

Molecular insights into mitochondrial transcription and its role in DNA replication

Viktor Posse

Department of
Medical Biochemistry and Cell Biology
Institute of Biomedicine
Sahlgrenska Academy
University of Gothenburg

Gothenburg, Sweden, 2017



UNIVERSITY OF
GOTHENBURG

Molecular insights into
mitochondrial transcription
and its role in DNA replication

© 2016 Viktor Posse
viktor.posse@medkem.gu.se

ISBN 978-91-629-0024-3 (PRINT)
ISBN 978-91-629-0023-6 (PDF)
<http://hdl.handle.net/2077/48657>

Printed in Gothenburg, Sweden 2016
Ineko AB

Abstract

The mitochondrion is an organelle of the eukaryotic cell responsible for the production of most of the cellular energy-carrying molecule adenosine triphosphate (ATP), through the process of oxidative phosphorylation. The mitochondrion contains its own genome, a small circular DNA molecule (mtDNA), encoding essential subunits of the oxidative phosphorylation system. Initiation of mitochondrial transcription involves three proteins, the mitochondrial RNA polymerase, POLRMT, and its two transcription factors, TFAM and TFB2M. Even though the process of transcription has been reconstituted *in vitro*, a full molecular understanding is still missing. Initiation of mitochondrial DNA replication is believed to be primed by transcription prematurely terminated at a sequence known as CSBII. The mechanisms of replication initiation have however not been fully defined. In this thesis we have studied transcription and replication of mtDNA.

In the first part of this thesis we demonstrate that the transcription initiation machinery is recruited in discrete steps. Furthermore, we find that a large domain of POLRMT known as the N-terminal extension is dispensable for transcription initiation, and instead functions in suppressing initiation events from non-promoter DNA. Additionally we demonstrate that TFB2M is the last factor that is recruited to the initiation complex and that it induces melting of the mitochondrial promoters.

In this thesis we also demonstrate that POLRMT is a non-processive polymerase that needs the presence of the elongation factor TEFM for processive transcription. TEFM increases the affinity of POLRMT for an elongation-like RNA-DNA template and decreases the probability of premature transcription termination. Our data also suggest that TEFM might be of importance for mitochondrial replication initiation, since it affects termination at CSBII.

In the last part of this thesis we study the RNA-DNA hybrids (R-loops) that can be formed by the CSBII terminated transcript. We characterize these R-loops and demonstrate that they can be processed by RNaseH1 to form replicative primers that can be used by the mitochondrial replication machinery.

Keywords: Mitochondrion, mtDNA, transcription, DNA replication

ISBN: 978-91-629-0024-3 (PRINT), 978-91-629-0023-6 (PDF)

Sammanfattning på svenska

Ritningen för hur en mänsklig cell ska byggas upp finns kodad i gener i cellens DNA. Ritningen i generna beskriver hur proteiner ska byggas upp och dessa proteiner genomför sedan de funktioner som krävs för att cellen ska fungera. Merparten av cellens DNA finns i cellkärnans kromosomer. Dock finns även en liten DNA-molekyl i en av cellens enheter som kallas mitokondrien. Mitokondrien brukar kallas för cellens kraftverk då denna enhet ansvarar för att ta vara på energin i maten vi äter. Både generna i cellkärnans kromosomer och mitokondriens DNA är livsviktiga för att cellen ska överleva. För att genernas ritningar ska kunna användas krävs att informationen omvandlas från DNA till protein. Det första steget i denna process kallas för transkription. En annan livsviktig process som involverar DNA är den så kallade DNA replikationen, vilken går ut på att kopiera en DNA-molekyl för att bilda två uppsättningar av densamma. Detta krävs för den naturliga processen när en cell ska dela sig och de två dottercellerna behöver varsin uppsättning av både kromosomerna och cellens mitokondriella DNA. I den här avhandlingen har vi studerat transkription och replikation av cellens mitokondriella DNA. Grundläggande forskning kring dessa processer är av stor vikt för att öka förståelsen för bland annat olika mitokondriella sjukdomar, normalt åldrande samt biverkningar från läkemedel mot HIV och hepatit-virus.

När vi har studerat mitokondriens transkription har vi framförallt använt metoder där vi återskapar delar av denna process i provrör. På detta sätt har vi i detalj kunnat studera hur maskineriet som sköter transkriptionen fungerar. Vi har bland annat lyckats bestämma hur olika proteiner i detta maskineri bidrar till att transkriptionen fungerar. Vi har även studerat och karaktäriserat en ny faktor, som vi visar är absolut nödvändig för lyckad transkription.

På samma sätt som med mitokondriens transkription kan vi återskapa vissa delar av DNA-replikationen i provrör. Dock har ingen tidigare lyckats återskapa den specifika uppstarten av mitokondriens DNA-replikation. I den här avhandlingen visar vi för första gången hur denna igångsättning kan återskapas i provrör. Vi visar även att utöver maskineriet för DNA-replikation har maskineriet för transkription en central roll i denna uppstartsprocess.

List of papers

This thesis is based on the following studies, referred to in the text by their Roman numerals.

