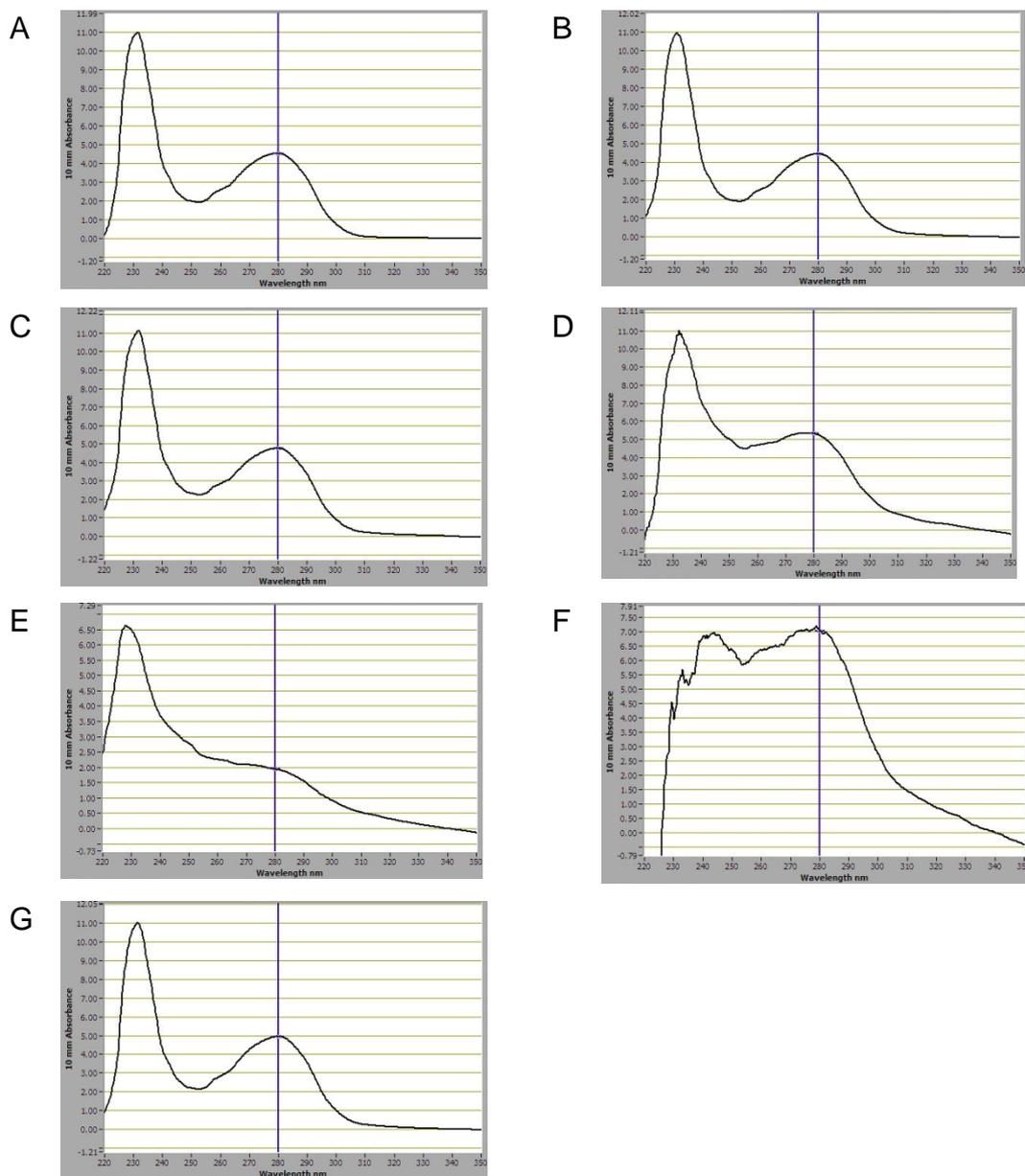
**Figure 3.15.** Absorbance profile for complex I sample with DDM in different NaCl concentrations. **A.** 100 mM NaCl A260/280: 0.58 **B.** 50 mM NaCl A260/280: 0.57 **C.** 35 mM NaCl A260/280: 0.60 **D.** 22.5 mM NaCl A260/280: 0.88 **E.** 11.3 mM NaCl A260/280: 1.16 **F.** 2 mM NaCl 260/280: 0.9 **G.** 100 mM NaCl (salt added to sample with 2 mM NaCl like panel **F.**) 260/280: 0.58

## 3.5 Preparation and optimization of cryo-EM grids

### 3.5.1 Blotting time optimization

The first parameter optimized during grids preparation was the blotting time. In most cryo-EM studies blotting time of grids varies in between 2-10 seconds (http://www.emdatabank.org/search.html). For the ovine complex I grids, a

much longer  blotting time was required. For a sample containing 0.2% DDM and applied on a Quantifoil holey carbon R2/2 Cu-300 grid, blotting times varying from 3 to 25 seconds were tested. Relatively long 20 seconds blot was found to be optimal (Fig. 3.16). However during later experiments it become clear that blotting time depends heavily on the type and concentration of solubilisation reagent, grid type and protein concentration. For the 0.2% Brij-35 solubilized sample at 5 mg/ml applied on R0.6/1 grids (like those used for the data collection) 34s blotting time was used. For samples in Amphipol A8-35 (1 mg/ml protein and R0.6/1 grids) usually ~15 s blotting time was optimal.

A

B
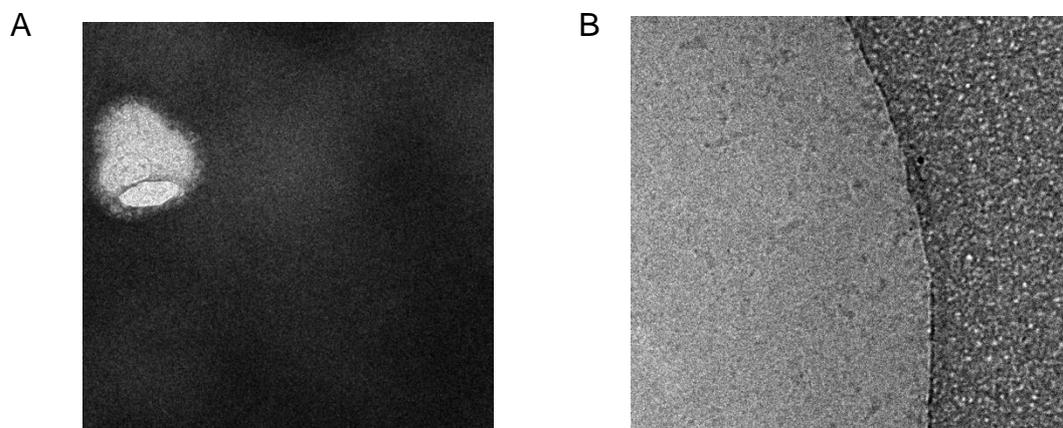


**Figure 3.16.**Cryo-EM micrographs of complex I in SizeEx buffer supplemented with 0.2% DDM using Quantifoil R2/2 grids. **A.** 13s blotting time, M=~5000X, the mark burned in a thick ice using the electron beam shows an edge of the grid hole **B.** 20s blotting time, M=30000 X, image shows part of the grid hole with particles around the carbon support edge.

### 3.5.2 Choice of electron microscopy grids

Initially Quantifoil holey carbon R2/2 Cu-300 grids were used for the microscope sample preparation. However, the high detergent and protein concentration and long blotting time needed were causing the so called "lens effect" caused by uneven ice thickness distribution in the holes (Flotenmeyer et al., 2007). As a result the ice in the center is much thinner than on the edges causing clustering of the protein particles only at the rim of the hole. To overcome this difficulty, grids of the smallest available hole diameter on the market (Quantifoil holey carbon R0.6/1 Cu-300) were used.
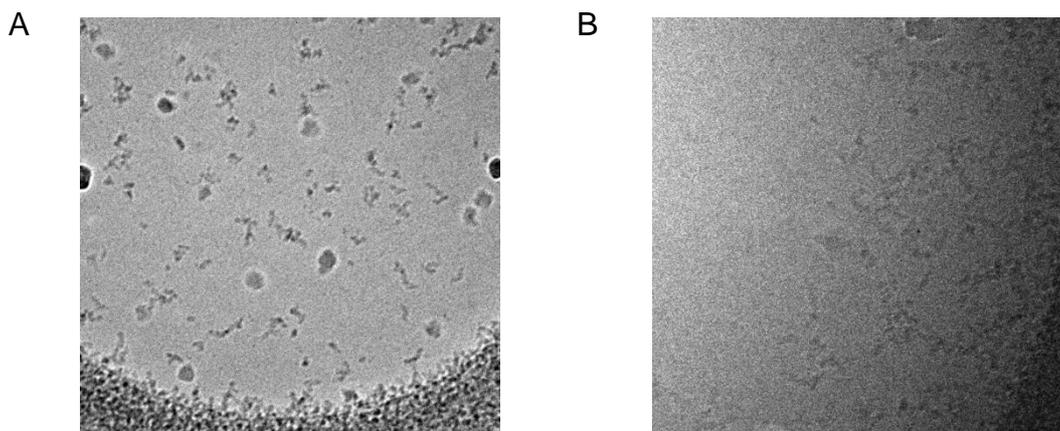
**Figure 3.17.** Cryo - EM micrographs of complex I in SizeEx buffer supplemented with 0.1% DDM **A.** Quantifoil R0.6/1 grid, M=30000 X **B.** Quantifoil R2/2 grid, M=30000 X. Protein concentration ~ 5 mg/ml.

*3.5.3 Screening and optimization of detergent conditions for complex I cryo-EM studies*

As mentioned in chapter 3.3 a number of detergents were found to keep complex I soluble outside the lipid membrane environment. Detergents that have proven to be good solubilisation reagents for complex I in size exclusion chromatography were used for the cryo-EM grids preparation.

After grid preparation, complex I in Triton X-100, DDM and LMNG was either broken into subcomplexes or aggregated (Fig. 3.18 A., C., D., E.). Disruption of the complex probably happened due to strong water-air interface interactions during grid blotting (see below). Detergents giving best results were Brij-35 and Cymal-7 (Fig. 3.18 B., F.). However, for two reasons it was decided to use the Brij-35 sample for the data collection. First, it gives good quality grids more consistently, second, it has a very low density (1.05 g/ml at 25 °C), so almost does not contribute to the background noise in the micrographs.
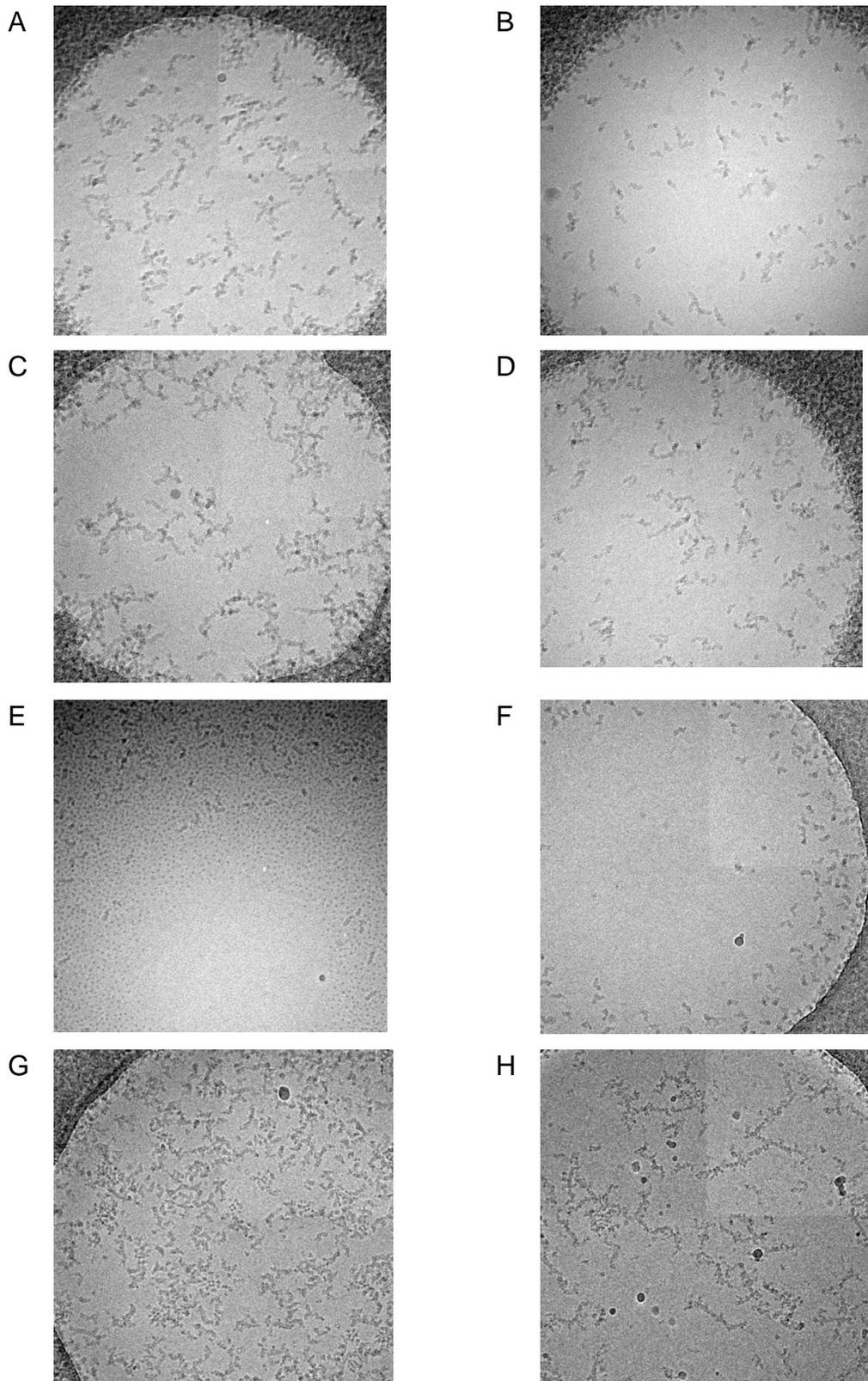
**Figure 3.18** Cryo-EM micrographs of complex I in different detergent buffers. **A.** 0.1% DDM **B.** 0.4% Cymal-7 **C.** 0.05% Triton X-100 **D.** 0.2% Triton X-100 **E. 5%** Brij-35, black dots presumably are detergent micelles **F.** Brij-35 0.2% **G.** LMNG 0.4% **H.** LMNG 0.2%. In all of the cases protein concentration is ~5 mg/ml, grid Quantifoil R0.6/1, nominal microscope magnification 30000 X.

Final optimization of the cryo-EM sample involved testing the optimal concentration of the Brij-35 detergent. Complex I gives a mono disperse size exclusion peak in SizeEx buffer containing only 0.02% of Brij-35 (~2x critical micelle concentration [CMC]). For the cryo-EM sample preparation, concentrations of the detergent from 0.04 to 0.2% were tested. It was noticed that around ~0.12% or less protein starts to show signs of aggregation. For that reason 0.2% Brij-35 was found to be optimal for complex I cryo-EM gird preparation.
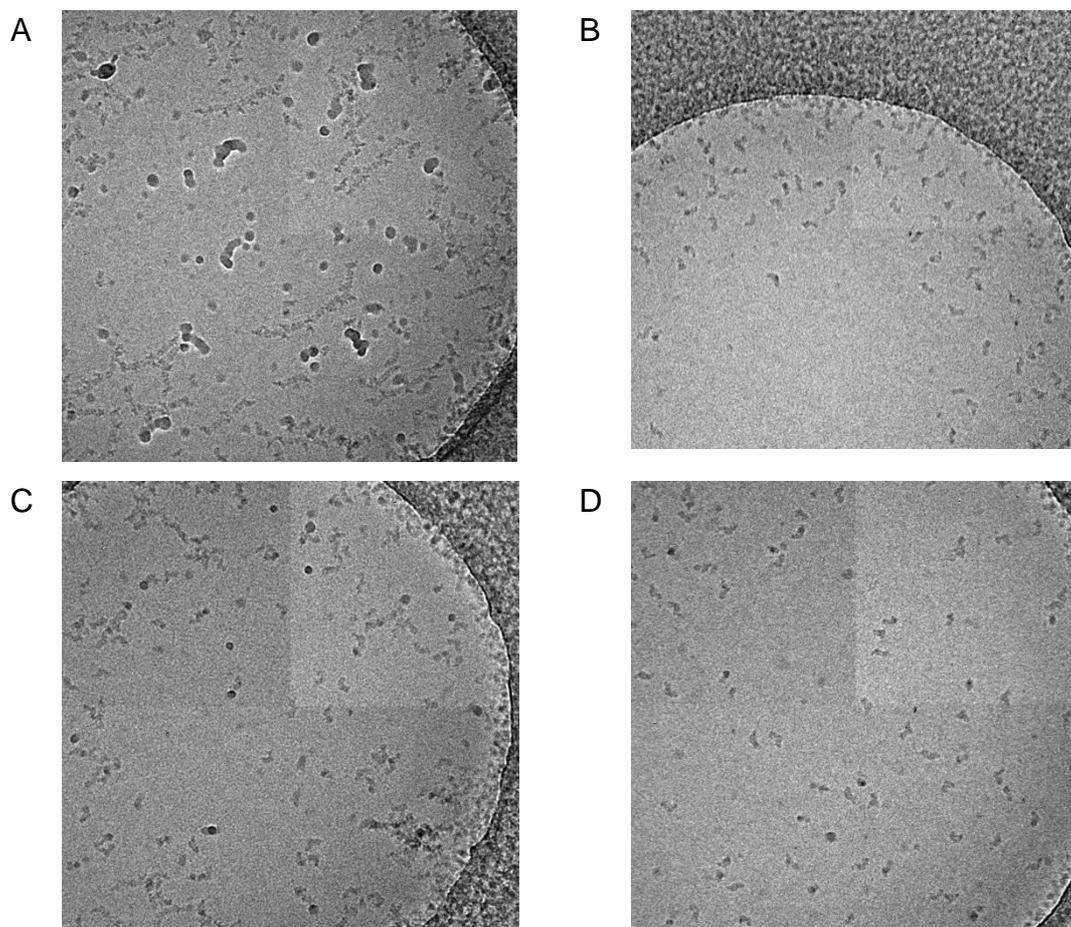


**Figure 3.19** Cryo-EM micrographs of complex I in Brij-35 of different concentrations. **A.** 0.04% **B.** 0.12% **C.** 0.14% **D.** 0.2%. In all of the cases protein concentration is ~5 mg/ml, grid Quantifoil R0.6/1, nominal microscope magnification 30000 X.

*3.5.4 Screening and optimization of amphipathic polymers conditions for complex I cryo-EM studies*

Complex I exchanged to Amphipol A8-35 and NV10 gave a single monodisperse peak when run over size exclusion column equilibrated in SizeEx buffer without detergent. Sample in NV10 as expected from the retention volume contained enzyme broken into sub-complexes (Fig 3.20.A, 3.14). Surprisingly complex I exchanged to Amphipol A8-35 also showed signs of degradation.
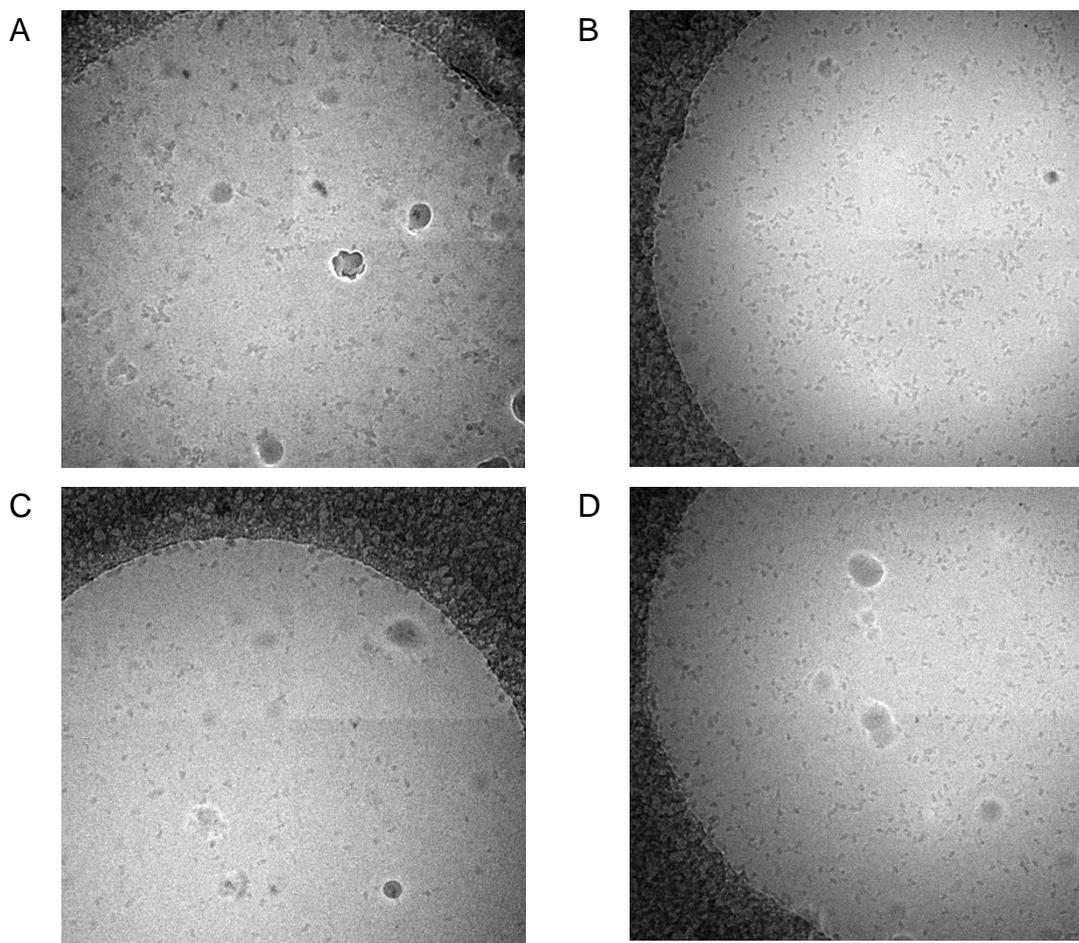


**Figure 3.20**Cryo-EM micrographs of complex I in amphipathic polymers **A.** NV10 pre-incubated with protein in 1:1 (w/w) ratio **B.** Amphipol A8-35 pre-incubated with protein in 3:1 (w/w) ratio **C.** Amphipol A8-35 pre-incubated with protein in 1:1 (w/w) ratio **D.** Amphipol A8-35 pre-incubated with protein in 1:3 (w/w) ratio. In all of the cases protein concentration is ~1 mg/ml, grid Quantifoil R0.6/1, nominal microscope magnification 30000 X.

From the size exclusion profile (Fig. 3.11) and cryo-EM micrographs (3.20.B-D) it was concluded that complex I in amphipol A8-35 breaks during sample

concentration or (more likely) during blotting. To test these hypotheses the sample was supplemented with a number of different additives just before application on the cryo-EM grid: $CaCl_2$ (final concentration 10mM), glycerol (final concentration 2%), amphipol A8-35 (final concentration 0.05%), CHAPS (final concentration 0.05%), Triton X-100 (final concentration 0.05%), LMNG (final concentration 0.05%), Cymal-7 (final concentration 0.05%), octyl-glucoside (final concentration 0.05%) and fluorinated octyl-maltoside (final concentration 0.05%)

G

H

I
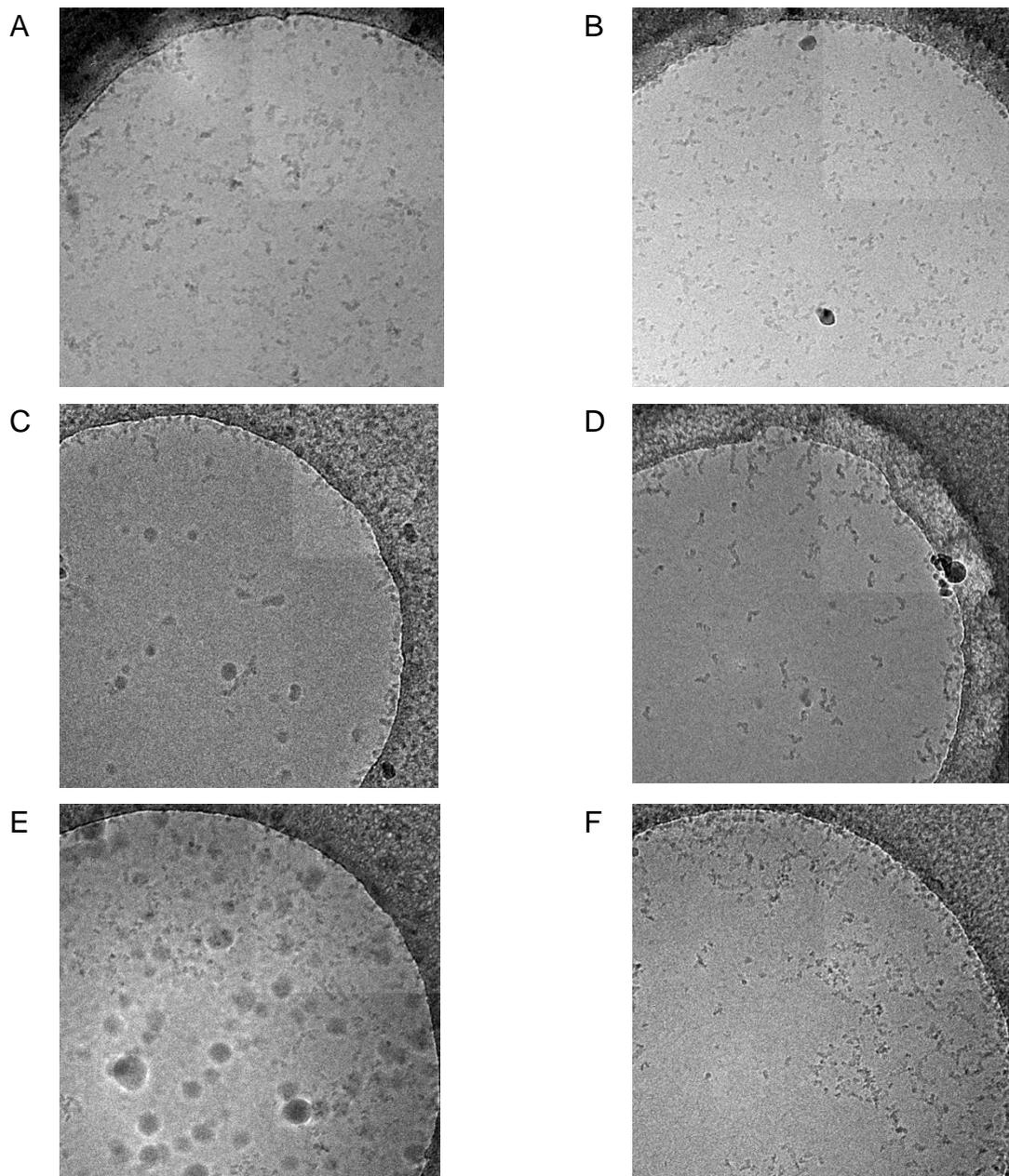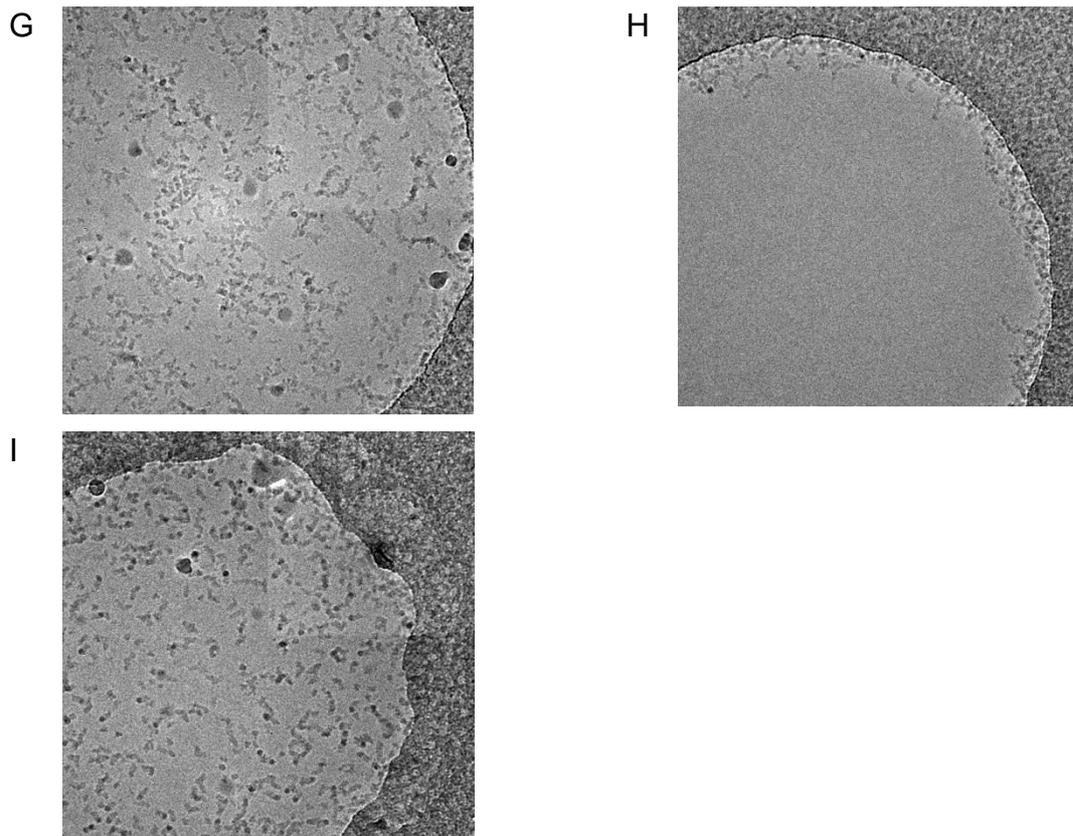
**Figure 3.21** Cryo-EM micrographs of Complex I in Amphipol A8-35 with different additives. **A.** 10mM $CaCl_2$ **B.** 2% glycerol **C.** 0.05% Triton X-100 **D.** 0.05% Chaps X-100 **E.** 0.05%LMNG **F.** 0.05% Cymal-7 **G.** 0.05% octylglucoside **H.** fluorinated octyl maltoside 0.2% **I.** Amphipol A8-35 0.05%. In all of the cases protein concentration is ~1 mg/ml, grid Quantifoil R0.6/1, nominal microscope magnification 30000 X.

The results indicated that the complex broke during the blotting procedure due to attraction of the protein particles to the air-water interphase (Glaeser et al., 2016). Some detergents in close-to-CMC concentrations were able to decrease the surface tension resulting in increase of the number of intact complex I particles. Most promising results for complex I were obtained with Chaps detergent.

Results obtained from grids containing complex I in Amphipol A8-35 supplemented with CHAPS detergent were very promising. However, data obtained from complex I in Brij-35 were still better and more reproducible.

## 3.6 Oxido:reductase activity of sample used for data collection.

DQ activity of complex I sample in Brij-35 was measured. The result is consistent with published studies (Okun, Zickermann, Zwicker, Schagger, & Brandt, 2000) and shows that polyoxyethylene-ether detergents act as specific inhibitors of complex I. In 0.1% Brij detergent (half the concentration used for grids preparation) complex I activity was completely inhibited similarly to 1µM piericidin A or rotenone (Letts, Degliesposti, et al., 2016).

## 3.7 Data collection and processing

Since the sample of ovine complex I in 0.2% Brij-35 detergent gave the most consistently cryo-EM grids of good quality, it was chosen for high-resolution data collection.

### 3.7.1 Data collection

Two datasets of ~2600 images total were collected in a total of ~72 hours. Micrographs were collected in standard FEI Falcon II system at a rate of 7Hz. Movie frames were recorded only for the first second of exposure. Frames from the second and third second were averaged to one frame per second. Thus, for each 3 s exposure 9 frames were collected. The data were collected at defocus ranging from ~2.3 to 3.4 µm. Even in micrographs recorded with low defocus (~2 µm) particles are clearly visible and S/N ratio did not limit particle picking.

### 3.7.2 Data processing

An average of ~100 complex I particles per image were picked automatically with RELION 1.4 using 3 s averaged micrographs. Contrast transfer functions (CTFs) were estimated using CTTFIND4 (Mindell and Grigorieff, 2003).On the basis of poor CTF models as well as visual inspection, bad micrographs were discarded. Low CTF figure of merit values consistently correlated with images blurred due to drift or ice contaminants. This resulted in the removal of ~7.5% of micrographs. For the remaining images, movie frames 1-7 (1 s) were whole-image drift corrected with MOTIONCORR (Li et al., 2013) followed by extraction

of ~241k particles. Finally, the defocus value for each particle was estimated and locally refined with the program Gctf (Zhang, 2016).
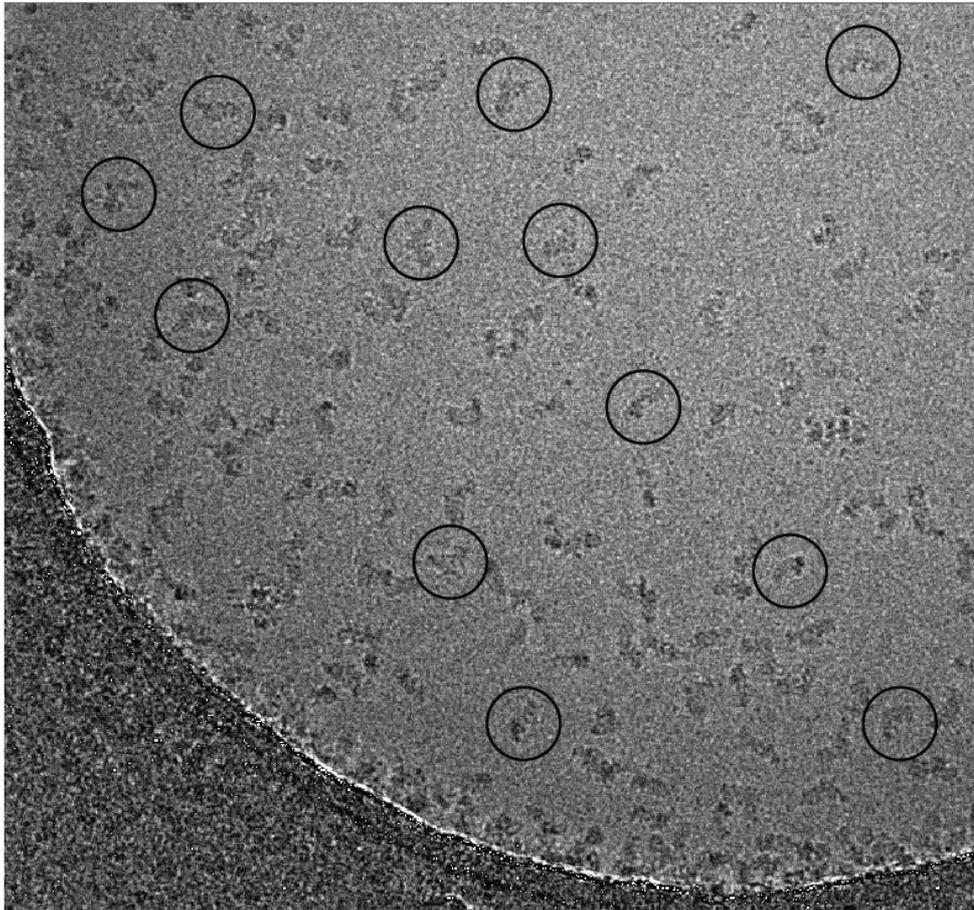


**Figure 3.22.** Typical micrograph obtained during automated data collection. Some representative complex I particles have been marked with black circles.

The particles were extracted using $296^2$ pixel box and sorted by reference-free 2D classification, resulting in ~171k particles selected from good 2D classes.
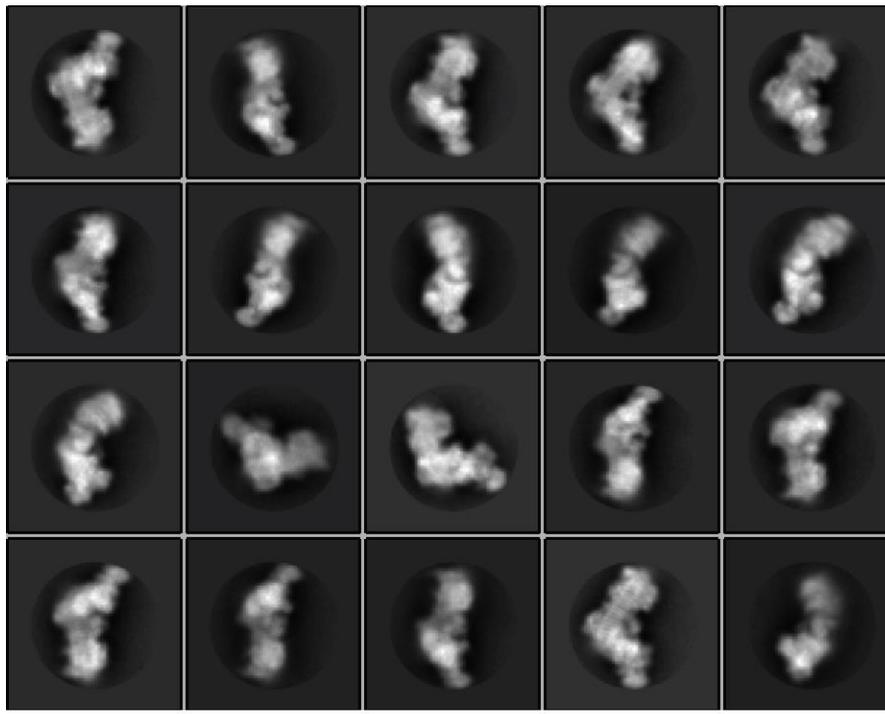
**414 Å**

**Figure 3.23** Examples of the 2D classes used for further processing.

These were used for 3D classification with a regularization parameter T of 8 and 30 Å low-pass filtered initial model from a previous low resolution model of the bovine enzyme (Vinothkumar et al., 2014). This resulted in ~130k particles of good quality, however, it was clear that the relative orientation between the two arms of the complex is slightly variable, producing 3D classes with either an "open" or "closed" angle between the arms (Fig. 3.24). A third class grouped mostly artefacts and broken particles yielding a low-resolution map with characteristic spurious density peaks. Particles in the "open" (~82k particles) and "closed" (~ 46k particles) conformations were selected for high-resolution map calculation (Fig.3.24).

For all high-resolution refinements, particles were re-extracted from the motion corrected micrographs with a $512^2$ pixel box (to allow for high-resolution CTF correction (Rosenthal & Henderson, 2003)). After initial auto-refinement, particle-based beam-induced motion correction and radiation-damage weighting (particle polishing) was performed (Fig. 3.25) (Scheres, 2014). Final

refinement was also performed in RELION, in this process two random half-datasets are refined independently in order to prevent the over fitting of noise. This allows so-called gold standard Fourier shell correlation (FSC) curves to be calculated such that a reliable estimate of overall map-resolution can be calculated (Scheres & Chen, 2012) (Fig.3.26). Finally, the two reconstructed maps were post-processed with a mask around the reconstructed volume to estimate the final resolution and to produce the final combined map. The map was then sharpened using a manually chosen negative B-factor (~100 $\text{Å}^2$) (Rosenthal and Henderson, 2003). After masking, the gold-standard Fourier Shell Correlation (FSC) fell below the 0.143 threshold at a resolution of 3.9 Å for class "open" and 4.6 for class "closed" (Fig.3.24 and 3.26).
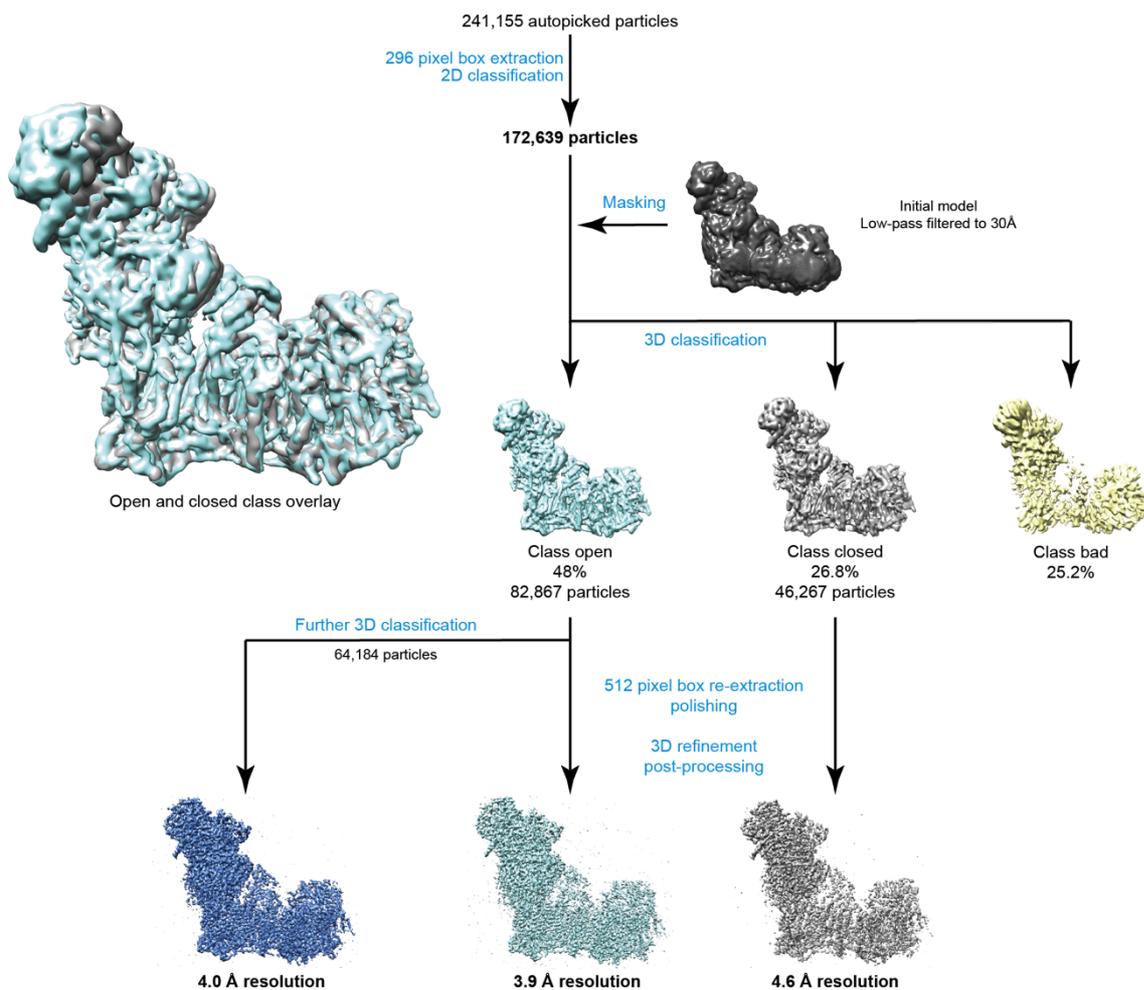


**Figure 3.24** Overview of 3D classification and refinement strategy.
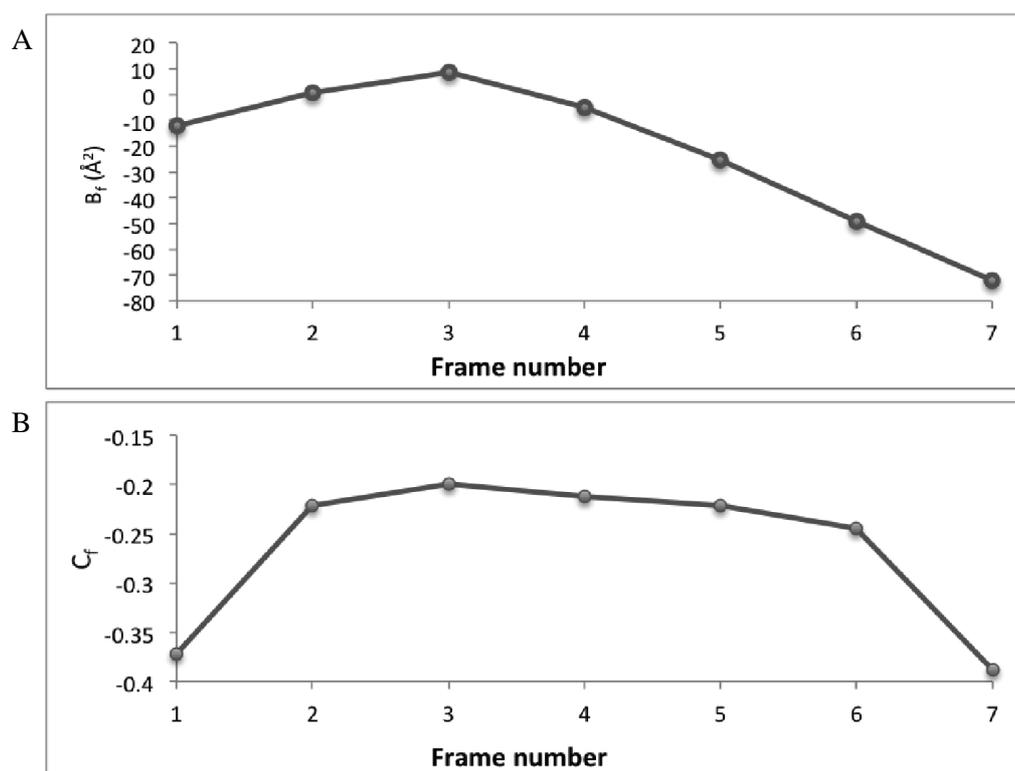
**Figure 3.25**Radiation-damage weighting from the RELION particle polishing procedure **A.** Relative B-factors (Bf) **B.** Intercepts (Cf).

This 3D class selection probably still allows for small variations in the conformation, therefore the local resolution varies within the maps, especially at the extremities of both arms. At the periphery of the molecule the resolution drops not only due to the usual decrease in the precision of particle alignments in these areas, but also due to differences in the protein conformation, greatest at the edges of the molecule. To overcome this limitation we performed 3D refinement focused on the peripheral and membrane domains of "open" class (since it gave the highest resolution) separately (with the subtraction of signal from the remaining parts of the complex (Bai et al., 2015)). This resulted in a 3.9 Å map of the peripheral arm (PD), very well resolved in all areas, including the edges of the domain (Fig. S2). The membrane domain (MD) refined to 4.1 Å, however, the map was more uniform and so better resolved for the distal part of the domain (near subunit ND5) as compared to the density from the refinement of the entire complex (Fig. 3.26 and S2).
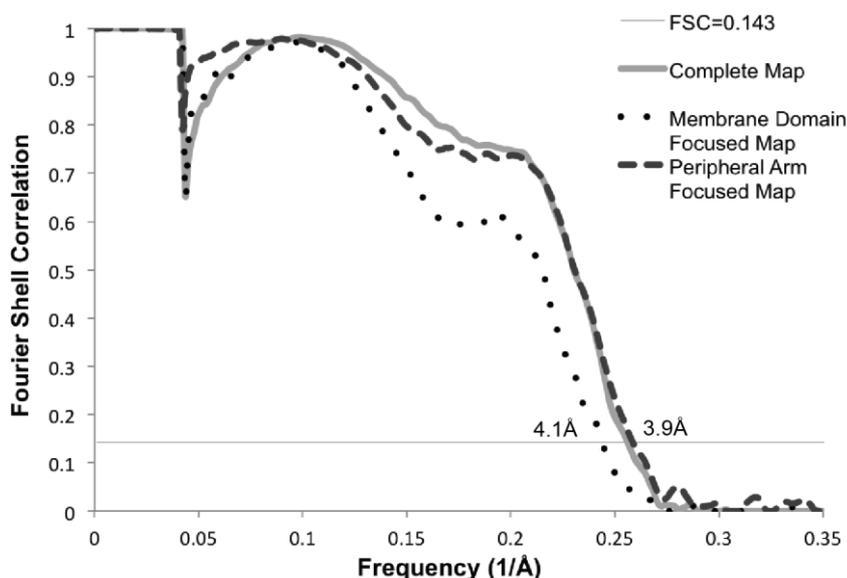
**Figure 3.26** Gold-standard (two halves of data refined independently) Fourier shell correlation (FSC) curves for the maps of the "open" complex I map (Complete Map) (resolution at FSC = 0.143 is 3.9 Å), membrane-arm-focused refinement (4.1 Å resolution) and peripheral-arm-focused refinement (3.9 Å resolution).

Higher quality refinement of the peripheral arm probably stems from the fact that high electron density of 8 FeS clusters helps in particle alignment. The least ordered part of the complex is the 42kDa subunit, loosely attached to the membrane domain. Extended 3D classification of the "open" class have been performed in order to identify the most homogeneous population, especially with respect to 42 kDa subunit. This class (64k particles) was refined to 4.0 Å, and the resulting density was used to model the 42 kDa subunit (Fig. 3.24 and S2). To assist with overall model building and refinement, several maps were carved around specific parts of the complex and combined into one map in UCSF Chimera (Pettersen et al., 2004): the peripheral arm from PD-focused refinement, the area around ND4/5 subunits (tip of the membrane domain) from the MD-focused refinement, the 42 kDa subunit density as above, and the rest of the complex from the overall 3.9 Å map for the "open" class (Fig. 3.27). The final model was refined against this combined map (Fig. 3.27).

A



B



75 kDa β9    30 kDa β5    TYKY Coil    ND2 TM3    ND4L TM2    ND5 TM3      PSST β-sheet

13 kDa Coil    18 kDa Coil    B17 Coil    PDSW α3    B9 TM    SGDH IMS Helix      39 kDa β-sheet
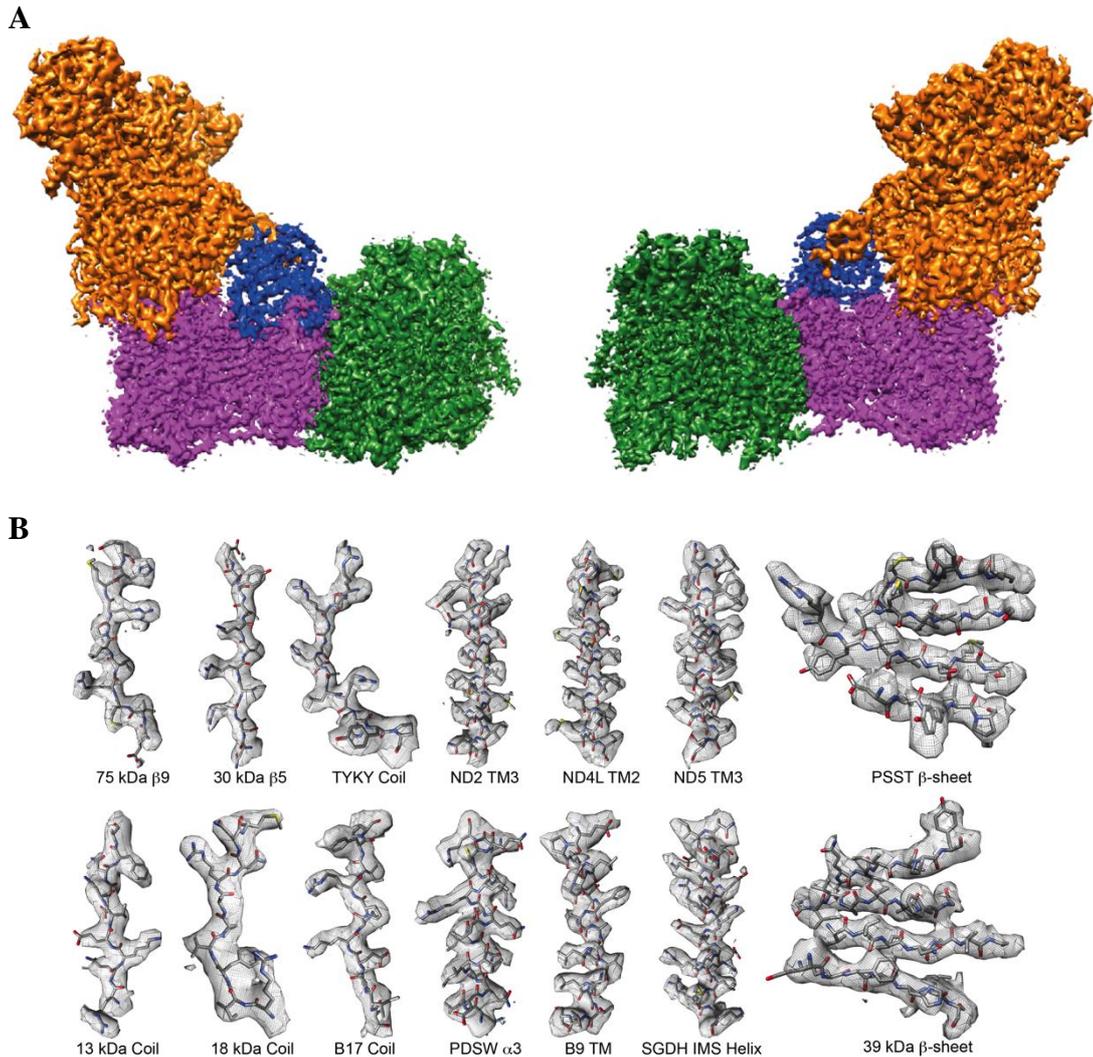
**Figure 3.27 A.** The final "combined" map was produced by combining maps with the best local resolution features, i.e. for peripheral arm – PA focused refinement map (orange), for the distal part of membrane domain – MD focused refinement map (green), for 42 kDa subunit – map from the selected homogenous complex I class (64k particles; blue) and the rest of the complex from the best map of the entire complex (magenta).**B.** Example coils, α-helices and β-sheets from core and supernumerary subunits. Cryo-EM density is shown with the model represented as sticks and coloured by atom with carbon in grey, oxygen in red, nitrogen in blue and sulphur in yellow. Panel B modified from one prepared by Dr James Letts.

