

From the Department of Cell and Molecular Biology  
Karolinska Institutet, Stockholm, Sweden

# **Structural-functional studies of mitochondrial matrix proteins**

Saba Shahzad



**Karolinska  
Institutet**

Stockholm 2018

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by AJ E-print AB

© Saba Shahzad, 2018

ISBN 978-91-7831-270-2

# Structural-functional studies of mitochondrial matrix proteins

## THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

**Saba Shahzad**

*Principal Supervisor:*

Dr. B. Martin Hällberg  
Karolinska Institutet  
Department of Cell and Molecular Biology  
Biomedicum

*Co-supervisor(s):*

Prof. Nils-Göran Larsson  
Karolinska Institutet  
Department of Medical Biochemistry and  
Biophysics  
Biomedicum

*Opponent:*

Dr. Christian Löw  
European Molecular Biology Laboratory (EMBL)  
Hamburg

*Examination Board:*

Assoc. Prof. Pål Stenmark  
Lund University  
Department of Experimental Medical Science

Prof. Nico Dantuma  
Karolinska Institutet  
Department of Cell and Molecular Biology

Prof. Maria Ankarcrona  
Karolinska Institutet  
Department of Neurobiology, Care Sciences and  
Society





*'Read! In the Name of your Lord who created- He has created man from a clot. Read! And your Lord is the Most Generous'. (Quran 96: 1-5)*



## ABSTRACT

The mitochondrion is the powerhouse of the eukaryotic cell. Most of the energy required to carry out cellular processes is generated inside mitochondria via the process of oxidative phosphorylation. The machinery required for oxidative phosphorylation is encoded by both the nuclear and the mitochondrial genome. The cellular energy production will thus collapse in the absence of mitochondrial gene expression. The mitochondrial RNA polymerase, POLRMT, together with two transcription factors, TFAM and TFB2m, initiate mitochondrial transcription. However, the exact mechanistic details of mitochondrial-transcription initiation are unclear. Furthermore, the transcription by POLRMT is non-processive, and it prematurely terminates after 150 nucleotides in a conserved sequence block region, CSBII. This indicates that accessory factors are required for a complete transcription event.

In Paper I, we investigated the transcription initiation in mitochondria and proposed a model. In addition, we were able to demonstrate that an N-terminal extension (NTE) in POLRMT plays a role in the transition from the initiation to the elongation phase, possibly by undergoing a conformational change. Upon the deletion of NTE, the POLRMT is hyperactive, and together with TFB2m, it can carry out non-specific transcription events. Thus, we conclude that the NTE is necessary for promoter-specific transcription initiation. Once the transcription is initiated, the POLRMT enters the elongation phase.

In Paper II, we were able to characterize the *in-vitro* role of a transcription-elongation factor, TEFM. TEFM increases the processivity of the POLRMT by allowing it to bypass road blocks, such as the CSBII, and DNA lesions, such as apurinic or apyrimidinic sites. We furthermore suggested that TEFM may be involved in the regulation of mitochondrial transcription and replication in mammalian mitochondria.

For mitochondrial homeostasis, the integrity of proteins in the mitochondrial matrix must be maintained, and this is achieved through the mitochondrial protein quality control (PQC) system. Lon is the major mitochondrial matrix protease and is essential for mitochondrial PQC. Lon belongs to the AAA<sup>+</sup>-protease family and is a homo-hexamer that uses energy from ATP hydrolysis to recognize, bind, and translocate its substrate into a proteolytic chamber.

In Paper III, we presented the structure of a full-length human mitochondrial Lon determined by single-particle cryo-EM to a resolution of 3.6 Å. We showed that the human Lon, in its ADP-bound form, has its six protomers arranged in an open helical conformation with an 8 Å translational shift. The highly flexible N-terminal domains of every first and fourth protomer dimerizes, thereby giving rise to a unique arrangement that strengthens the oligomerization. At the same time, the arrangement provides new structural motifs for inter-protomer communication. Based on our analysis, we propose a hand-over-hand model, using three protomers, for substrate translocation by Lon, and our results can be generalized to the broad family of LonA AAA<sup>+</sup> proteases.

**Keywords:** Mitochondria, mtDNA, transcription, protein quality control, Lon, AAA<sup>+</sup>, cryoEM

**ISBN:** 978-91-7831-270-2

## SAMMANFATTNING PÅ SVENSKA

Mitokondrierna är den eukaryota cellens kraftcentral. Större delen av den energi som används i den eukaryota cellen genereras från den ATP som produceras i mitokondrier med hjälp av oxidativ phosphorylering. Det proteinmaskineri som krävs för oxidativ fosforylering kodas i både mitokondriegenomet och i kärngenomet. Den oxidativa fosforyleringen och därigenom energiproduktionen kollapsar i frånvaro av mitokondriellt genuttryck. Det mitokondriella RNA-polymeraset, POLRMT, i komplex med två transkriptionsfaktorer, TFAM och TFB2m erfordras för initiering av mitokondriell transkription. De mekanistiska detaljerna för mitokondriell transkriptionsinitiering har dock länge varit oklara. Vidare har det visat sig att transkriptionen av POLRMT är icke-processiv och avslutas i förtid efter 150 nukleotider vid en konserverad sekvensblockregion, CSBII. Detta indikerade att ytterligare proteiner krävs för att få ett komplett mitokondriellt transkript.

I arbete I kunde vi demonstrera sammansättningsordningen för det mitokondriella transkriptionsinitieringsmaskineriet och föreslå en modell för transkriptionsinitiering. Dessutom kunde vi visa att en N-terminal extension (NTE) spelar en stor roll vid övergången från initieringsfasen till elongeringsfasen, eventuellt genom att genomgå en konformationsförändring. Vid borttagning av NTE blir komplexet POLRMT-TFB2m hyperaktivt och transkriberar icke-specifikt. Således visar vi att NTE är nödvändigt för promotorspecifikt transkriptionsinitiering. När transkriptionen väl har initierats, går POLRMT in i elongeringsfasen.

I arbete II bestämde vi *in vitro*-rollen för en transkriptionselongeringsfaktor, TEFM. TEFM ökar POLRMTs processivitet genom att möjliggöra transkription över CSBII och förbi DNA-förändringar såsom apuriniska eller apyrimidiniska lesioner. Vi föreslog dessutom att TEFM kan vara involverad i reglering av mitokondriell transkription och eventuellt indirekt även replikation i däggdjursmitokondrier.

För att bibehålla mitokondriell homeostas måste proteinerna i mitokondriematrix vara funktionella och detta uppnås genom det mitokondriella proteinkvalitetskontrollsystemet (PQC). Humant mitokondriellt Lon är det huvudsakliga mitokondriella matrisproteaset och Lon är väsentlig för mitokondriell PQC. Lon tillhör AAA+-proteasfamiljen och är en homo-hexamer som använder energi från ATP-hydrolys för att känna igen, binda och translokera substrat in i en proteolytisk kammare.

I arbete III beskriver vi strukturen av humant mitokondriellt Lon. Strukturen är bestämd med hjälp av s.k. singelpartikelkryoelektronmikroskopiteknik till en upplösning av 3,6 Å. Vi visar att humant Lon, i sin ADP-bundna form, har sina sex protomerer anordnade i en öppen spiralformad konformation med en 8 Å translation. De väldigt flexibla N-terminala domänerna hos den första och sista protomeren dimeriserar och ger därigenom upphov till ett unikt arrangemang som stärker oligomeriseringen och kommunikationen mellan protomererna. Baserat på vår analys föreslår vi att den Lon-katalyserade substratnedbrytningen följer en s.k. hand-over-hand-modell, med tre protomerer, och våra resultat kan generaliseras till den breda familjen av LonA AAA+-proteaser.

**Nyckelord:** Mitokondrier, mtDNA, transkription, kvalitetskontroll, Lon, AAA+, cryoEM

**ISBN:** 978-91-7831-270-2



## LIST OF SCIENTIFIC PAPERS

- I. The amino terminal extension of mammalian mitochondrial RNA polymerase ensures promoter specific transcription initiation.  
Posse V, Hoberg E, Dierckx A, **Shahzad S**, Koolmeister C, Larsson NG, Wilhelmsson LM, Hällberg BM and Gustafsson C;  
Nucleic Acids Research, 2014, 42; 3638-3647.
- II. TEFM is a potent stimulator of mitochondrial transcription elongation *in-vitro*.  
Posse V, **Shahzad S**, Falkenberg M, Hällberg BM and Gustafsson CM.  
Nucleic Acids Research, 2015, 43:2615-24.
- III. Structure and degradation mechanism of the Human Mitochondrial Lon protease  
**Shahzad S**, Hernandez CP, Falkenberg M and Hällberg BM  
*Manuscript*.

# CONTENTS

1	Introduction to Mitochondria .....	1
1.1	Origin of mitochondria .....	1
1.2	Structure of mitochondria .....	1
1.3	Mitochondrial biogenesis.....	2
1.4	Metabolism in mitochondria .....	3
1.5	Transport to mitochondria.....	3
1.6	The mitochondrial genome .....	4
2	Mitochondrial replication.....	7
2.1	Replication in bacteriophage T7 DNA polymerase .....	7
2.2	Replication in bacteria .....	7
2.3	Replication in mtDNA.....	7
3	Mitochondrial Transcription .....	9
3.1	Transcription in bacteria .....	9
3.2	Transcription in the nucleus of eukaryotes.....	9
3.3	Transcription in bacteriophage T7.....	10
3.4	Transcription in mitochondria.....	11
3.4.1	Proteins involved in mitochondrial transcription.....	12
3.4.2	Phases involved in mitochondrial transcription.....	20
3.4.3	Other proteins involved in mtDNA transcription .....	24
4	Post-transcriptional RNA processing.....	25
5	Mitochondrial Protein Regulation .....	26
5.1	PQC in mammalian mitochondria .....	26
5.2	The human mitochondrial Lon, LonP1 (LonM) .....	28
5.2.1	Evolution and location.....	29
5.2.2	Domain organization .....	30
5.2.3	Mechanism of action .....	31
5.3	26S proteasome .....	37
6	Methodology .....	38
6.1	Recombinant protein production in <i>E. coli</i> .....	38
6.2	Purification of proteins .....	38
6.3	Biochemical characterization of complexes .....	39
6.4	Structure determination.....	40
6.4.1	X-ray crystallography.....	40
6.4.2	Electron cryomicroscopy, cryoEM.....	42
6.5	Comparison of X-ray Crystallography and CryoEM .....	45
7	Summary of papers.....	47
7.1	Paper I: .....	47
7.2	Paper II: .....	49
7.3	Paper III:.....	51
8	Future perspectives.....	53
9	Acknowledgements .....	55

10 References..... 59



## LIST OF ABBREVIATIONS

A	Adenine
A-loop	Aromatic loop
AAA+	Atpase associated with various cellular activities
ADP	Adenosine diphosphate
ALAS-1	5-aminolevulinic acid synthase
Arg-finger	Arginine finger
ATP	Adenosine triphosphate
Bp	basepair
Bs	<i>Bacillus subtilis</i>
C	Cytosine
ClpXP	Casein lytic protease, where ClpX is the AAA+ and ClpP is the protease
cryo-ET	electron cryotomography
cryoEM	electron cryomicroscopy
CSB	Conserved sequence block region
CTD	C-terminal domain
D-Loop	Displacement loop
DNAP	DNA polymerase
<i>E. coli</i>	<i>Escherichia coli</i>
ELAC2	ElaC-ribonuclease 2 (gene)
EMSA	electrophoretic mobility shift assay
ES	Elongation scaffold
ETC	electron transport chain
G	Guanine
GF	Gel filtration
H-S	Heavy -strand
HhH	Helix-hairpin-Helix
His <sub>6</sub>	6x Histidine
HSP	Heavy strand promoter
i-AAA+	Intermembrane facing AAA+ protease
IMAC	Immobilized metal affinity chromatography
Kb	Kilobases
kDa	Kilodalton
L-S	Light strand
LonA	Lon of type A

LonB	Lon of type B
LonC	Lon of type C
LonM	Lon present in mitochondria, also known as LonP1
LonP1	Lon present in mitochondria, also known as LonM
LRPPRC	Leucine-rich pentatricopeptide repeat containing protein
LSP	Light strand promoter
m-AAA+	matrix facing AAA+ protease
MELAS	Mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes
MERF	Myoclonic epilepsy with ragged red fibers
mRNA	Messenger RNA
tRNA	Transfer RNA
rRNA	Ribosomal RNA
mtDNA	Mitochondrial DNA
MRPL12	Mitochondrial ribosomal protein L12
MRPP1	Mitochondrial ribonuclease P protein 1
MST	Microscale thermophoresis
Mta	<i>Meiothermus taiwanensis</i>
MTERF1	Mitochondrial transcription termination factor 1
mtSSB	Mitochondrial single-strand binding protein
NTD	N-Terminal domain
NTE	N-Terminal extension
NT-strand	Non-template DNA strand
OriH	Origin of replication for heavy strand
OriL	Origin of replication for light strand
OXPHOS	Oxidative phosphorylation
POLRMT	Mitochondrial RNA polymerase
PD	Protease domain
POL	Polymerase
PPR	Pentatricopeptide repeat domain
PQC	Protein quality control
PRORP1	Protein-only RNase P1
RNAP	RNA polymerase
S1	Sensor 1
S2	Sensor 2
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size-exclusion column

SPA	Single-particle analysis
ssDNA	Single-stranded DNA
StaR	Steroidogenic acute regulatory protein
T	Thymidine
T-strand	Template DNA strand
TAS	Termination-associated sequence
TCA	Tricarboxylic acid cycle
TEFM	Mitochondrial transcription elongation factor
TFAM	Mitochondrial transcription factor A
TFB2m	Mitochondrial transcription factor B2
TSS	Transcription start site
U	Uracil
UPR <sup>mt</sup>	Unfolded protein response in mitochondria
WA	Walker A
WB	Walker B



# 1 INTRODUCTION TO MITOCHONDRIA

The mitochondrion works not only as the ATP producer for eukaryotic cells but also plays an important role in many other cellular functions, including oxidation of fatty acids, biosynthesis of amino acids and heme, apoptosis, signal transduction pathways, and pathogen sensing. Due to the diverse roles played by the mitochondrion, a variety of metabolic, genetic, and age-related diseases have been linked to mitochondrial dysfunction. For this reason, a deeper understanding of the expression and regulation of mitochondrial gene expression is of immense importance for highlighting its role in basic human biology (Ngo and Davies, 2007).

## 1.1 ORIGIN OF MITOCHONDRIA

Many different names like *fila*, *korn*, *chondros*, *grains*, *faden*, *mito* (filaments) were attributed by the microscopists in the early 1850's to the organelle that we now call mitochondria (Lehninger, 1965).

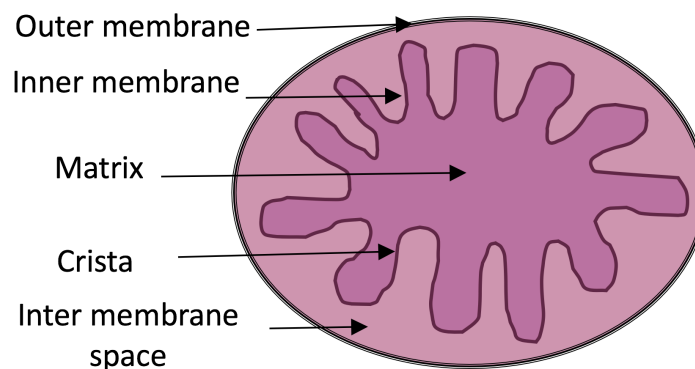
Finally, the term mitochondria was coined by Benda in 1898 (Scheduler, 2008). Two hypotheses were put forward to explain the origin of mitochondria. First, the autogenous origin hypothesis (direct filiation), which proposes that intracellular compartmentalization and functional specialization in a single eukaryotic cell resulted in a separate organelle i.e., the mitochondria. The second model states that around 1.6 billion years ago, the symbiotic relationship between a proto-eukaryotic cell and primitive prokaryotic cell (*alpha*-proteobacteria), capable of oxidative phosphorylation gave rise to mitochondria (Andersson et al., 2003; Gray et al., 1999). This symbiotic relationship equipped the modern eukaryotic cell with aerobic respiration capability (Gray et al., 1999). The discovery of a separate genome in mitochondria (Nass and Nass, 1963) supported the second model. However, during the course of evolution, only 15% of the modern mitochondrial genome can be traced back to its ancestral bacterial endosymbiont (Tsukihara et al., 1996). The mitochondria were initially linked to the transmission of genetic information, while it took almost 30-40 years of intense biochemical studies to characterize the main mitochondrial role as the 'powerhouse of cell'. Subsequent research highlighted its role in many other cellular processes like oxidation of fatty acids, biosynthesis of amino acids and heme, apoptosis, signal transduction pathways, and pathogen sensing (Polo et al., 2017).

## 1.2 STRUCTURE OF MITOCHONDRIA

Mitochondria have an inner and an outer membrane. The outer membrane makes a boundary wall, surrounding the whole mitochondria and the inner membrane convolutes to form finger-like projections that penetrate deep into the matrix and are known as cristae (Figure 1). These cristae were first observed by Palade and Sjöstrand in 1950s (Palade, 1952; Sjöstrand, 1953).

The outer membrane contains pore-forming proteins, porins that allow the free movement of uncharged molecules smaller than 5 kDa, while larger molecules are transported across the membrane through a group of assisting proteins called translocases. The presence of porins in

the outer membrane results in a minimal potential across the outer membrane area (De Pinto and Palmieri, 1992; Mannella, 1992). On the other hand, the inner membrane has a relatively high proportion of cardiolipin, which makes it impermeable to hydrophilic molecules. Thereby, special transport channels are required to regulate the transport of ion/molecules and metabolites over the inner membrane (Walther and Rapaport, 2009). This ion selectivity in the inner membrane enables the development of an electrochemical membrane potential of about 180 mV across the inner mitochondrial membrane area (Kühlbrandt, 2015) that can then use to drive ATP-synthesis through the action of ATP-synthase. Other than hosting special transporters, the inner mitochondrial membrane is studded with various proteins involved in the electron-transport chain and in ATP-synthesis (Palmer and Hall, 1972). The inner membrane surrounds the mitochondrial matrix that contains the enzymes involved in pathways for the oxidation of food i.e., the citric acid cycle,  $\beta$ -oxidation of fatty acids and the amino-acid metabolic pathways (Berg et al., 2012). In addition to this, the mitochondrial DNA and the machinery required for its replication, transcription, and translation is also present in the matrix. According to one estimate, the mitochondrial matrix has a protein concentration of up to 500 mg/ml which is equivalent to that in protein crystals (Kühlbrandt, 2015).



**Figure 1:** Schematic representation of Mitochondria.

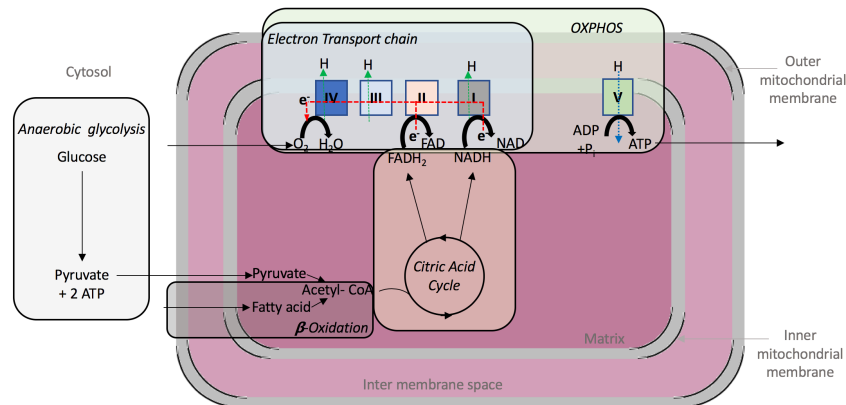
### 1.3 MITOCHONDRIAL BIOGENESIS

Early electron-microscope studies showed that mitochondria were sausage shaped, approximately 3-4  $\mu\text{m}$  long and 1  $\mu\text{m}$  wide. The number of mitochondria per cell varies from cell to cell generally depending on the activity and energy demand of the cell type. For example, a typical mammalian hepatocyte contains up to 1000-2000 mitochondria (Berg et al., 2012), occupying about 20% of cell volume. However, mitochondria are highly dynamic organelles that are continuously undergoing fission and fusion, hence the traditional approach of considering the mitochondrion as a single sausage-shaped unit is somewhat misleading. Therefore, the term used to describe the mitochondria in cells should rather be mitochondrial network instead of discussing them as individual units (Chan et al., 2004; Chen and Chan, 2004; Shaw and Nunnari, 2002). Mitochondrial fusion allows a homogenous distribution of proteins, metabolites, and mtDNA in the mitochondrial network of the cell. Mitochondrial fission allows an increase in mitochondrial mass in times of high-energy demands. Both

mechanisms contribute to the quality control of mitochondria where fusion allows the complementation of damaged components while fission allows sorting that may lead to the removal of damaged molecules through mitophagy (Baker et al., 2011).

## 1.4 METABOLISM IN MITOCHONDRIA

The mitochondrion is known as the ‘powerhouse of the cell’ due to its central role in producing ATP, the major energy currency of the cell. In the absence of mitochondria, cells are dependent on anaerobic respiration/glycolysis to generate all the required ATP. In anaerobic glycolysis, one glucose molecule generates only a net of 2 ATP molecules. In contrast, using mitochondrial respiration one glucose molecule yields 32 ATP molecules. The ATP is exported from mitochondria and is provided to all energy requiring reactions of the cell. The mitochondrial matrix, houses a plethora of enzymes required for the oxidation of fatty acids, carbohydrates, and amino acids. This process generates high-energy electron carriers such as NADH and FADH<sub>2</sub> (Figure 2) that then, in turn, transfer the electrons to the electron-transport chain.



**Figure 2:** Schematic representation of the ATP-production pathway in Mitochondria.

The electrons are transported through a series of four respiratory complexes embedded in the inner membrane to their final electron acceptor, oxygen. While the electrons move through the electron-transport chain, protons are pumped into the intermembrane space to create a mitochondrial membrane potential ( $\Delta\Psi$ ) gradient. Finally, to balance out the gradient, protons are channeled into the mitochondrial matrix through a rotor-type ATP synthase. This movement of protons induces conformational changes in the active site of ATP synthase and catalyze the conversion of ADP and P<sub>i</sub> into ATP (Berg et al., 2012).

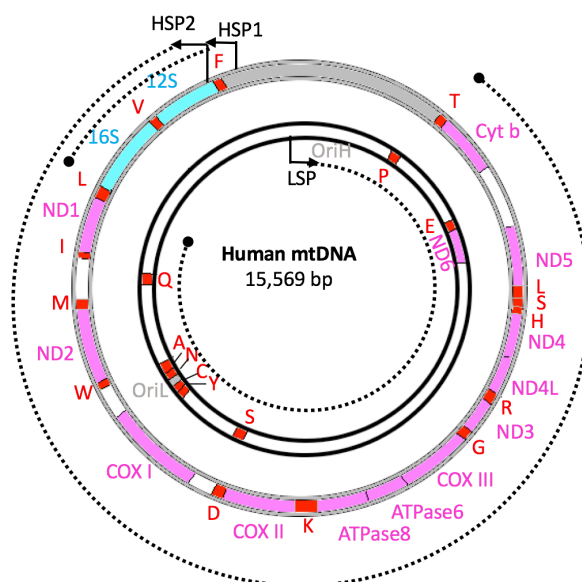
## 1.5 TRANSPORT TO MITOCHONDRIA

Human mitochondria require almost 1500 proteins and only 13 of them are mitochondrially encoded. All the remaining proteins are encoded by the nuclear DNA, translated by the cytosolic ribosomes and transferred to the mitochondria (Meisinger et al., 2008). Generally, the proteins destined for mitochondria have a basic N-terminal targeting sequence known as mitochondrial-targeting sequence (MTS), which is cleaved off after the protein enters the mitochondria. The translocation of the targeted proteins is carried out by the dynamic super

complexes formed by translocases of the outer membrane (TOM) and translocases of inner membrane (TIM) as seen by for example electron cryotomography (cryo-ET) (Gold et al., 2014). These two protein complexes span the intermembrane space and connect. Once the protein is delivered into the mitochondrial matrix. The tag is cleaved off and the protein is folded further by chaperones to function properly.

## 1.6 THE MITOCHONDRIAL GENOME

The human Mitochondrial DNA (mtDNA) is a small circular molecule made up of almost 17 kbp (16568 base pairs) (Bogenhagen and Clayton, 1974). mtDNA is relatively compact as compared to its bacterial ancestor (Andersson et al., 2003). One of the two strands is rich in guanines and the strands can be separated from each other in CsCl<sub>2</sub> gradient thereby, giving rise to the names heavy strand (H-S) and light strand (L-S) (Battey and Clayton, 1978). The genetic information is distributed on both strands and encodes 13 mRNAs, 22 tRNAs and two rRNAs (Anderson et al., 1981) (Figure 3).

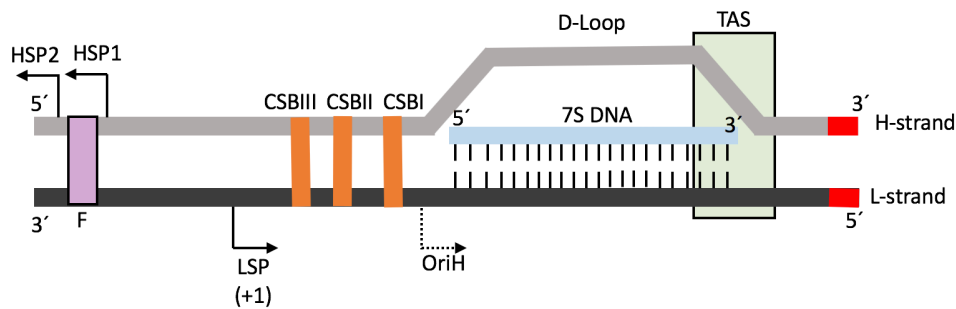


**Figure 3:** Color-coded schematic representation of human mtDNA. The transcription start site, TSS of the promoters present on their respective strands are represented by thick arrows pointing in the direction of the transcription. The transcript formed is represented by dotted line. The non-coding regions are represented in grey color while the regions coding for rRNA, mRNA and tRNA are shown in cyan, pink and red respectively.

There are two non-coding regions, one large and one small, located almost two-third parts away from the first one. The large non-coding region, NCR known as the control region is variable in different species but contains conserved patches of sequences. It has the promoter sequences for transcription, three conserved-sequence block regions, the origin of replication for the H-strand and a termination-associated sequence, TAS (Shadel and Clayton, 1997). Each strand has its own promoter, namely L-strand promoter, LSP and H-strand promoter, HSP and both are present in the NCR but transcribe in opposite directions giving rise to bidirectional transcription. (Montoya et al., 1982). Transcription from LSP is also responsible to produce a primer for replication of the H-strand (Chang and Clayton, 1985; Pham et al., 2006; Wong and Clayton, 1985). Another feature is a three-strand structure formed in this



region, called D-loop. Replication from OriH often undergoes premature termination giving rise to a ssDNA, referred to as 7S DNA. This 7S DNA often remains attached to the strand resulting in a three-stranded structure called as Displacement loop or D-loop (Figure 4).



**Figure 4:** Schematic representation of the D-loop region of human mtDNA. The TSS of LSP, HSP1 and HSP2 are marked by thick arrows. The origin of replication of the H-strand, OriH is marked by dotted line. The CSB region and 7S DNA are marked in orange and light blue, respectively.

The second non-coding region is located almost 11,000 bp downstream of HSP (Falkenberg, 2018), and contains the replication origin for the L-strand, OriL (Shadel and Clayton, 1997). The process of replication and transcription are discussed further in Chapter 2 and 3, respectively. Furthermore, the tRNAs of mitochondria have a different genetic code as compared to the cytoplasmic tRNAs. For example, in mitochondria, the stop codon UGA codes for tryptophan and AGG is instead the termination code while AGG codes for arginine in cytosolic translation (Anderson et al., 1981). Furthermore, mitochondrial ribosomes have a higher protein to RNA ratio (Agrawal and Sharma, 2012). Additionally in mitochondria, there are relatively few untranslated regions, no introns, and the open reading frames for proteins and rRNAs are often separated by tRNAs giving rise to the proposal of a tRNA punctuation model for mitochondrial gene expression (Hällberg and Larsson, 2014) as discussed in Chapter 4.

mtDNA is present in multiple copies per cell and mtDNA copy number generally depends on the energy demand of the cell. It may vary from 100,000 in human oocytes (Chen et al., 1995) to 3000 in human fibroblasts (Brown et al., 2011; Kukat et al., 2011). mtDNA is packed into nucleoprotein structures of 100 nm diameter known as **nucleoids** (Kukat et al., 2011). There is roughly one genome per nucleoid. The packaging protein is the mitochondrial transcription factor A, TFAM which binds the mtDNA and induces a U-turn. TFAM is also known to form homodimers. The binding, bending and dimerization induced by TFAM gives a compaction of mtDNA. TFAM is also an important factor for transcription, a role that will be discussed in Chapter 3. Several other proteins are also found in the nucleoid, such as the mitochondrial single-stranded DNA-binding protein (mtSSB), transcription factors and also some translational factors that might play a role in compaction (Kukat and Larsson, 2013).