- 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 CM.
Nucleic Acids Res. 2014. 42(6): 3638-47

- II. Human mitochondrial transcription factor B2 is required for promoter melting during initiation of transcription
Posse V and Gustafsson CM.
Manuscript

- III. TEFM is a potent stimulator of mitochondrial transcription elongation in vitro.
Posse V, Shahzad S, Falkenberg M, Hällberg BM and Gustafsson CM.
Nucleic Acids Res. 2015. 43(5): 2615-24

- IV. The molecular mechanism of DNA replication initiation in human mitochondria
Posse V, Al-Behadili A, Uhler JP, Falkenberg M and Gustafsson CM.
Manuscript

Related Publications

Mitochondrial transcription termination factor 1 directs polar replication fork pausing.

Shi Y*, **Posse V***, Zhu X, Hyvärinen AK, Jacobs HT, Falkenberg M and Gustafsson CM.

Nucleic Acids Res. 2016. 44(12): 5732-42.

*Equal contribution

POLRMT regulates the switch between replication primer formation and gene expression of mammalian mtDNA.

Kühl I, Miranda M, **Posse V**, Milenkovic D, Mourier A, Siira SJ, Bonekamp NA, Neumann U, Filipovska A, Polosa PL, Gustafsson CM and Larsson NG.

Sci Adv. 2016. 2(8): e1600963

Cross-strand binding of TFAM to a single mtDNA molecule forms the mitochondrial nucleoid.

Kukat C, Davies KM, Wurm CA, Spähr H, Bonekamp NA, Kühl I, Joos F, Polosa PL, Park CB, **Posse V**, Falkenberg M, Jakobs S, Kühlbrandt W and Larsson NG.

Proc Natl Acad Sci U S A. 2015. 112(36): 11288-9

Content

1. INTRODUCTION.....	1
1.1. The mitochondrion	1
1.1.1. Origin of mitochondria.....	1
1.1.2. Structure and dynamics of mitochondria.....	1
1.1.4. Metabolism	3
1.1.5. The mitochondrial genome.....	7
1.2. DNA transcription	10
1.2.1. A short introduction to DNA transcription.....	10
1.2.2. The T7 bacteriophage RNA polymerase	12
1.3. Mitochondrial transcription.....	14
1.3.1. The mitochondrial transcription machinery	15
1.3.2. Mitochondrial transcription associated proteins.....	20
1.3.3. R-loops in the CSB region.....	21
1.4. DNA replication	21
1.4.1. A short introduction to DNA replication.....	21
1.4.2. The T7 bacteriophage replisome	22
1.4.3. Initiation of DNA replication	23
1.5. Mitochondrial DNA replication	24
1.5.1. Models for replication of mitochondrial DNA.....	25
1.5.2. The mitochondrial DNA replication machinery.....	27
1.5.3. Additional mitochondrial DNA maintenance factors.....	28
1.6 Mitochondrial DNA in disease.....	29
2. SPECIFIC AIMS.....	30
3. RESULTS AND DISCUSSION	31
3.1. Paper I.....	31
3.2. Paper II	33
3.3. Paper III.....	35
3.4. Paper IV.....	36
4. CONCLUSIONS.....	40
5. FUTURE PERSPECTIVES.....	41
ACKNOWLEDGEMENTS.....	42
REFERENCES.....	43

Abbreviations

8-oxo-dG	8-oxo-2'-deoxyguanosine
A	Adenine
ADP	Adenosine diphosphate
AP-site	Apurinic/aprimidinic site
ATP	Adenosine triphosphate
ATP6 and 8	mtDNA encoded complex V subunits
bp	Base pair
BRE	B recognition element
C	Cytosine
CoA	Coenzyme A
Cox1-3	mtDNA encoded complex IV subunits
CSBI-III	Conserved sequence blocks I-III
CTD	C-terminal domain
Cytb	mtDNA encoded complex III subunit
Cyt c	Cytochrome c
DNA	Deoxyribonucleic acid
DPE	Downstream core promoter element
dATP	Deoxyadenosine triphosphate
ddCTP	(2', 3'-) Di-deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
dNTP	Deoxyribonucleoside triphosphate
dsDNA	Double stranded DNA
dTTP	Deoxythymidine triphosphate
D-loop	Displacement loop
FAD	Flavin adenine dinucleotide
FMV	Flavin mononucleotide
G4	G-quadruplex
G	Guanine
GTP	Guanosine triphosphate
HMG	High mobility group
HSP	Heavy strand promoter
H-strand	Heavy strand
IC	Initiation complex
Inr	Initiator element
kb	Kilo base pair

kDa	Kilo Dalton
LSP	Light strand promoter
L-strand	Light strand
mRNA	messenger RNA
mtDNA	Mitochondrial DNA
MTS	Mitochondrial targeting sequence
NAD	Nicotinamide adenine dinucleotide
NCR	mtDNA non-coding region
ND1-6	mtDNA encoded complex I subunits
NTD	N-terminal domain
NTE	N-terminal extension
OriH/O _H	Heavy strand origin or replication
OriL/O _L	Light strand origin or replication
OXPPOS	Oxidative phosphorylation
PEO	Progressive external ophthalmoplegia
P _i	Inorganic phosphate
PIC	Pre-initiation complex
Pol I-III	Eukaryotic nuclear RNA polymerases I-III
Poly(U)	Polyuridine
Poly(dT)	Polydeoxythymidine
PPR	Pentatricopeptide repeat
Q	Ubiquinone
R-loop	DNA hybridized RNA loop
RNA	Ribonucleic acid
RNAi	RNA interference
RNAP	RNA polymerase
rRNA	Ribosomal RNA
ssDNA	Single stranded DNA
T	Thymine
TAS	Termination associated sequence
TCA	Tricarboxylic acid
tRNA	Transfer-RNA
UTR	Untranslated region