The final map is of high quality, clearly showing large, medium and some small side-chains (Fig. 3.27 B). Carboxylates (Asp, Glu) have much lower density than other residues owing to early radiation damage, as observed previously (Allegretti, Mills, McMullan, Kuhlbrandt, & Vonck, 2014). Disulphide bridges also are subject to early damage, as in X-ray crystallography (Meents, Gutmann, Wagner, & Schulze-Briese, 2010). Map filtered to lower resolution is very similar

to the previous 5 Å resolution map for the bovine enzyme (Vinothkumar et al., 2014), suggesting that the mammalian complex I structure is very well conserved. One difference is that in the ovine complex the accessory 4TM subunit B14.7 is disordered in the detergent used for microscopy samples (Brij-35). It is likely to be disordered rather than detached as B14.7 was identified by mass-spectrometry in the sample used for EM (Table S2). For that reason the model for B14.7 subunit was build based on cryo-EM maps of ovine respiratory supercomplexes (Letts, Fiedorczuk, et al., 2016). In these maps all the subunits of complex I are well ordered, and so the final complex I model includes the poly-ALA structure of B14.7 based on 5.8 Å resolution map of the "tight" respirasome. Loss of B14.7 also results in the disorder of nearby C-terminal half of traverse helix HL and TM16 from ND5, as well as TM4 from ND6, which were also modelled as poly-ALA (designated as UNK residues in the model) based on the "tight" respirasome map. The density for the 42 kDa subunit is rather weak but this subunit clearly preserves the nucleoside kinase family fold, which allowed us to model most of it using Rosetta and visible large side-chains as a guide.

### 3.7.3 Separation of different complex I conformations

PD-focused, MD-focused maps significantly improved the quality of the protein density on the distal ends of complex I. This result also suggested that the initial 3D classification did not result in optimal separation of complex I conformations. Classes "open" and "closed" most likely contain particles that in fact belong to intermediate classes. In order to investigate this topic and obtain better separation of the particles the following approach was developed. Original classes "open" and "closed" were combined and 3D refined focusing on the peripheral domain (with the MD signal subtraction as previously (Bai et al., 2015)). Then the signal for the membrane domain was re-introduced and the particles were 3D classified masking only the membrane part of the protein without alignments. For best particles separation number of classes and regularization parameter tau (T) was empirically optimized to 10 and 20 respectively. Since positions of individual particles were fixed, they preserved orientations assigned during PA focused refinement and were separated based on the position of the MD (Fig. 3.28). The described protocol allowed for much

better separation of different complex I classes and resulted in five classes at final resolution varying between 4.0 - 4.5 Å. The highest resolution class of 4.0 Å (class 3) is very similar to the original class "open". The class most similar to class "closed" is class 1 of 4.5 Å resolution. In both cases the new classes have very similar resolution to the original "open" and "closed" classes however they have much less particles suggesting that they are much more homogenous (Table 4.1).
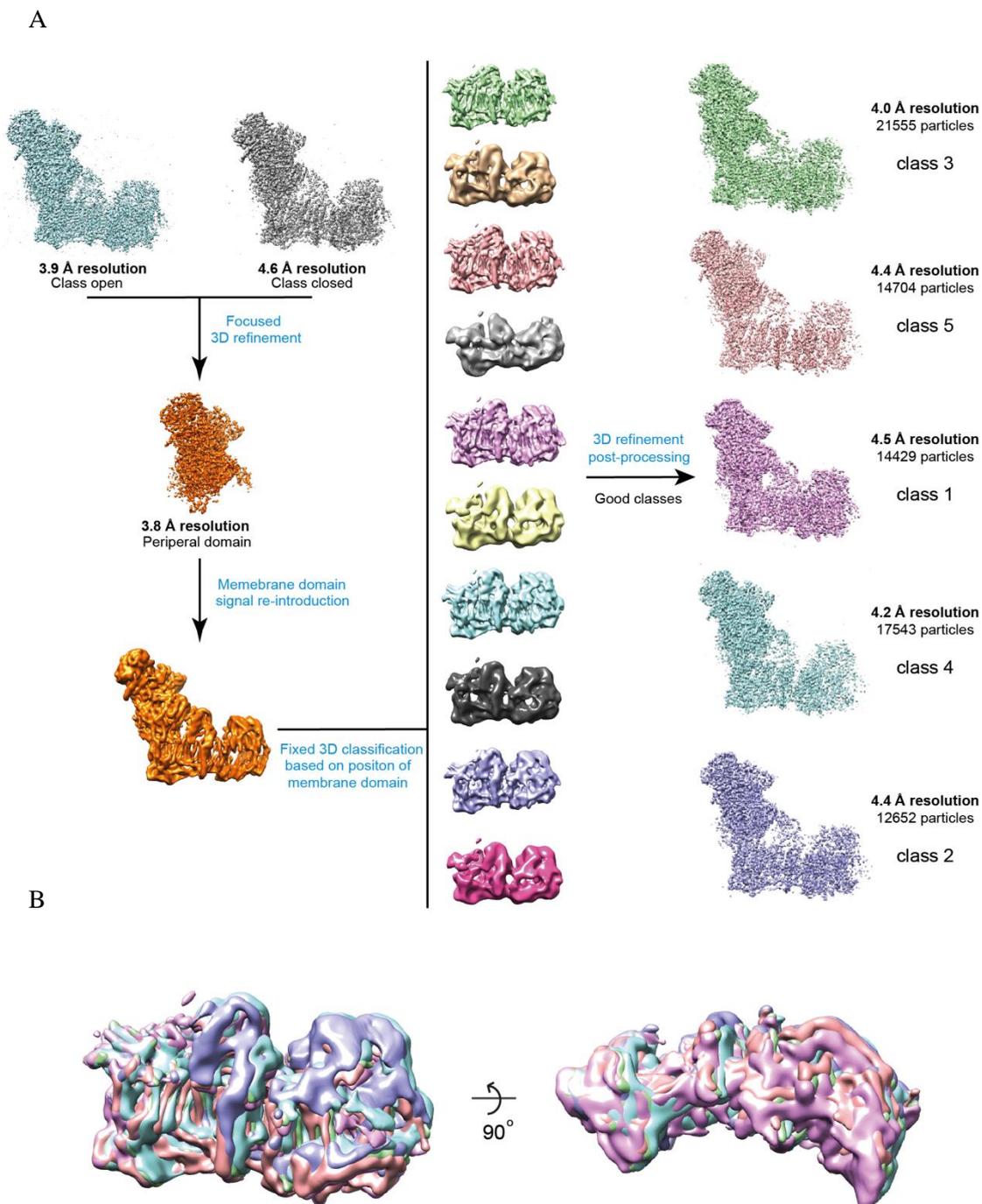
**Figure 3.28 A.** Overview of the improved 3D classification strategy and refinement strategy. **B.** Membrane domains of the five high resolution classes aligned to PA. Note that the membrane domain not only changes the position in wedge like movement (left) but also rotates (right). Five good classes were numbered 1 to 5 from most "closed" to most "open" class respectively (Chapter 4.2)

## 3.8 Cross - linking

Subunits assignment is fully consistent with the cross-linking experiments performed by the collaborators Dr Gianluca Degliesposti (MRC-LMB Cambridge, UK) and Dr Mark Skehel (MRC-LMB Cambridge, UK) (Chapter 2.8) and analysed by Dr James Letts (Fiedorczuk et al., 2016).

In total 218 unique cross-linked peptides were identified, of which 87 were between residues of different subunits (inter subunit), 73 were between residues within the same subunit (intra-subunit), and 58 were clear false positives (Table 3.8).

| Cross - linker | Number of experiments | Detergent | Spacer length (Å) | Approx. max length (Å) | Total cross - links observed | Inter-subunit cross - links | Intra-subunit cross - links | False positive cross - links | False positive rate |
|---|---|---|---|---|---|---|---|---|---|
| DSS | 4 | DDM | 11.4 | 23.4 | 174 | 82 | 67 | 25 | 0.14 |
| DSS | 2 | LMNG | 11.4 | 23.4 | 72 | 14 | 42 | 16 | 0.22 |
| DSS | 2 | LDAO/DDM | 11.4 | 23.4 | 72 | 37 | 20 | 15 | 0.21 |
| BS3 | 1 | DDM | 11.4 | 23.4 | 25 | 10 | 10 | 5 | 0.20 |
| DSA | 1 | DDM | 8.9 | 20.9 | 14 | 8 | 4 | 2 | 0.14 |
| SDH | 1 | DDM | 13.5 | 25.4 | 28 | 19 | 5 | 4 | 0.14 |
| ADH | 1 | DDM | 11.1 | 23.0 | 26 | 12 | 6 | 8 | 0.31 |
| | | | | Total Observed | 411 | 182 | 153 | 75 | 0.18 |
| | | | | Total unique | 218 | 87 | 73 | 58 | 0.27 |

Table 3.8 Summary of cross-linking experiments

False positives were identified by comparison to all known biochemical and structural information on complex I and the cross-links that are considered false positives are either between residues that are too distant from each other (>32 Å after allowing for exposed side chain flexibility from their modelled position), located on opposite sides of the membrane or the reactive residues are buried and not solvent accessible in the intact structure. Many of the false positive cross-links are found on unstructured coils at the edges of the complex I structure, indicating that they likely result from transient interactions between different molecules of complex I during the reaction (inter-complex cross-links). True positive cross-links were more likely to be observed in more than one experiment. Some high-scoring cross-links were observed between disordered termini or loops of subunits that could not be modelled in our structure, hence the accurate determination of distance for these cross-links was not possible. Good quality cross-links were observed for all supernumerary subunits except

for B14.7, KFYI and AGGG. No cross-links were observed for any of the mitochondria encoded core subunits, which are buried in the membrane and coated with a layer of supernumerary subunits (Fig. 3.29).



**Figure 3.29** Inter-subunit cross-link schematic. Core subunits coloured in cyan, previously assigned supernumerary subunits in magenta, newly assigned or newly built regions of supernumerary subunits in green, poly-alanine regions in orange and un-modelled regions in red. Observed cross-links are indicated by dashed black lines between either blue circles (lysine reactive cross-links) or red circles (acid reactive cross-links). No cross-links were observed to the core subunits of the membrane arm and hence they were omitted for clarity. The horizontal black lines indicate the approximate boundaries of the inner mitochondrial membrane. Subunits B14.7, B15 and ASHI are shown as being behind the membrane boundaries as they are found on the opposite (far) side of the membrane arm. Figure prepared by Dr James Letts.

## 3.9 Model building and refinement

Structures of highly conserved core subunits were build based on *T. thermophilus* X-ray structure (R. Baradaran et al., 2013). Using SQWRL4 software (Krivov et al., 2009) the bacterial model was modified to match ovine sequence. This model was manually adjusted and re-built to fit the experimental cryo-EM density using COOT (Emsley & Cowtan, 2004).

Positioning of accessory subunits was more challenging since in most cases their exact position was not known (Vinothkumar et al., 2014). In order to assign all supernumerary subunits unambiguously, data from mass spectrometric identification of chemically cross-linked peptides (XL-MS) was analysed (Chapter 3.8, Fig. 3.30, Table S8-9). These data in conjunction with the EM map allowed for correct positioning of all the subunits. Many of the previous assignments of supernumerary subunits were based on indirect biochemical data, secondary structure prediction and homology (B13, B15, B18, PDSW, 18 kDa, 15 kDa, B14.5band B17.2(J. Zhu et al., 2015)(Letts, Fiedorczuk, et al., 2016)), and thus, were tentative. Cross-links involving these subunits now provided experimental data confirming their positions. For example, single trans membrane domain (STMD) subunit B15 was tentatively assigned in the bovine structure to be associated with ND4 adjacent to helix HL based on its presence in subcomplex Iβ (containing core subunits ND4 and ND5) and the presence of comparable density in *Y. lipolytica*, which harbours a homologue of B15 (NB5M)(J. Zhu et al., 2015). This assignment has now been experimentally verified by a cross-link between B15 and PDSW in the inter membrane space.

The assignment of subunits PDSW and B18 to the helical density in the IMS has been confirmed by cross-links between PDSW, B15, ESSS and ASHI. The STMD subunit MNLL was identified by a cross-link to the B14.5b supernumerary subunit, containing two TMHs. MNLL binds to ND4 in the membrane near the interface with ND2. STMD subunits SGDH, B17, B12 and ASHI are known components of subcomplex Iβ (L A Sazanov & Walker, 2000) so are likely to surround ND4 and ND5. The assignment of the long helix in the IMS to SGDH is supported by cross-links to B14.5b, PGIV and the 15 kDa subunit. After the helix ends the C-terminus of SGDH extends along the IMS side of the complex as a coil interacting with the 15 kDa subunit and reaching to ND3 and ND6. The short N-terminal region of SGDH interacts with both ND4 and ND5 in the matrix. Therefore, SGDH spans nearly the entire length of the membrane arm of complex I from ND3 to ND5 (Chapter 4.12).

STMD subunit B17 has been identified in an assembly intermediate that lacks ND4 and ND5 but contains the ND6, ND3, ND4L and ND2 core subunits instead (Lazarou, McKenzie, Ohtake, Thorburn, & Ryan, 2007). Our cross-linking data unambiguously place B17 adjacent to core subunit ND5. The N-terminus of B17 contains an α-helix in the matrix that cross-links to the adjacent α-helix of B22, which binds on the matrix side of ND5.

Two cross-links between the N-terminus of B12 and SDAP-β place B12 at the tip of the membrane arm, adjacent to AGGG. Subunit ASHI has also been found in an assembly intermediate that lacks subunits ND4 and ND5(Vogel et al., 2007). Cross-links from the C-terminus of ASHI and the C-terminus of PDSW in the IMS place ASHI on the same side of the membrane arm as B15 and B14.7, claiming the last remaining unassigned TM helix density on this side of complex I. ASHI has a large N-terminal globular domain, and cross-links are observed between this N-terminal domain of ASHI and B15 and ESSS in the matrix. The density in this region is sufficient only for a poly-alanine model for the N-terminal domain of ASHI. The TM helix is partially disordered but a partial atomic model was build for the C-terminal half of the helix interacting with ND5. The C-terminus of ASHI in the IMS forms an extended coil, which nearly wraps around the long C-terminal helix of B18.

Previously there was ambiguity between the positions of B9 and MWFE, both STMD subunits known to associate with core subunit ND1. Cross-links between the N-terminus of PGIV and the C-terminus of MWFE and the N-terminus of TYKY and the N-terminus of B9 unambiguously place these subunits into the structure.

The 10 kDa subunit is known from biochemical studies to be part of subcomplex Iλ, which contains only the hydrophilic arm of the complex (L A Sazanov & Walker, 2000)(Finel et al., 1994)and to be part of an assembly intermediate made up of the N-module (Lazarou et al., 2007). Cross-links between the 10 kDa subunit and the 24 kDa and 51 kDa core subunits locate it to a cleft between those subunits. Supernumerary subunit B14.5a is also known to be a component of subcomplex Iλ(Hirst, Carroll, Fearnley, Shannon, & Walker,

2003)(L A Sazanov et al., 2000), but has not thus far been identified in any assembly intermediates. Cross-links of B14.5a with B17.2, B16.6, 75 kDa, 49 kDa, 30 kDa and B13 across its entire length (from Lys22 to Lys101) unambiguously place it in extended density along the surface of the Q-module.

In addition to these newly assigned subunits, the positions of large extensions to the previously assigned subunits were also identified by cross-linking. The C-terminus of PGIV extends as a coil along the IMS side of the matrix arm until it reaches the cleft between ND2 and ND4. This C-terminal extension is confirmed by cross-links between the C-terminus of PGIV, B14.5b and SGDH (Fig. 3.30). The position of the long N-terminal coil of PDSW is confirmed by an intra-subunit cross-link between PDSW residues Lys6 and Lys126. The N-terminus of B15 contains a long α-helix and its assignment is supported by a cross-link between the N-termini of B15 and adjacent ASHI. This B15 helix interacts with complex III in the supercomplex (Letts, Fiedorczuk, et al., 2016).

Several rounds of re-building and refinement resulted in high quality model in terms of geometry (Molprobity score 2.6 (corresponding to average structure at 2.6 Å resolution), very high EMringer score of 1.96 (model to map correctness validated by EMringer should result in scores above 1.0 for well-refined structures with maps in the 3 to 4Å range) (Barad et al., 2015) and fit to density (Table 3.9 and S3, Fig. 3.30).

| Statistics of complex I structure determination | | | | |
|---|---|---|---|---|
| Data collection | | | | |
| EM | Titan Krios 300kV, FEI Falcon II | | | |
| Pixel size (Å) | 1.39 | | | |
| Defocus range (μm) | -0.5 to -3.5 | | | |
| Reconstruction (RELION) | Overall | Membrane domain | Peripheral Arm | 42 kDa focused |
| Accuracy of rotations (°) | 0.573 | 0.711 | 0.728 | 0.732 |
| Accuracy of translations (pixel) | 0.308 | 0.400 | 0.400 | 0.430 |
| Final resolution (Å) | 3.9 | 4.1 | 3.9 | 4.0 |
| | | | | |
| Model refinement (PHENIX) | Complete model | | | |
| Resolution limit (Å) | 3.9 | | | |
| Residue number | 9510 | | | |
| Map CC (whole unit cell) | 0.758 | | | |
| Map CC (around atoms) | 0.782 | | | |
| rmsd (bonds) | 0.009 | | | |
| rmsd (angles) | 1.54 | | | |
| All-atom clash score | 27.67 | | | |
| Ramachandran plot | | | | |
| Outliers (%) | 1.07 | | | |
| Allowed (%) | 12.8 | | | |
| Favored (%) | 86.13 | | | |
| Rotamer outliers (%) | 0.09 | | | |
| Molprobity score | 2.50 | | | |
| EMringer score | 1.96 | | | |

**Table 3.9** Statistics of refinement
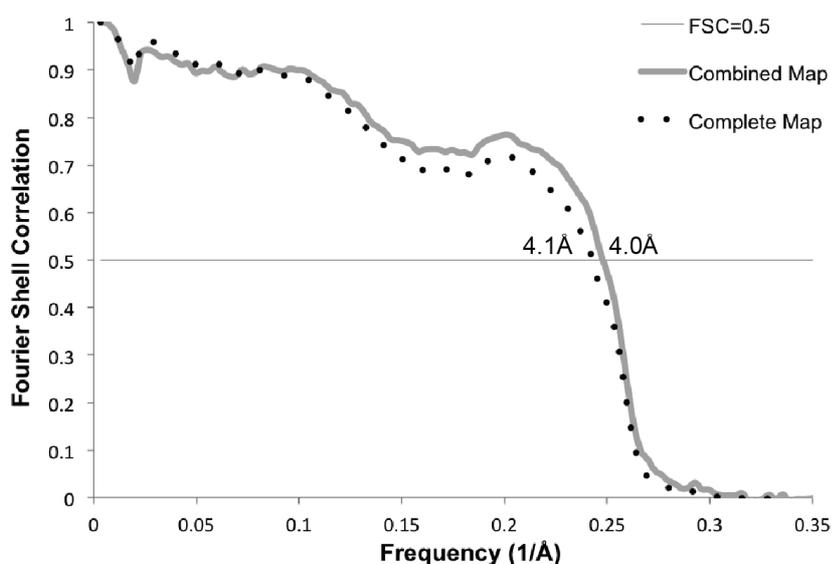


**Figure 3.30** FSC curve of the combined map vs. final model shows good agreement of the model with the map (FSC=0.5 at 4.0 Å resolution). FSC curve against the entire complex complete map, which was not used in refinement, is shown as a control

# Chapter 4. Discussion

## 4.1 Current knowledge on mammalian complex I structure.

The structure described in this work was published nearly simultaneously with the lower resolution (~ 4.2 Å) work on bovine complex I structure (J. Zhu et al., 2016). Our paper was submitted before the bovine work was published and so the identification and modelling of all supernumerary and core subunits in the structure was done by us independently. The bovine model has relatively low completeness (65% atomic for core subunits and 27% atomic for supernumerary subunits) due to lower resolution and so our model is the first nearly complete atomic model of mammalian complex I.

Later in 2016, a cryo-EM structure of porcine respiratory supercomplex at ~4.0 Å was reported, including an atomic model of complex I at 3.6 Å (Wu et al., 2016). This publication suffers from several drawbacks. These include the implausible proposal for complex III mechanism, dispensing with the well established Q-cycle. The paper describing the same supercomplex at lower resolution was published just a couple of months before (Gu et al., 2016) and remarkably, most EM maps describing the intermediate states of processing are of the wrong hand (mirror image) in both papers. Furthermore, even though the Cell paper was submitted one month after the publication of our structure, it purports to present the first structure of mammalian complex I, describing as novel all the complex I features we reported previously. In the porcine model (PDB 5GUP) the B-factors are set to random values and not refined, the FeS clusters geometry and environment are incorrect, the FMN isoalloxazine ring is flipped, and the statistics of model refinement or model completeness were not reported. Porcine PDBs were released only recently, 4-6 months after publications (Gu et al. and Wu et al. respectively) and we had to urge authors to

release focused EM maps, on which the models were actually built. Otherwise the structural features are well consistent with our model (backbone RMSD between ovine and porcine models for many subunits is only about 1 Å), showing the high degree of conservation of mammalian complex I structure. Despite higher overall resolution, there are no additional features observed in porcine structure, which were not reported for ovine enzyme (therefore the ovine structure remains the most complete fully refined model of mammalian complex I to date).

## 4.2 Model description

The structure of the 970 kDa complex comprises 45 different subunits, with a total of 78 transmembrane helices and a number of co-factors: FMN, NADPH, zinc and 8 iron–sulphur clusters. The core 14 subunits show conserved folds similar to bacterial homologues (L. A. Sazanov, 2015).

The ubiquinone chamber, like in the bacterial enzyme (R. Baradaran et al., 2013), is linked to the proton pumping modules by a central axis of charged residues suggesting a similar coupling mechanism. The structure shows that all the supernumerary subunits form a shell around the core and so probably mostly play a stabilizing role. Some of them, especially those containing additional co-factors (39 kDa, SDAPs, B14, B22 and 13 kDa) and phosphorylated residues (42 kDa, ESSS, MWFE, B14.5a, B14.5b and B16.6), may provide regulatory links to the redox status of the cell, lipid biosynthesis and mitochondrial homeostasis. Twelve identified bound lipid molecules likely stabilise interfaces between membrane subunits.

Intensive particles sorting and classification enabled distinguishing of five classes which differ by the relative position of complex I arms and accessibility of the ubiquinone binding site. In the most "open" classes it is blocked, suggesting the de-active state of the enzyme (Chapter 4.3) (Zickermann et al., 2015)(Fiedorczuk et al., 2016). However a subset of particles shows accessible, likely active conformations (Chapter 4.2) as in bovine (J. Zhu et al., 2016) and later porcine models (Wu et al., 2016).

The atomic model discussed in this chapter was build based on the so-called Combined Map (Chapter 3.6.2) (Fiedorczuk et al., 2016).

### 4.2.1 Core subunits

The architecture of the seven core subunits in the peripheral arm in *O. aries* is very similar to *T. thermophilus* complex used as an initial reference during model building (R. Baradaran et al., 2013). There are small relative shifts and rotations but the overall rigid body structure of the mitochondrial enzyme mirrors bacterial complex I.
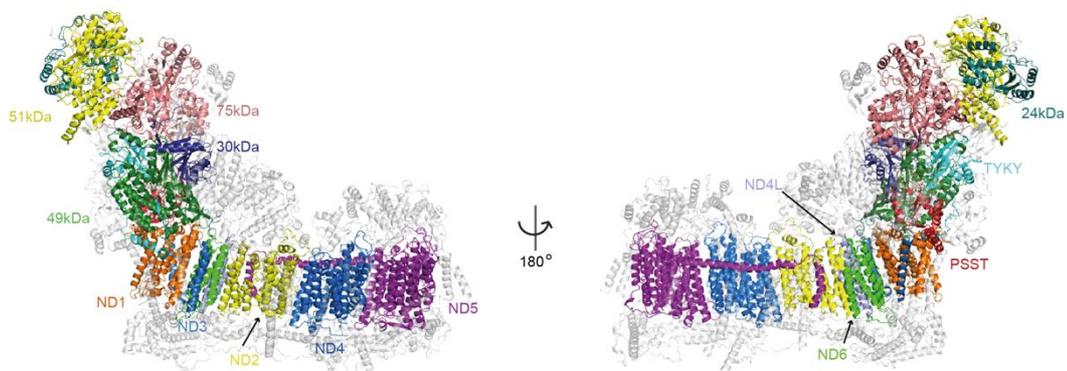


**Figure 4.1.** Structure depicted as a cartoon, with core subunits coloured and labelled, and supernumerary subunits in grey and transparent.

**Core peripheral arm subunits**

The peripheral arm of complex I (PA) can be divided into two functional modules: the N module responsible for NADH binding and transfer of electrons to the Q site and the Q module accommodating ubiquinone and triggering the coupling mechanism leading to proton translocation (in the P module).

The core of the N module consists of three conserved subunits (Fig. 4.2.A):

- 51 kDa subunit (harbouring the FMN, the [4Fe-4S] cluster N3)
- 24 kDa subunit (harbouring the [2Fe-2S] cluster N1a)
- 75 kDa subunit (harbouring the [2Fe-2S] cluster N1b and the [4Fe-4S] clusters N4 and N5)

The core of the Q module comprises four conserved proteins (Fig. 4.2.A):

- TYKY subunit (harbouring the [4Fe-4S] clusters N6a and N6b)
- PSST subunit (harbouring the [4Fe-4S] cluster N2)

- 30 kDa subunit (the only of the core subunits not binding ligands or co-factors)
- 49 kDa subunit (building the reduction site for ubiquinone)
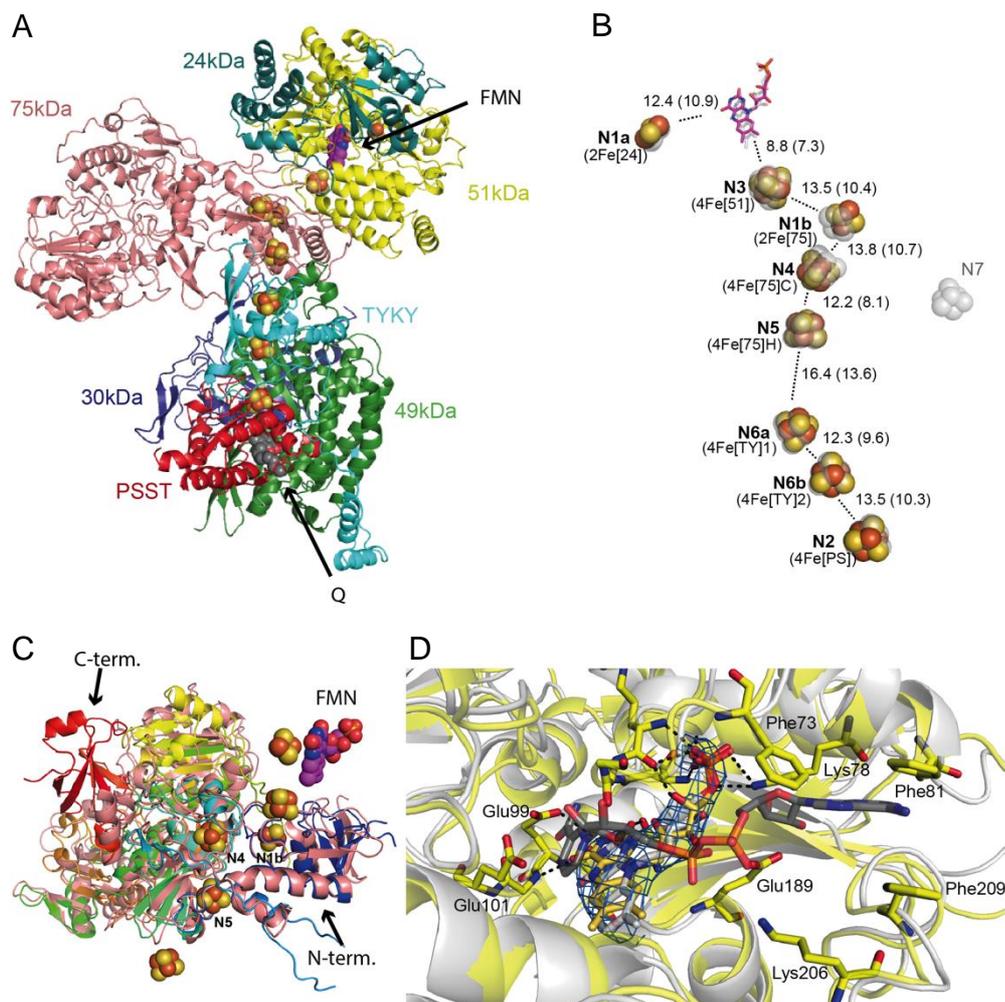
**Figure 4.2 A.** Structure of PA core subunits depicted as a cartoon coloured and labelled. Co-factors and quinone depicted as spheres. Q modelled based on *T. thermophilus* structure. **B.** Fe–S clusters are shown as spheres with centre-to-centre and edge-to-edge (in brackets) distances indicated in Å, overlaid with transparent grey depictions from *T. thermophilus*. Both traditional and structure-based (in brackets) nomenclature for clusters is shown. **C.** 75 kDa subunit (pink) overlay with *T. thermophilus* analogue (rainbow). FeS clusters co-ordinated by 75 kDa and termini of bacterial analogue labelled. **D.** NADH-binding site (overlay with *T. thermophilus* structure in grey, containing NADH). Cryo-EM density for flavin mononucleotide (FMN) is shown in blue. Key residues involved in interactions with FMN and NADH are shown as sticks. NADH modelled based on *T. thermophilus* reduced complex I structure (PDB 3IAM).

The core subunits of the mammalian and bacterial complex I vary mostly by the N and C terminal extensions. Compared to *T. thermophilus*, the biggest difference is folding of the C-terminus of the 75 kDa subunit (Fig. 4.2.C). Instead of a terminal β-barrel (*T. thermophilus*), the ovine molecule has two short α-helices separated by a roughly 30 amino acids loop. Extensions and

loops in TYKY, PSST and 30 kDa subunits are adjusted to better interact with supernumerary subunits. The N-terminal extension of the 49 kDa subunit (missing in *T. thermophilus*) is running along the membrane domain interacting extensively with the proton pumping module. This part is poorly resolved in the ovine model but is clearly visible in the bovine and porcine models (Chapter Comparison of the recent complex I structures) (J. Zhu et al., 2016).

The chain of the mitochondrial electron transfer starts from harvesting of two electrons from NADH (Fig. 4.2.B, 4.2.A) (L. A. Sazanov & Hinchliffe, 2006). The electron donor is oxidized at the FMN site at the 51 kDa subunit. Electrons are transferred along the FeS chain to the terminal N2 cluster and eventually to ubiquinone. All edge-to-edge distances in the redox chain are up to 14 Å, which is a maximal distance for electron transfer within physiological time scales (Fig. 4.2.B) (Page et al., 1999). Iron sulphur clusters are commonly regarded to be "stepping stones" for the electrons on the way from NADH to quinone. Most of them are co-ordinated by four cysteines in tetrahedral or planar geometry. However some clusters have a different environment, suggesting that they may have more prominent functions.

The N5 cluster is coordinated by 3 cysteines and one conserved histidine. It is separated by 13.6 Å from the next N6a cluster (the longest distance in the chain) and has a very low potential (Bridges, Bill, & Hirst, 2012). For that reason it is regarded to be a bottleneck on the path and can be considered as a control point for the overall rate of transferred electrons (L. A. Sazanov, 2015). The environment of the N5 cluster is composed of many polar and chargeable residues. Some of them like Asp104[75-kDa] directly interact with the supernumerary 13 kDa subunit, the only complex I subunit binding metal ion, $Zn^{2+}$ and proposed to be a redox sensor of the system (Fiedorczuk et al., 2016).

Another uniquely coordinated cluster is the terminal [4Fe-4S] N2 cluster. It is the last in the line and donates electrons to quinone (Fig. 4.2.B). It is coordinated by two conserved consecutive cysteines (54[PSST] and 55[PSST]), leading to strained geometry. The resulting flexibility of those cysteines may help link the conformation of neighbouring protein chains and redox potential of

the cluster. In *T. thermophilus* helices connected to this cluster shift upon cluster reduction (Berrisford & Sazanov, 2009)which may suggest catalytically driven conformation change in this area. Arg85[49-kDa], which undergoes rare symmetrical dimethylation (Carroll, Ding, Fearnley, & Walker, 2013), is in direct contact with the N2 cluster and is adding some hydrophobicity to the cluster's surrounding, probably also affecting its redox potential.

Function of the N1a cluster is not entirely understood. It is placed off the main electron transfer path but in a distance that enables electron transfer. It plays a structural role stabilizing the 24 kDa subunit (Birrell et al., 2013). However it may also assist in regulation of reactive oxygen species production at the flavin site (Chapter 1.4.3) (Esterhazy, King, Yakovlev, & Hirst, 2008)(L. A. Sazanov & Hinchliffe, 2006) (P. Holt et al., unpublished).

One of the most interesting features of the Q site of the ovine model is the conformation of theβ1-β2[49-kDa] loop. It is positioned in a way that disables ubiquinone access close enough to the N2 cluster to be reduced. The head group of the molecule can get only as close as 20Å, a distance too far to accept electrons. This conformation of the β1-β2[49-kDa] loop is very similar to the one described in the *Y. lipolytica* crystal structure (Fig. 4.3.A) (Zickermann et al., 2015). It has been proposed that this conformation reflects the de-active state of the enzyme. The recently published cryo-EM structures of the bovine and porcine complex I present states of the enzyme where the β1-β2[49-kDa] loop is retracted, enabling reduction of Q. This conformation is consistent with the crystal structure of *T. thermophilus* (R. Baradaran et al., 2013) and was proposed to represent the active form of the enzyme (Fig. 4.3.A). However, one of the low-resolution classes obtained after extensive re-classification of particles used in the ovine study (Chapter 3.6.3) also shows such a state of the enzyme. Based on this finding it can be assumed that the sample used for the cryo-EM studies of the complex contained particles in the de-active and active conformation in uneven proportion (Fig 3.27). Alternatively, it is possible that different observed conformations reflect intermediate steps in the catalytic cycle (randomly driven by thermal energy before the protein is snap-frozen). In this

case the movement of the β1-β2$^{49\text{-kDa}}$ loop into the cavity may help to eject reduced ubiquinol into the membrane upon the completion of the catalytic cycle.

The Q cavity is mainly created by four loops: β1-β2$^{49\text{-kDa}}$, TMH5-6$^{ND1}$, TMH1-2$^{ND3}$ and α2-β2$^{PSST}$. Except for several hydrophobic amino acids in β1-β2$^{49\text{-kDa}}$ the cavity is mostly covered with polar residues (Fig. 4.3.B).

In the ovine structure we can see that Arg77 at a very long loop α2-β2$^{PSST}$ is partially blocking the Q tunnel; interestingly this arginine undergoes a very rare hydroxylation modification (Carroll et al., 2013). The exact role of this amino acid is not well understood.
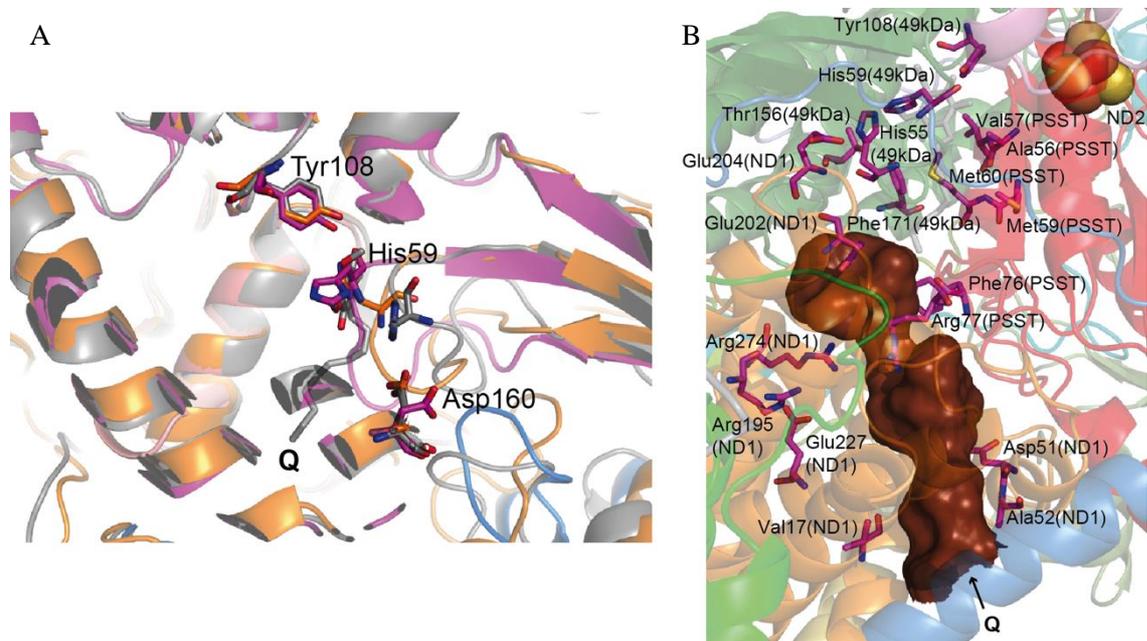


**Figure 4.3** Quinone-binding site. **A.** Key β1–β2$^{49\text{-kDa}}$ loop (magenta) deviates from bacterial structure (grey) and is more similar to *Y. lipolytica* (orange), clashing with the decyl-ubiquinone (Q) head group position in *T. thermophilus* **B.** Environment surrounding the Q cavity (brown surface, entrance point indicated by an arrow), with some of the functionally important residues shown as sticks and labelled subunit names in brackets.

**Core membrane domain subunits**

Seven core subunits encoded by mitochondrial DNA are embedded in the inner mitochondrial membrane (L. A. Sazanov, 2015):

- ND1 subunit (composing quinone binding site, the first proton translocation site and involved in coupling between the Q and P modules)

- ND3 subunit (composing the first proton translocation site and involved in coupling between the Q and P modules)

- ND4L subunit (composing the first proton translocation site and involved in coupling between the Q and P modules)

- ND6 subunit (composing the first proton translocation site and involved in coupling between the Q and P modules)

- ND2 subunit (composing the second proton translocation site)

- ND4 subunit (composing the third proton translocation site)

- ND5 subunit (composing the fourth proton translocation site)
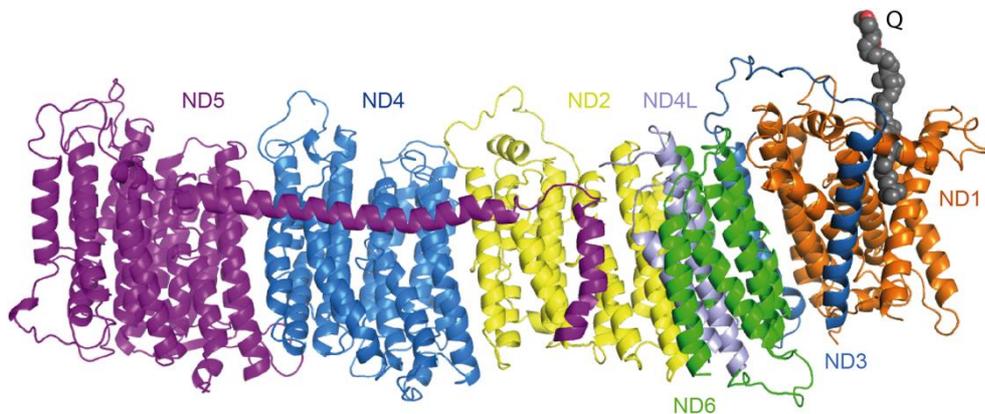


**Figure 4.4** Structure of MA core subunits depicted as a cartoon, coloured and labelled. Note that the Q reduction site is over 10Å above the membrane level and entry site in ND1 subunit. The long chain ubiquinone modelled based on the *T. thermophilus* structure is shown in grey.

Most of the membrane arm comprises of three homologous anitiporter-like subunits: ND2, ND4 and ND5. Each is build of 14 trans membrane helices (TMH), except of the ND2 subunit, which in metazoa lost first three TM helices. In spite of that, for clarity reasons, I will refer to it as 14 TM helices protein so numbering of ND2 helices follows 14 helices system and is consistent with ND4 and ND5. The cavity that arose as a result of deletion of 1-3 TMH of ND2 has

not been filled by any supernumerary subunits (Fiedorczuk et al., 2016), neither evolved to interact with other respirasome complexes (Letts, Fiedorczuk, et al., 2016). It is densely covered with specifically bound lipids molecules, including cardiolipin, which help to stabilize this site.

Distal ND5 subunit contains a long amphipathic lateral helix spanning ND4 and ND2 subunits and ends with TMH 16, locked in between ND2, ND4L and B14.7 subunits. This amphipathic lateral helix most likely does not play catalytic role (Chapter 1.4.3) (R. Baradaran et al., 2013)(G. Belevich et al., 2011) but acts as a scaffold stabilizing the trans membrane part of the complex from the matrix side. All three anitporter-like subunits are responsible for proton pumping across the inner mitochondrial membrane. Structure of this part of the protein is consistent with bacterial models. It confirms that the $H^+$ pumping machinery is very conserved. Each of the antiporter-like subunits contain symmetry related two half-channels composed by TMH 4-9 and TMH 9-13. They are open respectively to the matrix and IMS, and are connected in the middle of the membrane by conserved charged residues, which enable proton translocation (Efremov & Sazanov, 2011). This architecture of MD indicates that each antiporter-like subunit translocates one proton per catalytic cycle (total three). One more proton is likely transferred by the so-called E channel (R. Baradaran et al., 2013) with a similar architecture of two half-channels. It comprises ND3, ND4L and ND6 subunits contributing to the IMS half-channel and TMH 2-6 of ND1 forming the matrix side half-channel (Fig. 4.4). That gives a total number of four channels, each translocating one proton per one molecule of NADH oxidised.

The half channels are connected through protonatable residues in the middle of the membrane (L. A. Sazanov, 2015): Lys in TMH7 that belongs to the matrix half channel, Lys or His from TM8 that connect both half channels and Lys or Glu from TMH12 of IMS half-channel. Additionally, transport may be facilitated by conserved glutamic acids in TMH5. They are positioned in a way that enables interactions (affect pKa's) with Lys of TM7 and Lys/Glu of TMH12 from the neighbouring subunit. This feature is present in bacteria as well.

In the E channel possible analogues of Lys TMH7 and Glu TMH5 from ND2, ND4 and ND5 are respectively Glu192 and Glu143 of ND1 (Fig. 4.5). The antiporter role of TM12 is most likely fulfilled by TMH3 of ND6. Essential Tyr60 of ND6 and Glu34 of ND4L are placed in close proximity and also maybe involved in proton translocation or coupling mechanism. Those key residues are all included into a conserved continuous axis of charged and polar amino acids that extends from the Q site to the end of the membrane domain (Fig. 4.5).

It has been recently suggested that additional input channels for protons from the matrix may exist (L. A. Sazanov, 2015). Protonation of the central axis could happen through conserved Glu of TMH5 and along central axis of TM8 in the input channels of ND2, ND4 and ND5 pumping modules (Fig. 4.5).

Preserved from bacteria are also π–bulge like distortions in the TMH of the membrane domain (R. Baradaran et al., 2013). They are present at TMH3 of ND6 and TM helices 8 of ND2, ND4, and ND5. Distortions are formed at glycine pairs 62-63 and 239-240 in ND6 and ND4 respectively, serine pair 249-250 in ND5 and tryptophan 167 flanked by two glycine pairs in ND2.
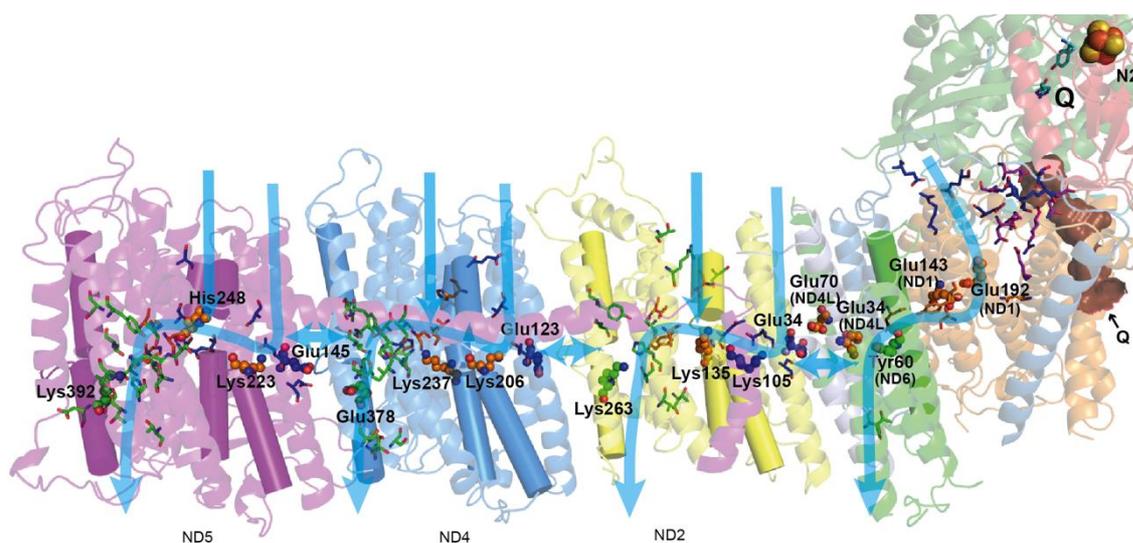
**Figure 4.5.** Structure of the core subunits of ovine complex I, coloured as in Fig. 1b, with polar residues in proton channels shown as sticks, with carbon in blue, orange and green for input, connecting and output parts, respectively. Key residues, Glu (TM5), Lys (TM7), Lys/His (TM8) and Lys/Glu (TM12) from the antiporters and the corresponding residues in the E-channel (near Q site), are shown as small spheres and labelled. These residues sit on flexible loops in discontinuous transmembrane helices shown as cylinders. Polar residues linking the E-channel to the Q cavity (brown) are shown in magenta. Tyr108[49-kDa] and His59[49-kDa] are shown in cyan near the position of bound Q in bacteria. Possible proton translocation pathways are indicated by blue arrows.

It has been reported that conserved cysteine 39 in TMH1-2[ND3] loop becomes accessible to chemical modifications only in the D state (de-active) of the enzyme (A. Galkin et al., 2008). This indicates that Cys39[ND3] or even the whole loop is important for the active/de-active state transition. It is difficult to judge if Cys39 is available for any modifications in our model because this residue is partially covered by the poly-alanine chain of ND6 built at low resolution. However, the comparison of the model with other mitochondrial complex I structures indicates a big displacement of the loop correlated with the movement of PA (Fig. 4.28).

*4.2.2 Supernumerary subunits*

Supernumerary (also called accessory) subunits are crucial to the structure, stability and assembly of mitochondrial complex I. Although they are not directly involved in catalytic turnover of the enzyme, they are proposed to play

regulatory and redox sensing roles (Fiedorczuk et al., 2016). Accessory subunits are solely encoded by the nuclear genome, translated in the cytoplasm and imported into mitochondria. The number and composition of accessory subunits varies across the phylogenetic tree (Gabaldon et al., 2005). The general rule is that their number grows throughout evolutionary time, from none in bacteria to 30 unique proteins in mammals. Most supernumerary subunits in complex I associate with either matrix or membrane domain. The exception is B16.6 that mostly binds to the MD but also has a long N-terminal extension interacting with the PA. In the previously published maps (Vinothkumar et al., 2014)(Zickermann et al., 2015), supernumerary subunits were impossible to assign due to low resolution. The 3.9Å ovine map enabled assignment of all of the subunits, additionally confirmed by extensive cross-linking experiments (Chapter 3.8) (Fiedorczuk et al., 2016). Based on the location we can divide the supernumerary subunits into three groups (Fig. 4.6):

- Subunits associated with the peripheral domain (10 kDa, 13 kDa, 18 kDa, 39 kDa, B8, B13, B14, B14.5a, B17.2, SDAP-α)
- Subunits associated with the membrane domain (42 kDa, 15 kDa, B18, B22, PDSW, PGIV, SDAP-β)
- Subunits associated with the membrane domain containing trans membrane motifs (B9, B12, B14.5b, B15, B16.6, B17, AGGG, ASHI, ESSS, KFYI, MNLL, MWFE, SGHD)

**Figure 4.6A.** Cryo-EM density coloured by subunit, with core subunits in grey (left–right view) **B.** Structure depicted with core subunits in grey and supernumerary subunits coloured and labelled (left–right, IMS–matrix views). Amphipathic helices at the 'heel' of the complex, probably attached to the lipid bilayer, are indicated as AH.

## Supernumerary subunits associated with the peripheral domain

B17.2 and B14.5a (together with TM B16.6 subunit) subunits stabilize the PA/MD interface. They contain N-terminal amphipathic α-helices bound at the surface of the membrane (Fig. 4.6.B and 4.7-8). The N-terminus of B17.2 is formed by two short perpendicular α-helices partially buried in the matrix side of the IMM, very close to the ND1 and PSST core subunits at the "back" of complex I. The α-helices are followed by a coil that at the level of the PSST subunit that becomes a three-strand β-sheet, almost parallel to the membrane

plane. From that point the protein snakes through the groove between the 39 kDa and TYKY subunits, and underneath the 13 kDa subunits up to 75 kDa subunit. Thus it connects the mitochondrial membrane, Q and N- modules (Fig. 4.7 and 4.8). The B17.2 subunit is incorporated into complex I at a very late stage of maturation (Chapter 1.4.2). It associates in the localization previously occupied by a homologous assembly factor B17.2L (Lazarou et al., 2007). Upon addition of the N module B17.2L leaves the complex, B17.2 subunit binds and 13 kDa (already bound to N module) wraps on top of the B17.2 C-terminal coil (Chapter 1.4.2).
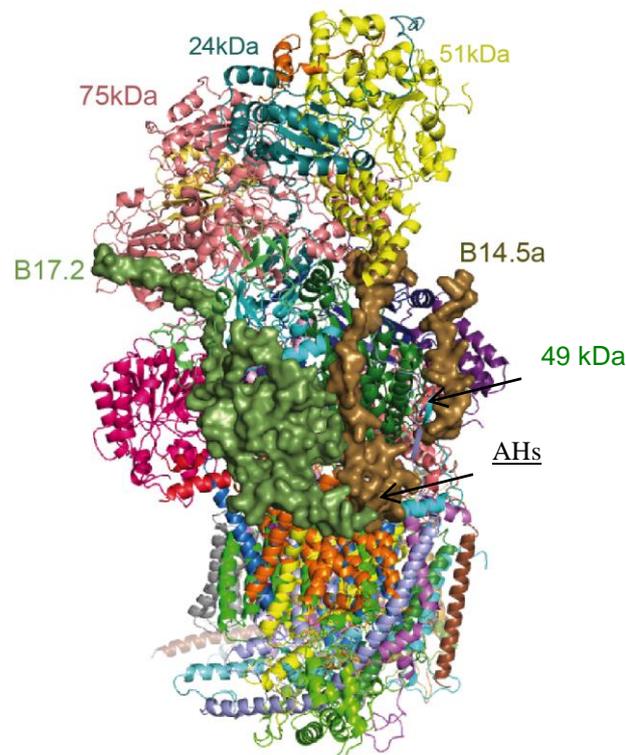


**Figure 4.7** Back view of complex I. The surfaces of the B17.2 and B14.5a subunits are shown. AH indicates position of the amphipathic C-termini of the proteins.

The B14.5a subunit just like B17.2 bridges the surface of IMM, Q and N modules. It starts with a short amphipathic α-helix on the N-terminus (Fig. 4.6.B). Then it winds between the PSST and 49 kDa subunits up to the N module where it turns in the direction of SDAP-α subunit. At the level of B13 the density for B14.5a disappears because residues 71-89 form a flexible loop. The C-terminus of the protein runs down the Q module between the B13, 30 kDa

and 49 kDa subunits, to finish not far from where the N-terminal α-helix is localized. Thus B14.5a subunit wraps around the whole Q module from the 49 kDa solvent exposed site (Fig. 4.7).

The space between amphipathic α-helices of B17.2, B14.5b and PSST is accommodated by a cardiolipin molecule that probably helps to stabilize this region (Fig. 4.23).
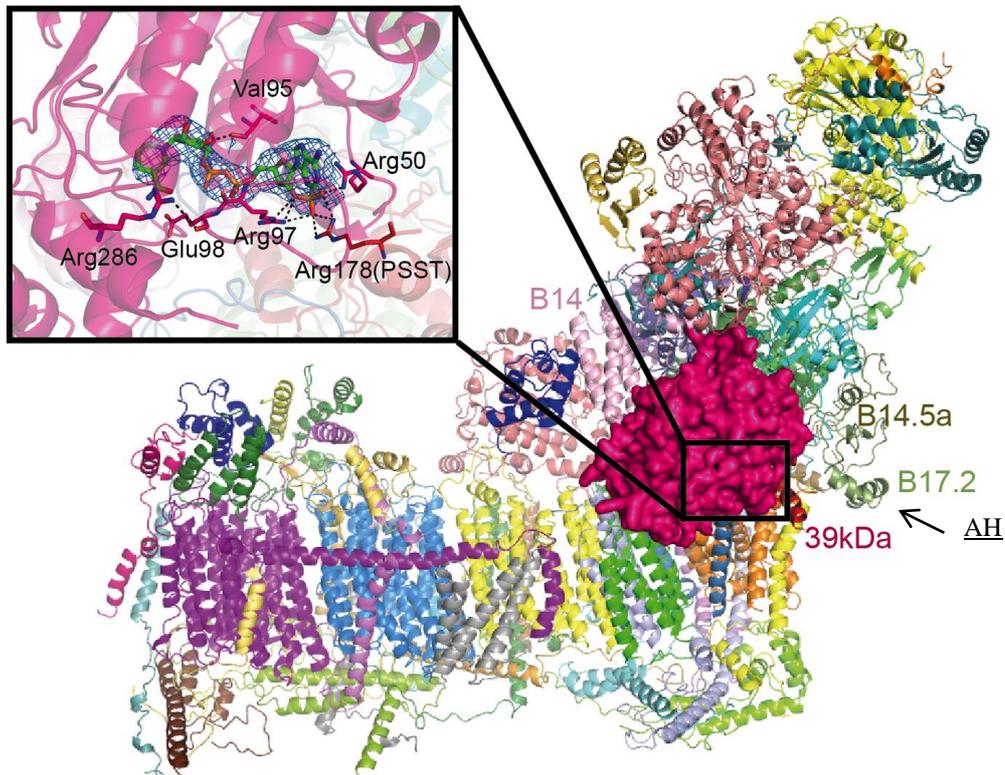


**Figure 4.8**Side view of complex I. The surface of the 39 kDa subunit is shown. Notice the position of the B17.2 subunit. Top left panel shows the NADPH binding site with the interacting residues. NAPH is shown with experimental cryo-EM density carved to within 5 Å.