## 2 MITOCHONDRIAL REPLICATION

During mitochondrial replication, the two strands of DNA are replicated simultaneously by the replication machinery, which is composed of a DNA polymerase, a helicase and several accessory proteins such as single-strand binding protein, topoisomerases and processivity factors.

### 2.1 REPLICATION IN BACTERIOPHAGE T7 DNA POLYMERASE

The T7 bacteriophage replication is well studied and can be reconstituted *in-vitro* using four subunits: T7 DNA polymerase, a combined helicase and primase, single strand binding protein and the host thioredoxin that act as a processivity factor and increases the processivity of the T7 DNA polymerase up to 100-fold (Hamdan and Richardson, 2009).

### 2.2 REPLICATION IN BACTERIA

The replication machinery of *E.coli* is well studied. The genome has one unique origin of replication called OriC, which is recognized by the origin recognition protein, DnaA. DnaA wraps around the DNA followed by the loading of the helicase DnaB with the help of its assistance proteins, DnaC. DnaA unwinds and exposes ssDNA, which is read by the primase DnaG to synthesize RNA primers. Once the RNA primers are available, the DNA polymerase, DNA polymerase III initiates DNA replication (Berg et al., 2012).

### 2.3 REPLICATION IN MTDNA

The DNA replication machinery in mitochondria can be reconstituted *in-vitro* using the following four factors: a single-subunit DNA polymerase (POL $\gamma$ A), a dimeric processing factor (POL $\gamma$ B), a hexameric helicase known as TWINKLE and a single-strand binding protein (mtSSB (Korhonen et al., 2004)). The replisome machinery is presumably similar to the bacteriophage T7 DNA polymerase since the POL $\gamma$ A and the helicase are distantly related to the bacteriophage T7 counterparts (Shutt and Gray, 2006), while the mtSSB has a homolog in *E. coli* (Tiranti et al., 1993). Regarding the mechanism of replication of mtDNA, a strand displacement model was first proposed in 1972 based on electron microscopy analysis (Robberson et al., 1972). This model was later well supported by biochemical observations (Clayton, 2003) and recently by genome-wide mapping of the mt-SSB protein (Miralles et al., 2014). According to this model, the primase, mitochondrial RNA polymerase (POLRMT) synthesizes an RNA primer that is used for the replication of the H-strand at OriH present in the D-loop region. The parental H-strand is displaced and covered by the mtSSB protein. The replisome continues thereby creating a nascent H-strand. After almost 11,000 bases, the ssDNA of the OriL region in the parental H-strand is exposed. The stem-loop formation of the OriL attracts the primase, POLRMT to synthesize a short RNA that will be used by the replisome to replicate the L-strand. Therefore, the synthesis of both strands is continuous in mtDNA and the replication of the H-strand is required to replicate the L-strand (Falkenberg, 2018).

Interestingly, most of the time the replication events are not always fully completed, and a pre-termination event occurs just after 650 nucleotides, which corresponds to the termination-associated sequence, TAS region in mtDNA. The helicase, TWINKLE is shown to have low occupancy in this region (Ikeda et al., 2015; Milenkovic et al., 2013). Later experiments show that mtDNA copy number is linked to the levels of TWINKLE in the cell, implying that the replication in mtDNA can be regulated at a pre-termination level rather than an initiation level (Falkenberg, 2018).

Several other proteins also play a role in mtDNA replication. For example, TOP1MT, the mitochondrial topoisomerase to reduce supercoiling in mtDNA (Stewart et al., 1998) (Falkenberg, 2018); RNase H1, for digesting the RNA primers (Holmes et al., 2015) and DNA ligase III for ligating the Okazaki fragments (Puebla-Osorio et al., 2006).

### 3 MITOCHONDRIAL TRANSCRIPTION

Transcription is generally divided into three phases known as initiation, elongation, and termination. During initiation, a specific region, usually upstream of the transcription start site (TSS) is recognized by the transcriptional machinery. This recognition causes the binding of the RNA polymerase (RNAP) to the TSS, followed by separation of the two strands at TSS. This is known as the transcription bubble formation. Once the two strands are separated, RNA starts to be synthesized and the transcriptional machinery enters the elongation phase until the machinery encounters a termination signal and is released. The transcriptional machinery can be as simple as a single protein subunit as common in viruses or highly complex involving multiple subunits, typically more than 10 in mammalian nuclear transcription (Berg et al., 2012). In the following section, the transcription in bacteria, eukaryotic nucleus, and in bacteriophages is summarized briefly, while the transcription in mitochondria will be discussed in depth.

#### 3.1 TRANSCRIPTION IN BACTERIA

In *E. coli*, the transcriptional machinery is composed of a multi-subunit RNAP, where the core comprises two  $\alpha$ , one  $\beta$  and one  $\beta'$  and a transient component,  $\sigma$  that identifies the promoter sequence upstream the TSS (-10 to -35) and recruits the subunits of core RNAP to the TSS. Once transcription is initiated, the  $\sigma$  subunit leaves the complex. The elongation is carried out by the core polymerase until it reaches the termination signal. In *E. coli*, the termination signal is rather simple and is composed of a symmetrical inverted repeat of a GC-rich sequence followed by a stretch of AT. The GC rich invert sequence is capable to form hairpin structures that halt the transcribing RNAP. The AT-rich region transcribes into a stretch of uracil in the newly transcribed RNA strand. As the affinity between adenines and uracils present in template DNA and nascent RNA respectively is weak, the RNA is released, and the termination event is completed. Factors like  $\rho$  can also be involved in the termination phase where they bind the RNA transcript and interfere with the active RNAP thus terminating the transcription (Berg et al., 2012).

#### 3.2 TRANSCRIPTION IN THE NUCLEUS OF EUKARYOTES

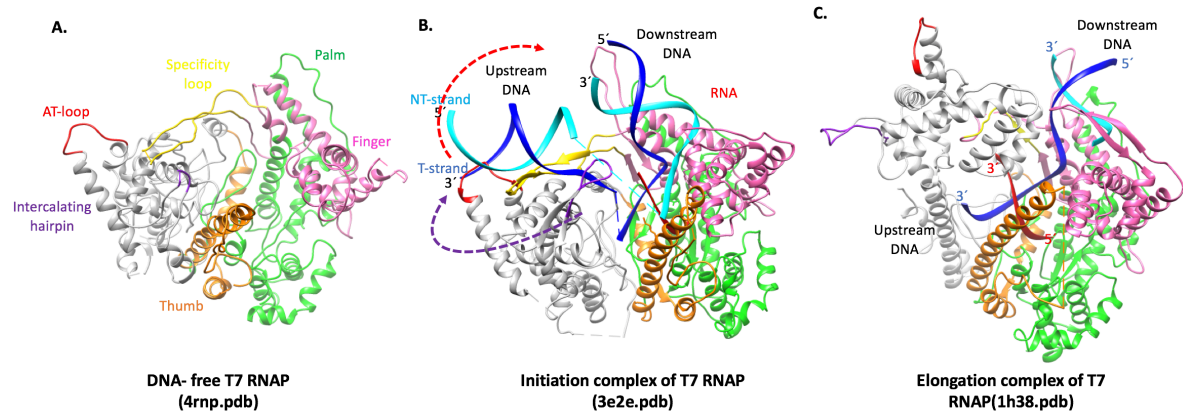
Nuclear transcription in eukaryotes is performed by three different types of RNAP: Pol I, Pol II, Pol III that mainly transcribe rRNA genes, protein-coding genes and tRNA genes, respectively (Berg et al., 2012). All three types of transcription machineries are composed of multi-subunit RNAP, which are similar to bacterial RNAP, a TATA-box binding protein and multiple transcription factors that assist at various stages of transcription. Pol II, responsible for transcribing the over-whelming amount of mRNAs, comprises of multiple DNA recognition elements in its promoters (TATA-box, recognition sites for TFIIB and initiator elements) and have six transcription factors (TFIIA, TFIIB, TFIID, TFIIIE, TFIIF and TFIIH) that makes the Pol II pre-initiation complex to be of megadalton size. This complex is not fully active and can only carry out basal transcription. The rate of transcription is modulated by gene-specific activators, repressors, mediators, enhancers and corepressors that provide a

second level of complexity to nuclear transcription. This is in addition to the proteins that are required for standard DNA maintenance. Termination of Pol I requires an 18 nt sequence that is recognized by a specific termination protein while the mechanism of termination in the other nuclear RNAPs is not well understood yet (Berg et al., 2012).

### 3.3 TRANSCRIPTION IN BACTERIOPHAGE T7

Bacteriophages have the least complex RNAP, since it is a single subunit protein. Still, it is capable of performing all steps of transcription by itself *in-vivo* and *in-vitro*. It is closely related to single subunit DNAP and RNAP, like *E. coli* DNAP and reverse transcriptase and is entirely different from the bacterial or eukaryotic RNA polymerases (Steitz, 2009; Tunitskaya and Kochetkov, 2002). Series of structural studies have been carried out in past, that have increased our understanding of single subunit and factor independent RNAPs. The structures of T7 RNAP available are DNA-free form (Sousa et al., 1993), promoter DNA-bound form (Cheetham et al., 1999; Cheetham and Steitz, 1999), initiation transcription complex with 3nt RNA-Bound complex (Cheetham and Steitz, 1999; ) (Durniak et al., 2008), early elongation complex (Tahirov et al., 2002) and of elongation complex (Yin and Steitz, 2002). T7 RNAP comprises of two domains, a regulatory N-terminal domain, NTD and catalytic C-terminal domain, CTD. The CTD can be divided into three domains: palm, thumb (326-411) and finger (554-784) domain similar to a cupped right hand. The template ssDNA binds at a deep cleft formed between the finger and thumb domains and reaches the active site in the palm domain (Figure 5). It is in the **palm** domain where the polymerization reaction between the rNTPS to form RNA takes place. The **Finger** domain separates the downstream DNA strands of the transcription bubble and the **thumb** domain acts as a clamp to prevent the disassociation of the complex and at the same time allows the complex to slide on the DNA. T7 RNAP binds the promoter sequence spanning -17 to +6. Furthermore, T7 RNAP recognizes and binds the promoters region through two of its structural elements: 1) the **AT-rich recognition loop** (93-101), a motif responsible for making specific contacts with DNA elements in -17 to -13 region and 2) the **Specificity loop** (739-770) that has positively charged residues at its tip and interacts with the region around the TSS and therefore, increase the promoter specificity. Once T7 RNAP binds the promoter, another important structural motif, the **intercalating hairpin** (232-242) separates the two strands around the TSS (-4 to +3) and a structure called the **open initiation complex** is formed. The separation of the two strands provides space for bond formation between the incoming NTPs (Figure 5B). During synthesis of the first 2 to 8 nucleotides, T7 RNAP remains bound to the promoter and translocate along the template strand (Cheetham et al., 1999). This dragging of the RNAP on the TSS induces topoisomerase stress in the DNA-RNA hybrid which induces the refolding of the NTD and places it away from the DNA. This step also releases the promoter and thus the involved transcription factors which result in unstructured promoter binding elements. The specificity loop is moved to create an **RNA exit pore** for funneling of the 5' end in nascent RNA (Figure 5C). The binding of RNA in the exit pore increases the stability of the transcription elongation complex and makes it more processive (Jeruzalmi and Steitz, 1998; Tahirov et al., 2002). The transition from initiation to elongation acts as a major barrier and

can result in several non-productive and abortive cycles of transcription (Tang et al., 2009). The **termination** modes in T7 RNAP can be classified into two categories. **Class 1**, which closely resembles the  $\rho$ -independent *E.coli* RNAP termination, involving a GC-rich hairpin followed by a poly(U) stretch (Hartvig and Christiansen, 1996). **Class 2**, on the other hand, have a seven bp DNA sequence, optionally followed by poly(U) stretch (Lyakhov et al., 1998). Therefore, the structure of a single subunit is well equipped with all the structural elements required to carry out a successful transcription in bacteriophages (Borkotoky and Murali, 2018).



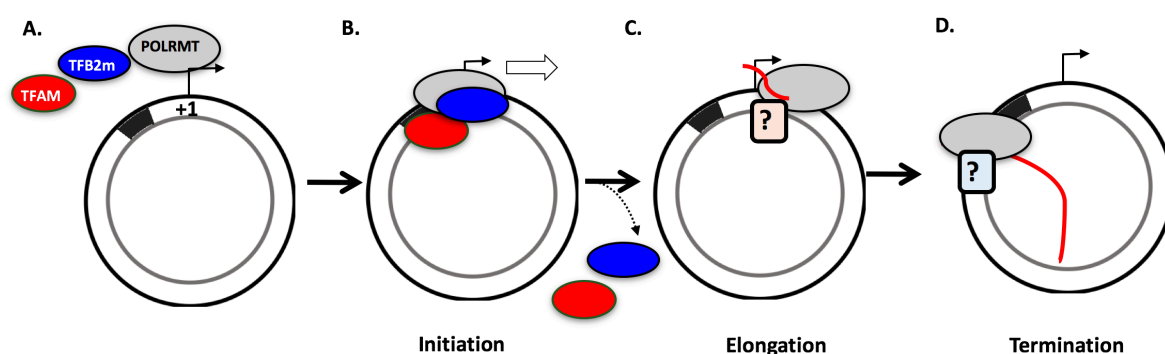
**Figure 5:** Conformational changes happening in T7 RNA P during transition to different phases. (A) Structure of DNA-free T7 RNAP (4rnp.pdb). Important regions are colored and labelled. See text for details (B) Structure of T7 RNAP during initiation complex (3e2e.pdb). Important regions are colored and labelled as in 5A. (C) Structure of T7 RNAP during elongation complex (1h38.pdb). Important regions are colored and labelled as in 5A. Note the movement of structural motifs involved in promoter binding (AT-loop in red) and melting (Intercalating hairpin in purple) upon transition from initiation to elongation phase.

### 3.4 TRANSCRIPTION IN MITOCHONDRIA

The mtDNA is a small circular molecule where the genetic information is encoded by both strands, therefore implying the requirement of at least one promoter per strand. There is consensus on a single promoter for the L-strand, **LSP**, while there is no agreement on an exact number of heavy strand promoters, **HSP**. In 1981, two independent studies found that the rRNAs transcribed by HSP are synthesized at a 50-100 times higher rate than other transcripts encoded by the same H-strand. Two reasons were put forward, firstly it could be a premature termination event downstream of the rRNA or it could be two promoters for HSP (Gelfand and Attardi, 1981; Ojala et al., 1981). Later *in-vivo* experiments supported the presence of two HSPs: HSP1 and HSP2 (Martin et al., 2005; Montoya et al., 1982). HSP2 is located almost 100bp downstream of HSP1. HSP1 initiates 19 nt upstream of the tRNA<sup>Phe</sup> gene and produce transcripts that cover tRNA<sup>Phe</sup>, 12S rRNA, tRNA<sup>Val</sup> and 16S rRNA. While HSP2 instead of stopping at 16S rRNA, supposedly transcribe the entire H-strand (Gustafsson et al., 2016; Montoya et al., 1982). H-strand codes for all rRNA, mRNAs and more than half of tRNAs. In addition to this, transcription from LSP is linked to replication, as it will generate promoters required for H-strand replication. Moreover, the *in-vivo* rate of L-strand transcription is found to be 2-3 times higher than that of H-strand (Cantatore and Attardi, 1980). Transcription initiation has been seen for all the three promoters *in-vitro*,

where HSP2 activity seems to be a marginal (Bogenhagen et al., 1984; Lodeiro et al., 2012). **This thesis will focus on the transcription events of LSP.**

Transcription from mtDNA is carried out by the mitochondrial RNA polymerase, **POLRMT**, a single subunit protein, which is distantly related to bacteriophage T7 RNA polymerase (Masters et al., 1987; Tiranti et al., 1997). Interestingly, unlike bacteriophage T7 RNAP, POLRMT cannot carry out transcription on its own and is dependent on factors at each stage of the transcription. It has been shown that two factors; mitochondrial transcription factor A, TFAM and mitochondrial transcription factor B2, TFB2m are required to initiate transcription from LSP (Gaspari et al., 2004; Morozov et al., 2014; Morozov et al., 2015; Posse and Gustafsson, 2017; Sologub et al., 2009b). In short, the transcription from mtDNA at LSP is initiated when the mitochondrial transcript factor A, TFAM, recognizes a promoter -12 to -14 region upstream the TSS (TSS is referred at +1). There, it binds and bends the DNA and recruits POLRMT and TFB2m to the TSS. TFB2m separates the two strands at TSS and allows the *de novo* synthesis of RNA (Figure 6).



**Figure 6:** Schematic representation of transcription in Mitochondria. (A) The proteins: POLRMT, TFAM, TFB2m will bind the Transcription start site (TSS) marked by +1 on LSP (B) During initiation, TFAM will recognize promoter and bind. TFAM will recruit the other two proteins, POLRMT and TFB2m. TFB2m will melt the dsDNA at the TSS, allowing the RNA to be synthesized. The exact order of binding of the two proteins is unknown. (C) During elongation, the initiation factors are released and the POLRMT enters the elongation phase. The step should be processive to transcribe the full length of L-strand. (D) At the end of cycle, the POLRMT and transcribed RNA is released and event is terminated.

The POLRMT then enters the elongation phase in which the initiation factors (TFAM, TFB2m) leave the complex while the processivity factor, transcription elongation factor, TEFM, binds the POLRMT and allows the synthesis of the near-genomic length transcripts from LSP. MTERF1 is reported to be the protein that completely blocks the transcription from the L-strand (Asin-Cayuela et al., 2005; Shi et al., 2016; Terzioglu et al., 2013).

### 3.4.1 Proteins involved in mitochondrial transcription

Biochemical studies together with structural investigation of truncated POLRMT, transcription initiation complex and transcription elongation complex, along with structures of mitochondrial transcription factor TFAM, TFB2m and TEFM have shed light on some of



the major aspects of mitochondrial transcription (Table 1). The factors involved and the phases of transcription are briefly discussed in the next section (Table 1).

#### 3.4.1.1 Mitochondrial RNA polymerase, (POLRMT)

POLRMT was first identified in yeast as Rpo41, an important factor required for maintenance of the mitochondrial genome (Greenleaf et al., 1986). Later, it was found that mitochondrial RNAP from both yeast and human are dependent on extra factors for carrying on transcription even if they are related to the single subunit T7 RNAP. Human POLRMT can initiate promoter-specific initiation *in-vitro* in the presence of two factors, TFAM and TFB2m.

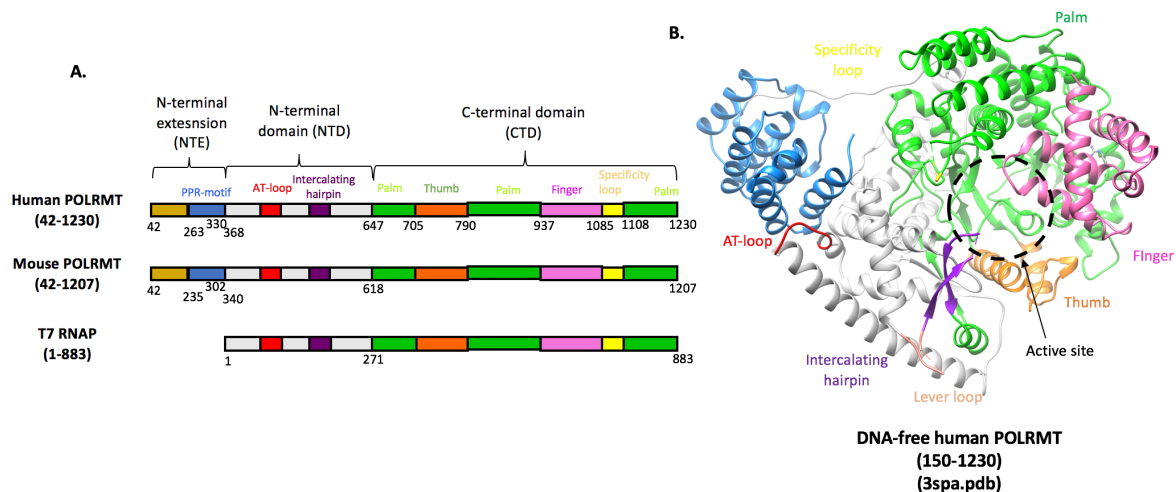
Human POLRMT is a protein of 1230 residues encoded by the nuclear genome. The initial 41 amino acids are used for targeting to mitochondria and the mature polypeptide is composed of 42-1230 amino acids. The structure of a **truncated apo POLRMT** (105-1230) shows that it has a unique extension on the N-terminus of the protein that is not present in the T7 RNA polymerase (Figure 7). This unique part can be divided into a non-structured N-Terminal extension, NTE (42-218), and a well-structured pentatricopeptide domain, PPR-domain (219-355), which is in addition to the T7 RNAP resembling parts: the N-terminal domain, NTD (356-623) and the C-terminal domain CTD (623-1230) (Ringel et al., 2011); 3SPA.pdb).

**Table 1:** Summary of the structures of the macromolecules involved in mitochondrial transcription.

Macromolecules	Details	Constructs used	PDB ID	Reference
<b>TFAM</b>	Bound to LSP (-39 to -12)	43-246	3TMM	(Ngo et al., 2011)
	Bound to LSP (-39 to -12)	43-246	3TQ6	(Rubio-Cosials et al., 2011)
	Bound to HSP (-29 to -8)	43-246	4NOD	(Ngo et al., 2014)
	Bound to a random sequence	43-246	4NNU	(Ngo et al., 2014)
<b>TFB2m</b>	Apo truncated protein	Initial 62 AA removed, and 268-294 loop replaced by GSSG linker	6ERO	(Hillen et al., 2017a)
<b>TEFM</b>	NTD	57- 134	5OL9	(Hillen et al., 2017b)
	CTD	153-356	OL8	(Hillen et al., 2017b)
<b>POLRMT</b>	Apo truncated	105-1230	3SPA	(Ringel et al., 2011)

<b>Initiation complex</b>	Proteins bound to LSP (-39 to +11) with bubble -4 to +3	TFAM (43-245), POLRMT (104-1230) and TFB2m (lacks 62 N-terminal residues, replaced long loop 268-294 with GSSG linker	6ERP	(Hillen et al., 2017a)
<b>Initiation complex</b>	Proteins bound to HSP1 (-39 to +11) with bubble -4 to +3	TFAM (43-245), POLRMT (104-1230) and TFB2m (lacks 62 N-terminal residues, replaced long loop 268-294 with GSSG linker	6ERQ	(Hillen et al., 2017a)
<b>Elongation complex</b>	POLRMT bound to DNA-RNA scaffold	POLRMT (150-1230)	4BOC	(Schwinghammer et al., 2013)
	POLRMT and TEFM bound to DNA-RNA scaffold	POLRMT (150-1230) CTD of TEFM (153-356)	5OLA	(Hillen et al., 2017b)

The **CTD** resembles the CTD of T7 RNAP and has a cupped right-hand architecture containing thumb (711-790), finger (954-1044) and the active site containing palm domain hinting that the catalytic mechanism is overall conserved. Locally, the **finger** domain that mediates catalysis through its O-helix and separates the downstream dsDNA after TSS, is rotated by 25° and thus has a ‘clenched’ conformation. The **thumb residues** (711-790) that act as a clamp and increase the processivity of POLRMT, has a partial density in the solved structure (Figure 7).



**Figure 7:** Domain architecture of Mitochondrial RNA polymerase, POLRMT. (A) Schematic representation of the domains compared in human POLRMT, mouse POLRMT and T7 RNAP. N-terminal extension including the PPR motif is missing in T7 RNAP. 1-42 region of human and mouse POLRMT is the mitochondrial targeting sequence, MTS and is removed when the proteins enter mitochondria (B) Structure of human POLRMT (150-1230; 3spa.pdb)

The **NTD** is partially similar but the structural motifs are repositioned. Particularly, the **AT-rich recognition loop**, responsible for promoter recognition is now involved in protein interactions with a unique pentatricopeptide repeat domain, PPR. The sequestering of AT-loop hints that this region can no longer play its destined role of promoter recognition. The second promoter recognition element, the **specificity loop**, makes specific contacts with the major groove upstream of TSS has no density in the available crystal structure. The third motif, responsible for separating the two strands at TSS in T7 RNAP, is the **intercalating hairpin**. This hairpin is repositioned 7Å away from the active site. Mutational analysis showed that variants of POLRMT with deleted intercalating hairpins can only initiate transcription from pre-melted templates (a dsDNA strands with mismatched NTPs at TSS to separate the two strands artificially), while the exposing of AT-loop by deleting the PPR region did not make POLRMT independent of TFAM. The overall structure explains the dependency of POLRMT on the transcriptional factors, TFAM and TFB2m.

Regarding the **NTE and PPR** domains that are found in POLRMT and absent in bacteriophage T7 RNAP, PPR-domains are commonly present in RNA-binding and processing proteins in plants and animals (Lightowers and Chrzanowska-Lightowers, 2008; Rodeheffer et al., 2001; Small and Peeters, 2000). In the POLRMT structure, the PPR is connected to the NTD by a proline-rich linker that serves as a hinge to move the PPR domain with respect to NTD. However, the exact role of NTE remains unknown. There is missing density for the NTE in the apo-POLRMT structure. However, the structure suggested that the unique N-terminal region of POLRMT will undergo a conformational change during transcription (Ringel et al., 2011).

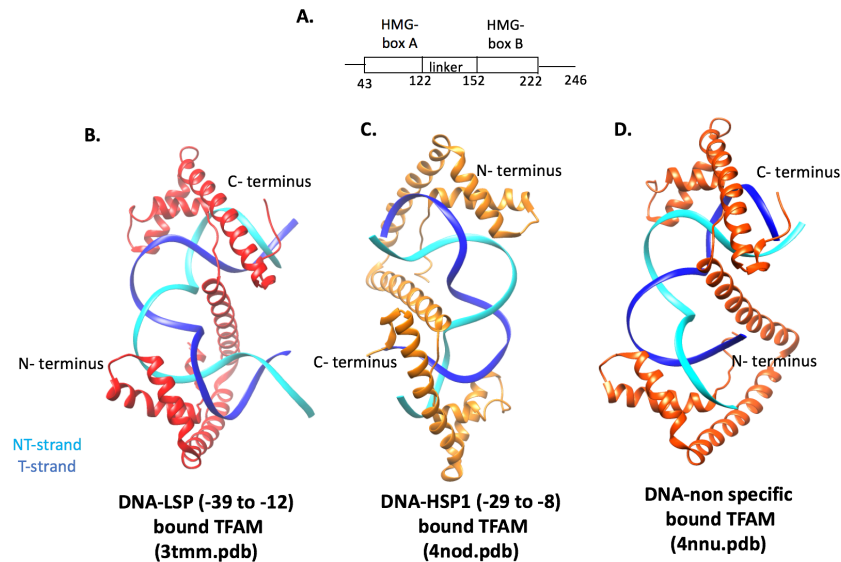
In yeast, the removal of the NTE results in mitochondrial genome instability. However, the phenotype can be rescued if the NTE is expressed separately (Wang and Shadel, 1999). Another study reported that loss of NTE in yeast results in the decrease of promoter-specific transcription but triggers the transition from initiation to elongation (Paratkar et al., 2011). Therefore, the NTE seems to be involved in promoter specificity. This was also supported in mouse RNAP, (mPOLRMT) where the N-terminal region plays a similar role, since removal of it in mPOLRMT makes a tight complex with TFB2m and makes the complex hyperactive. Specifically, transcription can be initiated from non-promoter templates and is no longer dependent on TFAM (Posse et al. 2014; Paper I). Based on the above analysis, it was proposed that TFAM recognizes and binds the DNA first and the POLRMT is recruited to the TSS by the interaction between NTE of POLRMT and TFAM (Posse et al., 2014; Paper I). The observation was backed up by another study (Morozov et al., 2014; Posse et al., 2014). The exact structural details of the interaction of NTE are unknown. In a similar study, an interaction was reported between mPOLRMT and upstream DNA around position -50 to -60 (Posse et al., 2014; Paper I) and this were later mapped to position -49 (Morozov et al., 2015). The region was shown to interact with NTD of the POLRMT (Posse et al., 2014; Paper I).

