Human mitochondrial protein complexes revealed by large-scale coevolution analysis and deep learning-based structure modeling

Jimin Pei, Jing Zhang, and Qian Cong*

McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center at Dallas

6001 Forest Park Rd., Dallas, TX, USA, Texas, U.S.A. 75390

Correspondence: qian.cong@utsouthwestern.edu

Abstract

Recent development of deep-learning methods has led to a breakthrough in the prediction accuracy of 3-dimensional protein structures. Extending these methods to protein pairs is expected to allow large-scale detection of protein-protein interactions and modeling protein complexes at the proteome level. We applied RoseTTAFold and AlphaFold2, two of the latest deep-learning methods for structure predictions, to analyze coevolution of human proteins residing in mitochondria, an organelle of vital importance in many cellular processes including energy production, metabolism, cell death, and antiviral response. Variations in mitochondrial proteins have been linked to a plethora of human diseases and genetic conditions. RoseTTAFold, with high computational speed, was used to predict the coevolution of about 95% of mitochondrial protein pairs. Top-ranked pairs were further subject to the modeling of the complex structures by AlphaFold2, which also produced contact probability with high precision and in many cases consistent with RoseTTAFold. Most of the top ranked pairs with high contact probability were supported by known protein-protein interactions and/or similarities to experimental structural complexes. For high-scoring pairs without experimental complex structures, our coevolution analyses and structural models shed light on the details of their interfaces, including CHCHD4-AIFM1, MTERF3-TRUB2, FMC1-ATPAF2, ECSIT-NDUFAF1 and COQ7-COQ9, among others. We also identified novel PPIs (PYURF-NDUFAF5, LYRM1-MTRF1L and COA8-COX10) for several proteins without experimentally characterized interaction partners, leading to predictions of their molecular functions and the biological processes they are involved in.

Key words

mitochondrial proteins; coevolution analysis; protein-protein interactions; RoseTTAFold; AlphaFold2

Introduction

Recent advances in Deep-Learning (DL) techniques for structure prediction based on protein-sequence alignments have led to a breakthrough in structural genomics. These state-of-the-art methods, which are now sensitive enough to work with shallow alignments, can predict protein 3D structure to atomic accuracy [1-3]. The methods can be applied on the whole-proteome scale [4], with the 3D structures for almost all human proteins being recently "determined" by DeepMind using Alphafold2. While the quality of these 3D structures remains to be validated by the scientific community, they unquestionably facilitate functional characterization of these proteins and interpretation of disease-causing mutations in them [4].

One of the next research directions where these DL methods can transform is to determine protein-protein interactomes. Using statistical analyses of coevolutionary signal between residues from different proteins, we have shown that the interface between interacting proteins in Bacteria can be accurately predicted from alignments of diverse sequences on the proteome scale [5]. This type of *in silico* protein-protein interaction screen appears to be more accurate than large-scale experimental screens such as yeast-two-hybrid and affinity-purification-mass-spectrometry [5]. The application of similar methods to human sequences were hindered by limitations in sensitivity for shallower Eukaryotic sequence alignments. However, we expect the increased performance of current DL methods will enable proteome-wide screen of protein-protein interaction (PPI) in human.

Mitochondria are dynamic membrane-bound organelles found in almost all eukaryotic cells. They carry a variety of essential cellular functions ranging from energy production and metabolism to regulation of cell death and immune response [6-10]. Mitochondria evolved from an endosymbiotic relationship between ancestral bacteria and eukaryotes [11]. While these organelles maintain their own genetic materials and transcriptional and translational machineries, most proteins targeted to mitochondria are now encoded in the nuclear genome. Technology advances in genomics, transcriptomics, proteomics, and metabolomics have extended the repertoire of mitochondrion-associated proteins, identified their interaction partners, and expanded the knowledge about their functions [12-16]. The MitoCarta3.0 database has summarized 1136 mitochondrial proteins in human and classified them according to their pathways [17]. These mitochondrial proteins are located in four compartments: the mitochondrial outer membrane (MOM), the intermembrane space (IMS), the mitochondrial inner membrane (MIM), and the matrix. Characterizations of mitochondrial proteins and

their complexes are crucial to our understanding of how mitochondria regulate various cellular activities and the mechanisms of a diverse range of mitochondrion-related diseases [18].

As a pilot study, we applied RoseTTAFold [1] and Alphafold2 [2] to identify interacting protein pairs and predict their interfaces among human mitochondrial proteins. We showed that a combination of these methods can identify interacting partners and model protein complexes on a large-scale with high precision. With these methods, we captured many known PPIs that were supported by multiple experiments and/or protein complex structures. In addition, we obtained detailed information about the interfaces of known PPIs without experimentally determined complex structures and identified novel PPIs for several proteins without experimentally characterized interaction partners. These findings provided structural insight into the mechanisms of some disease-causing mutations and functional predictions of several poorly characterized proteins.

Results and discussion

In silico screen can identify interacting proteins and predict their interface at high precision

We designed a computational procedure (flowchart in Figure 1A, see Materials and methods) to perform *in silico* PPI screen for 605099 pairs of 1118 mitochondrial proteins from the MitoCarta3.0 database [17] using RoseTTAFold [1] and AphaFold2 [2]. About 5% of mitochondrial protein pairs were excluded from this study because their combined lengths were beyond our GPU's compute capability. We used the best inter-residue score (see Materials and methods) between a pair of proteins to indicate their interaction confidence. Precision-recall curves based on true PPIs (positive controls) and random protein pairs (negative controls, see Materials and methods) evaluate the performance of these methods. The RoseTTAFold 2-track model, a fast version of RoseTTAFold without 3D representation of the proteins inside the model, is remarkably better than Direct Coupling Analysis (DCA), a statistical method we used previously in coevolution-based Bacteria PPI screens [5] (Figure 1B). The area under the precision-recall curve scores (AUC) for DCA (0.014) is much lower than the AUC for RoseTTAFold (0.113). In addition, we observed that residues in the mitochondria-targeting peptides of different proteins can show high contact probability, although they are not expected to be involved in protein-protein interface as they are cleaved after being transferred to mitochondria. Excluding the mitochondrion-targeting peptides

81 before computing the highest inter-residue contact probability for a pair of proteins further improved

82 the accuracy of the screen (Figure. 1C) (AUC improved from 0.113 to 0.132).

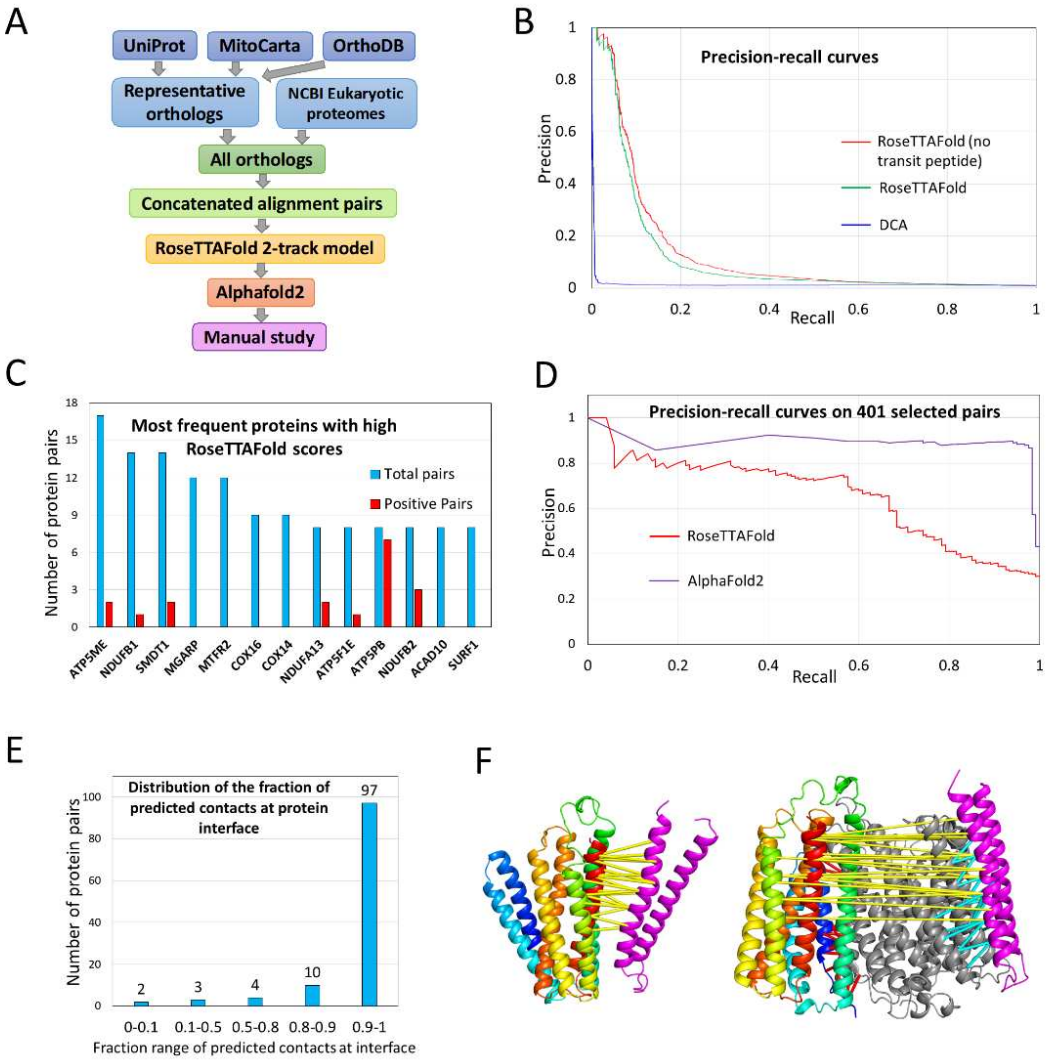

Figure 1. **A**. A flowchart of the procedure used in this study to perform PPI prediction and structural modeling. **B**. The precision-recall curves of three methods in detecting positive PPIs of mitochondrial protein pairs. DCA - Direct Coupling Analysis method, RoseTTAFold (no transit peptide) is the RoseTTAFold method used on residue pairs excluding those involving any residue in mitochondrial transit peptides. **C.** Counts of high-scoring interacting partners (contact probability score >0.9) (blue bars) for proteins with the highest number of such partners. The red bars (if any) show the number of true positives among the top-scoring pairs. **D.** Precision-recall curves for RoseTTFold and AlphaFold2 on the 401 pairs of selected PPI predictions. **E.** Distribution of the fraction of predicted contacts that are mapped to experimentally determined complex interfaces. **F.** The left panel shows the AlphaFold2 model of MT-CO2 (magenta) and MT-CO3 (rainbow) complex, with the top 20 RoseTTAFold-predicted contact pairs shown as yellow lines connecting their C-α atoms. The right panel shows the AlphaFold2 model of the MT-CO1 (gray), MT-CO2 (magenta) and MT-CO3 (rainbow) complex. Top contact pairs are shown in yellow, cyan, and red lines for MT-CO2/MT-CO3, MT-CO1/MT-CO2 and MT-CO1/MT-CO3 pairs, respectively.