The 39 kDa subunit is presumably involved in securing the right position of PA over the lipid bilayer and in sensing the redox state of the environment (Fiedorczuk et al., 2016). It belongs to the short-chain dehydrogenase/reductase family (Fearnley & Walker, 1992) although it has been shown not to display any catalytic activity (Abdrakhmanova, Zwicker, Kerscher, Zickermann, & Brandt, 2006). The protein binds close to the junction between MD and PA, interacting with the PSST and TYKY subunits (Fig. 4.8). It can possibly provide a support to PA through a hydrophobic loop (amino acids

187-194[39-kDa]) localized very close to the membrane. In the native environment that part may interact with the membrane surface. The α/β short-chain dehydrogenase/reductase motif of the protein contains an NADP(H)-binding Rossmann fold with a parallel seven-strand β-sheet accommodating a NADP(H) molecule. The co-factor plays a structural and stabilizing role but is not likely to be redox-functional since the nicotinamide ring of the molecule is 45Å away from the nearest FeS cluster. The NADP(H) molecule is trapped in between the central seven-strand β-sheet and a short peripheral α-helix, thus is absolutely crucial for the protein integrity. The co-factor is also coordinated by an Arg178[PSST] linked to PSST helix that interacts with the N2 cluster and undergoes conformational change upon its reduction (Berrisford & Sazanov, 2009). It is possible that such architecture allows the complex to "sense" the redox state of the matrix through NADPH/NADP⁺ ratio and communicate the information to N2 cluster. The 39 kDa subunit also interacts with the N-terminal α-helix of the 13 kDa subunit, another subunit proposed to be a redox sensor.
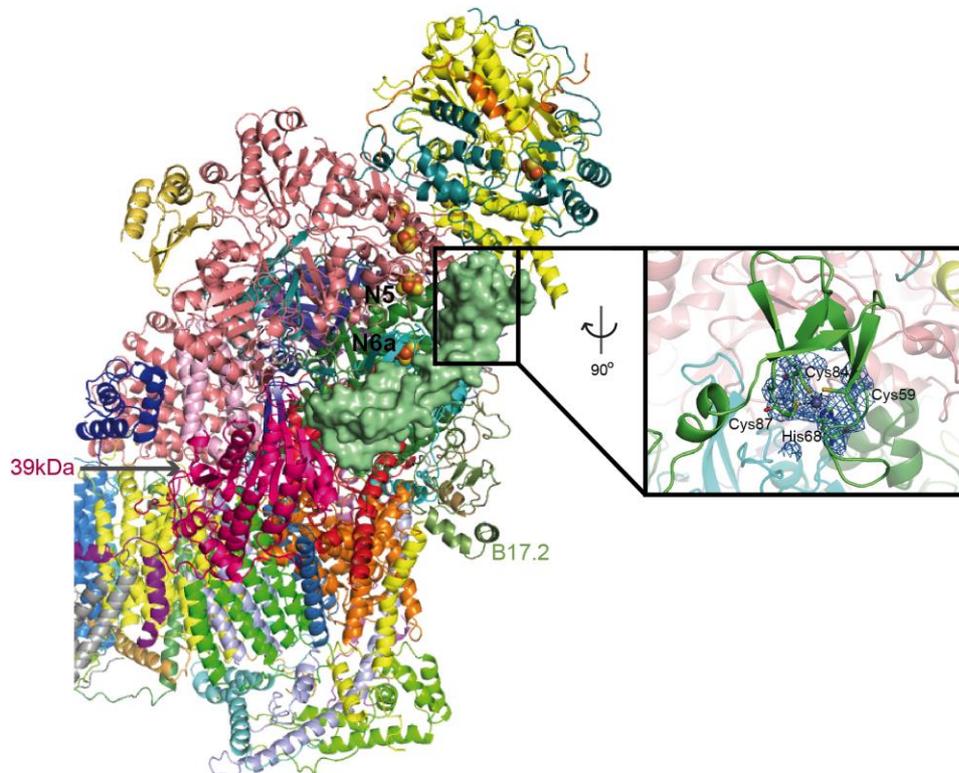


**Figure 4.9** Back/side view of complex I. The surface of the 13 kDa subunit is shown. Right panel shows the zinc-binding site with the interacting residues. The closest FeS clusters to the metal binding site are N5 and N6a are labelled. Zinc is shown with experimental cryo-EM density carved to within 5 Å.

At the N-terminus the 13 kDa subunit binds to the B17.2 subunit and forms an exposed coil running to the 75 kDa subunit linking the Q and N modules. At the C-terminus the protein forms two beta-sheets coordinating a $Zn^{2+}$ ion through four conserved residues: three cysteines and a histidine (Fig. 4.9). The metal ion is 13.4Å away from the N6a cluster and 15.3Å from the N5 cluster. It is not likely that $Zn^{2+}$ takes part in electron transfer. Although it is in the appropriate distance, it has been shown to be redox inactive (Oteiza, 2012). The $Zn^{2+}$ is likely to play a structural role, although some influence on the N5 and N6a clusters cannot be ruled out. It has been shown that loss of the13 kDa homologue in *Y. lipolytica* results in loss of EPR signal for the N6a cluster, destabilization of the N5 cluster and loss of the B17.2 subunit homologue (Kmita et al., 2015). As discussed above the 13 kDa subunit clamps on top of B17.2 and extensively interacts with TYKY, stabilizing both. Its absence would cause disassociation of the B17.2 subunit and exposure of the N6a cluster to the solvent. Zinc-binding proteins are known to often be redox state sensors (Maret, 2005) and conserved Tyr86[13-kDa] of the zinc-binding domain interacts with the residues coordinating the N5 cluster, so any changes at the metal binding site of the 13 kDa subunit may directly influence the cluster. Additionally, the13 kDa subunit interacts with the NADPH-binding 39 kDa subunit, another possible sensor for the redox state of the mitochondrial matrix. The unusual coordination, environment and interacting partners of the N5 cluster perhaps enable it to "sense" the redox state of the whole system or downstream clusters and adjust the electron flow.

The B13 subunit forms a three-helix bundle placed on the 30 kDa subunit with an extended C-terminal coil anchored on the interface of the 30 kDa, 49 kDa and 18 kDa subunits (Fig. 4.10). It does not seem to play any clear structure-stabilizing role like the B17.2 or 13 kDa subunits. On the other hand in the "closed" complex I map it comes very close to 42 kDa subunit and thus may help stabilise this state. It has been also implicated previously in the A/D transition (Ushakova, Duarte, Vinogradov, & Videira, 2005).
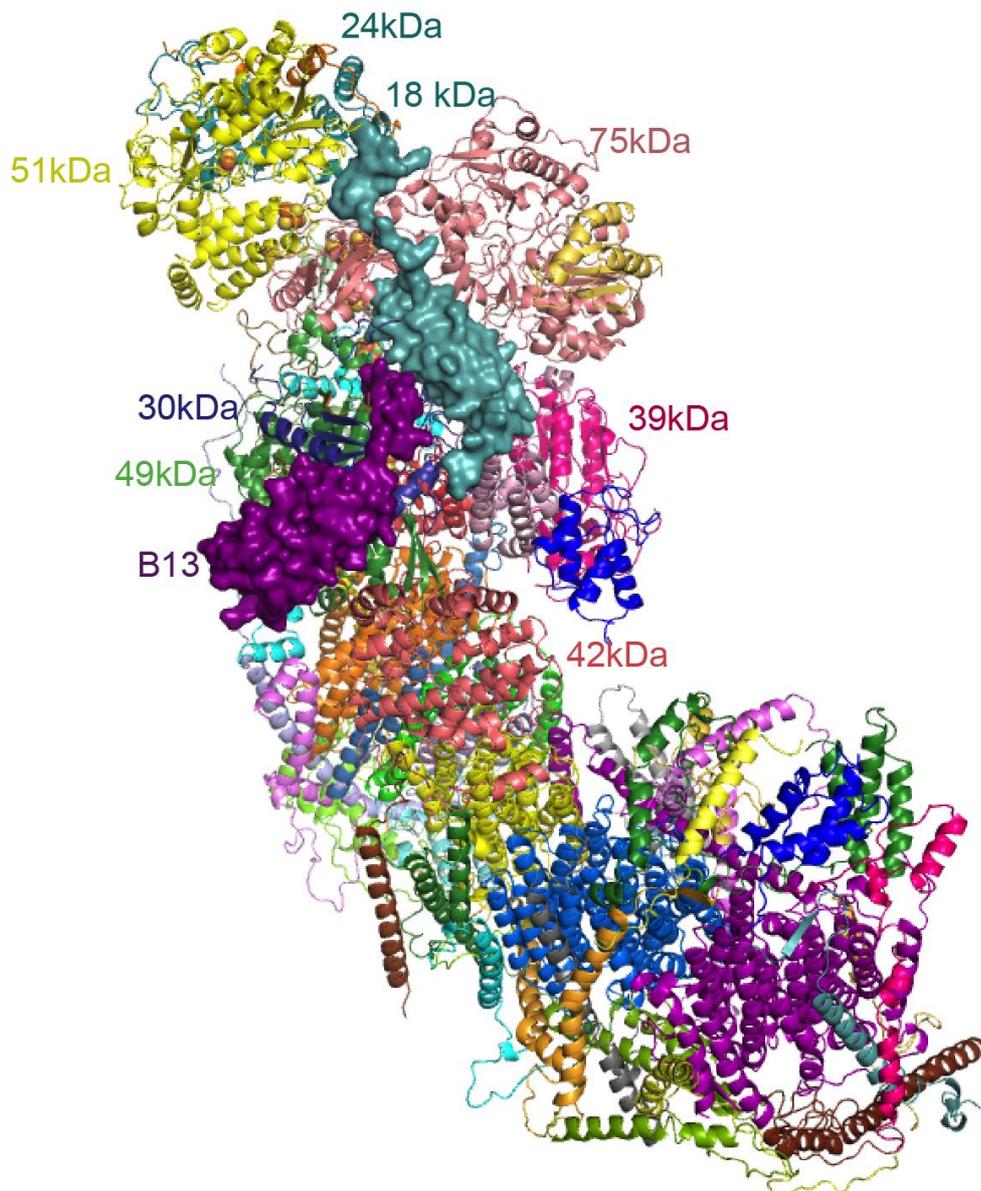
**Figure 4.10** Matrix side view of complex I. The surface of the B13 and 18 kDa subunit is shown.

On the opposite side of the 42 kDa subunit, also bound to the 30 kDa subunit, an ACP (acyl carrier protein)/LYR (leucine-tyrosine-arginine motif family) couple, the dimer of SDAP-α and B14 subunits, is located. It forms a large sub-domain placed between the 42 kDa and 39 kDa subunits (Fig. 4.11). There are two copies of ACP in mammalian complex I, SDAP-α and -β. Both interact with LYR family proteins: B14 and B22 respectively. Although it is not clear if SDAP proteins have any catalytic activity when incorporated into complex I, they definitely play a stabilizing role. In general ACPs are involved in fatty acid

production (Byers & Gong, 2007). They work as shuttles, transporting growing acyl chain between different enzymes. The chain is attached to a phosphopantetheine moiety covalently bound to Ser44. During the transport, the acyl chain is hidden in a hydrophobic cavity of the three ACP α-helices. Upon interaction with a fatty acid synthesis enzyme, the chain "flips out" and is accommodated into the cognate enzyme active site. The map clearly displays the density for phosphopantetheine in the "flipped out" position for both ACP/LYR pairs. Additionally for SDAP-α we see a 12-carbon long extension carried by phosphopantetheine predicted by mass-spectroscopy experiments (Carroll et al., 2005).

The B14 subunit has a similar fold to the B13 subunit, a three α-helix bundle and extended C-terminal coil docking it between the 39 kDa, 75 kDa and 18 kDa subunits (Fig. 4.11). The N-terminal bundle interacts with the TMH1-2[ND3] loop shown to change conformation upon A/D transition (Babot & Galkin, 2013). This fact together with the presented structure explains why deletion of the B14 subunit yields fully assembled (of course lacking also the SDAP-α subunit) but inactive complex (Angerer et al., 2014). It is tempting to speculate that since the acyl chain is required for the coupling of the ACP/LYR pair, in this way the OXPHOS system in mitochondria may be linked to the production of lipids.
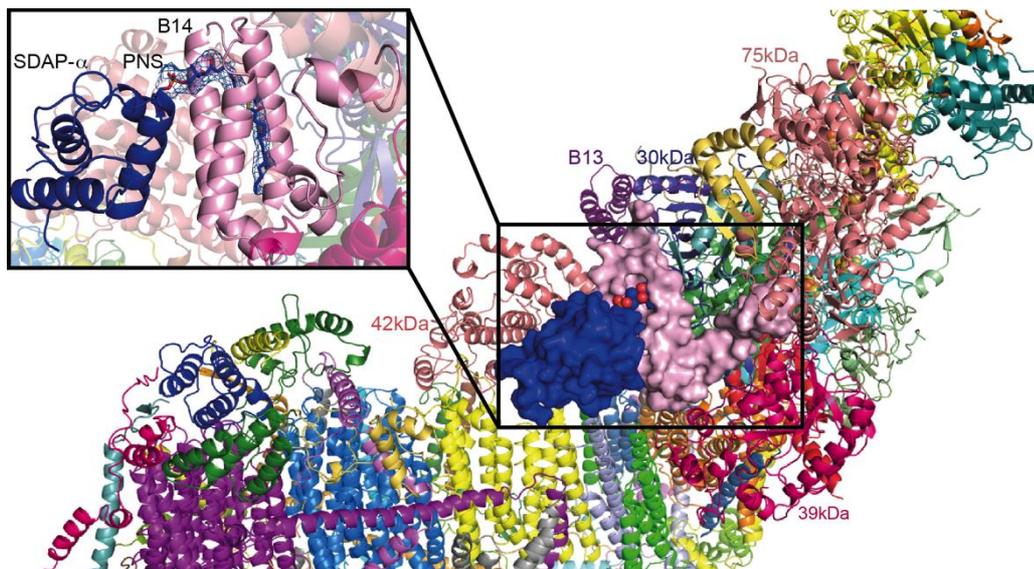
**Figure 4.11** Matrix/side view of complex I. The surface of the SDAP-α and B14 subunits is shown with PNS shown as spheres. Left-top panel shows PNS in SDAP-α site with the experimental density for the co-factor. PNS is shown with experimental cryo-EM density carved to within 5 Å. Notice that the angle of the panel differs slightly to better show the acyl chain.

The 18 kDa subunit is also interacting extensively with other subunits and improving integrity and stability of the enzyme. It has two long coils on the N- and C-termini, wrapping around the B14 subunit and binding between the 75 kDa, 51 kDa and 24 kDa subunits respectively (Fig. 4.10). The central part of the protein contains a four-stranded β-sheet, buried deeply in a cleft between the 49 kDa, 75 kDa and 30 kDa subunits. Additionally, a long β2-β3 loop runs under the 75 kDa subunit to the TYKY and 39 kDa subunits. That explains why patients with certain mutations in the 18 kDa subunit accumulate late stage complex I subassemblies lacking the N-module (Chapter 4.4.3) (Lazarou et al., 2007).

B8 is a small protein with a thioredoxin fold (Brockmann et al., 2004), which binds to the C-terminal domain of the 75 kDa subunit (Fig. 4.12). It is the only matrix subunit that contains two conserved cysteines positioned accurately for disulphide bond creation (Cys23 and Cys57), although the reducing environment of the mitochondrial matrix likely prevents the residues from forming a bridge. Intriguingly, patients suffering Parkinson's disease lose the B8

subunit from complex I (Keeney, Xie, Capaldi, & Bennett, 2006). That can be correlated with accumulation of oxidative damage to the enzyme in Parkinson's patients and possibly indicate the role of B8 subunit as another redox sensor.

The last supernumerary subunit attached to the peripheral domain of complex I is the 10 kDa subunit, located on the very top of the peripheral domain (Fig. 4.12). It has a long coiled structure that snakes through the groove between the 24 kDa and 51 kDa subunits. In our model we were able to build only the C-terminal half of the protein. It starts from the tip of the N module and ends at the 75 kDa subunit. The remaining part of the peptide is disordered. Most likely, like other coiled subunits it stabilizes this region. It has been reported that the 10 kDa subunit in complex I from rat's tissues can be present in two isoforms: short or long (Bridges, Mohammed, Harbour, & Hirst, 2017). Mass spectrometry data revealed the presence of only the short form in the ovine cryo-EM sample (Fiedorczuk et al., 2016).
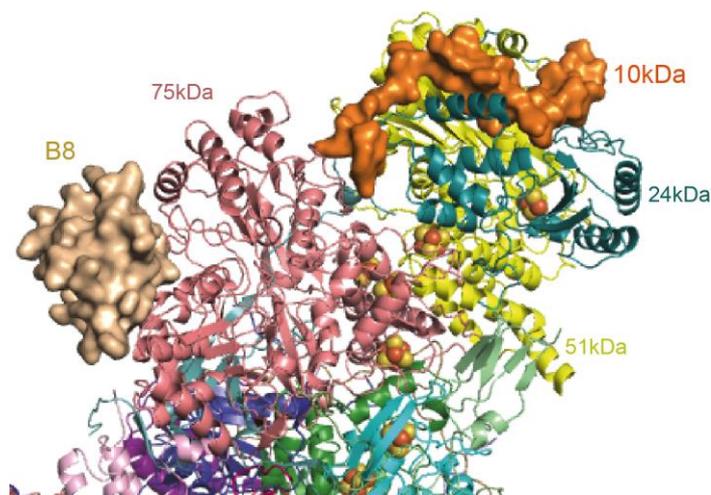


**Figure 4.12** Side view of the complex I N-module. The surfaces of the B8 and 10 kDa subunits are shown.

**Subunits associated with the membrane domain**

Soluble proteins that associate with mitochondrial complex I are located on both sides of the membrane. On the matrix side the 42 kDa, B22 and SDAP-β subunits form proximal and distal bulges visible even in very low-resolution

maps or 2D classes (Chapter 3.6.2 and Fig. 1.15). In the inter membrane space 15 kDa, B18, PDSW and PGIV subunits stabilize the complex.

**Matrix side**

On the matrix side close to the tip of the MD another ACP/LYR protein pair (SDAP-β/B22) together with the N-terminal domains of the B15, ASHI, B12, AGGG, B17 and SGDH form the distal bulge of the TM arm. The whole structure is formed from highly intertwined protein chains and the PNS ligand (Fig. 4.13). The SDAP-β/B22 pair is the main element of this structure and it extensively interacts with the ND5 subunit around proton entry sites (Fig. 4.14). Thus it may influence the pumping activity of the module.
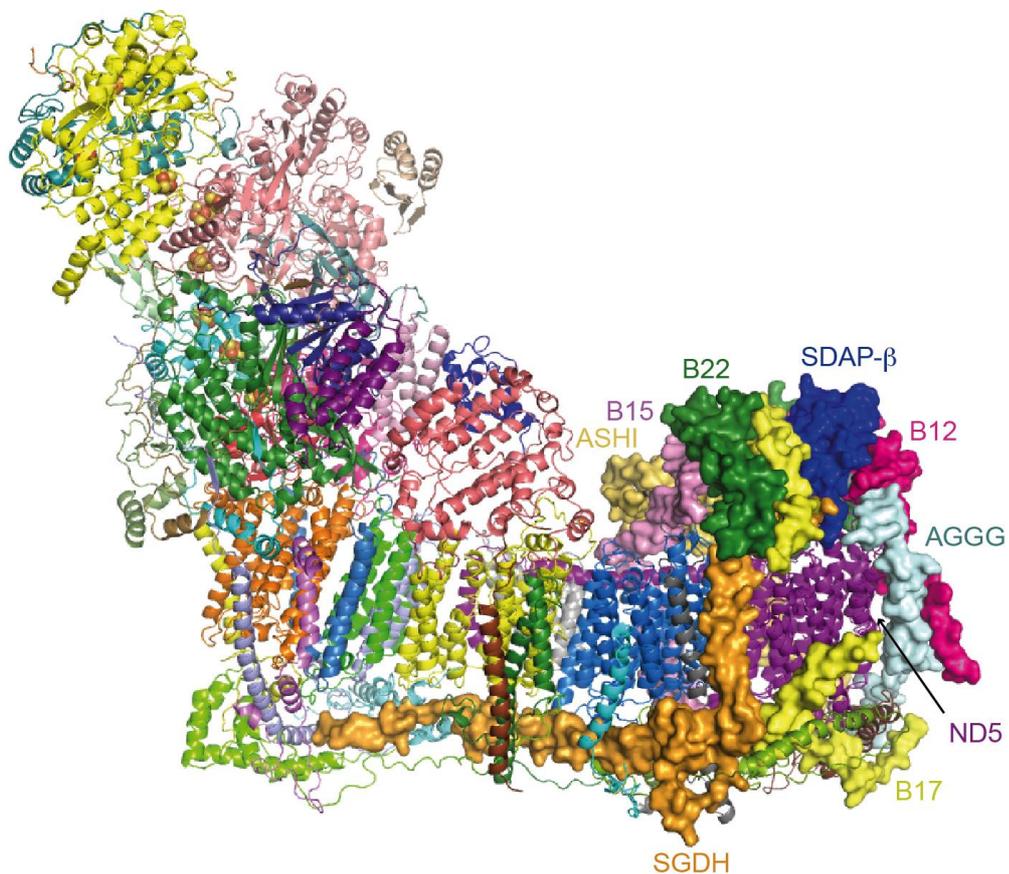


**Figure 4.13**Side view of complex I. The surface of the SGDH, B17, AGGG, B12, SDAP-α, B22, B15 and ASHI subunits composing the distal bulge on the matrix side of complex I is presented.

The B22 subunit (like the B14 subunit) belongs to the LYR family. Based on our results it is possible to understand the importance of the LYR motif for heterodimer formation. Leucine (residue 34 in B14 and 18 in B22) helps to form

a hydrophobic pocket accommodating the PNS-acyl chain, tyrosine (35 in B14 and 19 in B22) is responsible for stabilizing the alpha-helices forming the cavity and interacting with ACPs, finally arginine (36 in B14 and 20 in B22) forms a salt bridge with conserved Asp64 of the SDAP subunit. Previously, it was expected that the interaction between ACP and LYR proteins in complex I is formed solely by the phosphate of the PNS and arginine of the LYR motif (Angerer, 2015).
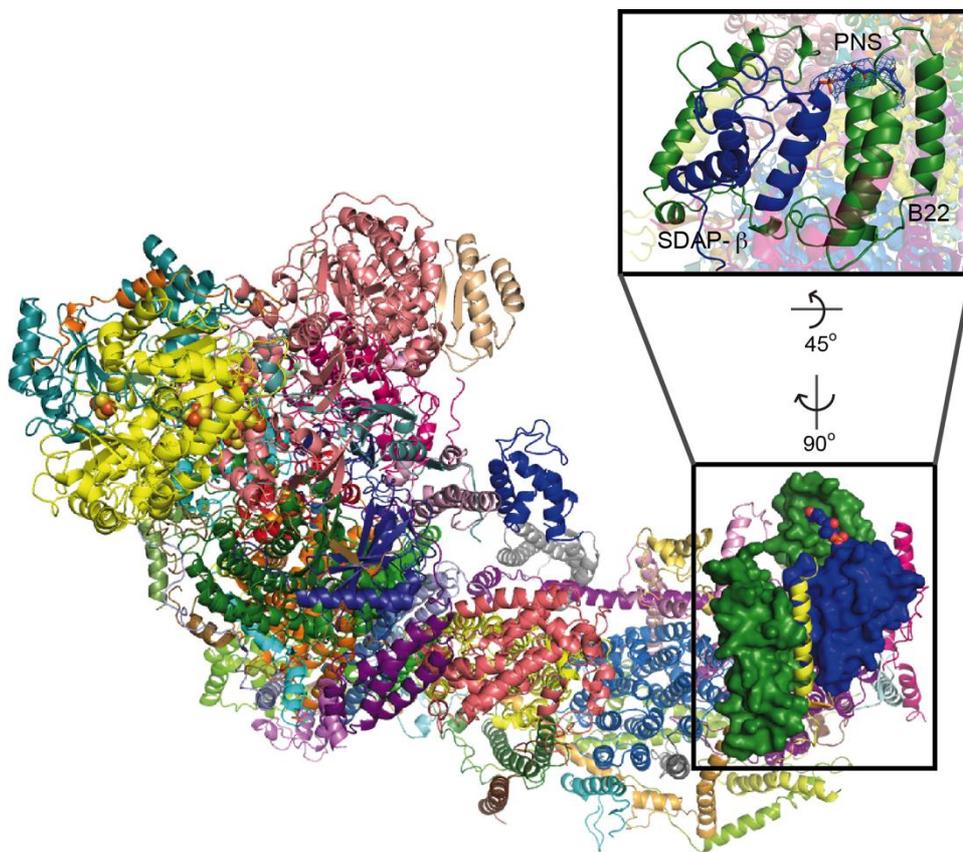


**Figure 4.14**Matrix/side view of complex I. The surface of the B22 and SDAP-β subunits composing the distal bulge on the matrix side of complex I is presented. Top-right panel shows PNS in SDAP-β site with the experimental density for the co-factor. The co-factor is shown with cryo-EM density carved to within 5 Å.

The more proximal bulge on the membrane arm of complex I is the42 kDa subunit. It belongs to a nucleoside kinase family and presents a characteristic globular fold (Johansson et al., 2001). It has a central alpha/beta nucleoside kinase fold with a parallel five-strand beta-sheet. It resides on the ND2 subunit

docked by three extensions, very close to PA (Fig. 4.15). It is stabilized by the B14.5b and KFYI of the TM supernumerary subunits and possibly the B13 and SDAP-α/B14 pair attached to the PA via the 30 kDa subunit. In the "closed" class map B13 and SDAP-α/B14 dimer come to within 3Å from 42 kDa subunit which enables direct interaction and possibly regulation of PA hinge-like movement. Any catalytic function of the 42 kDa subunit has not been described in the literature so far, although it contains all the key residues and the active site of the protein is accessible (Johansson et al., 2001).
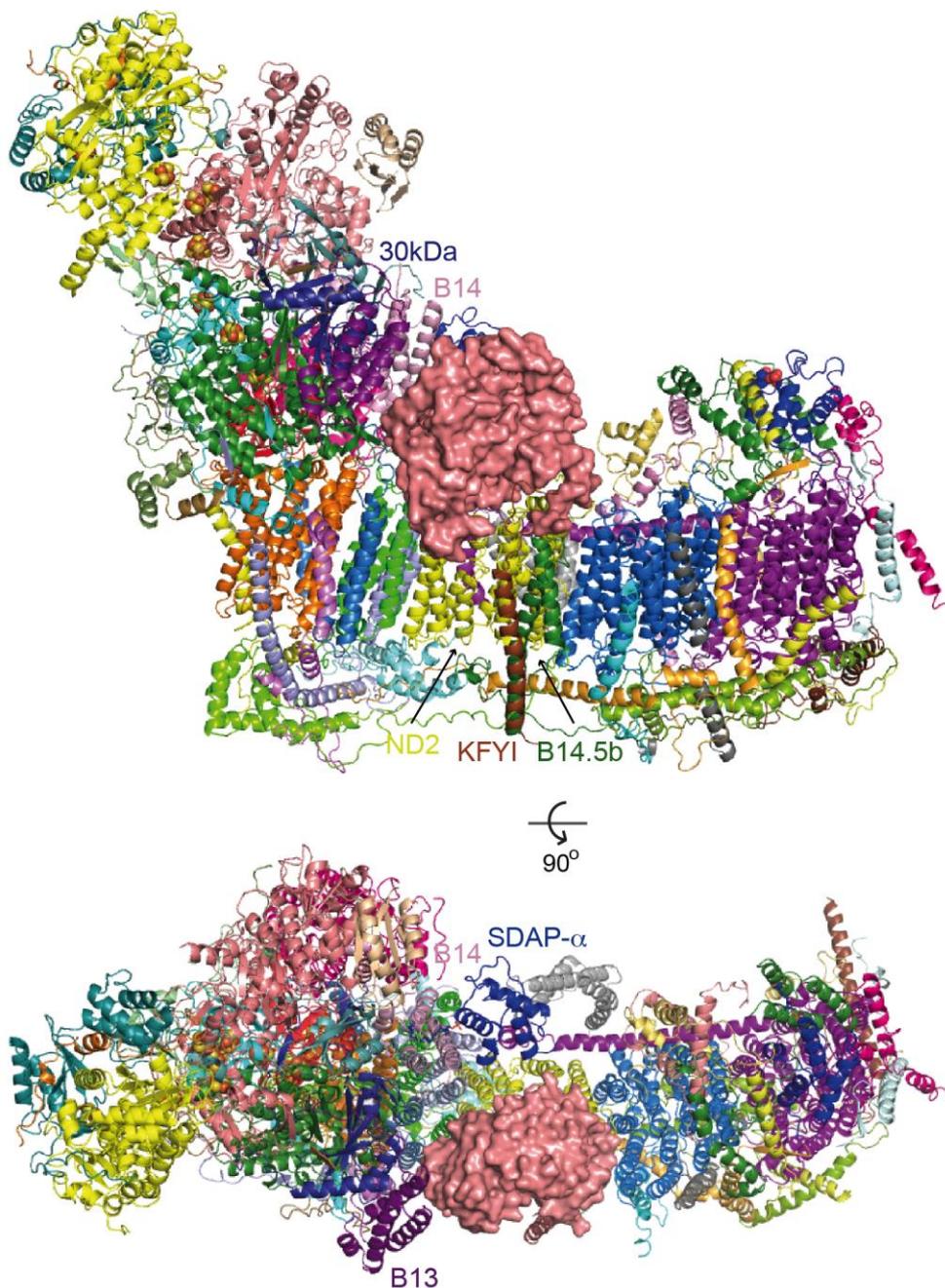
**Figure 4.15** Side (top) and matrix (bottom) views of complex I. The surface of the 49 kDa subunit is presented.

**Inter membrane site**

Three subunits of complex I (PGIV, 15 kDa and B18) contain Cx9C (CHCH) domains that stabilize their helix turn helix motifs (Szklarczyk et al., 2011).

116
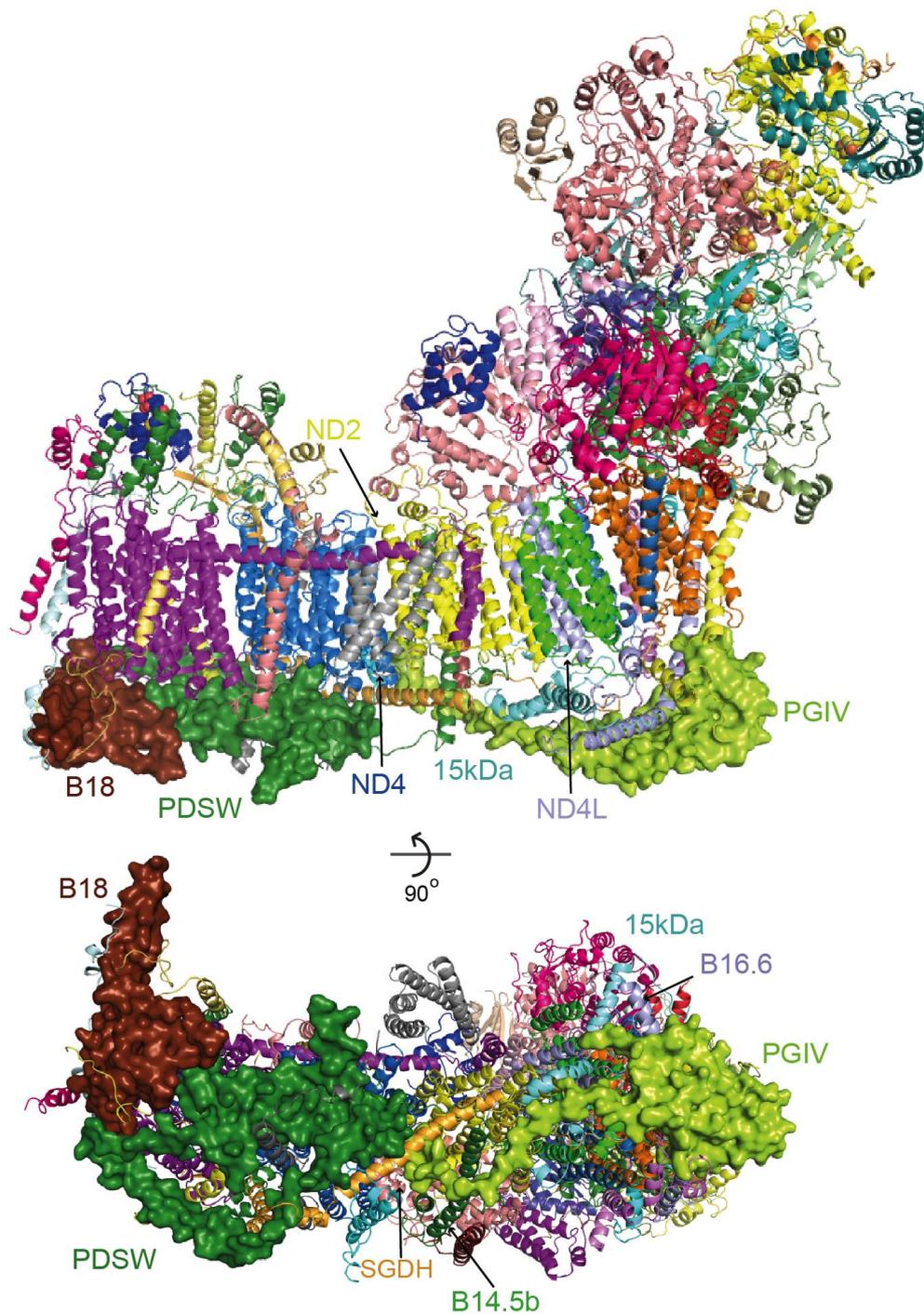
**Figure 4.16.** Side (top) and IMS (bottom) views of complex I. The surface of the B18, PDSW, PGIV and 15 kDa subunits are presented.

The two CHCH domains of PGIV form the 'heel' of mammalian complex I. Short α-helices lock the B16.6 subunit and a long C-terminus runs on the surface of the IMS site all the way to the ND2/ND4 interface where it binds a hydrophobic

Phe-Phe-Trp motif. This way it works as a "clamp", holding tightly the scaffolding subunits (B14.5b, SGDH, 15 kDa and B16.6) (Fig. 4.16).

Another structurally important Cx9C containing protein is the 15 kDa subunit. It contains a CHCH motif that stabilizes two short α-helices placed on the ND2 subunit form the IMS site. The subunit winds itself between the core ND2 and ND4L subunits and the supernumerary B16.6, SGDH, PGIV subunits. The15 kDa subunit is clearly very important for integrity of this region since it makes many extensive interactions with a number of partners (Fig. 4.16).

The B18 subunit is placed directly underneath the ND5 subunit. One of its Cx9C motif α-helices is exceptionally long and protrudes almost 30Å away from the complex. It is pointing in the direction of complex IV in the supercomplex (Letts, Fiedorczuk, et al., 2016). The loops of the CHCH motif and N-terminal coil of the B18 subunit are placed close to another protein containing di-sulphide bridges, the PDSW subunit (Fig. 4.16).

The PDSW subunit contains four cysteines placed in positions that enable disulphide bond formation. The bonds presumably stabilize the protein fold, resulting in stabilization of the interactions between ND5-ND4 and surrounding accessory subunits. Additionally, it contains a conserved cysteine in position 154 on the C-terminal α-helix that faces a conserved (in mammals) Cys112 of the ESSS subunit and so together they are likely to form the only inter-subunit disulphide bond in the complex (Fig. 4.17). The C-terminal α-helix (alike in the B18 subunit) is involved in the supercomplex formation interacting with complex III (Fig. 4.16)(Letts, Fiedorczuk, et al., 2016).
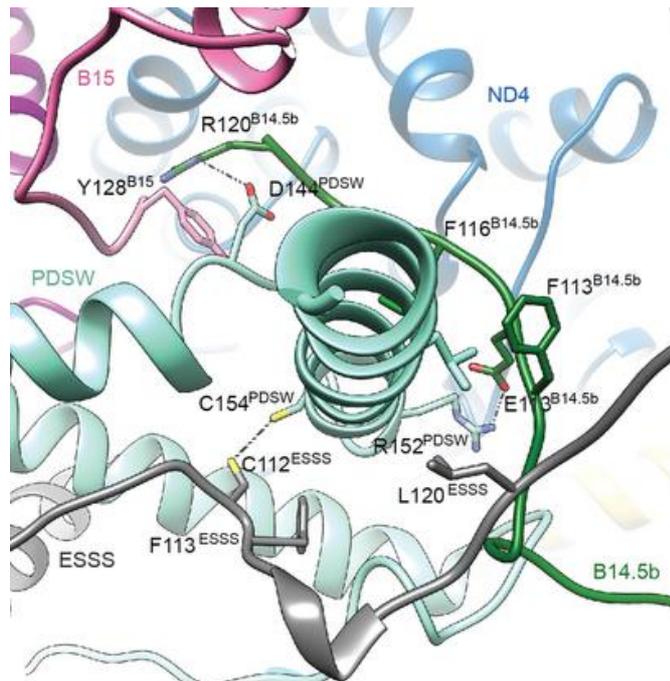
**Figure 4.17.** Close up of the C-terminal helix of supernumerary subunit PDSW at the centre of the membrane arm on the IMS side. This helix extends away from complex I and is encircled by the C termini of supernumerary subunits B14.5b, ESSS and B15. The side chains of residues involved in stabilizing interactions are shown. A possible disulfide bond between PDSW (Cys154) and ESSS (Cys112) and stabilizing salt bridges are indicated by dashed lines. Figure prepared by Dr James Letts.

Most of the proteins in the intermembrane space contain helix-turn-helix motifs stabilized by sulphur bridges. This feature and their location make them very important for complex I integrity. The PGIV and B15 subunits have probably overtaken the function of β-hairpin/helix motifs (βH) that stabilizes MD's periplasmic side in bacteria but are mostly missing in mammals (Efremov & Sazanov, 2011)(R. Baradaran et al., 2013). The B18 and PDSW subunits additionally may play a regulatory role in supercomplex formation (Letts, Fiedorczuk, et al., 2016). Upon a change in the redox conditions, disulphide bonds of CHCH motif can break and the interactions between CHCH α-helices and complexes IV and III may be disrupted.

**Subunits associated with the membrane domain containing transmembrane motifs**

Most of the supernumerary proteins interacting with the MD are defined by a single transmembrane α-helix that separates the N- and C-terminal ends. They form eighteen transmembrane helices, which in addition to TMHs of the core subunits give a total number of 78 α-helices embedded in the internal mitochondrial membrane. Accessory subunits in the membrane create a cage around the core that presumably increases stability. In all cases, they have the C-terminus located in the IMS, determined by the mechanism of insertion (Stuart, 2002).

The TM helices of the B15, ASHI, B12, AGGG, B17 and SGDH subunits surround the tip of the MD, stabilize it and contribute to the distal bulge (Fig. 4.13).

The highly tilted (~60° relative to the plane of the membrane) TM helix of B17 is unique because all the other TM helices of the complex are approximately perpendicular to the plane of the membrane. It has a C-terminal extension that together with a β-hairpin of the ND5 subunit makes a three-stranded anti-parallel β-sheet, expected to stabilize the IMS side of the ND5 subunit. On the matrix side it has a long alpha helix interacting with SDAP-β and B22 that stabilizes this ACP/LYR pair. N-terminus and TM helix of B17 are connected through a flexible linker with a short α-helix. That part of the protein is very flexible and therefore disordered in the cryo-EM map and impossible to build (Fig. 4.18).

The SGDH subunit binds between the ND4 and ND5 subunits, in a groove filled with two cardiolipin molecules (Fig. 4.18). The C-terminus of the protein wraps around the PDSW chain and continues along the ND4 subunit in the direction of the ND2 subunit. The SGDH and PDSW subunits are interlocked and together build two long α-helices traversing across the three antiporter-like subunits from ND5 to ND2,stabilizing this region from the IMS side. The N-terminal end of the SGDH subunits ends with a coil anchored deep in the "heel" of the enzyme.

The B12 subunit is placed at the distal end of the MD. It forms the most distal TM α-helix that interacts with complex IV in the supercomplex (Letts, Fiedorczuk, et al., 2016). The TM helices are also slightly tilted with respect of the membrane plane. The C-terminal part of the B12 subunit forms an unstructured coil, binding in proximity of the B18 and ND5 subunits. The N-terminus of the protein forms two parallel α-helices stabilizing the position of the SDAP-β subunit (Fig. 4.18).

The TM helix of AGGG is locked in betweenthetheB12 and ND5 subunits. It is leaning at a similar angle as the B12 subunit but in the opposite direction. The short N-terminus binds to the ND5 subunit and the C-terminal end can be traced as a coil running along the external α-helix of the B18 subunit interacting with complex IV (Fig. 4.18) (Letts, Fiedorczuk, et al., 2016).
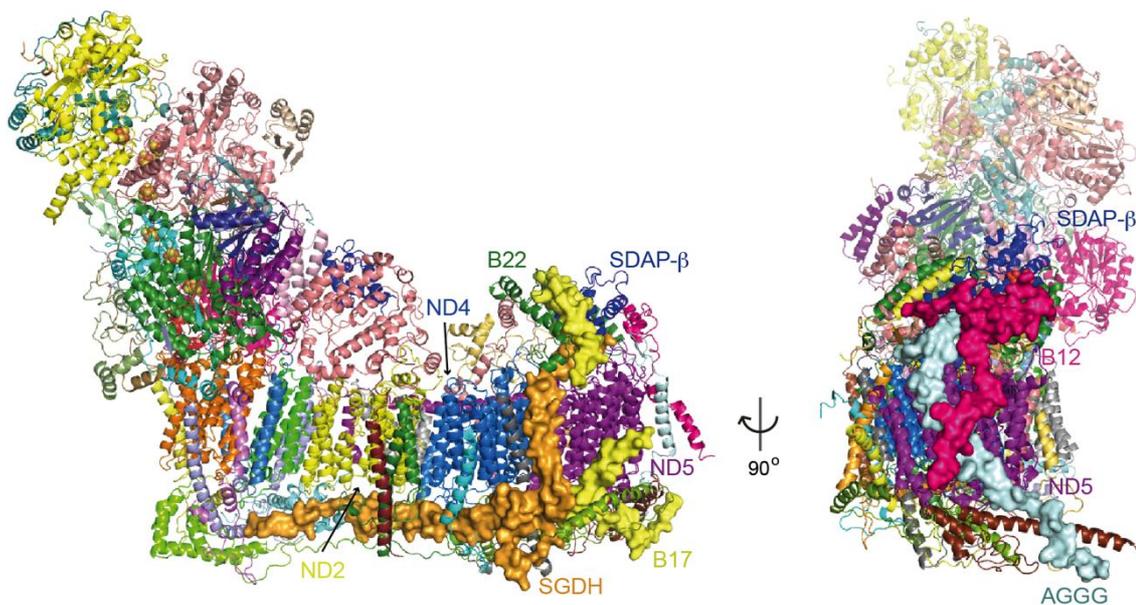


**Figure 4.18.** Left side and front views of complex I. The surface of the SGDH, B17, B12 and AGGG subunits presented.

The ASHI subunit binds to the ND5 subunit across the B17side. It has a relatively big globular N-terminal domain forming extensive interactions with the ND4, B22 and mostly B15 subunits. This part of the protein is disordered in the ovine map, for that reason it was only possible to trace the backbone of the peptide. The distortion may be caused by increased flexibility of the amphipathic α-helix of the ND5 subunit in the sample used. The C-terminus

forms a coil that runs between the ND5 and ND4 subunits interacting with both (Fig. 4.19).

The B15 subunit stabilizes a long lateral helix of the ND5 locking it to the ND4 subunit. The N-terminus forms two α-helices connected by a flexible linker. The first helix lays on the surface of the ND4 subunit close to the proton input channels, the second one runs away from the centre of the MD and interacts extensively with the B22 subunit. Those two α-helices lock the whole N-terminal domain of ASHI in between them (Fig. 4.19).



**Figure 4.19** Right side view of complex I. The surface of the ASHI, B15 and B14.7 subunits presented

The MWFE subunit is bound at the very back of the complex I MD domain. It strongly interacts with the ND1 subunit through a TMH placed anti-parallel to the TMH1 of the ND1 subunit, which builds the entry point for the ubiquinone channel. The N-terminal domain of the MWFE subunit is very short and barely sticks above the membrane level, where it is locked in between the B17.2, TYKY and ND1 subunits. The C-terminus of the subunit forms a short α-helix interacting extensively with other proteins building the "heel" of the complex,

namely accessory subunits B16.6, PGIV and the core ND1 and ND6subunits (Fig. 4.20).

The B16.6 subunit is one of the most interesting supernumerary subunits of mammalian complex I. It has a very long α-helix that runs continuously through the lipid bilayer, kinks and continues through the "heel" of complex I to the opposite side of the enzyme. The long α-helix is followed by a linker and a short helix locked in between the ND1, ND6, ND3 and MWFE subunits with a motif of large side chains and hydrophobic amino acids: Phe-Ile-Trp-Tyr-Thr. Close to the C-terminus of the B16.6 subunit a lipid binding site has been identified (Fig. 4.23). The N-terminus of the protein is a coil that stretches almost to the N module. The B16.6 subunit is the only protein of complex I that spans both arms of the enzyme. Its location suggests that it has an important stabilizing role, however it is also found to be a cell death regulator (also called GRIM19) (Lu & Cao, 2008). It can be hypothesized that upon miss-assembly or degradation of complex I, free B16.6 subunit may be released to the cytoplasm and promotes a cascade that leads to destruction of the cell.
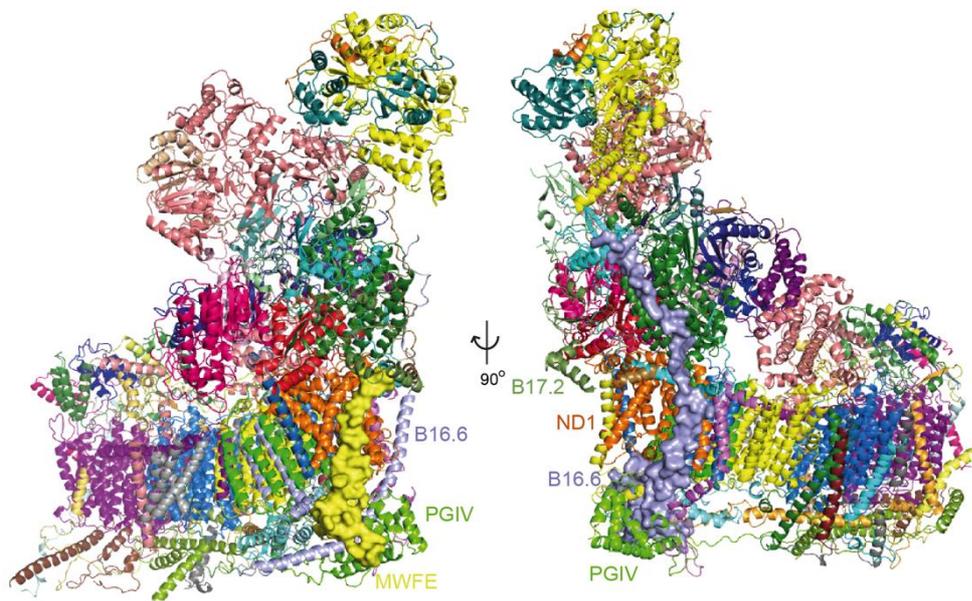


**Figure 4.20.**Back view of complex I. The surface of the MWFE and B16.6 subunits is presented.

Near the TMH of the B16.6, the B9 subunit is located. It has a trans membrane helix that interacts with the last TMH of the ND1 subunit and a short

amphipathic α-helix at the N-terminus proposed to anchor and stabilize this side of the complex in the membrane. The C-terminus is built of an α-helix followed by a coil, mostly interacting with the PGIV subunit (Fig. 4.20).

The B14.5b subunit is the only supernumerary subunit with two TMHs. The first of the helices is of the length (~25Å)approximating the internal mitochondrial lipid bilayer. The second is twice longer and protrudes deep into the intermembrane space. The protein binds next to the void of the ND2 subunit with both ends on the IMS site. The N-terminus, followed by the short TMH forms a short coil that circles the long TM helix and ends binding in between the ND2 and SGDH subunits, very close to a lipid molecule that was assigned as phosphatidylcholine (Fig. 4.21 and 4.23).

The KFYI subunit binds to the B14.5b subunit. It has a very short C-terminus, that is practically two α-helix turns extension of the TMH and a several amino acids long coil on the matrix site interacting with the 42 kDa subunit (Fig. 4.21).

**Figure 4.21** Left-IMS side view of complex I. The surface of the B9, KFYI and B14.5b subunits presented.

The ESSS subunit tightly binds between two transmembrane helices of the ND4 and SGDH subunits. On the IMS site, it has a short linker connecting its TMH to a 25 amino acids long α-helix placed in between loops and an α-helix of the PDSW subunit, almost perpendicularly to the membrane plane (Fig. 4.21). The protein ends with a coil that wraps the three helix bundle of the PDSW subunits and holds them together. The 38-amino acids long N-terminus of the ESSS subunit is not resolved well in our map but can be traced in the bovine and porcine maps (J. Zhu et al., 2016). It forms a coil that transverses along the longer axis of the membrane arm almost to the ND4L subunit. With a long extension on the N-terminus and α-helix/coil motif on the IMS the ESSS subunit spans two sides of the MD of complex I.

The MNLL subunit is one of the shortest complex I subunits. In the TM part it is attached to the ND4 subunit in between the ESSS and B14.5b subunits. The N-terminal amphipathic α-helix is positioned on the surface of the IMM in a way that enables stabilization of this region. The C-terminus forms a coil that wraps around the SGDH and PDSW subunits (Fig. 4.22).
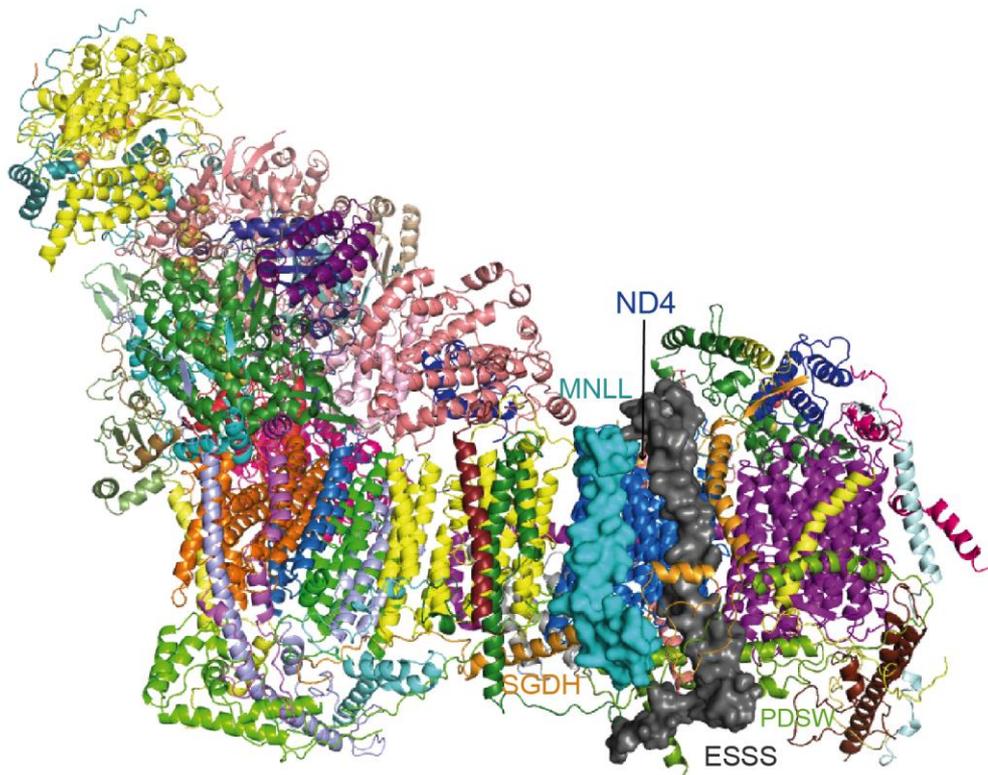


**Figure 4.22** IMS/side view of complex I. The surfaces of the MNLL and ESSS are presented.

Although mass spectrometry analysis of the sample used for cryo-EM grid preparation confirmed the presence of the B14.7 subunit (Table S2), maps revealed that the region of the map where it binds is disordered. The reason for the B14.7 subunit instability is not entirely clear but most likely it is due to use of Brij-35 for the grid preparation. For that reason for initial building of this subunit lower resolution maps of mitochondrial respirasome were used (Letts, Fiedorczuk, et al., 2016). However for a more thorough analysis of this region the bovine or porcine complex I maps can be used (J. Zhu et al., 2016)(Wu et al., 2016). The B14.7 subunit is a four transmembrane helix subunit with both termini on the IMS side (Fig. 4.19). It interacts with complex I through TMHs I

and IV. Helix I is in contact with the last TM helix of the ND5 subunit and helix IV binds to lateral helices of the ND5 and ND2 subunits. The protein belongs to the Tim17 protein family, which groups mitochondrial protein translocases (Tim17, Tim22 and Tim23) but family members are also found in peroxisomes and plastids (Murcha, Wang, & Whelan, 2012). To my knowledge any transporting activity of the B14.7 subunit has not been shown so far. The protein is likely to be important for complex I assembly or supercomplex formation, as it interacts extensively with complex III (Letts, Fiedorczuk, et al., 2016)(Wu et al., 2016).