As described above in the mitochondrial replication section, POLRMT also plays a role in replication by acting as a primase for H-strand. This is supported by biochemical data where depletion of POLRMT leads to a decrease in 7S DNA and mtDNA levels. At low POLRMT level, replication is favored over transcription as indicated by the priority to transcribe L-strand instead of H-strand (Kühl et al., 2016). However, the mechanism of POLRMT dependent OriH initiation and the mechanism of primer formation are not comprehensively understood.

mPOLRMT can also initiate transcription from ssDNA independent of transcription factors but it is not very processive and produce transcripts of 20-100 nucleotides. Because of this characteristic feature, POLRMT can work as a transcribing primase of L-Strand form OriL (discussed in the replication section). **OriL** forms stem-loop regions where the single strand loop region contains a stretch of poly d(T). There is no structural information for the complex of POLRMT with the OriL region. It can be speculated that the stem-loop OriL is protruding out and can provide a binding surface for the POLRMT, making it TFAM independent. Melting is not required on ssDNA, therefore, there is no need for TFB2m (Fusté et al., 2010; Wanrooij et al., 2008). But the exact mechanism of interaction remains to be investigated.

#### *3.4.1.2 Mitochondrial transcriptional factor A, TFAM*

TFAM was first identified to be responsible for promoter-specific initiation by POLRMT (Fisher and Clayton, 1985) and the DNA binding region was mapped to -12 to -40 bp upstream of TSS. Other studies mapped the TFAM binding site on LSP using DNase protection I assay and reported to be between -10 and -35 (Dairaghi et al., 1995; Fisher et al., 1987). TFAM belongs to the High Mobility Group (HMG) box containing family of proteins, where TFAM has two HMG boxes separated by a linker region and has a C-terminal tail. In addition to specific interactions in the LSP promoter region, TFAM was also found to bind DNA non-specifically similar to its yeast homolog, Abf2 (Bakkaiova et al., 2016). Structural investigations of TFAM bound to DNA includes crystal structures of TFAM with the LSP promoter region (Ngo et al., 2011; Rubio-Cosials et al., 2011), the HSP promoter region (Ngo et al., 2014) and a non-specific DNA region (Ngo et al., 2014) (Figure 8). The structures suggest that first HMG box, HMG1 binds and bends DNA at 90° and the second HMG box, HMG2 does the same, giving rise to a U-turn in dsDNA (Hällberg & Larsson, 2011). Interestingly, the TFAM bound to HSP1 appeared in an opposite direction, suggesting that the transcriptional machinery has different topology at LSP and HSP1 (Ngo et al., 2014). Two later studies; one biochemical and one structural, however has argued against this theory and suggested similar organization of transcriptional machinery on both promoters (Hillen et al., 2017a; Morozov and Temiakov, 2016). Furthermore, these later publications argues that the discrepancy occurs since the structural study of TFAM-HSP1 (Ngo et al., 2014) only used a partial sequence of HSP1 (-29 to -8) as compared to -39 to -16, which was proposed by (Fisher et al., 1987).



**Figure 8:** Domain architecture of human TFAM. (A) Schematic representation of domains presents in human TFAM, Two High mobility groups (HMG) boxes are separated by a linker region. At the C-terminus there is a flexible tail (B) Structure of TFAM bound to DNA corresponding to LSP (-39 to -12). HMG1 binds the proximal part of DNA while HMG2 binds the distal part. Both bend the DNA altogether to 180°. Two groups have reported the structure (3tmm.pdb and 3tq6.pdb) (C) Structure of TFAM bound to DNA corresponding to HSP1 (-29 to -8). HMG1 binds the distal part of DNA while HMG2 binds the proximal part (4nod.pdb) (D) TFAM binds and bends a non-specific DNA in the similar manner

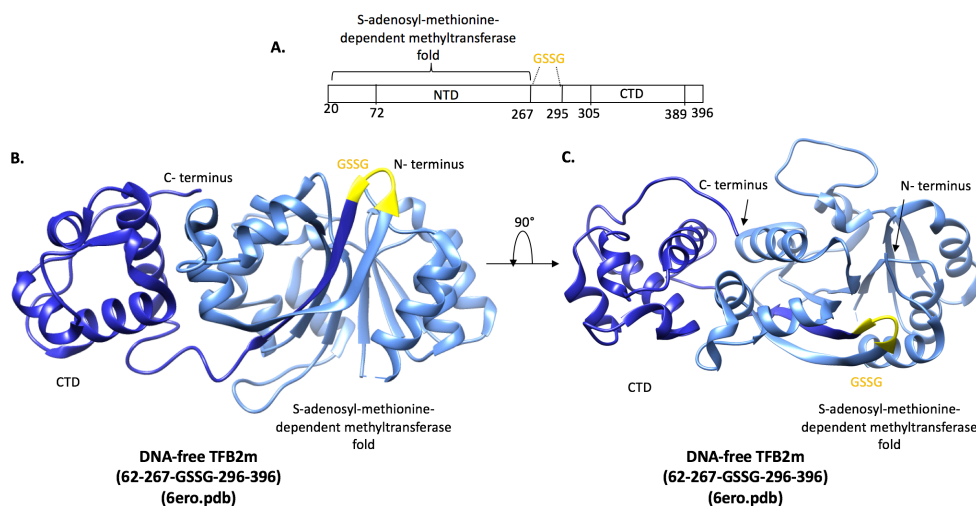
Given the reports that TFAM is making physical contact with POLRMT, the TFAM has to be positioned precisely and in this aspect the placement at -39 to -16 on HSP1 might be right. Earlier, the C-terminal tail of TFAM was found to be crucial for transcription initiation as removal of a C-terminal tail of TFAM (222-246 amino acids) severely impairs promoter-dependent transcription. The yeast homolog, Abf2 lacks this c-terminal tail and solely acts as a DNA packaging factor. Upon adding the C-terminal tail of TFAM to Abf2, the Abf2 was able to initiate promoter-specific transcription in a human system, suggesting that the TFAM C-terminal tail -POLRMT interaction, is required for promoter recognition (Morozov et al., 2014). Disruption of the TFAM gene in mouse heart ((Li et al., 2000; Wang et al., 1999) (Wredenberg et al., 2002), muscle (Wredenberg et al., 2002), pancreas (Silva et al., 2000) and nerve cells (Sörensen et al., 2001) leads to tissue-specific depletion of mtDNA and mtDNA transcripts causing severe respiratory chain deficiency. This shows that TFAM is indispensable for proper cell functioning. Furthermore, it has been shown that TFAM is present in saturating amounts to cover the DNA *in-vivo* (Kanki et al., 2004; Kukat et al., 2011; Takamatsu et al., 2002). Arguably saturating levels of TFAM will make transcription impossible, an observation further supported by the *in-vivo* methylation protection and crosslinking experiment that suggest the presence of partial naked regions in mtDNA (Ghivizzani et al., 1994; Lott et al., 2013; Pardue et al., 1984).

#### 3.4.1.3 Mitochondrial transcription factor B2, TFB2m

The second transcription factor in mammalian mitochondria is TFB2m. TFB2m and its homolog TFB1m were first identified on the basis of sequence similarity to the yeast mitochondrial transcription factor, Mtf1, the structure of which is similar to

methyltransferases, involved in maturation of ribosomal RNA (Falkenberg et al., 2002; McCulloch et al., 2002). Both were found to be tentative participants in *in-vitro* transcription experiments where TFB2m was found to have 10-fold higher activity than TFB1m (Falkenberg et al., 2002). TFB1m, on the other hand, was found to retain its role as a mitochondrial rRNA methyl transferase as knocking down of TFB1m results in impaired translation (McCulloch and Shadel, 2003; Metodiev et al., 2009). TFB2m has evolved into a *bona fide* mitochondrial transcription factor, where it is believed to be required for promoter melting (Litonin et al., 2010; Sologub et al., 2009a). TFB2m is not required for transcription on pre-melted templates (Sologub et al., 2009a) or ssDNA templates (Posse and Gustafsson, 2017) underlining its proposed role in promoter melting.

TFB2m is a 45 kDa nuclear-encoded protein comprised of 396 protein residues with a putative MTS containing 19 aminoacids. Recently, the structure of a truncated and modified TFB2m was reported (Hillen et al., 2017a). It has an N-terminal domain, NTD (97-267) and C-terminal domain, CTD (295-396), which are separated by a long linker loop (268-294). The NTD resembles rRNA methyltransferases while the CTD forms a helical domain with the disordered tail. To enable structure determination by X-ray crystallography, the long linker loop was replaced by a GSSG linker and 62 residues from N-terminus of TFB2m were removed (Figure 9). Replacement of the long linker region doesn't seem to have an effect on the TFB2m function while the removal of 62 residues results in a protein which has lost 40% of activity. In the yeast case, the long linker region in Mtf1 is linked to Mtf1 and Rpo41 interactions (Cliften et al., 1997). The structure of modified TFB2m shows that it resembles the mouse TFB1m structure with some divergence that includes a loop in the NTD that contributes to the positive surface formed by TFB2m at the interface with DNA (Metodiev et al., 2009) (Hillen et al., 2017a; McCulloch et al., 2002).



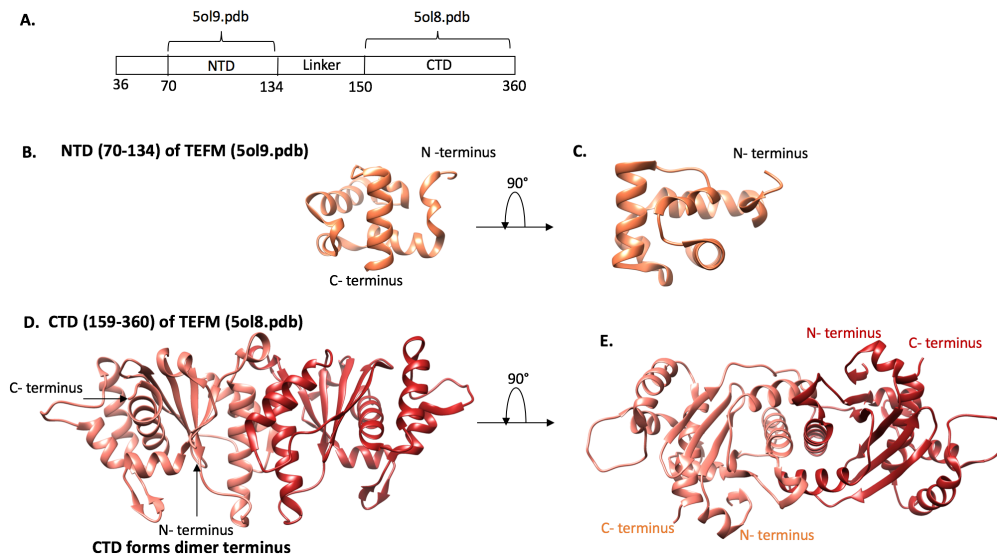
**Figure 9:** Domain architecture of human TFB2m. (A) Schematic representation of domains presents in human TFB2m, N-terminal domain is S-adenosyl methionine dependent methyl transferase domain and the C-terminal globular domain. Both are linked by a long flexible linker (267-295), which was replaced by a GSSG stretch for crystallography purpose. Also, the initial flexible 62 residues were removed (B) Structure of human TFB2m, NTD and CTD domain is colored in light and dark blue respectively. The GSSG linker is shown in yellow (6ero.pdb) (C) The view in B is rotated to 90° to see the top view.



#### 3.4.1.4 Mitochondrial transcription elongation factor, TEFM

TEFM was first discovered in RNA interference-mediated knockdown experiments where its depletion resulted in severe respiratory-chain dysfunction due to dramatic decrease in distal transcription from the promoters. It was proposed to interact with the C-terminal part of the POLRMT suggesting that it is an accessory subunit of POLRMT (Minczuk et al., 2011). The *in-vivo* experiments were confirmed up by the *in-vitro* transcription experiments showing that the addition of TEFM dramatically stimulates the processivity of POLRMT (Agaronyan et al., 2015); Posse et al., 2015 i.e. Paper II). During transcription from LSP, the POLRMT prematurely terminates at a conserved sequence block II region, CSBII (Pham et al., 2006) due to the formation of G-quadruplexes. Addition of TEFM to the transcriptional machinery completely abolishes this termination event (Agaronyan et al., 2015); Posse et al., 2015 i.e. Paper II). Further details about TEFM role in transcription are covered in Paper II.

TEFM is a nuclear encoded 36 kDa protein with a 36 AA of MTS, which upon cleavage gives the mature protein containing an N-terminal domain, NTD (35-135) and a C-terminal domain, CTD (150-360) joined by a flexible linker region (135-150). NTD is predicted to be a helix hairpin helix (HhH) motif and the CTD has an RNase H fold. This was also confirmed by a recent structure of TEFM where NTD and CTD were separately crystallized (Hillen et al., 2017b)(Figure 10). The CTD has an RNase H-like fold comprising of a central  $\beta$ -sheet and four C-terminal  $\alpha$ -helices, where the last helix forms a dimeric interface between two TEFMs. The fold is commonly found in the family of Holliday junction resolvases (Górecka et al., 2013; Wyatt and West, 2014). Even though it has an RNase H like fold, it does not have an endonuclease activity (Hillen et al., 2017b).



**Figure 10:** Domain architecture of human TEFM. (A) Schematic representation of domains presents in human TEFM, N-terminal domain and the C-terminal domain linked by a long flexible linker (134-150). (B) Structure of NTD of human TEFM (50l9.pdb) (C) Structure of dimer of CTD of human TEFM in two views (50l8.pdb).

Interestingly, the CTD structure is similar to the *Schizosaccharomyces pombe* holiday-junction resolvase, Cce1, which plays an important role in yeast mitochondrial genome replication (Ceschini et al., 2001; Lockshon et al., 1995). The globular NTD is unique to

TEFM and is made up of a Helix hairpin Helix motif (HhH), commonly present in DNA binding proteins. The structure of NTD resembles the HhH domain of the bacterial *tex* protein. It was found that the NTD is dispensable for TEFM activity (Hillen et al., 2017b).

The structure of POLRMT bound to TEFM and DNA-RNA elongation scaffold (Hillen et al., 2017b) shows how the TEFM is interacting with the other components in the elongation phase and the structural mechanism is discussed in the 3.4.2.2 section covering mitochondrial transcriptional elongation.

POLRMT from a thermophilic fungus, *Chaetomium thermophilum*, is 40% identical to its human counterpart. Bioinformatics study shows that there is an 80 amino acid stretch towards the c-terminus of the fungal POLRMT that is similar to regions strongly conserved in TEFM. This C-terminal extension is not found in human POLRMT. This suggests that the functionality of the mammalian TEFM protein could be a component of *Chaetomium thermophilum* POLRMT and other fungal POLRMT, possibly hinting the inbuilt multi-functional and self-sufficient role of POLRMT in the regulation of mtDNA transcription. However, the structural-functional relationship of this family of fungal POLRMTs remains to be explored.

#### 3.4.1.5 Mitochondrial termination factor 1, MTERF1

Mitochondrial termination factor 1, MTERF1, belongs to the MTERF family, in which all members are involved in mitochondrial gene expression. MTERF1 is found to be associated with transcription and has been shown to play an important role in replication (Hyvärinen et al., 2007; Shi et al., 2016). This protein specifically binds the dsDNA downstream of the 16S rRNA region inside the tRNA<sup>Leu</sup> gene (Kruse et al., 1989; Shang and Clayton, 1994). It was first identified as the terminator of both LSP and HSP transcripts and regulator of H-strand transcription that ensures that rRNA have an increased level of transcription which remains part of transcripts. On the contrary, another experiments, showed that MTERF1 is terminator from LSP and has no effect on rRNA and mRNA levels (Terzioglu et al., 2013). The structure of MTERF1 bound to tRNA<sup>Leu</sup> nucleotide sequence reveals that it is a strictly helical protein which wraps around the tRNA and partially unwinds it by flipping out three nucleotides (Jiménez-Menéndez et al., 2010; Yakubovskaya et al., 2010).

### 3.4.2 Phases involved in mitochondrial transcription

Putting all these players in the context of transcription, each phase is further discussed in the sections below.

#### 3.4.2.1 Mitochondrial transcription initiation

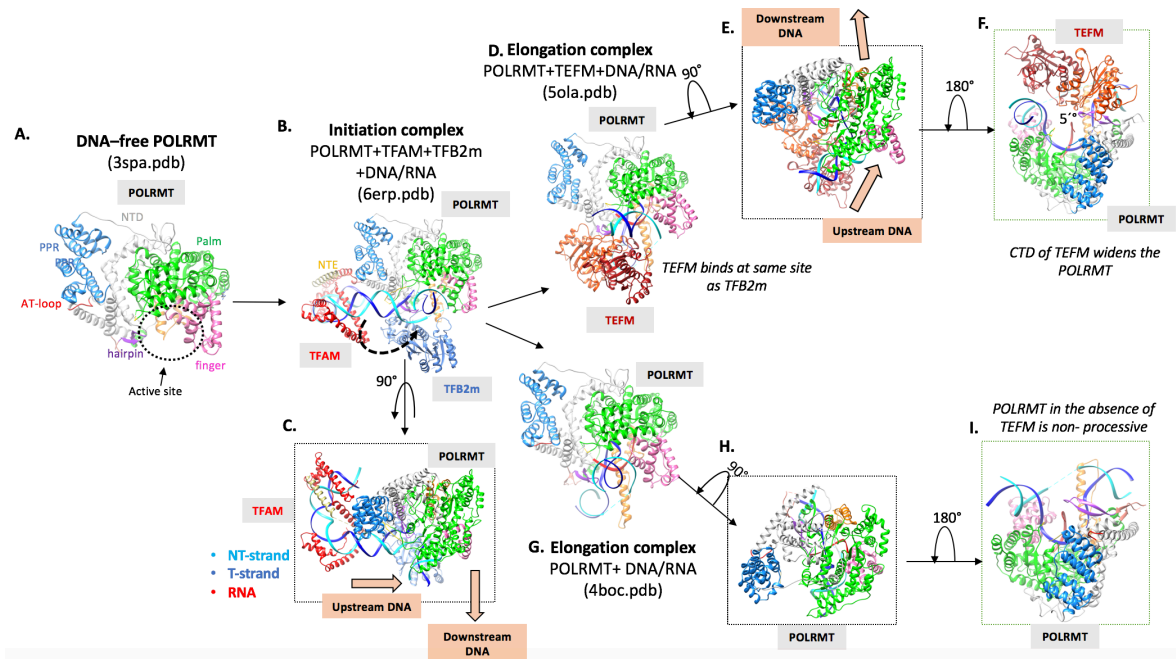
A recently published **structure of the transcription initiation complex** on LSP or HSP1 contained a pre-melted dsDNA template with the promoter region -35 to +10 and a bubble around the TSS -4 to +3 bubble (where +1 is TSS), TFAM (43-245), POLRMT (104-1230) and TFB2m (lacks 62 N-terminal residues, replaced long loop 268-294 with GSSG linker) (Hillen et al., 2017a). In short, during initiation the TFAM binds the upstream promoter -35



to +10 containing a region, bends the DNA, the distal part of the TFAM is contacted by the part of NTE in POLRMT which helps it to recruit it to TSS. The upstream dsDNA passes underneath the PPR domain towards the active site of the POLRMT, where the +1 TSS is unfolded by the TFB2m, which is supporting from underneath. TFB2m contacts the NT strand around TSS, thereby repositioning the intercalating hairpin and pushing the POLRMT outwards. The specificity loop of the POLRMT makes contacts with the major groove of the upstream DNA. The downstream dsDNA leaves the complex at 90° to the active site underneath the palm domain of the POLRMT (Figure 11).

In detail, the TFAM recognizes the promoter region, binds and bends the DNA, which makes a scaffold for the POLRMT. In addition to this, the distal HMG box of TFAM is approached by the tether helix (123-245) in POLRMT that allows precise positioning of POLRMT on the TSS. In the structure, TFAM does not make any contact with TFB2m. The C-terminus of TFB2m interacts with POLRMT causing a 25° rotation in the NTD. These interactions are at two points: the loop in the NTD of POLRMT, the lever loop (588-604) and the intercalating hairpin. The intercalating hairpin is moved by 7 Å which positions it between the two DNA strands. This confirms a previous interaction shown between mouse POLRMT and TFB2m. Furthermore, the interactions strengthen upon removal of the initial NTE and PPR domain of mouse POLRMT (Posse et al., 2014; Paper 1). Additionally, TFB2m holds on to the NT strand in the TSS region and therefore, stabilizes the melting of the dsDNA at the TSS. This is similar to the case of the bacterial sigma factor in bacterial transcription, even though it does not have any structural similarity (Murakami and Darst, 2003; Zhang et al., 2012). Other than this, the specificity loop appears to be ordered and runs along the major groove of DNA upstream of the TSS, hinting that it could interact with DNA. However, it is not established yet if the contact between DNA and protein is sequence-specific from the resolution obtained in the crystal structure.

As discussed above in the TFB2m section, the TFB2m construct used for crystallization has its initial 62 residues and a long loop region 268-294 replaced by a GSSG linker. The replacement of the long loop region has a pronounced effect and results in 40% reduction in protein activity (Hillen et al., 2017a). The initial 62 residues of TFB2m are believed to be interacting with the priming nucleotide (Morozov et al., 2015), thus implying that the contacts during initiation cannot be restricted to above-identified interactions and more remains open for further investigations. Moreover, a recent paper found out that a bubble in the TSS region makes the POLRMT, TFB2m independent and the transcription is similar to transcription from ssDNA (as happens in OriL). However, the exact mechanism of promoter melting by TFB2m remains unknown from the given data. It is agreed upon that TFB2m, like Mtf1 in the yeast system, plays a role during initiation and once the dsDNA is melted, TFB2m disassociates from the complex and POLRMT enters the elongation phase (Mangus et al., 1994).



**Figure 11:** Comparison of structures of POLRMT in different phases of transcription. (A) Structure of DNA-free POLRMT (150-2130). NTE (42-218) is missing in the crystal structure. Other important regions: AT-loop (red) is involved in promoter recognition, Hairpin (purple) is involved in DNA melting the +1 transcription start site, Palm (green), finger (pink) and thumb (orange) are involved in catalytic process (3spa.pdb). (B) Structure of Initiation complex (POLRMT, TFAM; TFB2m, DNA-LSP; 6erp.pdb). Upon DNA binding, there is not much change in the overall structure of POLRMT as TFAM recognizes the promoter region and bends it, POLRMT interacts with TFAM using its tether helix (123-245). TFB2m stabilizes the active site from bottom and interacts with the non-template (NT) strand at TSS. (C) View in B is rotated to see the upstream and downstream DNA path. (D) Structure of Elongation complex (POLRMT+TEFM+ES; 50la.pdb). On transition from initiation complex to elongation complex TFB2m leaves the complex and TEFM binds at its site. (E) View in D is rotated to see the DNA exit path. Note that the upstream DNA is now entering inside the complex using the downstream DNA path of Initiation complex and downstream DNA has a new exit path. (F) View in E is rotated to see the TEFM interactions. Dimer of TEFM sits on the downstream DNA and widens the POLRMT to give path to existing RNA. (G) Structure of Elongation complex in the absence of TEFM (POLRMT+ES; 4boc.pdb). (H) View in G is rotated to see the DNA path which is not very visible in this structure. (I) View in H is rotated to compare the view with TEFM bound EC in fig F. Note that TEFM opens up the POLRMT. Compare F and I.

### 3.4.2.2 Mitochondrial transcription elongation

To investigate the structure of POLRMT in its elongation phase of transcription, a DNA-RNA hybrid with 9nt RNA mimicking the oligonucleotides in an RNA polymerase elongation phase (elongation scaffold, ES) was used (Schwinghammer et al., 2013). In this study, the scaffold was incubated with  $\Delta 150$ POLRMT. A major change in the elongation state as compared to the structure of the presumed initiation state of POLRMT is the path of the DNA. The upstream dsDNA now passes underneath the palm subdomain, a path that was followed by the dsDNA in the initiation complex and the downstream dsDNA is leaving at  $180^\circ$  between thumb and lever loop of the POLRMT, which gives thumb domain a chance to increase the processivity of the POLRMT. The template strand is held by the POLRMT while the NT strand is kept away from the TSS because of a different path of the DNA in the elongation complex. Interestingly, the intercalating hairpin now unzips the two strands in the downstream region. It is not understood how the upstream region of dsDNA relative to TSS is separated as that should be the first step for any RNAP. The emerging RNA (5'-end) appears to go towards the PPR domain, a path also confirmed by mapping (Schwinghammer

et al., 2013). Since only 9nt RNA was used in the structure, the exact path of RNA cannot be traced. The structure reveals striking difference from the T7 RNAP, which undergoes a conformational change in its promoter binding NTD while human POLRMT rearranges the DNA path.

From the available structure of initiation and elongation complex, it can be speculated what happens during the transition from initiation to elongation phase: As the POLRMT translocate along the dsDNA, the distance between the POLRMT and TFAM will increase causing the tether helix (123-245) of POLRMT to detach from TFAM causing TFAM release. In a later step, the upstream dsDNA will be repositioned and this movement will possibly release the TFB2m factor that is no longer required. Emerging RNA will be separated from the downstream dsDNA by the intercalating hairpin. Still, some questions that remain to be answered are exactly what trigger the release of TFB2m? How is the path of DNA changed? Where would the NTE of POLRMT bind and how is the upstream dsDNA separated?

A recent structure of the elongation complex with TEFM shows interactions between POLRMT, TEFM and ES (Hillen et al., 2017b). During transcriptional elongation with bound TEFM, the homodimers CTDs of TEFM sits on the top of the active site where it appears to stabilize the upstream and downstream dsDNA and at the same time stabilizes the unwound NT strand at the transcription active site. This stabilization of NT strand is not possible in the absence of TEFM. (Schwinghammer et al., 2013). Downstream dsDNA is held like a clamp between TEFM and POLRMT, a structure that is observed in other transcription of bacterial and eukaryotic transcription systems (Bernecky et al., 2017; Liu and Steitz, 2017; Martinez-Rucobo et al., 2011). On top of making interactions with nucleic acid, TEFM also stabilizes the intercalating hairpin and crosslinking studies suggest that it may interact with the specificity loop at the RNA exit path (Hillen et al., 2017b). This interaction can narrow the RNA exit path. In order to obtain the crystal structure many flexible parts were removed as for example the NTD and the linker of TEFM. Especially the linker was found to be indispensable for TEFM activity (Hillen et al., 2017b). Moreover, the NTE of POLRMT, that is shown to interact with TEFM (from crosslinking experiments) was removed in order to obtain crystals. Furthermore, the RNA exit path cannot be tracked as RNA with 5 nt tail was used. It will be interesting to know if RNA is going towards PPR domain and how this path is structurally controlled.

#### 3.4.2.3 Mitochondrial transcription Termination

During transcription termination, POLRMT should stop RNA synthesis and leave the mtDNA. Termination of transcription in mammalian mitochondria is not well understood. One factor, MTERF1, can act on both the L-strand and H-strand. The structure of MTERF1 suggests that it binds the downstream tRNA<sup>Leu</sup> region specifically, partly unwind and bend it causing a roadblock for transcriptional machinery (Jiménez-Menéndez et al., 2010; Yakubovskaya et al., 2010). How this bending stops an active POLMRT is unknown. The factors involved in termination from HSP are unknown (Hillen et al., 2018)

### 3.4.3 Other proteins involved in mtDNA transcription

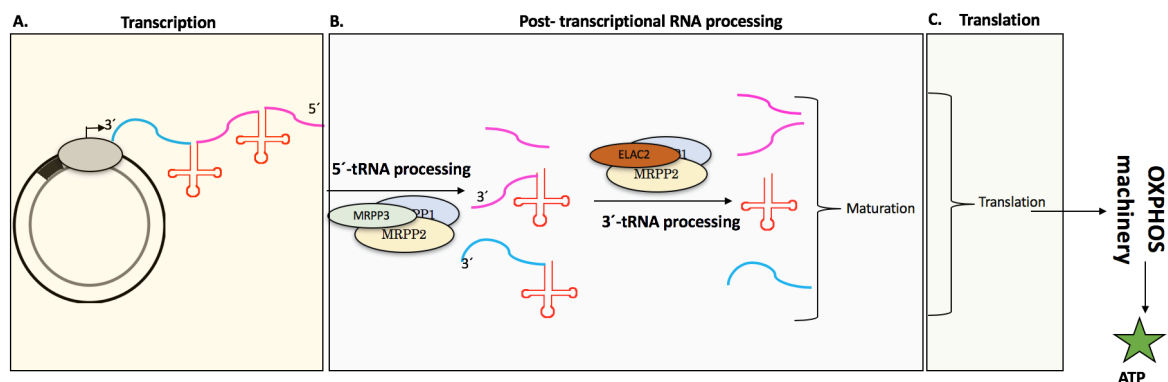
As discussed above, mtDNA transcription is not entirely understood and new factors are constantly being discovered. One such factor is mitochondrial ribosomal protein L12, **MRPL12** that holds on to POLRMT both *in-vivo* and *in-vitro*. One study finds that it stimulates the transcription from both strands (Litonin et al., 2010; Surovtseva et al., 2011; Wang et al., 2007) while another contradicts it and found no effect of MRPL12 on transcription *in-vitro*, even in the presence of mtDNA extracts (Litonin et al., 2010).