The majority of the top 100 pairs (70 out of 100) with the highest RoseTTAFold contact probability are PPIs supported by either experimental structural complexes or multiple incidences (≥ 3) in the BioGRID [19] database. They include 37 pairs involving mitochondrial ribosomal subunits [20, 21] and more than 10 pairs involving oxidative phosphorylation (OXPHOS) complex I (CI, NADH:ubiquinone oxidoreductase complex) subunits or assembly factors [22, 23]. Multiple PPIs were also found between the subunits and assembly factors of OXPHOS complex II (CII, succinate-Q oxidoreductase complex) [24], complex III (CIII, cytochrome bc1 complex), complex IV (CIV, ubiquinol-cytochrome c oxidase complex), and complex V (CV, ATP synthase complex) [23], respectively. Other protein pairs from the same complexes identified as top hits of our *in silico* screen include SUCLG1-SUCLG2 and SUCLG1-SUCLA2 from the succinyl-CoA synthetase complexes [25], PDHA1-PDHB and PDHA2-PDHB from the pyruvate dehydrogenase complex [26], ETFA-ETFB from the electron transfer flavoprotein complex [27], BCKDHA-BCKDHB from the branched-chain alpha-keto acid dehydrogenase complex [28], GATB-GATC and GATC-QRSL1 from the glutaminyl-tRNA amidotransferase complex [29], and TIMM9-TIMM10 from the translocase of the inner membrane (TIM) complex [30].

One known artifact of statistical methods for inter-protein coevolution analysis occurs for some proteins showing a high coevolution signal with many other proteins [5]. We observed the same artifact implementing RoseTTAFold on mitochondrial PPIs (Figure. 1C). 13 proteins exhibit a contact probability greater than 0.9 for 8 or more different interaction partners. Except for ATP5PB, interactions for most of these pairs are not supported by experiments (either experimental complex structures or BioGRID count $\geq$3) and could be false positives. To penalize such hubs of false positives, we adjusted the contact probability between a pair of proteins by a weight measuring each protein's top contact probability with other proteins (see Materials and methods). We ranked the candidate PPIs by both the original contact probability and the adjusted contact score and selected 401 pairs as potential interacting partners (see Materials and methods, supplementary Table S1).

We modeled the 3D structure for these 401 pairs using AlphaFold2 [2]. Surprisingly, although AlphaFold2 was not trained to distinguish true PPIs from false ones, it exhibits remarkable performance on this task. It ranks the majority of the true positives among the 401 candidate PPIs higher than other pairs, achieving high precision (Figure 1D): 136 pairs among the 401 candidate PPIs had a top contact probability above 0.5 by Alphafold2, and 118 of them are considered true positives according to our criteria (supported by experimental structure or BioGRID count ≥ 3), indicating a precision of at least 87%. Those top hits not meeting our criteria could still be true PPIs, as some of them are supported by

low-throughput experimental studies. Interesting examples are discussed in the following sections. At this level of precision, we recovered 118 of the 1563 true PPIs in the positive control set, indicating a recall of 7.5%.

It is important to note that applying AlphaFold2 to all pairs of mitochondrial proteins was computationally intractable (need ~100,000 GPU hours) , and reported performance was achieved by the sequential application of the fast RoseTTAFold 2-track model and AlphaFold2. We expect a better recall could be achieved by applying AlphaFold2 to all pairs of mitochondrial proteins. To probe the recall of AlphaFold2 for true PPIs, we applied it to all 473 pairs of mitochondrial proteins with confident experimental evidence (BioGRID reference count >= 3) for their interaction, and 120 pairs (25.4%) had top AlphaFold2 contact probability above 0.5 (Supplementary Table S2). We deposited the Alphafold2 models of these protein pairs at https://osf.io/g37mz/ The high accuracy of these AlphaFold2 models makes them a valuable resource to understand complex function, guide experimental design and explain disease mechanisms.

Out of the 401 interacting pairs we predicted *de novo* in our screen, 116 have known complex structure or close templates in the Protein Data Bank. We used these cases to evaluate the accuracy of residues contributing to the protein interface as suggested by pairwise RoseTTAFold scores . Most of these RoseTTAFold contacts show near perfect agreement with the experimental structures (Figure 1E). Only two pairs of proteins have low agreement between the contact residues predicted by RoseTTAFold and the interfaces defined by experimental complexes. One of the poorly predicted pairs is between MT-CO2 (COX2) and MT-CO3 (COX3), two mitochondrion-encoded CIV subunits. RoseTTAFold assigns high contact probability between the second transmembrane segment of MT-CO2 and the fourth and seventh transmembrane segments of COX3, in agreement with the AlphaFold2 structural model of MT-CO2 and MT-CO3 (Figure 1F, left panel). However, in experimental CIV structures, the transmembrane segments of MT-CO2 and MT-CO3 are separated by MT-CO1 and do not make contacts. We hypothesized that this wrong prediction was because we modeled a complex of MT-CO2 and MT-CO3 in the absence of MT-CO1. Therefore, we tried to model MT-CO1, MT-CO2, and MT-CO3 together by AlphaFold2, which generated a complex model in agreement with experimental structures (Figure 1F, right panel). The RoseTTAFold-predicted contacts between MT-CO1 and MT-CO2 (shown in red) and between MT-CO1 and MT-CO3 (shown in cyan) agree with this trimeric complex structure.

Coevolution analysis revealed protein interaction interfaces for known protein complexes without experimental 3D structures

Several top-ranking protein pairs from our screen were found to have experimental support for their interactions but lack experimentally determined structures of the complexes. Our coevolution analysis supports the interactions between these protein pairs and provides detailed structural information on their interaction interfaces. Some of these pairs are described below.

CHCHD4 and AIFM1

AIFM1 (Apoptosis-Inducing Factor 1, Mitochondria, also named AIF) is a mitochondrial flavoprotein with multiple functions in both cell survival and cell death. It is essential for maintaining mitochondrial integrity in healthy cells and plays an important role in the apoptosis pathway [31, 32]. Anchored to the MIM by an N-terminal transmembrane segment, the majority of AIFM1localizes to the IMS. AIFM1 exhibits FAD-dependent NADH oxidase activity [33] and encompasses two Rossmann-like FAD-binding domains (Pfam: Pyr_redox_2) and a C-terminal mixed alpha-beta domain (Pfam: Reductase_C) [32]. Upon apoptotic stimuli, AIFM1 is cleaved to produce a C-terminal soluble fragment that is translocated to the nucleus to trigger downstream apoptotic events. AIFM1 is also essential for cell viability as its downregulation results in the loss of CI subunits. The pro-survival activity of AIFM1 is tied to its interaction with CHCHD4, a CHCH domain-containing protein located in the IMS [34]. CHCHD4 (the ortholog of yeast protein Mia40) is a crucial component of the redox-regulated MIA machinery responsible for the import of a set of cysteine-containing protein substrates, such as COX17, COX19, MICU1 and COA7 [35]. CHCHD4 functions as a chaperone and catalyzes the formation of disulfide bonds in these substrate proteins, including some respiratory chain subunits and proteins with a variety of other functions such as redox regulation and antioxidant response [35].

We identified CHCHD4-AIFM1 as one of the top scoring PPI pairs with both AlphfaFold2 and RoseTTAFold contact probabilities above 0.99. The AlphaFold2-predicted complex structure between them is shown in Figure 2A. Their interaction is primarily mediated by the N-terminus of CHCHD4, which is missing in the solution structure of CHCHD4 and could be disordered on its own [36]. The predicted interacting residues of CHCHD4 are not from the CHCH domain (residue range: 45 to 109) with two $(CX_9C)_2$ motifs [37, 38]. Instead, our predicted interaction site is consistent with experimental studies describing the N-terminal CHCHD4 27-residue segment being sufficient for its interaction with AIFM1

[34]. Our prediction also mapped the interaction surface on AIFM1, which has eluded experimental studies thus far. The AIFM1 surface is mainly on the edge β-strand (residue 504-508) in the C-terminal Reductase_C domain (Figure 2A). The N-terminal segment of CHCHD4 forms a β-hairpin that interacts with this edge β-strand (Figure 2A). More than 20 missense mutations in AIFM1 (magenta positions in Figure 2A) have been reported in association with several diseases and have been classified as pathogenic/likely pathogenic [39-42]. Interestingly, one likely pathogenic mutation from AIFM1 (Y560H) is mapped near the interaction site (magenta sidechain shown in Figure 2A), and it is predicted to interact with two hydrophobic residues (I12 and F14) on CHCHD4 (sidechains shown in green spheres in Figure 2A). Interactions between these hydrophobic residues could contribute to the binding energy between CHCHD4 and AIFM1 and could explain potential deleterious effects for this mutation.

MTERF3 and TRUB2

MTERF3 (mitochondrial transcription termination factor 3, also named mitochondrial transcription termination domain containing 1 (MTERFD1)) resides in the mitochondrial matrix and is a negative regulator of mitochondrial DNA transcription [43]. MTERF3 also regulates mitochondrial ribosome biogenesis in vertebrates [44] and is a modular factor in mitochondrial protein synthesis [45]. The crystal structure of human MTERF3 revealed that it adopts a right-handed superhelix fold with an overall shape of a half donut [46]. The superhelix consists of tandem repeats of three α-helices named the MTERF-motif [46]. It was proposed that MTERF3 could interact with DNA using the concave side of the half-doughnut that harbors positively charged residues [46].

Recently, MTERF3 was found to colocalize with other proteins in mitochondrial RNA granules, which play crucial roles in post-transcriptional modifications of mitochondrial RNAs and the assembly of the mitochondrial ribosome [47, 48]. MTERF3 is a component of the multi-functional pseudouridine synthase module [48] within the RNA granule together with enzyme components responsible for pseudouridylation of 16S rRNA and mitochondrial mRNAs. TRUB2 is such a pseudouridine synthase that has been shown to interact with MTERF3, and depletion of MTERF3 led to significant decrease in TRUB2 as well as other components of the module [48]. Our coevolution analysis yields high contact probability scores (>0.99) by both AlphaFold2 and RoseTTAFold for MTERF3 and TRUB2, consistent with the experimentally verified physical interaction between them. Furthermore, the structure model of MTERF3 with TRUB2 sheds light on the interaction mode. Interestingly, coevolving residue pairs form

two potential interaction sites involving distal ends of the MTERF3 superhelix domain (Figure 2B). Such interactions convert the half-donut shape of MTERF3 to a ring-like complex structure with TRUB2 that maintains an empty space in the middle. This empty space could accommodate the binding of RNA molecules in the RNA granule. Such a structure could be important for maintaining the pseudouridine synthase module within the RNA granule and could play a role in delivering RNA substrates to the

209 pseudouridine synthase, TRUB2, for efficient reactions of pseudouridylation.