**Lipid Molecules**

We have observed twelve lipid molecules specifically bound to complex I. Four of them have four alkyl carbon groups and the rest two. Molecules with four alkyl groups were assigned as cardiolipins (CL). Cardiolipin is one of the most abundant lipid types in mitochondria (Paradies et al., 2014). It constitutes 20% of all lipid molecules and was shown to be very important for activity of mitochondrial proteins (Planas-Iglesias et al., 2015). Two chain lipids were assigned as phosphatidylcholine (PC) and phosphatidylethanolamine (PE) depending on the intensity of the cryo-EM density in the lipids' head groups. CL, PC and PE were identified as the most common fatty acids bound to mitochondrial complex I in numbers of: 16, 8 and 23 molecules per one molecule of the enzyme (Shinzawa-Itoh et al., 2010). Many of the identified lipids (including CL) occupy the cavity resulting from the evolutional loss of the N-terminal α-helices of the ND2 subunit. This fact indicates that either lipids compensate for the stability loss due to missing helices or the ND2 subunit evolved to accommodate specific regulatory lipids. Other cardiolipins are found in a large gap between the antiporter-like ND4 and ND5 subunits and at the "heel" of the complex interacting with the amphipathic helices of B17.2, B14.5a and PSST subunits. PCs and PEs were identified at the ND1/ND6 and ND4/ND2/B14.5b interfaces (Fig. 4.23). The tightly bound lipids presumably stabilize the structure, prevent potential proton leaks and possibly play regulatory roles. Previously, acyl chains specifically bound to complex I were only observed in the crystal structure of the *E. coli* MD (Efremov & Sazanov, 2011). However, the density for the head groups of these molecules was

blurred, preventing the identification as lipids or detergent, and the binding sites were different.
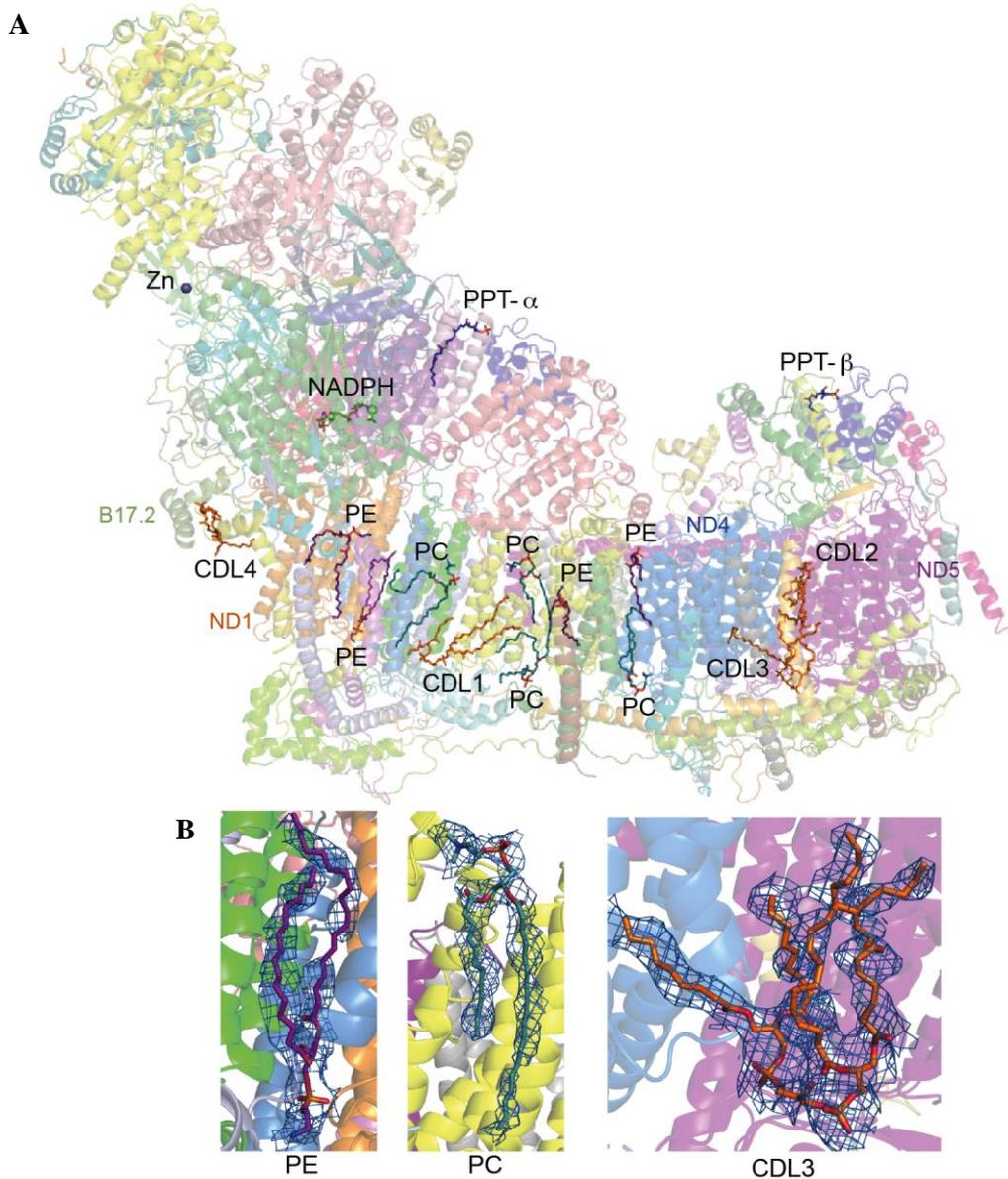


**Figure 4.23A.** Side view of complex I, with cofactors including lipids shown as sticks. CDL, cardiolipin; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PPT, phosphopantetheine. **B.** Lipids phosphatidylethanolamine, phosphatidylcholine and cardiolipin. Lipids are shown with experimental cryo-EM density carved to within 5 Å.

## 4.3 Different conformations of ovine complex I

Initial data processing revealed the presence of complex I in two different conformations ("open" and "closed") (Chapter 3.6.2). Extensive re-classification of all particles composing these two states allowed extracting three additional classes. The final result were five classes at resolution from 4 to 4.5Å.

| Class number | Resolution (Å) | Particles number | Rotation of PA (°) | Wedge-like motion of PA (°) | Q cavity |
|---|---|---|---|---|---|
| Class 5 | 4.4 | 14704 | Clock wise 2° | Yes 4° | Blocked |
| Class 4 | 4.2 | 17543 | Clock wise 4° | Yes 3.5° | Poorly resolved |
| Class 3 (analogue of "open") | 4.0 | 21555 | Clock wise 3° | Yes 3° | Blocked |
| Class 2 | 4.4 | 12653 | Clock wise 2° | No | Poorly resolved |
| Class 1 (analogue of "closed") | 4.5 | 14429 | Reference class | Reference class | Accessible |

**Table 4.1** Comparison of complex I conformations after re-classification. Rotation angles were measured between B17.2 and SDAP-α subunits. Wedge-like motion degrees were measured between ND1 and B8 subunits.

Re-classification of particles revealed that class 3 with ~21,5 k particles is almost identical to the original "open" class with ~82.9 k particles and reaches a similar resolution of 4.04 Å (3.9 Å class "open"). The "closed" class (4.6 Å, ~46.3 k particles) is analogous to class 1 (4.5 Å, ~14.4 particles). This result clearly shows the power of the extensive 3D re-classification approach that was taken (Chapters 2.6.2). From the ~129.2 k particles classified originally as the "open" and "closed" classes only ~60.7 k was enough to obtain maps of similar resolution. The remaining 53% of particles were re-assigned to different conformations (classes 2, 4 and 5) or artefacts (Chapter 3.6.3). Such an approach to classification (which one might call "focus-revert-and-classify") should be applicable to other cryo-EM studies, when differences between the conformations are relatively small globally, but significant locally and functionally.
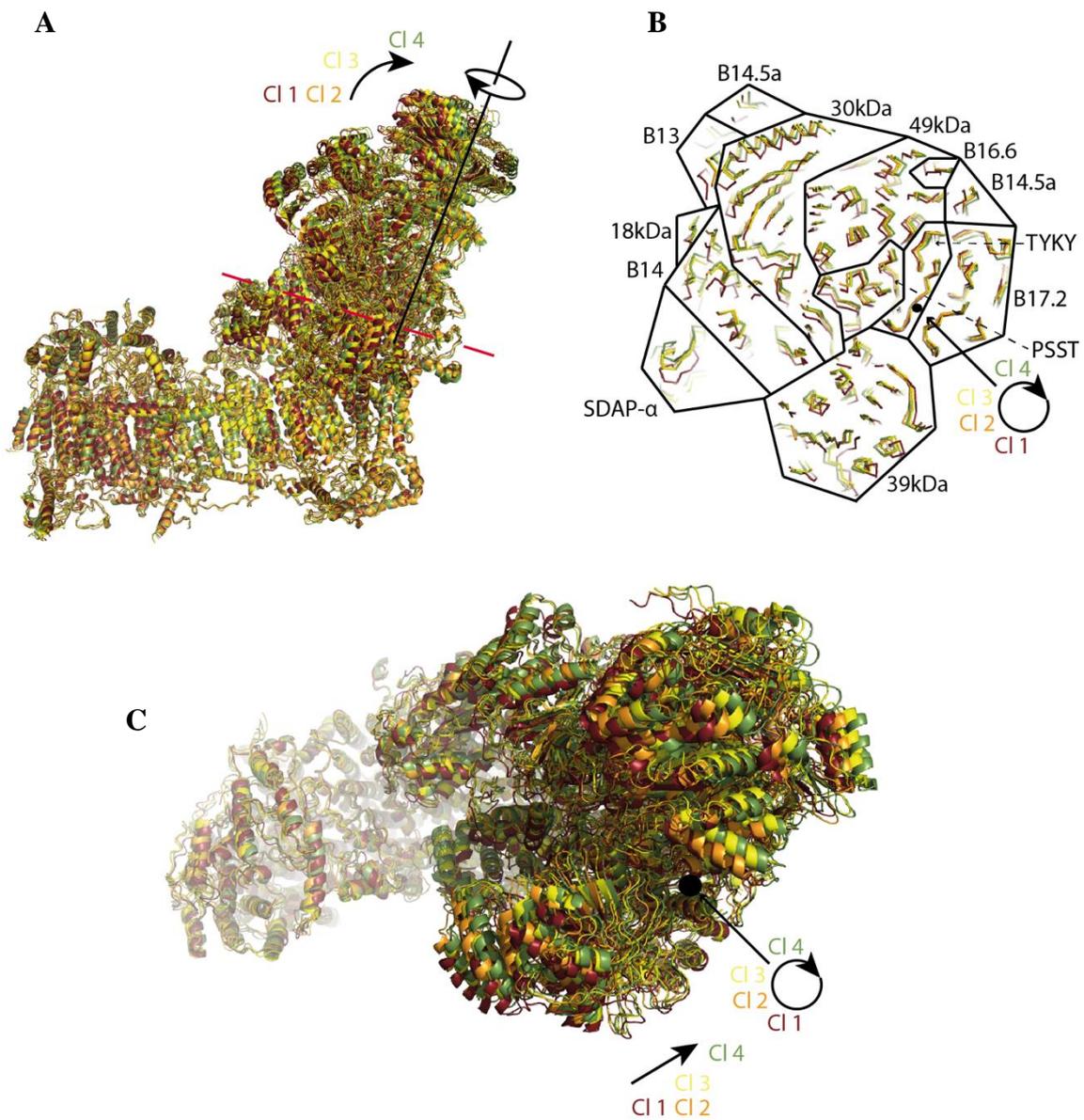
**Figure 4.24** Comparison of classes 1-4 of the new classification. **A.** Side view of complex I with rotation axis and direction of PA wedge-like movement. The red dashed line shows the location of the cross-section presented on panel B; solid line presents the rotation axis of the PA; arrows show direction of the rotation and movement of the PA **B.** Cross section of complex I peripheral arm. The arrow shows the direction of rotation; the black dot indicates the axis of rotation. **C.** Matrix side view of complex I. Arrows indicate the direction of the rotation and hinge like movement of the PA; point indicates axis of rotation. Structures depicted as cartoons (A. and C.) or ribbons (B.). Class 1 in red, class 2 in orange, class 3 in yellow, class 4 in green. Structures aligned by the core subunits ND1 - ND4.

Separated classes differ by the position of the PA over the MD. Class 1 is the most closed and anti clock wise rotated (matrix view) in contrast class 4 is most

open and clock wise rotated (Table 4.1). Class 2 is very similar to class 1, it has a similar overall position of the peripheral arm over the membrane arm, however less anti clockwise rotated. Class 4 has the most open conformation with a strong clockwise rotation, similar to the 4 Å resolution class 3 but with a larger rotation. Class 5 is the most different from all other classes in regard to PA position. It has an intermediate rotation, similar to class 2 but very strong hinge-like displacement in the opening direction.
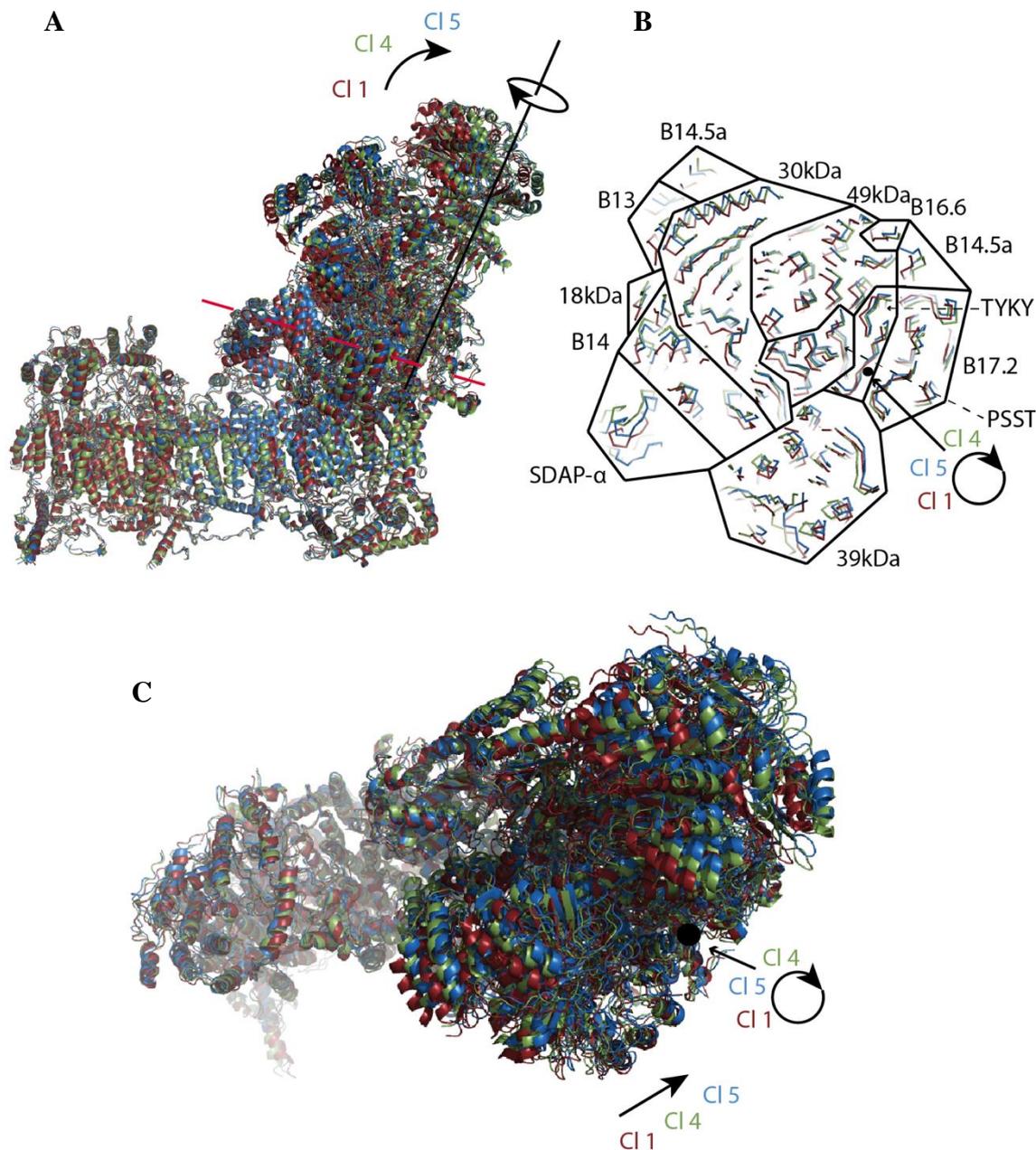
**Figure 4.25** Comparison of classes 1, 4 and 5 of the new classification. **A.** Side view of complex I with rotation axis and direction of PA movement. Red dashed line shows cross-section presented on panel B; solid line presents rotation axis of the PA; arrows show the direction of the rotation and movement of the PA. **B.** Cross section of complex I peripheral arm. Arrow is showing direction of rotation; the black dot shows the axis of rotation. **C.** Matrix side view of complex I. Arrows indicate direction of the rotation and hinge-like movement of the PA; dot indicates axis of rotation. Structures depicted as cartoons (A. and C.) or ribbons (B.). Class 1 in red, 4 in green and class 5 in blue. Structures aligned by the core subunits ND1 - ND4.

The new, more homogenous maps revealed correlation between the positions of the PD over the MD and conformation of the β1-β2$^{49\text{-kDa}}$ loop. Although all the maps in the region of the Q binding contain relatively high level of noise it is clear that the class 3 (analogue to "open") and class 5 (most open) have the β1-β2$^{49\text{-kDa}}$ loop blocking the Q site, but the class 1 (analogue to "closed") is more similar to the bovine Class 2 or porcine, or *T. thermophilus* models in this region, where the N2 cluster is accessible for oxidation. Additionally, in the ubiquinone site an elongated density indicates that it is occupied. It is not possible to judge at this level of resolution what this blob belongs to. Most probably it is a molecule of detergent Brij-35, which has been shown to be very efficient Q-site inhibitor of the mammalian complex I (Chapter 3.6)(Okun et al., 2000).
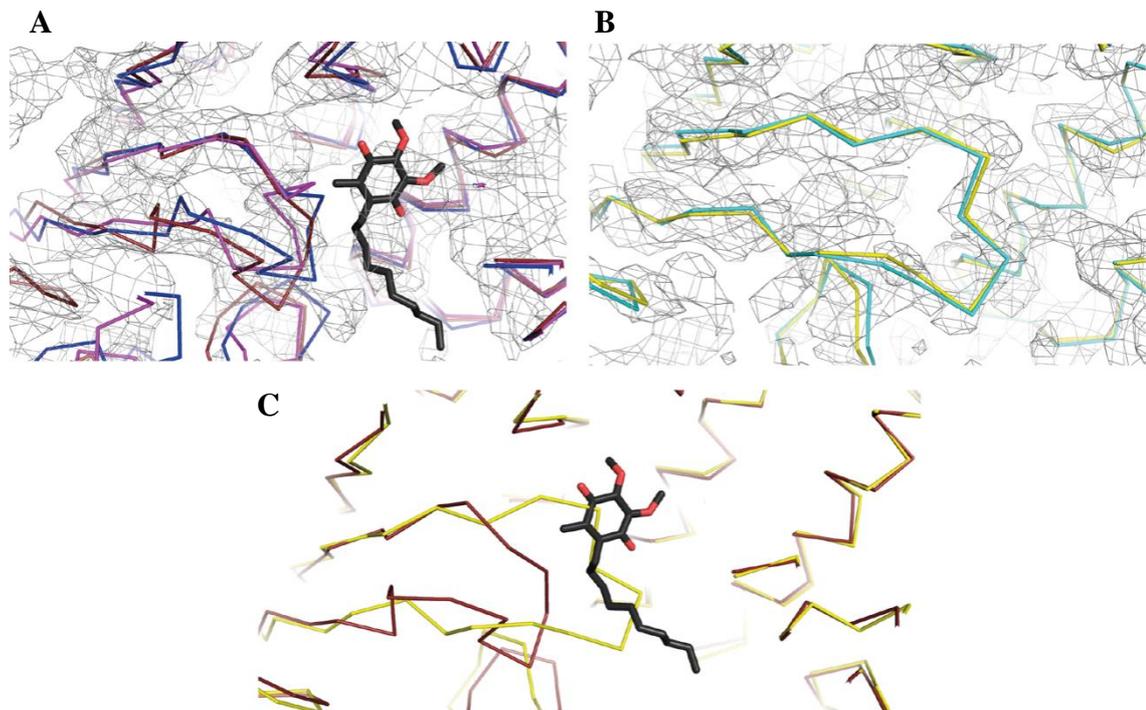


**Figure 4.26** Presentation of the Q site with the crucial β1-β2$^{49\text{-kDa}}$ loop in different models represented as ribbons. **A.** Bovine class 2 in blue, porcine model in magenta and ovine class 1 in orange with the experimental density from class 1 map in grey mesh. Structures may represent the active state of the enzyme. Quinone in dark grey was modelled based on the *T. thermophilus* structure; note that it aligns very well with the extra density in the Q chamber. **B.** Ovine class "open" in cyan and ovine class 3 in yellow with the experimental density from the class 3map in grey mesh. Structures may represent the de-active state of the enzyme or different stages in the catalytic cycle. **C.** Ovine class 3 in yellow and ovine class 1 in orange superimposed with quinone in grey, modelled based on the *T. thermophilus* structure. All the models were aligned by 49 kDa subunit and depicted as ribbons.

The two remaining classes, 2 and 4 have β1-β2$^{49\text{-kDa}}$ loop map regions poorly resolved and the position of the β1-β2$^{49\text{-kDa}}$ loop is difficult to judge. It may be caused by presence of particles in both states: with β1-β2$^{49\text{-kDa}}$ loop inserted and retracted.

Global changes in the PA positioning and Q chamber accessibility are clearly visible. Inspection of the MD also shows some changes in the position of the distal ND5 and surrounding subunits, however they are much less prominent (Fig. 4.27).
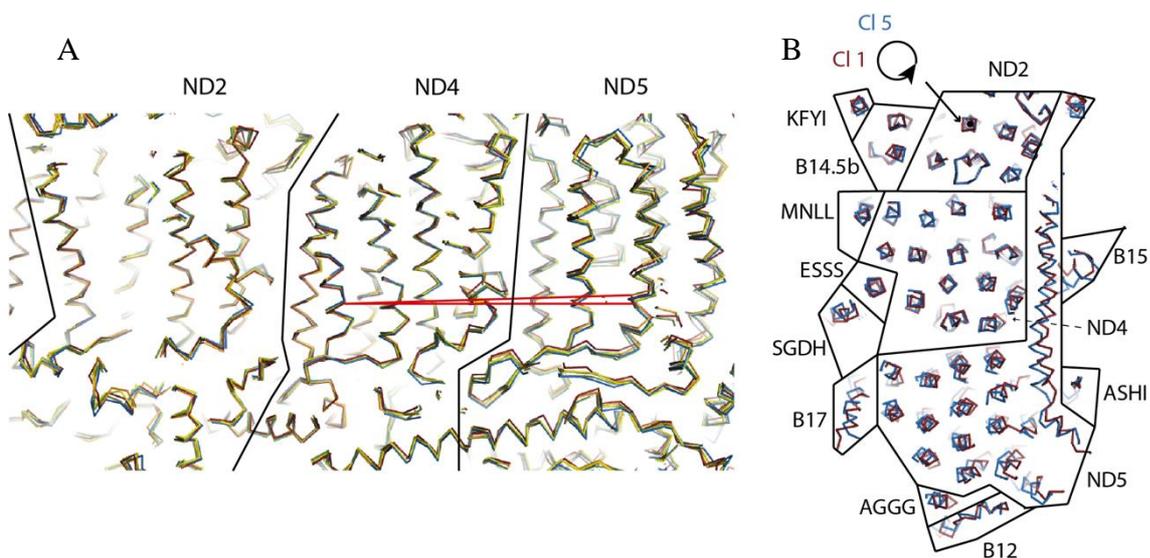


**Figure 4.27** Comparison of the membrane domain of classes 1-5 of the new classification. **A.** Side view across the membrane domain of complex I with change in the ND5 subunit position presented with the red lines connecting the corresponding point in the ND4 and ND5 subunits in classes 1 and 4, lines form and angle of approximately 1.5°. Notice how well TMHs in the ND2 and ND4 subunits are aligned compared to the ND5 subunit. **B.** Matrix view of complex I membrane cross-section. Class 1 in red and class 5 in blue. Structures aligned by the core subunits ND1-ND4.

Compared to class 1, classes 2, 3 and 4 have the ND5 subunit "kinked" in the direction of the IMS by respectively 0.5, 1.5, 1.8 degrees. The degree of the ND5 subunit displacement correlates with the change in PA position. Class 4 is the most different form class 1 in terms of PA position (rotation and hinge-like movement) and have the ND5 subunit shifted to IMS by the highest degree, whereas class 2 is the most similar in terms of PA and ND5 subunit

arrangement. Class 5 is again slightly different than the other classes. It has the ND5 subunit rotated clockwise by ~2 degrees (looking from the IMS) rather than "kinked". This is very interesting since the position of PA in this class also stands out comparing to the rest due to weak rotation and strong hinge-like opening. However two important points have to be highlighted. Firstly, due to very small changes, the exact direction of the ND5 subunit movement is difficult to evaluate unambiguously. Displacement can be caused by both rotation and the "kink" of the ND5 subunit. Secondly, reported degrees of the ND5 subunit displacement may differ depending on which amino acid residue will be used as a reference. Nevertheless this correlation of the ND5 subunit position change with overall position of the PA arm and accessibility of Q chamber is most intriguing.

## 4.4 Comparison of recently published complex I structures

### 4.4.1 Overall conformations

Recently in addition to the ovine mitochondrial complex I, two other high resolution models of the enzyme, from *B. Taurus* at 4.2 Å resolution (J. Zhu et al., 2016) and *S. scrofa* at 3.6 Å (Wu et al., 2016) and one medium resolution from *Ovis aries* at 5.8 Å (Letts, Fiedorczuk, et al., 2016) were published. As expected from the sequence identity (Chapter 3.1) all the models are very similar. Apart from some alterations in the primary sequences, a few differences in amino acids tracing are likely to come mainly from lower resolution at the peripheries of the maps. Global changes in the position of PA, distal part of the MD and the position of β1-β2[49-kDa] loop are also present (Table 4.2).

| Class number | Resolution (Å) | Rotation of PA (°) | Wedge-like motion of PA (°) | Rotation of ND5 subunit (°) | Q cavity |
|---|---|---|---|---|---|
| Bovine Class 1 | 4.3 | Clock wise 2° | No | Clock wise 1.3 | Poorly resolved |
| Bovine Class 2 | 4.4 | Anti-clock wise 2° | Yes 2° | Anti-clock wise 2.5 | Accessible |
| Bovine Class 3 | 5.6 | Clock wise 3° | Yes 3° | Clock wise 4° | Poorly resolved |
| Porcine respirasome | 3.6 | Clock wise 3° | Yes 2° | Anti-clock wise 1.5° | Accessible |
| Ovine respirasome (analogue to ovine class 2) | 5.8* | Clock wise 2° | No | No | Poorly resolved |
| Ovine class 1 | 4.5 | Reference class | Reference class | Reference class | Accessible |

**Table 4.2** Comparison of complex I conformations from different studies. Rotation angles were measured between B17.2 and SDAP-α subunits. Hingelike motion degrees were measured between ND1 and B8 subunits for all the models except porcine where B8 and 49 kDa subunits were used. *Overall resolution reported for the supercomplex, it can be underestimated for the complex I region of the map.

Comparison of the structures reveals distinct conformational states of the molecule defined by hinge and rotation like movements of the complex arms and displacement of ND5 (Table 4.2). It is possible that they correspond to active/de-active (A/D) conformations (Maklashina, Kotlyar, & Cecchini, 2003) or different states of the molecule during the catalytic cycle. Nevertheless this finding confirms the presence of global changes in the enzyme's structure. Complex I from bovine and ovine mitochondria was prepared for high-resolution studies in Cymal-7 and Brij-35 detergents respectively. The porcine and medium resolution ovine structures come from the enzyme purified as respirasome and in case of the first one refined to high resolution focusing only on the complex I density of the particles.

Interestingly the overall conformations presented in the bovine and porcine studies are somewhat different to conformations obtained from the ovine complex I sample (Chapter 4.2). The bovine study presents three forms of the enzyme. Class 1 with the position of PA over the MD and PA rotation similar to the ovine classes 1 and 2. Class 2 is even more closed and anti clockwise

rotated than ovine class 1 and Class 3 is analogous to ovine class 3. The most striking difference between studies is that complex I particles from the bovine study mostly allocate to the "closed" conformation, in contrast to the ovine where they are allocated more to the "open" classes. Thus most "closed" bovine Class 2 and most "open" ovine classes 4 and 5 are unique and do not have analogues (Fig. 4.28). All the classes also vary noticeably by the position of the distal ND5 and surrounding subunits. In this case change is more visible than in five ovine classes (Chapter 4.2). Anti clockwise rotation of the tip (around the ND5 subunit) of MD (looking from IMS) correlates with anti clockwise rotation of the peripheral arm (Fig. 4.28.D).
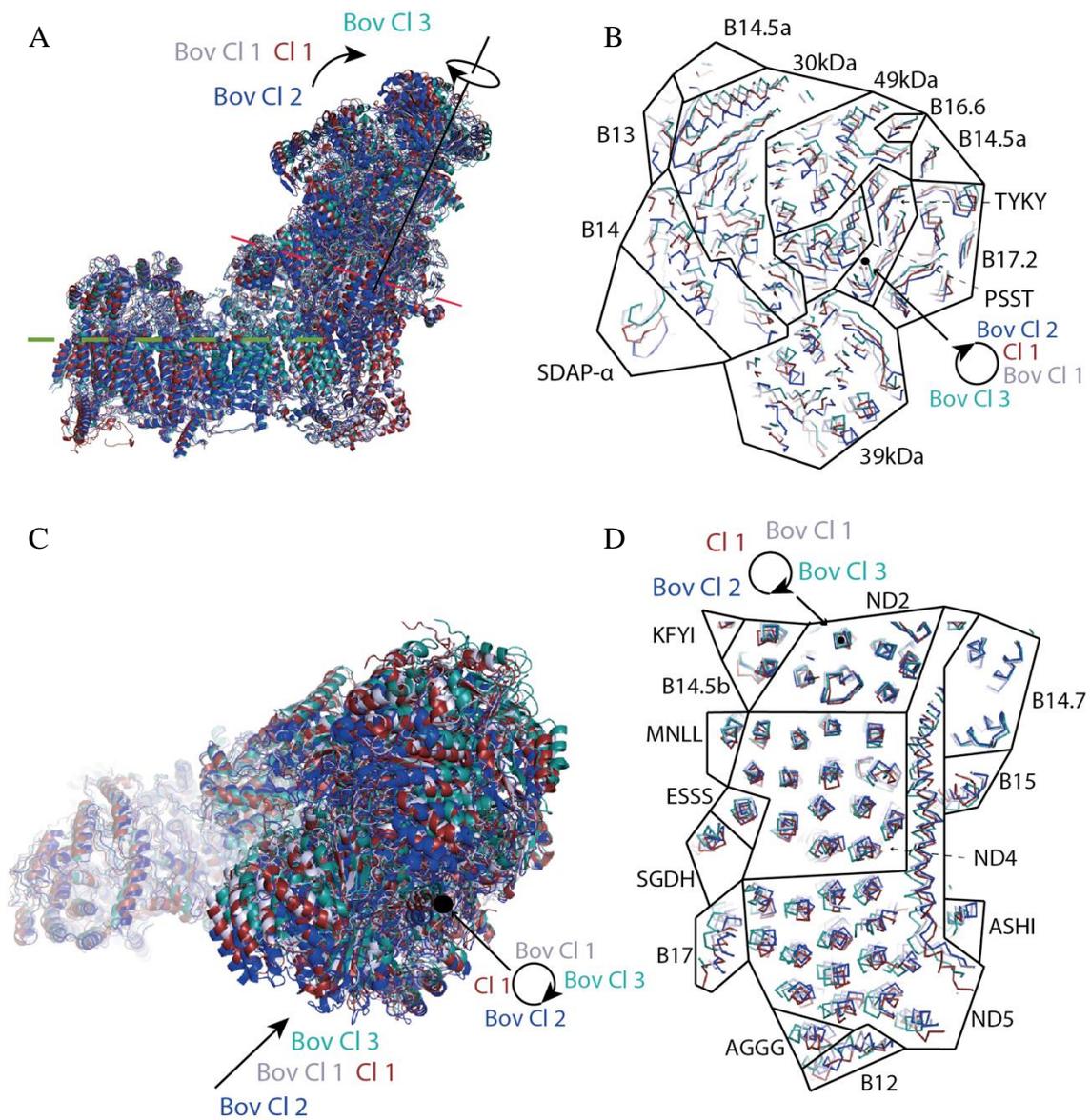
**Figure 4.28** Comparison of class 1 of the new classification and three classes published in the bovine complex I study. **A.** Side view of complex I with rotation axis and direction of PA movement. Red dashed line shows cross section presented on panel B; green dashed line shows cross section on panel D; solid line presents rotation axis of the PA; arrows show direction of the rotation and movement of the PA. **B.** Cross section of complex I peripheral arm. Arrow is showing direction of rotation; black dot shows the axis of rotation. **C.** Matrix side view of complex I. Arrows indicate direction of the rotation and wedge like movement of the PA; black dot indicates the axis of rotation. **D.** Cross section of complex I membrane domain showing direction of rotation of the distal part of membrane domain; black dot shows the axis of rotation. Structures depicted as cartoons (A. and C.) or ribbons (B. and D.). Class 1 in red, bovine class 1 in pale blue, bovine class 2 in blue and bovine class 3 in teal. Structures aligned by the core subunits ND1-ND4.

The porcine structure is a result of focused refinement of the complex I and III region in a digitonin-solubilized $I_1III_2IV$ supercomplex. The enzyme displays a closed conformation, similar to bovine class 2 structure but is more clockwise rotated. Compared to the ovine maps, it is most similar to class 1 with a stronger clockwise rotation and moved closer to the MD (Table 4.2). Interestingly the rotation axis of the PA is on the opposite side comparing to all other structures (Fig. 4.29 B). Also the membrane domain is slightly rotated in the anti clockwise direction looking from the matrix side of mitochondria (Fig. 4.29 D)
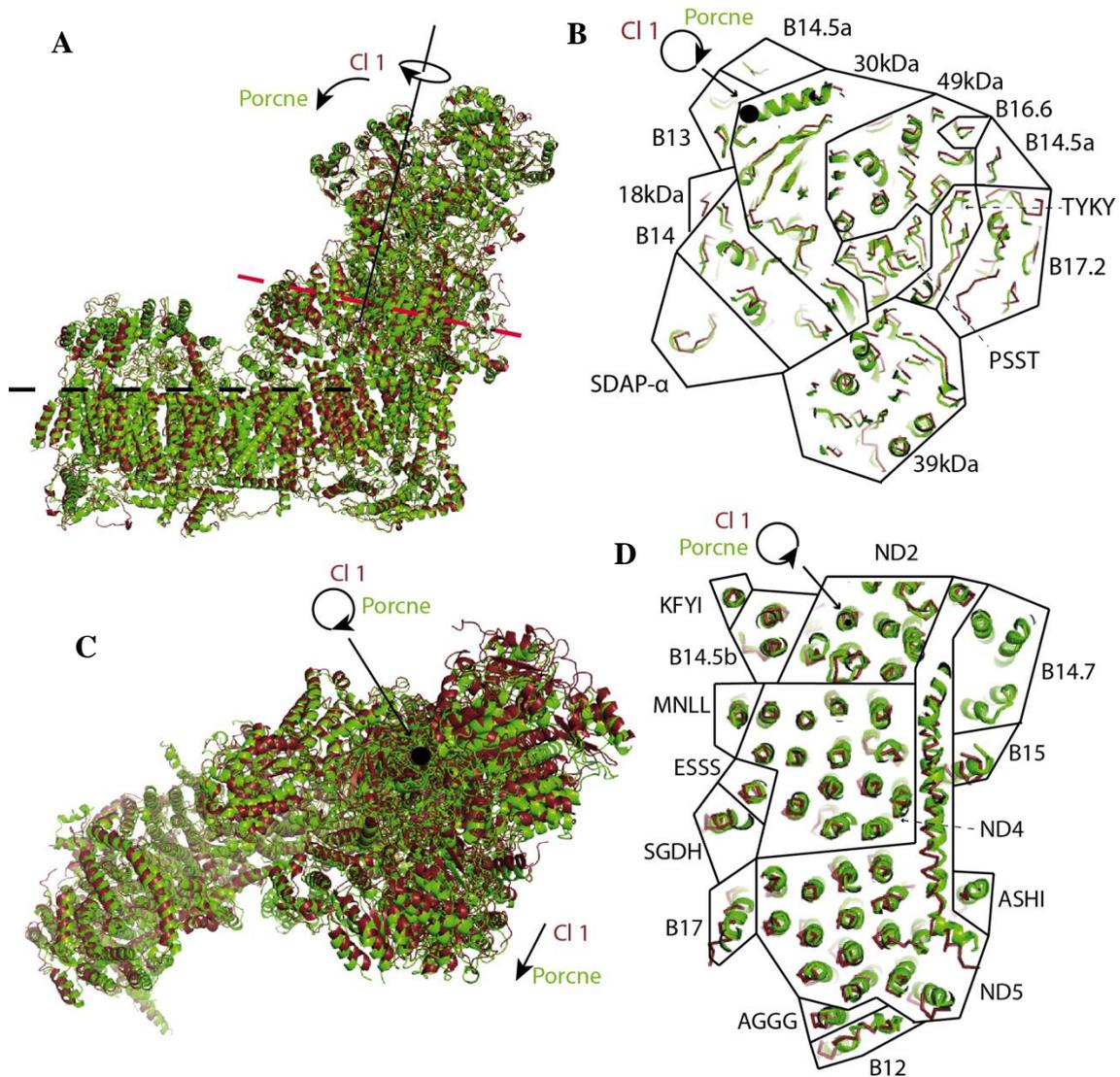
**Figure 4.29** Comparison of class 1of the new ovine classification and porcine model based on mitochondrial respirasome study. **A.** Side view of complex I with rotation axis and direction of PA movement. Red dashed line shows cross section presented on panel B; black dashed line shows cross section on panel D; solid line presents rotation axis of the PA; arrows show direction of the rotation and movement of the PA. Note that the rotation axis is on the opposite side of the PA comparing to other structures (Fig. 4.28 and 4.24-4.25) **B.** Cross section of complex I peripheral arm. Arrow is showing direction of rotation; black dot shows the axis of rotation. **C.** Matrix side view of complex I. Arrows indicate the direction of rotation and hinge-like movement of the PA; black dot indicates the axis of rotation. **D.** Cross section of complex I membrane domain showing direction of rotation of the distal part of membrane domain; the black dot shows the axis of rotation. Structures depicted as cartoons (A. and C.) or ribbons and cartoons (B. and D.). Class 1 in brown, porcine model in green. Structures aligned by the core subunits ND1-ND4.

The ovine complex I model from the respirasome study is currently at the lowest resolution. It is most similar to moderately "closed" ovine class 2. The position of the PA is very close and almost no differences in the MD were observed. The model was build based on ~5.8 Å resolution map which allowed modelling of main chains only. It is possible that due to low number of particles different complex I conformations were difficult to separate.

Importantly all the studies confirm that the global "closed" conformation of complex I correlates with an accessible Q site of the enzyme. The question if the "open" conformation is correlated with inserted β1-β2$^{49\text{-kDa}}$ loop in species other than ovine remains unclear, however this theory is worthy of further investigation (Chapter 4.3.2).

An interesting fact is that in the high resolution maps of the "closed" complex I conformation (ovine class 2 and porcine), the N-terminal extension of the 49 kDa subunit followed by the β1$^{49\text{-kDa}}$ is well ordered and possible to trace along the MD on the matrix side of the enzyme up to the 42 kDa subunit. In the ovine "open" map, this extension is disordered. Based on this observation it could be proposed that the N-terminus of the 49 kDa subunit may play a role in the open/closed transition. However it is unlikely because the following β1–2$^{49\text{-kDa}}$ strands are positioned identically in all the structures, it interacts extensively with many other MD subunits, remains disordered in the ovine class 1 map and is missing in the bacterial enzyme ruling out its catalytic function. Nevertheless, the 49 kDa extension interacts with the TMH2-3$^{ND3}$, which changes position upon A/D transition, thus increased flexibility of this coil cannot be ruled out (Fig 4.29).

In the TM domain the TMH5-6$^{ND1}$ loop, building the ubiquinone chamber, is poorly resolved in the bovine and ovine studies but is better defined in porcine (Wu et al., 2016). That indicates a requirement for some flexibility in this region, possibly linked to the A/D transition (4.3.2) (J. Zhu et al., 2016). In the bovine and porcine models the loop is retracted, making contact with the β1–β2$^{49\text{-kDa}}$ loop (4.29.A). Another crucial element linked to the A/D conversion is the α1-α2$^{ND3}$ loop (Babot & Galkin, 2013). In the bovine Class 2 ("closed") and porcine

models it is relatively well resolved but is less defined (more flexible) in the *O. aries* "open" conformation. In the bovine and porcine models the loop goes in between the ND1, 49 kDa and PSST subunits making contacts with the β1–β2[49-kDa], so that key Cys39[ND3] is buried in the protein interior. In the ovine model ("open" state) this contact is not so extensive. This cysteine becomes accessible to chemical modification only in the D form of the enzyme (A. Galkin et al., 2008), perhaps reflecting the possibility that "closed" conformations resemble the A form, while "open" state is closer to the D form (Fig. 4.29.B).

It has been proposed that upon quinone reduction, four protons are translocated through four channels, one each in antiporter-like subunits ND5, ND4 and ND2 plus one in ND1+ND4L+ND6 subunits (Baradaran 2013). All characteristic features of the transmembrane part of complex I are present in all mammalian models. Just like in the bacterial structures (R. Baradaran et al., 2013)(Efremov & Sazanov, 2011) TMH7 and TMH12 of ND2, ND4, ND5 contain a loop insertion in the middle of the membrane, while TMH8 has a π–bulge a at similar position. The long chain of charged and polar residues running along the central axis of the TM domain is well preserved as well. Maps enable modelling of conserved glutamates, aspartates, lysines and histidines. Those key residues are proposed to be crucial for the energy transduction from the Q-site to the very end of membrane domain. Conformational changes of chargeable residues may propagate through the central axis altering local pKas and accessibility of side chains central to proton transport (L. A. Sazanov, 2015)(R. Baradaran et al., 2013). In all models big cavities possibly filled with water and providing entry and exit points for protons are present.

The main difference in the proton-pumping sites that can be found between the models is the position of the ND5 subunit. Its movement is correlated with the direction of the PA rotation (Table 4.2, Fig. 4.28-29). Comparing to ovine class 1, looking from the matrix side PA and ND5 subunit in accord rotate in clock or anti clockwise directions which correlates with "closing" and "opening" of the complex (please note that in the porcine model it is a clockwise/anticlockwise relationship but the pivot point of the PA in the porcine model is on the opposite

side so the overall direction of the changes remains the same). Perhaps it is important for the catalytic mechanism of the enzyme.

It will require further investigation why in the bovine and porcine models the distal pumping module is rotating and in the ovine classes (exception class 5) it "kinks". The exact character and importance of these transitions is difficult to discuss due to the moderate resolution of the maps and relatively small changes in this region. Higher resolution models of various conformations may allow us in future to discern smaller conformational changes around key residues in other pumping modules.

*4.4.2 Conformational variability of complex I – possible links to the active/de-active state transition or to the catalytic cycle.*

Based on our and other recently published structures, it can be proposed that when the ubiquinone binding site is empty, the inserted β1–β2[49-kDa] loop covers the N2 cluster and key Tyr108[49-kDa], possibly preventing ROS production at the Q-site in the presence of NADH and absence of Q. In this model, the "retracting" loop movement would be initiated by the approaching ubiquinone molecule (or any entity/inhibitor binding at this site). Accommodation of a ligand would bring β1–β2[49-kDa] and TMH5-6[ND1] loops together and may stabilise the interface between PA and MD, through the interaction mediated by the long α1-α2[ND3] loop (better ordered in the "closed" state). Stacking interactions between the quinine head group and His59[49-kDa] may shift an entire 49 kDa β1–β2 sheet by about 2 Å (a transition observed between 'open' and 'closed' classes). The initial move would facilitate "embracing" interactions of B13 and SDAP-α pair with the 42 kDa subunit, stabilizing complex I in the "closed" conformation. Upon quinol (reduced quinone) release the 49 kDa/ND1/ND3 connection is lost and B13/SDAP-α/42 kDa interaction is not strong enough to keep the arms together, which results in enzyme's return to the initial "open" conformation. This is of course one of many possible mechanisms for "open"/"closed" state transition and more studies have to be conducted to confirm or rule it out.

These conformational changes may describe the transition between de-active and active states of the enzyme (respectively from "open" to "closed") or reflect the transitions that complex I undergoes during the catalytic cycle, with the "open" state occurring when quinol is ejected from the cavity (possibly helped by β1–β2$^{49\text{-kDa}}$ loop moving in) and "closed" state adopted once quinone is bound. Indeed, the A/D transitions and the catalytic cycle may follow the same conformational trajectories, except that in the D state conformation the enzyme is trapped by some protein modification (like that of Cys39$^{ND3}$) requiring several turnovers to be released during activation (Vinogradov, 1998).

A
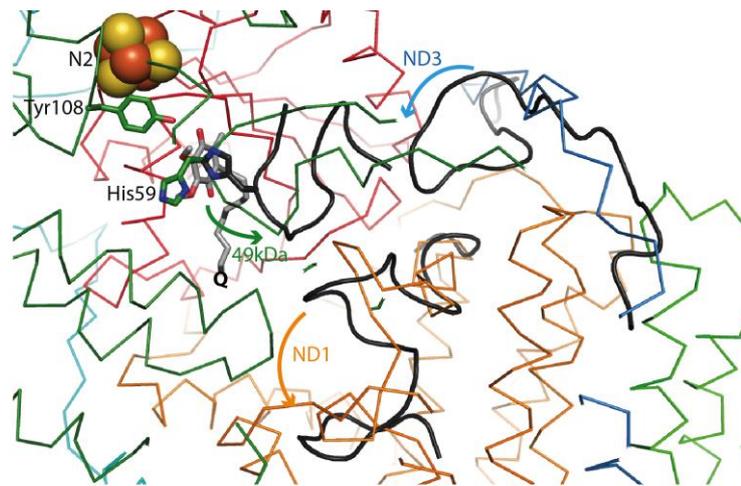


B
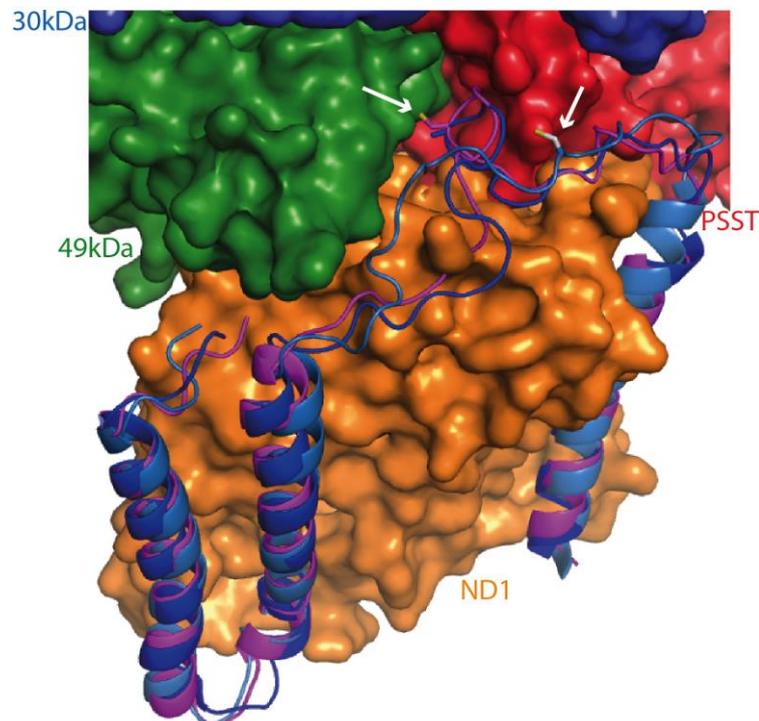
**Fig. 4.30 A.** Presentation of changes in positions of crucial loops of the 49 kDa, ND1 and ND3 subunits between "open" and "closed" states of complex I. Arrows indicate hypothetical movement of the loops upon ubiquinone binding. Presented as ribbon with the structure of the "open" ovine complex I in colour and the crucial loops from the bovine Class 2 ("closed" state)complex I in black; those loops were build very similarly in the porcine model. N2 cluster presented as spheres, crucial Tyr108$^{49\text{-kDa}}$ as sticks, quinone modelled based on the *T. termophilus* structure as sticks in gray. Structures were aligned by theND3 subunit to model the TMH1-2$^{ND3}$ and TMH5-6$^{ND1}$ loops and by the 49 kDa subunit to model the β1–β2$^{49\text{-kDa}}$ loop. **B.** Presentation of the TMH1-2$^{ND3}$ position in the bovine, ovine and porcine high-resolution models. The ND3 subunit presented as cartoon, ovine in light blue, bovine in blue and porcine in magenta. Arrows indicate position of the conserved Cys39 (Chapter 4.1.1) Subunits 30 kDa, 49 kDa, ND1 and PSST presented as surface and coloured.

**4.5 Disease causing mutations**

Mitochondrial diseases are the most common group of inherited metabolic disorders. Mostly they are caused by isolated complex I deficiency. The defects result from inherited mutations in mitochondrial or nuclear genes encoding complex I subunits or assembly factors. Additionally, oxidative stress and mitochondrial division can give rise to new mutations over an individual's lifetime. Dysfunction of the complex is mostly visible in organs with high-energy demand like the brain. Complex I has been implicated in many neurodegenerative disorders but also linked to cancer and ageing.

*4.5.1 Overview*

Complex I deficiencies are genetic disorders that are caused by mutations in genes in the nuclear DNA (nDNA) and mitochondrial DNA (mtDNA) that encode mitochondrial complex I subunits or proteins involved in its assembly. About 30% of mitochondrial diseases affecting energy metabolism can be traced to mutations in one of the subunits of complex I (Kirby et al., 1999). The diseases are characterized by defects in oxidative phosphorylation; however, symptoms vary widely in nature and severity from lethal to mild disorders. In general prognoses are poor. About 50% of patients diagnosed with complex I deficiency die within first two years of life and only 25% reach the age of ten (Koene et al., 2012). The first signs are usually observed shortly after birth and affect the central neurological system. The most common clinical presentations are: fatal neonatal lactic acidosis, infantile-onset Leigh syndrome, childhood-onset mitochondrial encepalomyopathy, lactic acidosis and stroke like episodes (MELAS) syndrome, Leber hereditary optic neuropathy (LHON) (Fassone & Rahman, 2012).

44 different proteins and a cohort of assembly factors acting in a strictly regulated manner are required for proper functioning of mitochondrial complex I (Chapter 1.4.2) (Sanchez-Caballero, Guerrero-Castillo, & Nijtmans, 2016). Thus even single mutations in any of these proteins can affect enzyme performance. To the numerous vulnerable genes directly responsible for complex I formation

we can also add those involved in metabolism of related co-factors and ligands (Shinzawa-Itoh et al., 2010).

In patients suffering from complex I deficiency, mutations have been found in all of the core subunits and in most accessory subunits (Rodenburg, 2016). Diseases are caused by a wide range of mutations in nuclear or mitochondrial DNA, including substitutions, insertions, deletions, inversions, duplications etc. In this chapter I will focus only on single amino acid mutations that cause complex I deficiency. It is important to remember that disease causing mutations are placed in regions relevant but not absolutely crucial for complex I activity. They only cause a decrease in complex I activity. Embryos with mutations in amino acids crucial for complex I function are eliminated in the early stages of development. Single amino acid substitutions in patients suffering from complex I deficiency have been identified in 27 subunits (14 core and 13 supernumerary) (Table S5). They are mostly localized around electron transfer pathways and proton pumping sides.
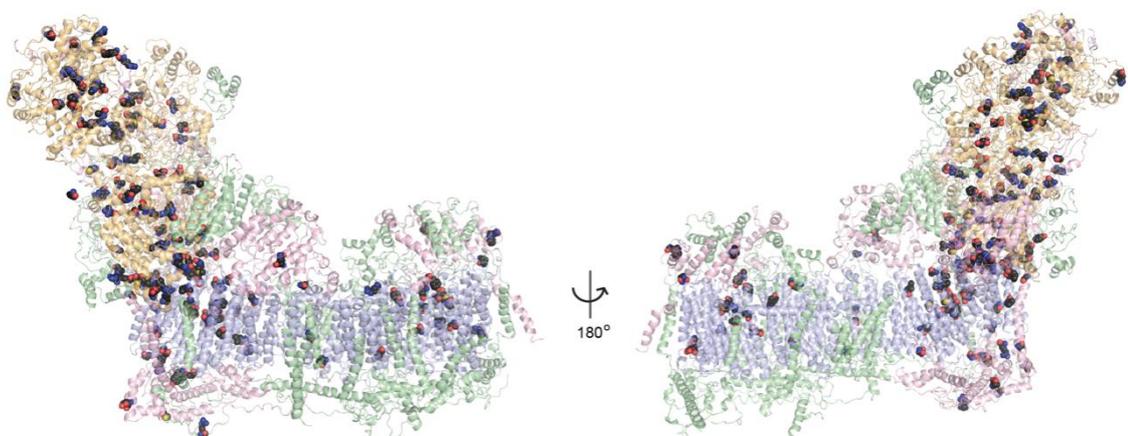


**Figure 4.31** Overview of complex I deficiency causing mutations. Complex I depicted as transparent cartoon with core PA coloured in orange, core MD in blue, SN subunits bearing mutations in pink and mutations free subunits in green.