Leucine-rich pentatricopeptide containing protein, **LRPPRC** was also linked to transcription both *in-vivo* and *in-vitro* experiments, where it was reported to stimulate mtDNA transcription and interact directly with POLRMT (Liu et al., 2011). Knock-out of LRPPRC from mouse heart did not affect the transcriptional level, on the other hand, the mRNA transcripts were less stable with decreased polyadenylation and dysfunctions in translation were observed (Ruzzenente et al., 2012). Mutations in LRPPRC causes cytochrome c oxidase deficiency, which is held responsible for the neurodegenerative disorder, French Canadian variant of Leigh syndrome (Merante et al., 1993).

## 4 POST-TRANSCRIPTIONAL RNA PROCESSING

The transcription of the L-strand and H-strand results in long polycistronic near-genomic length transcripts, which are processed to yield individual RNA molecules. It is suggested that RNA processing occurs co-transcriptionally (Antonicka et al., 2013; Jourdain et al., 2013; Lee et al., 2013). Processing of mitochondrial transcripts involves initial cleavage to yield individual RNAs, further maturation of each RNA molecule i.e. polyadenylation of mRNA, CCA additions to tRNA tails, other modifications in tRNA and nucleotide modification in rRNAs (Shokolenko and Alexeyev, 2017) (Figure 12).

A tRNA punctuation model was proposed by (Ojala et al., 1981), suggesting that since most RNAs in the polycistronic transcript are flanked by tRNA, the processing starts at the tRNAs (Figure 12). The 5'-processing is carried out by the mtRNase P enzymes, which comprise of three subunits of MRPP1-3. In plants, the protein-only RNase P1 (PRORP1) can carry this process independently. On the other hand, the recent structure of MRPP3 suggests that the active site of MRPP3 is distorted that explains the dependency of this protein on the other two factors MRPP1 and MRPP2 to carry out the cleavage at 5'-end of tRNA, supposedly by inducing a conformational change in MRPP3 (Reinhard et al., 2015). A recent study suggests that MRPP1-2 is not only required for MRPP3 activation to carry out 5'-processing of tRNA but it serves as a platform for tRNA during 3'-processing by ELAC2 and even further CCA addition steps (Reinhard et al. 2017). Silencing of ELAC2 causes accumulation of unprocessed transcripts *in vivo* (Brzezniak et al., 2011; Sanchez et al., 2011). After the 3'-end processing of tRNAs, the maturation of the tRNAs are continued with the addition of a Cytosine-Cytosine-Adenine (CCA) addition to the 3'-end. This reaction is performed by the tRNA nucleotidyl transferases (TRNT1) (Nagaike et al., 2001).



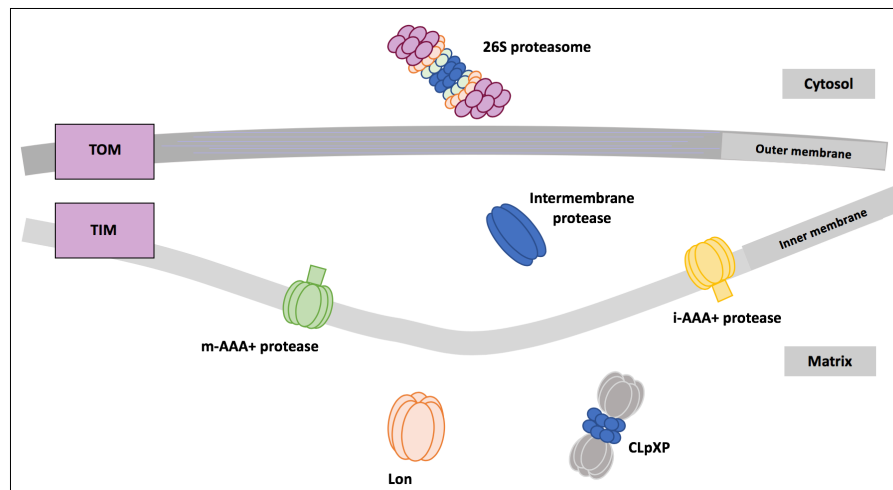
**Figure 12:** Post-transcriptional RNA processing. (A) mtDNA is transcribed giving rise to long polycistronic RNA. (B) According to tRNA punctuation model, tRNA will be processed first at 5'-end by RNase P composed of MRPP1, MRPP2, and the nuclease MRPP3. The 3'-ends will be processed by RNaseZ comprising of ELAC2 with the tRNAs still bound on MRPP1/2, resulting in individual RNAs still bound to MRPP1/2. (C) The RNAs will then be matured by different proteins. (D) Matured RNAs will function in the mitochondrial translation to generate membrane-bound components in the OXPHOS machinery, which is required for the generation of ATP.

## 5 MITOCHONDRIAL PROTEIN REGULATION

As described in the introduction, mtDNA encodes only 13 proteins. These 13 proteins constitute crucial parts of the ATP-production machinery, the remaining 1000-1500 proteins required to carry out the important mitochondrial functions are imported from the cytosol. Once imported, they fold in the mitochondrial matrix and are often organized into multimeric complexes to carry out their respective metabolic functions. All this emphasizes that mitochondrial proteostasis relies heavily on the import of proteins and maintaining their quality by a protein quality control (**PQC**) **mechanism**. Both mechanisms ensure that the mitochondrial proteome is adjusted to the cellular demands to maintain a healthy mitochondrial population as is crucial for cell physiology and survival (Baker Michael et al., 2013). The impairment of any of these mechanisms will affect the normal functioning of the mitochondria. The **mitochondrial PQC machinery** consists of an elaborate network of molecular **chaperones** and protein degradation factors, known as **proteases** that continually monitor and maintain the integrity of the proteome. **Chaperones** facilitate *de-novo* folding of the proteins but also refold the proteins affected by stress conditions (Baker and Haynes, 2011). Two of the chaperones Hsp60 (also known as HPD1) and Hsp70 (also known as HSPA and mortalin) have been shown to be involved in protein translocation and refolding reactions in mitochondria (Voos, 2013). Hsp78 is required at elevated temperatures and is necessary for the solubilization of aggregated proteins *in-vivo* (Neupert and Herrmann, 2007). **Proteases**, on the other side, degrade the unfolded and/or damaged proteins. Often, these protein-degrading machineries are ATP dependent and belong to the **AAA+** superfamily (*Atpase Associated with various cellular activities*) but there are some ATP-independent proteases also. The **AAA+** mitochondrial proteases forms a hollow chamber connected to the outside through narrow axial pores through which they translocate the substrate by using chemical energy from ATP hydrolysis (Gur et al., 2013).

### 5.1 PQC IN MAMMALIAN MITOCHONDRIA

In mammalian mitochondria, proteases are highly diverse and are present in each compartment of the organelle i.e., in the matrix, in the intermembrane space and in the inner membrane where their active site is oriented towards the matrix, known as **m-AAA+** proteases or facing the intermembrane space, known as **i-AAA** proteases (Quirós et al., 2015). Damaged membrane-associated subunits of electron transport chain, ETC are degraded by the **m-AAA+** and **i-AAA+** proteases (Hornig-Do et al., 2012; Stiburek et al., 2012; Zurita Rendón and Shoubbridge, 2012). The intermembrane space has several ATP-independent proteases that take part in the PQC mechanism. For example, HTRA2, a trimeric protein, is highly conserved from bacteria to mammals and depletion of it causes a drop in mitochondrial membrane potential, causing reduction in ATP levels. HTRA2 degrades oxidized proteins (Quirós et al., 2015)(Figure 13 and Table 2).



**Figure 13:** ATP-dependent proteolytic machines in metazoan cell. 26S proteasome works in cytosol. i-AAA+ proteases and m-AAA+ proteases are present in the inner membrane of the mitochondria. ClpXP and Lon are the major AAA+ proteases present in mitochondrial matrix.

The ATP dependent proteases in the mammalian mitochondrial matrix are ClpXP and Lon. The Serine protease, **ClpXP** (ClpX is the ATPase subunit and ClpP is the protease subunit) together with the matrix peptide exporter HAF-1, have been reported to be involved in the mitochondrial unfolded protein response (UPR<sup>mt</sup>) (Bezawork-Geleta et al., 2015). The mitochondrial UPR is a stress response that, in the advent of stress in mitochondria, causes upregulation of the transcription of nuclear-encoded chaperone genes to promote the protein homeostasis within mitochondria. It is well studied in *Caenorhabditis elegans* and *Drosophila melanogaster* but remains poorly understood in mammals (Bezawork-Geleta et al., 2015).

**Table 2:** Examples of proteases found in mitochondria.

Mitochondrial location	Symbol	Function	Type	Reference
Matrix	LonP1	PQC and mitochondrial biogenesis	ATP dependent	(Quirós et al., 2014b)
Matrix	Casein lytic protease XP (ClpXP)	PQC	ATP dependent	(Gispert et al., 2013)
Intermembrane space	HTRA2 (or OMI)	PQC and apoptosis	ATP independent	(Clausen et al., 2011)
Inner membrane facing interspace (i-AAA+)	YME1L1	PQC and mitochondrial biogenesis	ATP dependent	(Stiburek et al., 2012)
Inner membrane facing matrix (m-AAA+)	AFG3L2	PQC and mitochondrial biogenesis	ATP dependent	(Quirós et al., 2015)

On the other hand, it is clear that Lon is the major degrader of misfolded proteins in the matrix (Bezawork-Geleta et al., 2015) This study will focus on the Lon function as a protease.

## 5.2 THE HUMAN MITOCHONDRIAL LON, LONP1 (LONM)

In the human mitochondrial matrix, LonP1 is the main AAA+ protease. The highly conserved LonP1 selectively degrades proteins that have been damaged due to destabilizing mutations or oxidative modifications (Pinti et al., 2016a; Quirós et al., 2015). The proteins associated with Human LonP1 are summarized in Table 3. Recently data suggested that LonP1 is solely responsible for degradation of misfolded protein in human mitochondria and while it doesn't act on aggregated proteins, it prevents the formation of toxic aggregates (Venkatesh et al., 2012). It is also known to act on certain regulatory proteins (Pinti et al., 2015). *In-vitro* experiments with purified TFAM and human LonP1 has shown that a phosphorylated TFAM, which is no longer able to bind DNA, is degraded by human LonP1 (Lu et al., 2013). The same observation was reported in *Drosophila* cells (Matsushima et al., 2010). Therefore, human LonP1 can regulate mtDNA replication and transcription. Recently, a study showed that upon stress in mitochondria, a mitochondrial UPR<sup>mt</sup> response is induced which causes simultaneous transcriptional repression and increase in Lon-mediated degradation of MRPP3, a pre-tRNAs processing enzyme (Münch and Harper, 2016). This implies that human LonP1 also regulates mitochondrial gene expression.

**Table 3:** Proteins shown to be associated with LonP1. Adapted from (Pinti et al., 2015)

Proteins	Comment	References
<b>TCA cycle</b>		
Aconitase	Degraded by Lon	(Bota and Davies, 2002)
Glutaminase C	Degraded by Lon	(Kita et al., 2012)
<b>Oxidative phosphorylation</b>		
COX4-1	Degraded by Lon	(Fukuda et al., 2007)
SDH5	Degraded by Lon	(Bezawork-Geleta et al., 2014)
<b>Maintenance of mtDNA</b>		
mtDNA POL $\gamma$	Physically interact with nucleoids	(Bogenhagen et al., 2008)
TWINKLE	Physically interact with nucleoids	(Bogenhagen et al., 2008)
TFAM	Degraded by Lon when phosphorylated	(Lu et al., 2013)
MRPP3	Degraded by Lon	(Münch and Harper, 2016)
<b>Other proteases</b>		
StaR	Degraded by Lon	(Granot et al., 2007)
ALAS-1	Degraded by Lon	(Tian et al., 2011)
Cystathionine $\beta$ -synthase	Degraded by Lon	(Teng et al., 2013)



The majority of Lon is found to be localized in matrix while a small part is also found to be associated with inner membrane where nucleoids are present (Pinti et al., 2015). In nucleoids, human LonP1 is reported to interact with mtDNA by recognizing ssDNA in D-loop and *in-vivo* is found to bind the G-rich consensus sequence in the LSP and to the RNA transcribed from that region (Pinti et al., 2015), suggesting human LonP1 is involved in maintenance of DNA integrity (He et al., 2018).

Given the importance of human LonP1, dysfunction of Lon is linked to mitochondrial disorders causing loss of mitochondrial function, reduced cell proliferation, early embryonic lethality and apoptosis (Bota et al., 2005; Cheng et al., 2013; Gibellini et al., 2014; Quirós et al., 2014a) while overexpression promotes proliferation of cancer cells by remodeling their mitochondrial function. Imbalanced levels of Lon are linked to neurological, cardiac and hepatic disorders and aging (Pinti et al., 2015; Pinti et al., 2016a; Van Dyck and Langer, 1999). Despite its huge importance and clinical role, little is known about the structure-function relationship of human LonP1 as compared to its bacterial homologues.

### 5.2.1 Evolution and location

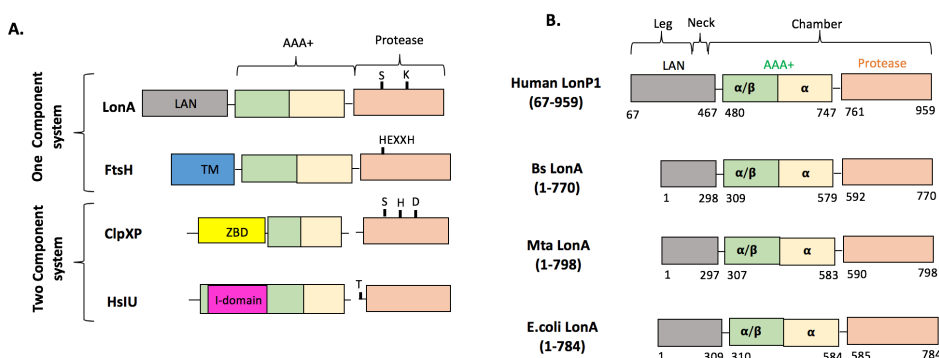
Lon is highly conserved throughout evolution and its homologues have been found in all kingdoms of life. Its high degree of conservation suggests that its presence is crucial for cell survival. Lon was first identified in *E. coli* as an ATP-dependent protease with a Ser-Lys catalytic dyad and was named La because upon UV irradiation, the *E. coli* mutants that were deficit in the Lon gene, appeared to be longer than the ones that had this gene (Charette et al., 1981; Chung and Goldberg, 1981; Howard-Flanders et al., 1964).

In humans, both mitochondrial and peroxisomal forms of Lon are encoded by the nuclear genome and the respective *LonP1* and *LonP2* genes are present on chromosome 19 and 16. Peroxisomal specific, LonP2 is not well explored, however it has shown to degrade proteins damaged by hydrogen peroxide (Pinti et al., 2016b).

The mitochondrial Lon, LonP1 is reported to have three isoforms in humans that are based on alternative splicing. The canonical isoform 1 is the longest (1-959) while isoform 2 and 3 have the 42-105 and the 1-196 residues missing respectively. The physiological function of the isoforms 2 and 3 is unknown. In the canonical isoform 1, the initial 67 residues are reported to be part of MTS (<https://www.uniprot.org/uniprot/P36776>). While all the studies (biochemical, kinetic, structural) up till now have used initial 114 amino acids as the MTS sequence (Fishovitz et al., 2017; He et al., 2018; Lee et al., 2018). Based on the recent structural reports of 114-959 construct, that the initial N-terminal residues are involved in oligomerization (Kereïche et al., 2016), it can be presumed that the residues 62-114 are of significant importance and there is a need to further explore this region (discussed in Paper III).

## 5.2.2 Domain organization

Human LonP1 has 30-40% sequence homology to bacterial Lon and together they belong to the LonA subclass. In this subclass, each protomer has a species-specific N-terminal domain, known as LAN, followed by an ATPase domain made up of an  $\alpha/\beta$ - and  $\alpha$ -subdomains and finally a C-terminal protease domain. This is similar to FtsH and other m-AAA+ proteases where a single polypeptide contains all the domains (Baranek et al., 2011; Lee et al., 2011) and is in contrast to other AAA+ proteases such as ClpXP (ClpX is the AAA+ ring and the ClpP is the protease) and the 26S proteasome where different functionalities reside on different polypeptides (Figure 14) (Baker and Sauer, 2012; Bard et al., 2018).



**Figure 14:** Cartoon representation of the various AAA+ proteases. There are two main types: One Component system that comprises of Lon (matrix) and FtsH (membrane) that have their ATPase on the same polypeptide where LonA has a species-specific Lan domain and FtsH has a transmembrane domain. The Two component system that comprises of ClpX and HslIU, have their ATPase and the peptidase activity on separate polypeptides. ClpX has a Zinc binding domain and HslIU has an accessory domain inserted in its AAA+ domain. The active proteolytic residues are marked.

LonP1, like bacterial Lons, form homohexamers where the LAN domains recognize and bind the substrate which upon ATP hydrolysis by the AAA+ module is threaded into the chamber composed of the AAA+ module and protease domain. Upon arrival to the protease domain, the substrate is hydrolyzed and finally expelled out of the chamber. The chamber is the main site of action where both protein substrates and ATP are hydrolyzed (Vieux et al., 2013). Several crystal structures of the chamber or the individual domains from the bacterial (LonA, ATP dependent, soluble), archaeal (LonB, ATP dependent, transmembrane) and ATP-independent bacterial homologs (LonC, soluble) are reported in order to understand the structural basis of Lon's function and are summarized in Table 4.

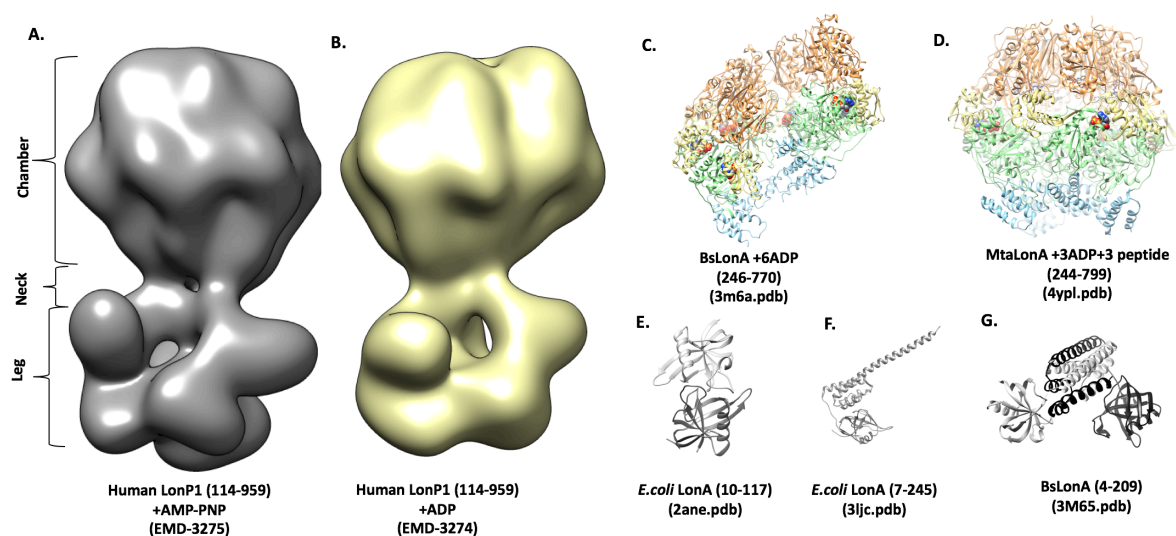
**Table 4:** List of published structures of Lon available in RCSB.

	Species	Construct	Comment	PDB ID	Reference
LonA	<i>E. coli</i>	Proteolytic domain (584–784 with S679A)	Hexameric, Open to surface, Ser-Lys dyad identified	1RR9	(Botos et al., 2004b)
	<i>E. coli</i>	Alpha domain (491–584)	Monomer, S2 identified	1QZM	(Botos et al., 2004a)
	<i>E. coli</i>	LAN domain (1–119)	Dimer with a unique fold	2ANE	(Li et al., 2005)
	<i>H. sapiens</i>	Proteolytic domain (756–959)	Hexameric, Ser-Lys dyad in different conformation than <i>E. coli</i> (1rr9.pdb)	2X36	(García-Nafria et al., 2010)
	<i>E. coli</i>	LAN domain (7–245)	monomer	3LJC	(Li et al., 2010)
	<i>Bacillus subtilis</i>	LAN domain (4–209)	monomer	3M65	(Duman and Löwe, 2010)
	<i>Bacillus subtilis</i>	AAA+ domain and protease domain (246–770) +6ADP	Hexamer, First Chamber structure	3M6A	(Duman and Löwe, 2010)
	<i>Brevibacillus thermoruber</i>	Alpha subdomain (492–585)	Monomer, suggests DNA binding residues	4GIT	(Lee et al., 2014)
	<i>Meiothermus taiwanensis</i> (Mta)	Chamber (244–799) +3ADP	Hexamer, three protomer model suggested	4YPL	(Lin et al., 2016)
	<i>Meiothermus taiwanensis</i> (Mta)	Chamber (244–799) +bortezomib	Hexamer, three protomer model suggested	4YPM	(Lin et al., 2016)
	<i>Meiothermus taiwanensis</i> (Mta)	243–490	monomer	4YPN	(Lin et al., 2016)
	<i>Meiothermus taiwanensis</i> (Mta)	Alpha and Proteolytic domain, 490–780	Hexamer in the presence of Mg 2+ ions	5E7S	(Lin et al., 2016)
	<i>Meiothermus taiwanensis</i> (Mta)	Alpha and Proteolytic domain, 490–780	Hexamer in the presence of Mg 2+ ions	5E7S	(Lin et al., 2016)
	<i>H. sapiens</i>	114–959 with S855A mutation+ AMPNP	CryoEM, 15 Å, first 3D representation, Closed planar arrangement in chamber	EMD3275	(Kereïche et al., 2016)
	<i>H. sapiens</i>	114–959 with S855A mutation+ ADP	CryoEM, 20 Å, first 3D representation, Open helical arrangement in chamber	EMD3274	(Kereïche et al., 2016)
LonB	<i>M. jannaschii</i>	Proteolytic domain (456–640)	Dimer, Ser-Lys-Asp, Identified as catalytic triad	1XHK	(Im et al., 2004)
	<i>Archaeoglobus fulgidus</i>	Proteolytic domain (416–620)	Hexamer, disordered active site	1ZOE	(Botos et al., 2005)
	<i>Thermococcus onnurineus</i>	18–635 + ADP	Hexamer with loose and tight binding protomers	3K1J	(Cha et al., 2010)
LonC	<i>Meiothermus taiwanensis</i>	1–108+ 218–710	Hexamer, Atpase independent Lon, LAN disordered	4FW9	(Liao et al., 2013)
	<i>Meiothermus taiwanensis</i>	35–385	Hexamer, LAN ordered	4FWV	(Li et al., 2013)

### 5.2.3 Mechanism of action

Interesting is the first three-dimensional low-resolution model for Human LonP which hints on the unique 3D organization of this protein (Kereïche et al., 2014). The hexameric Lon can be divided into three parts; head, neck and leg region (Figure 15). The Head or the chamber is composed of the protease and the Atpase domain of six protomers while the LAN domains

arrange in trimers of dimers to form three legs that are connected to the head by a neck region. The structure agrees with the previous observations that the LAN domains recognize and bind the substrate which is translocated into the chamber upon ATP hydrolysis and degraded by the protease domain (Kereïche et al., 2014).



**Figure 15:** Available Structures of LonA. (A) CryoEM structure of human LonP1 (114-959) bound to AMP-PNP (EMD-3275) (B) CryoEM structure of human LonP1 (114-959) bound to ADP (EMD-3274) (C) Crystal structure of chamber of Bs LonA (246-770) bound to ADP. The hexamers are arranged in helical conformation (3m6a.pdb). (D) Crystal structure of chamber of Mta LonA (244-799) bound to ADP and inhibitor. The hexamers are arranged in planar ring conformation (4ypl.pdb) (E) Crystal structure of LAN domain of *E. coli* LonA (10-117) (2ane.pdb). (F) Crystal structure of LAN domain of *E. coli* LonA (7-245) (3ljc.pdb). (G) Crystal structure of LAN domain of Bs LonA (4-209) (3m65.pdb).

To better understand each step of substrate degradation by Lon, the mechanism can be divided into the following four stages: 1) Substrate recognition and binding 2) substrate translocation 3) Substrate degradation 4) Substrate ejection. Little is known about the substrate ejection. However, the first three phases of substrate degradation by Lon will be briefly discussed below:

#### 5.2.3.1 Substrate recognition

Substrate recognition is an important step and human LonP1 should be able to discriminate between its targets and non-targets. The precise mechanism is not well understood for human LonP1. It has shown to act on DNA-free TFAM but the recognition sequence has not been identified (Lu et al., 2013).

In the eukaryotic 26S proteasome, the protein substrate to be degraded is recognized by a ubiquitination tag while in others, a small protein-sequence stretch on the substrate known as degron, is used to recognize and bind the substrate. In *E. coli*, the Clp and FtsH often recognize an 11 amino-acid stretch called the SsrA tag (AANDENYALAA). The tag, when added to other proteins, can be recognized by Clp *in-vivo* and the proteins are degraded (Smith et al., 1999). Other than this, five recognition motifs were recognized by broad scanning using a mutant variant of ClpP, three were located near the C-terminus and two were located on the N-terminus (reviewed in (Gur et al., 2013)).

For Bacterial Lons, numerous studies report a variety of target regions or degrons but there is not any coherence. A misfolded  $\beta$ -galactosidase substrate is recognized by a region between 28-49 on the N-terminus called  $\beta$ -20 (Gur and Sauer, 2008b). UmuD, a DNA repair protein involved in SOS response, is identified by the 15-29 residue region in the N-terminus (Gonzalez et al., 1998). SoxS, a transcription activator, is also identified by a 21 residues on the N-terminus (Shah and Wolf, 2006). Similarly, a regulatory protein, SulA, is identified by a degron called Sul20 (Wohlever et al., 2014). In yet another case, attaching a SsrA tag, which is specific for ClpP, makes the protein targeted by Lon (Gur and Sauer, 2008a). Yet another type of degradation by Lon requires adaptors. For example, in *Yersinia pestis*, a small heat shock protein Q, HspQ is believed to interact with Lon with its C-terminus that enhances the activity of Lon towards a range of substrates including YmoA, RsuA and Y039 (He et al., 2018). DNA has been shown to be an adapter for some DNA binding proteins like TrfA and RepE which can only be degraded when Lon is complexed with DNA (Karlłowicz et al., 2017). On the other hand, the search for a degron on other regulatory proteins like RcsA (Chung and Goldberg, 1981; Gottesman and Zipser, 1978) and DnaA were unsuccessful (Liu et al., 2018). In the latter case, it was proposed that Lon interacts with DnaA on two sites, one used to hold on Lon while other is where the degradation is initiated.

Another study suggested that Lon recognizes specific surface determinants or folds (Ondrovicová et al., 2005). Another study backed up this finding and suggested that Lon recognizes its substrate based on hydrophobicity, which are normally hidden when folded and exposed when unfolded. This can be a good explanation for the huge variety seen in Lon substrates. This study also suggested, based on kinetic experiments, that the Lon hexamer contains more than one recognition site for protein substrates (Gur and Sauer, 2008b).

Yet recently it was shown that human LonP1 failed to identify the substrates of bacterial Lon even though human LonP1 has 30-40 % sequence similarity to bacterial Lon, indicating a species-adapted substrate specificity (He et al., 2018). More work is required to understand the substrate recognition requirements for human LonP1.

