CHAPTER 2

Materials and Methods
In the research presented here, several approaches have been used to study the structure, stability and function of the ubiquinone reduction (Q) module. The materials and methods utilized in the entire study are enlisted below.

2.1 Phylogenetic analysis

Protein sequences of subunits NDUFS2, 3, 7 and 8 available from bacteria to mammals were downloaded from UniProtKB (http://uniprot.org) protein database. The sequences for each have been aligned using ClustalW, the program that is part of the MEGA: Molecular Evolutionary Genetics Analysis software [Thompson et al, 1994; Tamura et al, 2011]. Phylogenetic analysis was performed with the MEGA 5.5 software [Tamura et al, 2011; Hall BG, 2013], using the neighbor joining statistical method with 1000 bootstrap replications. The phylogenetic trees generated were visualized in FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/) [Rambaut, 2009].

2.2 Molecular modelling

2.2.1 Core subunits of the Q module

The Q module is made up of four core subunits NDUFS2, 3 7 & 8, the constituents of complex-I common from bacteria to mammals. The protein sequences of these four core subunits of Q module were acquired from the UniProtKB protein databank (Entry number: O75306, O75489, O75251, O00217). A similarity search carried out using the BLAST server [Altschul et al, 1990] to choose model templates, identified the bacterial complex (PDB ID: 2FUG, 3I9V, 3IAM) of the organism Thermus thermophilus in PROTEIN DATABANK (PDB) (http://www.rcsb.org). These had 40-50% sequence identity with the corresponding four subunits of human Complex-I Q module [Berrisford et al, 2009; Saznov et al, 2006]. Subunits NDUFS2 and NDUFS7 were modelled using D and F chains of 3I9V (resolution: 3.1 A°) whereas subunits NDUFS3 and NDUFS8 were modelled using E and G chains of 3IAM (3.1A°), templates were chosen based on higher sequence similarity and better resolution of the crystal structure. The signal sequences for subunits were predicted using MITOPROT: prediction of mitochondrial targeting sequences server [Claros et al, 1996]. Signal sequences were not part of modelling (Table 2.1).
Table 2.1: Signal sequences are shown predicted by MITOPROT for the four core subunits. These regions were excluded from modelling.

<table>
<thead>
<tr>
<th>Subunit</th>
<th>Signal sequence</th>
<th>Cleavage site</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDUFS2</td>
<td>MAALRALCGGRVAAQVLRP</td>
<td>21</td>
</tr>
<tr>
<td>NDUFS3</td>
<td>MAAAARGLWWRGILGASALTRGTGRPSV</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>LLLPVRE</td>
<td></td>
</tr>
<tr>
<td>NDUFS7</td>
<td>MAVLSAPGLRGFRILGRSSVGPARGV</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>HQSVAQDGGPVSTQPALPKARA</td>
<td></td>
</tr>
<tr>
<td>NDUFS8</td>
<td>MRCLTTPMLLRALAARAGPPGGRSLHS</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>SAVAA</td>
<td></td>
</tr>
</tbody>
</table>

The molecular models were prepared using the software MODELLERv9.10 [Eswar et al, 2007; 2008]. 20 models were generated in each case. Energy minimization of both wild type and mutants was carried out by using GROMACS v.4.5 [Hess et al, 2008] by the steepest descent minimization for 100ps with maximum force field cut off being 1 KJ/mol.

2.2.2 Modelling Q module assembly

The subunit assembly of Q module was modelled by comparing with the arrangement of chains D, E, F and G (PDB ID: 2I9V) using the alignment program in molecular modelling suite PRIME v3.1 (Schrödinger). Ten initial models of Q module along with three iron sulphur clusters were first generated. The models were then energy minimized and prepared in the “protein preparation wizard” of Maestro 9.3.

2.3 Evaluation of the generated models

The molecular models, subunits as well as the entire Q module, were evaluated using Discrete Optimized Protein Energy (DOPE) score, ERRAT (version 2.0) [Colovos et al, 1993], PDBsum [Laskowski, 2009], ProSA-web Protein structure analysis [Wiederstein et al, 2007] and RMSD based on Cα overlap between target and template.

The DOPE score is an independent assessment of the accuracy of the output models in the MODELLERv9.10. The program assigns a score for a model by considering the positions of all the non-hydrogen atoms. Lower scores correspond to predicted models that are more accurate. The ERRAT program verifies the quality of the model. This program plots error values as a function of position in the sequence.
by sliding a nine residue window along the sequence. The error function is based on the statistics of non-bonded atom-atom interactions in the template structure. PDBSum is a pictorial database providing an overview of the model structure in terms of Ramachandran plot statistics, main chain parameters etc. ProSA is a tool widely used to check errors in 3D protein models. An overall quality score or Z score is estimated and shown in a plot where scores estimated from experimentally determined structures in PDB are plotted. The Z score is an indication of the overall quality of model and measures the deviation of the total energy of the modelled structure from an energy distribution derived from random conformations.

The 20 models generated for each of the 4 subunits of the Q module and the 10 models generated for the entire Q module were evaluated with the above parameters. The ones with the best statistics were utilized for further study.

2.4 In silico mutagenesis of the subunits

The in silico mutants were generated using the program FoldX (Table 2.2) [Schymkowitz et al, 2005].

Table 2.2: List of in silico mutations generated using the FoldX program. All these mutations are known to result in Leigh or Leigh like syndrome.