1. Introduction

1.1. The mitochondrion

1.1.1. Origin of mitochondria

The mitochondrion is a membrane-enclosed organelle of the eukaryotic cell. A key function of the mitochondrion is to produce most of the energy-carrying molecule adenosine triphosphate (ATP) in the cell, through a process denoted as oxidative phosphorylation (OXPHOS). With the high levels of ATP generated by the OXPHOS system, this molecule functions as a cellular transporter for chemical energy, which is liberated when it breaks into adenosine diphosphate (ADP) and inorganic phosphate (P_i) (Alberts, 2008; Berg et al., 2012). A widely accepted theory states that the mitochondrion originated as an α -proteobacterium that invaded an early eukaryotic cell and kept living inside this cell in a symbiotic relationship with its host (Gray et al., 1999). One of the best evidences of this theory is the fact that mitochondria contain a separate genome (mtDNA) (Nass and Nass, 1963a; Nass and Nass, 1963b), derived from the α -proteobacterium (Andersson et al., 2003). The once free-living α -proteobacterium in time went through a heavy reduction of this genome as most genes either were transferred to the nuclear genome or lost. After millions of years of evolution, the previous α -proteobacterium contained only a remnant of a full genome and had become an organelle, the mitochondrion, not able to live outside the host cell. Despite this substantial reduction of the coding sequence, the mitochondrion still retains a small but essential fraction of the endosymbiont genome (Gray et al., 1999; Andersson et al., 2003).

1.1.2. Structure and dynamics of mitochondria

The mitochondrion is enclosed by two membranes, the inner membrane and the outer membrane. The outer membrane surrounds the organelle whereas the inner membrane is folded to form structures denoted as cristae (Figure 1) (Palade, 1952; Palade, 1953; Sjostrand, 1953). These cristae are the structures where oxidative phosphorylation to form ATP takes place (Palmer and Hall, 1972). The

inner compartment of the mitochondrion, the matrix, houses other metabolic processes such as the citric acid cycle, β -oxidation of fatty acids, and amino acid metabolism pathways (Berg et al., 2012), as well as the mitochondrial genome and the genome maintenance machinery (Gustafsson et al., 2016).

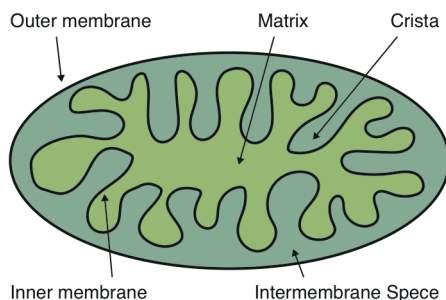


Figure 1. Two membranes termed the outer and the inner membrane surround the mitochondrion. The heavily folded inner membrane creates structures denoted as cristae. The membranes surround the inner compartment of the organelle, the matrix.

As the mitochondrion still contains its own coding genome, the mitochondrial proteome has a dual origin; nuclear and mitochondrial. The number of proteins in mammalian mitochondria is estimated to be around 1500, 99 % of which are encoded by the nuclear genome and only 13 proteins are encoded by the mitochondrial genome (Meisinger et al., 2008). All nuclear encoded factors present in the mitochondrion are translated in the cytoplasm and transported into the mitochondrion by protein import machineries present in the mitochondrial membranes (Figure 2) (Dolezal et al., 2006). Most of these proteins contain an N-terminal mitochondrial targeting sequence (MTS) that is removed upon import. The MTS can vary in length but is on average 30 amino acids and forms an amphiphilic α -helix with a high content of hydroxyl containing, hydrophobic and basic (positively charged) amino acids (Teixeira and Glaser, 2013). The mitochondrial genome is replicated and transcribed purely by nuclear encoded factors, whereas the translation machinery is of dual origin with nuclear encoded proteins and mitochondrial encoded RNA components (Figure 2) (Gustafsson et al., 2016).