210

211 **FMC1 and ATPAF2**

212 A functional connection between human mitochondrial proteins FMC1 (formerly named C7orf55) and

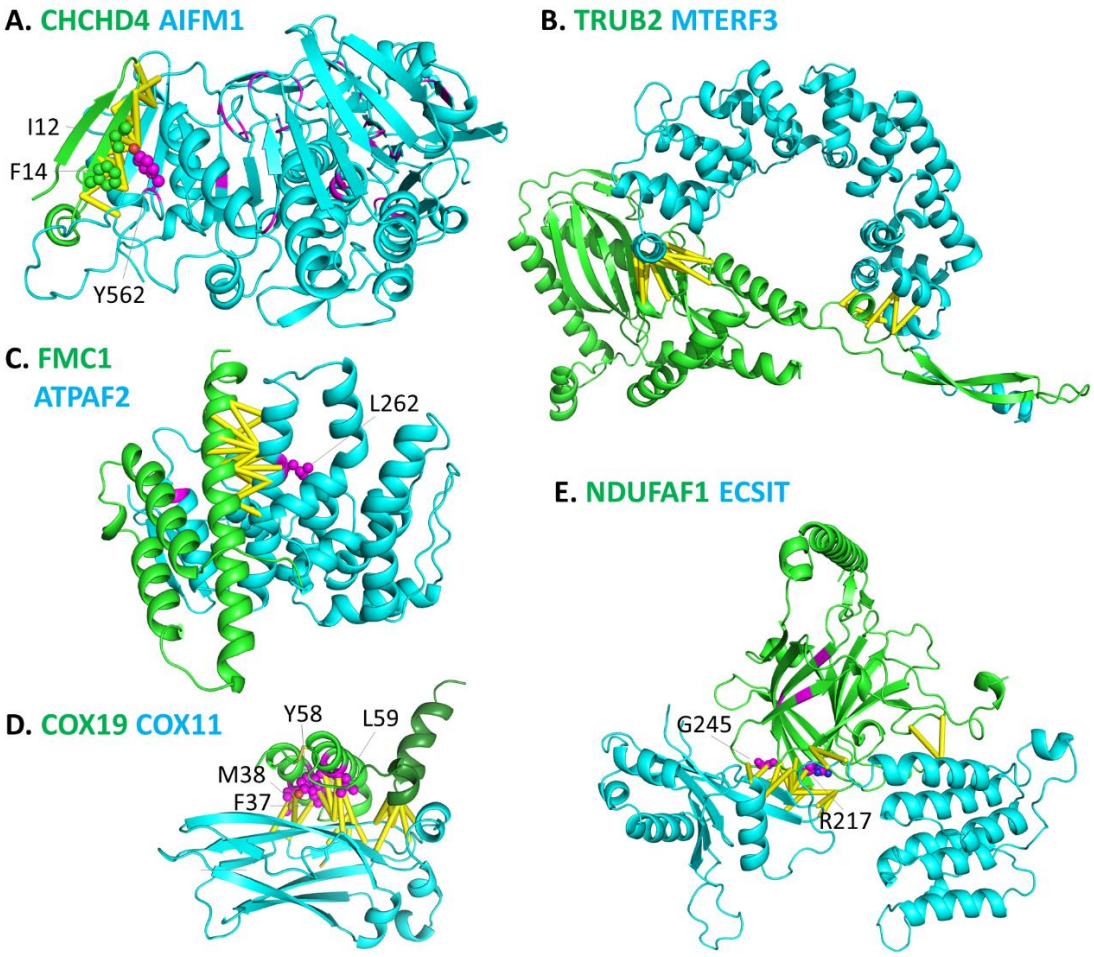

Figure 2. AlphaFold2 structural models of protein complexes without experimentally determined 3D structures. **A**. CHCHD4-AIFM1. Residues involving missense disease-causing mutations are colored in magenta, and those near the interaction interface have their sidechains shown in spheres. Sidechains of the two hydrophobic residues of CHCHD4 interacting with AIFM1 Y562 are shown in green spheres. **B**. TRUB2-MTERF3. **C**. FMC1-ATPAF2. Residues involving missense disease-causing mutations are colored in magenta, and those near the interaction interface have their sidechains shown in spheres **D**. COX19-COX11. Sidechains of the YL and FM signature sequences are shown in magenta spheres. The CHCH domain is shown in light green and the C-terminal two helices are shown in dark green. **E.** NDUFAF1-ECSIT. Residues of missense disease-causing mutations are colored in magenta, and those near the interaction interface are shown in spheres.

ATPAF2 (ATP synthase F1 complex assembly factor 2) was recently predicted by a computational tool (CLustering by Inferred Co-expression, CLIC) [49]. The interaction between these two proteins was also experimentally validated [49]. FMC1 belongs to the LYRM family of proteins [50, 51], and ATPAF2 is one of the assembly factors of CV [52]. This interaction was also found in two high-throughput PPI studies (reported in BioGRID) [14, 16]. Our PPI screen ranked the FMC1 and ATPAF2 pair among the top hits with both AlphaFold2 and RoseTTAFold contact probabilities above 0.99. The FMC1 protein is predicted to adopt a three-helix bundle fold, typical of the LYRM family proteins. The interaction site on FMC1 is mapped to the third core α-helix of such a fold (Figure 2C). The structure model of ATPAF2 contains a small 3-stranded β-sheet at the N-terminus, while the rest of the protein is mainly α-helical. The interaction site on ATPAF2 is mapped to the last α-helix and includes one residue (magenta sidechain shown in spheres in Figure 2C) with a likely pathogenic mutation (L262P) reported in the ClinVar database [41].

COX19 and COX11

COX19 is a conserved protein found in mitochondria of a diverse range of eukaryotic organisms including fungi, plants, and animals [53]. Like CHCHD4, COX19 belongs to a family of IMS proteins that contain the CHCH domain with a pair of CX9C signatures [37, 38]. The conserved cysteines in the CX9C motifs form two disulfide bonds in a pair of short antiparallel α-helices. Recently, COX19 was found to interact with COX11, a single-pass transmembrane protein involved in copper transfer to the Cu_B center of the COX1 subunit in the cytochrome c oxidase complex [54]. The interaction with Cox11 is required for maintaining stable levels of Cox19 in mitochondria [54]. Both COX19 and COX11 are classified as copper chaperones and CIV assembly factors [17, 55].

Our analysis revealed strong AlphaFold2 and RoseTTAFold contact probability scores (>0.99) between COX19 and COX11. Predicted interacting residues on COX19 are mainly mapped to the α-helical hairpin of the CHCH domain with the twin CX9C motifs (Figure 2D). The CX9C motifs in COX19 proteins are characterized by two conserved YL diads, giving rise to the refined motif signatures of Cx_6YLxC [54]. In the yeast system, these YL diads were proved to be crucial for the binding to COX11 based on mutagenesis studies [54]. This binding mode is consistent with our model, which showed high contact probabilities for residues corresponding to yeast YL diads in the human COX19 protein, where the first YL diad is changed to FM and the second YL is maintained (magenta residues in Figure 2D). The

Cx_6YLxC α-hairpin of COX19 mainly interacts with the second, third, and sixth core β-strands of the immunoglobulin domain of COX11 [56] (Figure 2D). The COX19 structural model also includes two small α-helices (colored in dark green in Figure 2D) C-terminal to the Cx_6YLxC helix hairpin (colored in lighter green). Residues in the turn between these two C-terminal α-helices are also predicted to interact with COX11 residues from the loop before the first core β-strand and the loop between the second and third β-strands (Figure 2D).

ECSIT and NDUFAF1

ECSIT (Evolutionarily Conserved Signaling Intermediate in Toll pathways) was originally identified as a protein involved in the Toll and bone morphogenetic protein signaling pathways [57, 58]. Later studies showed that ECSIT is a CI assembly factor that interacts with other assembly factors such as NDUFAF1 and ACAD9 [59, 60]. The interaction between ECSIT and NDUFAF1 is strongly supported by AlphaFold2 and RoseTTAFold (contact probability >0.99). The AlphaFold2-predicted ECSIT structure has an N-terminal domain consisting of α-helical repeats and a C-terminal alpha+beta domain with α-helices sandwiching a mainly anti-parallel β-sheet of six β-strands (cyan structure in Figure 2E). The AlphaFold2 model of NDUFAF1 has an N-terminal α-helix with mainly hydrophilic residues and a C-terminal β-sandwich with a jelly roll topology (green structure in Figure 2E). Multiple NDUFAF1-interacting sites of were predicted in ECSIT, including the β-hairpin formed by the third and fourth core β-strands of the second domain, the C-terminal end of the first α-helix of the second domain, the last α-helix of the first domain, and the loop in between the two domains. The interaction surface on NDUFAF1 is mainly mapped to the loop regions at one end between the two β-sheets of the jelly roll domain. Two pathogenic mutations responsible for Mitochondrial complex I deficiency, nuclear type 11 (R217C and G245R, colored in magenta in Figure 2E) in NDUFAF1 [61] lie in the predicted interface between NDUFAF1 and ECSIT, and they are predicted to interact with residues in ECSIT with high contact probability.

COQ3 and COQ6

COQ3 [62] and COQ6 [63] are two enzymes in the ubiquinone (coenzyme Q, CoQ) synthesis pathway, an essential component of the mitochondrial oxidative phosphorylation machinery. CoQ is synthesized by

the combined actions of at least nine proteins (COQ1–9) [64]. COQ6 is a monooxygenase that converts the C5-hydroxylation of 3-decaprenyl-4-hydroxybenzoic acid (DHB) to 3,4-dihydroxy-5-decaprenylbenzoic acid (DHDB) [65]. DHDB is then converted to 3-methoxy-4-hydroxy-5-decaprenylbenzoic acid by the S-adenosyl-L-methionine-dependent methyltransferase COQ3 [62], which adopts a classical Rossmann-like fold [66].

While the interaction between COQ3 and COQ6 was supported by several experimental studies [14, 67, 68], their complex remains undetermined. The COQ3-COQ6 interface modeled by AlphaFold2 is mostly restricted to the two α-helices before and after the first core β-strand in COQ3 and the loop region between two C-terminal long α-helices of COQ6 (Figure 3A). Since COQ6 and COQ3 are involved in consecutive steps in CoQ biosynthesis, their interaction in a protein complex could facilitate the transfer of the product of COQ6 to the active site of COQ3 for efficient catalysis. Several missense mutations in COQ6 have been associated with the autosomal recessive disorder of Primary coenzyme Q10 deficiency-6 (OMIM: 614650) [69]. One of them, R360L (sidechain shown in magenta spheres in Figure 3A), involves an arginine residue near the interaction interface. In the structure model, R360 forms a salt bridge with D402 of COQ6 that showed high contact probability with a couple of residues from COQ3 including the positively charged residue R166. The presence of these charged residues among others (such as K202 from COQ6 and D191 from COQ3) in the predicted interface highlight the importance of electrostatic interactions in the COQ3-COG6 complex and elucidates a possible molecular mechanism for Primary coenzyme Q10 deficiency-6.