<table>
<thead>
<tr>
<th>Subunit</th>
<th>Mutation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDUFS2</td>
<td>R228Q+S413P</td>
<td>Loeffen J et al 2001</td>
</tr>
<tr>
<td></td>
<td>R138Q+R333Q</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M292T+R118Q</td>
<td>Tuppen HA et al 2010</td>
</tr>
<tr>
<td></td>
<td>M292T+M443K</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M292T+E18K</td>
<td></td>
</tr>
<tr>
<td>NDUFS3</td>
<td>T145I+R199W</td>
<td>Benit P et al 2004</td>
</tr>
<tr>
<td>NDUFS7</td>
<td>V122M</td>
<td>Triepels RH et al 1999</td>
</tr>
<tr>
<td></td>
<td>R145H</td>
<td>Lebon S et al 2007</td>
</tr>
<tr>
<td>NDUFS8</td>
<td>P85L+R138H</td>
<td>Loeffen J et al 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Procaccio V et al 2004</td>
</tr>
</tbody>
</table>
2.5 Molecular Dynamics (MD) simulation of individual subunits

The molecular models of wild-type and mutant proteins were used to perform molecular dynamics simulation using GROMACS v.4.5 with the OPLS-AA/L all-atom force field. The subunits were solvated with SPC water model using the genbox program of GROMACS suite. The default cubic boxes of GROMACS with dimensions: 12.03 nm (NDUFS2), 10.55 nm (NDUFS3), 9.38 nm (NDUFS7) and 7.49 nm (NDUFS8) were used. Sodium and chloride ions were added to each system depending on the requirement for charge neutralization (Table 2.3).

<table>
<thead>
<tr>
<th>Subunit</th>
<th>Ions added to neutralize the system</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Na⁺</td>
</tr>
<tr>
<td><strong>NDUFS2</strong></td>
<td></td>
</tr>
<tr>
<td>w-t</td>
<td>4</td>
</tr>
<tr>
<td>R228Q+S413P</td>
<td>5</td>
</tr>
<tr>
<td>R138Q+R333Q</td>
<td>6</td>
</tr>
<tr>
<td>M292T+R118Q</td>
<td>5</td>
</tr>
<tr>
<td>M292T+M443K</td>
<td>3</td>
</tr>
<tr>
<td>M292T+E18K</td>
<td>2</td>
</tr>
<tr>
<td><strong>NDUFS3</strong></td>
<td></td>
</tr>
<tr>
<td>w-t</td>
<td>5</td>
</tr>
<tr>
<td>T145I+R199W</td>
<td>6</td>
</tr>
<tr>
<td><strong>NDUFS7</strong></td>
<td></td>
</tr>
<tr>
<td>w-t</td>
<td>-</td>
</tr>
<tr>
<td>V122M</td>
<td>-</td>
</tr>
<tr>
<td>R145H</td>
<td>-</td>
</tr>
<tr>
<td><strong>NDUFS8</strong></td>
<td></td>
</tr>
<tr>
<td>w-t</td>
<td>6</td>
</tr>
<tr>
<td>P85L+R138H</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 2.3: Ions introduced to neutralize the system for each subunit and its *in silico* mutant.
Energy minimization was performed by the steepest descent method for 50000 steps (the minimization tolerance was set to 1000 kJ/mol-nm). Equilibration was carried out in 2 steps using conditions NVT and NPT, respectively. Long range electrostatics was computed using the Particle Mesh Ewald (PME) method [Darden et al, 1993] and Lennard-Jones energy cut off was set to 1.0 nm. Bond lengths were constrained with the LINCS algorithm [Hess et al, 1997]. Simulations of 15 ns duration were performed on the wild type and mutant structures at constant temperature of 300 K maintained by modified Berendsen thermostat coupling [Berendsen et al, 1984] and at a constant pressure of 1 bar by Parrinello-Rahman pressure coupling. The time step employed was 2 fs and coordinates were saved every 2 ps for analysis of MD trajectory. Analyses were performed with the tools available in the GROMACS utilities. RMSD, RMSF, Radius of gyration (Rg), average number of inter and intra-molecular hydrogen bonds formed, solvent accessible area and secondary structure prediction along the trajectory (DSSP) were estimated.

2.6 Docking of n-decyl-ubiquinone (DBQ) in Q module

The molecular model of Q module prepared as already described was used to dock n-decyl-ubiquinone (DBQ) using Glide 5.8 (Schrödinger) [Friesner et al, 2004]. The model structure was imported and a centroid receptor grid was generated around the residue Y141 known to interact with DBQ [Angerer et al, 2012]. The ligand molecule DBQ (CID 2971) was downloaded from PubChem compound database [Balton et al, 2008] and prepared in the LigPrep v2.5 of the Schrödinger suite. This was then docked in the Q module and ten poses were generated. Out of these poses, the most likely one was chosen based on parameters such as glide gscore, glide emodel value and essential interactions were confirmed by experimental mutagenesis data [Angerer et al, 2012].

2.7 In silico mutation analysis

Residue scanning wizard of BioLuminate 1.0 in Schrödinger suite was used to study the structural effects of mutations on the Q module. For individual mutations the difference in the stability of mutant compared to wild-type protein was estimated. Similarly, the difference in binding affinity of each mutant for the three iron-sulphur clusters, the DBQ and other interacting subunits in comparison with corresponding wild-type proteins was also considered.
2.8 MD simulation of subunit complex with bound nDBQ

In order to determine the effects of mutations on the binding of n-DBQ, MD simulations were performed on the subunit complex of NDUFS2 and 7. The mutants of NDUFS7, V122M and R145H were the only ones considered in this study as they have been further verified by experimentation.