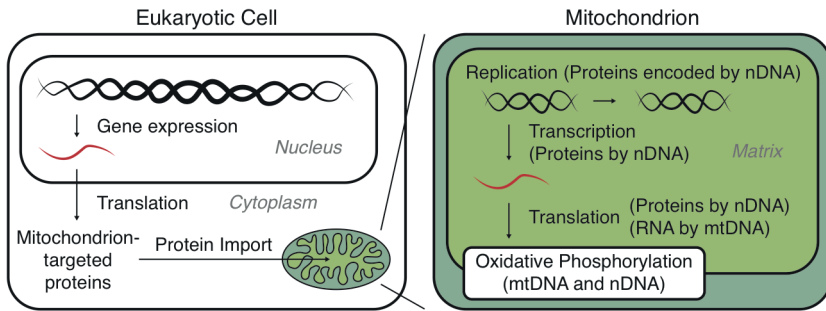


Figure 2. The proteome of the mitochondrion is of a dual origin. Left hand side: The vast majority of all mitochondrial proteins are encoded by the nuclear genome (DNA in black and mRNA in red), translated in the cytoplasm and directed to the mitochondrion with a mitochondrial targeting signal (MTS). These proteins are imported into the mitochondrion and localized to the correct compartment by specific import machineries. Right hand side: The human mitochondrial genome is replicated, transcribed and translated purely by nuclear (nDNA) encoded protein factors, whereas the RNA components of mitochondrial translation are mtDNA encoded. All proteins encoded by the human mitochondrial genome are membrane embedded subunits of the OXPHOS system.

The number of mitochondria varies per cell but generally reflects the energy demand of the specific cell type. As an example, liver cells contain in the range of 1000-2000 mitochondria (Alberts, 2008). The mitochondrion is a highly dynamic structure continuously undergoing both fission and fusion. Hence, the traditional view of this organelle as static individual structures is misleading and often the mitochondria of a cell are referred to as a mitochondrial network more than distinct units (Shaw and Nunnari, 2002; Chen and Chan, 2004). The fission and fusion machineries affect the distribution of genomes, metabolites and proteins in the mitochondrial network. These machineries also contribute to quality control by rescuing damaged mitochondria or discarding ineffective ones through autophagy (Twig et al., 2008; Youle and van der Bliek, 2012).

1.1.4. Metabolism

β -oxidation of fatty acids

Fatty acid degradation for energy extraction takes place on the mitochondrial outer membrane and in the matrix. Fatty acids are activated on the outer membrane through a reaction with ATP and coenzyme A (CoA) to form an intermediary metabolite where the fatty acid is conjugated to CoA through a

thioester bond. The activated metabolite is translocated to the matrix where continued degradation takes place. In the matrix fatty acids go through a series of oxidation, hydration and thiolysis reactions to form acetyl-CoA and a CoA-conjugated, two carbon atoms shorter, fatty acid. The shortened fatty acid can go through a new round of oxidation whereas the acetyl-CoA can enter the citric acid cycle for continued energy extraction. The degradation of fatty acids is termed β -oxidation, simply because the oxidation happens at the β -carbon, i.e. the second carbon from the carboxylic group of the fatty acid (Berg et al., 2012).

The citric acid cycle

The citric acid cycle, also known as the tricarboxylic acid (TCA) or Krebs cycle, is an enzymatic cycle harvesting the energy from acetyl-CoA. The acetyl-CoA that enters the citric acid cycle can have multiple origins. One pathway is through the oxidation of fatty acids as described above. A second pathway is from the degradation of glucose through glycolysis. The process of glycolysis takes place in the cytosol and generates two molecules of pyruvate per molecule of glucose. Pyruvate is further processed to form acetyl-CoA that can enter the citric acid cycle. This degradation of pyruvate releases one carbon atom in the form of carbon dioxide and transfers two electrons to a molecule of nicotinamide adenine dinucleotide (NAD) to form NADH (Berg et al., 2012). Acetyl-CoA enters the citric acid cycle by conjugation to oxaloacetate, which forms citric acid. Through a series of further enzymatic reactions the two carbon atoms from the acetyl group are released in the form of carbon dioxide generating one molecule of guanosine triphosphate (GTP) and eight high-energy electrons carried by three molecules of NADH and one oxidized flavin adenine dinucleotide (FAD) in the form of FADH₂. The redox potential of the electron carrying molecules of NADH and FADH₂ that are produced through the degradation of pyruvate and through the citric acid cycle can be further harvested through the OXPHOS system (Berg et al., 2012).

Oxidative phosphorylation

The OXPHOS system is composed of five membrane embedded enzyme complexes, four of which compose the electron transport chain or respiratory chain. This chain of protein complexes pass the electrons derived from food metabolites to oxygen, producing water. During this process protons (H⁺) are pumped from the matrix to the intermembrane space creating the mitochondrial membrane potential ($\Delta\Psi$). The fifth complex, the ATP synthase, uses this membrane potential to drive the synthesis of ATP (Alberts, 2008; Berg et al., 2012). The entire process is summarized in Figure 3.