COQ7 and COQ9

293 COQ7 is an oxidoreductase that catalyzes the hydroxylation of 2-polyprenyl-3-methyl-6-methoxy-1,4-
294 benzoquinol to form 3-demethylubiquinol, the penultimate step in CoQ biosynthesis [70]. While the

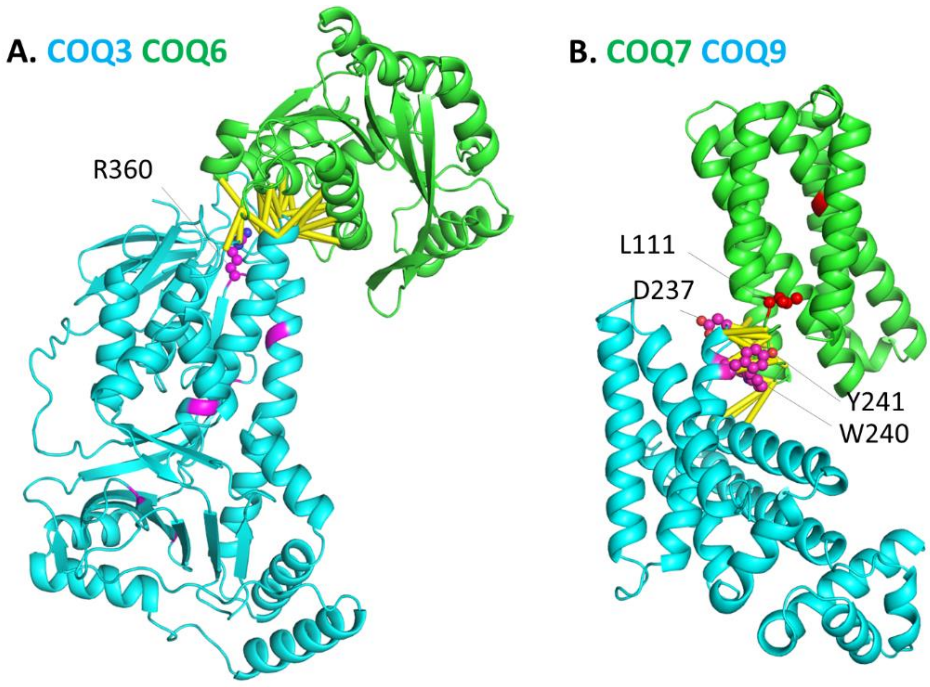

Figure 3. AlphaFold2 structural models of protein complexes in the ubiquinone biosynthesis pathway. **A.** COQ3-COQ6. Missense disease-causing mutations in COQ3 are colored in magenta, and the sidechain of R360 near the interaction interface is shown in spheres. **B.** COQ7-COQ9. Residues involving missense disease-causing mutations in COQ7 are colored in red, and sidechain of L111 near the interaction interface is shown in spheres. Sidechains are shown in magenta spheres for COQ9 residues whose mutagenesis affected the binding to COQ7.

295 structures of human COQ7 and its close homologs have not been solved, HHpred similarity searches
296 suggest that it is evolutionarily related to a large family of heme oxygenases and ribonucleotide
297 reductases with a Ferritin-like helix-bundle fold [70]. AlphaFold2 predicts COQ7 to adopt a helix-bundle
298 fold consisting of six α-helices, consistent with the HHpred results. The lipid-binding protein COQ9
299 functions in CoQ biosynthesis through its interaction with COQ7 and the stabilization of the whole CoQ
300 biosynthetic complex [71].

301 Recent structural studies on COQ9 revealed that it adopts an ancient fold of the bacterial TetR
302 family transcriptional regulators [71]. While the structure of the COQ7-COQ9 complex has not been
303 determined, mutagenesis studies suggested that the interaction site on COQ9 corresponds to a
304 conserved surface patch comprising mostly of the region between the seventh and eighth α-helices [71].

Such experimental results are consistent with our predictions, which mapped the interacting residues from COQ9 to this surface patch (Figure 3B). Several COQ9 residues, such as D237, Y240, and W241 (sidechains shown as magenta spheres in Figure 3B), show high contact probabilities with residues of COQ7. Their mutations (W240K, W240D, Y241K, and D237K) maintained normal melting temperature of

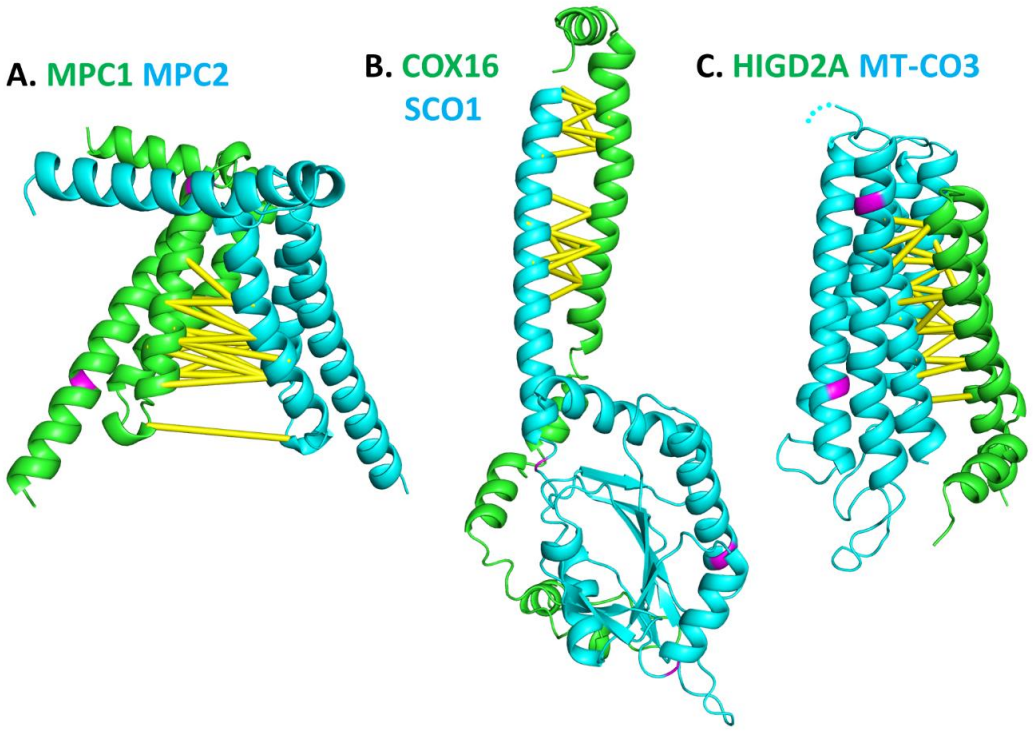

Figure 4. **AlphaFold2 structural models of protein complexes with predicted contacts in transmembrane segments. A.** MPC1-MPC2. **B.** COX16-SCO1. **C.** HIGD2A-MT-CO3. Residues involving missense disease-causing mutations are colored in magenta, and those near the interaction interface have their sidechains shown in spheres.

COQ9, while they abolished the interaction with COQ7 [71]. Our coevolution analysis further suggests the interaction site of COQ7 to be the C-terminal end of the second core α-helix and the loop region between the second and third core α-helices (Figure 3B). Missense mutations in COQ7 have been associated with the disorder of Coenzyme Q10 deficiency, primary 8 (OMIM: 616733) [72, 73]. One such mutation L111P (sidechain shown as red spheres in Figure 3B) is mapped near the interaction interface of COQ7 and COQ9, suggesting that its deleterious effect could be related to the destabilization of the COQ7-COQ9 complex.

Other protein complexes supported by coevolution analysis

Besides the above examples, we identified several high-scoring PPIs supported by experiments. They include the mitochondrial pyruvate carrier complex, a heterodimeric transporter formed by MPC1 and MPC2 [74, 75]. MPC1 and MPC2 are distantly related proteins with three transmembrane segments. HHpred results suggest that MPC proteins (Pfam: MPC) are remotely related to PQ-loop family of sugar transporters [76], and they have been classified in the same Pfam clan (MtN3-like). Consistent with this classification, the AlphaFold2 models of MPC1 and MPC2 adopt the same fold as the PQ-loop transporters [76, 77]. In addition, the right-handed three transmembrane segments of MPC1 and MPC2 form a pseudosymmetric dimer arranged in the same fashion as PQ-loop transporters (Figure 4A). MPC1 and MPC2 exhibit high contact probability mainly between the second transmembrane segment of MPC1 and the first transmembrane segment of MPC2 (Figure 4A). Two other predicted PPIs with contacting residues mainly from transmembrane segments are the pair of COX16 and SCO1 (both CIV assembly factors) [78] (Figure 4B) and the pair of HIGD2A (a CIV assembly factor) and MT-CO3 (a core subunit of CIV) [79, 80] (Figure 4C).

Coevolution analysis identified potential interaction partners for poorly characterized mitochondrial proteins

About 20% of mitochondrial proteins are uncharacterized and do not have clear function annotations [14]. Studies on PPIs by experimental or computational methods can help elucidate their cellular functions [49], as elucidated by the following examples.

Identification of PYURF as a potential subunit of the NDUFAF5 hydroxylase complex

Affinity enrichment mass spectrometry (AE-MS) was used in a previous study to identify PPI of 50 select mitochondrial uncharacterized proteins (MXPs) [14]. One MXP, formerly named C17orf89, was found to interact with NDUFAF5 and play an important role in CI assembly [14]. This protein has since been renamed NDUFAF8. We identified the NDUFAF8-NDUFAF5 complex in our screen with the contact probability of 0.911 by RoseTTAFold and the contact probability of 0.444 by AlphaFold2. We also identified another MXP named PYURF that likely interact with NDUFAF5. The PYURF-NDUFAF5 pair has a AlphaFold2 score of 0.998 and a RoseTTAFold score of 0.959. Predicted interaction with NDUFAF5

suggests that the function of the small uncharacterized protein PYURF could be related to CI assembly as well.