#### 5.2.3.2 Substrate translocation

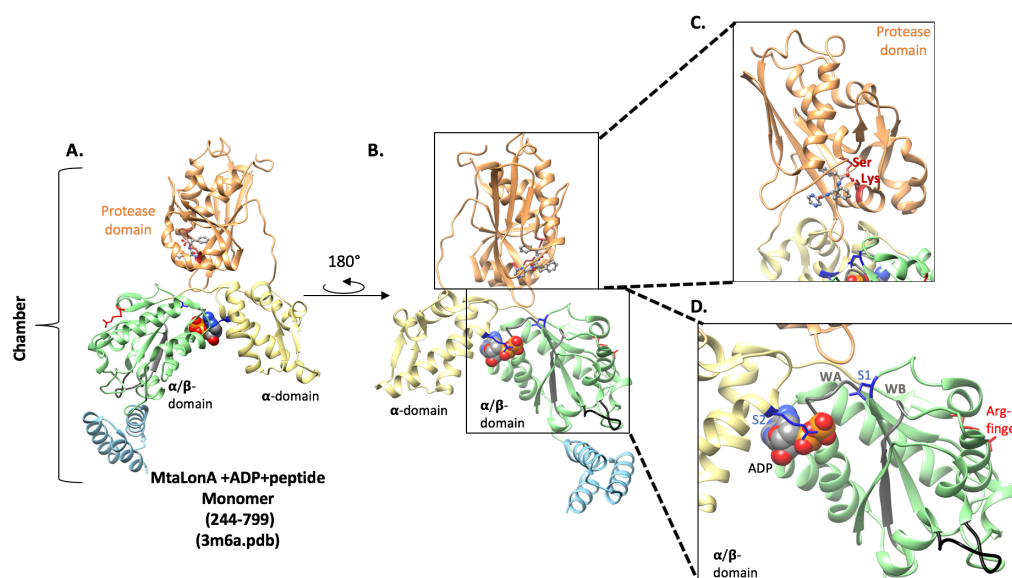
Once the substrate is recognized and bound by the LAN domains, its flexible region is brought close to the entrance of chamber from where the substrate is threaded inside by using energy from ATP hydrolysis by the ATPase domain. In addition to ATPase hydrolysis, the ATPase domain has flexible loops, termed as axial pore loops that have shown to directly interact with these substrates and thread it inside the chamber. All the protomers in the hexamers communicate in order to successfully translocate the substrate inside the chamber (Lin et al., 2016; Olivares et al., 2018; Sauer and Baker, 2011). The mechanism of ATP binding, ATP hydrolysis, intra- and inter-monomer communication along with the concerted action of loops are briefly discussed below.

Like the degron, which terminus is threaded inside the chamber of human LonP1, is unknown. Other proteases like FtsH and Clp bind the N-terminus of SsrA tagged lambda protein and degrades on C-terminus. It was found that Clp can efficiently translocate a polypeptide from the N-terminus to the C-terminus as it can in the opposite direction (Gur et al., 2012). The 26S proteasome binds the N-terminal ubiquitinated lysine's of the substrate and initiates degradation at the C-terminus (Bard et al., 2018) reviewed in (Gur et al., 2013). Still the proteases like VAT, that work in conjunction with the 20S proteasome, acts on unstructured N- or C-termini (Ripstein et al., 2017).

Once the flexible terminus of the substrate is placed at the entrance of the chamber, it should unfold and translocate inside, the mechanical energy required for this job is provided by the ATP hydrolysis in the ATPase domain. The ATPase domain has two sub domains,  $\alpha/\beta$  and  $\alpha$ -subdomains which are connected by an Aromatic-loop (A-loop) (Figure 16). It is at the interface of the two subdomains underneath the A-loop where the nucleotide-binding pocket is found. The structural motifs that are involved in the nucleotide binding, nucleotide hydrolysis and coupling ATP hydrolysis to conformational changes to carryout mechanical work, are well-conserved in the AAA+ machinery and well-studied in literature (Wendler et al., 2012).

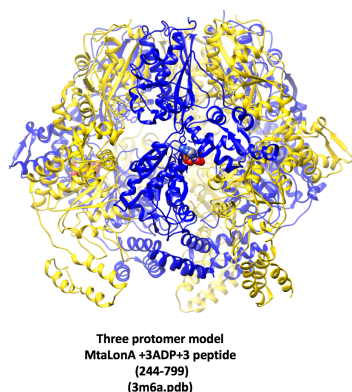
For nucleotide binding and hydrolysis, the known motifs are Walker A (WA, also known as P-loop), Walker B (WB), Sensor 1 (S1), Sensor 2 (S2), Arginine finger (Arg-finger) and Aromatic-loop (A-loop). The **WA** motif (-GPPGVK-) stabilizes the  $\beta$ - and  $\gamma$ -phosphates of the bound nucleotide. The **WB** (-LIDEVDK-) is involved in nucleotide hydrolysis and coordinate  $Mg^{2+}$  ion that is required for catalysis. The **S1** motif, usually a conserved Thr or Asn, places a nucleophilic water close to the  $\gamma$ -phosphate of the ATP and facilitates nucleophilic attack. The **S2** motif is mostly an Arg that helps to discriminate between ATP and ADP by hydrogen bonding to the terminal phosphate of ATP and is linked to conformational changes between the ATPase subdomains. The inter-subunit communication is carried out by the **Arg-finger** which is provided by the neighboring ATPase subunit. This Arg-finger protrudes out to reach the nucleotide binding pocket of the neighboring protomer, where it can coordinate with the  $\gamma$ -phosphate of ATP and can cause the ATP hydrolysis that is the basis of the power stroke. Moreover, the Arg-finger is in an ideal location as it is directly connected to the nucleotide-binding pocket and the flexible loops of the harboring protomer; therefore, any conformational change occurring in the substrate-binding regions can be transduced through the Arg-fingers to the rest of the protomers in the chamber (Wendler et al., 2012; Zhao et al., 2016).





**Figure 16:** Crystal structure of monomer of Mta LonA (244-799) bound to ADP and inhibitor, Bortezomib (4ypl.pdb, Chain B) (A) Monomer in one view and rotated in (B) to see the structural motifs involved in the AAA+ and protease domain. (C) Close up of protease domain, the active Ser and Lys residue are marked in red. An inhibitor can also be seen in the active site. (D) Close up view of the ADP bound active site in AAA+ domain. Important structural motifs are highlighted. Conserved pore loop is shown in black. See text for details.

The **conserved pore loops** protrude from the ATPase domains towards the central cavity of the chamber and are reported to physically interact with the incoming substrate. To investigate whether all loops of oligomers operate in concerted fashion to unfold and translocate protein or only one binds at a time, an elegant method was described by Sauer and his colleagues where they study the loops of Clp by generating a single chain hexamer of Clp (Iosefson et al., 2015). Six copies of the ClpX gene were cloned into a single gene and this strategy allowed to control the insertion of desired mutations within one protomer out of a hexamer. They found that a single subunit can promote efficient translocation of a substrate and is not dependent on the sequential activity of other protomers of hexamer (Iosefson et al., 2015). This suggests that the protomer follows a **hand-over-hand model** to translocate the substrate. This is supported by latest structural reports of many proteases: 26S proteasome (Bard et al., 2018), Vps4-vta1 (Monroe et al., 2017), Hsp104 aggregase (Gates et al., 2017), YME1 (Puchades et al., 2017), VAT protein (Ripstein et al., 2017). However, a crystallographic study of bacterial Lon, MtaLonA bound to three ADPs and an inhibitor at proteolytic active site suggested a **three protomer model**, where three non-neighboring protomers hydrolyse ATP simultaneously and cause a concerted movement of the loops (Lin et al., 2016). The sequential movements of loops in two sets thus allow the unfolding and translocations of the substrate inside the chamber. This is in contradiction to the canonical hand-over-hand model reported by many different methods for AAA+ machinery. In the single particle analysis of full-length human LonP1 included as a manuscript in this thesis, a hand-over-hand model where three protomers provide a hand for substrate translocation is suggested (Paper III).



**Figure 17:** The three protomer model based on the crystal structure of hexamer of Mta LonA (244-799) bound to ADP and inhibitor, Bortezomib (4ypl.pdb). Three protomers bound to ADP are shown in blue and ADP-free are shown in yellow.

Even though the structural motifs for nucleotide binding and hydrolysis are well conserved in human LonP1 and similar to bacterial Lon in structure, reports suggest that human LonP1 can have a different kinetic mechanism compared to bacterial Lons (Fishovitz et al., 2017).

#### 5.2.3.3 Substrate proteolysis

When the terminus of the substrate reaches the protease domain, it is degraded through the action of the Ser-Lys dyad in the proteolytic site. Biochemical, mutational and structural data have suggested that a Ser-Lys dyad is sufficient for catalytic activity in Lon (Rotanova et al., 2006). This is different from other serine protease where a Ser-Lys-Asp triad is needed (Botos et al., 2004b). The exact proteolytic location is unknown; however, one mechanistic study has demonstrated that Lon generally cleaves between hydrophobic residues present between highly charged patches of the polypeptide chain (Ondrovicová et al., 2005). The structure of the catalytic domain of *E. coli* suggests that polypeptides are oriented to place their si-face of the peptide bond in front of the serine hydroxyl group to allow nucleophilic attack (Botos et al., 2004b).

Regarding the communication between the ATPase and the protease domain, there is one crystal structure of the bacterial Lon, MtaLonA bound to an inhibitor at the proteolytic site; the structure shows that a  $\beta$ -hairpin structure that emerges from the protease domain, reaches the nucleotide binding pocket to communicate with the ATPase hydrolysis directly. It also have a comparatively hydrophobic tip that can guide the substrate to the proteolytic active site (Lin et al., 2016). However, the exact mechanism of intra- and inter-protomer communication remains to be explored.

Since the protease subunits of two-component system proteases do not hydrolyze ATP, it can be presumed that ATP hydrolysis is used only for translocation of the substrate to the proteolytic site and that the further cleavage of the peptide bond doesn't require ATP (Ripstein et al., 2017).



### 5.3 26S PROTEASOME

The 26S proteasome is the most elaborate proteasome found in eukaryotic nucleus and in the cytoplasm. It's also a AAA+ machinery where 20S forms the core particles forming a sequestered chamber in the form of barrel shape while the two 19S regulatory components bind to one or both ends of the chamber (Bard et al., 2018). 19S particles are responsible for recruiting the ubiquitinated protein, deubiquitinate them followed by unfolding and translocation through a central pore into the proteolytic chamber for degradation. The regulatory part has two components; a lid with 9 subcomponents where the one has deubiquitinating activity (Rpn11) and the base made upon 10 subunits out of which six are AAA+ ATPase (Rpt1-6). This ATPase forms a hetero-hexameric ring where each subunit has highly conserved nucleotide binding pocket and pore loops similar to Lon. The mechanism of action is believed to be that the after ubiquitinated substrate is recruited, the ATPase hydrolyses ATP and threads the flexible initiation region of the substrate into the chamber which is translocated by interaction with the pore loops to the proteolytic site inside the chamber. The proposed mechanism of translocation is hand over hand model that is supported by many cryoEM and substrate bound hexameric AAA+-motors. (Bard et al., 2018).

## 6 METHODOLOGY

### 6.1 RECOMBINANT PROTEIN PRODUCTION IN *E. COLI*

In order to carry on structural studies a large amount of pure and homogenous protein is required. However, this is normally not possible through harvesting proteins from natural resources. Therefore, the proteins are recombinantly expressed in other systems, like bacterial, yeast and animal cells. The protocols for recombinant expression in *E. coli* strains are well established (Graslund et al., 2008). It is convenient, quick and cost-effective most of the time as compared to other expression systems. Some of the major concerns in *E. coli* production system is protein inactivity that may arise as the protein may adopt a stable but erroneous fold not suitable for activity, or it may require additional molecules for protein folding that are not available in this system or erroneous di-sulphide bond formation or if the expressed protein is toxic to the cell. Most of the time this can be fine-tuned. Still in cases where to yield active proteins post translational modifications are required which are not possible in the simple bacterial system, other expression systems are preferred (reviewed in (Rosano and Ceccarelli, 2014).

The process requires cloning the required gene of protein into an expression vector using standard molecular biology techniques. The recombinant vector is then typically transformed into special strain of *E. coli* that also have a controlled expression of the T7 RNAP (involved in transcription). In normal conditions, a lac repressor remains bound to the lac promoter of T7-RNAP gene and is released only when the cells are induced by isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). This release causes the transcription and translation of T7-RNAP which then causes the transcription of the gene of interest located on the transformed plasmid. This is followed by the translation of the protein of interest (Studier et al., 1990) (Figure 18).

### 6.2 PURIFICATION OF PROTEINS

Once the protein is produced in the *E. coli* cells, it can be harvested by lysing the cells and purifying the protein by different protein purification methods. Specialized tags like (His)<sub>6</sub>, Maltose binding protein, MBP; Chitin binding domain, CBD and Strep-tag can be added to either end of the protein to facilitate both the purification and production procedure (Kimple et al., 2013). Other properties like DNA binding and pI of the protein can be exploited to further purify the protein. The quality of the protein is constantly checked by running the proteins on a denaturing gel (SDS-PAGE gel). External conditions like buffer, temperature, salt, pH also play a role in maintaining the protein fold intact and thus preventing it from aggregation.

In the current study, the following method was used as a standard: The proteins were expressed using a cleavable (His)<sub>6</sub> tag on the N-terminus in the *E. coli* system using standard expressing condition. Cells were lysed by homogenization, centrifuged at high speed to get

rid of cell debris. His-tagged protein is separated from all the other proteins present in the soluble fraction by using Immobilized metal affinity chromatography (IMAC). IMAC is based on the interaction between primarily histidine and a resin-bound metal (Jerker et al., 1975). In this case, the His<sub>6</sub> tag of the protein interacts with the immobilized Ni<sub>2+</sub> ion bound to the stationary phase in the column. This step will allow the binding of His-tagged protein to the Ni-loaded column, while most other proteins will be washed away. The bound protein is released either by changing pH or by adding a competitive molecule like Imidazole. The tag of the protein can be cleaved now and can be subjected to additional rounds of purification. The current work involved DNA binding proteins, therefore DNA and nucleotides were the common contaminants and their presence can be assessed by absorbance measurement at 260 and 280 nm. The value of A<sub>260/280</sub> is a quick indicative of nucleic acid presence (Schowen, 1993). Values higher than 1 usually indicate that DNA or RNA is bound, while values between 0.5 and 0.7 suggests that the proteins are free of DNA. Purification on heparin or ion exchange columns help in getting rid of bound oligonucleotides. Usually a final step of purification is size exclusion chromatography, SEC (or gel filtration column, GF) which separates the proteins based on size and shape. Here, the proteins are allowed to pass through a porous material where proteins bigger than the bead size will come out first while the proteins smaller than the bead size will be entering resin beads and will take longer time to elute. The elution is further monitored by the UV meter to measure absorption at 280 nm. The obtained chromatogram is indicative of the oligomeric state and quality of the protein. SEC works as a robust technique to study not only the homogeneity of the protein, but protein-DNA interaction and protein-protein interactions. Other qualitative technique used throughout purification is running the protein on the denaturing sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). In SDS-PAGE the protein is denatured by adding reducing agent ( $\beta$ -mercaptoethanol or Dithiothreitol, DTT) that will break the disulfide bonds and SDS that will unfold the proteins and confer a negative charge to all the proteins, causing the proteins to migrate in a 1D gel system based on molecular weight. The gels are later stained with Coomassie blue stain, whose sulfonic acid groups will interact with the positive amine groups of protein causing the proteins to be stained in white background (Kimple et al., 2013; Nes, 1999).

### **6.3 BIOCHEMICAL CHARACTERIZATION OF COMPLEXES**

To investigate interactions between protein-protein and protein-DNA/RNA, the following methods were used: Usually a quick qualitative run is done to analyze the interaction between components followed by a preparative and quantitative method. In the current study, native gels, electrophoretic mobility shift assays (EMSA) or analytical gel filtration column were used for this purpose while the quantitative measurements were done using microscale thermophoresis technique (MST). Complex preparations for structural studies were prepared on preparative gel filtration columns.

Native gels (polyacrylamide or agarose) are run in non-reducing and non-denaturing environment, where the protein migration will depend on the pI of the protein complex, the

size of the protein complex and the pH used. Smaller and basic protein complexes will run faster as compared to larger and less charged protein complexes. The gels can be stained later for visualization. EMSA are based on the same principle as native gels but they are used for studying protein-DNA/RNA interactions. Free-DNA/RNAs will migrate faster as compared to the protein-bound DNAs, as a result a shift in the band between the free and protein-bound DNA will be apparent. These gels can be stained both for protein and nucleic acid for visualization. The gel system requires small amounts of sample, many conditions can be tested in one run, and can be used for doing stoichiometric calculations. However, the movement is based on the percentage of gels and running conditions like (temperature, buffer type, pH of buffer) and often needs further optimization. Another problem is to extract the formed complex from the gel (Arndt et al., 2012). A second approach, analytical GF often validates the gel-based measurements and are more indicative of the quality of the sample. It separates based on size as the complex is bigger in size than the individual protein. Although it is an efficient method, the measurement of thermodynamics, kinetics and affinity are provided by other quantitative biophysical methods, such as MST. MST is used to measure the interaction between two components where one of them is fluorescently labeled. The principle is based on the directed movement of fluorescently labeled molecules in a micro-temperature gradient, the movement of the labelled molecule will be different in bound and free state, as the complex molecule will have a different hydration shell, charge and size as compared to the non-bound particles. These changes can be measured and affinity can be calculated. The technique is different from other biophysical methods like, isothermal titration calorimetry, ITC which requires large amount of sample, and surface plasmon resonance, SPR which requires immobilization of one of the ligands. The detection limit is in the range of pM to mM range of dissociation constants. One drawback of MST is that the fluorescent labeling can result in nonspecific binding (Jerabek-Willemsen et al., 2014).

Generally, for structural studies, once the complex is established qualitatively, the next step is to produce it in enough amounts for structural studies and that can be done by performing a preparative purification on a gel-filtration column.

## **6.4 STRUCTURE DETERMINATION**

Three-dimensional structural knowledge of the proteins provides the basis for understanding the function of macromolecules. This study employed x-ray crystallography and cryoEM. Both will be briefly discussed below and all the used methods are summarized in Figure 18.

### **6.4.1 X-ray crystallography**

X-ray crystallography is the most widely used method to determine three-dimensional structure of proteins. In X-ray crystallography, crystals are exposed to intense and monochromatic x-rays. If the crystals have a highly ordered and repetitive network of molecules they will diffract the incident beam. The diffracted beams that follow Bragg's law  $n\lambda = 2d\sin\theta$  will constructively interfere giving rise to a diffraction pattern on the detector. This diffraction pattern has the intensity information *i.e.* squares of amplitudes but the phase

angles are lost. This problem of losing the phase information is known as phase problem, and strategies such as multiple-isomorphous replacement and anomalous dispersion can be employed to obtain the phase information. Once the phase problem is solved the initial electron density map can be calculated through Fourier transformation. If adequate resolution is obtained a macromolecular model can be built into the electron density. The model is thereafter refined to fit the data as accurate as possible. Parameters like the R-factor are used as an indicative tool for measurement of the quality of fit between model and data (Rhodes, 2006a).

**Protein crystallization** is the basic requirement for structure determination by x-ray crystallography. It is normally based on vapor-diffusion principle where the concentrated protein sample is mixed with a crystallization solution (a precipitant) in a  $\mu\text{l}$  drop volume and is kept close to a reservoir containing higher precipitant concentration in a closed system. This will allow the system to reach equilibrium. Slowly the water will vaporize from the drop, the protein concentration in the drop will gradually increase and move towards nucleation in ideal conditions. This is the condition where the attractive forces between the molecules are maximal and they start arranging themselves in arrays of crystals, which can grow to give rise to full crystals (Rhodes, 2006b).

But proteins and especially larger macromolecular complexes are large, flexible and dynamic molecules and naturally resist crystallization. Therefore, in most of the cases to increase the crystallizability of the proteins, the proteins have to be redesigned by removing flexible loop regions and often dividing into smaller domains and then determining the structure of each domain separately. These regions can be predicted by *in-silico* and wet-lab based methods such as limited proteolysis. The larger the size and the complexity of the macromolecule to be crystallized, the smaller is the chance for crystals that will diffract to high resolution. In addition to engineering the protein, multiple constructs of proteins with different lengths and tags are screened simultaneously and entered into the crystallization pipeline. This increases the chance of obtaining crystals.

Availability of commercial screens and robotic systems have made the initial screening procedure fast and convenient but finding the right crystallization condition is still the bottleneck (Rhodes, 2006b). Moreover, the challenge is not only to get a crystal, but a good diffraction quality crystal (visibly big, large, no cracks). Therefore, once a hit (the initial crystal) is obtained, the crystallization condition is optimized systematically to get big and better diffracting crystals. It begins with tweaking the concentration of the components in the crystallizing condition (salt, precipitant, pH of buffer) systematically and broadening out to test new additives (amino acids, polyamines, reducing agents) in the crystallization conditions. Additional parameters like temperature, drop size and drop volume will add a second layer of variety to the optimization procedure. Another commonly used technique is seeding that uses pieces of crystals as nucleation points to grow new crystals of the macromolecule. Dehydrating fully grown crystals is yet another strategy, where the fully-grown crystals are washed in range of solutions with the aim of reducing the solvent content

of the crystal, that will possible shrink the crystal lattice and thereby enhance the crystal packing resulting in a better diffracting crystal. Variations added can be the type of dehydrant, ratio, temperature, time. Optimizing crystals mostly involve doing all the above steps in parallel but systematically (Wheeler et al., 2012). In short, just obtaining a good quality diffracting crystal can be a challenge in itself and can take several years. It should be noted that the crystal structure is one static conformation of the macromolecule and doesn't represent the dynamic state.

#### **6.4.2 Electron cryomicroscopy, cryoEM**

**cryoEM** is another technique that have undergone tremendous development and received immense attention in the last few years. There are three variants of it: single-particle analysis (SPA), electron tomography and electron crystallography. Single-particle cryoEM is the main method used in Paper III and will be briefly discussed below.

The principle of SPA is based on transmission electron microscopy where a thin layer of vitrified sample is exposed to a more or less coherent beam of electrons. The electrons will interact with the specimen giving rise to both elastic and inelastic scattering. The lens system converts the elastically scattered electrons to an image that can be recorded on a detector. Images are processed to yield a three-dimensional reconstruction of the particle (Frank, 1975; Kuhlbrandt, 2014; Wang, 2015). The following steps are involved in a cryoEM experiment:

##### *6.4.2.1 Sample preparation*

The aim for sample preparation is to obtain thin layer of vitrified sample through which the electron beam can transmit. This can be obtained by applying the sample to an electron microscopy grid followed by removing excessive sample by blotting. The grid is then rapidly plunged frozen in liquid ethane which is at -180°C resulting in the fixation of the particles in their native conformation state. Ethane has a higher heat capacity than liquid nitrogen which will take comparatively more time to cool the grid and can cause crystallization of ice and sample. Therefore, ethane is preferred (Carroni and Saibil, 2016; Dubochet and McDowell, 1981).

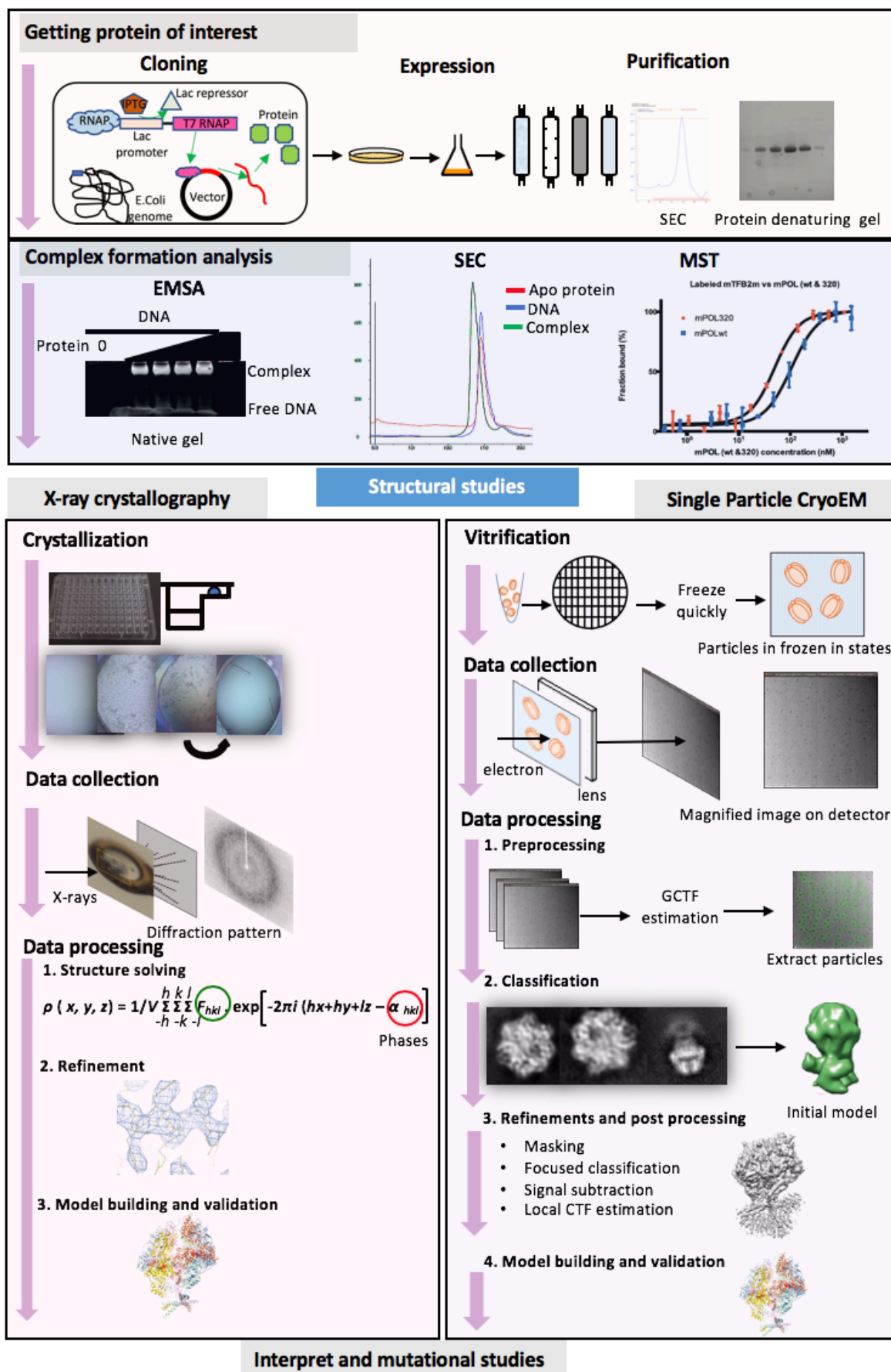
Before applying the sample, the surface properties of the grid should be changed by glow-discharging the grid. During glow-discharging the grid is placed in an evacuated chamber connected to power supply. As the current is passed, the air inside the chamber get ionized resulting in the deposition of negatively charged ions on the surface of the carbon grid. This grid can now hold on to the hydrophilic biological samples. Sometimes extra layer of thin continuous carbon is applied to evenly distribute the particles on the grid but this may decrease the signal to noise ratio from the protein as the carbon itself scatters electrons. There is huge variety of commercially available grids with different types of coating, mesh and hole size. Vitrification is semi-automated these days by using robotics that mainly control the vitrifying condition. Usually multiple grids with different blotting conditions and concentrations of the sample are prepared and screened for best ice and sample on the microscope before the actual data collection step (Aebi and Pollard, 1987).

#### *6.4.2.2 Aligning the microscope*

The microscope is fine tuned for best performance. This involves gun alignment, C2 aperture centering, beam tilt, objective aperture centering, lens astigmatism and gain reference for detector which is done before screening the grid (Sun and Li, 2010).

#### *6.4.2.3 Screening the grid and test data collection*

The vitrified grid is exposed to electrons in a screening microscope (100-200 KeV) with a good detector. The aim is to look for nice distribution of particles in reasonably thin ice. Once good grids are obtained, a small test dataset is collected, processed and analyzed based on parameters like, resolution, sample homogeneity, signal to noise ratio, orientation distributions and number of particles. If the data looks reasonable, more data should be collected from the same grid. If not that then the procedure should be repeated all over (Carroni and Saibil, 2016; Herzik et al., 2017).



**Figure 18:** Summary of the methods used in the current study. See text for details



## 6.5 COMPARISON OF X-RAY CRYSTALLOGRAPHY AND CRYOEM

**Table 5:** Comparison of X-ray crystallography and single particle Cryo-EM.

Characteristic	X-ray crystallography	CryoEM
<b>Sample</b>	Require crystals	Require thin layer of vitrified sample
<b>Source</b>	X-rays give diffraction pattern according to Bragg's law	Electrons give direct images of the particles
<b>Information obtained</b>	No focus lens for x-rays so phase information is lost	Use of focus lens, phases are also recorded
<b>Sample requirements</b>	Purity highly important	Purity and homogeneity help, heterogeneity can potentially be removed computationally
<b>Sample concentration</b>	Above 10 mg/ml concentration of sample required	0.5-1 mg/ml of sample required
<b>Sample volume</b>	35 $\mu$ l for one screening experiment	3 -10 $\mu$ l for one screening experiment
<b>State</b>	Static conformation	Different dynamic states can be captured at the same time
<b>Accuracy</b>	As high concentration of protein is used, non-biological conformations can exist	Biologically relevant and close to native states are achieved
<b>Additional assessment</b>	Complementary methods require to ensure the state observed is not crystalline artifact	Not required usually
<b>Size limitation</b>	The smaller and more compact the macromolecule, the higher is the chance for high-quality diffraction	The larger the particle, the higher is the chance to obtain high resolution structural information
<b>Resolution</b>	Atomic resolution up till 0.5 Å. Typically 2-3 Å.	Typically, 2.5-8 Å resolution
<b>Data Processing</b>	Fast and well-developed methods	Time consuming, Methods are developing.
<b>Bottle neck</b>	Crystallization and achieving well diffracting crystals	Obtaining good grids and data processing
<b>Cost</b>	Affordable	Quite expensive data collection and processing time
<b>Target</b>	Preferred for capturing static conformation of < 100 kDa at molecular resolution	Preferred for capturing dynamic states of big machineries > 250 kDa at medium to high resolution.