Desmond Molecular Dynamics systems v3.1 [Guo et al, 2010] with Optimized Potentials for Liquid Simulations (OPLS) all atom force field 2005 [Kaminski et al, 2001, Jorgensen et al, 1996] was used to perform molecular dynamic simulations on w-t complex of NDUFS2 and 7, w-t without Fe-S cluster and n-DBQ bound mutant proteins V122M and R145H. A complex of NDUFS2 and 7 was modelled using Prime v.3.1 [Jacobson et al, 2004] n-DBQ was docked at the interface of NDUFS2 and 7 using Glide v.5.8 [Friesner et al, 2004]. Suitable pose was chosen depending on the Glide gscore and the presence of the essential hydrogen bond between Y141 of NDUFS2 and head carbonyl group of n-DBQ. Mutants were prepared in the mutagenesis wizard of Maestro v9.3. Modelled protein bound with n-DBQ were prepared in the protein preparation wizard of Maestro v 9.3. Preparation of protein structures included addition and optimization of hydrogen atoms, generating metal binding states of the Fe-S cluster and restrained minimization using impref. The prepared structures were then uploaded in Desmond setup wizard and were solvated with SPC water model in an orthorhombic periodic boundary box so as to minimize system volume. Systems were neutralized using appropriate number of counterions. Energy of the prepared system was minimized up to maximum 1000 steps using steepest descent method until a gradient threshold (25 kcal/mol/Å°) is reached followed by LBFGS (Low-memory Broyden-Fletcher-Goldfarb-Shanno quasi-Newtonian minimizer) until a convergence threshold of 1 kcal/mol/Å° was achieved. The systems were equilibrated with the default parameters in Desmond v3.1 and MD simulations were carried out for 5 ns at a constant temperature of 300 K and a constant pressure of 1 bar with a time step of 2 fs. Long range electrostatic interactions were calculated by the smooth particle mesh Ewald method and a 9 Å cut-off radius was used for Coulombic short range interaction cut-off method.

The quality of simulations in terms of total energy, potential energy, temperature, pressure and volume were analyzed. The root mean square deviation
(RMSD), fluctuation of residues (RMSF), hydrogen bond between ligand and residues and distances were calculated by the Simulation Event Analysis module in Desmond v3.1. All the figures for the molecular modelling, docking and simulation studies were prepared in the PyMOL Molecular Graphics System, Version 1.5.0.4.

Cloning, expression and purification of the four core subunits of the Q module and the ND1 gene was attempted. Due to several problems associated with either cloning or expression of \( NDUFS2 \), \( NDUFS8 \) and ND1 (cloning is described in Appendix), further characterization and study of these proteins was not successful. Thus, the other two proteins \( NDUFS3 \) and 7 cloned were utilized for detailed study. The cloning, expression, purification and further biophysical studies performed on the purified proteins are described ahead.

2.9 RNA isolation, cDNA preparation, primer design and PCR amplification

Total RNA was isolated from the human colorectal adenocarcinoma cell line HT29 (1x10^6 cells) using Trizol® Reagent (Life Technologies, Cat#10296-010) as per the manufacturer’s instructions. Purified RNA samples were analyzed by denaturing agarose gel electrophoresis and concentration was spectrophotometrically determined using Nanodrop (Thermo Scientific, USA). One \( \mu \)g of purified RNA was used for the preparation of cDNA using the SuperScript™ III First Strand Synthesis System (Life Technologies, Cat#18080-051).

2.9.1 Preparation of \( NDUFS3 \) clone

Suitable primers (NDUFS3F: 5’ ATC ATA TGG CGG CGG CGG C 3’ & NDUFS3R: 5’ TGC TCG AGC TAC TTG GCA TCA GGC TTC 3’) were designed based on the RNA sequences downloaded from from National centre for biotechnology information NCBI (http://www.ncbi.nlm.nih.gov/) for amplification of the full length \( NDUFS3 \) ORF. The PCR reactions were set in a 50 \( \mu \)l volume containing 1X Pfu buffer (20 mM Tris-HCl pH 8.8 at 25°C, 10 mM KCl, 10 mM \((NH_4)_2SO_4\), 2 mM MgSO_4, 0.1% Triton X-100, & 0.1 mg ml^{-1} nuclease free BSA), 1 unit of Pfu polymerase, 200 \( \mu \)M each of dNTP and forward and reverse primers and 100 ng of the amplified cDNA. Cycling conditions were 95°C for 3 min, followed by 30 cycles of 95°C for 10 s, 50°C for 45 s, 72°C for 1 min and final extension of 72°C for 20 min with hold of 25°C forever. The amplicon of desired size (808 bp) was gel
extracted using the QIAQuick Gel Purification kit (Qiagen, Cat#28704), which was used for TA cloning in pGEM-T vector (Promega).

2.9.2 Preparation of NDUFS7 clone

Suitable primers (NDUFS7F: 5’ GCC ATA TGG CGG TGC TGT CAG CTC 3’ & NDUFS7R: 5’ TGC TCG AGC TAC CTG CGG TAC CAG ATC 3’) were designed based on the RNA sequences downloaded from NCBI (http://www.ncbi.nlm.nih.gov/) for amplification of the full length NDUFS7 ORF. The PCR reactions were set in a 50 μl volume containing 1X Pfu buffer (20 mM Tris-HCl pH 8.8 at 25°C, 10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1% Triton X-100, & 0.1 mg ml⁻¹ nuclease free BSA), 1 unit of Pfu polymerase, 200 µM each of dNTP and forward and reverse primers and 100 ng of the amplified cDNA. Cycling conditions were 95°C for 3 min, followed by 30 cycles of 95°C for 10 s, 50°C for 45 s, 72°C for 1 min and final extension of 72°C for 20 min with hold of 25°C forever. The amplicon of desired size (656 bp) was gel extracted using the QIAQuick Gel Purification kit (Qiagen, Cat#28704), which was used for TA cloning in pGEM-T vector (Promega).