A general genotype-phenotype correlation has not been clearly established so far (Tucker, Compton, Calvo, & Thorburn, 2011). It is not well understood how mutations in certain complex I related genes translate to specific clinical phenotypes. An oft-cited example of this ambiguity is a mutation in position Arg340 of the ND4 subunit. Substitution to histidine causes LHON syndrome

(Wallace et al., 1988) but mutation to serine results in Leigh syndrome (Deschauer et al., 2003). Until now a better understanding of complex I dysfunction caused by single substitutions has been obstructed by a lack of high-resolution models for mitochondrial complex I.

Information about disease causing subunits in complex I were acquired from HGDM, ClinVar and Mitomap databases and exact references for all of them can be found in Table S5. Throughout this chapter amino acid numbering will be consistent with the sequence of the mature human proteins. Names of nuclear encoded human subunit homologues are in brackets in bold in headers of each paragraph of 4.5.2-4.5.4 next to the bovine names used throughout the chapter. Information about the nomenclature of subunits and ISCs can be found in Tables S6 and S7.

*4.5.2 Core Subunits Peripheral Arm*

**75 kDa (NDUFS1) subunit**
The most frequent outcome of mutations in the 75 kDa subunit is Leigh syndrome (Table S5). Val48Asp, Gly143Glu, Val205Ala, Leu208Val, Arg218Trp are localized in the close proximity of iron-sulphur (FeS) clusters (Table S6). They build loops crucial for the proper binding and stability of the clusters. Val48 is located ~14Å (all distances are measured from the α-carbon of the amino acid to the edge of FeS cluster, unless indicated) away from the N3 and N1b clusters in a loop interacting directly with cluster-coordinating chains. Gly143 is also ~14Å away from the N1b cluster. It is part of a loop followed by a short β-sheet and a coil that binds the N1b cluster. Val205 is located in a loop between the N3 and N1b clusters 5.4 and 3.9Å away respectively. Leu208 also forms part of this loop. Arg218 is located ~11Å away from clusters N5 and N6a. The η-nitrogens at the side chain of Arg218 are at distances allowing direct interaction (~3Å) with the main chain carbonyl oxygens of cysteines binding these clusters (Cys108$^{75-kDa}$ and Cys80$^{PSST}$).

Interestingly, in the case of amino acids located ten or more Ångstrom away from FeS clusters the disease-causing mutations involve substitutions to

residues of a more extreme nature, involving amino acids of very different chemical characteristics (Val48Asp, Gly143Glu, Arg218Trp). Such alterations are likely to disrupt secondary structure elements. In the case of residues located very close (~5Å) to catalytically important centres, even a change to chemically similar amino acid decreases the efficiency of the enzyme (Val205Ala, Leu208Val).

Other mutations such as: Asp229Gly, Arg385Cys, Gln499Lys, Thr572Ala, Asp596Asn, Tyr672His/Cys and Met684Val seem to cause complex I dysfunction due to improper complex assembly. They all are placed in the region of the 75 kDa subunit distal C-terminal subdomain at loops, β-strands or α-helices that are central for proper folding of this region.
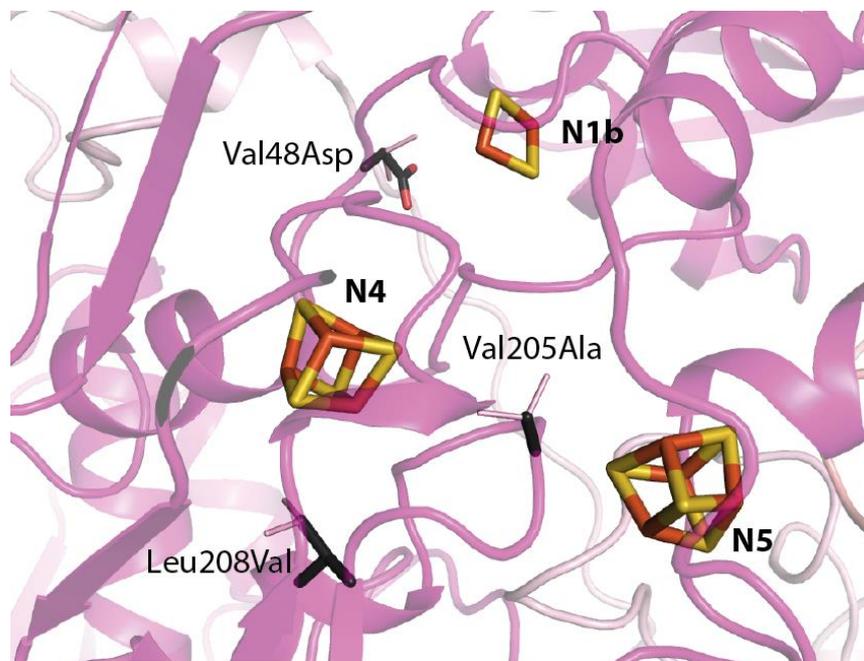


**Figure 4.32** Examples of mutations in the 75 kDa subunit. The 75 kDa subunit pictured as cartoon in magenta, natural variants of amino acids as light pink sticks and disease causing substitutions as black sticks. Co-factors are labelled.

**51 kDa (NDUFV1) subunit**

The 51 kDa subunit is one of the PA subunits with the highest number of reported single amino acid mutations – 17 (Table S5). Many are located around the FMN binding site where NADH is oxidized and electron transfer starts: Arg68Gly, Tyr184Cys, Cys186Gly, Ala191Val. Others like Glu357Lys, Pro102Leu, Glu194Lys, Arg366His/Cys, Thr403Met, affect FeS cluster binding

or like Ser36Pro, Lys91Glu, Arg127Trp, Arg179Pro, Pro232Arg, Ala321Val are likely to affect the stability of the whole NADH oxidizing module. Mutations that are not placed closely to the catalytically important centers of the subunit and are more likely to affect assembly are: Ser36Pro, Lys91Glu, Arg127Trp, Arg179Pro, Pro232Arg, Ala321Val.

Arg68 is one of the amino acids building the FMN binding pocket. The nitrogen from its peptide bond is at hydrogen bond distance (2.65 Å away) from one of the terminal oxygen atoms from the ribitol group of FMN. The side chain of arginine, although pointing in the opposite direction, is important for the stability of this region due to interactions with surrounding amino acids. Tyr184 has a bulky side chain contributing to the FMN binding site architecture with the hydroxyl group of the tyrosyl only 3 Å away from the flavin ring. A substitution to cysteine would completely change the hydrophobic properties of this region. Cys186 does not interact directly with the ligand but mutation to glycine probably changes positions of the nearby residues, which are crucial for FMN and NADH binding, including Glu189. Similarly, Ala191, which does not interact directly with FMN or cluster N3, is located very close to FMN coordinating residues and interacts with Pro203 whose side chain is only 3.4Å from the N3 cluster.

Pro102 is located in a loop between FMN and cluster N1a. The function of this cluster is not entirely known. It has been suggested to have a role in preventing reactive oxygen species (ROS) production (L. A. Sazanov & Hinchliffe, 2006)(Leonid A Sazanov, 2007)(Esterhazy et al., 2008) or purely in assembly/stability (Birrell et al., 2013). The side chain of Pro102 is only 3.7 Å away from N1a and a change to leucine may cause cluster displacement. Glu194 is in a short α-helix placed in between two coils, one stabilizing FMN and the other binding cluster N3. Change of the local charge in this case may lead to displacement of relevant side chains in this area. Additionally, Glu194 is located in the interface between the 51 kDa,24 kDa and 75 kDa subunits, directly interacting with Arg177 of the 75 kDa subunit. Glu357 is separated by only one residue from the conserved Cys359 that coordinates cluster N3. Thus, Glu357Lys very likely will affect positioning of the protein main chain. The same

explanation can be used for the substitutions in Arg366 and Thr403 situated right next to the N3 cluster-coordinating Cys365 and Cys405 respectively.

Other substitutions in the 51 kDa subunit are playing destabilizing roles or perhaps affecting FMN binding and FeS cluster stability in a more indirect way. Lys91 is located in the periphery of the protein. Interestingly, its positively charged side chain is placed close to a coil followed by an α-helix stabilizing FMN. Change of the charge on the side chain may cause displacement of the coil and affect the FMN binding site. Positioning of the Lys91 side chain can also be disturbed by mutation in the proximate Arg127, which is additionally placed at the end of an α-helix central for 51 kDa subunit integrity. The same can be said about Pro232 and Ala321, which are located in the α-helices that follow or are followed by loops that come into close proximity of the FMN site. Mutations in Ser36 and Arg179 are positioned away from the catalytically important locations thus most likely have an effect on the subunit's assembly. Ser36 is located in an outer α-helix of the 51 kDa subunit that makes interactions with the 10 kDa and C-terminal extension of the24 kDa subunit. Arg179 is found in a β-strand that also interacts with the 10 kDa and 24 kDa subunits and additionally is followed by a loop that influences stabilization of FMN (however the co-factor binding site is almost 10 residues away).
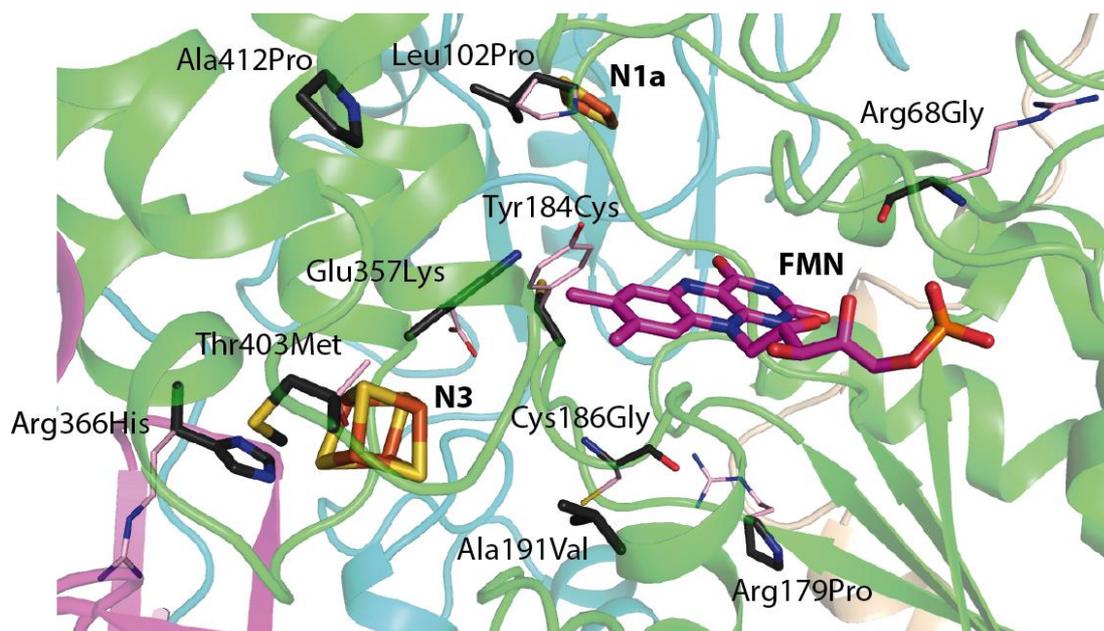
**Figure 4.33** Examples of mutations in the 51 kDa subunit. The 75 kDa subunit pictured as cartoon in magenta, 51 kDa subunit in green, 24 kDa subunit in cyan, natural variants of amino acids as light pink sticks and disease causing substitutions as black sticks. Co-factors are labelled.

### 49 kDa (NDUFS2) subunit

The Arg85Gln, Arg105Gln, Ala191Val mutations possibly influence electron transfer from the terminal N2 cluster to ubiquinone.

In mammalian complex I, Arg85 undergoes an unusual dimethylation (Rhein, Carroll, Ding, Fearnley, & Walker, 2013). The guanido group of the side chain is in direct contact with cluster N2. It has been proposed that it contributes to the relatively high reduction potential of the cluster (L. A. Sazanov & Hinchliffe, 2006)(Hirst & Roessler, 2015). Mutation in this position will likely impact the redox potential of the cluster. Arg105 similarly to Arg85 is placed very close to N2,with one nitrogen of the side chain only 2.8 Å away from the nearest sulphur of the cluster and its mutations have a drastic effect on the properties of this redox center (Grgic, Zwicker, Kashani-Poor, Kerscher, & Brandt, 2004). Although Ala191 does not influence the N2 cluster directly, it is placed next to His190, which was proposed to be the redox-Bohr group associated with the cluster (Zwicker et al., 2006) and like the arginines, defines the environment for the N2 cluster.

Other mutations of the 49 kDa subunit are localized in regions important for the subunit's stability or interactions with other subunits. Asp77Val is found in the middle of a β-strand that is part of a β-sheet interacting with the 30 kDa, PSST and ND1 subunits. Glu115Lys is in an α-helix located in the centre of the 49 kDa subunit. Arg195Gln and Pro196Gln are on the interface with TYKY. Met259Thr is part of a short loop that interacts with the 30 kDa, TYKY, B13 and 14.5b subunits. Arg290Gln is in a coil that probably stabilizes interactions with TYKY and B13. Arg300Gln is placed in the middle of an α-helix near coils of B16.6, TYKY and B14.5a. Ser380Pro and Ala393Ser are in a loop and β-sheet, forming interactions with the 30 kDa subunit. Met410Lys and Asp413Asn are in a loop followed by an α-helix that comes in contact with TMH5-6$^{ND1}$, which in turn interacts with the mechanistically relevant β1-β2$^{49-kDa}$.
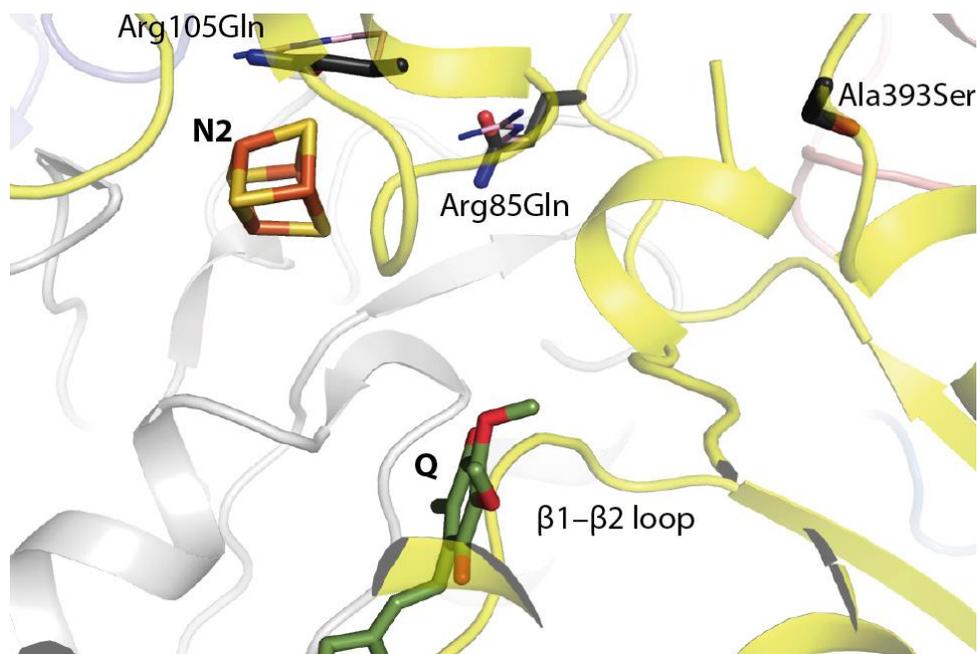


**Figure 4.34** Examples of mutations in the 49 kDa subunit. The 49 kDa subunit pictured as cartoon in yellow, PSST subunit in grey, natural variants of amino acids as light pink sticks and disease causing substitutions as black sticks. Co-factors are labelled. Quinone position modelled on *T. thermophilus* structure. Notice β1-β2$^{49-kDa}$ loop clashing with Q head group.

**TYKY (NDUFS8) subunit**

The N-terminal end of the TYKY subunit is composed of two connected amphipathic α-helices placed on the surface of the internal mitochondrial

membrane (IMM), on the matrix side. One possible role of those α-helices is to stabilize the position of the PA with regards to the lipid bilayer (Fiedorczuk et al., 2016). This makes them important for the regulation of the proposed hinge/rotation like move of the PA (J. Zhu et al., 2016)(Fiedorczuk et al., 2016). The relevance of these α-helices can also be demonstrated from the analysis of disease causing substitutions. Arg20Trp, Glu29Gln, Arg43Trp, Pro45Leu are all located either in or in close proximity to these helices. Additionally, this region of TYKY is in contact with the B16.6, ND1, B9, B14.5a subunits and a cardiolipin molecule.

Other mutations, Arg68His, Arg104His, Gly120Ser, Ala125Asp likely cause destabilization of iron-sulphur clusters. Arg68 and Arg104 are found in coils in direct contact with loops binding cluster N5. Gly120 and Ala125 are placed right next to Cys119 and Cys126, responsible for coordination of clusters N6b and N6a, respectively.

The role of Pro51Leu and Arg60Cys mutations can be assigned to mis-assembly of the subunits. Pro51 is at the loop localized between the B17.2 and PSST subunits and Arg60 can be found in a loop making the interface with the 49 kDa subunit.
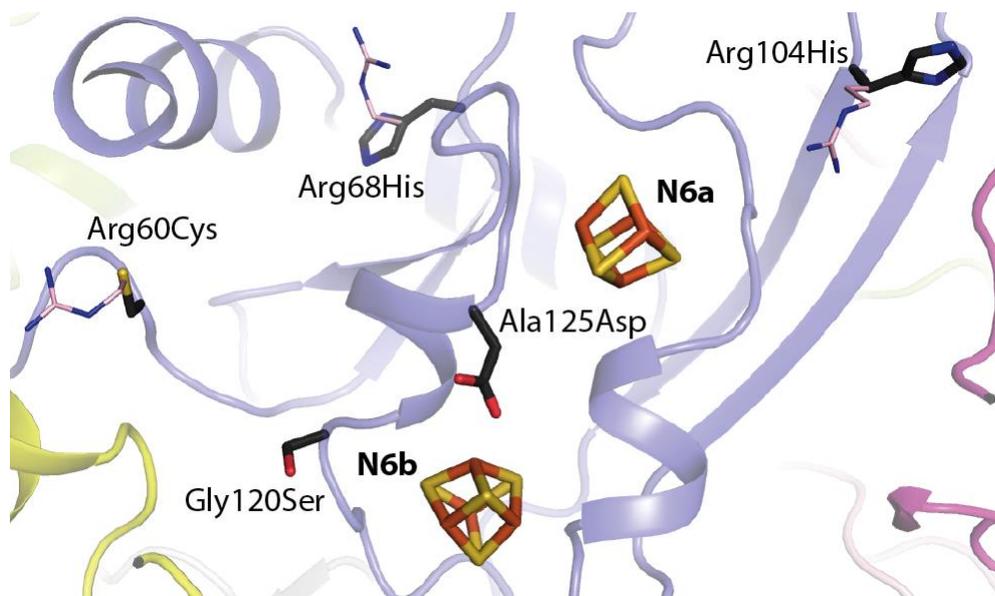
**Figure 4.35** Examples of mutations in the TYKY subunit. The TYKY subunit pictured as cartoon in blue, natural variants of amino acids as light pink sticks and disease causing substitutions as black sticks. Co-factors labelled.

## PSST (NDUFS7) subunit

Val84Met and Arg107His are found in two parallel β-strands that are followed by loops coordinating the terminal N2 cluster. This four strand β-sheet makes up a great part of the PSST subunit and is central for the stability of this region at the base of the PA.

## 30 kDa (NDUFS3) and 24 kDa (NDUFV2) subunits

The 30 kDa and 24 kDa subunits are the only subunits of the core peripheral domain that do not take part directly in electron transfer, but rather provide a platform for subunits assembly (Friedrich & Scheide, 2000). Identified disease causing mutations Thr109Ile, Arg163Trp and Pro187Leu of the 30 kDa subunit are located mostly at the interfaces with other subunits. Thr109 is located in the middle of the large 5-strand β-sheet. The sheet forms most of the 30 kDa subunit and is central for stability of this region. Arg163 and Pro187 are in the long C-terminal coil of the 30 kDa subunit that interacts with numerous neighbouring subunits. Arg163 is found in a loop that directly contacts the 49 kDa and 18 kDa subunits. Pro187 is found at the end of a "hairpin-like" loop inserted in between the PSST, 49 kDa and B14 subunits. The single Lys177Arg

substitution in the 24 kDa subunit is located at the very periphery of the PA and its role is not clear.

*4.5.3 Core Subunits Membrane Domain*

**ND1 subunit**

Mutations in this subunit cause a broad spectrum of clinical features, including Alzheimer and Parkinson Disease, MELAS or Leigh syndrome. Interestingly Ala52Thr mutation is one of the most frequently encountered mutations in LHON syndrome, identified in approximately 13% of cases (together with Ara340His[ND4] [69%] and Met64Val[ND6] [14%] they cause almost 95% of cases). The broad variety of the diseases caused by substitutions in the ND1 region can be explained by the subunit's involvement in several different functions. It contains the quinone-binding chamber, creates the proton input channel for the first translocation module and directly enables coupling: it links the Q-chamber with proton pumping modules. Disease causing mutations may affect all of those functions. Those that probably impact the N2 cluster accessibility to quinone are: Val11Met, Glu24Lys, Leu28Met, Tyr30His, Met31Val, Ala52Thr, Met53Ile, Glu59Lys, Tyr277Cys. The transfer of the information about quinone reduction/protonation from the Q to P module may be affected by: Val208Leu, Arg195Gln, Glu214Lys, Tyr215His. Mutations Ser110Asn, Ala112Thr, Glu143Lys, Gly131Ser, Ala132Thr are most likely to affect proton pumping activity of the first proton channel (so-called E-channel in ND1+ND4L+ND6). Additionally, a number of substitutions, Thr164Ala, Thr240Met, Leu285Pro, Leu289Met, cause dysfunction, probably by affecting the structure's stability.

Three α-helices are building the triangular entry point of the quinone channel: TMH1, amphipathic α-helix 2 and TMH2. Val11, Glu24, Leu28, Tyr30, Met31 are placed in the first TM helix and Ala52, Met53 and Glu59 on the amphipathic α-helix 2. Mutation at any of these positions can partially obstruct access to the Q-site either by destabilizing secondary structure elements or introducing a side chain blocking it directly. Tyr277 is on one of the loops forming the Q-channel so substitution to cysteine will affect the architecture of the path. Additionally,

amphipathic α-helix 2 of the quinone entry site and the side chain of Tyr277 are both in contact with the amphipathic α-helix of PSST stabilizing this region.

The quinone reduction site and the first proton pumping module are linked by a number of polar and charged residues that are proposed to link the sites via conformational changes of side chains (L. A. Sazanov, 2015). Val208, Arg195, Glu214 and Tyr215 are located in this area and mutations in these positions are likely to disturb conformational coupling.

Mutations in positions Ser110, Gly131, Ala132, Glu143 are all localized around the E-channel and may affect translocation of protons. Especially important seems to be the conserved residue Glu143, connecting both half-channels of the proton translocation unit (L. A. Sazanov, 2015).

Structurally important mutations in Thr164 and Thr240 are placed in the "heel" of complex I from the IMS. Thr164 is located in the centre of a short α-helix stabilizing TMHs 2-6 and Thr240 and Thr240 can be found very close to the C-terminus of TMH6. Mutations in this region may indirectly affect quinone entry site or stability of ND1 and its interactions with supernumerary subunits. Substitutions Leu285, Leu289 are located at the C-terminal TMH of ND1 facing the protein interior. This is the region involved in interactions with the ND3 subunit and mutations may influence this interaction and/or the core fold of ND1.
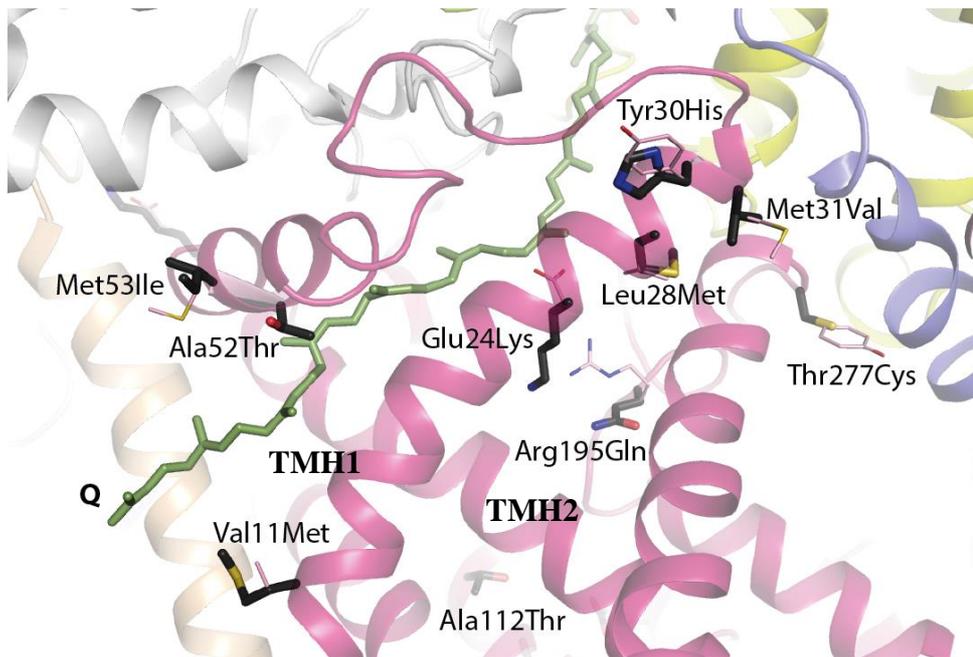
**Figure 4.36** Examples of mutations in the ND1 subunit. The ND1 subunit pictured as cartoon in red, natural variants of amino acids as light pink sticks and disease causing substitutions as black sticks. Quinone acyl chain (green) modelled based on *T. thermophilus* structure. TMHs building entry site for ubiquinone and ubiquinone are labelled.

### ND6 subunit

Mutations in the ND6 subunit can be classified into three groups, those affecting proton translocation directly: Gly36Ser, Tyr59Cys, Met63Val, Met64Ile, Met64Val (14% of all LHON patients), Leu60Ser; those affecting proton translocation indirectly: Ile26Met, Ala72Val, Ala74Val; and those likely to have an effect on complex I structural stability: Asn117Asp, Ser132Leu.

Tyr59 is a conserved amino acid proposed to be crucial for proton translocation through the membrane in the IMS half-channel of the E-channel. In *E. coli,* mutations of this residue to cysteine or phenylalanine decrease complex I benzoquinone reduction activity by approximately 48 or 57% respectively (Kao et al., 2005). The function of the Tyr59 or other neighbouring key residues can also be affected by the mutations in the nearby Gly36, Leu60, Met64 and Met63. Most of these residues sit on the highly conserved TMH3, which contains □-bulge in the middle near Tyr59, and may be directly involved in

propagation of conformational changes from the Q-site towards antiporter-like subunits.

Residues Ile26, Ala72, Ala74 are located on the same or adjacent TMHs to TMH3 (like Tyr59), and interestingly they are all located at the very ends of helices on the matrix side of the membrane. They may either influence the nearby (and crucial for catalytic turnover), TMH1-2$^{ND3}$ loop (Babot & Galkin, 2013) or more likely, even minor disturbances to the structure of key TMH3$^{ND6}$ are detrimental to the coupling mechanism. Mutations in Asn117 and Ser132 are in the loops in the IMS, involved in stabilization of this region. Asn117 is in a loop interacting with ND5, ND4L and ND4; Ser132 is placed in the centre of the complex I "heel" making contacts with several supernumerary subunits.
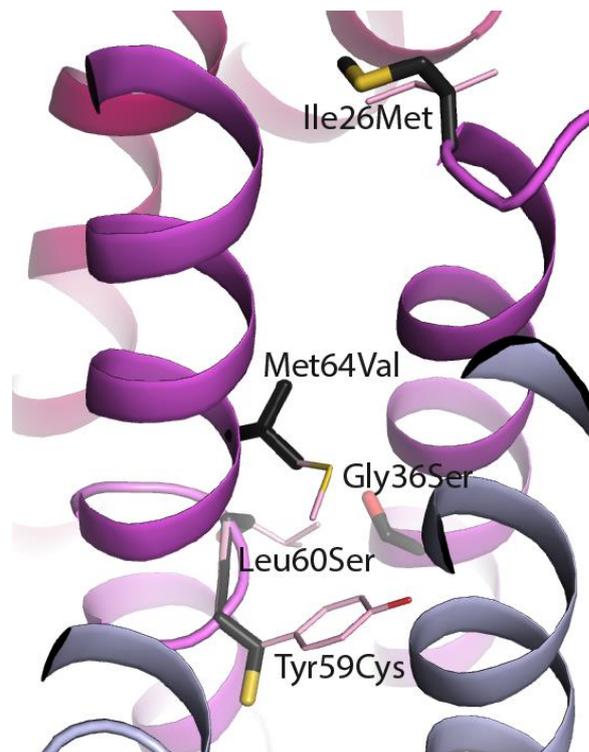


**Figure 4.37** Examples of mutations in the ND6 subunit. The ND6 subunit pictured as cartoon in purple, ND4L in grey and ND3 in red, natural variants of amino acids as light pink sticks and disease causing substitutions as black sticks.

**ND3 subunit**

Most of the mutations in the ND3 subunit, Ser34Pro, Ser45Pro, Gly47Ala, Ile60Thr, Glu66Asn are located in putative catalytically important regions.

Ser34, Ser45, Gly47 are all located in the long loop of ND3, that traverses the membrane domain on the matrix side of the mitochondria. Although none of these amino acids are conserved (Efremov & Sazanov, 2011) they may affect the catalytic turnover of complex I indirectly, since the loop plays a role in the A/D transition of the enzyme (Babot & Galkin, 2013) and stabilises the PA/MD interface.

Glu66 is on a TMH next to key TMH3$^{ND6}$ and so any disturbance in this area may restrict/change conformational space available to TMH3$^{ND6}$ (Fiedorczuk et al., 2016).

The role of Ile60 is difficult to predict. It is placed off the proton translocation site with the side chain at the interface with ND4L and ND6 subunits, thus it probably plays a structural role and substitution to the polar threonine may affect complex assembly.

**ND4L subunit**

Val65Ala and Cys32Arg are both located close to the sites crucial for proton pumping (Fiedorczuk et al., 2016). Val65 is situated next to the protuberance (π-bulge) of the TMH3$^{ND6}$ very close to the key residues of ND4L and ND6. Cys32 is located next to the key Glu34 and Tyr69$^{ND6}$ (Fiedorczuk et al., 2016).

**ND2 subunit**

Mutations of ND2: Ile57Met, Leu71Pro, Gly259Ser, Ala331Ser are most likely influencing protein stability.

Non-conserved residues Ile57, Leu71 and Ala331 are placed at the interface with ND4L and B14.5b respectively. Their side chains are facing the neighbouring subunits and are very likely to be involved in inter-subunit interactions.

Gly259 is in the loop in the centre of broken TMH12, close to the key Lys263. Although this amino acid residue is not likely to be involved in the coupling mechanism, mutation in this position may influence the nearby relevant residues.

**ND4 subunit**

All the reported mutations of ND4: Ile165Thr, Val313Ile, Arg340Ser, Arg340His, Tyr409His are located on the periphery of the subunit in regions likely important for complex I stability, not catalysis.

Ile165 is placed at the C-terminal end of the α-helix forming the main interaction site with ND2. Although this helix contains amino acids important for proton translocation (R. Baradaran et al., 2013), Ile165 is located very far from the crucial central axis of the membrane. Val313 can be found on TMH12 also far from the proton translocation site. The amino acid is not conserved (Efremov & Sazanov, 2011) so it is not likely to play any catalytic role. Arg340 (one of the three common LHON mutations) is found on the TMH12-13 loop on the matrix side of the membrane. The loop is in contact with the TMH10-11 loop and B15 subunit, indicating that the mutation may have a destabilizing effect. Additionally, Arg340 interacts with the negatively charged C-terminus of the key TM7$^{ND4}$, so the mutation may impede catalytically relevant movements of this discontinuous helix. Tyr409 interacts with subunit ASHI, and so the mutation may be destabilising.

**ND5 subunit**

Mutations Phe124Leu, Glu145Gly, Ala171Val, Ala236Thr, Met237Leu, Ser244Cys and Asp393Asn/Gly most likely affect catalytic function of the ND5 subunit. Other substitutions: Tyr159His, Gln434Arg and Gly465Glu probably play a more structural role.

Phe124 is placed in the middle of TMH4. The phenyl ring of the amino acid is pointing into the centre of the pumping module and is in direct contact with Ser250 and Thr251 of the π-bulge at TMH8, which are proposed to be crucial for the input half-channel of the ND5 proton-pumping module. Glu145 is one of

the key residues accepting protons from the matrix side of the membrane and passing it to the proton translocation chain (Fiedorczuk et al., 2016). Ala171 is located on TMH7, close to the cavity making an entry point for protons. Change to a larger and hydrophobic side chain may affect accessibility of the key residues to protons. The same would apply to Ala236 and Met237 situated almost opposite to Ala171 on the broken TMH8. Ser244 located on the π-bulge containing TMH8 is one of the amino acids linking the input and output paths of the ND5 translocation chain. Asp393 is placed next to the key Lys392 and builds the proton outlet cavity in this module. Very likely mutations in all of the indicated positions influence proton-pumping efficiency.

Tyr159 directly interacts with ND4 residues on the matrix site of the membrane, so a mutation in this region probably affects stability of the protein. Gln434 is found on the loop that is in direct contact with B12 and B22 subunits; it is possible that substitution in this region may disturb its stability and influence interaction with complex IV in the supercomplex (Letts, Fiedorczuk, et al., 2016). Gly465 is found in TMH13 directly opposite to the loop linking the two half helices of TMH12. Mutation to a bulkier and charged residue like glutamate most likely will strongly interfere with the role of TMH12 in proton pumping.
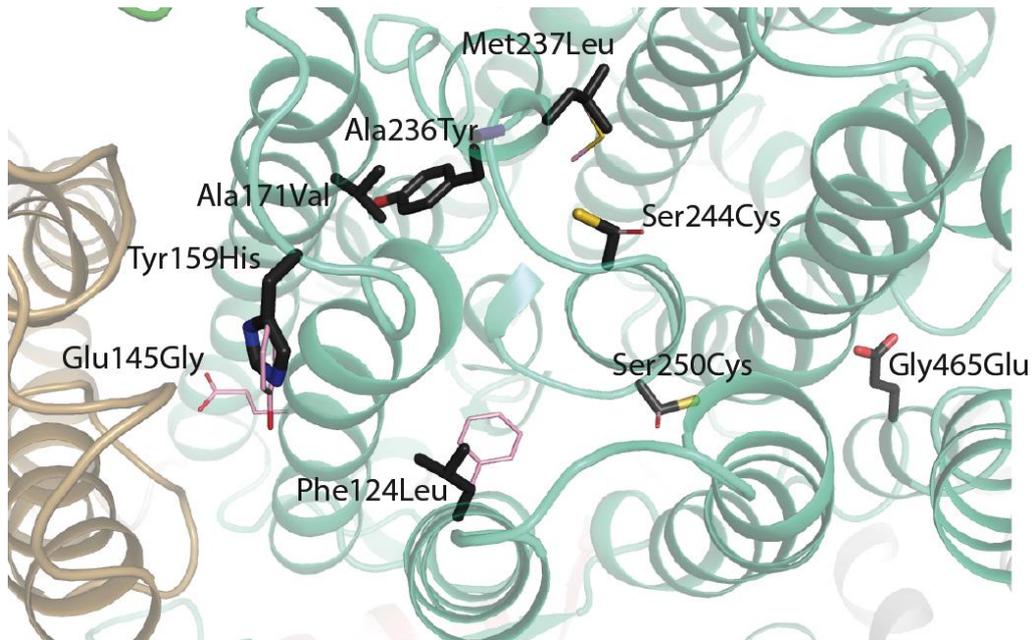
**Figure 4.38** Examples of mutations in the ND5 subunit. The ND5 subunit pictured as cartoon in pale green, ND4 in brown, natural variants of amino acids as light pink sticks and disease causing substitutions as black sticks.

## Membrane domain – summary

One striking observation from the review of all the disease causing point mutations in the MD core subunits is that they are localized mostly around the first (ND1, ND6) and the last (ND5) proton-pumping modules. Substitutions in ND2 and ND4 are much less frequent and mostly found on the periphery of the subunits in the regions not likely important for catalysis. The simplest explanation for this observation is that blocking the coupling mechanism at the level of ND5, ND4, ND2 will reduce complex I efficiency by 25, 50 and 75% respectively. However in the studies on the bacterial enzyme it has been shown that mutations in the most distal pumping module also affect the pumping efficiency of the more proximal ones (Michel, Deleon-Rangel, Zhu, Van Ree, & Vik, 2011). Therefore one can argue that the lower rate of observed mutations in ND2 and ND4 is also caused by the fact that they not only affect the proton pumping activity of a certain module but also cause disassembly or functional disconnection of the entire membrane arm translating into severe loss of complex I function. In ND1 and ND6 mutations are located around the E-channel and Q-site where charged and polar residues form large clusters and in

the case of mutation in a single residue another one may compensate for its function.

*4.5.4 Supernumerary subunits*

**13 kDa (NDUFS6) subunit**

The Cys87Tyr substitution is the only known single amino acid disease-causing mutation for this subunit. It has been previously discussed in the ovine complex I study (Fiedorczuk et al., 2016). The 13 kDa subunit is the only supernumerary complex I protein binding a metal ion, $Zn^{2+}$. Zinc-binding proteins are known to be redox active sensors (Maret, 2005) and it has been proposed that the 13 kDa subunit may play such a role in complex I (Fiedorczuk et al., 2016). Conserved Cys87 is one of the residues co-ordinating the zinc ion. It has also been reported that in *Yarrowia lipolytica* the loss of the 13 kDa subunit results in destabilization of the N5 cluster (Kmita et al., 2015). The structure revealed that the metal binding domain of the 13 kDa subunit indeed interacts with residues co-ordinating the N5 cluster. Additionally the protein also interacts with the NADPH binding 39 kDa subunit, which is also proposed to be a sensor for the redox state of the mitochondrial matrix. Mutation of Cys87 to another amino acid not only destabilizes the zinc-binding site, it also decreases electron transfer efficiency by disturbing the N5 cluster and affecting regulatory capabilities of the system (Kmita et al., 2015)(Fiedorczuk et al., 2016).
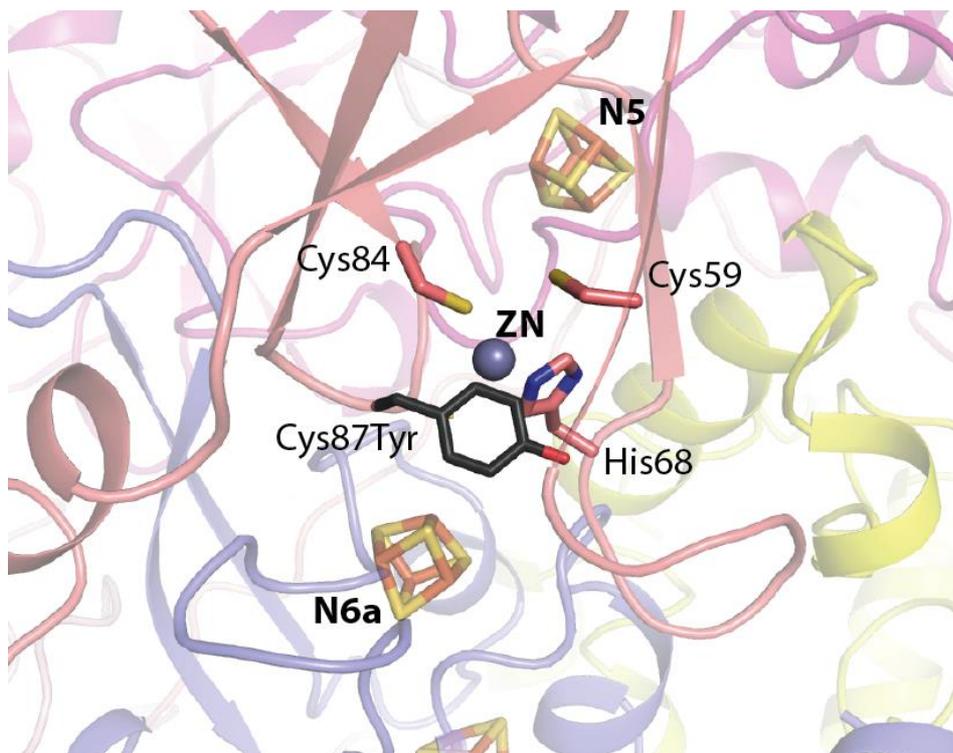
**Figure 4.39** Mutation in the 13 kDa subunit. The 13 kDa subunit pictured as cartoon in pink, 75 kDa in magenta, TYKY in blue, natural variants of amino acids as light pink sticks and disease causing substitution as black sticks.

### 18 kDa (NDUFS4) subunit

A reported disease-causing mutation in the 18 kDa is Asp77His. This substitution is located in a loop anchored between the neighbouring 30 kDa, 51 kDa, 75 kDa core subunits. The change in charge of the residue is likely to affect stability of the complex in this region or cause improper assembly.

### 10 kDa (NDUFV3) subunit

From the three mutations reported in the literature only one is in the region covered by the ovine model. Gly60Asp is located in a coil in the interface between 75 kDa, 51 kDa, 24 kDa and 18 kDa subunits and the mutation possibly destabilizes this region. The Lys22Asn mutation is in the disordered part of both the ovine and bovine models and Glu242Lys is present only in the long isoform of the 10 kDa subunit absent in our sample (Bridges et al., 2017). Mass spectrometry experiments performed on the ovine sample indicated the

presence only of the short (75 amino acids long) form of the protein (Letts, Degliesposti, et al., 2016).

## 39 kDa (NDUFA9) subunit

The reported mutation Arg286Pro has already been described in a previous ovine model study (Fiedorczuk et al., 2016). Arg286 is one of the amino acids coordinating the NADPH molecule bound in the centre of the 39 kDa subunit. Mutation to proline will cause loss of the co-factor and distortion of the whole subunit (van den Bosch et al., 2012), which results in decreased complex I activity (Abdrakhmanova et al., 2006). Very interestingly, NADPH is also coordinated by Arg178[PSST] that is located on an α-helix that extends to directly contact the N2 cluster. Thus loss of the NADPH molecule caused by the Arg286Pro mutation may influence the stability of the terminal FeS cluster. The direct connection between these two co-factors can possibly provide a mechanism of electron transfer regulation. A change in the reducing environment of the mitochondrial matrix may oxidize NADPH to $NADP^+$, which likely results in a change in the positioning of the co-factor. This distortion may influence the position or potential of the N2 cluster through Arg178-α-helix[PSST].
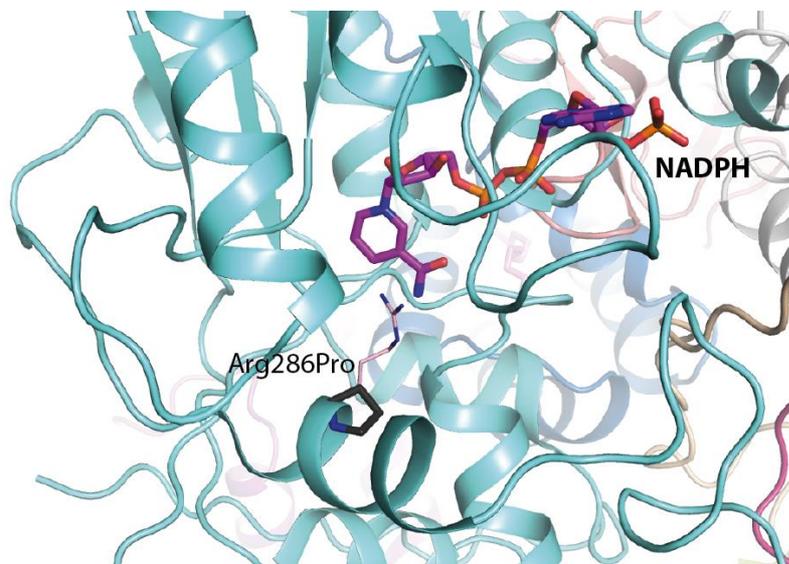


**Figure 4.40** Mutation in the 39 kDa subunit. The 39 kDa subunit pictured as cartoon in cyan, natural variants of amino acids as light pink sticks and disease causing substitution as black sticks. Co-factor is labelled.

**42 kDa (NDUFA10) and B22 (NDUFB9) subunits**

Both subunits are found on the matrix side of the membrane domain. The 42 kDa subunit directly binds the ND2 subunit, very close to the MD-PA junction and B22 is in the distal part, binding to the ND5 subunit.

The 42 kDa subunit has a nucleotide kinase fold, although to our knowledge no catalytic activity has been reported so far. It has been proposed that the subunit has an important regulatory role for complex I hinge/rotation-like arms movement (Fiedorczuk et al., 2016).

The disease causing mutation Gly64Glu is located at the interface with the ND2 and KFYI subunits, so it may decrease stability of the complex in this region. The Gln107Arg mutation is found in the central α-helix of the subunit, within the nucleotide kinase fold and most likely affects the structure.

The B22 subunit belongs to the LYR protein family and together with the interacting partner SDAP-β creates part of a bulge on the distal side of the membrane domain. The proteins interact through the phosphopantetheine moiety (PNS) covalently bound to Ser44 of SDAP-β and anchored into a hydrophobic pocket built by the three N-terminal helices of B22 (Fiedorczuk et al., 2016). Two mutations of B22 are reported to cause complex I deficiency (Table S5). Leu63Pro occurs at the last α-helix of the bundle, accommodating the PNS co-factor. Mutation to proline in the middle of the helix may affect the structure of the PNS hydrophobic pocket. If the B22 protein is not able to bind PNS, the SDAP/B22 complex likely falls apart and the whole region gets destabilized. The hetero dimer is placed directly on the terminal antiporter-like ND5 module, so that these changes may affect proton pumping. The second mutation Tyr96Cys is located in the long coil of B22 that connects the N-terminal bundle with the C-terminal part of the protein. The region around Tyr96 is interacting directly with the ND5, ND4 and SGDH subunits. Thus substitution to a very different amino acid most likely influences subunit assembly in this area.
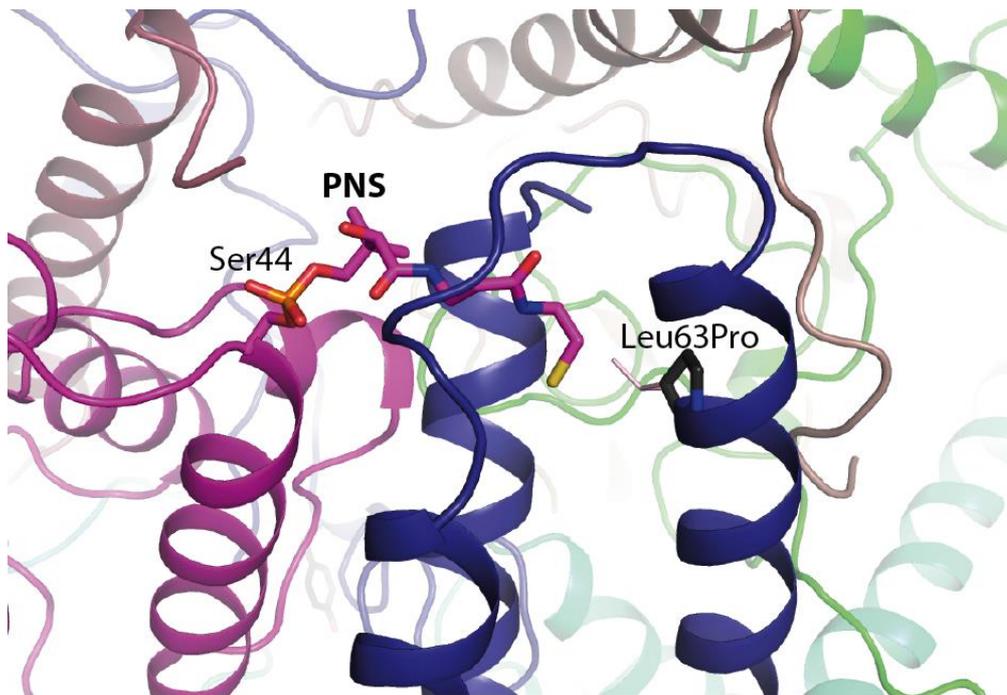
**Figure 4.41** Mutation in the B22 subunit. The B22 subunit pictured as cartoon in blue, SDAP-β in magenta, natural variants of amino acids as light pink sticks and disease causing substitution as black sticks. PNS co-factor and covalently bound Ser44$^{\text{SDAP-β}}$ are labelled.

## MWFE (NDUFA1), PGIV (NDUFA8), B16.6 (NDUFA13), B12 (NDUFB3) and 15 kDa (NDUFS5) subunits

These subunits are all located in the membrane domain of complex I on the IMS side. Presumably, they mostly play stabilizing roles, important for enzyme integrity. Disease causing mutations are mainly located on the interfaces with other complex I subunits.

Disease causing substitutions in MWFE are: Gly8Arg, at the interface site with ND1; Gly32Arg and Arg37Ser, at a loop interacting with PGIV, ND1 and B16.6; Tyr43His, at a short helix interacting with PGIV and ND6. In each of these cases, mutation results in an amino acid residue of very different chemical and structural nature.

PGIV has two single amino acid mutations reported: Glu108Lys, located at one of the α-helices building the heel of the complex (interestingly in sheep lysine is

the natural variant in this position (Letts, Degliesposti, et al., 2016)); and Arg134Gln, located at the start of a long coil connecting quinone and pumping modules.
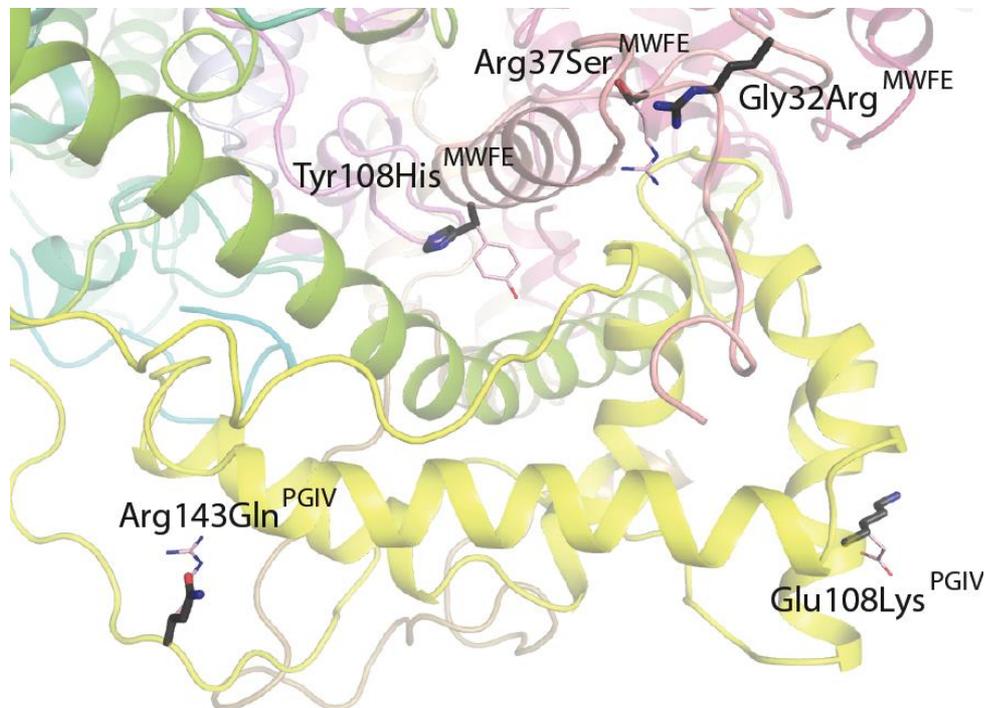


**Figure 4.42** Mutations in the PGIV and MWFE subunits. The PGIV subunit pictured as cartoon in yellow, MWFE in pink, B16.6 in green, natural variants of amino acids as light pink sticks and disease causing substitution as black sticks.

For each of the B16.6, B12 and 15 kDa subunits to my knowledge only one single point mutation has been reported in the literature so far (Table S5.). Lys4Asn[B16.6] is located at a start of the N-terminal coil, where it binds to the B14.5a subunit. Pro95Ser[15-kDa] is at the C-terminus of the subunit, which binds to and locks the C-terminus of B16.6 to the complex. Trp21Arg[B12] is located on the N-terminal coil of the subunit. The hydrophobic side chain of tryptophan is in contact with B22 and AGGG subunits. Mutation to arginine is likely to affect assembly in this area.
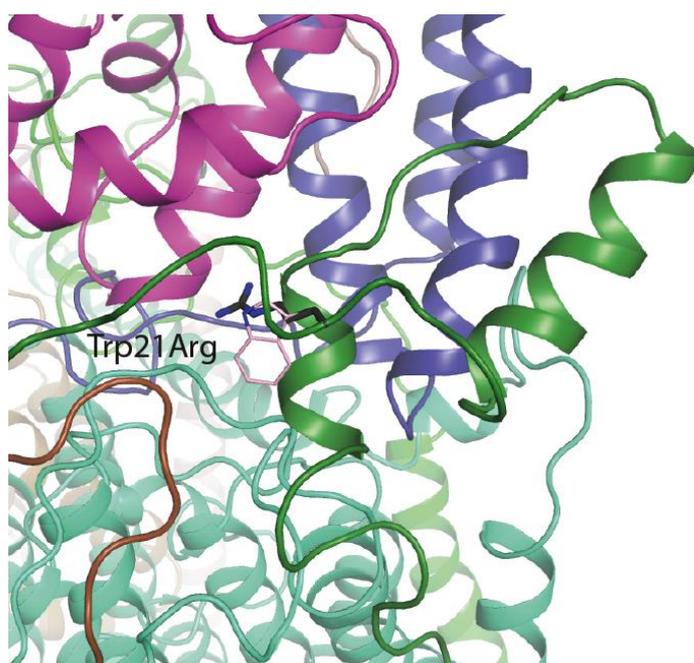
**Figure 4.43** Mutation in the B12 subunit. The B12 subunit pictured as cartoon in green, SDAP-β in magenta, B22 in blue, natural variant of amino acid as light pink sticks and disease causing substitution as black sticks.