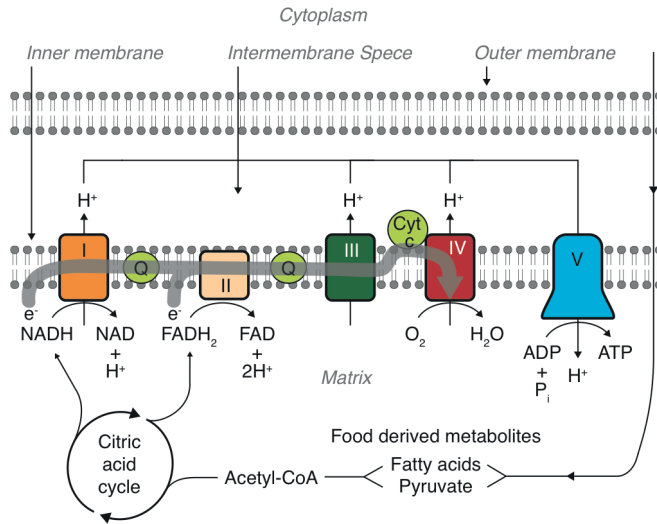


Figure 3. Food metabolites derived from fatty acids and glucose (via pyruvate) enter the citric acid cycle via conversion to acetyl-CoA. The citric acid cycle harvests high-energy electrons from acetyl-CoA to form NADH and FADH₂. These molecules enter the electron transport chain at complex I and complex II of the OXPHOS system respectively. The electrons are transported through these complexes and via ubiquinone (Q) to complex III and further via cytochrome c (Cyt c) to complex IV. In complex IV, the electrons are transferred to the final acceptor O₂, which is converted to H₂O. Complexes I, III and IV generate a proton gradient by pumping protons from the matrix to the intermembrane space. This gradient is used by complex V to drive the synthesis of ATP.

The first complex of the electron transport chain, complex I or NADH ubiquinone oxidoreductase, is the entry point for the electrons of NADH. Mammalian complex I is a 45 subunit, 970 kDa complex consisting of seven mtDNA encoded as well 38 nuclear encoded subunits. This enzyme harvests electrons from NADH and transports these electrons through the enzyme via the initial acceptor of flavin mononucleotide (FMN) and a series of iron-sulfur (Fe-S) clusters to the final acceptor of this step, the membrane embedded molecule of ubiquinone, also known as coenzyme Q (Q). The transport of the NADH electrons drives the pumping of four protons from the matrix side of the membrane to the intermembrane space. In accepting the electrons from complex I, Q is reduced to form QH₂ (Saraste, 1999; Berg et al., 2012; Fiedorczuk et al., 2016).

The FADH₂ molecules formed by the citric acid cycle enter the electron transport chain at complex II, the succinate ubiquinone reductase. Complex II is also a membrane embedded enzymatic complex, however in contrast to complex

I it does not contain any mtDNA encoded subunits. Complex II transfers the electrons from FADH₂ through a series of iron-sulfur clusters to the acceptor of Q, but in this process it does not generate any translocation of protons as in the case with complex I (Saraste, 1999; Berg et al., 2012).

The third complex of the electron transport chain, complex III or cytochrome bc₁ contain two types of cytochromes. Cytochromes are heme-containing electron transporting proteins. In addition to its heme groups, complex III also contains an iron-sulfur cluster. Mammalian complex III contains 11 subunits, where one is coded for by mtDNA. The complex accepts electrons from QH₂ and transports these electrons via the iron-sulfur cluster and its heme groups to the electron acceptor, the soluble protein cytochrome c. During this process two protons are translocated from the matrix to the intermembrane space (Saraste, 1999; Berg et al., 2012).

Cytochrome c delivers the electrons from complex III to the intermembrane space side of complex IV, the cytochrome c oxidase. Mammalian complex IV contains 13 subunits where three catalytic core subunits are encoded by the mitochondrial genome and ten subunits are encoded by the nuclear genome. Complex IV contains two heme groups and three copper ions needed for its activity. Through these metal centers, complex IV transfers the electrons from cytochrome c to the final electron acceptor oxygen (O₂), forming water (H₂O). The need for oxygen in this process is what makes an organism aerobic. Through the transfer of electrons, complex IV pumps four protons from the matrix to the intermembrane space (Saraste, 1999; Berg et al., 2012).

The last step in the OXPHOS chain is where complex V, the ATP synthase, uses the proton gradient generated by the electron transport chain to drive the synthesis of ATP. The mammalian complex V has a molecular mass of more than 500 kDa and contains two mtDNA-encoded subunits. It is structured in two large modules; the membrane embedded F₀ module and the F₀-anchored soluble F₁ module that protrudes into the mitochondrial matrix. The two modules have distinct functions, as F₀ is a proton channel and F₁ functions as an ATPase. The energy from the proton diffusion through the F₀ module drives the F₁ synthesis of ATP from the substrates of ADP and P_i (Saraste, 1999; Berg et al., 2012).

Further metabolic pathways

The mitochondrion also plays a central role in the metabolism of amino acids. The first step of amino acid degradation involves the removal of the amino group to form ammonium ions (NH₄⁺). The ammonium ions can later be turned into urea through the urea cycle. The urea cycle takes place in two different

compartments, with some steps in the mitochondrial matrix and other steps in the cytosol. The urea that is formed can be excreted and discarded, whereas the carbon chain part of amino acids enters the citric acid cycle at various steps (depending on the amino acid). In the biosynthesis of amino acids, intermediates of the citric acid cycle are used as precursors (Berg et al., 2012).