2.10 TA cloning and sub-cloning in pET bacterial expression vector

The purified amplicon was A-tailed using Taq polymerase in 1X Thermopol buffer (NEB) containing 200 µM dATP and 5 units of Taq DNA polymerase (NEB) at 72°C for 20 minutes. The A-tailed amplicons were then cloned into pGEM-T vector (Promega), followed by chemical transformation into E. coli DH5α (Invitrogen). The plasmids from the positive colonies, screened through colony PCR, were purified by the standard alkaline lysis method. Full length NDUFS3 cDNA ORF or NDUFS7 cDNA ORF from pGEM-T vector was further subcloned into pET-28b(+) vector (Novagen) between NdeI and XhoI sites. Sequences were confirmed by DNA sequencing using BigDye™ Terminator Cycle Sequencing Ready Reaction Kit v3.1 (ABI, Cat#4337457) in an automated 3730 DNA analyzer (ABI). Appropriate plasmids were transformed into E. coli BL21(DE3) for protein expression.
2.11 Site-directed mutagenesis

2.11.1 NDUFS3

The double mutant T145I-R199W is known to cause Leigh syndrome. A two step approach was used to prepare the double mutant (T145I-R199W) of NDUFS3. Two sets of primers were designed for the same. The SDMF1 5’CGG ATC CGT GTG AAG ATC TAC ACA GAT GAG CTG3’, SDMR1 5’CAG CTC ATC TGT GTA GAT CTT CAC ACG GAT CCG3’ and SDMF2 5’CTT CGA GGG ACA TCC TTT CTG GAA AGA CTT TCC TCT ATC3’, SDMR2 5’GAT AGA GGA AAG TCT TCT TCA TGG AAA TGG TGT CCC TCG AAG3’. Site-directed mutagenesis was carried out with the help of Phusion™ Site-Directed Mutagenesis kit (Finnzymes, Cat#F541). The PCR reactions were set in a 50 μl volume containing the w-t NDUFS3 gene cloned in pGEM-T vector, 1X HF buffer, 200 μM each of dNTP and forward and reverse primers, 0.02U/μl Phusion Hot Start DNA polymerase. Cycling conditions were 95°C for 3 min, followed by 30 cycles of 95°C for 10 s, 55°C for 5 min, 68°C for 6 min and final extension of 72°C for 20 min with hold of 25°C forever. In the first PCR cycle the first set of primers (SDMF1 and SDMR1) were used. The purified amplicons were phosphorylated and ligated to circularize the plasmid and transformed into E. coli DH5α cells. The plasmids were sequenced and mutant plasmids were used for a second round of PCR with the second set of primers (SDMF2 and SDMR2). The PCR protocol was repeated and the plasmids were sequenced to choose the suitable construct. The double mutant of NDUFS3 was further sub-cloned in the pET-28b(+) vector (Novagen) between NdeI and XhoI restriction enzyme sites and the plasmids were transformed in E coli BL21(DE3) for protein expression. Sequences were confirmed by DNA sequencing using BigDye™ Terminator Cycle Sequencing Ready Reaction Kit v3.1 (ABI, Cat#4337457) in an automated 3730 DNA analyzer (ABI) at each step.

2.11.2 NDUFS7

Two mutants containing point mutations V122M and R145H were prepared. Suitable primers were designed for the same. SDMF1 5’GAT CAT GGC CGG CAC ACT CAC CAA CAA GAT GCC 3’, SDMR1 5’ATG ACG TCG GAC TGG CGC GGC GGG CTG 3’ and SDMF2 5’CCG CAC TAC GTG GTC TCC ATG GGG AG 3’, SDMR2 5’CTC CGG CAT CTG GTC GTA GAC CTT GCG AA 3’. SDMF1 and SDMR1
were used for V122M mutation and SDMF2 and SDMR2 were used for R145H mutation. Site-directed mutagenesis was done with the help of Phusion™ Site-Directed Mutagenesis kit (Finnzymes, Cat#F541). The PCR reactions were set in a 50 μl volume containing the w-t NDUFS7 gene cloned in pGEM-T vector, 1X HF buffer, 200 μM each of dNTP and forward and reverse primers, 0.02U/μl Phusion Hot Start DNA polymerase. Cycling conditions were 95°C for 3 min, followed by 30 cycles of 95°C for 10 s, 55°C for 5 min, 68°C for 6 min and final extension of 68°C for 20 min with hold of 25°C forever. The purified amplicons were phosphorylated and ligated to circularize the plasmid and transformed into E. coli DH5α cells. The plasmids were sequenced to confirm the mutagenesis. Both the mutants of NDUFS7 were further sub-cloned in the pET-28b(+) vector (Novagen) between NdeI and XhoI restriction enzyme sites and the plasmids were transformed in E.coli BL21(DE3) for protein expression. Sequences were confirmed by DNA sequencing using BigDye™ Terminator Cycle Sequencing Ready Reaction Kit v3.1 (ABI, Cat#4337457) in an automated 3730 DNA analyzer (ABI) at each step.

2.12 Protein expression and solubility

A fresh sterile plate of LB containing 60 μg/ml Kanamycin was streaked using the glycerol stock of BL21(DE3) E. coli cells transformed with suitable gene cloned in pET-28b(+) expression vector. The plate was incubated at 37°C for 14 hours. A single colony from the plate was further inoculated in 5 ml sterile LB containing 60 μg/ml kanamycin. The culture was incubated for 12-14 hours at 37°C with a shaking of 150 rpm.