NDUFAF5 adopts a Rossmann-like fold with seven β-strands and belongs to the family of S-adenosylmethionine-dependent methyltransferases [81]. However, instead of being a methyltransferase, it catalyzes the hydroxylation of an arginine residue in of the CI subunit NDUFS7, and this posttranslational modification is crucial in the early stage of CI assembly [81]. As an essential assembly factor, NDUFAF5 has mutations causing the disease of mitochondrial complex I deficiency (nuclear type 16, OMIM: 618238) [82]. Mapping of the predicted interface residues onto the structure models of NDUFAF8, NDUFAF5 and PYURF suggests that NDUFAF8 and PYURF occupy distinct interaction sites on

355 NDUFAF5 (Figure 5A) that are distal from each other. NDUFAF5 (cyan in Figure 5A) residues coevolving
356 with NDUFAF8 (orange in Figure 5A) is mapped to an α-helix before the last β-strand and is far from the
357 active site. A likely pathogenic mutation (F55L, sidechain of F55 shown in blue spheres in Figure 5A) in
358 NDUFAF8 [83] was mapped near its interaction site with NDUFAF5.

359 NDUFAF5 residues coevolving with PYURF (green in Figure 5A) are mapped near the active site

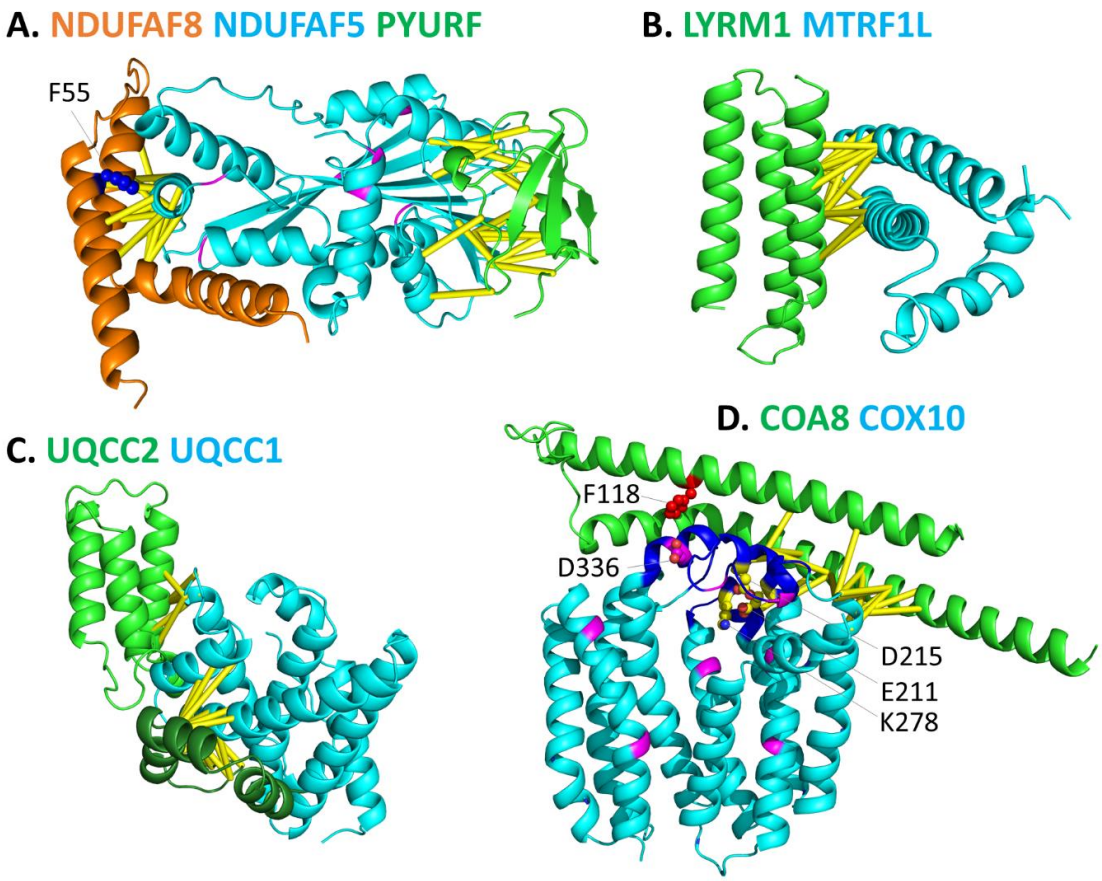

Figure 5. AlphaFold2 structural models of protein complexes involving proteins with poorly characterized proteins. A. NDUFAF8-NDUFAF5-PYURF. Residues involving missense disease-causing mutations of NDUFAF5 are colored in magenta. Sidechain of NDUFAF8 F55 involving a missense disease-causing mutation is shown in blue spheres. **B.** LYRM1-MTRF1L. **C.** UQCC1-UQCC2. **D.** COA8-COX10. The CAP domains of COX10 are colored in blue. COX10 residues involving missense disease-causing mutations are colored in magenta, and those near the interaction interface have their sidechains shown in spheres. D215, E211, and K278 of COX10 (with carbon atoms colored in yellow) are conserved charged residues likely involved in catalysis. The residue (F118) involving a missense disease-causing mutation in COA8 has its sidechain shown in red spheres.

360 to an edge β-strand, the α-helix before it and the Rossmann crossover loop after it. PYURF is a small
361 protein predicted to adopt a Trm112p-like fold [84]. Remote homologs of PYURF include Rieske-like

362 ferredoxins and the Trm112 proteins in archaea and eukaryotes [85]. Interestingly, Trm112 proteins are
363 subunits of multifunctional methyltransferases that form complexes with and activate the tRNA
364 methyltransferases such as Trm11 and Trm9. Structures of the recently solved archaeal Trm112-Trm11
365 complexes [86, 87] revealed similar interaction sites as predicted by AlphaFold2 in PYURF and NDUFAF5.
366 Our coevolution analysis as well as the remote homology of PYURF to Trm112 proteins suggests that
367 PYURF is a subunit that forms a complex with the NDUFAF5 hydroxylase and could play an essential role
368 in the catalytic activity. The involvement of PYURF in CI assembly was supported by a recent genome-
369 wide CRISPR death screen that shows the phenotype of CI deficiency when the PYURF gene was
370 disrupted [88].

371

372 **Identification of potential interaction partners and interfaces of LYRM proteins**

373 The LYRM (leucine/tyrosine/arginine motif) family proteins (also called LYR proteins) are small proteins
374 with diverse functions. LYRM proteins are found to associate with OXPHOS complexes as CI subunits
375 (NDUFB9 (LYRM3) and NDUFA6 (LYRM6)) , CI assembly factor (LYRM2) [89], CII assembly factors
376 (SDHAF1 (LYRM8) and SDHAF3 (LYRM10)) [90], CIII assembly factor (LYRM7) [91], and CV assembly
377 factor (FMC1) [49]. Other LYRM proteins are involved in Fe–S cluster biosynthesis (LYRM4, also known as
378 ISD11) [89], mitoribosome assembly and translation (AltMIEF1 (also named AltMiD51 or L0R8F8)) [92-
379 94], and regulation of the electron transferring flavoprotein complex (LYRM5) [50, 51]. LYRM proteins
380 contain the LYR domain that adopts a three-helix bundle fold [93]. Our top predictions include several
381 known interactions involving LYRM proteins, such as FMC1-ATPAF2 (described above), LYRM4-NSF1 that
382 is important for Fe–S cluster biosynthesis, NDUFB9/LYRM3-NDUFB3 in CI, and NDUFA5-NDUFS3 in CI.
383 We identified NDUFA5 (Pfam: ETC_C1_NDUFA5), an accessory subunit of CI, as a distant member of the
384 LYRM family. Remote homology between NDUFAF5 and other LYRM proteins is supported by HHpred
385 search results, the presence of LYR motif in NDUFA5, and the three-helix bundle fold of NDUFA5 in CI
386 structures [95, 96].

387 The functions of LYRM1 and LYRM9, two LYR-domain containing proteins, remain to be
388 elucidated [94, 97]. In our PPI screen LYRM1 is predicted to interact with MTRF1L (mitochondrial
389 translational release factor 1-like) (AlphaFold2 score: 0.902, RoseTTAFold score: 0.955). MTRF1L is
390 responsible for translation termination upon detection of certain termination codons [93]. In the
391 structural model of the LYRM1-MTRF1L complex, the interface residues in LYRM1 are mapped to the

first α-helix of the LYR domain with a three-helix bundle fold. The interface residues from MTRF1L are mapped to a N-terminal α-helical domain [98] (Figure 5B). RoseTTAFold also reported modestly high contact probability scores for the LYRM1-LYRM9 pair (score: 0.862) and the LYRM9-MRPL57 (mitochondrial ribosomal protein 63) pair (score: 0.907), albeit the AlphaFold2 scores for them are low (below 0.2). AltMIEF1, a small protein encoded by an open reading frame in the 5' region of the MIEF1 gene, was the only LYRM protein previously found to function in translation and ribosomal assembly [92]. Our study suggests that the uncharacterized LYRM1 and LYRM9 could also play important roles in regulation of translation and other functions related to mitochondrial ribosomes.

HHpred sequence similarity searches and Alphafold2 structural models also suggest the presence of a divergent LYR domain in UQCC2, a small protein that interacts with UQCC1 and functions in the assembly of CIII [99]. UQCC1 and UQCC2 are orthologs to fungal CIII assembly factors Cbp3p and Cbp6p, respectively. AlphaFold2 structural model predicts that UQCC1 adopts an α-helical bundle fold with six major α-helices, consistent with the experimental structure of its bacterial homolog [100]. Our analysis predicted high probability for interaction between UQCC1 and UQCC2. The predicted UQCC1-UQCC2 interface in UQCC2 is mapped to the N-terminal LYR domain (colored in lighter green in Figure 5C) that adopts a three-helix bundle fold, as well as regions C-terminal (colored in darker green in Figure 5C) to the LYR domain. Our findings suggest UQCC2 as a new member of the LYRM family that functions in CIII assembly [101].

Identification of COA8 as a potential subunit of the COX10 enzyme complex

COA8, previously named APOPT1 (apoptogenic-1), was recently discovered to function in CIV assembly. Loss-of-function mutations in COA8 [102, 103] caused the disease phenotype of leukoencephalopathy associated with mitochondrial cytochrome *c* oxidase (COX) deficiency, and gene knockout of COA8 in mouse [104] and knockdown in *Drosophila* [105] resulted in reduced cytochrome c oxidase activity and levels. The functional mechanism of COA8 in CIV assembly has not been revealed. Our analysis identified high contact probability scores (>0.99 by both AlphaFold2 and RoseTTAFold) between COA8 and COX10. As an essential CIV assembly factor, COX10 catalyzes the farnesylation of the vinyl group of heme B to produce heme O, the first step of the mitochondrial heme A biosynthesis required for CIV assembly.