As mentioned in table, both structural techniques have their pros and cons and complement each other. cryoEM can be used to get information of the overall architecture and it can at the same time guide to pinpoint the flexible parts.

## 7 SUMMARY OF PAPERS

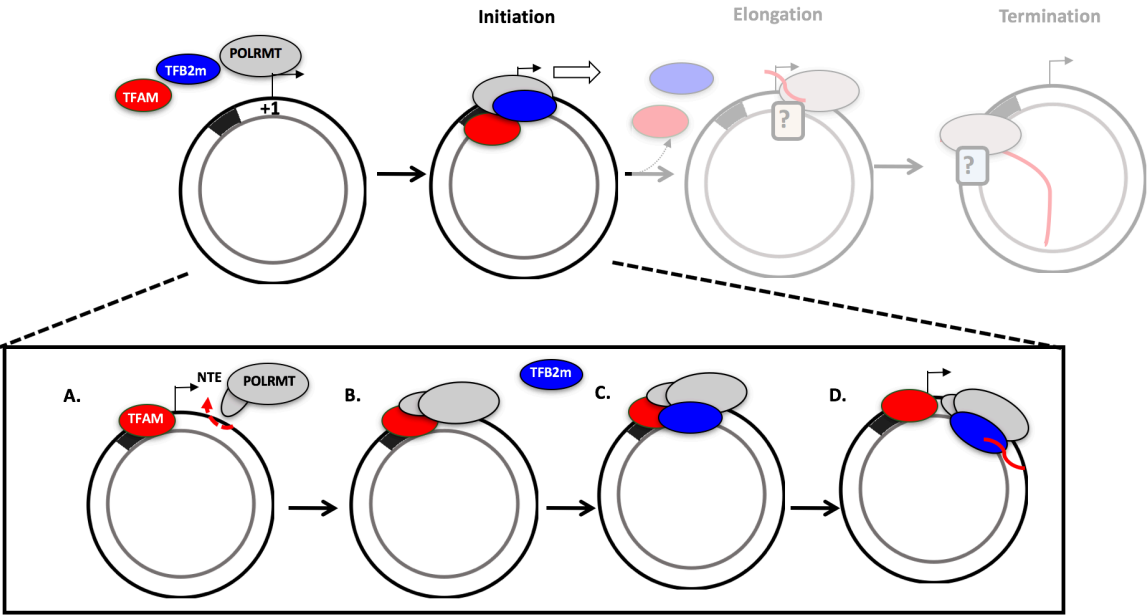
### 7.1 PAPER I:

#### **The amino terminal extension of mammalian mitochondrial RNA polymerase ensures promoter specific transcription initiation**

This part of the PhD project was aimed at understanding mitochondrial transcription and this paper addresses the transcriptional machinery that is involved during initiation of mitochondrial transcription. The catalytic core of the human mitochondrial RNA polymerase (POLRMT) is structurally similar to the bacteriophageal T7 RNA polymerase, implying that the mechanism of RNA synthesis is similar. However, POLRMT has a unique N-terminal extension (NTE) with unknown function. A previous crystal structure of a POLRMT construct that lacks the initial 104 amino acids in the construct showed only density from residue 218. Biochemical studies of the above construct showed that it is active on dsDNA while the construct 200 and 368 could only carry out transcription in the presence of a pre-melted templates, suggesting a role of the NTE in transcription initiation (Ringel et al., 2011). A similar truncation in yeast polymerase, Rpo41 affects its binding to the transcription factor Mtf1 (Paratkar et al., 2011). The function of NTE in mammalian mitochondrial transcription remained largely unknown.

To address the possible roles of this NTE in regulation, a project was initiated where the NTE of mouse POLRMT was truncated to different lengths and its effect on transcription was investigated by biophysical and biochemical methods. Out of many mutants that were designed, only one mutant where the initial 320 amino acids were removed (denoted as mPOL320), was found to be active in *in-vitro* transcriptional assays. A microscale thermophoresis (MST) experiment was done on different relevant complexes. MST measures the change in migration of partner-free and partner-bound protein induced by a temperature gradient. Upon titration of the labelled transcription factor B2, TFB2m, to the wtPOL and mPOL320, we found that the mPOL320 has increased affinity for TFB2m and it forms a tight complex. The complex can be purified on a gel filtration column. The *in-vitro* experiments showed that mPOL320 clearly increases the transcriptional activity and the truncation makes the POLRMT independent of the promoter-recognition factor, TFAM causing it to be hyper active and transcribe even from non-specific DNA sequences. To further examine if the NTE truncation is having an effect on initiation complex formation a DNase I foot-printing experiment was performed, which showed that NTE is having a negligible effect on the protection pattern. It was shown that only TFAM protects a region upstream from the transcription start site at position -10 to -40. Neither POLRMT nor TFB2M in isolation nor when added together could create a footprint. In the presence of all three components, a protection at -50 and -60 region was also noticed, which is in complement to a -40 to +10 footprint reported in previous studies (Gaspari et al., 2004). With the same experiment, the binding order of proteins required for transcription initiation were determined to be TFAM followed by the complex of POLRMT-TFB2m. The results obtained in this study differ from previous reports where a similar truncation in yeast Rpo41 will yield a non-active polymerase

(Paratkar et al., 2011), while in mouse POLRMT, removal of NTE will still result in an active polymerase, even over active. Removal of NTE affects the interaction of POLRMT with its transcriptional factor, TFAM and TFB2m implying that NTE plays an important regulatory role and ensures promoter-specific transcription.



**Figure 19:** Summary of results of Paper I.

## 7.2 PAPER II:

### **TEFM is a potent stimulator of mitochondrial transcription elongation *in-vitro***

This part of the project addresses the mitochondrial transcriptional machinery during its elongation phase and studies the potential role of a novel protein factor involved in the elongation phase. The transcription elongation factor (TEFM) was identified in 2011, by a RNAi mediated knockdown in human cells, as a protein required in the formation of promoter distal transcripts by interacting with the catalytic core of the mitochondrial RNA polymerase, POLRMT (Minczuk et al., 2011). To further investigate the biochemical and structural role of TEFM in mitochondrial transcription, a study was initiated as a PhD project where *in-vitro* transcription assays, microscale thermophoresis, MST and DNase I foot printing techniques were used to answer the biological question. The project was in collaboration with Claes Gustafson from GU.

Using *in-vitro* transcription assays, we were able to show that only by using initiation transcriptional machinery i.e. POLRMT with two initiation transcription factors TFAM and TFB2m, transcription can be initiated from LSP transcription start site but 65 % of these transcription events are prematurely abolished just after a short RNA of 120 nucleotides is synthesized. However, in the presence of TEFM, this premature termination is abolished. The termination region corresponds to the conserved sequence block II region (CSBII) which in the mitochondrial genome transcribes to a G-quadruplex. G-quadruplexes are guanine rich sequences that stack in planes to form bulky, stable regions. TEFM assists in the transcription of such putative termination points. Further experiments showed that TEFM also allows the mitochondrial transcription machinery to bypass normal DNA lesions like apurinic or apyrimidinic sites (AP-site) and 8-oxo-2'-deoxyguanosine (8-oxo-dG) sites.

Complex formation experiment on size exclusions chromatography suggested that the TEFM and POLRMT can interact in the absence of RNA-DNA and interactions is stronger in the presence of RNA transcribing DNA. This affinity was measured quantitatively by Microscale thermophoresis experiment, MST. By DNase I foot printing experiments, it was showed that TEFM is involved even at the initiation stage of mitochondrial transcription as the foot print was altered in the presence of TEFM.

All these experiments led us to formulate that TEFM is a potent transcription elongation factor and it increases the POLRMT affinity to DNA template and make it less prone to termination events. In addition, POLRMT completely abolishes the termination event at CSBII region, suggesting that TEFM could be involved in generating the replication primers for OriH transcription. A similar report at that time suggested that TEFM could act as a control between replication and transcription (Agaronyan et al., 2015). In our paper, we did mention that this could be linked but if it directly acts as a switch needs to be addressed with

further

experiments.

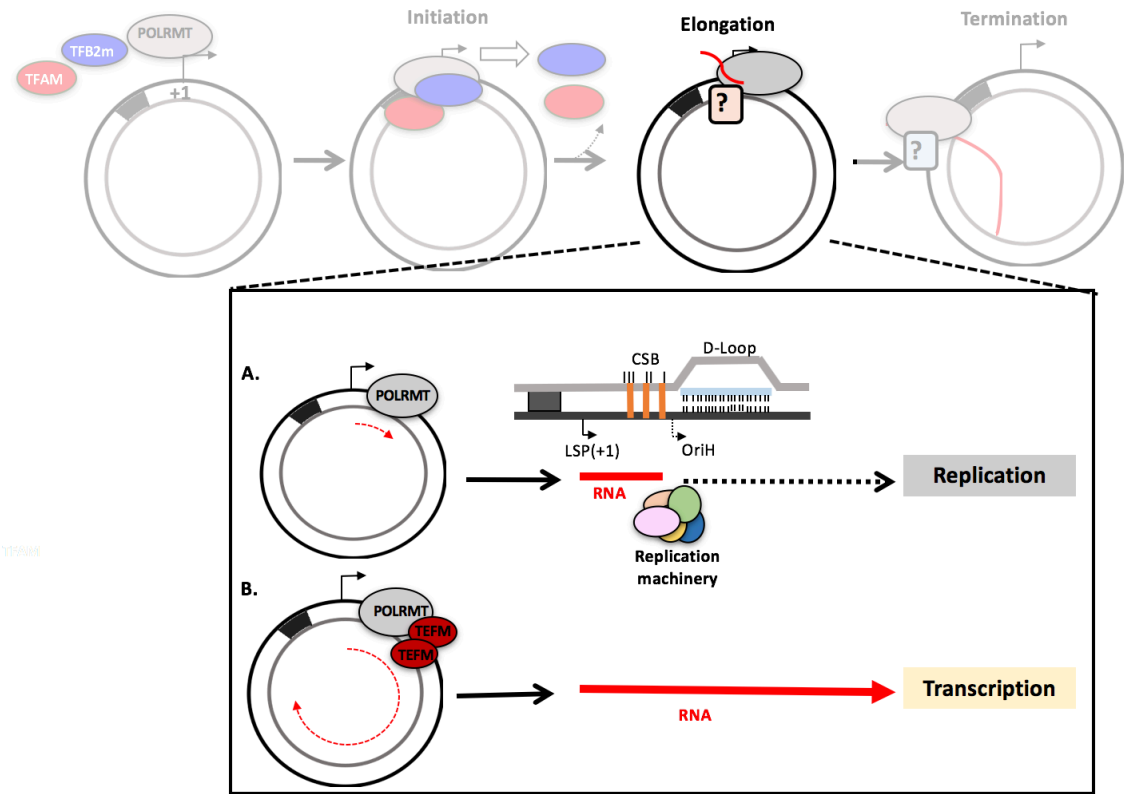


Figure 20: Summary of results of Paper II.

### 7.3 PAPER III:

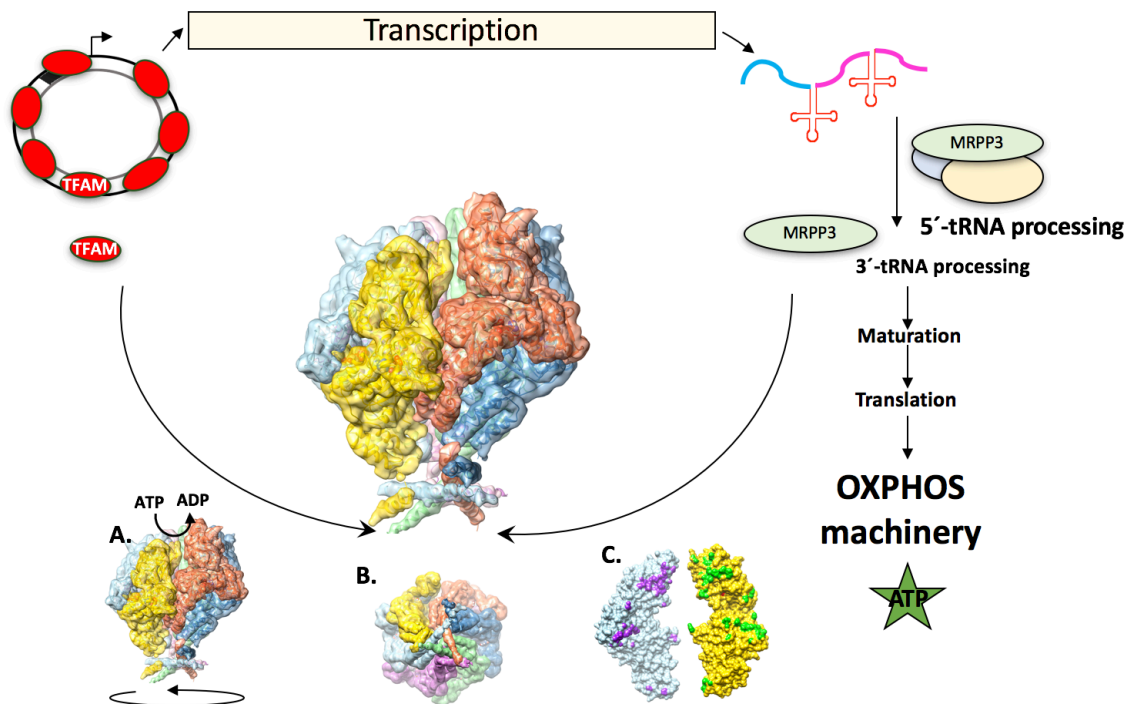
#### Structure and degradation mechanism of the Human Mitochondrial Lon protease

Human mitochondrial Lon is AAA<sup>+</sup> protease (ATPase Associated with diverse cellular Activities) and performs the crucial role of degrading misfolded, oxidatively modified proteins in human mitochondria. Furthermore, it is involved in the metabolic regulation by degrading proteins involved in replication, repair and transcription of the mitochondrial genome (Bota et al., 2005). For instance, it has been shown to regulate the level of TFAM, the transcription factor in human mitochondria and specifically recognizes and degrades the DNA-free TFAM (Lu et al., 2013). Another study showed that upon mitochondrial stress followed by the triggering of the mitochondrial unfolded protein response, Lon specifically degrades MRPP3 (Münch and Harper, 2016). MRPP3 is the nuclease subunit in the mitochondrial RNase P complex that performs the initial step in mitochondrial RNA processing. Moreover, Lon has been implicated in multisystem developmental disease, CODAS syndrome (Cerebral, Ocular, Dental, Auricular Skeletal) (Peter et al., 2018; Pinti et al., 2016a; Strauss et al., 2015). LonP1 is a homo-hexamer that uses energy from ATP hydrolysis to recognize, bind, and translocate its substrate into a proteolytic chamber. Earlier studies of individual domains and truncated proteins of bacterial and archaeal homologues, along with the low-resolution structure (15-20 Å) of human LonP1 have suggested the three-dimensional architecture of human LonP1. However, the absence of a structure of a full-length eukaryotic mitochondrial Lon have always hampered our understanding.

For this purpose, human mitochondrial Lon, human LonP1 was recombinantly expressed, purified as a hexamer, vitrified on electron microscopy grids and screened for single-particle electron cryoEM microscopy (cryo-EM). Extensive data collections, optimisations and data processing has resulted in a high-resolution electrostatic potential map of human mitochondrial Lon with information up to 3.6 Å. The atomic model of LonP1 protease was built and refined. The structure shows that human mitochondrial Lon is a homo-hexamer where the promoters are arranged in an open helical conformation with a slight translational shift. Additionally, we found that the legs have a unique dimeric arrangement, where the LAN domains of every first and fourth protomer is interacting thereby giving rise to an extra pore in the neck region which acts as the first gate to the chamber entrance. Accompanied by the ATPase and protease assays, we are able to say how LonP1 may capture substrates and translocate it through the chamber towards the proteolytic domain. A crystal structure of a bacterial homologue suggests that Lon uses three of its protomers synergistically and the sequential movements of non-neighbouring protomers in two sets unfold the substrate (Lin et al., 2016). This is in contradiction to what is reported for the other ATP-dependent proteases : 26S proteasome (de la Pena et al., 2018), Vps4-vta1 (Monroe et al., 2017), Hsp104 aggregase (Gates et al., 2017), YME1 (Puchades et al., 2017), VAT protein (Ripstein et al., 2017). Our single-particle analysis of a full-length human LonP1 suggests that it exhibits a combination of both, Lon exhibits hand over hand model where three protomers provide the hand. We propose that ATP is hydrolyzed by all the protomers in order, where one ATP

hydrolysis event by one protomer moves the other protomers through extensive communication which is not restricted to the chamber. The structure is the first high-resolution structure of human mitochondrial Lon.

Given the consideration of Lon as one of the potential therapeutic target and potential biomarker for cancer diagnosis(Lu, 2017) (Bota and Davies, 2016).The availability of the high resolution structure is one step forward toward understanding the dynamic machinery. Moreover, its only through the cryoEM that the dynamics of the AAA+ machinery are captured and additionally have guided to the static part of the protein which can be solved by x-ray crystallography.



**Figure 22:** Summary of results of Paper III.



## 8 FUTURE PERSPECTIVES

The mitochondrion is the power house of the cell and abnormal functioning of the mitochondrion is linked to variety of metabolic, genetic, and age-related diseases. The aim of the current PhD was to explore the structural-functional relationships of mammalian mitochondrial matrix proteins. There are almost 1000-1500 proteins present in human mitochondria and the structures of only 190 proteins (or parts of proteins) are reported in the structural data bank, PDB. With 97% of them being solved by X-ray crystallography. X-ray crystallography is an excellent and so far, the only available technique to get high atomic resolution information for the static thermodynamically favourable fold of a small protein. This information is mandatory for a pharmaceutical drug design. But proteins are dynamic machines with flexible regions busy in carrying on basic functions that are required for life. In the present era, single particle cryoEM has emerged as a strong tool to study big dynamic machineries at high resolution giving rise to endless possibilities.

The specific aim of this thesis was to understand protein machinery involved in transcription and quality control. With Paper I, the proteins involved in first step of transcription i.e. initiation were studied and a model for the binding order of the proteins was proposed. The proposed model was later supported by the follow-up studies, especially the structural report of the initiation complex (Hillen et al., 2018). Although the structural report is intriguing, there are some basic questions that are not answered. The reason perhaps lies in the truncations made in order to be able to crystallise the proteins. For example, 20-62 residues of TFB2m, which are involved in DNA priming are removed in the published structure. Interestingly, the study reported that removal of these residues does not affect the function of the protein which is to separate the dsDNA at the transcription start site. But later, instead of a normal dsDNA - that truncated TFB2m is able to separate, a pre-melted dsDNA was used in the structure. As a result, only a few core contacts were recorded. Now with the availability of cryoEM, there is a possibility to capture a transcribing transcription complex. It must be noted that the complex is at border-line size-wise for cryoEM and another challenge will be to capture a stable homogenous complex.

In Paper II, the elongation phase of transcription was investigated. We were able to characterise a putative transcription elongation factor that turned out to be critical for the processivity of the mitochondrial RNA polymerase. A later study reported the structure of transcription elongation complex with TEFM. Once again as per demand of the technique, X-ray crystallography, all the flexible regions of the protein were removed and only the core of the complex was crystallised. As a result, there are still unaddressed questions that need further investigation. This complex would be interesting to be explored with cryoEM.

In Paper III, the cryoEM structure of human mitochondrial Lon protease, LonP1 is reported. The dynamic structure is intriguing but demands more mutational, biochemical and structural work to fully understand the mechanism. Still it is the first report of a high-resolution full-

length structure of human LonP1 that can work as a guide to answer mechanistic questions about substrate recognition, substrate translocation and ATP hydrolysis. It would be interesting to see if the wt-type human LonP1 with ADP bound has a similar structure as the mutant LonP1 we were able to obtain a high-resolution reconstruction of. Now with the conditions of purification, data collection and processing being established, there is a possibility to capture the protease in different states of action. Given the involvement of Lon in cancer and proposal of Lon being used as a potential biomarker for cancer and target for other therapeutics, the research work can be extended to drug design.

To conclude, the method of cryoEM together with the unexplored large mitochondrial machineries, it can be certainly said that the future for mitochondrial research holds promise. Both structural techniques x-ray crystallography and cryoEM can complement each other to understand the structural-functional relationship of a multitude of interesting mitochondrial macromolecular complexes.

## 9 ACKNOWLEDGEMENTS

PhD is a journey to explore the unknown from known. It's a process with many critical stages of reading, planning, learning, performing, processing, analyzing and writing. The one person who guides you through all these critical steps to reach your goal is your supervisor. I would like to take this opportunity to thank **Dr. Martin Hällberg**, my supervisor for making me what I am now, a Researcher. I did not only acquire skills from him, I learnt to be resilient, thoughtful, and patient. I learnt to do good meaningful science. I learned to plan, prioritize and stay focus. If it was planning an experiment or buying a Halloween costume for my child, his door was always open. How can I not mention the tireless synchrotron trips or the weekend experiments and the statement 'night is still young'. I guess, just saying Thank you will not be merely enough but still THANK YOU MARTIN. I will miss working with you.

I would like to thank my co-supervisor **Dr. Nils-Göran Larsson** for arranging excellent mito-retreats and seminar series.

A wishful thank to our collaborators from Gothenburg University, **Dr. Maria Falkenberg**, **Dr. Claes Gustafsson**, **Dr. Viktor Posse** and soon to be **Dr. Carlos Pardo Hernandez**. It was pleasure working with you.

I would like to take an opportunity to also thank **Matti Nikkola**, the director of PhD studies in KI, for all the moral support. It's good to know that there is a person who can always lend an ear and share an experienced advice. A big thank goes to **Dr. Sara Sandin**, my mentor for not only your scientific advice and professional guidance but for welcoming and practically assisting me in my research trip to NTU, Singapore.

The scientific quest would have been impossible without the technical expertise of **Dr. Julian Conrad** and **Dr. Marta Corroni**, from SciLife Lab, Stockholm. Thanks Julian, especially for helping me understand the ice thickness on grid, the counting mode on K2 and no doubt for professional career advice. Marta, I owe a big thanks to you not only for arranging the resourceful cryoEM course but also for critical scientific advice on my projects. I wish, I could learn more from both of you.

My thesis would be incomplete, if I will not thank the person who introduced me to structural biology, **Dr. Christina Divne**. It was the passion with which Christina explained the structures that I just fell in love with the field. Thanks, Christina for arranging not only a good course but welcoming me in your scientific team at MBB, KI. **Dr. Rosaria Gandini** and **Dr. Tien Chen tye** (TC), you were the source of practical scientific knowledge for me. I still remember the first synchrotron trip to Lund, and how tirelessly, we worked for it. Especially Rosaria, I missed you.

To all the ex-colleagues who have found new ventures in life: **Dr. Agnes Zimmer**, you were the first one with whom I shared the lab in my early PhD days. Hats off to the management, everything at its place, labelled and categorized. Your presentations in the group meetings were an inspiration. No doubt, you have chosen the best field for yourself.

**Dr. Karin Wallden**, a friend, a colleague, a mentor, all in one package. Always, willing to help with critical and sincere advice. I will cherish the memories of booking clashes on the Akta system and the clogging of homogenizer by a cleaning brush.

**Dr. Linda Reinhard** and **Dr. Mikalai Lapkouski**, even though we didn't share the lab physically but still whatever I learned in those virtual group lab meetings and journal clubs, is an asset. How can I not mention, my then PhD fellow student and now **Dr. Sagar Sridhara**. I will always remember in good words the time we spent together, the course in Uppsala, the discussions, fikas, lunches and I will remember, you asking me every now and then 'how many crystals do you have in total?' and my reply 'Crystallization and a good resolution are two different things, wish for good resolving crystal'.

I would also like to thank my office colleagues at KI for creating a productive workplace; **Christer Hoog's**, **Lena Strom's** and **Laura Baranello's** group members. Thanks Lena for being there for everyone, whether it's a lab organizational issue or a halftime presentation. Thanks Laura, for arranging the seminar series and the welcoming 'Hi'. A special thanks to **Sonata Valentinienė**, my lunch and office mate and always with a smile. I will miss that all-around discussions with you. I wish we could have spent more time together. **Dr. Anika Wiegard** and **Dr. Anna Kuznetsova** for thoughtful writing advices.

It is not exaggeration to say that it takes a village to raise a PhD. To, all my off-work friends, **Sara, Arshi, Amna, Taiba, Sobia and Madeeha**. Thank you for the moral and physical support, whether it was arranging weekend dinners or family trips, it definitely relieved the stress a lot and revitalize me for forthcoming tasks.

My lovely sister, my blood and friend, **Ms. Kiran Shahzad**, always there in every possible way. Having you is like having a part of my parents and childhood with me all the time. My brother Sohail, Adil and sister Hina, thanks for the support.

My extended family members: my father-in-law, **Muhammad Akbar** kept my spirits high, retelling me our 'family' aspiration to earn a PhD. My mother-in-law, **Khursheed Begum**, I knew I could always trust in your sincere prayers. My sisters-in-law, **Mrs. Yasmin Tariq, Shama, Maria** and especially **Tahira**, if you would not have travelled to Sweden, for couple of times to be with us for six-months, I would not have been able to upkeep my ever-challenging research responsibilities.

Loves of my life, my children: **Amna** (10 yrs. old), **Ahmad** (8 yrs. old) and **Wali** (3 yrs. old), if it was anything that could make me relax in a second, was hug from you. Your smell could just make every worry blow away and yes, I would proudly say that my kids endured the PhD along with me. They will surely miss the weekend trips to lab.

The second person after my supervisor, without whom my PhD was impossible and unmanageable, is my half-soul, my husband **Dr. Nadeem Akbar**. THANK YOU for ALL the support, words are not enough. It was demanding and stressful but you know ‘We did it together’ and I believe it’s just the beginning of a new journey.

Two things are completely matchless in this world- The nurturing love of parents, and an innocent love of a child. To those who don’t know me, I was born in Pakistan to an-average educated family. My Mother, **Riffat Tahira**, who sought her kids to educate, her prayers and being always there with a shoulder to vent myself in a moment of need. My father, **Mr. Shahzad Ahmad**, a policeman preferred to get us (me and my siblings) educated rather than investing in business and I think he made the right choice. Educating a woman is educating a family and it is an investment for unborn future generations. Words lose their values, but customs I follow, to say thank you with a forehead kiss, on much awaited day when they can see me in the black gown receiving a degree from the world-renowned university, **Karolinska Institute**.

Lastly thanks **Sweden** for warmly welcoming enthusiastic optimistic people with open hands.

Thank you, **reader**, for spending your precious moments.