Autoinduction method was used to obtain soluble expression of the protein. This method as described by Studier [Studier, 2005] utilizes the Rich (ZYM-5052) autoinduction medium for soluble expression of proteins. In short, the media comprises of the following solutions:

Stock solutions

- Sterile 1 M MgSO₄ stock solution
- Sterile 50 X 5052 solution

For 100 ml sequentially dissolve 2.5 g glucose and 10 g α-lactose in 75 ml water and finally add 25 g glycerol.
Sterile 25 X M solution
For 200 ml sequentially dissolve in 80 ml water 3.6 g Na$_2$SO$_4$ anhydrous, 13.4 g NH$_4$Cl anhydrous, 17.0 g KH$_2$PO$_4$ anhydrous and 17.7 g Na$_2$HPO$_4$ anhydrous. Make up the volume to 200 ml.

ZY medium
For 1000 ml medium, dissolve 10 g N-Z-amine AS (or tryptone), 5 g Yeast extract in 937 ml water.

1000x metal solution
For 100 ml, autoclaved solutions of 50 ml 0.1 M FeCl$_3$-6H$_2$O (in 0.1 M HCl), 2 ml 1 M CaCl$_2$, 1 ml 1 M MnCl$_2$-4H$_2$O, 1 ml 1 M ZnSO$_4$-7H$_2$O, 1 ml 0.2 M CoCl$_2$-6H$_2$O, 2 ml 0.1 M CuCl$_2$-2H$_2$O, 1 ml 0.2 M NiCl$_2$-6H$_2$O, 2 ml 0.1 M Na$_2$MoO$_4$-5H$_2$O, 2 ml 0.1 M Na$_2$SeO$_3$-5H$_2$O and 2 ml 0.1 M H$_3$BO$_3$ were mixed in 36 ml water.

All the above solutions were autoclaved seperately at 120°C for 15 minutes. The Rich medium for autoinduction was prepared by mixing the above stock solutions aseptically in the following ratios

### Table 2.4: Composition of the Rich ZY medium.

<table>
<thead>
<tr>
<th>Stock solution</th>
<th>For 1000 ml of Rich ZY medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>1M MgSO$_4$</td>
<td>1 ml</td>
</tr>
<tr>
<td>1000X metals</td>
<td>1 ml</td>
</tr>
<tr>
<td>50 X 5052</td>
<td>20 ml</td>
</tr>
<tr>
<td>25 X M</td>
<td>40 ml</td>
</tr>
<tr>
<td>Kanamycin (60 mg ml$^{-1}$)</td>
<td>1 ml</td>
</tr>
<tr>
<td>ZY medium</td>
<td>Make up the volume to 1000 ml</td>
</tr>
</tbody>
</table>

The expression protocol in short involves, inoculating 1 ml of the primary culture in fresh LB medium with Kanamycin (5 ml) and incubation at 37°C to activate dormant cells. In the meanwhile, autoinduction media is prepared by adding required amount of ZY medium, 50 X 5052 solution, 25 X M solution and Kanamycin (Table 2.4) in a sterile flask under aseptic conditions. 10 ml of the primary culture was added to 1000 ml of the Rich ZYM autoinduction medium. The inoculated medium was kept at 37°C shaker for 2 hours. The culture was then incubated at 16°C for 16 hours.
Following prolonged incubation, the cells were harvested by spinning at 4000 rpm for 15 minutes at 4°C. The cell pellet was re-dissolved in suitable amounts of Lysis Buffer (50 mM Tris-Cl pH 8.5, 100 mM NaCl, 0.01% IGEPAL® CA-630 and 1 mg/ml lysozyme) for NDUFS3 and (20 mM sodium phosphate buffer pH 7.0, 300 mM NaCl, 0.01% IGEPAL® CA-630 and 1 mg/ml lysozyme) for NDUFS7 and was kept in ice for 30 minutes with intermittent shaking. Cells were lysed by sonication at 60% power in Esquire Biotech Ultrasonic homogenizer with a pulse of 6 s followed by pause of 8 s for 10 minutes. The sonicated solution was centrifuged at 10,000 rpm for 45 minutes at 4°C. The pellet and supernatant were separated. The expression profile was checked on 12% SDS-PAGE gel followed by Coomassie Brilliant Blue R-250 staining.

2.13 Purification of NDUFS3 and NDUFS7

2.13.1 NDUFS3

The protein from the crude cell lysate was purified in two steps of column chromatography, anion exchange followed by size exclusion chromatography. In short, the supernatant obtained from the lysed cells was centrifuged at 10,000 rpm for 15 minutes at 4°C. It was then used for loading directly onto a Q-Sepharose (Sigma-Aldrich, Cat#Q1126) column pre-equilibrated with 50 mM Tris-Cl pH 8.5, 100 mM NaCl and 0.01% IGEPAL® CA-630. The column was washed with the above buffer till the eluate showed no absorbance at 280 nm. The column was subjected to a NaCl gradient from 0.1 M to 0.5 M and final wash was given with the buffer containing 1 M NaCl. Absorbance of the fractions was recorded at 280 nm. The fractions showing OD >0.2 were run on 12% SDS-PAGE gel to detect the presence of the protein. The fractions showing the protein with desired molecular weight (~35 kDa) were pooled and concentrated in a Labconco vacuum concentrator. The OD\textsubscript{280} was noted and concentration of the protein solution was checked using Bradford assay [Bradford, 1976]. The Sephadex G-200 column (Sigma-Aldrich, Cat#G-200-120) was pre-equilibrated with 50 mM Tris-Cl pH 8.5, 300 mM NaCl and 0.01% IGEPAL® CA-630. The protein was loaded onto the column and washed with three column volumes of equilibration buffer. Absorbance of the fractions was measured at 280 nm. The fractions showing an OD >0.2 were run on 12% SDS-PAGE gel to identify protein
fractions. Both w-t and mutant (T145I-R199W) proteins were purified by this protocol.