The predicted interaction between COA8 and COX10 suggests that COA8 could be a subunit of the heme farnesyltransferase complex. The AlphFold2 model of the COA8-COX10 complex revealed that

COA8 forms two long α-helices, and COX10 is a multi-pass transmembrane protein with 9 transmembrane segments, consistent with its homology to proteins in the UbiA prenyltransferase family (Pfam: UbiA) [106, 107] (Figure 5D). The predicted interface residues are mainly mapped to the middle of the second α-helix in COA8 and in between the second and third transmembrane segments of COX10 (Figure 5D), which is part of the cap domains (colored in blue in Figure 5D) previously defined as the connections between transmembrane helices near the active site [107]. Several missense mutations of COX10 have been reported in mitochondrial disorders characterized by cytochrome c oxidase deficiency. Most of the pathogenic mutations in COX10 (colored magenta) are not close to the predicted interface between COA8 and COX10. Most of the pathogenic mutations in COA8 resulted in loss-of-function by introducing premature stop codons or frame shifts [108]. One recently found missense pathogenic mutation in COA8, F118S [103] (sidechain shown in red spheres in Figure 5D), lies in the predicted interface of the COA8-COX10 complex and is close to residue D336 in COX10 that also harbors pathogenetic mutations (D336G and D336V) [109]. Our results suggest that those pathogenetic mutations may cause disease phenotypes by disrupting the proper interaction between COA8 and COX10.

Materials and methods

Generation of protein sequence alignments

The eukaryotic proteomes were downloaded from the NCBI genome database. It consists of 49102568 proteins from 2568 representative or reference genomes. The list of human mitochondrial proteins was obtained from the MitoCarta3.0 database [17]and their sequences were obtained from the UniProt database [42]. For each human mitochondrial protein (called a target protein), the corresponding orthologous group at the eukaryotes level in OrthoDB [110] was identified. This group of orthologous proteins were clustered by CD-HIT [111] at 40% identity level. For each CD-HIT cluster, we selected one representative sequence that showed the best BLAST [112] score to the target protein. The representative sequences were then used as queries to search against the NCBI eukaryotic proteomes to identify homologous proteins to the target protein.

For each organism we identified up to one protein (best hit, if available) that shows the highest sequence similarity to the target protein from the combined BLAST hits found by multiple representative sequences. These best hits together with the target protein are considered as the expanded orthologous group for the target protein. This group of sequences were subject to multiple sequence alignment by MAFFT (with the --auto option) [113]. To construct the joint alignment for any two target proteins, we identified the intersection of the organisms for their orthologous groups and merged the sub-alignments containing only these organisms for the two proteins. Positions that are gaps in the human proteins were removed. Predictions of coevolving residues by deep learning neural network were carried out for protein pairs that have a combined length of less than 1500 amino acid residues.

Selection of top predictions

We applied the RoseTTAFold 2-track model, a faster but inferior model than the 3-track one, to each concatenated alignment [ref]. RoseTTAFold predicts the probability density for the distance between each residue pair. We summed up the probability for distance bins below 10Å and used it as the contact probability (ranging from 0 to 1) of a residue pair. Residue pairs involving the last 10 residues in the first sequence and the first 10 residues of the second sequence were discarded as they represent artificially continuous segment in the merged alignment and could have biased scores. To further improve the performance, we also removed residue pairs involving the mitochondrial target sequences (defined in the UniProt Feature fields) in either protein. The best contact probability of the remaining residue pairs was reported as the contact probability score for the protein pair.

We observed that some proteins enriched in top hits with high contact probability scores could be false positives. To downgrade the influence of these proteins, we designed a score to penalize proteins involving many high-scoring pairs. For each protein, we recorded up to 20 contact probability scores of its highest scoring pairs with contact probability scores more than 0.5. If the number of pairs (N) with scores more than 0.5 for a protein is less than 20, we extended the list of its top scores by adding $20 - N$ pseudo scores of 0.5 at the end, such that the list consists of 20 numbers. The average of the 20 numbers in this score list is then used a penalty weight, and its minimum value is 0.5. The new adjusted score for any protein pair is calculated as the original score divided by the maximum of the two penalty weights of the two proteins.

We selected the top 200 pairs ranked by the original contact probability score and the top 350 pairs ranked by the adjusted score. Their union consists of 401 pairs of proteins. AlphaFold2 was used to build structural models for these 401 pairs. AlphaFold2 also produces probability distribution for residue-residue distances, and we computed contact probability as the sum of probability for distance bins below 10 Å. The highest contact probability between residues of two proteins was used to indicate the contact probability for this protein pair.

Performance evaluation

For precision-recall analysis, we excluded protein pairs that are potentially paralogs (defined as sequence identity >40%) as they could have artificially high contact scores by sharing sequences in their alignments. For the rest of the pairs, we considered a pair of proteins to be positive cases of PPI if it has been reported by three or more experimental studies in the BioGRID database, or if both proteins have significant sequence similarity (>30% sequence identity) to two interacting chains in an experimentally determined structure of a protein complex. A total of 1563 positive pairs (473 supported by BioGRID and 1193 supported by experimental complexes) were found in our dataset. Negative cases of PPIs are those protein pairs without experimental structures and were never reported in the BioGRID database. For precision-recall analysis, we used the set of positive pairs and a randomly selected subset of negative pairs that are 100 times more than the positive pairs.

For the PPIs supported by known experimental complex structures, we evaluated the agreement of top-scoring residue pairs provided by RoseTTAFold against the experimentally determined interface residues. BLAST was used to detect proteins with known structures for each human mitochondrial protein. The BLAST alignment was used to map the residues in a target protein to those residues in its homologs with known structures. A residue pair predicted by RoseTTAFold (defined as those with contact probability >0.5) is considered to lie in the interaction interface of an experimental complex if the corresponding aligned residue pair in the complex structure has an inter-residue non-hydrogen atom pair within 10Å. The fraction of true interacting residue pairs was then calculated as the number of residue pairs lying in the experimental interface divided by the total number of residue pairs reported by RoseTTAFold.

Supplementary materials

Supplementary Table S1. RoseTTAFold and AlphaFold2 scores of top-scoring protein pairs by RoseTTAFold.

Supplementary Table S2. RoseTTAFold and AlphaFold2 scores for protein pairs with BioGRID interaction count ≥ 3.

Acknowledgement

We would like to thank Minkyung Beak for sharing the RoseTTAFold script prior to its publication, Aaron Guillory for help in maintaining the computer cluster used in this study, Lisa N. Kinch and Nick V. Grishin for helpful discussions and suggestions. Qian Cong is a Southwestern Medical Foundation scholar.

Conflicts of interest

None declared.

Figure legends

Figure 1. **A**. A flowchart of the procedure used in this study to perform PPI prediction and structural modeling. **B**. The precision-recall curves of three methods in detecting positive PPIs of mitochondrial protein pairs. DCA - Direct Coupling Analysis method, RoseTTAFold (no transit peptide) is the RoseTTAFold method used on residue pairs excluding those involving any residue in mitochondrial transit peptides. **C.** Counts of high-scoring interacting partners (contact probability score >0.9) (blue bars) for proteins with the highest number of such partners. The red bars (if any) show the number of true positives among the top-scoring pairs. **D.** Precision-recall curves for RoseTTAFold and AlphaFold2 on the 401 pairs of selected PPI predictions. **E.** Distribution of the fraction of predicted contacts that are mapped to experimentally determined complex interfaces. **F.** The left panel shows the AlphaFold2 model of MT-CO2 (magenta) and MT-CO3 (rainbow) complex, with the top 20 RoseTTAFold-predicted contact pairs shown as yellow lines connecting their C-α atoms. The right panel shows the AlphaFold2 model of the MT-CO1 (gray), MT-CO2 (magenta) and MT-CO3 (rainbow) complex. Top contact pairs are

shown in yellow, cyan, and red lines for MT-CO2/MT-CO3, MT-CO1/MT-CO2 and MT-CO1/MT-CO3 pairs, respectively.

Figure 2. AlphaFold2 structural models of protein complexes without experimentally determined 3D structures. **A**. CHCHD4-AIFM1. Residues involving missense disease-causing mutations are colored in magenta, and those near the interaction interface have their sidechains shown in spheres. Sidechains of the two hydrophobic residues of CHCHD4 interacting with AIFM1 Y562 are shown in green spheres. **B**. TRUB2-MTERF3. **C**. FMC1-ATPAF2. Residues involving missense disease-causing mutations are colored in magenta, and those near the interaction interface have their sidechains shown in spheres **D**. COX19-COX11. Sidechains of the YL and FM signature sequences are shown in magenta spheres. The CHCH domain is shown in light green and the C-terminal two helices are shown in dark green. **E.** NDUFAF1-ECSIT. Residues of missense disease-causing mutations are colored in magenta, and those near the interaction interface are shown in spheres.

Figure 3. AlphaFold2 structural models of protein complexes in the ubiquinone biosynthesis pathway. **A.** COQ3-COQ6. Missense disease-causing mutations in COQ3 are colored in magenta, and the sidechain of R360 near the interaction interface is shown in spheres. **B.** COQ7-COQ9. Residues involving missense disease-causing mutations in COQ7 are colored in red, and sidechain of L111 near the interaction interface is shown in spheres. Sidechains are shown in magenta spheres for COQ9 residues whose mutagenesis affected the binding to COQ7.

Figure 4. **AlphaFold2 structural models of protein complexes with predicted contacts in transmembrane segments. A.** MPC1-MPC2. **B.** COX16-SCO1. **C.** HIGD2A-MT-CO3. Residues involving missense disease-causing mutations are colored in magenta, and those near the interaction interface have their sidechains shown in spheres.

Figure 5. AlphaFold2 structural models of protein complexes involving proteins with poorly characterized proteins. A. NDUFAF8-NDUFAF5-PYURF. Residues involving missense disease-causing mutations of NDUFAF5 are colored in magenta. Sidechain of NDUFAF8 F55 involving a missense disease-causing mutation is shown in blue spheres. **B.** LYRM1-MTRF1L. **C.** UQCC1-UQCC2. **D.** COA8-COX10. The CAP domains of COX10 are colored in blue. COX10 residues involving missense disease-causing mutations are colored in magenta, and those near the interaction interface have their sidechains shown in spheres. D215, E211, and K278 of COX10 (with carbon atoms colored in yellow) are conserved

564 charged residues likely involved in catalysis. The residue (F118) involving a missense disease-causing
565 mutation in COA8 has its sidechain shown in red spheres.