## 10 REFERENCES

- Aebi, U., and Pollard, T.D. (1987). A glow discharge unit to render electron microscope grids and other surfaces hydrophilic. *J Electron Microscop Tech* 7, 29-33.
- Agaronyan, K., Morozov, Y.I., Anikin, M., and Temiakov, D. (2015). Mitochondrial biology. Replication-transcription switch in human mitochondria. *Science* 347, 548-551.
- Agrawal, R.K., and Sharma, M.R. (2012). Structural aspects of mitochondrial translational apparatus. *Curr Opin Struct Biol* 22, 797-803.
- Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., *et al.* (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290, 457-465.
- Andersson, S.G., Karlberg, O., Canbäck, B., and Kurland, C.G. (2003). On the origin of mitochondria: a genomics perspective. *Philos Trans R Soc Lond B Biol Sci* 358, 165-177; discussion 177-169.
- Antonicka, H., Sasarman, F., Nishimura, T., Paupe, V., and Shoubridge, E.A. (2013). The mitochondrial RNA-binding protein GRSF1 localizes to RNA granules and is required for posttranscriptional mitochondrial gene expression. *Cell Metab* 17, 386-398.
- Arndt, C., Koristka, S., Bartsch, H., and Bachmann, M. (2012). Native polyacrylamide gels. *Methods in molecular biology (Clifton, NJ)* 869, 49.
- Asin-Cayuela, J., Schwend, T., Farge, G., and Gustafsson, C.M. (2005). The Human Mitochondrial Transcription Termination Factor (mTERF) Is Fully Active in Vitro in the Non-phosphorylated Form. *Journal of Biological Chemistry* 280, 25499-25505.
- Baker, B.M., and Haynes, C.M. (2011). Mitochondrial protein quality control during biogenesis and aging. *Trends Biochem Sci* 36, 254-261.
- Baker Michael, J., Palmer Catherine, S., and Stojanovski, D. (2013). Mitochondrial protein quality control in health and disease. *British Journal of Pharmacology* 171, 1870-1889.
- Baker, M.J., Tatsuta, T., and Langer, T. (2011). Quality control of mitochondrial proteostasis. *Cold Spring Harb Perspect Biol* 3.
- Baker, T.A., and Sauer, R.T. (2012). ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim Biophys Acta* 1823, 15-28.
- Bakkaiova, J., Marini, V., Willcox, S., Nosek, J., Griffith, Jack D., Krejci, L., and Tomaska, L. (2016). Yeast mitochondrial HMG proteins: DNA-binding properties of the most evolutionarily divergent component of mitochondrial nucleoids. *Bioscience Reports* 36, e00288.
- Baranek, M., Grabsztunowicz, M., Sikora, B., and Jackowski, G. (2011). [ATP-dependent FtsH and Lon chloroplast proteases]. *Postepy Biochem* 57, 101-108.
- Bard, J.A.M., Goodall, E.A., Greene, E.R., Jonsson, E., Dong, K.C., and Martin, A. (2018). Structure and Function of the 26S Proteasome. *Annual Review of Biochemistry* 87, 697-724.
- Battey, J., and Clayton, D.A. (1978). The transcription map of mouse mitochondrial DNA. *Cell* 14, 143-156.

- Berg, J., M, Tymoczko, J., L, and Stryer, L. (2012). *Biochemistry* (New York: W. H. Freeman).
- Bernecky, C., Plitzko, J.M., and Cramer, P. (2017). Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat Struct Mol Biol* 24, 809-815.
- Bezawork-Geleta, A., Brodie, E.J., Dougan, D.A., and Truscott, K.N. (2015). LON is the master protease that protects against protein aggregation in human mitochondria through direct degradation of misfolded proteins. *Scientific Reports* 5, 17397.
- Bezawork-Geleta, A., Saiyed, T., Dougan, D.A., and Truscott, K.N. (2014). Mitochondrial matrix proteostasis is linked to hereditary paraganglioma: LON-mediated turnover of the human flavinylation factor SDH5 is regulated by its interaction with SDHA. *FASEB J* 28, 1794-1804.
- Bogenhagen, D., and Clayton, D.A. (1974). The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial deoxyribonucleic acid. *J Biol Chem* 249, 7991-7995.
- Bogenhagen, D.F., Applegate, E.F., and Yoza, B.K. (1984). Identification of a promoter for transcription of the heavy strand of human mtDNA: in vitro transcription and deletion mutagenesis. *Cell* 36, 1105-1113.
- Bogenhagen, D.F., Rousseau, D., and Burke, S. (2008). The layered structure of human mitochondrial DNA nucleoids. *J Biol Chem* 283, 3665-3675.
- Borkotoky, S., and Murali, A. (2018). The highly efficient T7 RNA polymerase: A wonder macromolecule in biological realm. *International Journal of Biological Macromolecules* 118, 49-56.
- Bota, D.A., and Davies, K.J. (2002). Lon protease preferentially degrades oxidized mitochondrial aconitase by an ATP-stimulated mechanism. *Nat Cell Biol* 4, 674-680.
- Bota, D.A., and Davies, K.J. (2016). Mitochondrial Lon protease in human disease and aging: Including an etiologic classification of Lon-related diseases and disorders. *Free Radic Biol Med* 100, 188-198.
- Bota, D.A., Ngo, J.K., and Davies, K.J. (2005). Downregulation of the human Lon protease impairs mitochondrial structure and function and causes cell death. *Free Radic Biol Med* 38, 665-677.
- Botos, I., Melnikov, E.E., Cherry, S., Khalatova, A.G., Rasulova, F.S., Tropea, J.E., Maurizi, M.R., Rotanova, T.V., Gustchina, A., and Wlodawer, A. (2004a). Crystal structure of the AAA+ alpha domain of E. coli Lon protease at 1.9A resolution. *J Struct Biol* 146, 113-122.
- Botos, I., Melnikov, E.E., Cherry, S., Kozlov, S., Makhovskaya, O.V., Tropea, J.E., Gustchina, A., Rotanova, T.V., and Wlodawer, A. (2005). Atomic-resolution crystal structure of the proteolytic domain of *Archaeoglobus fulgidus* lon reveals the conformational variability in the active sites of lon proteases. *J Mol Biol* 351, 144-157.
- Botos, I., Melnikov, E.E., Cherry, S., Tropea, J.E., Khalatova, A.G., Rasulova, F., Dauter, Z., Maurizi, M.R., Rotanova, T.V., Wlodawer, A., *et al.* (2004b). The catalytic domain of *Escherichia coli* Lon protease has a unique fold and a Ser-Lys dyad in the active site. *J Biol Chem* 279, 8140-8148.
- Brown, T.A., Tkachuk, A.N., Shtengel, G., Kopek, B.G., Bogenhagen, D.F., Hess, H.F., and Clayton, D.A. (2011). Superresolution Fluorescence Imaging of Mitochondrial Nucleoids



- Reveals Their Spatial Range, Limits, and Membrane Interaction. *Molecular and Cellular Biology* 31, 4994-5010.
- Brzezniak, L.K., Bijata, M., Szczesny, R.J., and Stepien, P.P. (2011). Involvement of human ELAC2 gene product in 3' end processing of mitochondrial tRNAs. *RNA Biol* 8, 616-626.
- Cantatore, P., and Attardi, G. (1980). Mapping of nascent light and heavy strand transcripts on the physical map of HeLa cell mitochondrial DNA. *Nucleic Acids Research* 8, 2605-2626.
- Carroni, M., and Saibil, H.R. (2016). Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods (San Diego, Calif)* 95, 78-85.
- Ceschini, S., Keeley, A., McAlister, M.S.B., Oram, M., Phelan, J., Pearl, L.H., Tsaneva, I.R., and Barrett, T.E. (2001). Crystal structure of the fission yeast mitochondrial Holliday junction resolvase Ydc2. *The EMBO Journal* 20, 6601-6611.
- Cha, S.S., An, Y.J., Lee, C.R., Lee, H.S., Kim, Y.G., Kim, S.J., Kwon, K.K., De Donatis, G.M., Lee, J.H., Maurizi, M.R., *et al.* (2010). Crystal structure of Lon protease: molecular architecture of gated entry to a sequestered degradation chamber. *EMBO J* 29, 3520-3530.
- Chan, C., Harland, M.L., Webb, S.E., Chen, J., Miller, A.L., and Barritt, G.J. (2004). Evaluation, using targeted aequorins, of the roles of the endoplasmic reticulum and its (Ca<sup>2+</sup>++Mg<sup>2+</sup>)ATP-ases in the activation of store-operated Ca<sup>2+</sup> channels in liver cells. *Cell Calcium* 35, 317-331.
- Chang, D.D., and Clayton, D.A. (1985). Priming of human mitochondrial DNA replication occurs at the light-strand promoter. *Proc Natl Acad Sci U S A* 82, 351-355.
- Charette, M.F., Henderson, G.W., and Markovitz, A. (1981). ATP hydrolysis-dependent protease activity of the lon (capR) protein of *Escherichia coli* K-12. *Proc Natl Acad Sci U S A* 78, 4728-4732.
- Cheetham, G.M.T., Jeruzalmi, D., and Steitz, T.A. (1999). Structural basis for initiation of transcription from an RNA polymerase–promoter complex. *Nature* 399, 80.
- Cheetham, G.M.T., and Steitz, a.T.A. (1999). Structure of a Transcribing T7 RNA Polymerase Initiation Complex. *Science* 286, 2305.
- Chen, H., and Chan, D.C. (2004). Mitochondrial dynamics in mammals. *Curr Top Dev Biol* 59, 119-144.
- Chen, X., Prosser, R., Simonetti, S., Sadlock, J., Jagiello, G., and Schon, E.A. (1995). Rearranged mitochondrial genomes are present in human oocytes. *American Journal of Human Genetics* 57, 239-247.
- Cheng, C.W., Kuo, C.Y., Fan, C.C., Fang, W.C., Jiang, S.S., Lo, Y.K., Wang, T.Y., Kao, M.C., and Lee, A.Y. (2013). Overexpression of Lon contributes to survival and aggressive phenotype of cancer cells through mitochondrial complex I-mediated generation of reactive oxygen species. *Cell Death Dis* 4, e681.
- Chung, C.H., and Goldberg, A.L. (1981). The product of the lon (capR) gene in *Escherichia coli* is the ATP-dependent protease, protease La. *Proc Natl Acad Sci U S A* 78, 4931-4935.
- Clausen, T., Kaiser, M., Huber, R., and Ehrmann, M. (2011). HTRA proteases: regulated proteolysis in protein quality control. *Nature Reviews Molecular Cell Biology* 12, 152.
- Clayton, D.A. (2003). Mitochondrial DNA replication: what we know. *IUBMB Life* 55, 213-217.

- Cliften, P.F., Park, J.Y., Davis, B.P., Jang, S.H., and Jaehning, J.A. (1997). Identification of three regions essential for interaction between a sigma-like factor and core RNA polymerase. *Genes Dev* 11, 2897-2909.
- Dairaghi, D.J., Shadel, G.S., and Clayton, D.A. (1995). Addition of a 29 residue carboxyl-terminal tail converts a simple HMG box-containing protein into a transcriptional activator. *J Mol Biol* 249, 11-28.
- de la Pena, A.H., Goodall, E.A., Gates, S.N., Lander, G.C., and Martin, A. (2018). Structures of the substrate-engaged 26S proteasome reveal the mechanisms for ATP hydrolysis-driven translocation. *bioRxiv*.
- De Pinto, V.D., and Palmieri, F. (1992). Transmembrane arrangement of mitochondrial porin or voltage-dependent anion channel (VDAC). *J Bioenerg Biomembr* 24, 21-26.
- Dubochet, J., and McDowell, A.W. (1981). VITRIFICATION OF PURE WATER FOR ELECTRON MICROSCOPY. *Journal of Microscopy* 124, 3-4.
- Duman, R.E., and Löwe, J. (2010). Crystal structures of *Bacillus subtilis* Lon protease. *J Mol Biol* 401, 653-670.
- Durniak, K.J., Bailey, S., and Stetiz, T.A. (2008). The Structure of a Transcribing T7 RNA Polymerase Complex Captured During Its Transition from Initiation to Elongation. *Science (New York, NY)* 322, 553-557.
- Falkenberg, M. (2018). Mitochondrial DNA replication in mammalian cells: overview of the pathway. *Essays Biochem* 62, 287-296.
- Falkenberg, M., Gaspari, M., Rantanen, A., Trifunovic, A., Larsson, N.G., and Gustafsson, C.M. (2002). Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. *Nat Genet* 31, 289-294.
- Fisher, R.P., and Clayton, D.A. (1985). A transcription factor required for promoter recognition by human mitochondrial RNA polymerase. Accurate initiation at the heavy- and light-strand promoters dissected and reconstituted in vitro. *J Biol Chem* 260, 11330-11338.
- Fisher, R.P., Topper, J.N., and Clayton, D.A. (1987). Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements. *Cell* 50, 247-258.
- Fishovitz, J., Sha, Z., Chilakala, S., Cheng, I., Xu, Y., and Lee, I. (2017). Utilization of Mechanistic Enzymology to Evaluate the Significance of ADP Binding to Human Lon Protease. *Front Mol Biosci* 4, 47.
- Frank, J. (1975). Averaging of low exposure electron micrographs of non-periodic objects. *Ultramicroscopy* 1, 159-162.
- Fukuda, R., Zhang, H., Kim, J.-w., Shimoda, L., Dang, C.V., and Semenza, Gregg L. (2007). HIF-1 Regulates Cytochrome Oxidase Subunits to Optimize Efficiency of Respiration in Hypoxic Cells. *Cell* 129, 111-122.
- Fusté, J.M., Wanrooij, S., Jemt, E., Granycome, C.E., Cluett, T.J., Shi, Y., Atanassova, N., Holt, I.J., Gustafsson, C.M., and Falkenberg, M. (2010). Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication. *Mol Cell* 37, 67-78.
- García-Nafria, J., Ondrovicová, G., Blagova, E., Levnikov, V.M., Bauer, J.A., Suzuki, C.K., Kutejová, E., Wilkinson, A.J., and Wilson, K.S. (2010). Structure of the catalytic domain of the human mitochondrial Lon protease: proposed relation of oligomer formation and activity. *Protein Sci* 19, 987-999.

- Gaspari, M., Falkenberg, M., Larsson, N.G., and Gustafsson, C.M. (2004). The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells. *EMBO J* 23, 4606-4614.
- Gates, S.N., Yokom, A.L., Lin, J., Jackrel, M.E., Rizo, A.N., Kendsersky, N.M., Buell, C.E., Sweeny, E.A., Mack, K.L., Chuang, E., *et al.* (2017). Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104. *Science (New York, NY)* 357, 273-279.
- Gelfand, R., and Attardi, G. (1981). Synthesis and turnover of mitochondrial ribonucleic acid in HeLa cells: the mature ribosomal and messenger ribonucleic acid species are metabolically unstable. *Molecular and Cellular Biology* 1, 497.
- Ghivizzani, S.C., Madsen, C.S., Nelen, M.R., Ammini, C.V., and Hauswirth, W.W. (1994). In organello footprint analysis of human mitochondrial DNA: human mitochondrial transcription factor A interactions at the origin of replication. *Molecular and Cellular Biology* 14, 7717.
- Gibellini, L., Pinti, M., Beretti, F., Pierri, C.L., Onofrio, A., Riccio, M., Carnevale, G., De Biasi, S., Nasi, M., Torelli, F., *et al.* (2014). Sirtuin 3 interacts with Lon protease and regulates its acetylation status. *Mitochondrion* 18, 76-81.
- Gispert, S., Parganlija, D., Klinkenberg, M., Dröse, S., Wittig, I., Mittelbronn, M., Grzmil, P., Koob, S., Hamann, A., Walter, M., *et al.* (2013). Loss of mitochondrial peptidase Clpp leads to infertility, hearing loss plus growth retardation via accumulation of CLPX, mtDNA and inflammatory factors. *Human Molecular Genetics* 22, 4871-4887.
- Gold, V.A.M., Ieva, R., Walter, A., Pfanner, N., van der Laan, M., and Kühlbrandt, W. (2014). Visualizing active membrane protein complexes by electron cryotomography. *Nature Communications* 5, 4129.
- Gonzalez, M., Frank, E.G., Levine, A.S., and Woodgate, R. (1998). Lon-mediated proteolysis of the Escherichia coli UmuD mutagenesis protein: in vitro degradation and identification of residues required for proteolysis. *Genes Dev* 12, 3889-3899.
- Gottesman, S., and Zipser, D. (1978). Deg phenotype of Escherichia coli lon mutants. *J Bacteriol* 133, 844-851.
- Granot, Z., Kobiler, O., Melamed-Book, N., Eimerl, S., Bahat, A., Lu, B., Braun, S., Maurizi, M.R., Suzuki, C.K., Oppenheim, A.B., *et al.* (2007). Turnover of mitochondrial steroidogenic acute regulatory (StAR) protein by Lon protease: the unexpected effect of proteasome inhibitors. *Mol Endocrinol* 21, 2164-2177.
- Graslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., *et al.* (2008). Protein production and purification. *Nature Methods* 5, 135.
- Gray, M.W., Burger, G., and Lang, B.F. (1999). Mitochondrial evolution. *Science* 283, 1476-1481.
- Greenleaf, A.L., Kelly, J.L., and Lehman, I.R. (1986). Yeast RPO41 gene product is required for transcription and maintenance of the mitochondrial genome. *Proc Natl Acad Sci U S A* 83, 3391-3394.
- Gur, E., Ottofueling, R., and Dougan, D.A. (2013). Machines of destruction - AAA+ proteases and the adaptors that control them. *Subcell Biochem* 66, 3-33.

- Gur, E., and Sauer, R.T. (2008a). Evolution of the *ssrA* degradation tag in *Mycoplasma*: Specificity switch to a different protease. *Proceedings of the National Academy of Sciences* *105*, 16113.
- Gur, E., and Sauer, R.T. (2008b). Recognition of misfolded proteins by Lon, a AAA(+) protease. *Genes Dev* *22*, 2267-2277.
- Gur, E., Vishkautzan, M., and Sauer, R.T. (2012). Protein unfolding and degradation by the AAA+ Lon protease. *Protein Sci* *21*, 268-278.
- Gustafsson, C.M., Falkenberg, M., and Larsson, N.-G. (2016). Maintenance and Expression of Mammalian Mitochondrial DNA. *Annual Review of Biochemistry* *85*, 133-160.
- Górecka, K.M., Komorowska, W., and Nowotny, M. (2013). Crystal structure of RuvC resolvase in complex with Holliday junction substrate. *Nucleic Acids Research* *41*, 9945-9955.
- Hamdan, S.M., and Richardson, C.C. (2009). Motors, switches, and contacts in the replisome. *Annu Rev Biochem* *78*, 205-243.
- Hartvig, L., and Christiansen, J. (1996). Intrinsic termination of T7 RNA polymerase mediated by either RNA or DNA. *The EMBO Journal* *15*, 4767-4774.
- He, L., Luo, D., Yang, F., Li, C., Zhang, X., Deng, H., and Zhang, J.R. (2018). Multiple domains of bacterial and human Lon proteases define substrate selectivity. *Emerg Microbes Infect* *7*, 149.
- Herzik, M.A., Wu, M., and Lander, G.C. (2017). Achieving Better Than 3 Å Resolution By Single Particle Cryo-EM At 200 keV. *bioRxiv*.
- Hillen, H.S., Morozov, Y.I., Sarfallah, A., Temiakov, D., and Cramer, P. (2017a). Structural Basis of Mitochondrial Transcription Initiation. *Cell* *171*, 1072-1081.e1010.
- Hillen, H.S., Parshin, A.V., Agaronyan, K., Morozov, Y.I., Graber, J.J., Chernev, A., Schwinghammer, K., Urlaub, H., Anikin, M., Cramer, P., *et al.* (2017b). Mechanism of Transcription Anti-termination in Human Mitochondria. *Cell* *171*, 1082-1093.e1013.
- Hillen, H.S., Temiakov, D., and Cramer, P. (2018). Structural basis of mitochondrial transcription. *Nat Struct Mol Biol* *25*, 754-765.
- Holmes, J.B., Akman, G., Wood, S.R., Sakhuja, K., Cerritelli, S.M., Moss, C., Bowmaker, M.R., Jacobs, H.T., Crouch, R.J., and Holt, I.J. (2015). Primer retention owing to the absence of RNase H1 is catastrophic for mitochondrial DNA replication. *Proceedings of the National Academy of Sciences* *112*, 9334.
- Hornig-Do, H.T., Tatsuta, T., Buckermann, A., Bust, M., Kollberg, G., Rötig, A., Hellmich, M., Nijtmans, L., and Wiesner, R.J. (2012). Nonsense mutations in the COX1 subunit impair the stability of respiratory chain complexes rather than their assembly. *The EMBO Journal* *31*, 1293.
- Howard-Flanders, P., Simson, E., and Theriot, L. (1964). The excision of thymine dimers from DNA, filament formation and sensitivity to ultraviolet light in *Escherichia coli* K-12. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* *1*, 219-226.
- Hyvärinen, A.K., Pohjoismäki, J.L., Reyes, A., Wanrooij, S., Yasukawa, T., Karhunen, P.J., Spelbrink, J.N., Holt, I.J., and Jacobs, H.T. (2007). The mitochondrial transcription termination factor mTERF modulates replication pausing in human mitochondrial DNA. *Nucleic Acids Res* *35*, 6458-6474.

- Hällberg, B.M., and Larsson, N.-G. (2014). Making Proteins in the Powerhouse. *Cell Metabolism* *20*, 226-240.
- Ikeda, M., Ide, T., Fujino, T., Arai, S., Saku, K., Kakino, T., Tyynismaa, H., Yamasaki, T., Yamada, K.-i., Kang, D., *et al.* (2015). Overexpression of TFAM or Twinkle Increases mtDNA Copy Number and Facilitates Cardioprotection Associated with Limited Mitochondrial Oxidative Stress. *PLoS ONE* *10*, e0119687.
- Im, Y.J., Na, Y., Kang, G.B., Rho, S.H., Kim, M.K., Lee, J.H., Chung, C.H., and Eom, S.H. (2004). The active site of a Lon protease from *Methanococcus jannaschii* distinctly differs from the canonical catalytic Dyad of Lon proteases. *J Biol Chem* *279*, 53451-53457.
- Iosefson, O., Nager, A.R., Baker, T.A., and Sauer, R.T. (2015). Coordinated gripping of substrate by subunits of a AAA+ proteolytic machine. *Nature Chemical Biology* *11*, 201.
- Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H.M., Duhr, S., Baaske, P., and Breitsprecher, D. (2014). MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure* *1077*, 101-113.
- Jerker, P., Jan, C., Ingmar, O., and Greta, B. (1975). Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* *258*, 598.
- Jeruzalmi, D., and Steitz, T.A. (1998). Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *The EMBO Journal* *17*, 4101-4113.
- Jiménez-Menéndez, N., Fernández-Millán, P., Rubio-Cosials, A., Arnan, C., Montoya, J., Jacobs, H.T., Bernadó, P., Coll, M., Usón, I., and Solà, M. (2010). Human mitochondrial mTERF wraps around DNA through a left-handed superhelical tandem repeat. *Nature Structural & Molecular Biology* *17*, 891.
- Jourdain, A.A., Koppen, M., Wydro, M., Rodley, C.D., Lightowlers, R.N., Chrzanowska-Lightowlers, Z.M., and Martinou, J.C. (2013). GRSF1 regulates RNA processing in mitochondrial RNA granules. *Cell Metab* *17*, 399-410.
- Kanki, T., Nakayama, H., Sasaki, N., Takio, K., Alam, T.I., Hamasaki, N., and Kang, D. (2004). Mitochondrial nucleoid and transcription factor A. *Ann N Y Acad Sci* *1011*, 61-68.
- Karłowicz, A., Wegrzyn, K., Gross, M., Kaczynska, D., Ropelewska, M., Siemiątkowska, M., Bujnicki, J.M., and Konieczny, I. (2017). Defining the crucial domain and amino acid residues in bacterial Lon protease for DNA binding and processing of DNA-interacting substrates. *J Biol Chem* *292*, 7507-7518.
- Kerešiche, S., Kováčik, L., Bednár, J., Pevala, V., Kunová, N., Ondrovičová, G., Bauer, J., Ambro, L., Bellová, J., Kutejová, E., *et al.* (2016). The N-terminal domain plays a crucial role in the structure of a full-length human mitochondrial Lon protease. *Sci Rep* *6*, 33631.
- Kerešiche, S., Kováčik, L., Pevala, V., Ambro, L., Bellová, J., Kutejová, E., and Raška, I. (2014). Three-dimensional reconstruction of the S885A mutant of human mitochondrial Lon protease. *Folia Biol (Praha)* *60 Suppl 1*, 62-65.
- Kimple, M.E., Brill, A.L., and Pasker, R.L. (2013). Overview of Affinity Tags for Protein Purification. *Current protocols in protein science / editorial board, John E Coligan [et al]* *73*, Unit-9.9.
- Kita, K., Suzuki, T., and Ochi, T. (2012). Diphenylarsinic acid promotes degradation of glutaminase C by mitochondrial Lon protease. *J Biol Chem* *287*, 18163-18172.
- Korhonen, J.A., Pham, X.H., Pellegrini, M., and Falkenberg, M. (2004). Reconstitution of a minimal mtDNA replisome in vitro. *EMBO J* *23*, 2423-2429.

- Kruse, B., Narasimhan, N., and Attardi, G. (1989). Termination of transcription in human mitochondria: identification and purification of a DNA binding protein factor that promotes termination. *Cell* 58, 391-397.
- Kuhlbrandt, W. (2014). The Resolution Revolution. *Science* 343, 1443-1444.
- Kukat, C., and Larsson, N.-G. (2013). mtDNA makes a U-turn for the mitochondrial nucleoid. *Trends in Cell Biology* 23, 457-463.
- Kukat, C., Wurm, C.A., Spahr, H., Falkenberg, M., Larsson, N.G., and Jakobs, S. (2011). Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. *Proc Natl Acad Sci U S A* 108, 13534-13539.
- Kühl, I., Miranda, M., Posse, V., Milenkovic, D., Mourier, A., Siira, S.J., Bonekamp, N.A., Neumann, U., Filipovska, A., Polosa, P.L., *et al.* (2016). POLRMT regulates the switch between replication primer formation and gene expression of mammalian mtDNA. *Sci Adv* 2, e1600963.
- Kühlbrandt, W. (2015). Structure and function of mitochondrial membrane protein complexes. *BMC Biol* 13, 89.
- Lee, A.Y., Chen, Y.D., Chang, Y.Y., Lin, Y.C., Chang, C.F., Huang, S.J., Wu, S.H., and Hsu, C.H. (2014). Structural basis for DNA-mediated allosteric regulation facilitated by the AAA+ module of Lon protease. *Acta Crystallogr D Biol Crystallogr* 70, 218-230.
- Lee, J.H., Ancona, V., and Zhao, Y. (2018). Lon protease modulates virulence traits in *Erwinia amylovora* by direct monitoring of major regulators and indirectly through the Rcs and Gac-Csr regulatory systems. *Mol Plant Pathol* 19, 827-840.
- Lee, K.W., Okot-Kotber, C., LaComb, J.F., and Bogenhagen, D.F. (2013). Mitochondrial ribosomal RNA (rRNA) methyltransferase family members are positioned to modify nascent rRNA in foci near the mitochondrial DNA nucleoid. *J Biol Chem* 288, 31386-31399.
- Lee, S., Augustin, S., Tatsuta, T., Gerdes, F., Langer, T., and Tsai, F.T.F. (2011). Electron Cryomicroscopy Structure of a Membrane-anchored Mitochondrial AAA Protease. *The Journal of Biological Chemistry* 286, 4404-4411.
- Lehninger, A., L (1965). *The Mitochondrion-Molecular Basis of Structure and Function* (New York: W. A. Benjamin).
- Li, H., Wang, J., Wilhelmsson, H., Hansson, A., Thorén, P., Duffy, J., Rustin, P., and Larsson, N.-G. (2000). Genetic modification of survival in tissue-specific knockout mice with mitochondrial cardiomyopathy. *Proceedings of the National Academy of Sciences* 97, 3467.
- Li, J.K., Liao, J.H., Li, H., Kuo, C.I., Huang, K.F., Yang, L.W., Wu, S.H., and Chang, C.I. (2013). The N-terminal substrate-recognition domain of a LonC protease exhibits structural and functional similarity to cytosolic chaperones. *Acta Crystallogr D Biol Crystallogr* 69, 1789-1797.
- Li, M., Gustchina, A., Rasulova, F.S., Melnikov, E.E., Maurizi, M.R., Rotanova, T.V., Dauter, Z., and Wlodawer, A. (2010). Structure of the N-terminal fragment of *Escherichia coli* Lon protease. *Acta Crystallogr D Biol Crystallogr* 66, 865-873.
- Li, M., Rasulova, F., Melnikov, E.E., Rotanova, T.V., Gustchina, A., Maurizi, M.R., and Wlodawer, A. (2005). Crystal structure of the N-terminal domain of *E. coli* Lon protease. *Protein Sci* 14, 2895-2900.