As elaborated in the oxidative phosphorylation section above, mitochondria contain numerous iron containing proteins, both heme group proteins and iron-sulfur cluster containing proteins. In fact, mitochondria are the major site for the production of prosthetic groups containing iron, and further, this organelle has been implicated to play a role in the regulation of iron homeostasis in the entire eukaryotic cell (Sheftel and Lill, 2009).

1.1.5. The mitochondrial genome

The genome of mitochondria, mtDNA, is in mammals strictly maternally inherited (Giles et al., 1980), as paternal sperm mtDNA is eliminated in the late pronucleus stage of fertilization (Kaneda et al., 1995). The genome is present in multiple copies, ranging between 1,000 and 10,000 in most cell types (Bogenhagen and Clayton, 1974; Shmookler Reis and Goldstein, 1983), and up to more than 100,000 copies in oocytes (Piko and Taylor, 1987). The mitochondrial genome is in mammals relatively small in size, both compared to the nuclear genome of the respective organism and the genome of the bacterial ancestor (Andersson et al., 2003). The protein coding genes left in mitochondria are supposedly too hydrophobic to be translated in the cytoplasm and subsequently imported into the organelle (von Heijne, 1986). In addition, almost all mitochondrial genomes retain a setup of the tRNA and rRNA genes required for mitochondrial translation. Interestingly, most genomes have been reduced to only contain just enough coding sequence for a functional translation machinery. For example, the human genome encodes two rRNAs and 22 tRNAs, which is substantially fewer RNA molecules than other translation machineries (Anderson et al., 1981). As a consequence, the mitochondrial ribosomes have a higher protein to RNA ratio than both its bacterial and cytosolic counterparts (Agrawal and Sharma, 2012). The translation machinery of mitochondria also differs from other systems in that the genetic code has been modified. As an example, in vertebrate mitochondria the stop codon UGA codes for tryptophan, AUA and AUU code for methionine (and translation start) instead of isoleucine and AGA and AGG are termination rather than arginine codons (Anderson et al., 1981; Hallberg and Larsson, 2014).

Human mtDNA contain only two non-coding sequences, one large denoted as the control region or the non-coding region (NCR) and one smaller found in a tRNA cluster about 5 kilo base pairs (kb) from the NCR (Figure 4). The NCR is highly variable between species, but contains conserved sequence patches of functional importance. These sequences include: the promoters for transcription; a region with three conserved sequence blocks (CSBI, II and III); the origin of replication for the H-strand (OriH); and a termination-associated sequence (TAS) (Figure 5) (Shadel and Clayton, 1997). There is one promoter responsible for the expression of each strand, hence they are denoted as the H-strand promoter, HSP, and the L-strand promoter, LSP (Montoya et al., 1982). Transcription initiated from LSP is also believed to be responsible for primer formation at OriH, and RNA to DNA transitions have been mapped to the CSB region (Chang and Clayton, 1985; Chang et al., 1985; Kang et al., 1997; Pham et al., 2006). In addition, premature transcription termination at CSBI results in an abundant non-coding poly-adenylated LSP transcript (7S RNA) of unknown function (Figure 5) (Ojala and Attardi, 1974; Jemt et al., 2015). The process of mitochondrial transcription will be further explained in the transcription section of this thesis. The NCR region also contains a triple stranded displacement loop structure (D-loop), formed by premature termination of OriH-dependent mtDNA replication events close to the TAS region. The terminated product, referred to as 7S DNA, covers most of the non-promoter containing part of the NCR (Figure 5). The second non-coding region of human mtDNA contains the origin of replication of the L-strand, OriL (Figure 4) (Shadel and Clayton, 1997).

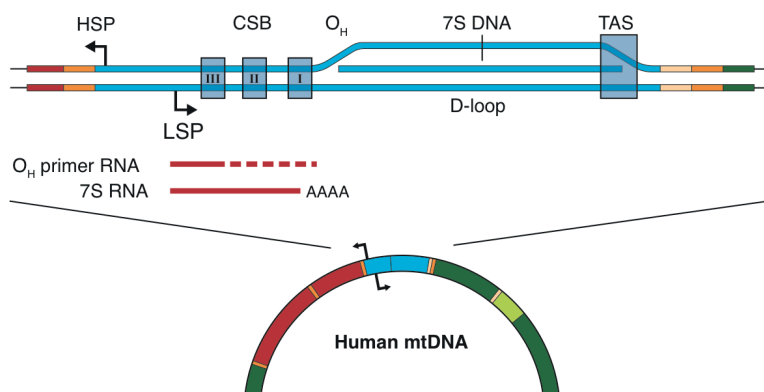


Figure 5. Schematic representation of the control region of mtDNA. The positions of the promoters (LSP and HSP), CSB elements (CSBIII, CSBII and CSBI), OriH (O_H), 7S DNA, D-loop and TAS region are indicated. The corresponding positions of the putative OriH primer and the polyadenylated 7S RNA are represented by their respective transcripts. Color-coding are the same as in figure 4.