2.13.2 NDUFS7

The supernatant obtained from the lysed cells was used for loading directly on Ni-NTA agarose (Life technologies, Cat#R109-15)) column pre-equilibrated with 20 mM sodium phosphate buffer pH 7.0, 300 mM NaCl, 20mM imidazole and 0.01% IGEPAL® CA-630. The column was washed with the above buffer till the eluate didn’t show any absorbance at 280 nm. The column was subjected to an imidazole gradient from 0.1 M to 0.4 M and final wash was given with buffer containing 0.5 M imidazole. Absorbance of the fractions was taken at 280 nm. The fractions showing an OD >0.2 were run on a 12% SDS-PAGE gel to detect the presence of the protein of interest. The fractions showing the protein with desired molecular weight (~20 kDa) were pooled and dialyzed to remove the imidazole and concentrated in a Labcono vacuum concentrator. The OD$_{280}$ and concentration of the concentrated protein was checked using Bradford’s assay [Bradford, 1976]. The Superose 12 column (GE Healthcare, Cat#G-17-5173-01) was pre-equilibrated with 20 mM sodium phosphate buffer pH 7.0, 150 mM NaCl and 0.01% IGEPAL® CA-630. The protein was loaded onto the column and was subjected to 2 column volumes of equilibration buffer washes. Absorbance of the fractions was taken at 280 nm. The fractions showing an OD >0.2 were run on 12% SDS-PAGE gel to detect the presence of the protein. Western Blot with anti-His tag antibody and MALDI TOF/TOF was used to confirm the presence of the protein and its molecular weight.

2.14 Western Blot and MALDI-TOF/TOF™

Western blots were prepared by electroblotting the SDS-PAGE gels onto PVDF membrane [Towbin et al, 1979]. The membrane was blocked with 5% fat-free skimmed milk in PBS for 60 minutes followed by 3 washes of 5 minutes each in PBS containing 0.05% TWEEN® 20. The membrane was further incubated with the monoclonal anti-polyhistidine antiboby (Sigma-Aldrich, Cat# H1029) at 1:1000 dilution in PBS containing 1% BSA overnight. The membrane was washed 3 times for 5 minutes each in PBS containing 0.05% Tween®20 and incubated with anti-mouse Ig-G (Fc-specific) Peroxidase conjugate antibody in PBS and Tween®20 at a
concentration of 1:6000 (Sigma-Aldrich, Cat#A0168). The membrane was further treated with Novex HRP Substrate (Invitrogen Cat#WP 20004) to visualize the bands.

For MALDI-TOF/TOF™ the concentration of the purified protein sample was determined by Bradford assay. 10µg ml⁻¹ sample was mixed with a saturated solution of α-cyanohydroxy cinapinic acid in a mixture of 0.1% TFA:ACN and spotted onto the MALDI plate. The plate was dried at room temperature and was analyzed with AB SCIEX TOF/TOF™ 5800 System for MALDI mass spectrometry imaging.

2.15 Biophysical characterization

2.15.1 Circular Dichroism (CD) Spectroscopy

CD spectra of the purified proteins were recorded using a Jasco J-815-150S (Jasco, Tokyo, Japan) spectropolarimeter connected to a Peltier Type CD/FL Cell circulating water bath at room temperature. Far UV spectra was recorded in a rectangular quartz cell of 1 mm path length in the range of 200-250 nm at a scan speed of 100 nm min⁻¹ with a response time of 1 s and a slit width of 1 nm. Purified w-t NDUFS3 at a concentration of 0.05 mg ml⁻¹ and its mutant (T145I-R199W) at a concentration of 0.1 mg ml⁻¹ was used for all the far-UV CD samples. Purified w-t NDUFS7 at a concentration of 0.08 mg ml⁻¹ was used for all the far-UV CD samples. Near UV CD spectrum was recorded in the range of 250-300 nm with a protein concentration of 1 mg ml⁻¹ for all the proteins. Each spectrum was recorded as an average of 5 scans.

Conformational transition studies of w-t, mutant NDUFS3 and w-t, mutant NDUFS7 at various conditions of pH, chemical denaturant and temperature were performed.

2.15.1.1 pH variation: The purified proteins were incubated against buffers of different pH for 4 hours before recording the spectra. The different buffers used were: 20 mM glycine-HCl (pH 1.0-3.0), 20 mM citrate-phosphate buffer (pH 4.0-6.0), 20 mM tris-HCl (pH 7.0-9.0), 20 mM glycine-NaOH (pH 10.0-12.0). The native tertiary structure of proteins was studied by recording the near UV CD spectrum in the range of 250-300 nm with a working concentration range of 1 mg ml⁻¹.

2.15.1.2. Temperature dependence: The CD spectra of the purified proteins were recorded by increasing the temperature of the samples at the rate of 2°C min⁻¹ within
the temperature range of 25-90°C. Ellipticity was recorded at a temperature interval of 5°C and incubation time of 5 minutes was maintained between 200-250 nm.

2.15.1.3. Gdn-HCl concentration: The purified proteins were incubated in Gdn-HCl in the concentration range of 0-4 M at pH 8.5 for 6 hours before recording the spectra. Buffer scans were subtracted from each spectrum for further analysis.

Results were determined in terms of the mean residue ellipticity (MRE). CD in millidegrees was converted to mean residue ellipticity by the formula:

\[(\theta)_\lambda = M\theta_\lambda / 10dc\]

where, M is the molecular mass, \(\theta_\lambda\) is the ellipticity in millidegrees, d is the cuvette path length in cm and c is the concentration of the protein in mg ml\(^{-1}\), r is the number of residues. Secondary structure content was estimated using the CDPro program (CDSSTR, CONTIN, SELCON3) [Sreerama et al, 2004].