## B14 (NDUFA6) and MNLL (NDUFB1) subunits

Mutations Ser3Asn[B14] on the matrix side and Met65Leu[MNLL] on the IMS side are located in the disordered parts of the map and hence cannot be modelled. They are not likely to play a catalytic role and it remains to be established if they have any structural role.

*4.5.5 Complex I-related diseases treatment and drug development perspectives*
Complex I deficiencies are caused by a variety of re-arrangements in the nuclear or mitochondrial genome (Tucker et al., 2011). The reasons why large deletions, insertions or non-sense mutation in complex I subunits or assembly factors cause respiratory diseases are easy to understand. They usually result in loss or destabilization of entire functional modules or complete loss of complex I. In this chapter I focused exclusively on single point missense mutations, though it is still very challenging to unambiguously link alterations in the genotype with the phenotype of the patients (Table S5).

170

There is no cure for complex I related diseases. The current best treatment is to optimize nutrition and reduce physiologic stress. Pharmacological treatment includes supplementation with CoQ, riboflavin, L-creatine, L-arginine, L-carnitine and vitamins: B1, C, E and folinic acid (reduced form of folic acid) (Parikh et al., 2009).

## 4.6 Coupling mechanism of mammalian complex I

The mechanism of coupling between electron transfer and proton translocation is still enigmatic. Conservation of key features from bacteria to mammals suggests that the basic mechanism is likely the same, with add-on "stabilisers" (disulphide - rich subunits) and "regulators" (Zn, NADPH and phosphopantetheine – containing subunits). The central axis of polar residues in the membrane probably plays a key role (Fig. 4.44). In each catalytic cycle, the negative charge stored either on Q or on nearby residues in the enclosed Q site may drive conformational changes in ND1 and the proton channel near the Q site, which would propagate via the central axis to channels in antiporter-like subunits ND2, ND4 and ND5, resulting in changes of pKa and accessibility of key residues. The net result would be the pumping of four protons per cycle, one per each channel. The observed conformation of loops in the Q site, of the presented ovine model built on the "Complete Map", likely reflects the D state. This conformation might also occur during normal function when quinol is ejected from the site into the lipid bilayer, if A/D transitions are related to conformations encountered during the catalytic cycle (Zickermann et al., 2015). Supernumerary subunits implicated in A/D transitions (39 kDa, B13 and SDAP-α/B14 pair) could also participate in catalytic conformational changes by interacting with the key TM1-TM2$^{ND3}$ loop flanking the Q site, and possibly through interactions with the 42 kDa subunit (Fig. 4.44). In the "closed" class, B13 and SDAP-α move towards 42 kDa (Chapter 4.3), hinting at such a possibility. Since the 42 kDa subunit is metazoan-specific (Ulrich Brandt, 2006), its role may be to "fine tune" movements during turnover. The traverse helix HL from ND5 appears to play mainly a stabilizing "stator" role (G. Belevich et al., 2011) rather than being a moving element (Efremov et al., 2010). Rigid disulphide-rich supernumerary subunits traversing the IMS side of the

membrane domain may represent another "stator" element unique to the mitochondrial enzyme (Chapter 4.2.2).
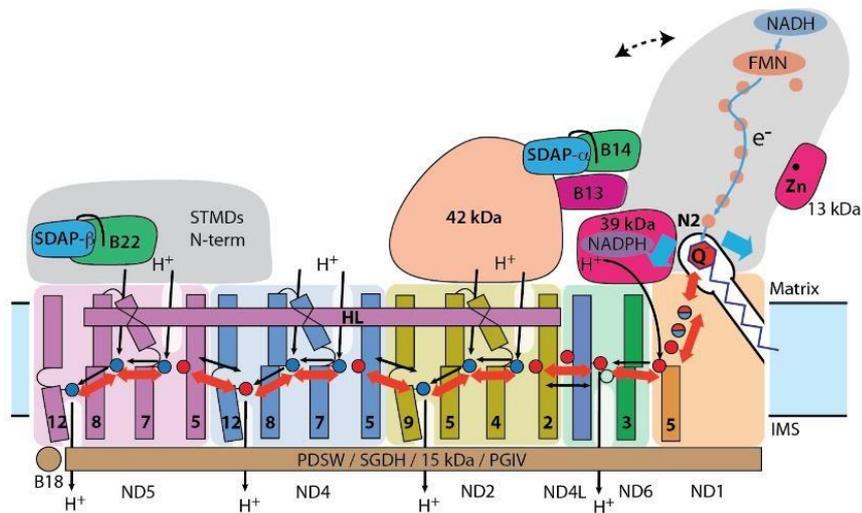


**Figure 4.44** Graphic of the coupling mechanism. Core and some putatively regulatory supernumerary subunits are shown. Conformational changes, indicated by red arrows, propagate from the Q site/E-channel to antiporter-like subunits via the central hydrophilic axis. Shifts of helices near the cluster N2 (blue arrows) may help initiate the process. ND5 helix HL and traverse helices from four supernumerary subunits on the IMS side may serve as stators. Dashed line indicates the shift of peripheral arm in the closed conformation. The NADPH-containing 39-kDa subunit and Zn-containing 13-kDa subunit are essential for activity and may serve as redox sensors. Both SDAP subunits interact with their LYR partners via flipped-out phosphopantetheine (black line). The net result of one conformational cycle, driven by NADH:ubiquinoneoxidoreduction, is the translocation of four protons across the membrane (black lines indicate possible pathways).

# 5. Summary and future work

Mitochondrial complex I is an entry point for electrons into the respiratory chain. It provides an environment for ubiquinone reduction and translocates four protons across the IMM per each molecule of NADH oxidized. Thus it plays a key role in the respiration process and its dysfunction is linked to many neurological disorders. X-ray crystallography studies on the bacterial enzyme provided great insights into the structure of the fourteen conserved core subunits (R. Baradaran et al., 2013). However, arrangement of the 30 remaining supernumerary subunits of the mammalian mitochondrial complex I was known only from previous low resolution studies (Vinothkumar et al., 2014).

This work presents the most complete and accurate reconstruction of mammalian mitochondrial complex I to date. Data demonstrate that the core of the complex is very similar to the bacterial counterpart, indicating the same basic mechanism. Supernumerary subunits most likely do not have a direct effect on the overall catalysis by the enzyme but rather stabilize and regulate its function. The 13 kDa and 39 kDa subunits are proposed to work as redox state sensors that may regulate electron flow through the peripheral arm. SDAP-α/B14 and SDAP-β/B22 paired via a phosphopantetheine moiety provide a possible linker between fatty acid production and OXPHOS respiration. Several supernumerary subunits may have evolved to provide contacts with other complexes within the respirasome. Disulphide-containing intertwined subunits lining the IMS side of the complex provide a rigid outer shell for the enzyme in the oxidizing environment of the IMS.

Five different conformations of the enzyme observed in this study potentially shed a light on different catalytic states of the enzyme. Overall they differ by the angle of the PA over the MD which seems to correlate with the accessibility of

the ubiquinone-binding site. This indicates that they may reflect the active and de-active states of the enzyme, or conformational states during the catalytic cycle. As proposed in Chapter 4.3.2, conformational changes of the crucial loops in the 49 kDa, ND1 and ND3 subunits in the mammalian enzyme are probably initiated by the binding of ubiquinone. Further large-scale conformational changes may require interactions of the 42 kDa, B13 and SDAP-α/B14 subunits. The models also give some hints about conformational changes in the proton pumping modules, specifically the ND5 subunit, during catalytic turnover.

The structure allows for investigating the molecular basis of mutations causing complex I related diseases (Chapter 4.4). The atomic model of mitochondrial complex I can help to develop new drugs for a better treatment of patients suffering from complex I deficiency (Anderson, 2003). The structure can provide a starting point for structure based, in silico design and in vitro testing of new types of molecules that may mitigate severe mitochondrial disorders. Currently treatment of complex I deficiencies involves mainly nutrition optimization (Parikh et al., 2009). With an atomic model of the complex it is possible to design synthetic compounds selectively modulating ROS production at the FMN site, electron transfer to ubiquinone or proton translocation. That makes complex I an important target for next generation anti-cancer drugs or antibiotics.

One of the most important questions in the field of complex I biology, how electron transfer is coupled to proton translocation, still remains unanswered (Chapter 1.4.3). It is generally accepted that coupling is triggered by Q reduction and requires long-distance conformational changes along the membrane domain but the exact mechanism remains elusive. To solve this puzzle, high-resolution structures of complex I in the presence of NADH and ubiquinone are required. Cryo-EM methods of focused refinements and subsequent classification presented in this dissertation may easily be adapted for investigation of complex I with bound NADH and ubiquinone-10 analogues.

This work gives hints how future structural work performed on isolated complex I can be improved. Most importantly, solubilization reagent should be chosen

carefully. Brij-35 proved to be very useful in production of 'good contrast' grids with even particles distribution, unfortunately it acts as a complex I inhibitor and likely was the reason for the disorder of the B14.7 subunit in the presented cryo-EM maps. In the described experiments a number of detergents have been tested. One that performed well in cryo-EM screens was Cymal-7. The utility of Cymal-7 has also been shown in bovine complex I studies (Vinothkumar et al., 2014)(J. Zhu et al., 2016). It was also presented that a good alternative to detergents are some amphipathic polymers (Flotenmeyer et al., 2007). The commonly used Amphipol A8-35 with low concentration (0.02%) additives like CHAPS is a very promising solubilization reagent that can be used in complex I structural studies (Chapter 3.3). However, specific activity of the complex in amphipols yet has to be tested.

Another approach would be studying complex I structure in more native environment, mimicking the mitochondrial lipid bilayer. That can be achieved by reconstituting purified complex I in nanodiscs or liposomes of desired lipid composition. However, using either of these systems one can face certain difficulties. Mitochondrial complex I has a very large MD which is at the upper limit of currently available membrane scaffold proteins used for nanodiscs formation (Hagn et al., 2013). Trapping such a big molecule can be challenging and there will be not much space left in the nanodisc for lipids molecules. Liposomes, on the other hand, can force the protein to adapt artificial ternary structure in the MD region due to the curvature of vesicles. This method will also introduce complications during data collection (low number of particles per image) and processing (challenging CTF correction due to the unknown Z-axis location of protein in thick ice). Additionally, the choice of lipid composition mimicking bilayers of the IMM around complex I in native mitochondria is very challenging. Those problems can be solved by using styrene-maleic acid (SMA) for membrane solubilization (Scheidelaar et al., 2015). It has the ability to solubilize membranes in the form of 'nanodisks' allowing extraction and purification of membrane proteins from their native environment in a single, detergent-free step. However, using SMA will require developing a novel purification protocol.

Once a better method for solubilization of the protein in an active state is developed, the sample can be incubated with NADH and ubiquinone and frozen to capture different conformational states of the enzyme. The idea would be to allow complex I to achieve full oxidoreductase activity with these substrates and then snap-freeze it for cryo-EM, so that different conformations can then be computationally sorted. Comparison of complex I states revealed by this study and future structures of the catalytically active enzyme are likely to give us a detailed catalytic mechanism of the enzyme's action. Performing a similar experiment with addition of different inhibitors can add to our knowledge about the complex I coupling mechanism by possibly trapping the molecule in certain catalytic states. The biggest group of complex I inhibitors are quinone binding site competitors (Esposti, 1998). Other inhibitors, presumably binding close to the proton translocation sites, like metformin (Viollet, 2014) or bivalent ions, especially $Zn^{2+}$ (Sharpley, 2006) may help to indicate precisely proton translocation paths. Although a number of these inhibitors has been known for years, still there is no mammalian complex I structure with inhibitor bound. For that reason, structural data about complex I with any of the inhibitors bound would be of great interest of the field.

Another powerful approach is a combination of structural biology with mutagenesis studies. Nuclear encoded mammalian complex I proteins can be mutated in order to investigate the function of presumably crucial amino acids at FMN or ubiquinone binding sites. Mutagenesis experiments targeting the mitochondria-encoded part of the enzyme will probably be more challenging and a better model here would be a bacterial homologue. Since, very likely, coupling and proton pumping mechanisms are in principle identical in bacteria and eukaryotes, the prokaryotic model would be easier to use for studies of mutations in the TM core part of the protein.

Complex I in mitochondria is physiologically found mostly in the form of supercomplexes (Schagger & Pfeiffer, 2001). This fact suggests that individual respiratory complexes should be studied not only in isolation but also together as a respirasome. Recent studies have presented moderate resolution structures of $I_1III_2IV_1$supercomplex (Letts, Fiedorczuk, et al., 2016)(Wu et al.,

2016) in detergent. To fully understand the OXPHOS mechanism, high-resolution models of all IMM complexes and supercomplexesare required. Thanks to the recent developments, single particle cryo-EM allows achievement of resolutions high enough to obtain maps interpretable on the atomic scale. Also here, the use of respiratory complex inhibitors may be very insightful. Blocking one of the enzymes in the supercomplex after it achieves full oxidoreductase activity can help to understand processes in other enzymes and their interactions. For example, inhibition of cytochrome *c* oxidase with KCN may help to capture complex I, functionally located upstream, in fully reduced conformation.

Single particle studies of supercomplexes just like isolated complex I require removing protein form the native lipid environment, which can cause irreversible damage. Again, it would be desirable to have a possibility to investigate proteins' structures in the native lipid bilayer. However, because of their substantial size, supercomplexes are even more difficult to reconstitute in lipid membrane substitute systems. Bypassing this obstacle may be possible with cryo-electron tomography, which currently does not offer high enough resolution, but hopefully in a few years will develop to permit scientists to look at proteins in the lipid membrane at high enough resolution.

Currently cryo-EM is developing very rapidly. Every month brings new improvements and tools enabling higher resolution structures in a faster and easier way. New types of data collection grids (Russo, 2015), improved processing software (Kimanius, 2016), detectors (http://www.gatan.com/k3-camera) or phase plates (Danev, 2017) are only examples of current developments in the field. With the new EM hardware and software combined with biochemical methods based on mutagenesis and use of specific inhibitors, understanding of complex I coupling mechanism and physiological relevance of mitochondrial supercomplexes seems to be just a matter of time.

In summary, the presented study provides insights into the structure and function of all mitochondrial complex I subunits and shows global changes in the enzyme's structure that may be assigned to different catalytic states of the

molecule. It allows for better understanding of function of the supernumerary subunits and mutations leading to complex I deficiency. It enlarges the scope of our knowledge regarding one of the most basic metabolic processes – cellular respiration. However, more structures at higher resolution and of different catalytic states are required. They will improve not only our understanding of the very basic processes in our cells – ATP production – but in time will hopefully affect the life and well-being of patients suffering from mitochondrial illnesses.

# Supplement

| Subunits: Human Nomenclature | Bovine | Ovine | Porcine | Subunits: Common 'Bovine' Nomenclature | Ovine | Porcine |
|---|---|---|---|---|---|---|
| **Core** | | | | | | |
| NDUFV1 | 97.5 | 98.0 | 97.5 | 51 kDa | 99.5 | 98.7 |
| NDUFV2 | 98.6 | 98.2 | 97.2 | 24 kDa | 99.5 | 97.7 |
| NDUFS1 | 97.4 | 97.3 | 97.2 | 75 kDa | 99.3 | 96.7 |
| NDUFS2 | 95.4 | 95.6 | 95.8 | 49 kDa | 99.8 | 98.6 |
| NDUFS3 | 94.7 | 95.2 | 93.9 | 30 kDa | 97.4 | 95.2 |
| NDUFS7 | 92.7 | 92.7 | 92.2 | PSST | 100 | 96.1 |
| NDUFS8 | 96.0 | 96.0[1] | 96.0 | TYKY | 100[1] | 98.3 |
| **Average Nuclear:** | **96.0** | **96.1** | **95.7** | | **99.4** | **97.3** |
| ND1 | 78.0 | 78.0 | 77.0 | ND1 | 93.7 | 89.6 |
| ND2 | 63.1 | 63.1 | 62.0 | ND2 | 91.9 | 75.8 |
| ND3 | 73.9 | 73.0 | 71.3 | ND3 | 94.8 | 90.4 |
| ND4 | 74.1 | 75.2 | 73.2 | ND4 | 92.8 | 86.5 |
| ND4L | 73.5 | 76.5 | 76.5 | ND4L | 93.9 | 86.7 |
| ND5 | 69.5 | 69.6 | 69.5 | ND5 | 90.6 | 81.7 |
| ND6 | 62.3 | 61.1 | 57.7 | ND6 | 93.7 | 88.6 |
| **Average Mito:** | **70.6** | **70.9** | **69.6** | | **93.1** | **85.6** |
| **Supernumerary** | | | | | | |
| NDUFS4 | 95.5 | 95.5 | 95.5 | 18 kDa | 98.5 | 95.5 |
| NDUFS5 | 74.3 | 70.5 | 79.0 | 15 kDa | 90.5 | 88.6 |
| NDUFS6 | 89.6 | 90.6[1] | 88.5 | 13 kDa | 97.9[1] | 93.8 |
| NDUFV3 | 86.7 | 88.0[1] | 29.3[2] | 10 kDa | 98.7[1] | 33.3[2] |
| NDUFA1 | 80.0 | 80.0 | 80.0 | MWFE | 95.7 | 81.4 |
| NDUFA2 | 93.9 | 93.9 | 94.9 | B8 | 100 | 94.9 |
| NDUFA3 | 83.1 | 83.1 | 78.3 | B9 | 98.8 | 85.5 |
| NDUFA5 | 87.0 | 87.0 | 87.8 | B13 | 98.3 | 94.8 |
| NDUFA6 | 90.6 | 90.6 | 89.0 | B14 | 98.4 | 92.9 |
| NDUFA7 | 88.4 | 87.5 | 90.2 | B14.5a | 96.4 | 92.0 |
| NDUFA8 | 87.7 | 87.1 | 90.6 | PGIV | 98.2 | 95.9 |
| NDUFA9 | 79.7 | 80.9 | 83.3 | 39 kDa | 96.2 | 87.8 |
| NDUFA10 | 80.6 | 79.7 | 79.7 | 42 kDa | 94.7 | 87.5 |
| NDUFA11 | 72.1 | 70.0 | 69.3 | B14.7 | 95.7 | 90.0 |
| NDUFA12 | 89.7 | 89.7 | 89.0 | B17.2 | 100 | 95.2 |
| NDUFA13 | 83.2 | 83.2 | 86.7 | B16.6 | 95.8 | 93.7 |
| NDUFAB1 | 97.7 | 97.7 | 96.6 | SDAP | 100 | 98.9 |
| NDUFB1 | 82.5 | 77.2 | 82.5 | MNLL | 94.7 | 84.2 |
| NDUFB2 | 90.3 | 90.3 | 88.9 | AGGG | 100 | 95.8 |
| NDUFB3 | 82.5 | 83.5 | 84.5 | B12 | 99.0 | 95.9 |
| NDUFB4 | 73.4 | 75.0 | 78.1 | B15 | 93.8 | 87.5 |
| NDUFB5 | 85.3 | 85.5 | 100 | SGDH | 95.8 | 85.3 |
| NDUFB6 | 78.0 | 78.0 | 71.7 | B17 | 97.6 | 86.6 |
| NDUFB7 | 86.8 | 85.3 | 85.3 | B18 | 98.5 | 93.4 |
| NDUFB8 | 85.4 | 86.1 | 85.4 | ASHI | 98.7 | 93.0 |
| NDUFB9 | 91.0 | 89.3 | 91.1 | B22 | 96.6 | 93.3 |
| NDUFB10 | 78.5 | 78.0 | 77.4 | PDSW | 97.7 | 93.1 |
| NDUFB11 | 85.6 | 85.6 | 83.2 | ESSS | 98.4 | 90.4 |
| NDUFC1 | 81.6 | 79.6 | 83.7 | KFYI | 98.0 | 93.9 |
| NDUFC2 | 73.3 | 75.0 | 73.3 | B14.5b | 93.3 | 85.0 |
| **Average:** | **84.5** | **83.9** | **83.1** | | **96.3** | **89.5** |

**Table S1.** Comparison of sequences of core and accessory subunits between three easily available organisms: *Sus scrofa, Bost taurus, Ovis aries* and human enzyme.Sequence identity of mammalian complexes I relative to human (left) and bovine (right). After Letts et al (Letts, Degliesposti, et al., 2016).

| Subunit | | Mature subunit size[1] | | Identification | | |
|---------|---------|------|------|---------|---------|---------|
| Bovine name | Human name | #aa | MW (kDa) | Initial preparation | Improved preparation | Sequence Database Accession Number |
| 51 kDa | NDUFV1 | 445 | 48.6 | YES | YES | Uniprot # W5PUX0 |
| 24 kDa | NDUFV2 | 217 | 23.8 | YES | YES | Uniprot # W5NRY1 |
| 75 kDa | NDUFS1 | 704 | 76.9 | YES | YES | Uniprot # W5QB34 |
| 49 kDa | NDUFS2 | 430 | 49.1 | YES | YES | Uniprot # W5PJ73 |
| 30 kDa | NDUFS3 | 228 | 26.4 | YES | YES | Uniprot # W5PB27 |
| PSST | NDUFS7 | 179 | 20.1 | YES | YES | Uniprot # W5PPP6 |
| TYKY | NDUFS8 | 176 | 20.2 | YES | YES | NCB Accession # XP_011972879.1 |
| ND1 | ND1 | 318 | 35.9 | YES | YES | Uniprot # O78747 |
| ND2 | ND2 | 347 | 39.1 | YES | YES | Uniprot # O78748 |
| ND3 | ND3 | 115 | 13.1 | NO | YES | Uniprot # O78753 |
| ND4 | ND4 | 459 | 52 | YES | YES | Uniprot # O78755 |
| ND4L | ND4L | 98 | 10.8 | NO | NO | Uniprot # O78754[2] |
| ND5 | ND5 | 606 | 68.4 | YES | YES | Uniprot # O78756 |
| ND6 | ND6 | 175 | 19.1 | NO | YES | Uniprot # O78757 |
| | | | | | | |
| 18 kDa | NDUFS4 | 133 | 15.3 | YES | YES | Uniprot # W5PE07 |
| 15 kDa | NDUFS5 | 105 | 12.4 | YES | YES | Uniprot # W5QFF9 |
| 13 kDa | NDUFS6 | 96 | 10.6 | YES | YES | NCB Accession # XP_011980592.1 |
| 10 kDa | NDUFV3 | 75 | 8.4 | NO | YES | NCB Accession # XP_011991231.1 |
| MWFE | NDUFA1 | 70 | 8.2 | NO | YES | NCB Accession # NP_001305903.1 |
| B8 | NDUFA2 | 98 | 10.9 | YES | YES | Uniprot # W5QAH8 |
| B9 | NDUFA3 | 83 | 9.2 | YES | YES | Uniprot # W5NYM7 |
| B13 | NDUFA5 | 115 | 13.1 | YES | YES | Uniprot # W5PNX7 |
| B14 | NDUFA6 | 127 | 14.9 | YES | YES | Uniprot # W5QC06 |
| B14.5a | NDUFA7 | 112 | 12.4 | YES | YES | NCB Accession # XP_004008614.1 |
| PGIV | NDUFA8 | 171 | 20 | YES | YES | Uniprot # W5PYA5 |
| 39 kDa | NDUFA9 | 344 | 39 | YES | YES | Uniprot # W5PI58 |
| 42 kDa | NDUFA10 | 320 | 36.8 | YES | YES | Uniprot # W5QBF5 |
| B14.7 | NDUFA11 | 140 | 14.6 | YES | YES | Uniprot # W5PAR2 |
| B17.2 | NDUFA12 | 145 | 17.1 | YES | YES | Uniprot # B9VGZ9 |
| B16.6 | NDUFA13 | 143 | 16.6 | YES | YES | NCB Accession # XP_004008450.1 |
| SDAP | NDUFAB1 | 88 | 10.1 | YES | YES | Uniprot # W5NQT7 |
| MNLL | NDUFB1 | 57 | 6.9 | YES | YES | NCB Accession # XP_004018002.1 |
| AGGG | NDUFB2 | 72 | 8.5 | YES | YES | Uniprot # W5PVD7 |
| B12 | NDUFB3 | 97 | 11 | YES | YES | Uniprot # W5Q5T4 |
| B15 | NDUFB4 | 128 | 15 | NO | YES | NCB Accession XP_004003003.1 |
| SGDH | NDUFB5 | 143 | 16.7 | YES | YES | Uniprot # W5QHN8 |
| B17 | NDUFB6 | 127 | 15.4 | YES | YES | Uniprot # W5PZE3 |
| B18 | NDUFB7 | 136 | 16.3 | YES | YES | Uniprot # W5P5V3 |
| ASHI | NDUFB8 | 158 | 18.8 | YES | YES | Uniprot # W5Q1B0 |
| B22 | NDUFB9 | 178 | 21.6 | YES | YES | Uniprot # W5PGA3 |
| PDSW | NDUFB10 | 175 | 20.8 | YES | YES | NCB Accession # XP_011999786.1 |
| ESSS | NDUFB11 | 125 | 14.4 | YES | YES | Uniprot # W5PWF1 |
| KFYI | NDUFC1 | 49 | 5.8 | NO | YES | NCB Accession # XP_004017292.1 |
| B14.5b | NDUFC2 | 120 | 14.2 | NO | YES | NCB Accession # XP_004019479.1 |

**Table S2.** Mass spectrometry identification of complex I subunits from the initial and improved preparations. Note that the absence of many subunits in the initial sample is probably due to suboptimal performance of the experiment. Sites of mitochondrial targeting sequence removal were determined experimentally by peptide identification and/or by sequence alignment with the bovine subunits.
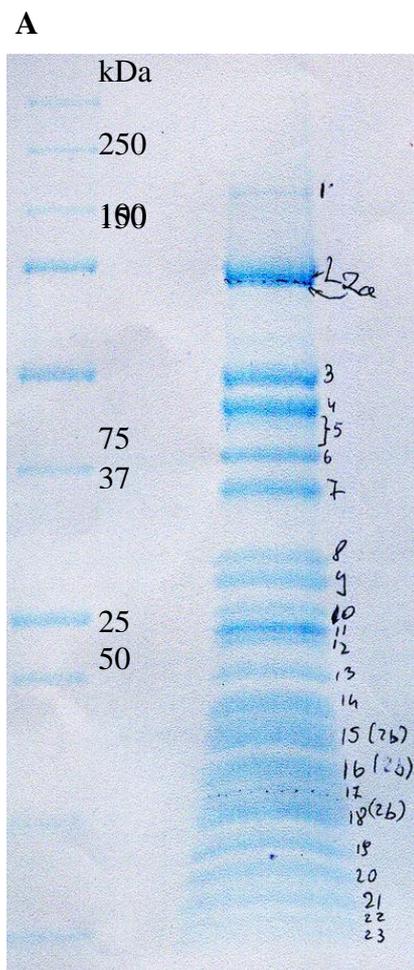
**B**

| Band # | Proteins identified in the band |
|--------|--------------------------------|
| 1 | 2-oxoglutarate dehydrogenase |
| 2 | 75 kDa subunit |
| 3 | 51 kDa subunit |
| 4 | 59 kDa subunit |
| 5 | ND5 |
| 6 | 42 kDa subunit |
| 7 | 39 kDa subunit |
| | ND4 |
| 8 | 30 kDa subunit |
| | ND2 |
| 9 | 30 kDa subunit |
| | ND1 |
| 10 | 24 kDa subunit |
| | Cytochrome c oxidase subunit 2 |
| 11 | PDSW |
| | B22 |
| 12 | TYKY |
| | PDSW |
| 13 | PSST |
| 14 | PGIV |
| | ASHI |
| 15 | 18 kDa subunit |
| | ESSS |
| | ASHI |
| | B17 |
| | B18 |
| 16 | B 17.2 |
| | B16.6 |
| 17 | sample failed to produce hits |
| 18 | B14 |
| | B14.5a |
| 19 | 13 kDa |
| | B12 |
| | B13 |
| 20 | B8 |
| | AGGG |
| | Cytochrome c oxidase subunit 6A2 |
| 21 | SDAP |
| | B9 |
| 22 | Sample failed to produce hits |
| 23 | 18 kDa |
| | MNLL |
| | Cytochrome c oxidase subunit 7A1 |

**Figure S1A.** SDS-PAGE of complex I from the initial DDM purification. Numbers represent bands that were cut out from the gel and analysed by tryptic peptide mass fingerprinting. Numbers (2a) and (2b) are side notes that should be ignored. **B.** Complex I subunits and contaminants identified by tryptic peptide mass fingerprinting. Numbers correspond to bands labelled on the gel Fig. S1A. In red protein contaminates.

182
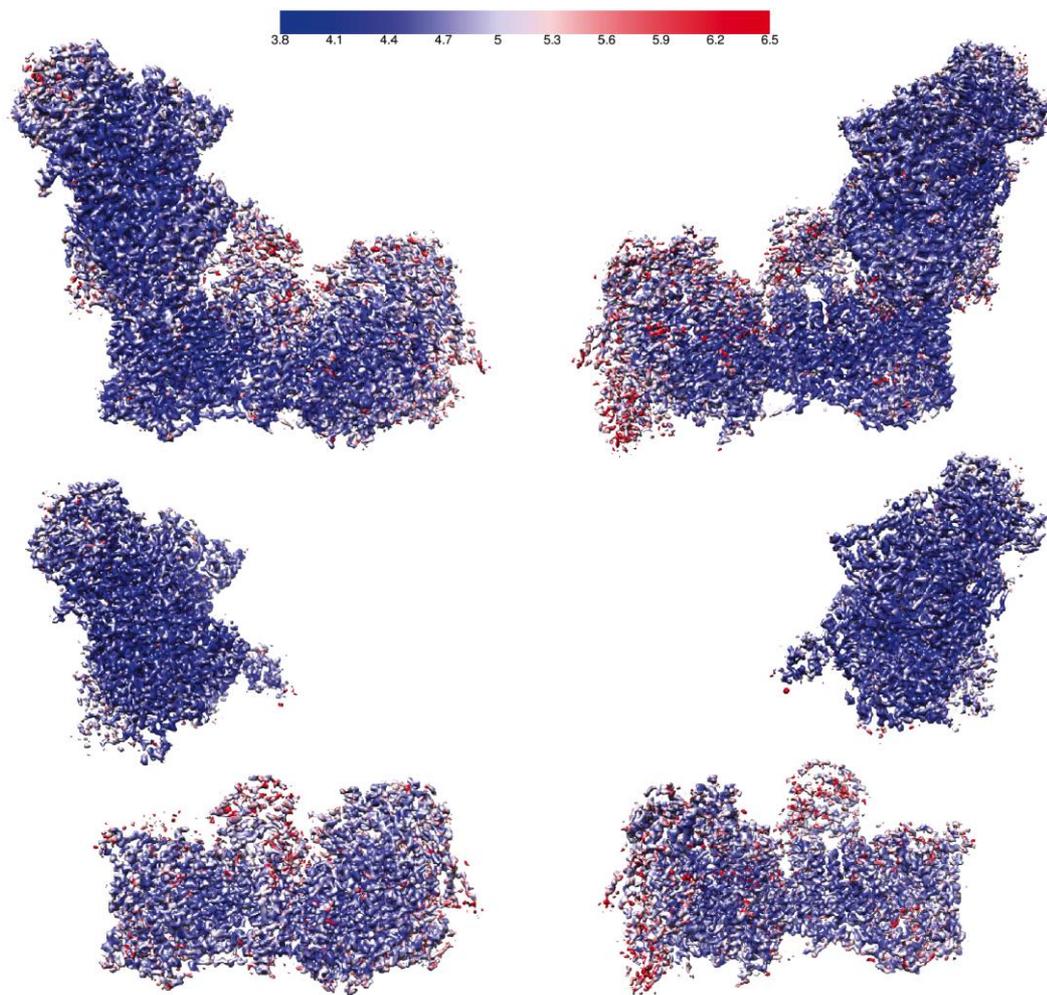
**Figure S2** Local resolution estimation by Resmap of A. Entire complex I B. Peripheral arm focused refinement map C. Membrane domain focused refinement map. Maps are coloured according to the shown resolution scale in Å.

| | Subunit Name Bovine / Human | Chain | Total residues / range built | Poly-ALA model | Un-modelled residues | % atomic model | TMHs | Cofactors / Notes |
|---|---|---|---|---|---|---|---|---|
| Core | 51 kDa / NDUFV1 | 1 | 445 / 7-438 | - | 1-6, 439-445 | 97.1 | - | FMN N3 (4Fe[51]) |
| | 24 kDa / NDUFV2 | 2 | 217 / 3-216 | - | 1-2, 217 | 98.6 | - | N1a (2Fe[24]) |
| | 75 kDa / NDUFS1 | 3 | 704 / 6-693 | - | 1-5, 694-704 | 97.7 | - | N1b (2Fe[75]) N4 (4Fe[75]C) N5 (4Fe[75]H) |
| | 49 kDa / NDUFS2 | 4 | 430 / 44-430 | - | 1-43 | 90.0 | - | - |
| | 30 kDa / NDUFS3 | 5 | 228 / 7-214 | - | 1-6, 215-228 | 91.2 | - | - |
| | PSST / NDUFS7 | 6 | 179 / 25-179 | - | 1-24 | 86.6 | - | N2 (4Fe[PS]) |
| | TYKY / NDUFS8 | 9 | 176 / 1-176 | - | - | 100 | - | N6a (4Fe[TY]1) N6b (4Fe[TY[2) |
| | ND1 | H | 318 / 1-318 | - | - | 100 | 8 | |
| | ND2 | N | 347 / 1-347 | - | - | 100 | 11 | (antiporter-like) |
| | ND3 | A | 115 / 1-115 | - | - | 100 | 3 | |
| | ND4 | M | 459 / 1-459 | - | - | 100 | 14 | (antiporter-like) |
| | ND4L | K | 98 / 1-86 | - | 87-98 | 87.8 | 3 | |
| | ND5 | L | 606 / 1-568 | 509-598 | 599-606 | 84.0 | 16 | (antiporter-like) |
| | ND6 | J | 175 / 1-108, 123-175 | 77-122 | - | 73.7 | 5 | - |
| Supernumerary | 10 kDa / NDUFV3 | a | 75 / 34-74 | - | 1-33, 75 | 54.7 | - | - |
| | 13 kDa / NDUFS6 | b | 96 / 1-95 | - | 96 | 99.0 | - | $Zn^{2+}$ |
| | 18 kDa / NDUFS4 | c | 133 / 11-133 | - | 1-10 | 92.5 | - | - |
| | 39 kDa / NDUFA9 | d | 345 / 1-252, 277-338 | 325-338 | 253-276, 339-345 | 87.0 | - | NADPH |
| | B8 / NDUFA2 | e | 98 / 13-96 | - | 1-12, 97-98 | 85.7 | - | Thioredoxin fold |
| | B13 / NDUFA5 | f | 115 / 4-115 | - | 1-3 | 97.4 | - | - |
| | B14 / NDUFA6 | g | 127 / 14-127 | - | 1-13 | 89.8 | - | LYR protein |
| | B14.5a / NDUFA7 | h | 112 / 1-71, 89-112 | - | 72-88 | 84.8 | - | - |
| | B17.2 / NDUFA12 | i | 145 / 1-144 | - | 145 | 99.3 | - | - |
| | SDAP-□ / NDUFAB1 | j | 88/ 3-87 | - | 1-2, 88 | 96.6 | - | Acyl carrier protein phosphopantetheine |
| | 42 kDa / NDUFA10 | k | 320 / 1-320 | 227-320 | - | 70.9 | - | Nucleoside kinase family |
| | 15 kDa / NDUFS5 | l | 105 / 1-95 | - | 96-105 | 90.5 | - | Quadruple $CX_9C$ 2 CHCH domains |
| | B9 / NDUFA3 | m | 83 / 4-83 | - | 1-3 | 96.4 | 1 | STMD |
| | B12 / NDUFB3 | n | 97 / 13-86 | 83-86 | 1-12, 87-97 | 73.2 | 1 | STMD |
| | B14.5b / NDUFC2 | o | 120 / 1-120 | - | - | 100 | 2 | - |
| | B15 / NDUFB4 | p | 128 / 17-73, 95-128 | 17-27, 74-94 | 1-16 | 62.5 | 1 | STMD |
| | B16.6 / NDUFA13 | q | 143 / 4-143 | - | 1-3 | 97.9 | 1 | Identical to GRIM-19, STMD |
| | B17 / NDUFB6 | r | 127 / 1-37, 63-118 | 63-71 | 38-62, 119-127 | 66.1 | 1 | STMD |
| | B18 / NDUFB7 | s | 136 / 5-122 | 5-15 | 1-4, 123-136 | 78.7 | - | Double $Cx_9C$ CHCH domain |
| | B22 / NDUFB9 | t | 178 / 9-174 | - | 1-8, 175-178 | 93.3 | - | LYR protein |
| | AGGG / NDUFB2 | u | 72 / 3-68 | - | 1-2, 69-72 | 91.7 | 1 | STMD |
| | ASHI / NDUFB8 | v | 158 / 1-84, 101-143 | 1-100 | 144-158 | 27.2 | 1 | STMD |
| | ESSS / NDUFB11 | w | 125 / 38-123 | - | 1-37, 124-125 | 68.8 | 1 | STMD |
| | KFYI / NDUFC1 | x | 49 / 1-48 | - | 49 | 98.0 | 1 | STMD |
| | MNLL / NDUFB1 | y | 57 / 4-56 | - | 1-3, 57 | 93.0 | 1 | STMD |
| | MWFE / NDUFA1 | z | 70 / 2-70 | - | 1 | 98.6 | 1 | STMD |
| | PDSW / NDUFB10 | Z | 175 / 3-173 | - | 1-2, 174-175 | 97.7 | - | - |
| | PGIV / NDUFA8 | Y | 171 / 1-171 | - | - | 100 | - | Quadruple $CX_9C$ 2 CHCH domains |
| | SDAP-□ / NDUFAB1 | X | 88 / 1-88 | - | - | 100 | - | Acyl carrier protein phosphopantetheine |
| | SGDH / NDUFB5 | W | 143 / 5-143 | - | 1-4 | 97.2 | 1 | STMD |
| | B14.7 / NDUFA11 | V | 140 / 7-125 | 7-125 | 21 | 0 | 4 | - |
| | Total | | 8516 / 8037 | 518 | 479 | 88.3 | 78 | |

**Table S3.** Summary of the model. Figure prepared by Dr James Letts.

184

| Subunit | Surface area, Å² | Buried area, Å² | ΔG$^{int}$ kcal/mol | N$_{HB}$ | N$_{SB}$ | Interacting subunits (Descending buried area order, core subunits in bold) |
|---|---|---|---|---|---|---|
| | | | Entire complex summary | | | |
| | 297384.9 | 259150.4 | -1838.5 | | | |
| | | | Core subunits | | | |
| **51 kDa** | 19017.7 | 6988.3 | -12.38 | 77 | 18 | **24 kDa,75 kDa**, 10 kDa, 18 kDa, B14.5a |
| **24 kDa** | 14271.6 | 5144.8 | -34.27 | 43 | 15 | **51 kDa**, 10 kDa, **75 kDa**, 18 kDa, 13 kDa |
| **75 kDa** | 29733.9 | 10317.9 | -23.97 | 102 | 43 | 18 kDa, **51 kDa**, B8, **49 kDa**, **24 kDa,30 kDa**, B17.2, B14, **TYKY**, 13 kDa, B14.5a, 39 kDa, 10 kDa |
| **49 kDa** | 17459.5 | 11703.0 | -39.24 | 93 | 37 | **30 kDa, TYKY, PSST**, B14.5a, **ND1**, B16.6, **75 kDa**, B13, **ND3**, 13 kDa, B14 |
| **30 kDa** | 14832.7 | 9972.2 | -40.37 | 79 | 34 | **49 kDa**, B13, 18 kDa, B14.5a B14 **PSST**, **75 kDa**, 39 kDa, **TYKY** |
| **PSST** | 9483.5 | 6258.3 | -26.86 | 47 | 31 | **49 kDa, ND1, TYKY, 30 kDa**, 39 kDa, **ND3**, B17.2, B14 |
| **TYKY** | 13499.4 | 10446.6 | -47.00 | 87 | 33 | **49 kDa**, B17.2, **PSST**, B14.5a, 13 kDa, **ND1**, **75 kDa**, B16.6, B9, **30 kDa**, 18 kDa, 39 kDa, MWFE |
| **ND1** | 16976.8 | 11161.2 | -98.77 | 61 | 21 | **ND3**, MWFE, **PSST, 49 kDa, ND6**, B16.6, **TYKY**, B9, PGIV, B17.2 |
| **ND2** | 16531.5 | 9336.0 | -106.34 | 19 | 1 | **ND4L**, 42 kDa, **ND4**, B14.5b, 15 kDa, **ND5**, SGDH |
| **ND3** | 10944.8 | 6994.6 | -81.41 | 15 | 0 | **ND1, ND6, ND4L, PSST**, 49 kDa, B9, 39 kDa, B14, B16.6 |
| **ND4** | 18758.1 | 11100.2 | -98.26 | 42 | 3 | **ND5**, ESSS, **ND2**, B15, SGDH, MNLL, PDSW, B22, PGIV, B14.5b, ASHI |
| **ND4L** | 6435.4 | 4638.4 | -64.53 | 7 | 2 | **ND6, ND2**, 15 kDa, **ND3** |
| **ND5** | 26832.2 | 10837.8 | -94.45 | 47 | 5 | **ND4**, AGGG, B22, PDSW, ASHI, B17, B18, B12, B15, **ND2**, SDAP-☐☐, SGDH, B14.7 |
| **ND6** | 12734.4 | 8272.0 | -95.71 | 19 | 2 | **ND4L, ND3, ND1**, 15 kDa, B16.6, MWFE |
| | | | Supernumerary subunits | | | |
| **10 kDa** | 4892.2 | 2471.0 | -15.11 | 23 | 9 | **51 kDa, 24 kDa**, 18 kDa, **75 kDa** |
| **13 kDa** | 7651.1 | 2991.2 | -10.11 | 16 | 6 | **TYKY**, B17.2, 39 kDa, **75 kDa, 49 kDa, 24 kDa** |
| **18 kDa** | 11083.8 | 5848.7 | -11.52 | 57 | 20 | **75 kDa, 30 kDa**, B14, 51 kDa, 39 kDa, 10 kDa, **TYKY, 24 kDa**, B17.2 |
| **39 kDa** | 16659.0 | 3814.3 | -5.70 | 20 | 11 | **PSST**, 13 kDa, **30 kDa**, B14, 18 kDa, **ND3, 75 kDa, TYKY** |
| **B8** | 5628.2 | 1196.4 | -2.44 | 13 | 10 | **75 kDa** |
| **B13** | 8653.4 | 2370.0 | -15.60 | 14 | 5 | **30 kDa, 49 kDa**, B14.5a |
| **B14** | 9627.2 | 4586.9 | -9.29 | 31 | 19 | SDAP-☐☐☐**30 kDa**, 18 kDa, **75 kDa**, 39 kDa, **ND3, 49 kDa, PSST** |
| **B14.5a** | 11473.6 | 5753.0 | -40.33 | 49 | 17 | **49 kDa, 30 kDa, TYKY, 75 kDa**, B17.2, B13, B16.6, **51 kDa** |
| **B17.2** | 12313.0 | 4593.5 | -25.36 | 40 | 9 | **TYKY**, 13 kDa, **75 kDa**, B14.5a, **PSST, ND1**, 18 kDa |
| **SDAP-☐** | 5667.7 | 731.3 | -3.93 | 7 | 12 | B14 |
| **SDAP-☐** | 6074.4 | 2554.8 | -9.23 | 22 | 26 | B22, B17, B12, **ND5**, AGGG |
| **42 kDa** | 16761.9 | 1942.1 | -19.89 | 2 | 0 | **ND2**, KFYI, B14.5b |
| **15 kDa** | 9228.8 | 5236.4 | -29.49 | 27 | 9 | SGDH, B16.6, **ND2, ND6, ND4L**, PGIV, B14.5b |
| **B9** | 8142.1 | 3420.4 | -27.55 | 19 | 8 | PGIV, **ND1**, B16.6, **TYKY, ND3**, SGDH |
| **B12** | 7188.3 | 2427.6 | -13.38 | 14 | 7 | **ND5**, AGGG, SDAP-☐☐, B22 |
| **B14.5b** | 10935.4 | 5873.6 | -39.62 | 31 | 8 | **ND2**, KFYI, PDSW, SGDH, PGIV, **ND4**, 15 kDa, ESSS, 42 kDa, B15 |
| **B15** | 10929.5 | 3772.9 | -15.63 | 16 | 0 | **ND4**, ASHI, B22, **ND5**, PDSW, B14.5b |
| **B16.6** | 14451.0 | 7839.8 | -51.27 | 42 | 12 | PGIV, 15 kDa, ND1, **49 kDa, ND6**, MWFE, B9, **TYKY**, B14.5a, **ND3**, SGDH |
| **B17** | 10137.0 | 4528.8 | -24.39 | 29 | 6 | B22, PDSW, **ND5**, B18, SDAP-☐☐, SGDH |
| **B18** | 9860.6 | 3512.3 | -6.33 | 24 | 6 | **ND5**, AGGG, ASHI, B17, PDSW |
| **B22** | 14868.0 | 5662.9 | -32.67 | 25 | 15 | **ND5**, B17, SDAP-☐☐, B15, SGDH, B12, **ND4**, ASHI |
| **AGGG** | 7273.2 | 3125.3 | -26.86 | 17 | 7 | **ND5**, B18, B12, SDAP-☐☐, ASHI |
| **ASHI** | 11797.4 | 2971.5 | -22.74 | 10 | 1 | B15, **ND5**, B18, **ND4**, AGGG, B22 |
| **ESSS** | 8835.5 | 4764.1 | -39.36 | 21 | 5 | **ND4**, PDSW, SGDH, B14.5b |
| **KFYI** | 5365.5 | 1349.6 | -12.98 | 0 | 2 | B14.5b, 42 kDa |
| **MNLL** | 5714.8 | 1719.4 | -14.42 | 9 | 5 | **ND4**, SGDH, PDSW |
| **MWFE** | 6638.2 | 3788.4 | -29.68 | 29 | 4 | **ND1**, PGIV, B16.6, **ND6, TYKY** |
| **PDSW** | 15770.6 | 7444.1 | -35.07 | 54 | 8 | ESSS, SGDH, **ND5**, B17, B14.5b, **ND4**, B18, B15, MNLL |
| **PGIV** | 14545.4 | 6552.6 | -38.16 | 49 | 28 | B16.6, B9, MWFE, SGDH, B14.5b, 15 kDa, **ND4, ND1** |
| **SGDH** | 14101.4 | 8031.0 | -55.12 | 40 | 10 | **ND4**, PDSW, 15 kDa, ESSS, B14.5b, MNLL, PGIV, B22, **ND2**, B17, **ND5**, B16.6, B9 |

**Table S4** Summary of interactions between subunits of ovine complex I. Analysis was performed using the PISA server (www.ebi.ac.uk/pdbe/pisa/). ΔG$_{int}$ indicates the solvation free energy gain upon formation of the assembly, N$_{HB}$ – number of hydrogen bonds at the interface and N$_{SB}$ – number of salt bridges at the interface

| Subunit name | | Location | Most frequently reported clinical feature | Mutation | Proposed reason for complex I disfunction | Reference |
|---|---|---|---|---|---|---|
| Human | Bovine | | | | | |
| NDUFS1 | 75 kDa | Peripheral domain/core | Leigh syndrome | Val48Asp | DE | (Haack, Madignier, et al., 2012) |
| | | | | Gly143Glu | DE | (Danhauser et al., 2011) |
| | | | | Val205Ala | DE | (Pagniez-Mammeri et al., 2009) |
| | | | | Leu208Val | DE | (Martin et al., 2005) |
| | | | | Arg218Trp | DA | (Benit et al., 2001) |
| | | | | Asp229Gly | DA | (Benit et al., 2001) |
| | | | | Arg385Cys | DA | (Hoefs et al., 2010) |
| | | | | Gln499Lys | DA | (Bugiani et al., 2004) |
| | | | | Thr572Ala | DA | (Ferreira et al., 2011) |
| | | | | Asp596Asn | DA | (Visch et al., 2004) |
| | | | | Tyr672His | DA | (Haack, Madignier, et al., 2012) |
| | | | | Tyr672Cys | DA | (Haack, Madignier, et al., 2012) |
| | | | | Met684Val | DA | (Benit et al., 2001) |
| NDUFV1 | 51 kDa | Peripheral domain/core | Leigh syndrome | Ser36Pro | DA | (Koene et al., 2012) |
| | | | | Arg68Gly | DE | (Marin et al., 2013) |
| | | | | *Lys91Glu | DA | (Calvo et al., 2010) |
| | | | | Pro102Leu | DE | (Lieber et al., 2013) |
| | | | | *Arg127Trp | DA | (Calvo et al., 2010) |
| | | | | Arg179Pro | DA | (Marin et al., 2013) |
| | | | | Tyr184Cys | DE | (Benit et al., 2001) |
| | | | | Cys186Gly | DA,DE | (Benit et al., 2001) |
| | | | | Ala191Val | DA,DE | (Schuelke et al., 2002) |
| | | | | Glu194Lys | DA,DE | (Benit et al., 2001) |
| | | | | *Pro232Arg | DA | (Calvo et al., 2010) |
| | | | | Ala321Val | DA | (Schuelke et al., 1999) |
| | | | | *Glu357Lys | DA,DE | (Calvo et al., 2010) |
| | | | | *Arg366His | DE | (Calvo et al., 2010) |
| | | | | Arg366Cys | DE | (Breningstall, Shoffner, & Patterson, 2008) |
| | | | | Thr403Met | DE | (Schuelke et al., 1999) |
| | | | | Ala412Pro | DE | (Benit et al., 2001) |
| NDUFS2 | 49 kDa | Peripheral domain/core | Leigh syndrome, hypertrophic cardiomyopaty | Asp77Val | DA | (Haack, Haberberger, et al., 2012) |
| | | | | Arg85Gln | DA,DE | (Tuppen et al., 2010) |
| | | | | Arg105Gln | DA,DE | (Tuppen et al., 2010) |

186

| | | | | Glu115Lys | DA | (Tuppen et al., 2010) |
|---|---|---|---|---|---|---|
| | | | | Ala191Val | DA,DE | (Bugiani et al., 2004) |
| | | | | Arg195Gln | DA | (Loeffen et al., 2001) |
| | | | | Pro196Gln | DA | (Loeffen et al., 2001) |
| | | | | Met259Thr | DA | (McFarland et al., 2009) |
| | | | | Arg290Gln | DA | (Haack, Haberberger, et al., 2012) |
| | | | | Arg300Gln | DA | (Tuppen et al., 2010) |
| | | | | Ser380Pro | DA | (Loeffen et al., 2001) |
| | | | | *Ala393Ser | DA | (Calvo et al., 2010) |
| | | | | Met410Lys | DA | (Tuppen et al., 2010) |
| | | | | Asp413Asn | DA | (Distelmaier et al., 2009) |
| NDUFS3 | 30 kDa | Peripheral domain/core | Leigh syndrome | Thr109Ile | DA | (Benit et al., 2004) |
| | | | | Arg163Trp | DA | (Benit et al., 2004) |
| | | | | Pro187Leu | DA | (Pagniez-Mammeri et al., 2009) |
| NDUFV2 | 24 kDa | Peripheral domain/core | Hypertrophic cardiomyopaty | Lys177Arg | DA | (Nishioka et al., 2010) |
| NDUFS8 | TYKY | Peripheral domain/core | Leigh syndrome, hypertrophic cardiomyopaty | Arg20Trp | DA | (Marina et al., 2013) |
| | | | | Glu29Gln | DA | (Haack, Haberberger, et al., 2012) |
| | | | | Arg43Trp | DA | (Haack, Haberberger, et al., 2012) |
| | | | | Pro45Leu | DA | (Loeffen et al., 1998) |
| | | | | Pro51Leu | DA | (Procaccio & Wallace, 2004) |
| | | | | Arg60Cys | DA | (Visch et al., 2006) |
| | | | | Arg68His | DE | (Loeffen et al., 1998) |
| | | | | Arg104His | DA,DE | (Procaccio & Wallace, 2004) |
| | | | | *Gly120Ser | DE | (Calvo et al., 2010) |
| | | | | Ala125Asp | DE | (Haack, Haberberger, et al., 2012) |
| NDUFS7 | PSST | Peripheral domain/core | Leigh syndrome | Val84Met | DE | (Loeffen et al., 1998) |
| | | | | Arg107His | DA,DE | (Lebon et al., 2007) |
| MT-ND6 | ND6 | Membrane domain/ core | Broad spectrum | Ser132Leu | DA, DP | (Yu-Wai-Man & Chinnery, 1993) |
| | | | | Asn117Asp | DA, DP | (Yu-Wai-Man & Chinnery, 1993) |
| | | | | Ala72Val | DP | (Ravn et al., 2001) |
| | | | | Ala74Val | DP | (Jun, Brown, & Wallace, 1994) |

| | | | | Met64Ile | DP | (Yu-Wai-Man & Chinnery, 1993) |
|---|---|---|---|---|---|---|
| | | | | Tyr59Cys | DP | (Yu-Wai-Man & Chinnery, 1993) |
| | | | | Leu60Ser | DP | (Yu-Wai-Man & Chinnery, 1993) |
| | | | | Gly36Ser | DP | (Yu-Wai-Man & Chinnery, 1993) |
| | | | | Ile26Met | DA, catalysis | (Yu-Wai-Man & Chinnery, 1993) |
| | | | | Met64Val | DP | (Macmillan, Johns, Fu, & Shoubridge, 2000) |
| MT-ND5 | ND5 | Membrane domain/ core | Broad spectrum | Phe124Leu | DP | (Taylor, Morris, Hutchinson, & Turnbull, 2002) |
| | | | | Glu145Gly | DP | (Liolitsa, Rahman, Benton, Carr, & Hanna, 2003) |
| | | | | Tyr159His | DA | (Yu-Wai-Man & Chinnery, 1993) |
| | | | | Ala171Val | DP | (Mayorov, Biousse, Newman, & Brown, 2005) |
| | | | | Ala236Thr | DP | (Naini et al., 2005) |
| | | | | Met237Leu | DP | (Liolitsa et al., 2003) |
| | | | | Ser244Cys | DP | (Crimi et al., 2003) |
| | | | | Asp393Asn | DP | (Sudo, Honzawa, Nonaka, & Goto, 2004) |
| | | | | Asp393Gly | DP | (Bugiani et al., 2004) |
| | | | | Gln434Arg | DA | (Yu-Wai-Man & Chinnery, 1993) |
| | | | | Gly465Glu | DA | (Howell, Halvorson, Burns, McCullough, & Paulton, 1993) |
| | | | | Ser250Cys | DP | (Crimi et al., 2003) |
| MT-ND4L | ND4L | Membrane domain/ core | Leber optic atrophy | Val65Ala | DP | (Brown et al., 2002) |
| | | | | Cys32Arg | DP | (Polyak et al., 1998) |
| MT-ND4 | ND4 | Membrane domain/ core | Broad spectrum | Ile165Thr | DA | (Leo-Kottler, Luberichs, Besch, Christ-Adler, & Fauser, 2002) |
| | | | | Val313Ile | DA | (Yu-Wai-Man & Chinnery, 1993) |
| | | | | Arg340Ser | DA | (Q. Wang et al., 2005) |
| | | | | Arg340His | DA | (Wallace et al., 1988) |
| | | | | Tyr409His | DA | (Vanniarajan et al., 2006) |