The mitochondrial nucleoid

Mitochondrial DNA is compacted into a protein-DNA structure denoted as the mitochondrial nucleoid. The mitochondrial nucleoid is on average around 100 nm in diameter and contains roughly one genome per structure (Kukat et al., 2011). The main protein component of the nucleoid is the mitochondrial transcription factor A protein, TFAM. TFAM binds DNA in a non-sequence specific manner, introducing bends in the DNA (Fisher et al., 1992). These features together with the cooperative mode of binding make it an effective DNA packaging factor (Farge et al., 2012; Kukat et al., 2015). The TFAM to DNA ratio influences both transcription and replication as high levels of TFAM, i.e. tight packing, inhibits both of these processes *in vitro*. Further, reconstituted nucleoids at physiological TFAM to DNA ratios have shown a high degree of packaging variability between individual DNA molecules, ranging from almost naked DNA to tightly compacted nucleoids (Farge et al., 2014). In addition, the TFAM protein level has been shown to correlate to the levels of mtDNA (Larsson et al., 1998; Ekstrand et al., 2004). Many other proteins have been found within the mitochondrial nucleoid structure; these include mtDNA genome maintenance factors and transcription factors, but also some translation factors (Bogenhagen, 2012).

1.2. DNA transcription

1.2.1. A short introduction to DNA transcription

During transcription, the genetic information stored in DNA is copied into single stranded RNA. The RNA molecules can then be further processed to form functional RNA molecules (e.g. tRNAs, rRNAs or non-coding RNAs) or in the case of a protein-coding sequence, mRNA that can be translated into the corresponding amino acid sequence. The transcription process is carried out by enzymes known as RNA polymerases (RNAP) that initiate transcription from certain DNA elements, promoters. These enzymes read the template strand in a 3' to 5' direction and simultaneously catalyze the building of a nascent RNA strand in a 5' to 3' direction. As the RNA strand is built based on the base pairing of nucleotides to the template strand, the nascent RNA will be identical to the non-template strand (also denoted as the coding strand) in its nucleotide sequence. Transcription machineries can be of differing complexity ranging

from single subunit polymerase systems of viruses to the multi-protein transcription factor and polymerase systems of eukaryotes (Alberts, 2008; Berg et al., 2012).

Transcription is generally divided into three phases: initiation, elongation and termination. During initiation the transcription machinery identifies and binds to the promoter element of DNA. In *E. coli*, this element consists of two common motifs found upstream of the transcription start site, known as the -10 and -35 sequence (10 and 35 base pairs upstream of the start site, respectively, with the first transcribed nucleotide denoted as +1). To initiate transcription the holoenzyme RNAP (consisting of two α , one β and one β' subunit, $\alpha_2\beta\beta'$) first must bind to the promoter. In order to do so a fifth subunit, the specificity factor known as the σ -factor, aids the core polymerase in recognition of the promoter sequence. When transcription has been initiated the process enters the elongation phase and the σ -factor is released. Transcription proceeds until the machinery reaches a termination signal. The simplest form of a termination signal consists of a palindromic guanine-cytosine (GC) rich DNA sequence followed by an adenine-thymine (AT) rich sequence, with the adenines in the template strand. The palindrome creates a hairpin structure that halts the transcribing polymerase. The following AT-rich region generates a stretch of uracils (poly(U)) in the nascent RNA strand and since the base pairing between RNA and DNA with uracil and adenine is weak, the RNA dissociates from the template strand and transcription is terminated. Termination of transcription can also be forced by protein factors. Such factors generally bind the nascent RNA strand and hinder further transcription elongation by the polymerase. In *E. coli*, the most common transcription termination factor is known as the ρ -protein (Berg et al., 2012).

The eukaryotic nuclear genome is transcribed by three different RNA polymerases (Pol I, Pol II and Pol III). The three polymerases are similar in structure and subunit composition, and also share structural similarities with the RNAP of *E. coli*. Pol I transcribes rRNA genes, Pol II protein-coding genes and Pol III tRNA genes (Berg et al., 2012). All three nuclear RNA polymerase machineries share some common features; they contain a multi subunit RNAP, a TATA-box binding protein, and multiple transcription factors. The Pol II promoter may contain multiple DNA elements such as the TATA-box, the BRE (recognition site for transcription factor IIB), the Inr (initiator element), and the DPE (downstream core promoter element). In contrast to *E. coli*, proteins other than the polymerase recognize all Pol II promoter elements. These proteins are known as transcription factors. For Pol II, there are six general transcription factors, TFIIA, TFIIB, TFIID, TFIIIE, TFIIF and TFIIH. These factors assist the RNAP in recruitment, DNA binding, DNA melting and start site selection

(Alberts, 2008; Berg et al., 2012; Vannini and Cramer, 2012). Together, Pol II and its transcription factors form a mega Dalton sized pre-initiation complex (PIC). In addition to the general transcription factors, Pol II is also influenced by gene specific activators and the multi protein activator-PIC bridging complex known as the Mediator (Alberts, 2008; Murakami et al., 2013; Robinson et al., 2016).