- Liao, J.H., Ihara, K., Kuo, C.I., Huang, K.F., Wakatsuki, S., Wu, S.H., and Chang, C.I. (2013). Structures of an ATP-independent Lon-like protease and its complexes with covalent inhibitors. *Acta Crystallogr D Biol Crystallogr* *69*, 1395-1402.
- Lightowlers, R.N., and Chrzanowska-Lightowlers, Z.M. (2008). PPR (pentatricopeptide repeat) proteins in mammals: important aids to mitochondrial gene expression. *Biochem J* *416*, e5-6.
- Lin, C.C., Su, S.C., Su, M.Y., Liang, P.H., Feng, C.C., Wu, S.H., and Chang, C.I. (2016). Structural Insights into the Allosteric Operation of the Lon AAA+ Protease. *Structure* *24*, 667-675.
- Litonin, D., Sologub, M., Shi, Y., Savkina, M., Anikin, M., Falkenberg, M., Gustafsson, C.M., and Temiakov, D. (2010). Human mitochondrial transcription revisited: only TFAM and TFB2M are required for transcription of the mitochondrial genes in vitro. *J Biol Chem* *285*, 18129-18133.
- Liu, B., and Steitz, T.A. (2017). Structural insights into NusG regulating transcription elongation. *Nucleic Acids Res* *45*, 968-974.
- Liu, J., Zeinert, R., Francis, L., and Chien, P. (2018). Lon recognition of the replication initiator DnaA requires a bipartite degron. *Mol Microbiol*.
- Liu, L., Sanosaka, M., Lei, S., Bestwick, M.L., Frey, J.H., Surovtseva, Y.V., Shadel, G.S., and Cooper, M.P. (2011). LRP130 protein remodels mitochondria and stimulates fatty acid oxidation. *J Biol Chem* *286*, 41253-41264.
- Lockshon, D., Zweifel, S.G., Freeman-Cook, L.L., Lorimer, H.E., Brewer, B.J., and Fangman, W.L. (1995). A role for recombination junctions in the segregation of mitochondrial DNA in yeast. *Cell* *81*, 947-955.
- Lodeiro, M.F., Uchida, A., Bestwick, M., Moustafa, I.M., Arnold, J.J., Shadel, G.S., and Cameron, C.E. (2012). Transcription from the second heavy-strand promoter of human mtDNA is repressed by transcription factor A in vitro. *Proceedings of the National Academy of Sciences* *109*, 6513.
- Lott, M.T., Leipzig, J.N., Derbeneva, O., Xie, H.M., Chalkia, D., Sarmady, M., Procaccio, V., and Wallace, D.C. (2013). mtDNA Variation and Analysis Using Mitomap and Mitomaster. *Curr Protoc Bioinformatics* *44*, 1.23.21-26.
- Lu, B. (2017). Mitochondrial Lon Protease and Cancer. *Adv Exp Med Biol* *1038*, 173-182.
- Lu, B., Lee, J., Nie, X., Li, M., Morozov, Y.I., Venkatesh, S., Bogenhagen, D.F., Temiakov, D., and Suzuki, C.K. (2013). Phosphorylation of human TFAM in mitochondria impairs DNA binding and promotes degradation by the AAA+ Lon protease. *Mol Cell* *49*, 121-132.
- Lyakhov, D.L., He, B., Zhang, X., Studier, F.W., Dunn, J.J., and McAllister, W.T. (1998). Pausing and termination by bacteriophage T7 RNA polymerase. Edited by M. Gottesman. *Journal of Molecular Biology* *280*, 201-213.
- Mangus, D.A., Jang, S.H., and Jaehning, J.A. (1994). Release of the yeast mitochondrial RNA polymerase specificity factor from transcription complexes. *J Biol Chem* *269*, 26568-26574.
- Mannella, C.A. (1992). The 'ins' and 'outs' of mitochondrial membrane channels. *Trends Biochem Sci* *17*, 315-320.

- Martin, M., Cho, J., Cesare, A.J., Griffith, J.D., and Attardi, G. (2005). Termination Factor-Mediated DNA Loop between Termination and Initiation Sites Drives Mitochondrial rRNA Synthesis. *Cell* *123*, 1227-1240.
- Martinez-Rucobo, F.W., Sainsbury, S., Cheung, A.C., and Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J* *30*, 1302-1310.
- Masters, B.S., Stohl, L.L., and Clayton, D.A. (1987). Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. *Cell* *51*, 89-99.
- Matsushima, Y., Goto, Y., and Kaguni, L.S. (2010). Mitochondrial Lon protease regulates mitochondrial DNA copy number and transcription by selective degradation of mitochondrial transcription factor A (TFAM). *Proc Natl Acad Sci U S A* *107*, 18410-18415.
- McCulloch, V., Seidel-Rogol, B.L., and Shadel, G.S. (2002). A human mitochondrial transcription factor is related to RNA adenine methyltransferases and binds S-adenosylmethionine. *Mol Cell Biol* *22*, 1116-1125.
- McCulloch, V., and Shadel, G.S. (2003). Human mitochondrial transcription factor B1 interacts with the C-terminal activation region of h-mtTFA and stimulates transcription independently of its RNA methyltransferase activity. *Mol Cell Biol* *23*, 5816-5824.
- Meisinger, C., Sickmann, A., and Pfanner, N. (2008). The mitochondrial proteome: from inventory to function. *Cell* *134*, 22-24.
- Merante, F., Petrova-Benedict, R., MacKay, N., Mitchell, G., Lambert, M., Morin, C., De Braekeleer, M., Laframboise, R., Gagné, R., and Robinson, B.H. (1993). A biochemically distinct form of cytochrome oxidase (COX) deficiency in the Saguenay-Lac-Saint-Jean region of Quebec. *Am J Hum Genet* *53*, 481-487.
- Metodiev, M.D., Lesko, N., Park, C.B., Cámara, Y., Shi, Y., Wibom, R., Hultenby, K., Gustafsson, C.M., and Larsson, N.G. (2009). Methylation of 12S rRNA is necessary for in vivo stability of the small subunit of the mammalian mitochondrial ribosome. *Cell Metab* *9*, 386-397.
- Milenkovic, D., Matic, S., Köhl, I., Ruzzenente, B., Freyer, C., Jemt, E., Park, C.B., Falkenberg, M., and Larsson, N.-G. (2013). TWINKLE is an essential mitochondrial helicase required for synthesis of nascent D-loop strands and complete mtDNA replication. *Human Molecular Genetics* *22*, 1983-1993.
- Minczuk, M., He, J., Duch, A.M., Ettema, T.J., Chlebowska, A., Dzionek, K., Nijtmans, L.G., Huynen, M.A., and Holt, I.J. (2011). TEFM (c17orf42) is necessary for transcription of human mtDNA. *Nucleic Acids Res* *39*, 4284-4299.
- Miralles, M., Garcia, M., Tejero, M., Bosch, A., and Chillón, M. (2014). Production of chimeric adenovirus. *Methods Mol Biol* *1089*, 231-243.
- Monroe, N., Han, H., Shen, P.S., Sundquist, W.I., and Hill, C.P. (2017). Structural basis of protein translocation by the Vps4-Vta1 AAA ATPase. *eLife* *6*, e24487.
- Montoya, J., Christianson, T., Levens, D., Rabinowitz, M., and Attardi, G. (1982). Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. *Proc Natl Acad Sci U S A* *79*, 7195-7199.
- Morozov, Y.I., Agaronyan, K., Cheung, A.C., Anikin, M., Cramer, P., and Temiakov, D. (2014). A novel intermediate in transcription initiation by human mitochondrial RNA polymerase. *Nucleic Acids Res* *42*, 3884-3893.



- Morozov, Y.I., Parshin, A.V., Agaronyan, K., Cheung, A.C., Anikin, M., Cramer, P., and Temiakov, D. (2015). A model for transcription initiation in human mitochondria. *Nucleic Acids Res* 43, 3726-3735.
- Morozov, Y.I., and Temiakov, D. (2016). Human Mitochondrial Transcription Initiation Complexes Have Similar Topology on the Light and Heavy Strand Promoters. *J Biol Chem* 291, 13432-13435.
- Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the whole story. *Current Opinion in Structural Biology* 13, 31-39.
- Münch, C., and Harper, J.W. (2016). Mitochondrial unfolded protein response controls matrix pre-RNA processing and translation. *Nature* 534, 710.
- Nagaike, T., Suzuki, T., Tomari, Y., Takemoto-Hori, C., Negayama, F., Watanabe, K., and Ueda, T. (2001). Identification and Characterization of Mammalian Mitochondrial tRNA nucleotidyltransferases. *Journal of Biological Chemistry* 276, 40041-40049.
- Nass, M.M., and Nass, S. (1963). Intramitochondrial fibers with DNA characteristics. I. Fixation and electron staining reactions. *J Cell Biol* 19, 593-611.
- Nes, W.D. (1999). Protein Purification: Principles, High-Resolution Methods, and Applications. Second Edition Edited by Jan-Christer Janson and Lars Rydén (Uppsala University). John Wiley & Sons: New York. 1988. 695 pp. ISBN 0-471-18626-0. *Journal of the American Chemical Society* 121, 1625-1625.
- Neupert, W., and Herrmann, J.M. (2007). Translocation of Proteins into Mitochondria. *Annual Review of Biochemistry* 76, 723-749.
- Ngo, H.B., Kaiser, J.T., and Chan, D.C. (2011). The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA. *Nat Struct Mol Biol* 18, 1290-1296.
- Ngo, H.B., Lovely, G.A., Phillips, R., and Chan, D.C. (2014). Distinct structural features of TFAM drive mitochondrial DNA packaging versus transcriptional activation. *Nat Commun* 5, 3077.
- Ngo, J.K., and Davies, K.J. (2007). Importance of the lon protease in mitochondrial maintenance and the significance of declining lon in aging. *Ann N Y Acad Sci* 1119, 78-87.
- Ojala, D., Montoya, J., and Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290, 470.
- Olivares, A.O., Baker, T.A., and Sauer, R.T. (2018). Mechanical Protein Unfolding and Degradation. *Annu Rev Physiol* 80, 413-429.
- Ondrovicová, G., Liu, T., Singh, K., Tian, B., Li, H., Gakh, O., Perecko, D., Janata, J., Granot, Z., Orly, J., *et al.* (2005). Cleavage site selection within a folded substrate by the ATP-dependent lon protease. *J Biol Chem* 280, 25103-25110.
- Palade, G.E. (1952). The fine structure of mitochondria. *Anat Rec* 114, 427-451.
- Palmer, J.M., and Hall, D.O. (1972). The mitochondrial membrane system. *Prog Biophys Mol Biol* 24, 125-176.
- Paratkar, S., Deshpande, A.P., Tang, G.Q., and Patel, S.S. (2011). The N-terminal domain of the yeast mitochondrial RNA polymerase regulates multiple steps of transcription. *J Biol Chem* 286, 16109-16120.

- Pardue, M.L., Fostel, J.M., and Cech, T.R. (1984). DNA-protein interactions in the *Drosophila virilis* mitochondrial chromosome. *Nucleic Acids Research* 12, 1991-1999.
- Peter, B., Waddington, C.L., Oláhová, M., Sommerville, E.W., Hopton, S., Pyle, A., Champion, M., Ohlson, M., Siibak, T., Chrzanowska-Lightowlers, Z.M.A., *et al.* (2018). Defective mitochondrial protease LonP1 can cause classical mitochondrial disease. *Human Molecular Genetics* 27, 1743-1753.
- Pham, X.H., Farge, G., Shi, Y., Gaspari, M., Gustafsson, C.M., and Falkenberg, M. (2006). Conserved sequence box II directs transcription termination and primer formation in mitochondria. *J Biol Chem* 281, 24647-24652.
- Pinti, M., Gibellini, L., Liu, Y., Xu, S., Lu, B., and Cossarizza, A. (2015). Mitochondrial Lon protease at the crossroads of oxidative stress, ageing and cancer. *Cell Mol Life Sci* 72, 4807-4824.
- Pinti, M., Gibellini, L., Nasi, M., De Biasi, S., Bortolotti, C.A., Iannone, A., and Cossarizza, A. (2016a). Emerging role of Lon protease as a master regulator of mitochondrial functions. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1857, 1300-1306.
- Pinti, M., Gibellini, L., Nasi, M., De Biasi, S., Bortolotti, C.A., Iannone, A., and Cossarizza, A. (2016b). Emerging role of Lon protease as a master regulator of mitochondrial functions. *Biochim Biophys Acta* 1857, 1300-1306.
- Polo, M., Alegre, F., Moragrega, A.B., Gibellini, L., Marti-Rodrigo, A., Blas-Garcia, A., Esplugues, J.V., and Apostolova, N. (2017). Lon protease: a novel mitochondrial matrix protein in the interconnection between drug-induced mitochondrial dysfunction and endoplasmic reticulum stress. *Br J Pharmacol* 174, 4409-4429.
- Posse, V., and Gustafsson, C.M. (2017). Human Mitochondrial Transcription Factor B2 Is Required for Promoter Melting during Initiation of Transcription. *J Biol Chem* 292, 2637-2645.
- Posse, V., Hoberg, E., Dierckx, A., Shahzad, S., Koolmeister, C., Larsson, N.G., Wilhelmsson, L.M., Hällberg, B.M., and Gustafsson, C.M. (2014). The amino terminal extension of mammalian mitochondrial RNA polymerase ensures promoter specific transcription initiation. *Nucleic Acids Res* 42, 3638-3647.
- Puchades, C., Rampello, A.J., Shin, M., Giuliano, C.J., Wiseman, R.L., Glynn, S.E., and Lander, G.C. (2017). Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing. *Science* 358.
- Puebla-Osorio, N., Lacey, D.B., Alt, F.W., and Zhu, C. (2006). Early Embryonic Lethality Due to Targeted Inactivation of DNA Ligase III. *Molecular and Cellular Biology* 26, 3935.
- Quirós, P.M., Bárcena, C., and López-Otín, C. (2014a). Lon protease: A key enzyme controlling mitochondrial bioenergetics in cancer. *Mol Cell Oncol* 1, e968505.
- Quirós, Pedro M., Español, Y., Acín-Pérez, R., Rodríguez, F., Bárcena, C., Watanabe, K., Calvo, E., Loureiro, M., Fernández-García, M.S., Fueyo, A., *et al.* (2014b). ATP-Dependent Lon Protease Controls Tumor Bioenergetics by Reprogramming Mitochondrial Activity. *Cell Reports* 8, 542-556.
- Quirós, P.M., Langer, T., and López-Otín, C. (2015). New roles for mitochondrial proteases in health, ageing and disease. *Nature Reviews Molecular Cell Biology* 16, 345.
- Reinhard, L., Sridhara, S., and Hällberg, B.M. (2015). Structure of the nuclease subunit of human mitochondrial RNase P. *Nucleic Acids Res* 43, 5664-5672.

- Rhodes, G. (2006a). Chapter 2 - An Overview of Protein Crystallography. In *Crystallography Made Crystal Clear* (Third Edition), G. Rhodes, ed. (Burlington: Academic Press), pp. 7-30.
- Rhodes, G. (2006b). Chapter 3 - Protein Crystals. In *Crystallography Made Crystal Clear* (Third Edition), G. Rhodes, ed. (Burlington: Academic Press), pp. 31-47.
- Ringel, R., Sologub, M., Morozov, Y.I., Litonin, D., Cramer, P., and Temiakov, D. (2011). Structure of human mitochondrial RNA polymerase. *Nature* *478*, 269-273.
- Ripstein, Z.A., Huang, R., Augustyniak, R., Kay, L.E., and Rubinstein, J.L. (2017). Structure of a AAA+ unfoldase in the process of unfolding substrate. *eLife* *6*, e25754.
- Robberson, D.L., Kasamatsu, H., and Vinograd, J. (1972). Replication of mitochondrial DNA. Circular replicative intermediates in mouse L cells. *Proc Natl Acad Sci U S A* *69*, 737-741.
- Rodeheffer, M.S., Boone, B.E., Bryan, A.C., and Shadel, G.S. (2001). Nam1p, a protein involved in RNA processing and translation, is coupled to transcription through an interaction with yeast mitochondrial RNA polymerase. *J Biol Chem* *276*, 8616-8622.
- Rosano, G.L., and Ceccarelli, E.A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology* *5*, 172.
- Rotanova, T.V., Botos, I., Melnikov, E.E., Rasulova, F., Gustchina, A., Maurizi, M.R., and Wlodawer, A. (2006). Slicing a protease: Structural features of the ATP-dependent Lon proteases gleaned from investigations of isolated domains. *Protein Science : A Publication of the Protein Society* *15*, 1815-1828.
- Rubio-Cosials, A., Sidow, J.F., Jiménez-Menéndez, N., Fernández-Millán, P., Montoya, J., Jacobs, H.T., Coll, M., Bernadó, P., and Solà, M. (2011). Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter. *Nat Struct Mol Biol* *18*, 1281-1289.
- Ruzzenente, B., Metodiev, M.D., Wredenberg, A., Bratic, A., Park, C.B., Cámara, Y., Milenkovic, D., Zickermann, V., Wibom, R., Hultenby, K., *et al.* (2012). LRPPRC is necessary for polyadenylation and coordination of translation of mitochondrial mRNAs. *EMBO J* *31*, 443-456.
- Sanchez, M.I., Mercer, T.R., Davies, S.M., Shearwood, A.M., Nygård, K.K., Richman, T.R., Mattick, J.S., Rackham, O., and Filipovska, A. (2011). RNA processing in human mitochondria. *Cell Cycle* *10*, 2904-2916.
- Sauer, R.T., and Baker, T.A. (2011). AAA+ proteases: ATP-fueled machines of protein destruction. *Annu Rev Biochem* *80*, 587-612.
- Scheduler, I., E (2008). *Mitochondria*, Second edn (New Jersey: John Wiley & Sons, Inc., Publication).
- Schowen, R.L. (1993). Principles of biochemistry 2nd ed. (Lehninger, Albert L.; Nelson, David L.; Cox, Michael M.). *Journal of Chemical Education* *70*, A223.
- Schwinghammer, K., Cheung, A.C., Morozov, Y.I., Agaronyan, K., Temiakov, D., and Cramer, P. (2013). Structure of human mitochondrial RNA polymerase elongation complex. *Nat Struct Mol Biol* *20*, 1298-1303.
- Shadel, G.S., and Clayton, D.A. (1997). Mitochondrial DNA maintenance in vertebrates. *Annu Rev Biochem* *66*, 409-435.

- Shah, I.M., and Wolf, R.E. (2006). Sequence requirements for Lon-dependent degradation of the *Escherichia coli* transcription activator SoxS: identification of the SoxS residues critical to proteolysis and specific inhibition of in vitro degradation by a peptide comprised of the N-terminal 21 amino acid residues. *J Mol Biol* 357, 718-731.
- Shang, J., and Clayton, D.A. (1994). Human mitochondrial transcription termination exhibits RNA polymerase independence and biased bipolarity in vitro. *J Biol Chem* 269, 29112-29120.
- Shaw, J.M., and Nunnari, J. (2002). Mitochondrial dynamics and division in budding yeast. *Trends Cell Biol* 12, 178-184.
- Shi, Y., Posse, V., Zhu, X., Hyvärinen, A.K., Jacobs, H.T., Falkenberg, M., and Gustafsson, C.M. (2016). Mitochondrial transcription termination factor 1 directs polar replication fork pausing. *Nucleic Acids Res* 44, 5732-5742.
- Shokolenko, I.N., and Alexeyev, M.F. (2017). Mitochondrial transcription in mammalian cells. *Frontiers in bioscience (Landmark edition)* 22, 835-853.
- Shutt, T.E., and Gray, M.W. (2006). Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends Genet* 22, 90-95.
- Silva, J.P., Köhler, M., Graff, C., Oldfors, A., Magnuson, M.A., Berggren, P.-O., and Larsson, N.-G. (2000). Impaired insulin secretion and  $\beta$ -cell loss in tissue-specific knockout mice with mitochondrial diabetes. *Nature Genetics* 26, 336.
- Sjostrand, F.S. (1953). The ultrastructure of the innersegments of the retinal rods of the guinea pig eye as revealed by electron microscopy. *J Cell Comp Physiol* 42, 45-70.
- Small, I.D., and Peeters, N. (2000). The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* 25, 46-47.
- Smith, C.K., Baker, T.A., and Sauer, R.T. (1999). Lon and Clp family proteases and chaperones share homologous substrate-recognition domains. *Proc Natl Acad Sci U S A* 96, 6678-6682.
- Sologub, M., Litonin, D., Anikin, M., Mustaev, A., and Temiakov, D. (2009a). TFB2 is a transient component of the catalytic site of the human mitochondrial RNA polymerase. *Cell* 139, 934-944.
- Sologub, M.I.u., Kochetkov, S.N., and Temiakov, D.E. (2009b). [Transcription and its regulation in mammalian and human mitochondria]. *Mol Biol (Mosk)* 43, 215-229.
- Sousa, R., Chung, Y.J., Rose, J.P., and Wang, B.-C. (1993). Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* 364, 593.
- Steitz, T.A. (2009). The structural changes of T7 RNA polymerase from transcription initiation to elongation. *Current opinion in structural biology* 19, 683.
- Stewart, L., Redinbo, M.R., Qiu, X., Hol, W.G.J., and Champoux, J.J. (1998). A Model for the Mechanism of Human Topoisomerase I. *Science* 279, 1534.
- Stiburek, L., Cesnekova, J., Kostkova, O., Fornuskova, D., Vinsova, K., Wenchich, L., Houstek, J., Zeman, J., and Fox, T.D. (2012). YME1L controls the accumulation of respiratory chain subunits and is required for apoptotic resistance, cristae morphogenesis, and cell proliferation. *Molecular Biology of the Cell* 23, 1010-1023.
- Strauss, K.A., Jinks, R.N., Puffenberger, E.G., Venkatesh, S., Singh, K., Cheng, I., Mikita, N., Thilagavathi, J., Lee, J., Sarafianos, S., *et al.* (2015). CODAS syndrome is associated with

- mutations of LONP1, encoding mitochondrial AAA+ Lon protease. *Am J Hum Genet* 96, 121-135.
- Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* 185, 60-89.
- Sun, J., and Li, H. (2010). Chapter Ten - How to Operate a Cryo-Electron Microscope. In *Methods in Enzymology*, G.J. Jensen, ed. (Academic Press), pp. 231-249.
- Surovtseva, Y.V., Shutt, T.E., Cotney, J., Cimen, H., Chen, S.Y., Koc, E.C., and Shadel, G.S. (2011). Mitochondrial ribosomal protein L12 selectively associates with human mitochondrial RNA polymerase to activate transcription. *Proc Natl Acad Sci U S A* 108, 17921-17926.
- Sörensen, L., Ekstrand, M., Silva, J.P., Lindqvist, E., Xu, B., Rustin, P., Olson, L., and Larsson, N.G. (2001). Late-onset corticohippocampal neurodepletion attributable to catastrophic failure of oxidative phosphorylation in MILON mice. *J Neurosci* 21, 8082-8090.
- Tahirov, T.H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W.T., Vassilyev, D.G., and Yokoyama, S. (2002). Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature* 420, 43.
- Takamatsu, C., Umeda, S., Ohsato, T., Ohno, T., Abe, Y., Fukuoh, A., Shinagawa, H., Hamasaki, N., and Kang, D. (2002). Regulation of mitochondrial D-loops by transcription factor A and single-stranded DNA-binding protein. *EMBO Rep* 3, 451-456.
- Tang, G.-Q., Roy, R., Bandwar, R.P., Ha, T., and Patel, S.S. (2009). Real-time observation of the transition from transcription initiation to elongation of the RNA polymerase. *Proceedings of the National Academy of Sciences* 106, 22175.
- Teng, H., Wu, B., Zhao, K., Yang, G., Wu, L., and Wang, R. (2013). Oxygen-sensitive mitochondrial accumulation of cystathionine  $\beta$ -synthase mediated by Lon protease. *Proc Natl Acad Sci U S A* 110, 12679-12684.
- Terzioglu, M., Ruzzenente, B., Harmel, J., Mourier, A., Jemt, E., López, M.D., Kukat, C., Stewart, J.B., Wibom, R., Meharg, C., *et al.* (2013). MTERF1 binds mtDNA to prevent transcriptional interference at the light-strand promoter but is dispensable for rRNA gene transcription regulation. *Cell Metab* 17, 618-626.
- Tian, Q., Li, T., Hou, W., Zheng, J., Schrum, L.W., and Bonkovsky, H.L. (2011). Lon peptidase 1 (LONP1)-dependent breakdown of mitochondrial 5-aminolevulinic acid synthase protein by heme in human liver cells. *J Biol Chem* 286, 26424-26430.
- Tiranti, V., Rocchi, M., DiDonato, S., and Zeviani, M. (1993). Cloning of human and rat cDNAs encoding the mitochondrial single-stranded DNA-binding protein (SSB). *Gene* 126, 219-225.
- Tiranti, V., Savoia, A., Forti, F., D'Apolito, M.F., Centra, M., Rocchi, M., and Zeviani, M. (1997). Identification of the gene encoding the human mitochondrial RNA polymerase (h-mtRPOL) by cyberscreening of the Expressed Sequence Tags database. *Hum Mol Genet* 6, 615-625.
- Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., and Yoshikawa, S. (1996). The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 Å. *Science* 272, 1136-1144.
- Tunitskaya, V.L., and Kochetkov, S.N. (2002). Structural-Functional Analysis of Bacteriophage T7 RNA Polymerase. *Biochemistry (Moscow)* 67, 1124-1135.

- Van Dyck, L., and Langer, T. (1999). ATP-dependent proteases controlling mitochondrial function in the yeast *Saccharomyces cerevisiae*. *Cell Mol Life Sci* 56, 825-842.
- Venkatesh, S., Lee, J., Singh, K., Lee, I., and Suzuki, C.K. (2012). Multitasking in the mitochondrion by the ATP-dependent Lon protease. *Biochim Biophys Acta* 1823, 56-66.
- Vieux, E.F., Wohlever, M.L., Chen, J.Z., Sauer, R.T., and Baker, T.A. (2013). Distinct quaternary structures of the AAA+ Lon protease control substrate degradation. *Proc Natl Acad Sci U S A* 110, E2002-2008.
- Voos, W. (2013). Chaperone-protease networks in mitochondrial protein homeostasis. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1833, 388-399.
- Walther, D.M., and Rapaport, D. (2009). Biogenesis of mitochondrial outer membrane proteins. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1793, 42-51.
- Wang, H. (2015). Cryo- electron microscopy for structural biology: current status and future perspectives. *Science China Life Sciences* 58, 750-756.
- Wang, J., Wilhelmsson, H., Graff, C., Li, H., Oldfors, A., Rustin, P., Br uning, J.C., Kahn, C.R., Clayton, D.A., Barsh, G.S., *et al.* (1999). Dilated cardiomyopathy and atrioventricular conduction blocks induced by heart-specific inactivation of mitochondrial DNA gene expression. *Nat Genet* 21, 133-137.
- Wang, Y., and Shadel, G.S. (1999). Stability of the mitochondrial genome requires an amino-terminal domain of yeast mitochondrial RNA polymerase. *Proc Natl Acad Sci U S A* 96, 8046-8051.
- Wang, Z., Cotney, J., and Shadel, G.S. (2007). Human mitochondrial ribosomal protein MRPL12 interacts directly with mitochondrial RNA polymerase to modulate mitochondrial gene expression. *J Biol Chem* 282, 12610-12618.
- Wanrooij, S., Fust , J.M., Farge, G., Shi, Y., Gustafsson, C.M., and Falkenberg, M. (2008). Human mitochondrial RNA polymerase primes lagging-strand DNA synthesis in vitro. *Proc Natl Acad Sci U S A* 105, 11122-11127.
- Wendler, P., Ciniawsky, S., Kock, M., and Kube, S. (2012). Structure and function of the AAA+ nucleotide binding pocket. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1823, 2-14.
- Wheeler, M.J., Russi, S., Bowler, M.G., and Bowler, M.W. (2012). Measurement of the equilibrium relative humidity for common precipitant concentrations: facilitating controlled dehydration experiments. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* 68, 111-114.
- Wohlever, M.L., Baker, T.A., and Sauer, R.T. (2014). Roles of the N domain of the AAA+ Lon protease in substrate recognition, allosteric regulation and chaperone activity. *Mol Microbiol* 91, 66-78.
- Wong, T.W., and Clayton, D.A. (1985). In vitro replication of human mitochondrial DNA: accurate initiation at the origin of light-strand synthesis. *Cell* 42, 951-958.
- Wredenberg, A., Wibom, R., Wilhelmsson, H., Graff, C., Wiener, H.H., Burden, S.J., Oldfors, A., Westerblad, H., and Larsson, N.-G. (2002). Increased mitochondrial mass in mitochondrial myopathy mice. *Proceedings of the National Academy of Sciences* 99, 15066.
- Wyatt, H.D.M., and West, S.C. (2014). Holliday Junction Resolvases. *Cold Spring Harbor Perspectives in Biology* 6, a023192.

Yakubovskaya, E., Mejia, E., Byrnes, J., Hambardjieva, E., and Garcia-Diaz, M. (2010). Helix unwinding and base flipping enable human MTERF1 to terminate mitochondrial transcription. *Cell* 141, 982-993.

Yin, Y.W., and Steitz, T.A. (2002). Structural Basis for the Transition from Initiation to Elongation Transcription in T7 RNA Polymerase. *Science* 298, 1387.

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E., and Ebright, R.H. (2012). Structural Basis of Transcription Initiation. *Science* 338, 1076.

Zhao, Z., De-Donatis, G.M., Schwartz, C., Fang, H., Li, J., and Guo, P. (2016). An Arginine Finger Regulates the Sequential Action of Asymmetrical Hexameric ATPase in the Double-Stranded DNA Translocation Motor. *Molecular and Cellular Biology* 36, 2514.

Zurita Rendón, O., and Shoubridge, E.A. (2012). Early complex I assembly defects result in rapid turnover of the ND1 subunit. *Human Molecular Genetics* 21, 3815-3824.