188

| | | | | Ser34Pro | DP | (McFarland et al., 2004) |
|---|---|---|---|---|---|---|
| | | | | Ser45Pro | DP | (McFarland et al., 2004) |
| MT-ND3 | ND3 | Membrane domain/ core | Leigh syndrome, leber optic atrophy | Gly47Ala | DP | (Sarzi et al., 2007) |
| | | | | Ile60Thr | DP | (Yu-Wai-Man & Chinnery, 1993) |
| | | | | Glu66Asn | DP | (Leshinsky-Silver et al., 2010) |
| MT-ND2 | ND2 | Membrane domain/ core | Broad spectrum | Ile57Met | DA | (Brown et al., 2001) |
| | | | | Leu71Pro | DA | (Hinttala et al., 2006) |
| | | | | Gly259Ser | DA | (Brown et al., 1992) |
| | | | | Ala331Ser | DA | (Q. Wang et al., 2005) |
| MT-ND1 | ND1 | Membrane domain/ core | Broad spectrum | Val11Met | DE | (Blakely et al., 2005) |
| | | | | Leu28Met | DE | (Gutierrez Cortes et al., 2012) |
| | | | | Tyr30His | DE | (Obayashi et al., 1992) |
| | | | | Met31Val | DE | (Hutchin & Cortopassi, 1995) |
| | | | | Met53Ile | DE | (Funalot et al., 2002) |
| | | | | Glu59Lys | DE | (Moslemi et al., 2008) |
| | | | | Ser110Asn | DP | (Yu-Wai-Man & Chinnery, 1993) |
| | | | | Gly131Ser | DP | (Kirby et al., 2004) |
| | | | | Ala132Thr | DP | (Yu-Wai-Man & Chinnery, 1993) |
| | | | | Glu143Lys | DP | (Yu-Wai-Man & Chinnery, 1993) |
| | | | | Thr164Ala | DA | (Simon et al., 2003) |
| | | | | Arg195Gln | coupling | (Moslemi et al., 2008) |
| | | | | Val208Leu | coupling | (Wray et al., 2013) |
| | | | | Glu214Lys | coupling | (Kirby et al., 2004) |
| | | | | Tyr215His | coupling | (Kirby et al., 2004) |

189

| | | | | Thr240Met | DA | (Yu-Wai-Man & Chinnery, 1993) |
|---|---|---|---|---|---|---|
| | | | | Tyr277Cys | DP | (Howell et al., 1991) |
| | | | | Leu285Pro | DA | (Howell et al., 1991) |
| | | | | Leu289Met | DA | (Yu-Wai-Man & Chinnery, 1993) |
| | | | | Glu24Lys | DP | (Blakely et al., 2005) |
| | | | | Ala52Thr | DP | (Howell et al., 1991) |
| | | | | Ala112Thr | DP | (Q. Wang et al., 2005) |
| NDUFS4 | 18 kDa | Peripheral domain/ supernumerary | Leigh syndrome, hypertrophic cardiomyopaty | Asp77His | DA | (Leshinsky-Silver et al., 2009) |
| NDUFS6 | 13 kDa | Peripheral domain/ supernumerary | Fatal lactic acidosis | Cys87Tyr | DA | (Spiegel et al., 2009) |
| NDUFV3 | 10 kDa | Peripheral domain/ supernumerary | Complex I deficiency | *Lys22Asn | not modelled | (Calvo et al., 2010) |
| | | | | *Gly69Asp | DA | (Calvo et al., 2010) |
| | | | | *Glu242Lys | not modelled | (Calvo et al., 2010) |
| NDUFA9 | 39 kDa | Peripheral domain/ supernumerary | Leigh syndrome | Arg286Pro | DA, catalysis | (van den Bosch et al., 2012) |
| NDUFA6 | B14 | Peripheral domain/ supernumerary | Complex I deficiency | *Ser3Asn | not modelled | (Calvo et al., 2010) |
| NDUFA1 | MWFE | Membrane domain/ supernumerary | Leigh syndrome | Gly8Arg | DA | (Fernandez-Moreira et al., 2007) |
| | | | | Gly32Arg | DA | (Potluri et al., 2009) |
| | | | | Arg37Ser | DA | (Fernandez-Moreira et al., 2007) |
| | | | | Tyr43His | DA | |
| NDUFA8 | PGIV | Membrane domain/ supernumerary | Complex I deficiency | Glu108Lys | DA | (Bugiani et al., 2004) |
| | | | | *Arg134Gln | DA | (Calvo et al., 2010) |
| NDUFA10 | 42 kDa | Membrane domain/ supernumerary | Complex I deficiency | Gly64Glu | DA | (Haack, Haberberger, et al., 2012) |
| | | | | Gln107Arg | DA | (Hoefs et al., 2011) |
| NDUFA13 | B16.6 | Membrane domain/ supernumerary | Encephalopathy | Lys4Asn | DA | (Maximo et al., 2005) |
| NDUFS5 | 15 kDa | Membrane domain/ supernumerary | Complex I deficiency | *Pro95Ser | DA | (Calvo et al., 2010) |

| NDUFB1 | MNLL | Membrane domain/ supernumerary | Complex I deficiency | *Met65Leu | not modelled | (Calvo et al., 2010) |
|--------|------|-------------------------------|----------------------|-----------|--------------|---------------------|
| NDUFB3 | B12  | Membrane domain/ supernumerary | Encephalomyo pathy, lactic acidosis | *Trp21Arg | DA | (Calvo et al., 2010) |
| NDUFB9 | B22  | Membrane domain/ supernumerary | Encephalomyo pathy | Leu63Pro | DA | (Haack, Haberberger, et al., 2012) |
|        |      |                               |                      | *Tyr96Cys | DA | (Calvo et al., 2010) |

**Table S5** Summary of the single amino acid substitutions in complex I subunits leading to complex I deficiencies. DA – disassembly, DE – disruption to electron transfer, DP – disruption to proton translocation, * - likely pathogenic.

| Module | Escherichia coli | | Thermus thermophilus | | Yarrowia lipolytica | | Bos taurus (Bovine) | | Homo sapiens[1] | Cofactors[2] |
|---|---|---|---|---|---|---|---|---|---|---|
| **Peripheral arm** | | | | | | | | | | Cofactors[2] |
| Dehydrogenase (N) | **NuoF** | | **Nqo1** | | **NUBM** | | **51 kDa** | | **NDUFV1** | FMN<br>N3 (4Fe[51]) |
| | 445[3] | 49.3[4] | 438 | 48.6 | 470 | 51.6 | 444 | 48.5 | | |
| | **NuoE** | | **Nqo2** | | **NUHM** | | **24 kDa** | | **NDUFV2** | N1a (2Fe[24]) |
| | 166 | 18.6 | 180 | 20.2 | 215 | 24.1 | 217 | 23.8 | | |
| | **NuoG** | | **Nqo3** | | **NUAM** | | **75 kDa** | | **NDUFS1** | N1b (2Fe[75])<br>N4 (4Fe[75]C)<br>N5 (4Fe[75]H)<br>(N7)[5] |
| | 907 | 100.2 | 783 | 86.5 | 694 | 75.2 | 704 | 77.0 | | |
| Connecting (Q) | **NuoD (NuoCD)[6]** | | **Nqo4** | | **NUCM** | | **49 kDa** | | **NDUFS2** | No cofactor |
| | 595 | 68.1 | 409 | 46.4 | 444 | 49.9 | 430 | 49.1 | | |
| | **NuoC** | | **Nqo5** | | **NUGM** | | **30 kDa** | | **NDUFS3** | No cofactor |
| | see above | | 207 | 23.4 | 251 | 29.2 | 228 | 26.4 | | |
| | **Nuol** | | **Nqo9** | | **NUIM** | | **TYKY** | | **NDUFS8** | N6a (4Fe{TY]1)<br>N6b (4Fe[TY]2) |
| | 180 | 20.5 | 182 | 20.1 | 198 | 22.3 | 176 | 20.2 | | |
| | **NuoB** | | **Nqo6** | | **NUKM** | | **PSST** | | **NDUFS7** | N2 (4Fe[PS]) |
| | 220 | 25.1 | 180 | 20.1 | 183 | 20.4 | 179 | 20.1 | | |
| **Membrane arm** | | | | | | | | | | TMH[7] |
| - | **NuoH** | | **Nqo8** | | **NU1M** | | **ND1** | | **ND1** | 8-9 |
| | 325 | 36.2 | 365 | 41.0 | 341 | 38.3 | 318 | 35.7 | | |
| Pumping (P) | **NuoA** | | **Nqo7** | | **NU3M** | | **ND3** | | **ND3** | 3 |
| | 147 | 16.5 | 119 | 13.1 | 128 | 14.5 | 115 | 13.1 | | |
| | **NuoJ** | | **Nqo10** | | **NU6M** | | **ND6** | | **ND6** | 5 |
| | 184 | 19.9 | 176 | 18.6 | 185 | 20.8 | 175 | 19.1 | | |
| | **NuoK** | | **Nqo11** | | **NULM** | | **ND4L** | | **ND4L** | 3 |
| | 100 | 10.8 | 95 | 10.0 | 89 | 9.8 | 98 | 10.8 | | |
| | **NuoN** | | **Nqo14** | | **NU2M** | | **ND2** | | **ND2** | 11-14 |
| | 485 | 52.0 | 427 | 44.4 | 469 | 53.3 | 347 | 39.3 | | |
| | **NuoM** | | **Nqo13** | | **NU4M** | | **ND4** | | **ND4** | 14 |
| | 509 | 56.5 | 469 | 49.2 | 486 | 54.5 | 459 | 52.1 | | |
| | **NuoL** | | **Nqo12** | | **NU5M** | | **ND5** | | **ND5** | 16-17 |
| | 613 | 66.5 | 606 | 65.1 | 655 | 73.7 | 606 | 68.3 | | |
| Total Core | 4876 | 540 | 4636 | 507 | 4808 | 538 | 4496 | 503.5 | | |

**Table S6.** Core subunits of complex I. Table modified from prepared by Dr James.

[1] *Homo sapiens* core subunit names are shown for reference, no structures of human Complex I have been elucidated

[2] The traditional nomenclature for Fe-S clusters (Nx, derived from initially described electron paramagnetic resonance (EPR) signatures), as well as the nomenclature proposed recently on the basis of re-assignment of EPR signals to structurally observed clusters, is shown. In the new nomenclature, clusters are named according to their nuclearity (2Fe or 4Fe), their subunit location (using bovine nomenclature) and when necessary, as ligated by four Cys (C) or three Cys and one His (H).

[3] Number of amino acids in the mature subunit

[4] MW of the mature subunit in kilo-Daltons

[5] Cluster N7 is present only in some bacteria (for example, *E. coli* and *T. thermophilus*)

[6] Subunits NuoC and NuoD are fused in *E. coli* and some other bacteria

[7] Number of transmembrane helices

| Module | Bos taurus | | Homo sapiens | Number of TMHs | Notes |
|---|---|---|---|---|---|
| **N**<br><br>51 kDa<br>24 kDa<br>75 kDa | 18 kDa<br>133[2] 15.3[3] | | NDUFS4 | | |
| | 13 kDa<br>96 10.5 | | NDUFS6 | | $Zn^{2+}$ binding |
| | 10 kDa<br>75 8.4 | | NDUFV3 | | |
| | B17.2<br>145 17.1 | | NDUFA12 | | Homologous to assembly factor B17.2L |
| | B8<br>98 10.9 | | NDUFA2 | | Thioredoxin fold |
| **Q**<br><br>49 kDa<br>30 kDa<br>PSST<br>TYKY | B14.5a<br>112 12.5 | | NDUFA7 | | |
| | SDAP-α<br>88 10.1 | | NDUFAB1 | | Acyl carrier protein |
| | 39 kDa<br>345 39.1 | | NDUFA9 | | NADPH-binding short-chain dehydrogenase/reductase family |
| | B13<br>115 13.2 | | NDUFA5 | | |
| | B14<br>127 14.9 | | NDUFA6 | | LYR protein |
| **ND1** | MWFE<br>70 8.1 | | NDUFA1 | | |
| | B16.6<br>143 16.5 | | NDUFA13 | | Identical to GRIM-19 in humans and cows |
| | B9<br>83 9.2 | | NDUFA3 | 1 | |
| | PGIV<br>171 20 | | NDUFA8 | | Quadruple $CX_9C$, 2 CHCH domains |
| **P**<br><br>ND2<br>ND3<br>ND4L<br>ND4<br>ND5<br>ND6 | B14.7<br>140 14.6 | | NDUFA11 | 4 | Related to TIM 17, 22, 23 |
| | B17<br>127 15.4 | | NDUFB6 | 1 | |
| | 42 kDa<br>320 36.7 | | NDUFA10 | | Nucleoside kinase family |
| | KFYI<br>49 5.8 | | NDUFC1 | 1 | |
| | 15 kDa<br>105 12.5 | | NDUFS5 | | Double $CX_9C$ |
| | ASHI<br>158 18.7 | | NDUFB8 | 1 | |
| | B14.5b<br>120 14.1 | | NDUFC2 | 2 | |
| | AGGG | | NDUFB2 | 1 | |

193

| Name | Amino acids[2] | MW[3] | Gene | [1] | Notes |
|---|---|---|---|---|---|
| | 72 | 8.5 | | | |
| MNLL | 57 | 6.7 | NDUFB1 | 1 | |
| SGDH | 143 | 16.7 | NDUFB5 | 1 | |
| B22 | 178 | 21 | NDUFB9 | | LYR protein |
| B18 | 136 | 16.3 | NDUFB7 | | Double CX$_9$C |
| SDAP-β | 88 | 10.1 | NDUFAB1 | | Acyl carrier protein |
| PDSW | 176 | 21 | NDUFB10 | | |
| B15 | 128 | 15.1 | NDUFB4 | 1 | |
| ESSS | 125 | 14.5 | NDUFB11 | 1 | |
| B12 | 97 | 11 | NDUFB3 | 1 | |

**Table S7** Supernumerary subunits of mitochondrial complex I. Table modified form prepared by Dr James Letts
[1] Assembly intermediate
[2] Number of amino acids in the mature subunit
[3] MW of the mature subunit in kilo-Daltons

| Subunit 1 | Residue | Subunit 2 | Residue | Cross - linker | Euclidian Distance (Å) | SAS (Å) | Number of times observed | Best Score |
|---|---|---|---|---|---|---|---|---|
| 51 kDa_NDUFV1 | Lys8 | 10 kDa_NDUFV3 | Lys6 | DSS | - | - | 1 | 39.92 |
| 51 kDa_NDUFV1 | Lys8 | 10 kDa_NDUFV3 | Lys17 | DSS | - | - | 1 | 32.05 |
| 51 kDa_NDUFV1 | Lys34 | 10 kDa_NDUFV3 | Lys6 | DSS | - | - | 2 | 40.70 |
| 51 kDa_NDUFV1 | Lys34 | 10 kDa_NDUFV3 | Lys17 | DSS | - | - | 2 | 37.83 |
| 51 kDa_NDUFV1 | Lys34 | 10 kDa_NDUFV3 | Lys22 | DSS | - | - | 2 | 30.38 |
| 51 kDa_NDUFV1 | Lys34 | 10 kDa_NDUFV3 | Lys23 | DSS | - | - | 3 | 39.59 |
| 51 kDa_NDUFV1 | Glu158 | 10 kDa_NDUFV3 | Glu31 | SDH | - | - | 1 | 28.79 |
| 51 kDa_NDUFV1 | Lys397 | 75 kDa_NDUFS1 | Lys87 | DSS | 20.8 | 29.0 | 2 | 31.34 |
| 51 kDa_NDUFV1 | Lys397 | 75 kDa_NDUFS1 | Lys88 | DSS | 10.1 | 12.3 | 1 | 31.68 |
| 24 kDa_NDUFV2 | Lys180 | 10 kDa_NDUFV3 | Lys17 | DSS | - | - | 1 | 36.92 |
| 24 kDa_NDUFV2 | Lys183 | 10 kDa_NDUFV3 | Lys23 | DSS | - | - | 1 | 28.87 |
| 75 kDa_NDUFS1 | Lys61 | 30 kDa_NDUFS3 | Lys223 | DSS | - | - | 3 | 40.12 |
| 75 kDa_NDUFS1 | Lys61 | B14.5a_NDUFA7 | Lys79 | DSS | - | - | 2 | 35.52 |
| 75 kDa_NDUFS1 | Glu297 | B14_NDUFA6 | Asp116 | ADH | 11.8 | 15.4 | 1 | 30.51 |
| 75 kDa_NDUFS1 | Glu297 | 39 kDa_NDUFA9 | Asp220 | SDH | 10.6 | 11.8 | 1 | 27.55 |
| 75 kDa_NDUFS1 | Asp486 | B8_NDUFA2 | Glu37 | SDH | 13.8 | 15.3 | 1 | 27.03 |
| 75 kDa_NDUFS1 | Asp504 | B8_NDUFA2 | Asp49 | SDH/ADH | 20.8 | - | 2 | 26.73 |
| 75 kDa_NDUFS1 | Lys516 | B8_NDUFA2 | Lys12 | DSS | - | - | 1 | 30.00 |
| 75 kDa_NDUFS1 | Lys598 | B14_NDUFA6 | Lys115 | DSS | 16.5 | 19.3 | 1 | 29.43 |
| 75 kDa_NDUFS1 | Glu624 | B8_NDUFA2 | Asp59 | SDH/ADH | 9.2 | 12.3 | 2 | 43.52 |
| 75 kDa_NDUFS1 | Lys650 | B8_NDUFA2 | Lys45 | DSS | 13.1 | 15.1 | 1 | 29.43 |
| 49 kDa_NDUFS2 | Lys334 | B16.6_NDUFA13 | Lys6 | DSS | 11.0 | 17.3 | 1 | 31.00 |
| 49 kDa_NDUFS2 | Lys334 | B14.5a_NDUFA7 | Lys39 | DSS/BS3/DSA | 9.8 | 9.9 | 7 | 33.87 |
| 49 kDa_NDUFS2 | Lys404 | B14.5a_NDUFA7 | Lys79 | DSS | - | - | 1 | 25.63 |
| 30 kDa_NDUFS3 | Lys20 | B14.5a_NDUFA7 | Lys79 | BS3 | - | - | 1 | 37.58 |
| 30 kDa_NDUFS3 | Lys20 | B14.5a_NDUFA7 | Lys92 | DSS/BS3 | 13.7 | 16.3 | 5 | 45.74 |
| 30 kDa_NDUFS3 | Glu158 | B14_NDUFA6 | Glu108 | ADH | 18.4 | 24.0 | 1 | 28.42 |
| 30 kDa_NDUFS3 | Glu158 | B14_NDUFA6 | Glu110 | SDH/ADH | 12.0 | 12.1 | 1 | 30.63 |
| 30 kDa_NDUFS3 | Asp178 | PSST_NDUFS7 | Glu109 | SDH | 12.6 | 14.4 | 1 | 34.49 |
| 30 kDa_NDUFS3 | Glu189 | B14_NDUFA6 | Glu108 | SDH | 22.0 | 35.4 | 2 | 35.35 |
| 30 kDa_NDUFS3 | Lys196 | 18 kDa_NDUFS4 | Lys56 | DSS | 9.0 | 10.4 | 1 | 31.17 |
| 30 kDa_NDUFS3 | Lys223 | B13_NDUFA5 | Lys45 | DSS | - | - | 1 | 29.47 |
| 30 kDa_NDUFS3 | Lys223 | 18 kDa_NDUFS4 | Lys53 | DSS | - | - | 1 | 31.96 |
| 30 kDa_NDUFS3 | Lys223 | B14.5a_NDUFA7 | Lys79 | DSS | - | - | 3 | 33.18 |
| 30 kDa_NDUFS3 | Lys224 | 18 kDa_NDUFS4 | Lys53 | DSS | - | - | 2 | 31.17 |
| 30 kDa_NDUFS3 | Lys224 | 18 kDa_NDUFS4 | Lys112 | DSS | - | - | 2 | 29.62 |
| PSST_NDUFS7 | Lys17 | 39 kDa_NDUFA9 | Lys78 | DSS | - | - | 2 | 30.05 |
| PSST_NDUFS7 | Lys17 | 39 kDa_NDUFA9 | Lys144 | DSS | - | - | 2 | 33.47 |
| TYKY_NDUFS8 | Lys15 | B9_NDUFA3 | Lys9 | DSS | 11.1 | 15.4 | 1 | 30.62 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| TYKY_NDUFS8 | Lys54 | B17.2_NDUFA12 | Lys43 | DSS | 19.4 | 19.9 | 1 | 29.83 |
| 18 kDa_NDUFS4 | Lys31 | B14_NDUFA6 | Lys68 | DSS | 11.1 | 13.0 | 1 | 27.97 |
| 18 kDa_NDUFS4 | Lys31 | B14_NDUFA6 | Lys120 | DSS | 12.1 | 12.9 | 1 | 33.32 |
| 15 kDa_NDUFS5 | Lys53 | SGDH_NDUFB5 | Lys117 | DSS | 8.9 | 10.9 | 1 | 30.45 |
| 15 kDa_NDUFS5 | Glu101 | B16.6_NDUFA13 | Glu91 | SDH/ADH | - | - | 2 | 25.99 |
| 15 kDa_NDUFS5 | Asp102 | B16.6_NDUFA13 | Glu88 | SDH | - | - | 1 | 29.29 |
| 15 kDa_NDUFS5 | Asp102 | B16.6_NDUFA13 | Glu91 | SDH | - | - | 1 | 28.28 |
| 13 kDa_NDUFS6 | Asp22 | 39 kDa_NDUFA9 | Glu162 | ADH | 23.6 | 27.4 | 1 | 32.67 |
| MWFE_NDUFA1 | Lys64 | PGIV_NDUFA8 | Lys18 | DSS/BS3/DSA | 7.9 | 11.9 | 6 | 35.25 |
| B13_NDUFA5 | Lys39 | B14.5a_NDUFA7 | Lys79 | DSS | - | - | 2 | 37.96 |
| B13_NDUFA5 | Lys45 | B14.5a_NDUFA7 | Lys79 | DSS/BS3/DSA | - | - | 7 | 44.62 |
| B13_NDUFA5 | Lys45 | B14.5a_NDUFA7 | Lys91 | DSS | 10.5 | 11.2 | 2 | 48.38 |
| B13_NDUFA5 | Lys45 | B14.5a_NDUFA7 | Lys92 | DSS/BS3/DSA | 14.7 | 21.2 | 7 | 45.83 |
| B13_NDUFA5 | Lys65 | B14.5a_NDUFA7 | Lys101 | DSS | 16.8 | 18.0 | 1 | 29.78 |
| B14.5a_NDUFA7 | Lys22 | B17.2_NDUFA12 | Lys53 | DSS | 16.3 | 19.2 | 1 | 43.39 |
| B14.5a_NDUFA7 | Lys32 | B17.2_NDUFA12 | Lys101 | DSS | 11.9 | 13.0 | 1 | 26.08 |
| B14.5a_NDUFA7 | Lys32 | B17.2_NDUFA12 | Lys43 | DSS | 19.3 | 25.5 | 3 | 37.85 |
| B14.5a_NDUFA7 | Lys32 | B17.2_NDUFA12 | Lys53 | DSS | 7.0 | 8.2 | 1 | 36.39 |
| B14.5a_NDUFA7 | Lys39 | B16.6_NDUFA13 | Lys6 | DSS | 10.9 | 13.3 | 1 | 42.73 |
| PGIV_NDUFA8 | Glu5 | B16.6_NDUFA13 | Glu91 | ADH | 15.8 | 18.3 | 1 | 30.86 |
| PGIV_NDUFA8 | Glu10 | B16.6_NDUFA13 | Glu91 | ADH | 10.8 | 12.6 | 1 | 34.43 |
| PGIV_NDUFA8 | Lys18 | B16.6_NDUFA13 | Lys78 | DSS | 19.4 | 24.9 | 1 | 35.3 |
| PGIV_NDUFA8 | Glu148 | B14.5b_NDUFC2 | Asp16 | SDH/ADH | 18.5 | 24.5 | 2 | 39.9 |
| PGIV_NDUFA8 | Glu148 | B14.5b_NDUFC2 | Glu99 | ADH | 14.4 | 16.1 | 1 | 32.81 |
| PGIV_NDUFA8 | Glu152 | B14.5b_NDUFC2 | Glu17 | SDH/ADH | 11.1 | 12.6 | 2 | 36.13 |
| PGIV_NDUFA8 | Glu152 | B14.5b_NDUFC2 | Asp100 | SDH/ADH | 21.7 | 29.4 | 2 | 36.24 |
| PGIV_NDUFA8 | Asp156 | B14.5b_NDUFC2 | Glu17 | SDH/ADH | 7.0 | 7.2 | 2 | 34.99 |
| PGIV_NDUFA8 | Lys158 | B14.5b_NDUFC2 | Lys95 | DSS | 12.7 | 13.5 | 3 | 44.15 |
| PGIV_NDUFA8 | Lys158 | B14.5b_NDUFC2 | Lys104 | DSS | 27.5 | 29.5 | 1 | 38.98 |
| PGIV_NDUFA8 | Lys158 | SGDH_NDUFB5 | Lys106 | DSS | 13.6 | 13.9 | 5 | 46.33 |
| 42 kDa_NDUFA10 | Lys87 | ESSS_NDUFB11 | Lys17 | DSS | - | - | 3 | 33.42 |
| MNLL_NDUFB1 | Lys29 | B14.5b_NDUFC2 | Lys25 | DSS | 11.9 | 12.2 | 1 | 23.68 |
| B12_NDUFB3 | Lys12 | SDAP_NDUFAB1 | Lys20 | DSS | - | - | 2 | 32.79 |
| B12_NDUFB3 | Lys19 | SDAP_NDUFAB1 | Lys20 | DSS | 6.4 | 6.9 | 1 | 36.87 |
| B15_NDUFB4 | Lys30 | ASHI_NDUFB8 | Lys25 | DSS/BS3/DSA | 22.7 | 26.5 | 6 | 39.21 |
| B15_NDUFB4 | Lys119 | PDSW_NDUFB10 | Lys170 | DSS/BS3/DSA | 26.7 | 30.7 | 7 | 38.09 |
| SGDH_NDUFB5 | Lys106 | B14.5b_NDUFC2 | Lys95 | DSS | 12.6 | 14.6 | 3 | 37.89 |
| SGDH_NDUFB5 | Lys106 | B14.5b_NDUFC2 | Lys104 | DSS | 14.3 | 15.8 | 8 | 44.04 |
| SGDH_NDUFB5 | Lys106 | B14.5b_NDUFC2 | Lys106 | DSS | 11.4 | 12.1 | 2 | 39.56 |
| SGDH_NDUFB5 | Lys106 | B14.5b_NDUFC2 | Lys107 | DSS | 19 | 21.1 | 2 | 33.65 |
| B17_NDUFB6 | Lys23 | B22_NDUFB9 | Lys120 | DSS | 10.6 | 13.3 | 1 | 31.11 |
| ASHI_NDUFB8 | Lys25 | ESSS_NDUFB11 | Lys17 | DSS | - | - | 6 | 34.91 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ASHI_NDUFB8 | Lys49 | ESSS_NDUFB11 | Lys17 | DSS/BS3/DSA | - | - | 1 | 27.74 |
| ASHI_NDUFB8 | Lys148 | PDSW_NDUFB10 | Lys126 | DSS | - | - | 1 | 31.02 |
| ASHI_NDUFB8 | Lys148 | PDSW_NDUFB10 | Lys167 | DSS | - | - | 1 | 29.02 |
| ASHI_NDUFB8 | Lys148 | PDSW_NDUFB10 | Lys170 | DSS | - | - | 2 | 32.55 |
| B22_NDUFB9 | Lys137 | ESSS_NDUFB11 | Lys17 | DSS/BS3/DSA | - | - | 6 | 36.24 |
| PDSW_NDUFB10 | Glu165 | ESSS_NDUFB11 | Glu109 | SDH | 17.1 | 19.1 | 1 | 32.16 |

**Table S8** Inter-subunit cross - links. Residues in red are not modelled in the structure. Solvent accessible surface distances (SAS) calculated using Xwalk are shown. If the SAS distance is less than 8 Å over the approximate maximum cross - link length the distance is shown in red. These are considered true cross - links because of the flexibility of the side chains and were each checked manually in the structure. False positive cross - links are too far apart in the structure and likely result from inter-complex cross - linking.Table modified form prepared by Dr James Letts.

| Subunit 1 | Residue | Subunit 2 | Residue | Cross - linker | Euclidian Distance (Å) | SAS (Å) | Number of times observed | Best Score |
|---|---|---|---|---|---|---|---|---|
| 51 kDa_NDUFV1 | Lys16 | 51 kDa_NDUFV1 | Lys8 | DSS | 8.7 | 10.5 | 2 | 38.06 |
| 51 kDa_NDUFV1 | Lys61 | 51 kDa_NDUFV1 | Lys78 | DSS | 13.6 | 24.7 | 1 | 30.07 |
| 51 kDa_NDUFV1 | Lys61 | 51 kDa_NDUFV1 | Lys84 | DSS | 15.0 | 17.9 | 4 | 46.78 |
| 24 kDa_NDUFV2 | Lys29 | 24 kDa_NDUFV2 | Lys36 | DSS | 6.7 | 7.7 | 2 | 35.59 |
| 24 kDa_NDUFV2 | Glu40 | 24 kDa_NDUFV2 | Glu32 | SDH | 11.8 | 11.8 | 1 | 38.17 |
| 24 kDa_NDUFV2 | Lys122 | 24 kDa_NDUFV2 | Lys167 | DSS | 17.2 | 21.9 | 1 | 28.76 |
| 75 kDa_NDUFS1 | Lys61 | 75 kDa_NDUFS1 | Lys75 | DSS | 17.5 | 19.7 | 3 | 32.37 |
| 75 kDa_NDUFS1 | Lys61 | 75 kDa_NDUFS1 | Lys85 | DSS | 11.4 | 11.6 | 1 | 35.52 |
| 75 kDa_NDUFS1 | Lys64 | 75 kDa_NDUFS1 | Lys85 | DSS | 12.0 | 18.9 | 1 | 32.99 |
| 75 kDa_NDUFS1 | Lys275 | 75 kDa_NDUFS1 | Lys686 | DSS | 12.9 | 13.1 | 1 | 27.89 |
| 75 kDa_NDUFS1 | Lys508 | 75 kDa_NDUFS1 | Lys516 | DSS | 8.1 | 8.1 | 7 | 38.03 |
| 49 kDa_NDUFS2 | Glu302 | 49 kDa_NDUFS2 | Glu227 | ADH | 16.6 | 24.4 | 1 | 32.51 |
| 30 kDa_NDUFS3 | Asp178 | 30 kDa_NDUFS3 | Glu189 | SDH/ADH | 22.8 | - | 1 | 30.87 |
| 30 kDa_NDUFS3 | Glu179 | 30 kDa_NDUFS3 | Glu189 | SDH | 17.9 | 32.3 | 1 | 29.17 |
| 30 kDa_NDUFS3 | Lys217 | 30 kDa_NDUFS3 | Lys223 | DSS | - | - | 2 | 42.78 |
| PSST_NDUFS7 | Lys17 | PSST_NDUFS7 | Lys32 | DSS | 12.1 | 18.2 | 2 | 36.22 |
| PSST_NDUFS7 | Lys17 | PSST_NDUFS7 | Lys102 | DSS | - | - | 1 | 25.57 |
| 10 kDa_NDUFV3 | Lys6 | 10 kDa_NDUFV3 | Lys17 | DSS | - | - | 2 | 34.46 |
| 10 kDa_NDUFV3 | Lys6 | 10 kDa_NDUFV3 | Lys23 | DSS | - | - | 1 | 25.17 |
| 10 kDa_NDUFV3 | Lys16 | 10 kDa_NDUFV3 | Lys23 | DSS | - | - | 1 | 26.12 |
| 18 kDa_NDUFS4 | Glu15 | 18 kDa_NDUFS4 | Asp91 | ADH | 23.5 | 26.1 | 1 | 35.81 |
| 18 kDa_NDUFS4 | Lys53 | 18 kDa_NDUFS4 | Lys109 | DSS | 18.7 | 27.4 | 1 | 38.56 |
| 18 kDa_NDUFS4 | Lys112 | 18 kDa_NDUFS4 | Lys126 | DSS | 17.0 | 18.4 | 1 | 33.81 |
| 18 kDa_NDUFS4 | Lys114 | 18 kDa_NDUFS4 | Lys126 | DSS | 16.9 | 19.5 | 2 | 31.85 |
| 15 kDa_NDUFS5 | Lys79 | 15 kDa_NDUFS5 | Lys84 | DSS | 15.6 | 25.3 | 1 | 32.98 |
| 15 kDa_NDUFS5 | Lys79 | 15 kDa_NDUFS5 | Lys87 | DSS | 19.3 | 23.1 | 3 | 40.95 |
| 15 kDa_NDUFS5 | Lys79 | 15 kDa_NDUFS5 | Lys90 | DSS | 22.9 | 30.0 | 1 | 37.70 |
| 13 kDa_NDUFS6 | Glu9 | 13 kDa_NDUFS6 | Asp19 | ADH | 10.5 | 26.9 | 1 | 24.92 |
| 13 kDa_NDUFS6 | Lys10 | 13 kDa_NDUFS6 | Lys33 | DSS | 6.5 | 7.4 | 1 | 28.36 |
| B8_NDUFA2 | Lys12 | B8_NDUFA2 | Lys74 | DSS/BS3 | - | - | 3 | 38.11 |
| B8_NDUFA2 | Lys12 | B8_NDUFA2 | Lys97 | DSS/BS3/DSA | - | - | 5 | 32.69 |
| B8_NDUFA2 | Lys38 | B8_NDUFA2 | Lys45 | DSS/BS3 | 15.2 | 15.7 | 5 | 42.23 |
| B8_NDUFA2 | Glu42 | B8_NDUFA2 | Glu91 | SDH/ADH | 10.1 | 11.4 | 2 | 36.53 |
| B8_NDUFA2 | Lys74 | B8_NDUFA2 | Lys97 | DSS/BS3 | - | - | 8 | 42.40 |
| B13_NDUFA5 | Lys6 | B13_NDUFA5 | Lys65 | DSS | 11.6 | 13.2 | 2 | 30.26 |
| B13_NDUFA5 | Lys24 | B13_NDUFA5 | Lys59 | DSS/BS3/DSA | 5.4 | 5.4 | 9 | 44.79 |
| B13_NDUFA5 | Lys39 | B13_NDUFA5 | Lys45 | DSS | 6.7 | 7.1 | 3 | 34.82 |
| B14_NDUFA6 | Lys56 | B14_NDUFA6 | Lys68 | DSS | 20.7 | 25.2 | 1 | 35.17 |
| B14_NDUFA6 | Lys67 | B14_NDUFA6 | Lys84 | DSS | 21.0 | 26.5 | 1 | 29.51 |
| B14_NDUFA6 | Lys68 | B14_NDUFA6 | Lys84 | DSS | 10.6 | 11.6 | 1 | 30.39 |
| B14_NDUFA6 | Lys84 | B14_NDUFA6 | Lys120 | DSS | 26.7 | 31.2 | 1 | 28.85 |
| B14.5a_NDUFA7 | Lys22 | B14.5a_NDUFA7 | Lys32 | DSS/BS3/DSA | 20.2 | 23.0 | 6 | 34.86 |
| B14.5a_NDUFA7 | Lys32 | B14.5a_NDUFA7 | Lys39 | DSS | 23.6 | 26.6 | 3 | 34.08 |
| B14.5a_NDUFA7 | Lys79 | B14.5a_NDUFA7 | Lys92 | DSS | - | - | 1 | 39.81 |
| PGIV_NDUFA8 | Lys40 | PGIV_NDUFA8 | Lys50 | DSS | 17.7 | - | 1 | 29.26 |
| PGIV_NDUFA8 | Lys40 | PGIV_NDUFA8 | Lys131 | DSS | 12.8 | - | 1 | 32.48 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| PGIV_NDUFA8 | Glu138 | PGIV_NDUFA8 | Glu152 | ADH | 26.2 | <span style="color:red">28.1</span> | 1 | 42.97 |
| 39 kDa_NDUFA9 | Lys78 | 39 kDa_NDUFA9 | Lys122 | DSS | 11.3 | 24.9 | 2 | 38.48 |
| 39 kDa_NDUFA9 | Asp79 | 39 kDa_NDUFA9 | Glu86 | SDH | 12.6 | 15.5 | 1 | 34.29 |
| 39 kDa_NDUFA9 | Lys140 | 39 kDa_NDUFA9 | Lys149 | DSS | 16.5 | - | 1 | 31.25 |
| 39 kDa_NDUFA9 | Lys140 | 39 kDa_NDUFA9 | Lys154 | DSS/BS3 | 11.0 | 13.5 | 8 | 38.22 |
| 39 kDa_NDUFA9 | Lys144 | 39 kDa_NDUFA9 | Lys149 | DSS | 13.8 | <span style="color:red">-</span> | 3 | 38.39 |
| 39 kDa_NDUFA9 | Lys149 | 39 kDa_NDUFA9 | Lys283 | BS3 | 14.2 | 20.7 | 1 | 41.33 |
| 42 kDa_NDUFA10 | Lys35 | 42 kDa_NDUFA10 | Lys185 | DSS | 20.2 | <span style="color:red">25.5</span> | 1 | 34.62 |
| 42 kDa_NDUFA10 | Lys35 | 42 kDa_NDUFA10 | Lys186 | DSS | 14.3 | 16.1 | 1 | 41.18 |
| 42 kDa_NDUFA10 | Lys37 | 42 kDa_NDUFA10 | Lys45 | DSS | 16.2 | 19.8 | 2 | 36.39 |
| 42 kDa_NDUFA10 | Lys37 | 42 kDa_NDUFA10 | Lys185 | DSS | 14.4 | 15.8 | 1 | 28.54 |
| 42 kDa_NDUFA10 | Lys37 | 42 kDa_NDUFA10 | Lys186 | DSS | 10.4 | 11.4 | 3 | 37.80 |
| 42 kDa_NDUFA10 | Lys40 | 42 kDa_NDUFA10 | Lys49 | DSS | 12.9 | 16.9 | 2 | 37.25 |
| 42 kDa_NDUFA10 | Lys40 | 42 kDa_NDUFA10 | Lys186 | DSS/BS3 | 16.8 | 18.9 | 7 | 43.10 |
| 42 kDa_NDUFA10 | Lys87 | 42 kDa_NDUFA10 | Lys193 | DSS | 18.1 | 21.0 | 1 | 40.29 |
| 42 kDa_NDUFA10 | Lys245 | 42 kDa_NDUFA10 | Lys255 | DSS | 20.4 | - | 1 | 25.53 |
| B15_NDUFB4 | Lys113 | B15_NDUFB4 | Lys119 | DSS | 18.6 | 22.9 | 2 | 31.92 |
| B18_NDUFB7 | Lys111 | B18_NDUFB7 | <span style="color:red">Lys123</span> | DSS | - | - | 1 | 25.82 |
| B18_NDUFB7 | Lys112 | B18_NDUFB7 | <span style="color:red">Lys123</span> | DSS | - | - | 1 | 27.64 |
| ASHI_NDUFB8 | Lys6 | ASHI_NDUFB8 | Lys25 | DSS | 19.9 | <span style="color:red">35.5</span> | 5 | 34.09 |
| ASHI_NDUFB8 | Lys6 | ASHI_NDUFB8 | Lys26 | DSS | 16.0 | 20.5 | 1 | 31.32 |
| B22_NDUFB9 | Lys51 | B22_NDUFB9 | Lys58 | DSS | 19.1 | 26.3 | 1 | 38.30 |
| PDSW_NDUFB10 | Lys6 | PDSW_NDUFB10 | Lys126 | DSS | 15.8 | 18.5 | 1 | 32.69 |
| B17.2_NDUFA12 | Lys43 | B17.2_NDUFA12 | Lys101 | DSS | 13.5 | 22.6 | 1 | 32.59 |
| B17.2_NDUFA12 | Lys43 | B17.2_NDUFA12 | Lys107 | DSS | 22.5 | <span style="color:red">25.9</span> | 1 | 30.15 |
| SDAP_NDUFAB1 | Lys12 | SDAP_NDUFAB1 | Lys20 | DSS | 11.7 | 11.7 | 1 | 37.22 |
| B14.5b_NDUFC2 | Lys104 | B14.5b_NDUFC2 | Lys107 | DSS | 6.2 | 7.6 | 1 | 26.21 |

**Table S9** Intra-subunit cross - links. Residues in red are not modelled in the structure. Solvent accessible surface distances (SAS) calculated using Xwalk are shown. If the SAS distance is less than 8 Å over the approximate maximum cross - link length the distance is shown in red. These are considered true cross - links because of the flexibility of the side chains and were each checked manually in the structure. False positive cross - links are too far apart in the structure and likely result from inter-complex cross - linking.Table modified form prepared by Dr James Letts

# References

Abdrakhmanova, A., Zickermann, V., Bostina, M., Radermacher, M., Schagger, H., Kerscher, S., & Brandt, U. (2004). Subunit composition of mitochondrial complex I from the yeast Yarrowia lipolytica. *Biochim Biophys Acta, 1658*(1-2), 148-156

Abdrakhmanova, A., Zwicker, K., Kerscher, S., Zickermann, V., & Brandt, U. (2006). 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*(12), 1676-1682

Abrahams, J. P., Leslie, A. G., Lutter, R., & Walker, J. E. (1994). Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. *Nature, 370*(6491), 621-628

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., . . . Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallogr. D, 66*(Pt 2), 213-221

Adrian, M., Dubochet, J., Lepault, J., & McDowall, A. W. (1984). Cryo-electron microscopy of viruses. *Nature, 308*(5954), 32-36.

Akram, M. (2014). Citric acid cycle and role of its intermediates in metabolism. *Cell Biochem Biophys, 68*(3), 475-478. doi:10.1007/s12013-013-9750-1

Alberts, B. (2015). *Molecular biology of the cell (6th edition)*. 2015.

Allegretti, M., Mills, D. J., McMullan, G., Kuhlbrandt, W., & Vonck, J. (2014). Atomic model of the F420-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector. *eLife, 3*, e01963

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol, 215*(3), 403-410

Anderson, A. C. (2003). The process of structure-based drug design. *Chem Biol, 10*(9), 787-797.

Anesti, V., & Scorrano, L. (2006). The relationship between mitochondrial shape and function and the cytoskeleton. *Biochim Biophys Acta, 1757*(5-6), 692-699

Angerer, H. (2015). Eukaryotic LYR Proteins Interact with Mitochondrial Protein Complexes. *Biology (Basel), 4*(1), 133-150. doi:10.3390/biology4010133

Angerer, H., Radermacher, M., Mankowska, M., Steger, M., Zwicker, K., Heide, H., . . . Zickermann, V. (2014). The LYR protein subunit NB4M/NDUFA6 of mitochondrial complex I anchors an acyl carrier protein and is essential for catalytic activity. *Proc Natl Acad Sci U S A, 111*(14), 5207-5212

Arteni, A. A., Zhang, P., Battchikova, N., Ogawa, T., Aro, E. M., & Boekema, E. J. (2006). Structural characterization of NDH-1 complexes of *Thermosynechococcus elongatus* by single particle electron microscopy. *Biochim. Biophys. Acta, 1757*(11), 1469-1475

Babot, M., & Galkin, A. (2013). Molecular mechanism and physiological role of active-deactive transition of mitochondrial complex I. *Biochem Soc Trans, 41*(5), 1325-1330

Bai, X. C., Rajendra, E., Yang, G., Shi, Y., & Scheres, S. H. (2015). Sampling the conformational space of the catalytic subunit of human gamma-secretase. *eLife, 4*. doi:10.7554/eLife.11182

Barad, B. A., Echols, N., Wang, R. Y.-R., Cheng, Y., DiMaio, F., Adams, P. D., & Fraser, J. S. (2015). Side-chain-directed model and map validation for 3D Electron Cryomicroscopy. *bioRxiv*

Baradaran, R. (2009). *Purification, characterisation and crystallisation of intact complex I from Thermus thermophilus.* (Doctor of Philosphy), University of Cambridge, Cambridge.

Baradaran, R., Berrisford, J. M., Minhas, G. S., & Sazanov, L. A. (2013). Crystal structure of the entire respiratory complex I. *Nature, 494*(7438), 443-448.

Baranova, E. A., Holt, P. J., & Sazanov, L. A. (2007). Projection structure of the membrane domain of Escherichia coli respiratory complex I at 8 A resolution. *J Mol Biol, 366*(1), 140-154

Belevich, G., Knuuti, J., Verkhovsky, M. I., Wikstrom, M., & Verkhovskaya, M. (2011). Probing the mechanistic role of the long alpha-helix in subunit L of respiratory Complex I from Escherichia coli by site-directed mutagenesis. *Molecular microbiology, 82*(5), 1086-1095

Belevich, I., Verkhovsky, M. I., & Wikstrom, M. (2006). Proton-coupled electron transfer drives the proton pump of cytochrome c oxidase. *Nature, 440*(7085), 829-832.

Benit, P., Chretien, D., Kadhom, N., de Lonlay-Debeney, P., Cormier-Daire, V., Cabral, A., . . . Rotig, A. (2001). Large-scale deletion and point mutations of the nuclear NDUFV1 and NDUFS1 genes in mitochondrial complex I deficiency. *Am. J. Hum. Genet., 68*(6), 1344-1352.

Benit, P., Slama, A., Cartault, F., Giurgea, I., Chretien, D., Lebon, S., . . . Rustin, P. (2004). Mutant NDUFS3 subunit of mitochondrial complex I causes Leigh syndrome. *J Med Genet, 41*(1), 14-17.

Berg, J. M., Tymoczko, J. L., & Stryer, L. (2002). *Biochemistry, Fifth Edition*: W. H. Freeman.

Berrisford, J. M., & Sazanov, L. A. (2009). Structural basis for the mechanism of respiratory complex I. *J Biol Chem, 284*(43), 29773-29783

Bertsova, Y. V., & Bogachev, A. V. (2004). The origin of the sodium-dependent NADH oxidation by the respiratory chain of Klebsiella pneumoniae. *FEBS Lett, 563*(1-3), 207-212

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., . . . Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res., 42*(Web Server issue), W252-258

Birrell, J. A., Morina, K., Bridges, H. R., Friedrich, T., & Hirst, J. (2013). Investigating the function of [2Fe-2S] cluster N1a, the off-pathway cluster in complex I, by manipulating its reduction potential. *Biochem J, 456*(1), 139-146

Blakely, E. L., de Silva, R., King, A., Schwarzer, V., Harrower, T., Dawidek, G., . . . Taylor, R. W. (2005). LHON/MELAS overlap syndrome associated with a mitochondrial MTND1 gene mutation. *European journal of human genetics : EJHG, 13*(5), 623-627

Böttcher, B., Scheide, D., Hesterberg, M., Nagel-Steger, L., & Friedrich, T. (2002). A novel, enzymatically active conformation of the Escherichia coli NADH:ubiquinone oxidoreductase (complex I). *J Biol Chem, 277*(20), 17970-17977

Brandt, U. (2006). Energy converting NADH:quinone oxidoreductase (complex I). *Annu Rev Biochem, 75*, 69-92.

Brandt, U. (2011). A two-state stabilization-change mechanism for proton-pumping complex I. *Biochim Biophys Acta, 1807*(10), 1364-1369

Breningstall, G. N., Shoffner, J., & Patterson, R. J. (2008). Siblings with leukoencephalopathy. *Semin Pediatr Neurol, 15*(4), 212-215

Bridges, H. R., Bill, E., & Hirst, J. (2012). Mossbauer spectroscopy on respiratory complex I: the iron-sulfur cluster ensemble in the NADH-reduced enzyme is partially oxidized. *Biochemistry, 51*(1), 149-158

Bridges, H. R., Fearnley, I. M., & Hirst, J. (2010). The subunit composition of mitochondrial NADH:ubiquinone oxidoreductase (complex I) from Pichia pastoris. *Molecular & cellular proteomics : MCP, 9*(10), 2318-2326

Bridges, H. R., Mohammed, K., Harbour, M. E., & Hirst, J. (2017). Subunit NDUFV3 is present in two distinct isoforms in mammalian complex I. *Biochim Biophys Acta, 1858*(3), 197-207

Brockmann, C., Diehl, A., Rehbein, K., Strauss, H., Schmieder, P., Korn, B., . . . Oschkinat, H. (2004). The oxidized subunit B8 from human complex I adopts a thioredoxin fold. *Structure, 12*(9), 1645-1654

Brown, M. D., Starikovskaya, E., Derbeneva, O., Hosseini, S., Allen, J. C., Mikhailovskaya, I. E., . . . Wallace, D. C. (2002). The role of mtDNA background in disease expression: a new primary LHON mutation associated with Western Eurasian haplogroup J. *Hum Genet, 110*(2), 130-138

Brown, M. D., Voljavec, A. S., Lott, M. T., Torroni, A., Yang, C. C., & Wallace, D. C. (1992). Mitochondrial DNA complex I and III mutations associated with Leber's hereditary optic neuropathy. *Genetics, 130*(1), 163-173.

Brown, M. D., Zhadanov, S., Allen, J. C., Hosseini, S., Newman, N. J., Atamonov, V. V., . . . Wallace, D. C. (2001). Novel mtDNA mutations and oxidative phosphorylation dysfunction in Russian LHON families. *Hum Genet, 109*(1), 33-39.

Brzoska, K., Meczynska, S., & Kruszewski, M. (2006). Iron-sulfur cluster proteins: electron transfer and beyond. *Acta Biochim Pol, 53*(4), 685-691.

Bugiani, M., Invernizzi, F., Alberio, S., Briem, E., Lamantea, E., Carrara, F., . . . Zeviani, M. (2004). Clinical and molecular findings in children with complex I deficiency. *Biochim Biophys Acta, 1659*(2-3), 136-147

Byers, D. M., & Gong, H. (2007). Acyl carrier protein: structure-function relationships in a conserved multifunctional protein family. *Biochem Cell Biol, 85*(6), 649-662

Calvo, S. E., Tucker, E. J., Compton, A. G., Kirby, D. M., Crawford, G., Burtt, N. P., . . . Mootha, V. K. (2010). High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nat Genet, 42*(10), 851-858

Campello, S., & Scorrano, L. (2010). Mitochondrial shape changes: orchestrating cell pathophysiology. *EMBO Rep, 11*(9), 678-684

Carroll, J., Ding, S., Fearnley, I. M., & Walker, J. E. (2013). Post-translational modifications near the quinone binding site of mammalian complex I. *J Biol Chem, 288*(34), 24799-24808

Carroll, J., Fearnley, I. M., Shannon, R. J., Hirst, J., & Walker, J. E. (2003). Analysis of the subunit composition of complex I from bovine heart mitochondria. *Molecular & cellular proteomics : MCP, 2*(2), 117-12

Carroll, J., Fearnley, I. M., Skehel, J. M., Runswick, M. J., Shannon, R. J., Hirst, J., & Walker, J. E. (2005). The post-translational modifications of the nuclear encoded subunits of complex I from bovine heart mitochondria. *Molecular & cellular proteomics : MCP, 4*(5), 693-699

Carroll, J., Fearnley, I. M., Skehel, J. M., Shannon, R. J., Hirst, J., & Walker, J. E. (2006). Bovine complex I is a complex of 45 different subunits. *J. Biol. Chem., 281*(43), 32724-32727

Cheng, Y., Grigorieff, N., Penczek, P. A., & Walz, T. (2015). A primer to single-particle cryo-electron microscopy. *Cell, 161*(3), 438-449

Conrad, M., Jakupoglu, C., Moreno, S. G., Lippl, S., Banjac, A., Schneider, M., . . . Brielmeier, M. (2004). Essential role for mitochondrial thioredoxin reductase in hematopoiesis, heart development, and heart function. *Mol Cell Biol, 24*(21), 9414-9423

Crimi, M., Galbiati, S., Moroni, I., Bordoni, A., Perini, M. P., Lamantea, E., . . . Comi, G. P. (2003). A missense mutation in the mitochondrial ND5 gene associated with a Leigh-MELAS overlap syndrome. *Neurology, 60*(11), 1857-1861

Crofts, A. R. (2004). The cytochrome bc1 complex: function in the context of structure. Annu Rev Physiol, 66, 689-733

Danev, R., Tegunov, D., & Baumeister, W. (2017). Using the Volta phase plate with defocus for cryo-EM single particle analysis. *eLife*,6, e23006.