2.15.2. Steady State Fluorescence Spectroscopy

The intrinsic fluorescence of the proteins was recorded using a Perkin Elmer LS50 fluorescence spectrophotometer connected to a Julabo F20 water bath. The spectra were recorded using a quartz cuvette at room temperature. The background emission due to the buffer or denaturants was subtracted for further analysis. The protein solutions were excited at 295 nm and the emission spectra were recorded between 300-400 nm setting the slit width 7 nm and speed 100 nm min\(^{-1}\). The effect of pH was recorded by incubating the protein in suitable buffers of different pH (1-12) as described in the section 2.15.1.1. Thermal unfolding was monitored by incubating the protein for 5 min at the desired temperature followed by recording the spectra. Thermal aggregation was studied by the Rayleigh scattering measurement on the same instrument under different conditions of protein concentration (25-200 µg ml\(^{-1}\)), in the presence of salt (5-355 mM NaCl) and detergent (0.05-0.16 mM IGEPAL® CA-630). Chemical unfolding was monitored by recording the spectra of the protein incubated in Gdn-HCl (0-6 M) for 6 hours at room temperature. The decomposition analysis of tryptophans was performed by the PFAST: Protein Fluorescence and Structural Toolkit [Shen et al, 2008; Hixon et al, 2009].
2.15.3 Solute Quenching studies

2.15.3.1. Steady-state Fluorescence Quenching

Fluorescence quenching measurement was done with quenchers: acrylamide (5 M), caesium chloride (5 M) prepared in milliQ water and KI (5 M) prepared in 0.2 M sodium thiosulphate by titrating 80 µg ml⁻¹ of the protein solution prepared in 50 mM Tris-Cl pH 8.5, 100 mM NaCl and 0.01% IGEPAL®CA-630. Quencher from the stock solutions was added till the final concentration of quencher in the protein solution reached 0.5 M. The fluorescence intensities were measured at the wavelength corresponding to maximum emission and volume correction was done before analyzing the quenching data [Lakowicz, 1983]. Stern-Volmer analysis of the quenching data was used to estimate $K_{sv}$ and modified Stern-Volmer plots for determining $f_a$ (fraction accessibility) [Lehrer et al., 1978]. The Stern-Volmer equation given below was used for the same.

$$
\frac{F_0}{F_c} = 1 + K_{sv}[Q] \tag{1}
$$

Where $F_0$ and $F_c$ represent the relative fluorescence intensities corrected for dilution in absence and presence of the quencher [Q]. $K_{sv}$ is the Stern-Volmer constant for the given quencher.

2.15.3.2. Fluorescence lifetime measurement

Lifetime measurements were recorded on Edinburgh Instruments FLS-920 single photon counting spectrofluorimeter. The excitation source was a laser pico second pulsed light emitting diode (model EPLED-295) and fluorescence was detected by a synchronization photomultiplier. Diluted Ludox solution was used for measuring the Instrument response function (IRF). The samples (1 mg ml⁻¹) were excited at 295 nm and emission was recorded at 342 nm, 346 nm and 356 nm for w-t, mutant and denatured protein samples. Slit widths of 15 nm each were used for the excitation and emission monochromators. The resultant decay curves were analyzed by a multiexponential iterative fitting program provided by Edinburgh Instruments. The average lifetimes after each step of quenching was calculated using the equations [Inokuti et al., 1965; Grinvald et al., 1974]

$$
\tau = \sum_i \alpha_i \tau_i / \sum_i \alpha_i \tag{2}
$$
\[ \langle \tau \rangle = \sum \alpha_i \tau_i^2 / \sum \alpha_i \tau_i \]  \hspace{1cm} (3)

Where \( i = 1, 2, ... \)

The average lifetimes \( \tau \) and \( \langle \tau \rangle \) were obtained by two different approaches and the plot of \( \tau_0 / \tau \) against quencher concentration \([Q]\) gave the \( K_{sv} \). The bimolecular quenching constant, \( k_q \) was calculated using equation given below [Lehrer, 1971].

\[ k_q = K_{sv} / \tau \]  \hspace{1cm} (4)

The dynamic and static components were resolved by using the equation shown below [Lacowicz et al, 1973]:

\[ F_0/F_c = (1 + K_{sv}[Q])(1 + K_s[Q]) \]  \hspace{1cm} (5)

Where \( K_{sv} \) is the Stern-Volmer (dynamic) quenching constant, \( K_s \) is the static quenching constant and \([Q]\) is the concentration of quencher. The dynamic quenching constant is the reflection of the degree to which the quencher achieves the encounter distance of the fluorophore and is determined by the fluorescence lifetime measurement fitted to the equation [Lacowicz et al, 1973]:

\[ \tau_0 / \tau = (1 + K_{sv}[Q]) \]  \hspace{1cm} (6)

where \( \tau_0 \) is the average lifetime in the absence of quencher and \( \tau \) is the lifetime in the presence of the quencher at concentration \([Q]\).

**2.15.4. Hydrophobic dye binding**

Binding of the hydrophobic dye 8-Anilino-1-naphthalene sulfonic acid (ANS) was studied by recording the emission spectra between 400-550 nm post excitation at 375 nm using a steady state spectrofluorimeter. ANS has been shown to bind to hydrophobic regions of partially unfolded proteins that are exposed to the solvent [Semisotnov et al, 1991]. The proteins were incubated at various pH (1-12), Gdn-HCl concentrations (0-6 M) and at different temperatures (25-90°C) to study the unfolding of polypeptide chain. 5 µl of 15 mM ANS was mixed with 2 ml of protein solution (0.05 mg ml\(^{-1}\)). Buffer spectrum with ANS present was subtracted in each case for further analysis.
2.15.5. Assays for studying protein aggregation

2.15.5.1. Rayleigh scattering

The measurement of scattering of the w-t and mutant proteins were carried out in the time drive module of the Perkin Elmer LS50 fluorescence spectrophotometer connected to a Julabo F20 water bath. Measurements were recorded at 28°C and 37°C. Buffer values were subtracted in each case for further analysis.