566

567 ## References

568 1. Baek, M., et al., *Accurate prediction of protein structures and interactions using a three-track*
569 *neural network.* Science, 2021.
570 2. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021.
571 3. Kinch, L.N., et al., *Topology evaluation of models for difficult targets in the 14th round of the*
572 *critical assessment of protein structure prediction.* Proteins, 2021.
573 4. Tunyasuvunakool, K., et al., *Highly accurate protein structure prediction for the human proteome.*
574 Nature, 2021.
575 5. Cong, Q., et al., *Protein interaction networks revealed by proteome coevolution.* Science, 2019.
576 **365**(6449): p. 185-189.
577 6. Osellame, L.D., T.S. Blacker, and M.R. Duchen, *Cellular and molecular mechanisms of*
578 *mitochondrial function.* Best Pract Res Clin Endocrinol Metab, 2012. **26**(6): p. 711-23.
579 7. West, A.P., et al., *Mitochondrial DNA stress primes the antiviral innate immune response.* Nature,
580 2015. **520**(7548): p. 553-7.
581 8. Green, D.R. and G. Kroemer, *The pathophysiology of mitochondrial cell death.* Science, 2004.
582 **305**(5684): p. 626-9.
583 9. Pagliarini, D.J. and J. Rutter, *Hallmarks of a new era in mitochondrial biochemistry.* Genes Dev,
584 2013. **27**(24): p. 2615-27.
585 10. Pfanner, N., B. Warscheid, and N. Wiedemann, *Mitochondrial proteins: from biogenesis to*
586 *functional networks.* Nat Rev Mol Cell Biol, 2019. **20**(5): p. 267-284.
587 11. Martin, W.F., S. Garg, and V. Zimorski, *Endosymbiotic theories for eukaryote origin.* Philos Trans
588 R Soc Lond B Biol Sci, 2015. **370**(1678): p. 20140330.
589 12. Moutaoufik, M.T., et al., *Rewiring of the Human Mitochondrial Interactome during Neuronal*
590 *Reprogramming Reveals Regulators of the Respirasome and Neurogenesis.* iScience, 2019. **19**: p.
591 1114-1132.
592 13. Koshiba, T. and H. Kosako, *Mass spectrometry-based methods for analysing the mitochondrial*
593 *interactome in mammalian cells.* J Biochem, 2020. **167**(3): p. 225-231.
594 14. Floyd, B.J., et al., *Mitochondrial Protein Interaction Mapping Identifies Regulators of Respiratory*
595 *Chain Function.* Mol Cell, 2016. **63**(4): p. 621-632.
596 15. Antonicka, H., et al., *A High-Density Human Mitochondrial Proximity Interaction Network.* Cell
597 Metab, 2020. **32**(3): p. 479-497 e9.
598 16. Huttlin, E.L., et al., *Architecture of the human interactome defines protein communities and*
599 *disease networks.* Nature, 2017. **545**(7655): p. 505-509.
600 17. Rath, S., et al., *MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle*
601 *localization and pathway annotations.* Nucleic Acids Res, 2021. **49**(D1): p. D1541-D1547.
602 18. Thompson, K., et al., *Recent advances in understanding the molecular genetic basis of*
603 *mitochondrial disease.* J Inherit Metab Dis, 2020. **43**(1): p. 36-50.
604 19. Oughtred, R., et al., *The BioGRID database: A comprehensive biomedical resource of curated*
605 *protein, genetic, and chemical interactions.* Protein Sci, 2021. **30**(1): p. 187-200.
606 20. Amunts, A., et al., *Ribosome. The structure of the human mitochondrial ribosome.* Science, 2015.
607 **348**(6230): p. 95-98.

21. Koripella, R.K., et al., *Structures of the human mitochondrial ribosome bound to EF-G1 reveal distinct features of mitochondrial translation elongation.* Nat Commun, 2020. **11**(1): p. 3830.
22. Grigorieff, N., *Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 A in ice.* J Mol Biol, 1998. **277**(5): p. 1033-46.
23. Signes, A. and E. Fernandez-Vizarra, *Assembly of mammalian oxidative phosphorylation complexes I-V and supercomplexes.* Essays Biochem, 2018. **62**(3): p. 255-270.
24. Zhou, Q., et al., *Thiabendazole inhibits ubiquinone reduction activity of mitochondrial respiratory complex II via a water molecule mediated binding feature.* Protein Cell, 2011. **2**(7): p. 531-42.
25. Huang, J. and M.E. Fraser, *Tartryl-CoA inhibits succinyl-CoA synthetase.* Acta Crystallogr F Struct Biol Commun, 2020. **76**(Pt 7): p. 302-308.
26. Patel, M.S. and L.G. Korotchkina, *Regulation of the pyruvate dehydrogenase complex.* Biochem Soc Trans, 2006. **34**(Pt 2): p. 217-22.
27. Henriques, B.J., et al., *Electron transfer flavoprotein and its role in mitochondrial energy metabolism in health and disease.* Gene, 2021. **776**: p. 145407.
28. Burrage, L.C., et al., *Branched-chain amino acid metabolism: from rare Mendelian diseases to more common disorders.* Hum Mol Genet, 2014. **23**(R1): p. R1-8.
29. Brown, J.R. and W.F. Doolittle, *Gene descent, duplication, and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases.* J Mol Evol, 1999. **49**(4): p. 485-95.
30. Qi, L., et al., *Cryo-EM structure of the human mitochondrial translocase TIM22 complex.* Cell Res, 2021. **31**(3): p. 369-372.
31. Susin, S.A., et al., *Molecular characterization of mitochondrial apoptosis-inducing factor.* Nature, 1999. **397**(6718): p. 441-6.
32. Hangen, E., et al., *Life with or without AIF.* Trends Biochem Sci, 2010. **35**(5): p. 278-87.
33. Miramar, M.D., et al., *NADH oxidase activity of mitochondrial apoptosis-inducing factor.* J Biol Chem, 2001. **276**(19): p. 16391-8.
34. Hangen, E., et al., *Interaction between AIF and CHCHD4 Regulates Respiratory Chain Biogenesis.* Mol Cell, 2015. **58**(6): p. 1001-14.
35. Dickson-Murray, E., et al., *The Mia40/CHCHD4 Oxidative Folding System: Redox Regulation and Signaling in the Mitochondrial Intermembrane Space.* Antioxidants (Basel), 2021. **10**(4).
36. Banci, L., et al., *MIA40 is an oxidoreductase that catalyzes oxidative protein folding in mitochondria.* Nat Struct Mol Biol, 2009. **16**(2): p. 198-206.
37. Banci, L., et al., *The coiled coil-helix-coiled coil-helix proteins may be redox proteins.* FEBS Lett, 2009. **583**(11): p. 1699-702.
38. Modjtahedi, N., et al., *Mitochondrial Proteins Containing Coiled-Coil-Helix-Coiled-Coil-Helix (CHCH) Domains in Health and Disease.* Trends Biochem Sci, 2016. **41**(3): p. 245-260.
39. Bano, D. and J.H.M. Prehn, *Apoptosis-Inducing Factor (AIF) in Physiology and Disease: The Tale of a Repented Natural Born Killer.* EBioMedicine, 2018. **30**: p. 29-37.
40. Pei, J. and N.V. Grishin, *The DBSAV Database: Predicting Deleteriousness of Single Amino Acid Variations in the Human Proteome.* J Mol Biol, 2021. **433**(11): p. 166915.
41. Landrum, M.J., et al., *ClinVar: improvements to accessing data.* Nucleic Acids Res, 2020. **48**(D1): p. D835-D844.
42. UniProt, C., *UniProt: the universal protein knowledgebase in 2021.* Nucleic Acids Res, 2021. **49**(D1): p. D480-D489.
43. Park, C.B., et al., *MTERF3 is a negative regulator of mammalian mtDNA transcription.* Cell, 2007. **130**(2): p. 273-85.
44. Wredenberg, A., et al., *MTERF3 regulates mitochondrial ribosome biogenesis in invertebrates and mammals.* PLoS Genet, 2013. **9**(1): p. e1003178.

45. Roberti, M., et al., *MTERF3, the most conserved member of the mTERF-family, is a modular factor involved in mitochondrial protein synthesis.* Biochim Biophys Acta, 2006. **1757**(9-10): p. 1199-206.
46. Spahr, H., et al., *Structure of mitochondrial transcription termination factor 3 reveals a novel nucleic acid-binding domain.* Biochem Biophys Res Commun, 2010. **397**(3): p. 386-90.
47. Antonicka, H. and E.A. Shoubridge, *Mitochondrial RNA Granules Are Centers for Posttranscriptional RNA Processing and Ribosome Biogenesis.* Cell Rep, 2015. **10**(6): p. 920-932.
48. Antonicka, H., et al., *A pseudouridine synthase module is essential for mitochondrial protein synthesis and cell viability.* EMBO Rep, 2017. **18**(1): p. 28-38.
49. Li, Y., et al., *CLIC, a tool for expanding biological pathways based on co-expression across thousands of datasets.* PLoS Comput Biol, 2017. **13**(7): p. e1005653.
50. Angerer, H., *Eukaryotic LYR Proteins Interact with Mitochondrial Protein Complexes.* Biology (Basel), 2015. **4**(1): p. 133-50.
51. Tang, J.X., et al., *Mitochondrial OXPHOS Biogenesis: Co-Regulation of Protein Synthesis, Import, and Assembly Pathways.* Int J Mol Sci, 2020. **21**(11).
52. De Meirleir, L., et al., *Respiratory chain complex V deficiency due to a mutation in the assembly gene ATP12.* J Med Genet, 2004. **41**(2): p. 120-4.
53. Nobrega, M.P., et al., *Characterization of COX19, a widely distributed gene required for expression of mitochondrial cytochrome oxidase.* J Biol Chem, 2002. **277**(43): p. 40206-11.
54. Bode, M., et al., *Redox-regulated dynamic interplay between Cox19 and the copper-binding protein Cox11 in the intermembrane space of mitochondria facilitates biogenesis of cytochrome c oxidase.* Mol Biol Cell, 2015. **26**(13): p. 2385-401.
55. Palumaa, P., *Copper chaperones. The concept of conformational control in the metabolism of copper.* FEBS Lett, 2013. **587**(13): p. 1902-10.
56. Banci, L., et al., *Solution structure of Cox11, a novel type of beta-immunoglobulin-like fold involved in CuB site formation of cytochrome c oxidase.* J Biol Chem, 2004. **279**(33): p. 34833-9.
57. Kopp, E., et al., *ECSIT is an evolutionarily conserved intermediate in the Toll/IL-1 signal transduction pathway.* Genes Dev, 1999. **13**(16): p. 2059-71.
58. Xiao, C., et al., *Ecsit is required for Bmp signaling and mesoderm formation during mouse embryogenesis.* Genes Dev, 2003. **17**(23): p. 2933-49.
59. Vogel, R.O., et al., *Cytosolic signaling protein Ecsit also localizes to mitochondria where it interacts with chaperone NDUFAF1 and functions in complex I assembly.* Genes Dev, 2007. **21**(5): p. 615-24.
60. Carneiro, F.R.G., et al., *An Essential Role for ECSIT in Mitochondrial Complex I Assembly and Mitophagy in Macrophages.* Cell Rep, 2018. **22**(10): p. 2654-2666.
61. Fassone, E., et al., *Mutations in the mitochondrial complex I assembly factor NDUFAF1 cause fatal infantile hypertrophic cardiomyopathy.* J Med Genet, 2011. **48**(10): p. 691-7.
62. Jonassen, T. and C.F. Clarke, *Isolation and functional expression of human COQ3, a gene encoding a methyltransferase required for ubiquinone biosynthesis.* J Biol Chem, 2000. **275**(17): p. 12381-7.
63. Ozeir, M., et al., *Coenzyme Q biosynthesis: Coq6 is required for the C5-hydroxylation reaction and substrate analogs rescue Coq6 deficiency.* Chem Biol, 2011. **18**(9): p. 1134-42.
64. Bentinger, M., M. Tekle, and G. Dallner, *Coenzyme Q--biosynthesis and functions.* Biochem Biophys Res Commun, 2010. **396**(1): p. 74-9.
65. Heeringa, S.F., et al., *COQ6 mutations in human patients produce nephrotic syndrome with sensorineural deafness.* J Clin Invest, 2011. **121**(5): p. 2013-24.
66. Medvedev, K.E., et al., *A Fifth of the Protein World: Rossmann-like Proteins as an Evolutionarily Successful Structural unit.* J Mol Biol, 2021. **433**(4): p. 166788.