Danhauser, K., Iuso, A., Haack, T. B., Freisinger, P., Brockmann, K., Mayr, J. A., . . . Prokisch, H. (2011). Cellular rescue-assay aids verification of causative DNA-variants in mitochondrial complex I deficiency. *Molecular genetics and metabolism, 103*(2), 161-166

Deschauer, M., Bamberg, C., Claus, D., Zierz, S., Turnbull, D. M., & Taylor, R. W. (2003). Late-onset encephalopathy associated with a C11777A mutation of mitochondrial DNA. *Neurology, 60*(8), 1357-1359

DiMaio, F., Song, Y., Li, X., Brunner, M. J., Xu, C., Conticello, V., . . . Baker, D. (2015). Atomic-accuracy models from 4.5-A cryo-electron microscopy data with density-guided iterative local refinement. *Nature methods, 12*(4), 361-365

Distelmaier, F., Koopman, W. J., van den Heuvel, L. P., Rodenburg, R. J., Mayatepek, E., Willems, P. H., & Smeitink, J. A. (2009). Mitochondrial complex I deficiency: from organelle dysfunction to clinical disease. *Brain : a journal of neurology, 132*(Pt 4), 833-842

Djafarzadeh, R., Kerscher, S., Zwicker, K., Radermacher, M., Lindahl, M., Schagger, H., & Brandt, U. (2000). Biophysical and structural characterization of proton-translocating NADH-dehydrogenase (complex I) from the strictly aerobic yeast *Yarrowia lipolytica*. *Biochim. Biophys. Acta, 1459*(1), 230-238.

Dudkina, N. V., Heinemeyer, J., Sunderhaus, S., Boekema, E. J., & Braun, H. P. (2006). Respiratory chain supercomplexes in the plant mitochondrial membrane. *Trends Plant Sci, 11*(5), 232-240

Dudkina, N. V., Kudryashev, M., Stahlberg, H., & Boekema, E. J. (2011). Interaction of complexes I, III, and IV within the bovine respirasome by single particle cryoelectron tomography. *Proc Natl Acad Sci U S A, 108*(37), 15196-15200

Dyall, S. D., Brown, M. T., & Johnson, P. J. (2004). Ancient invasions: from endosymbionts to organelles. *Science, 304*(5668), 253-257

Efremov, R. G., Baradaran, R., & Sazanov, L. A. (2010). The architecture of respiratory complex I. *Nature, 465*(7297), 441-445

Efremov, R. G., & Sazanov, L. A. (2011). Structure of the membrane domain of respiratory complex I. *Nature, 476*(7361), 414-420

Efremov, R. G., & Sazanov, L. A. (2012). The coupling mechanism of respiratory complex I - A structural and evolutionary perspective. *Biochim Biophys Acta*

Egerton, R. F. (2007). *Physical principles of electron microscopy : an introduction to TEM, SEM, and AEM*. New York: Springer Science+Business Media.

Emelyanov, V. V. (2001). Evolutionary relationship of Rickettsiae and mitochondria. *FEBS Lett, 501*(1), 11-18.

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta crystallogr. D, 60*(Pt 12 Pt 1), 2126-2132

Esposti, D. (1998). Inhibitors of NADH-ubiquinone reductase: an overview. *Biochim Biophys Acta,* 1364(2):222-35

Esterhazy, D., King, M. S., Yakovlev, G., & Hirst, J. (2008). Production of reactive oxygen species by complex I (NADH:ubiquinone oxidoreductase) from *Escherichia coli* and comparison to the enzyme from mitochondria. *Biochemistry, 47*(12), 3964-3971.

Fassone, E., & Rahman, S. (2012). Complex I deficiency: clinical features, biochemistry and molecular genetics. *J Med Genet, 49*(9), 578-590. doi:10.1136/jmedgenet-2012-101159

Fearnley, I. M., & Walker, J. E. (1992). Conservation of sequences of subunits of mitochondrial complex I and their relationships with other proteins. *Biochim. Biophys. Acta, 1140*(2), 105-134.

Fernandez-Moreira, D., Ugalde, C., Smeets, R., Rodenburg, R. J., Lopez-Laso, E., Ruiz-Falco, M. L., . . . Arenas, J. (2007). X-linked NDUFA1 gene mutations associated with mitochondrial encephalomyopathy. *Ann Neurol, 61*(1), 73-83

Ferreira, M., Torraco, A., Rizza, T., Fattori, F., Meschini, M. C., Castana, C., . . . Bertini, E. (2011). Progressive cavitating leukoencephalopathy associated with respiratory chain complex I deficiency and a novel mutation in NDUFS1. *Neurogenetics, 12*(1), 9-17

Fiedorczuk, K., Letts, J. A., Degliesposti, G., Kaszuba, K., Skehel, M., & Sazanov, L. A. (2016). Atomic structure of the entire mammalian mitochondrial complex I. *Nature, 538*(7625), 406-410

Finel, M. (1998). Does NADH play a central role in energy metabolism in Helicobacter pylori? *Trends Biochem. Sci., 23*(11), 412-413.

Finel, M., Majander, A. S., Tyynela, J., De Jong, A. M., Albracht, S. P., & Wikstrom, M. (1994). Isolation and characterisation of subcomplexes of the mitochondrial NADH:ubiquinone oxidoreductase (complex I). *Eur. J. Biochem., 226*(1), 237-242

Flotenmeyer, M., Weiss, H., Tribet, C., Popot, J. L., & Leonard, K. (2007). The use of amphipathic polymers for cryo electron microscopy of NADH:ubiquinone oxidoreductase (complex I). *J Microsc, 227*(Pt 3), 229-235

Friedman, J. R., & Nunnari, J. (2014). Mitochondrial form and function. *Nature, 505*(7483), 335-343

Friedrich, T., & Scheide, D. (2000). The respiratory complex I of bacteria, archaea and eukarya and its module common with membrane-bound multisubunit hydrogenases. *FEBS Lett., 479*(1-2), 1-5

Friedrich, T., Steinmuller, K., & Weiss, H. (1995). The proton-pumping respiratory complex I of bacteria and mitochondria and its homologue in chloroplasts. *FEBS Lett., 367*(2), 107-111

Fujimoto, M., & Hayashi, T. (2011). New insights into the role of mitochondria-associated endoplasmic reticulum membrane. *Int Rev Cell Mol Biol, 292*, 73-117

Funalot, B., Reynier, P., Vighetto, A., Ranoux, D., Bonnefont, J. P., Godinot, C., . . . Mas, J. L. (2002). Leigh-like encephalopathy complicating Leber's hereditary optic neuropathy. *Ann Neurol, 52*(3), 374-377

Gabaldon, T., Rainey, D., & Huynen, M. A. (2005). Tracing the evolution of a large protein complex in the eukaryotes, NADH:ubiquinone oxidoreductase (Complex I). *J. Mol. Biol., 348*(4), 857-870

Galkin, A., Dröse, S., & Brandt, U. (2006). The proton pumping stoichiometry of purified mitochondrial complex I reconstituted into proteoliposomes. *Biochim Biophys Acta, 1757*(12), 1575-1581

Galkin, A., Meyer, B., Wittig, I., Karas, M., Schagger, H., Vinogradov, A., & Brandt, U. (2008). 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*(30), 20907-20913

Gemperli, A. C., Dimroth, P., & Steuber, J. (2002). The respiratory complex I (NDH I) from Klebsiella pneumoniae, a sodium pump. *J Biol Chem, 277*(37), 33811-33817

Ghayee, H. K., & Auchus, R. J. (2007). Basic concepts and recent developments in human steroid hormone biosynthesis. *Rev Endocr Metab Disord, 8*(4), 289-300

Glaeser, R. M., Han, B. G., Csencsits, R., Killilea, A., Pulk, A., & Cate, J. H. (2016). Factors that Influence the Formation and Stability of Thin, Cryo - EM Specimens. *Biophysical Journal, 110*(4), 749-755

Gorman, G. S., Chinnery, P. F., DiMauro, S., Hirano, M., Koga, Y., McFarland, R., . . . Turnbull, D. M. (2016). Mitochondrial diseases. *Nat Rev Dis Primers, 2*, 16080

Gray, M. W., Burger, G., & Lang, B. F. (1999). Mitochondrial evolution. *Science, 283*(5407)

Grgic, L., Zwicker, K., Kashani-Poor, N., Kerscher, S., & Brandt, U. (2004). Functional significance of conserved histidines and arginines in the 49-kDa subunit of mitochondrial complex I. *J. Biol. Chem., 279*(20), 21193-21199

Grigorieff, N. (1998). Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 A in ice. *J. Mol. Biol., 277*(5), 1033-1046.

Grigorieff, N., & Harrison, S. C. (2011). Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. *Curr Opin Struct Biol, 21*(2), 265-273

Gu, J., Wu, M., Guo, R., Yan, K., Lei, J., Gao, N., & Yang, M. (2016). The architecture of the mammalian respirasome. *Nature, 537*(7622), 639-643

Guenebaut, V., Schlitt, A., Weiss, H., Leonard, K., & Friedrich, T. (1998). Consistent structure between bacterial and mitochondrial NADH:ubiquinone oxidoreductase (complex I). *J. Mol. Biol., 276*(1), 105-112

Guenebaut, V., Vincentelli, R., Mills, D., Weiss, H., & Leonard, K. R. (1997). Three-dimensional structure of NADH-dehydrogenase from *Neurospora crassa* by electron microscopy and conical tilt reconstruction. *J. Mol. Biol., 265*(4), 409-418

Guerrero-Castillo, S., Baertling, F., Kownatzki, D., Wessels, H. J., Arnold, S., Brandt, U., & Nijtmans, L. (2017). The Assembly Pathway of Mitochondrial Respiratory Chain Complex I. *Cell metab., 25*(1), 128-139

Gutierrez Cortes, N., Pertuiset, C., Dumon, E., Borlin, M., Hebert-Chatelain, E., Pierron, D., . . . Rocher, C. (2012). Novel mitochondrial DNA mutations responsible for maternally inherited nonsyndromic hearing loss. *Hum Mutat, 33*(4), 681-689

Haack, T. B., Haberberger, B., Frisch, E. M., Wieland, T., Iuso, A., Gorza, M., . . . Prokisch, H. (2012). Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing. J Med Genet, 49(4), 277-283

Haack, T. B., Madignier, F., Herzer, M., Lamantea, E., Danhauser, K., Invernizzi, F., . . . Prokisch, H. (2012). Mutation screening of 75 candidate genes in 152 complex I deficiency cases identifies pathogenic variants in 16 genes including NDUFB9. J Med Genet, 49(2), 83-89

Hagn, F., Etzkorn, M., Raschle, T., & Wagner, G. (2013). Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins. Journal of the American Chemical Society, 135(5), 1919–1925

Haider, M., Rose, H., Uhlemann, S., Kabius, B., & Urban, K. (1998). Towards 0.1 nm resolution with the first spherically corrected transmission electron microscope. Journal of Electron Microscopy, 47(5), 395-405

Hatefi, Y., Haavik, A. G., Fowler, L. R., & Griffiths, D. E. (1962). Studies on the electron transfer system. XLII. Reconstitution of the electron transfer system. *J Biol Chem, 237*, 2661-2669

Hatefi, Y., Haavik, A. G., & Griffiths, D. E. (1962). Studies on the electron transfer system. XL. Preparation and properties of mitochondrial DPNH-coenzyme Q reductase. *J Biol Chem, 237*, 1676-1680

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E., & Downing, K. H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J Mol Biol, 213*(4), 899-929

Herrmann, J. M., & Riemer, J. (2010). The intermembrane space of mitochondria. *Antioxid Redox Signal, 13*(9), 1341-1358

Hinttala, R., Smeets, R., Moilanen, J. S., Ugalde, C., Uusimaa, J., Smeitink, J. A., & Majamaa, K. (2006). Analysis of mitochondrial DNA sequences in patients with isolated or combined oxidative phosphorylation system deficiency. *J Med Genet, 43*(11), 881-886

Hirst, J. (2013). Mitochondrial complex I. *Annu Rev Biochem, 82*, 551-575

Hirst, J., Carroll, J., Fearnley, I. M., Shannon, R. J., & Walker, J. E. (2003). The nuclear encoded subunits of complex I from bovine heart mitochondria. *Biochim Biophys Acta, 1604*(3), 135-150

Hirst, J., & Roessler, M. M. (2015). Energy conversion, redox catalysis and generation of reactive oxygen species by respiratory complex I. *Biochim Biophys Acta*

Hoefs, S. J., Skjeldal, O. H., Rodenburg, R. J., Nedregaard, B., van Kaauwen, E. P., Spiekerkotter, U., . . . van den Heuvel, L. P. (2010). Novel mutations in the NDUFS1 gene cause low residual activities in human complex I deficiencies. *Molecular genetics and metabolism, 100*(3), 251-256

Hoefs, S. J., van Spronsen, F. J., Lenssen, E. W., Nijtmans, L. G., Rodenburg, R. J., Smeitink, J. A., & van den Heuvel, L. P. (2011). NDUFA10 mutations cause complex I deficiency in a patient with Leigh disease. *European journal of human genetics : EJHG, 19*(3), 270-274

Hofhaus, G., Weiss, H., & Leonard, K. (1991). Electron microscopic analysis of the peripheral and membrane parts of mitochondrial NADH dehydrogenase (complex I). *J. Mol. Biol., 221*(3), 1027-1043

Hoppe, W., Schramm, H. J., Sturm, M., Hunsmann, N., & Gassmann, J. (1976). 3-Dimensional Electron-Microscopy of Individual Biological Objects .3. Experimental Results on Yeast Fatty-Acid Synthetase. *Zeitschrift Fur Naturforschung Section a-a Journal of Physical Sciences, 31*(11), 1380-1390

Horder, T. J. (2006). Gavin Rylands de Beer: how embryology foreshadowed the dilemmas of the genome. *Nat Rev Genet, 7*(11), 892-898

Howell, N., Bindoff, L. A., McCullough, D. A., Kubacka, I., Poulton, J., Mackey, D., . . . Turnbull, D. M. (1991). Leber hereditary optic neuropathy: identification of the same mitochondrial ND1 mutation in six pedigrees. *Am J Hum Genet, 49*(5), 939-950

Howell, N., Halvorson, S., Burns, J., McCullough, D. A., & Paulton, J. (1993). When does bilateral optic atrophy become Leber hereditary optic neuropathy? *Am J Hum Genet, 53*(4), 959-963

Hutchin, T., & Cortopassi, G. (1995). A mitochondrial DNA clone is associated with increased risk for Alzheimer disease. *Proc Natl Acad Sci U S A, 92*(15), 6892-6895

Johansson, K., Ramaswamy, S., Ljungcrantz, C., Knecht, W., Piskur, J., Munch-Petersen, B., . . . Eklund, H. (2001). Structural basis for substrate specificities of cellular deoxyribonucleoside kinases. *Nat Struct Biol, 8*(7), 616-620

Jun, A. S., Brown, M. D., & Wallace, D. C. (1994). A mitochondrial DNA mutation at nucleotide pair 14459 of the NADH dehydrogenase subunit 6 gene associated with maternally inherited Leber hereditary optic neuropathy and dystonia. *Proc Natl Acad Sci U S A, 91*(13), 6206-6210

Kao, M. C., Di Bernardo, S., Nakamaru-Ogiso, E., Miyoshi, H., Matsuno-Yagi, A., & Yagi, T. (2005). Characterization of the membrane domain subunit NuoJ (ND6) of the NADH-quinone oxidoreductase from *Escherichia coli* by chromosomal DNA manipulation. *Biochemistry, 44*(9), 3562-3571

Karnkowska, A., Vacek, V., Zubacova, Z., Treitli, S. C., Petrzelkova, R., Eme, L., . . . Hampl, V. (2016). A Eukaryote without a Mitochondrial Organelle. *Curr Biol, 26*(10), 1274-1284

Karp, G. (2013). *Cell and molecular biology : concepts and experiments*. Hoboken, NJ: John Wiley

Kazak, L., Reyes, A., & Holt, I. J. (2012). Minimizing the damage: repair pathways keep mitochondrial DNA intact. *Nat Rev Mol Cell Biol, 13*(10), 659-671

Keeney, P. M., Xie, J., Capaldi, R. A., & Bennett, J. P., Jr. (2006). Parkinson's disease brain mitochondrial complex I has oxidatively damaged subunits and is functionally impaired and misassembled. *J Neurosci, 26*(19), 5256-5264.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc, 10*(6), 845-858

Kim, M. R., Chang, H. S., Kim, B. H., Kim, S., Baek, S. H., Kim, J. H., . . . Kim, J. R. (2003). Involvements of mitochondrial thioredoxin reductase (TrxR2) in cell proliferation. *Biochem Biophys Res Commun, 304*(1), 119-124

Kirby, D. M., Crawford, M., Cleary, M. A., Dahl, H. H., Dennett, X., & Thorburn, D. R. (1999). Respiratory chain complex I deficiency: an underdiagnosed energy generation disorder. *Neurology, 52*(6), 1255-1264

Kirby, D. M., McFarland, R., Ohtake, A., Dunning, C., Ryan, M. T., Wilson, C., . . . Taylor, R. W. (2004). Mutations of the mitochondrial ND1 gene as a cause of MELAS. *J Med Genet, 41*(10), 784-789

Kimanius, D., Forsberg, B., Scheres, S., Lindahl, E. (2016) Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. bioRxiv 059717

Kmita, K., Wirth, C., Warnau, J., Guerrero-Castillo, S., Hunte, C., Hummer, G., . . . Zickermann, V. (2015). Accessory NUMM (NDUFS6) subunit harbors a Zn-binding site and is essential for biogenesis of mitochondrial complex I. *Proceedings of the National Academy of Sciences of the United States of America, 112*(18), 5685-5690

Koene, S., Rodenburg, R. J., van der Knaap, M. S., Willemsen, M. A., Sperl, W., Laugel, V., . . . Smeitink, J. A. (2012). Natural disease course and genotype-phenotype correlations in Complex I deficiency caused by nuclear gene defects: what we learned from 130 cases. *J Inherit Metab Dis, 35*(5), 737-747

Krivov, G. G., Shapovalov, M. V., & Dunbrack, R. L., Jr. (2009). Improved prediction of protein side-chain conformations with SCWRL4. *Proteins, 77*(4), 778-795

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol., 305*(3), 567-580

Kuhlbrandt, W. (2014). Cryo - EM enters a new era. *eLife, 3*, e03678

Kuhlbrandt, W. (2015). Structure and function of mitochondrial membrane protein complexes. *BMC Biol, 13*, 89

Lazarou, M., McKenzie, M., Ohtake, A., Thorburn, D. R., & Ryan, M. T. (2007). Analysis of the assembly profiles for mitochondrial- and nuclear-DNA-encoded subunits into complex I. *Mol Cell Biol, 27*(12), 4228-4237

Lebon, S., Rodriguez, D., Bridoux, D., Zerrad, A., Rotig, A., Munnich, A., . . . Slama, A. (2007). A novel mutation in the human complex I NDUFS7 subunit associated with Leigh syndrome. *Molecular genetics and metabolism, 90*(4), 379-382

Leo-Kottler, B., Luberichs, J., Besch, D., Christ-Adler, M., & Fauser, S. (2002). Leber's hereditary optic neuropathy: clinical and molecular genetic results in a patient with a point mutation at np T11253C (isoleucine to threonine) in the ND4 gene and spontaneous recovery. *Graefes Arch Clin Exp Ophthalmol, 240*(9), 758-764

Leonard, K., Haiker, H., & Weiss, H. (1987). Three-dimensional structure of NADH: ubiquinone reductase (complex I) from *Neurospora* mitochondria

determined by electron microscopy of membrane crystals. *J. Mol. Biol., 194*(2), 277-286

Leshinsky-Silver, E., Lebre, A. S., Minai, L., Saada, A., Steffann, J., Cohen, S., . . . Lerman-Sagie, T. (2009). NDUFS4 mutations cause Leigh syndrome with predominant brainstem involvement. *Molecular genetics and metabolism, 97*(3), 185-189

Leshinsky-Silver, E., Lev, D., Malinger, G., Shapira, D., Cohen, S., Lerman-Sagie, T., & Saada, A. (2010). Leigh disease presenting in utero due to a novel missense mutation in the mitochondrial DNA-ND3. *Molecular genetics and metabolism, 100*(1), 65-70

Letts, J. A., Degliesposti, G., Fiedorczuk, K., Skehel, M., & Sazanov, L. A. (2016). Purification of Ovine Respiratory Complex I Results in a Highly Active and Stable Preparation. *J Biol Chem, 291*(47), 24657-24675

Letts, J. A., Fiedorczuk, K., & Sazanov, L. A. (2016). The architecture of respiratory supercomplexes. *Nature, 537*(7622), 644-648

Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., . . . Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo - EM. *Nature methods, 10*(6), 584-590

Lieber, D. S., Calvo, S. E., Shanahan, K., Slate, N. G., Liu, S., Hershman, S. G., . . . Mootha, V. K. (2013). Targeted exome sequencing of suspected mitochondrial disorders. *Neurology, 80*(19), 1762-1770

Lill, R., Dutkiewicz, R., Freibert, S. A., Heidenreich, T., Mascarenhas, J., Netz, D. J., . . . Muhlenhoff, U. (2015). The role of mitochondria and the CIA machinery in the maturation of cytosolic and nuclear iron-sulfur proteins. *Eur J Cell Biol, 94*(7-9), 280-291

Liolitsa, D., Rahman, S., Benton, S., Carr, L. J., & Hanna, M. G. (2003). Is the mitochondrial complex I ND5 gene a hot-spot for MELAS causing mutations? *Ann Neurol, 53*(1), 128-132

Lodish, H. F. (2008). *Molecular cell biology*.

Loeffen, J., Elpeleg, O., Smeitink, J., Smeets, R., Stockler-Ipsiroglu, S., Mandel, H., . . . van den Heuvel, L. (2001). Mutations in the complex I NDUFS2 gene of patients with cardiomyopathy and encephalomyopathy. *Ann Neurol, 49*(2), 195-201

Loeffen, J., Smeitink, J., Triepels, R., Smeets, R., Schuelke, M., Sengers, R., . . . van den Heuvel, L. (1998). The first nuclear-encoded complex I mutation in a patient with Leigh syndrome. *Am J Hum Genet, 63*(6), 1598-1608

Logan, D. C. (2006). The mitochondrial compartment. *J Exp Bot, 57*(6), 1225-1243

Lu, H., & Cao, X. (2008). GRIM-19 is essential for maintenance of mitochondrial membrane potential. *Mol Biol Cell, 19*(5), 1893-1902

Macmillan, C., Johns, T. A., Fu, K., & Shoubridge, E. A. (2000). Predominance of the T14484C mutation in French-Canadian families with Leber hereditary optic neuropathy is due to a founder effect. *Am J Hum Genet, 66*(1), 332-335

Maklashina, E., Kotlyar, A. B., & Cecchini, G. (2003). Active/de-active transition of respiratory complex I in bacteria, fungi, and animals. *Biochim Biophys Acta, 1606*(1-3), 95-103

Mannella, C. A. (2006). Structure and dynamics of the mitochondrial inner membrane cristae. *Biochim Biophys Acta, 1763*(5-6), 542-548

Marabini, R., Masegosa, I. M., San Martin, M. C., Marco, S., Fernandez, J. J., de la Fraga, L. G., . . . Carazo, J. M. (1996). Xmipp: An Image Processing Package for Electron Microscopy. *J Struct Biol, 116*(1), 237-240

Maret, W. (2005). Zinc coordination environments in proteins determine zinc functions. *J Trace Elem Med Biol, 19*(1), 7-12

Margulis, L. (1970). *Origin of Eukaryotic Cells*. New Haven: Yale University Press.

Marin, S. E., Mesterman, R., Robinson, B., Rodenburg, R. J., Smeitink, J., & Tarnopolsky, M. A. (2013). Leigh syndrome associated with mitochondrial complex I deficiency due to novel mutations In NDUFV1 and NDUFS2. *Gene, 516*(1), 162-167

Marina, A. D., Schara, U., Pyle, A., Moller-Hartmann, C., Holinski-Feder, E., Abicht, A., . . . Horvath, R. (2013). NDUFS8-related Complex I Deficiency Extends Phenotype from "PEO Plus" to Leigh Syndrome. *JIMD Rep, 10*, 17-22

Martin, M. A., Blazquez, A., Gutierrez-Solana, L. G., Fernandez-Moreira, D., Briones, P., Andreu, A. L., . . . Arenas, J. (2005). Leigh syndrome associated with mitochondrial complex I deficiency due to a novel mutation in the NDUFS1 gene. *Arch. Neurol., 62*(4), 659-661

Maximo, V., Botelho, T., Capela, J., Soares, P., Lima, J., Taveira, A., . . . Sobrinho-Simoes, M. (2005). Somatic and germline mutation in GRIM-19, a dual function gene involved in mitochondrial metabolism and cell death, is linked to mitochondrion-rich (Hurthle cell) tumours of the thyroid. *Br J Cancer, 92*(10), 1892-1898

Mayorov, V., Biousse, V., Newman, N. J., & Brown, M. D. (2005). The role of the ND5 gene in LHON: characterization of a new, heteroplasmic LHON mutation. *Ann Neurol, 58*(5), 807-811

McFarland, R., Kirby, D. M., Fowler, K. J., Ohtake, A., Ryan, M. T., Amor, D. J., . . . Thorburn, D. R. (2004). De novo mutations in the mitochondrial ND3 gene as a cause of infantile mitochondrial encephalopathy and complex I deficiency. *Ann Neurol, 55*(1), 58-64

McFarland, R., Tuppen, H., He, L., Blakely, E. L., Morris, A. A. M., Clarke, M., . . . Taylor, R. W. (2009). Recurrent mutations in the NDUFS2 gene causing isolated complex I deficiency in skeletal muscle. *Neuromuscular Disorders, 19*(8-9), 562-562

McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics, 16*(4), 404-405

McKenzie, M., & Ryan, M. T. (2010). Assembly factors of human mitochondrial complex I and their defects in disease. *IUBMB Life, 62*(7), 497-502

Meents, A., Gutmann, S., Wagner, A., & Schulze-Briese, C. (2010). Origin and temperature dependence of radiation damage in biological samples at cryogenic temperatures. *Proc Natl Acad Sci U S A, 107*(3), 1094-1099

Meyer, E. H. (2012). Proteomic investigations of complex I composition: how to define a subunit? *Front Plant Sci, 3*, 106

Mi, H. L., Endo, T., Schreiber, U., & Asada, K. (1992). Donation of Electrons from Cytosolic Components to the Intersystem Chain in the Cyanobacterium Synechococcus Sp Pcc-7002 as Determined by the Reduction of P700+. *Plant and Cell Physiology, 33*(8), 1099-1105

Michel, J., Deleon-Rangel, J., Zhu, S., Van Ree, K., & Vik, S. B. (2011). Mutagenesis of the L, M, and N Subunits of Complex I from Escherichia coli Indicates a Common Role in Function. *PLoS One, 6*(2), e17420

Milazzo, A. C., Leblanc, P., Duttweiler, F., Jin, L., Bouwer, J. C., Peltier, S., . . . Xuong, N. H. (2005). Active pixel sensor array as a detector for electron microscopy. *Ultramicroscopy, 104*(2), 152-159

Mishra, P., & Chan, D. C. (2014). Mitochondrial dynamics and inheritance during cell division, development and disease. *Nat Rev Mol Cell Biol, 15*(10), 634-646

Mitchell, P. (1975). The protonmotive Q cycle: a general formulation. *FEBS Lett, 59*(2), 137-139

Mitchell, P. (1979). Keilin's respiratory chain concept and its chemiosmotic consequences. *Science, 206*(4423), 1148-1159

Moparthi, V. K., & Hagerhall, C. (2011). The evolution of respiratory chain complex I from a smaller last common ancestor consisting of 11 protein subunits. *Journal of molecular evolution, 72*(5-6), 484-497

Morgan, D. J., & Sazanov, L. A. (2008). Three-dimensional structure of respiratory complex I from Escherichia coli in ice in the presence of nucleotides. *Biochim Biophys Acta, 1777*(7-8), 711-718

Moritz, K. M., Lim, G. B., & Wintour, E. M. (1997). Developmental regulation of erythropoietin and erythropoiesis. *Am J Physiol, 273*(6 Pt 2), R1829-1844

Moser, C. C., Farid, T. A., Chobot, S. E., & Dutton, P. L. (2006). Electron tunneling chains of mitochondria. *Biochim Biophys Acta, 1757*(9-10), 1096-1109

Moslemi, A. R., Darin, N., Tulinius, M., Wiklund, L. M., Holme, E., & Oldfors, A. (2008). Progressive encephalopathy and complex I deficiency associated with mutations in MTND1. *Neuropediatrics, 39*(1), 24-28

Murcha, M. W., Wang, Y., & Whelan, J. (2012). A molecular link between mitochondrial preprotein transporters and respiratory chain complexes. *Plant Signal Behav, 7*(12), 1594-1597

Murphy, M. P. (2009). How mitochondria produce reactive oxygen species. *Biochem. J., 417*(1), 1-13

Naini, A. B., Lu, J., Kaufmann, P., Bernstein, R. A., Mancuso, M., Bonilla, E., . . . DiMauro, S. (2005). Novel mitochondrial DNA ND5 mutation in a patient with clinical features of MELAS and MERRF. *Arch Neurol, 62*(3), 473-476

Nakamaru-Ogiso, E., Kao, M.-C., Chen, H., Sinha, S. C., Yagi, T., & Ohnishi, T. (2010). The membrane subunit NuoL(ND5) is involved in the indirect proton pumping mechanism of E. coli complex I. *J Biol Chem*

Nicholls, D. G., & Ferguson, S. J. (2002). Bioenergetics 3.

Nishioka, K., Vilarino-Guell, C., Cobb, S. A., Kachergus, J. M., Ross, O. A., Hentati, E., . . . Farrer, M. J. (2010). Genetic variation of the mitochondrial complex I subunit NDUFV2 and Parkinson's disease. *Parkinsonism Relat Disord, 16*(10), 686-687

Obayashi, T., Hattori, K., Sugiyama, S., Tanaka, M., Tanaka, T., Itoyama, S., . . . et al. (1992). Point mutations in mitochondrial DNA in patients with hypertrophic cardiomyopathy. *Am Heart J, 124*(5), 1263-1269

Ohnishi, S. T., Salerno, J. C., & Ohnishi, T. (2010). Possible roles of two quinone molecules in direct and indirect proton pumps of bovine heart NADH-quinone oxidoreductase (complex I). *Biochim Biophys Acta, 1797*(12), 1891-1893

Ohnishi, T. (1998). Iron-sulfur clusters/semiquinones in complex I. *Biochim Biophys Acta, 1364*(2), 186-206

Okun, J. G., Zickermann, V., Zwicker, K., Schagger, H., & Brandt, U. (2000). Binding of detergents and inhibitors to bovine complex I - a novel purification procedure for bovine complex I retaining full inhibitor sensitivity. *Biochim. Biophys. Acta, 1459*(1), 77-87

Oteiza, P. I. (2012). Zinc and the modulation of redox homeostasis. *Free Radic Biol Med, 53*(9), 1748-1759

Page, C. C., Moser, C. C., Chen, X., & Dutton, P. L. (1999). Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Nature, 402*(6757), 47-52

Pagniez-Mammeri, H., Lombes, A., Brivet, M., Ogier-de Baulny, H., Landrieu, P., Legrand, A., & Slama, A. (2009). Rapid screening for nuclear genes mutations in isolated respiratory chain complex I defects. *Molecular genetics and metabolism, 96*(4), 196-200

Paradies, G., Paradies, V., De Benedictis, V., Ruggiero, F. M., & Petrosillo, G. (2014). Functional role of cardiolipin in mitochondrial bioenergetics. *Biochim Biophys Acta, 1837*(4), 408-417

Parikh, S., Saneto, R., Falk, M. J., Anselm, I., Cohen, B. H., Haas, R., & Medicine Society, T. M. (2009). A modern approach to the treatment of mitochondrial disease. *Curr Treat Options Neurol, 11*(6), 414-430

Peng, G., Fritzsch, G., Zickermann, V., Schagger, H., Mentele, R., Lottspeich, F., . . . Michel, H. (2003). Isolation, characterization and electron microscopic single particle analysis of the NADH:ubiquinone oxidoreductase (complex I) from the hyperthermophilic eubacterium *Aquifex aeolicus. Biochemistry, 42*(10), 3032-3039

Peng, L., Yamamoto, H., & Shikanai, T. (2010). Structure and biogenesis of the chloroplast NAD(P)H dehydrogenase complex. *Biochim Biophys Acta*

Perkins, G. A., & Frey, T. G. (2000). Recent structural insight into mitochondria gained by microscopy. *Micron, 31*(1), 97-111

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry, 25*(13), 1605-1612

Planas-Iglesias, J., Dwarakanath, H., Mohammadyani, D., Yanamala, N., Kagan, V. E., & Klein-Seetharaman, J. (2015). Cardiolipin Interactions with Proteins. *Biophysical Journal, 109*(6), 1282-1294

Polyak, K., Li, Y., Zhu, H., Lengauer, C., Willson, J. K., Markowitz, S. D., . . . Vogelstein, B. (1998). Somatic mutations of the mitochondrial genome in human colorectal tumours. *Nat Genet, 20*(3), 291-293

Porterfield, J. Z., & Zlotnick, A. (2010). A simple and general method for determining the protein and nucleic acid content of viruses by UV absorbance. *Virology, 407*(2), 281-288

Potluri, P., Davila, A., Ruiz-Pesini, E., Mishmar, D., O'Hearn, S., Hancock, S., . . . Procaccio, V. (2009). A novel NDUFA1 mutation leads to a progressive mitochondrial complex I-specific neurodegenerative disease. *Molecular genetics and metabolism, 96*(4), 189-195

Procaccio, V., & Wallace, D. C. (2004). Late-onset Leigh syndrome in a patient with mitochondrial complex I NDUFS8 mutations. *Neurology, 62*(10), 1899-1901

Radermacher, M., Wagenknecht, T., Verschoor, A., & Frank, J. (1987). Three-dimensional reconstruction from a single-exposure, random conical tilt

series applied to the 50S ribosomal subunit of Escherichia coli. *J Microsc, 146*(Pt 2), 113-136

Ragan, C. I., & Racker, E. (1973). Resolution and reconstitution of the mitochondrial electron transport system. IV. The reconstitution of rotenone-sensitive reduced nicotinamide adenine dinucleotide-ubiquinone reductase from reduced nicotinamide adenine dinucleotide dehydrogenase and phospholipids. *J Biol Chem, 248*(19), 6876-6884

Ravn, K., Wibrand, F., Hansen, F. J., Horn, N., Rosenberg, T., & Schwartz, M. (2001). An mtDNA mutation, 14453G-->A, in the NADH dehydrogenase subunit 6 associated with severe MELAS syndrome. *European journal of human genetics : EJHG, 9*(10), 805-809

Rhein, V. F., Carroll, J., Ding, S., Fearnley, I. M., & Walker, J. E. (2013). NDUFAF7 methylates arginine 85 in the NDUFS2 subunit of human complex I. *J Biol Chem, 288*(46), 33016-33026

Rizzuto, R., De Stefani, D., Raffaello, A., & Mammucari, C. (2012). Mitochondria as sensors and regulators of calcium signalling. *Nat Rev Mol Cell Biol, 13*(9), 566-578

Robinson, J. M. (2008). Reactive oxygen species in phagocytic leukocytes. *Histochem Cell Biol, 130*(2), 281-297

Rodenburg, R. J. (2016). Mitochondrial complex I-linked disease. *Biochim Biophys Acta, 1857*(7), 938-945

Rohou, A., & Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol, 192*(2), 216-221

Rosenthal, P. B., & Henderson, R. (2003). Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *J. Mol. Biol., 333*(4), 721-745

Rossmann, M. G., Moras, D., & Olsen, K. W. (1974). Chemical and biological evolution of nucleotide-binding protein. *Nature, 250*(463), 194-199

Rost, B., Yachdav, G., & Liu, J. (2004). The PredictProtein server. *Nucleic Acids Res., 32*(Web Server issue), W321-326

Rouault, T. A. (2015). Mammalian iron-sulphur proteins: novel insights into biogenesis and function. *Nat Rev Mol Cell Biol, 16*(1), 45-55

Ruska, E. (1987). The Development of the Electron-Microscope and of Electron-Microscopy. *Bioscience Reports, 7*(8), 607-629

Russo CJ, Passmore LA. (2014) Ultrastable gold substrates for electron cryomicroscopy. Science. 346(6215), 1377-1380

Sanchez-Caballero, L., Guerrero-Castillo, S., & Nijtmans, L. (2016). Unraveling the complexity of mitochondrial complex I assembly: A dynamic process. *Biochim Biophys Acta, 1857*(7), 980-990

Sanz, A., & Stefanatos, R. K. (2008). The mitochondrial free radical theory of aging: a critical view. *Curr Aging Sci, 1*(1), 10-21

Saraste, M. (1999). Oxidative phosphorylation at the fin de siecle. *Science, 283*(5407), 1488-1493

Sarewicz, M., & Osyczka, A. (2015). Electronic connection between the quinone and cytochrome C redox pools and its role in regulation of mitochondrial electron transport and redox signaling. *Physiol Rev, 95*(1), 219-243

Sarzi, E., Brown, M. D., Lebon, S., Chretien, D., Munnich, A., Rotig, A., & Procaccio, V. (2007). A novel recurrent mitochondrial DNA mutation in ND3 gene is

associated with isolated complex I deficiency causing Leigh syndrome and dystonia. *Am J Med Genet A, 143a*(1), 33-41

Sazanov, L. A. (2007). Respiratory complex I: mechanistic and structural insights provided by the crystal structure of the hydrophilic domain. *Biochemistry-Us, 46*(9), 2275-2288

Sazanov, L. A. (2015). A giant molecular proton pump: structure and mechanism of respiratory complex I. *Nat Rev Mol Cell Biol, 16*(6), 375-388

Sazanov, L. A., & Hinchliffe, P. (2006). Structure of the hydrophilic domain of respiratory complex I from Thermus thermophilus. *Science, 311*(5766), 1430-1436

Sazanov, L. A., Peak-Chew, S. Y., Fearnley, I. M., & Walker, J. E. (2000). Resolution of the membrane domain of bovine complex I into subcomplexes: implications for the structural organization of the enzyme. *Biochemistry-Us, 39*(24), 7229-7235

Sazanov, L. A., Peak-Chew, S. Y., Fearnley, I. M., & Walker, J. E. (2000). Resolution of the membrane domain of bovine complex I into subcomplexes: implications for the structural organization of the enzyme. *Biochemistry, 39*(24), 7229-7235

Sazanov, L. A., & Walker, J. E. (2000). Cryo-electron crystallography of two sub-complexes of bovine complex I reveals the relationship between the membrane and peripheral arms. *J Mol Biol, 302*(2), 455-464

Schagger, H., & Pfeiffer, K. (2001). The ratio of oxidative phosphorylation complexes I-V in bovine heart mitochondria and the composition of respiratory chain supercomplexes. J Biol Chem, 276(41), 37861-37867

Scheidelaar S, Koorengevel MC, Pardo JD, Meeldijk JD, Breukink E, Killian JA. Molecular Model for the Solubilization of Membranes into Nanodisks by Styrene Maleic Acid Copolymers. Biophysical Journal. 2015;108(2):279-290. doi:10.1016/j.bpj.2014.11.3464.

Scheres, S. H. (2012). RELION: implementation of a Bayesian approach to cryo - EM structure determination. J Struct Biol, 180(3), 519-530

Scheres, S. H. (2014). Beam-induced motion correction for sub-megadalton cryo - EM particles. *eLife, 3*, e03665. doi:10.7554/eLife.03665

Scheres, S. H., & Chen, S. (2012). Prevention of overfitting in cryo - EM structure determination. *Nature methods, 9*(9), 853-854

Schuelke, M., Detjen, A., van den Heuvel, L., Korenke, C., Janssen, A., Smits, A., . . . Smeitink, J. (2002). New nuclear encoded mitochondrial mutation illustrates pitfalls in prenatal diagnosis by biochemical methods. *Clin Chem, 48*(5), 772-775

Schuelke, M., Smeitink, J., Mariman, E., Loeffen, J., Plecko, B., Trijbels, F., . . . van den Heuvel, L. (1999). Mutant NDUFV1 subunit of mitochondrial complex I causes leukodystrophy and myoclonic epilepsy. *Nat Genet, 21*(3), 260-261

Sharma, L. K., Lu, J., & Bai, Y. (2009). Mitochondrial respiratory complex I: structure, function and implication in human diseases. *Curr Med Chem, 16*(10), 1266-1277

Sharpley, M. S., Hirst J. (2006). The inhibition of mitochondrial complex I (NADH:ubiquinone oxidoreductase) by Zn2+. *J Biol Chem*. 281(46), 34803-9

Shinzawa-Itoh, K., Seiyama, J., Terada, H., Nakatsubo, R., Naoki, K., Nakashima, Y., & Yoshikawa, S. (2010). Bovine heart NADH-ubiquinone oxidoreductase

contains one molecule of ubiquinone with ten isoprene units as one of the cofactors. *Biochemistry, 49*(3), 487-492

Simon, D. K., Friedman, J., Breakefield, X. O., Jankovic, J., Brin, M. F., Provias, J., . . . Tarnopolsky, M. A. (2003). A heteroplasmic mitochondrial complex I gene mutation in adult-onset dystonia. *Neurogenetics, 4*(4), 199-205

Spiegel, R., Shaag, A., Mandel, H., Reich, D., Penyakov, M., Hujeirat, Y., . . . Shalev, S. A. (2009). Mutated NDUFS6 is the cause of fatal neonatal lactic acidemia in Caucasus Jews. *European journal of human genetics : EJHG, 17*(9), 1200-1203

Steuber, J., Schmid, C., Rufibach, M., & Dimroth, P. (2000). Na+ translocation by complex I (NADH:quinone oxidoreductase) of Escherichia coli. *Molecular microbiology, 35*(2), 428-434

Straka, J. G., Rank, J. M., & Bloomer, J. R. (1990). Porphyria and porphyrin metabolism. *Annu Rev Med, 41*, 457-469

Stroh, A., Anderka, O., Pfeiffer, K., Yagi, T., Finel, M., Ludwig, B., & Schagger, H. (2004). Assembly of respiratory complexes I, III, and IV into NADH oxidase supercomplex stabilizes complex I in *Paracoccus denitrificans. J Biol Chem, 279*(6), 5000-5007

Stroud, D. A., Surgenor, E. E., Formosa, L. E., Reljic, B., Frazier, A. E., Dibley, M. G., . . . Ryan, M. T. (2016). Accessory subunits are integral for assembly and function of human mitochondrial complex I. *Nature, 538*(7623), 123-126

Stuart, R. (2002). Insertion of proteins into the inner membrane of mitochondria: the role of the Oxa1 complex. *Biochim Biophys Acta, 1592*(1), 79-87

Sturtz, L. A., Diekert, K., Jensen, L. T., Lill, R., & Culotta, V. C. (2001). A fraction of yeast Cu,Zn-superoxide dismutase and its metallochaperone, CCS, localize to the intermembrane space of mitochondria. A physiological role for SOD1 in guarding against mitochondrial oxidative damage. *J Biol Chem, 276*(41), 38084-38089

Sudhof, T. C. (2012). Calcium control of neurotransmitter release. *Cold Spring Harb Perspect Biol, 4*(1), a011353

Sudo, A., Honzawa, S., Nonaka, I., & Goto, Y. (2004). Leigh syndrome caused by mitochondrial DNA G13513A mutation: frequency and clinical features in Japan. *J Hum Genet, 49*(2), 92-96

Sun, F., Huo, X., Zhai, Y., Wang, A., Xu, J., Su, D., . . . Rao, Z. (2005). Crystal structure of mitochondrial respiratory membrane protein complex II. *Cell, 121*(7), 1043-1057

Szklarczyk, R., Wanschers, B. F., Nabuurs, S. B., Nouws, J., Nijtmans, L. G., & Huynen, M. A. (2011). NDUFB7 and NDUFA8 are located at the intermembrane surface of complex I. *FEBS Lett, 585*(5), 737-743

Taylor, R. W., Morris, A. A., Hutchinson, M., & Turnbull, D. M. (2002). Leigh disease associated with a novel mitochondrial DNA ND5 mutation. *European journal of human genetics : EJHG, 10*(2), 141-144

Thannickal, V. J., & Fanburg, B. L. (2000). Reactive oxygen species in cell signaling. *Am J Physiol Lung Cell Mol Physiol, 279*(6), L1005-1028

Tocilescu, M. A., Zickermann, V., Zwicker, K., & Brandt, U. (2010). Quinone binding and reduction by respiratory complex I. *Biochim Biophys Acta, 1797*(12), 1883-1890

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., . . . Yoshikawa, S. (1996). The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 A. *Science, 272*(5265), 1136-1144

Tucker, E. J., Compton, A. G., Calvo, S. E., & Thorburn, D. R. (2011). The molecular basis of human complex I deficiency. *IUBMB Life, 63*(9), 669-677

Tuppen, H. A., Hogan, V. E., He, L., Blakely, E. L., Worgan, L., Al-Dosary, M., . . . Taylor, R. W. (2010). The p.M292T NDUFS2 mutation causes complex I-deficient Leigh syndrome in multiple families. *Brain : a journal of neurology, 133*(10), 2952-2963

Ushakova, A. V., Duarte, M., Vinogradov, A. D., & Videira, A. (2005). The 29.9 kDa subunit of mitochondrial complex I is involved in the enzyme active/de-active transitions. *J Mol Biol, 351*(2), 327-333

van den Bosch, B. J., Gerards, M., Sluiter, W., Stegmann, A. P., Jongen, E. L., Hellebrekers, D. M., . . . Smeets, H. J. (2012). Defective NDUFA9 as a novel cause of neonatally fatal complex I disease. *J Med Genet, 49*(1), 10-15

van Heel, M., Gowen, B., Matadeen, R., Orlova, E. V., Finn, R., Pape, T., . . . Patwardhan, A. (2000). Single-particle electron cryo-microscopy: towards atomic resolution. *Q. Rev. Biophys., 33*(4), 307-369

Vanniarajan, A., Rajshekher, G. P., Joshi, M. B., Reddy, A. G., Singh, L., & Thangaraj, K. (2006). Novel mitochondrial mutation in the ND4 gene associated with Leigh syndrome. *Acta Neurol Scand, 114*(5), 350-353

Vartak, R., Deng, J., Fang, H., & Bai, Y. (2015). Redefining the roles of mitochondrial DNA-encoded subunits in respiratory Complex I assembly. *Biochim Biophys Acta, 1852*(7), 1531-1539

Verkhovskaya, M., & Bloch, D. A. (2013). Energy-converting respiratory Complex I: on the way to the molecular mechanism of the proton pump. *Int J Biochem Cell Biol, 45*(2), 491-511

Verkhovsky, M. I., Jasaitis, A., Verkhovskaya, M. L., Morgan, J. E., & Wikstrom, M. (1999). Proton translocation by cytochrome c oxidase. *Nature, 400*(6743), 480-483

Vinogradov, A. D. (1998). Catalytic properties of the mitochondrial NADH-ubiquinone oxidoreductase (complex I) and the pseudo-reversible active/inactive enzyme transition. *Biochim. Biophys. Acta, 1364*(2), 169-185

Vinothkumar, K. R., & Henderson, R. (2016). Single particle electron cryomicroscopy: trends, issues and future perspective. Q Rev Biophys, 49, e13

Vinothkumar, K. R., Zhu, J., & Hirst, J. (2014). Architecture of mammalian respiratory complex I. Nature

Viollet B, Guigas B, Sanz Garcia N, Leclerc J, Foretz M, Andreelli F. Cellular and molecular mechanisms of metformin: an overview. Clinical Science (London, England : 1979). 2012;122(6):253-270.

Visch, H. J., Koopman, W. J., Leusink, A., van Emst-de Vries, S. E., van den Heuvel, L. W., Willems, P. H., & Smeitink, J. A. (2006). Decreased agonist-stimulated mitochondrial ATP production caused by a pathological reduction in endoplasmic reticulum calcium content in human complex I deficiency. *Biochim Biophys Acta, 1762*(1), 115-123

Visch, H. J., Rutter, G. A., Koopman, W. J., Koenderink, J. B., Verkaart, S., de Groot, T., . . . Willems, P. H. (2004). Inhibition of mitochondrial Na+-Ca2+ exchange

restores agonist-induced ATP production and Ca2+ handling in human complex I deficiency. *J Biol Chem, 279*(39), 40328-40336

Vogel, R. O., Dieteren, C. E., van den Heuvel, L. P., Willems, P. H., Smeitink, J. A., Koopman, W. J., & Nijtmans, L. G. (2007). Identification of mitochondrial complex I assembly intermediates by tracing tagged NDUFS3 demonstrates the entry point of mitochondrial subunits. *J Biol Chem, 282*(10), 7582-7590

Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., . . . Bricogne, G. (2011). Data processing and analysis with the autoPROC toolbox. *Acta crystallogr. D, 67*(Pt 4), 293-302

Walker, J. E. (1992). The NADH - ubiquinone oxidoreductase (complex I) of respiratory chains. *Q. Rev. Biophys., 25*(3), 253-324

Walker, J. E. (2013). The ATP synthase: the understood, the uncertain and the unknown. *Biochem Soc Trans, 41*(1), 1-16

Wallace, D. C., Singh, G., Lott, M. T., Hodge, J. A., Schurr, T. G., Lezza, A. M., . . . Nikoskelainen, E. K. (1988). Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science, 242*(4884), 1427-1430.

Wallin, I. E. (1923). The mitochondria problem. *American Naturalist, 57*, 255-261

Walther, D. M., & Rapaport, D. (2009). Biogenesis of mitochondrial outer membrane proteins. *Biochim Biophys Acta, 1793*(1), 42-51

Wang, C., & Youle, R. J. (2009). The role of mitochondria in apoptosis*. *Annu Rev Genet, 43*, 95-118

Wang, Q., Li, R., Zhao, H., Peters, J. L., Liu, Q., Yang, L., . . . Guan, M. X. (2005). Clinical and molecular characterization of a Chinese patient with auditory neuropathy associated with mitochondrial 12S rRNA T1095C mutation. *Am J Med Genet A, 133a*(1), 27-30

Watt, I. N., Montgomery, M. G., Runswick, M. J., Leslie, A. G., & Walker, J. E. (2010). Bioenergetic cost of making an adenosine triphosphate molecule in animal mitochondria. *Proc. Natl. Acad. Sci. U S A, 107*(39), 16823-16827

Weerakoon, D. R., & Olson, J. W. (2008). The Campylobacter jejuni NADH:ubiquinone oxidoreductase (complex I) utilizes flavodoxin rather than NADH. *J Bacteriol, 190*(3), 915-925

Wiesner, R. J., Ruegg, J. C., & Morano, I. (1992). Counting target molecules by exponential polymerase chain reaction: copy number of mitochondrial DNA in rat tissues. *Biochem Biophys Res Commun, 183*(2), 553-559

Wikstrom, M. (1984). Two protons are pumped from the mitochondrial matrix per electron transferred between NADH and ubiquinone. *FEBS Lett., 169*(2), 300-304

Wray, C. D., Friederich, M. W., du Sart, D., Pantaleo, S., Smet, J., Kucera, C., . . . Van Hove, J. L. (2013). A new mutation in MT-ND1 m.3928G>C p.V208L causes Leigh disease with infantile spasms. *Mitochondrion, 13*(6), 656-661

Wu, M., Gu, J., Guo, R., Huang, Y., & Yang, M. (2016). Structure of Mammalian Respiratory Supercomplex I1III2IV1. *Cell, 167*(6), 1598-1609 e1510

Xia, D., Yu, C. A., Kim, H., Xian, J. Z., Kachurin, A. M., Zhang, L., . . . Deisenhofer, J. (1997). Crystal structure of the cytochrome $bc_1$ complex from bovine heart mitochondria. *Science, 277*(5322), 60-66

Yakovlev, G., Reda, T., & Hirst, J. (2007). Reevaluating the relationship between EPR spectra and enzyme structure for the iron sulfur clusters in NADH:quinone oxidoreductase. *Proc Natl Acad Sci U S A, 104*(31), 12720-12725.

Ye, H., & Rouault, T. A. (2010). Human iron-sulfur cluster assembly, cellular iron homeostasis, and disease. *Biochemistry-Us, 49*(24), 4945-4956

Yip, C.-Y., Fearnley, I. M., Harbour, M. E., Jayawardena, K., & Sazanov, L. A. (2010). Purification and characterisation of complex I from Paracoccus denitrificans. *Bba-Bioenergetics, 1797*, 25-25

Yip, C. Y., Harbour, M. E., Jayawardena, K., Fearnley, I. M., & Sazanov, L. A. (2011). Evolution of respiratory complex I: "supernumerary" subunits are present in the alpha-proteobacterial enzyme. *J. Biol. Chem., 286*(7), 5023-5033

Yu-Wai-Man, P., & Chinnery, P. F. (1993). Leber Hereditary Optic Neuropathy. In R. A. Pagon, M. P. Adam, H. H. Ardinger, S. E. Wallace, A. Amemiya, L. J. H. Bean, T. D. Bird, N. Ledbetter, H. C. Mefford, R. J. H. Smith, & K. Stephens (Eds.), *GeneReviews(R)*

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J Struct Biol, 193*(1), 1-12

Zhu, J., King, M. S., Yu, M., Klipcan, L., Leslie, A. G., & Hirst, J. (2015). Structure of subcomplex Ibeta of mammalian respiratory complex I leads to new supernumerary subunit assignments. *Proc Natl Acad Sci U S A, 112*(39), 12087-12092

Zhu, J., Vinothkumar, K. R., & Hirst, J. (2016). Structure of mammalian respiratory complex I. *Natur*

Zhu, S., & Vik, S. B. (2015). Constraining the Lateral Helix of Respiratory Complex I by Cross - linking Does Not Impair Enzyme Activity or Proton Translocation. *J Biol Chem, 290*(34), 20761-20773

Zickermann, V., Wirth, C., Nasiri, H., Siegmund, K., Schwalbe, H., Hunte, C., & Brandt, U. (2015). Structural biology. Mechanistic insight from the crystal structure of mitochondrial complex I. *Science, 347*(6217), 44-49

Zwicker, K., Galkin, A., Drose, S., Grgic, L., Kerscher, S., & Brandt, U. (2006). The Redox-Bohr group associated with iron-sulfur cluster N2 of complex I. *J. Biol. Chem., 281*(32), 23013-23017