2.15.5.2 Thioflavin-T (Th-T) binding assay

The amyloidophilic Thioflavin-T dye shows a characteristic increase in fluorescence upon binding to amyloid fibrils. Binding of Thioflavin-T was studied by recording the fluorescence emission spectra between 450-600 nm with an excitation at 442 nm using a steady state fluorescence spectrofluorimeter [Groenning, 2010; Chang et al, 2009; Khurana et al, 2005]. In short, a stock solution of 2 mM Th-T was prepared in 150 mM NaCl and 100 mM sodium phosphate buffer (pH 7.0) and filtered through 0.22 µm Millipore filter. A fresh working solution was prepared by adjusting the dye concentration to 200 µM. Protein samples (25-100 µg ml⁻¹) were incubated at 25°C and 37°C at pH 8.5 for 1 hour and further used for the assay. A 250 µl aliquot of sample solution was mixed with a 250 µl Th-T working solution and the spectra were recorded. Buffer spectrum with Th-T was subtracted in each case for further analysis.

2.15.5.3 Congo red (CR) assay

CR is similar to Th-T and binds to β-rich structures, inducing an increase in absorption and a red shit in the CR absorption band from 490 to 540 nm. Binding of Congo red was studied by recording the absorbance spectra on a UV-Vis spectrophotometer between 380-700 nm [Chang et al, 2009; Klunk et al 1999]. In short a 2 mM CR stock solution prepared in 150 mM NaCl and 5 mM potassium phosphate buffer (pH 7.5) was filtered through 0.22 µM Millipore filter. A fresh working solution was prepared by adjusting dye concentration to 200 µM. Protein samples were incubated at 25°C and 37°C for 1 hour. A 250 µl protein sample (25-100 µg ml⁻¹) was mixed with 50 µl of the working stock solution and 700 µl of the above buffer and the UV spectrum was recorded. Buffer spectrum with CR was subtracted in each case for further analysis.
2.16 Fe-S cluster detection

W-t *NDUFS7* possesses an iron sulphur cluster N2. A spectrum was recorded on a UV-visible spectrophotometer between 300-800 nm for the w-t, V122M, R145H and w-t after removal of the Fe-S cluster for the detection of this cluster. Buffer spectrum was subtracted in each case for further analysis. The Fe-S cluster was removed from the w-t by incubating the protein in 5 mM DTT and 2 mM EDTA for 1 hour. Far and near-UV CD spectra were recorded for the w-t, R145H, V122M with Fe-S cluster and then with w-t without Fe-S cluster to detect structural changes.

2.17 n-DBQ binding assay

*NDUFS7* possess a partial binding site for nDBQ. In order to explore the binding affinity of n-DBQ to protein; a fluorescence based assay was designed. The w-t and V122M, R145H mutant proteins in 20 mM sodium phosphate buffer pH 7.4, 100 mM NaCl and 0.01% IGEPAL-CA630 were placed in a quartz cuvette and maintained at a desired temperature (28°, 32°, 37° and 42°C) with a Julabo circulating water bath. A 0.1 mM stock solution of n-DBQ was prepared in dimethylsulfoxide (DMSO). The n-DBQ solution was added in 12-14 aliquots (3-5 µl) each and fluorescence spectra were recorded with each addition. Samples were excited at 295 nm and spectra were recorded between 310-400nm. The fluorescence intensity at 341 nm (λ_{max} of w-t and mutant proteins) was considered for further analysis. Corrections were made to compensate for the dilution effect upon addition of nDBQ. Each data point was an average of 3 independent sets of experiments with standard deviation less than 5%.

The association constants were calculated by the method of Chipman *et al* (Chipman *et al*, 1967). The abscissa intercept of the plot of log [C]_{r} against log \{(ΔF)/(Fc-F_∞)\}, where [C]_{r} is the free concentration of n-DBQ, yielded pKa value for protein-nDBQ interaction according to

\[
\log \left[ \frac{(F_0-F_C)}{(F_C-F_∞)} \right] = \log K_a + \log \left[ [C]_{r}[P]_{t} \left( \frac{ΔF}{ΔF_∞} \right) \right] \quad (1)
\]

where, Fc is the fluorescence intensity of protein at any point during the titration, [P]_{t} is the total protein concentration, ΔF_∞ is the change in fluorescence intensity at saturation of binding, [C]_{t} is the total n-DBQ concentration given by,

\[
[C]_{r} = \{[C]_{t} - [P]_{t} \left( \frac{ΔF}{ΔF_∞} \right) \} \quad (2)
\]
Free energy changes of association ($\Delta G$) were determined by:

$$\Delta G = -RT \ln K_a$$  \hspace{1cm} (3)

Temperature dependence of the association constants was used to determine thermodynamic parameters. Change in enthalpy ($\Delta H$) was determined from the Van’t Hoff plot by equation:

$$\ln K_a = (-\frac{\Delta H}{RT}) + \frac{\Delta S}{R}$$  \hspace{1cm} (4)

where, $\Delta H$ is enthalpy change, $R$ is gas constant, $\Delta S$ is entropy change and $T$ is the absolute temperature. The entropy change was obtained from the equation

$$\Delta G = \Delta H - T \Delta S$$  \hspace{1cm} (5)