67. Wan, C., et al., *Panorama of ancient metazoan macromolecular complexes.* Nature, 2015. **525**(7569): p. 339-44.
68. Yen, H.C., et al., *Characterization of human mitochondrial PDSS and COQ proteins and their roles in maintaining coenzyme Q10 levels and each other's stability.* Biochim Biophys Acta Bioenerg, 2020. **1861**(7): p. 148192.
69. Acosta, M.J., et al., *Coenzyme Q biosynthesis in health and disease.* Biochim Biophys Acta, 2016. **1857**(8): p. 1079-1085.
70. Nakai, D., et al., *Mouse homologue of coq7/clk-1, longevity gene in Caenorhabditis elegans, is essential for coenzyme Q synthesis, maintenance of mitochondrial integrity, and neurogenesis.* Biochem Biophys Res Commun, 2001. **289**(2): p. 463-71.
71. Lohman, D.C., et al., *Mitochondrial COQ9 is a lipid-binding protein that associates with COQ7 to enable coenzyme Q biosynthesis.* Proc Natl Acad Sci U S A, 2014. **111**(44): p. E4697-705.
72. Freyer, C., et al., *Rescue of primary ubiquinone deficiency due to a novel COQ7 defect using 2,4-dihydroxybensoic acid.* J Med Genet, 2015. **52**(11): p. 779-83.
73. Wang, Y., et al., *Pathogenicity of two COQ7 mutations and responses to 2,4-dihydroxybenzoate bypass treatment.* J Cell Mol Med, 2017. **21**(10): p. 2329-2343.
74. Herzig, S., et al., *Identification and functional expression of the mitochondrial pyruvate carrier.* Science, 2012. **337**(6090): p. 93-6.
75. Bricker, D.K., et al., *A mitochondrial pyruvate carrier required for pyruvate uptake in yeast, Drosophila, and humans.* Science, 2012. **337**(6090): p. 96-100.
76. Lee, Y., et al., *Structural basis for the facilitative diffusion mechanism by SemiSWEET transporter.* Nat Commun, 2015. **6**: p. 6112.
77. Xu, Y., et al., *Structures of bacterial homologues of SWEET transporters in two distinct conformations.* Nature, 2014. **515**(7527): p. 448-452.
78. Aich, A., et al., *COX16 promotes COX2 metallation and assembly during respiratory complex IV biogenesis.* Elife, 2018. **7**.
79. Hock, D.H., et al., *HIGD2A is Required for Assembly of the COX3 Module of Human Mitochondrial Complex IV.* Mol Cell Proteomics, 2020. **19**(7): p. 1145-1160.
80. Timon-Gomez, A., et al., *Distinct Roles of Mitochondrial HIGD1A and HIGD2A in Respiratory Complex and Supercomplex Biogenesis.* Cell Rep, 2020. **31**(5): p. 107607.
81. Rhein, V.F., et al., *NDUFAF5 Hydroxylates NDUFS7 at an Early Stage in the Assembly of Human Complex I.* J Biol Chem, 2016. **291**(28): p. 14851-60.
82. Simon, M.T., et al., *Novel mutations in the mitochondrial complex I assembly gene NDUFAF5 reveal heterogeneous phenotypes.* Mol Genet Metab, 2019. **126**(1): p. 53-63.
83. Alston, C.L., et al., *Pathogenic Bi-allelic Mutations in NDUFAF8 Cause Leigh Syndrome with an Isolated Complex I Deficiency.* Am J Hum Genet, 2020. **106**(1): p. 92-101.
84. Cheng, H., et al., *ECOD: an evolutionary classification of protein domains.* PLoS Comput Biol, 2014. **10**(12): p. e1003926.
85. Purushothaman, S.K., et al., *Trm11p and Trm112p are both required for the formation of 2-methylguanosine at position 10 in yeast tRNA.* Mol Cell Biol, 2005. **25**(11): p. 4359-70.
86. van Tran, N., et al., *Evolutionary insights into Trm112-methyltransferase holoenzymes involved in translation between archaea and eukaryotes.* Nucleic Acids Res, 2018. **46**(16): p. 8483-8499.
87. Wang, C., et al., *Structural and functional insights into Archaeoglobus fulgidus m2G10 tRNA methyltransferase Trm11 and its Trm112 activator.* Nucleic Acids Res, 2020. **48**(19): p. 11068-11082.
88. Arroyo, J.D., et al., *A Genome-wide CRISPR Death Screen Identifies Genes Essential for Oxidative Phosphorylation.* Cell Metab, 2016. **24**(6): p. 875-885.

89. Boniecki, M.T., et al., *Structure and functional dynamics of the mitochondrial Fe/S cluster synthesis complex.* Nat Commun, 2017. **8**(1): p. 1287.
90. Na, U., et al., *The LYR factors SDHAF1 and SDHAF3 mediate maturation of the iron-sulfur subunit of succinate dehydrogenase.* Cell Metab, 2014. **20**(2): p. 253-66.
91. Sanchez, E., et al., *LYRM7/MZM1L is a UQCRFS1 chaperone involved in the last steps of mitochondrial Complex III assembly in human cells.* Biochim Biophys Acta, 2013. **1827**(3): p. 285-93.
92. Rathore, A., et al., *MIEF1 Microprotein Regulates Mitochondrial Translation.* Biochemistry, 2018. **57**(38): p. 5564-5575.
93. Brown, A., et al., *Structures of the human mitochondrial ribosome in native states of assembly.* Nat Struct Mol Biol, 2017. **24**(10): p. 866-869.
94. Dibley, M.G., et al., *The Mitochondrial Acyl-carrier Protein Interaction Network Highlights Important Roles for LYRM Family Members in Complex I and Mitoribosome Assembly.* Mol Cell Proteomics, 2020. **19**(1): p. 65-77.
95. Kampjut, D. and L.A. Sazanov, *The coupling mechanism of mammalian respiratory complex I.* Science, 2020. **370**(6516).
96. Gu, J., et al., *The architecture of the mammalian respirasome.* Nature, 2016. **537**(7622): p. 639-43.
97. Angerer, H., *The superfamily of mitochondrial Complex1_LYR motif-containing (LYRM) proteins.* Biochem Soc Trans, 2013. **41**(5): p. 1335-41.
98. Shin, D.H., et al., *Structural analyses of peptide release factor 1 from Thermotoga maritima reveal domain flexibility required for its interaction with the ribosome.* J Mol Biol, 2004. **341**(1): p. 227-39.
99. Tucker, E.J., et al., *Mutations in the UQCC1-interacting protein, UQCC2, cause human complex III deficiency associated with perturbed cytochrome b protein expression.* PLoS Genet, 2013. **9**(12): p. e1004034.
100. Ndi, M., et al., *Structural basis for the interaction of the chaperone Cbp3 with newly synthesized cytochrome b during mitochondrial respiratory chain assembly.* J Biol Chem, 2019. **294**(45): p. 16663-16671.
101. Fernandez-Vizarra, E. and M. Zeviani, *Mitochondrial complex III Rieske Fe-S protein processing and assembly.* Cell Cycle, 2018. **17**(6): p. 681-687.
102. Sharma, S., et al., *Cavitating Leukoencephalopathy With Posterior Predominance Caused by a Deletion in the APOPT1 Gene in an Indian Boy.* J Child Neurol, 2018. **33**(6): p. 428-431.
103. Melchionda, L., et al., *Mutations in APOPT1, encoding a mitochondrial protein, cause cavitating leukoencephalopathy with cytochrome c oxidase deficiency.* Am J Hum Genet, 2014. **95**(3): p. 315-25.
104. Signes, A., et al., *APOPT1/COA8 assists COX assembly and is oppositely regulated by UPS and ROS.* EMBO Mol Med, 2019. **11**(1).
105. Brischigliaro, M., et al., *Knockdown of APOPT1/COA8 Causes Cytochrome c Oxidase Deficiency, Neuromuscular Impairment, and Reduced Resistance to Oxidative Stress in Drosophila melanogaster.* Front Physiol, 2019. **10**: p. 1143.
106. Ren, S., et al., *Structural and Functional Insights into an Archaeal Lipid Synthase.* Cell Rep, 2020. **33**(3): p. 108294.
107. Cheng, W. and W. Li, *Structural insights into ubiquinone biosynthesis in membranes.* Science, 2014. **343**(6173): p. 878-81.
108. Hedberg-Oldfors, C., et al., *COX deficiency and leukoencephalopathy due to a novel homozygous APOPT1/COA8 mutation.* Neurol Genet, 2020. **6**(4): p. e464.

797 109. Antonicka, H., et al., *Mutations in COX10 result in a defect in mitochondrial heme A biosynthesis*
798 *and account for multiple, early-onset clinical phenotypes associated with isolated COX deficiency.*
799 Hum Mol Genet, 2003. **12**(20): p. 2693-702.
800 110. Kriventseva, E.V., et al., *OrthoDB v10: sampling the diversity of animal, plant, fungal, protist,*
801 *bacterial and viral genomes for evolutionary and functional annotations of orthologs.* Nucleic
802 Acids Res, 2019. **47**(D1): p. D807-D811.
803 111. Fu, L., et al., *CD-HIT: accelerated for clustering the next-generation sequencing data.*
804 Bioinformatics, 2012. **28**(23): p. 3150-2.
805 112. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search*
806 *programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.
807 113. Katoh, K. and D.M. Standley, *MAFFT multiple sequence alignment software version 7:*
808 *improvements in performance and usability.* Mol Biol Evol, 2013. **30**(4): p. 772-80.

809