1.2.2. The T7 bacteriophage RNA polymerase

On the other end of the complexity scale of transcribing polymerases is the single subunit RNAP of bacteriophage T7. This single subunit RNAP of 99 kDa requires no additional factors for successful promoter specific transcription *in vitro* or *in vivo*. The T7 RNAP is homologous to the DNA polymerase I family and shares no sequence or structural homology with multi subunit RNA polymerases (Sousa et al., 1993). The DNA polymerase I family of polymerases have what is often referred to as a right-handed structure, with thumb, fingers and palm domains. The catalytic site of these polymerases is found in the palm domain, in a template DNA housing cleft between the fingers and the thumb domain (Steitz et al., 1994). This active site contains two magnesium ions coordinated by two strictly conserved aspartic acid residues (Doublet et al., 1998).

The T7 RNAP consensus promoter sequence contains 23 conserved base pairs spanning from -17 to +6 (Oakley et al., 1979; Dunn and Studier, 1983). The polymerase recognizes and binds to the promoter mainly through two structural elements, the AT-rich recognition loop and the specificity loop. The AT-rich recognition loop binds to an AT-rich sequence at -17 to -13 and the specificity loop binds to the start site proximal promoter parts. A third element, the intercalating hairpin, contains an amino acid residue (Val237), which helps to separate the two DNA strands at position -4 to promote DNA melting creating what is known as the open complex. Both the AT-rich recognition loop and the intercalating hairpin are found in the N-terminal domain of the enzyme whereas the specificity loop is found in the fingers domain, in the C-terminal part of the polymerase. This latter loop stabilizes the open promoter-polymerase complex by interacting with the single stranded portion of the template strand (Cheetham et al., 1999).

DNA polymerases of the DNA polymerase I family contain a steric gate at the active site, which helps to discriminate against ribonucleotides. In most DNA polymerase I enzymes the steric gate is composed of a glutamic acid that blocks

nucleotides with a 2'-OH group (i.e. ribonucleotides) from entering the active site. In T7 RNAP the glutamic acid has been replaced by glycine. In addition, a histidine residue provides a hydrogen bond to the 2'-OH group, making the presence of a ribonucleotide more energetically favorable as compared to a deoxyribonucleotide, hence it functions as a RNA polymerase instead of a DNA polymerase (Cheetham and Steitz, 1999).

Once the open complex has been formed transcription can be initiated by the formation of a phosphodiester bond between the initiating +1 nucleotide and the incoming +2 nucleotide (Cheetham et al., 1999). T7 RNAP, as most RNA polymerases goes through cycles of abortive transcription, producing oligonucleotides between two and eight residues. After reaching the eighth nucleotide, T7 RNAP will go through promoter release and can thus transition from the initiation to the elongation phase (Martin et al., 1988; Briebe and Sousa, 2001). As the polymerase goes through this transition from initiation to elongation the enzyme undergoes substantial conformational changes, mostly seen in the N-terminal domain. These structural changes mainly involve promoter-binding elements of the T7 RNAP, explaining promoter release upon transition to elongation. The intercalating hairpin becomes unstructured and is no longer needed for the enzymatic activity, whereas the specificity loop is moved to form and accommodate the RNA exiting tunnel, where it contacts the protruding transcribed RNA. Although the elongating polymerase has undergone substantial structural rearrangement as compared to the initiating enzyme, the core polymerase domain remains largely intact (Tahirov et al., 2002; Durniak et al., 2008).

T7 RNAP can terminate transcription at two different types of termination signals. Class I termination signals closely resemble the ρ -independent termination of *E. coli* RNAP where a GC-rich hairpin followed by a poly(U) stretch effectively halts the RNA polymerase and decreases the RNA-DNA hybrid stability resulting in transcription termination (Hartvig and Christiansen, 1996). The second termination signal, class II, consists of a seven base pair DNA sequence, optionally followed by a poly(U) stretch. In the case where a poly(U) stretch is missing, the termination site is instead predominantly a pause site. This type of pause site is efficiently turned into a termination site in the presence of the T7 RNAP inhibitor T7 lysozyme (He et al., 1998; Lyakhov et al., 1998).

1.3. Mitochondrial transcription

Both strands of mtDNA are transcribed from their respective promoter (HSP and LSP) producing near genome sized polycistronic transcripts (Montoya et al., 1982). The majority of all protein coding genes are flanked by tRNA molecules, which are excised by RNase P and ELAC2 to form the individual mRNAs (Ojala et al., 1981; Hallberg and Larsson, 2014). This mode of creating individual mRNA transcripts is referred to as the punctuation model. After tRNA excision, non-tRNA transcripts are rapidly polyadenylated and in some cases, the addition of an A creates the translation stop codon for that particular transcript (Ojala et al., 1981).

The mitochondrial transcription machinery consists of four proteins; the mitochondrial RNA polymerase (POLRMT), the transcription factors A and B2 (TFAM and TFB2M respectively), and the transcription elongation factor (TEFM) (Gustafsson et al., 2016). Three of these proteins, POLRMT, TFAM, and TFB2M, are needed for promoter specific transcription initiation (Falkenberg et al., 2002; Litonin et al., 2010; Shi et al., 2012). Both mitochondrial promoters contain a high-affinity binding site for TFAM, situated upstream of the transcription start site, between positions -12 and -40 (Fisher et al., 1987). Bound at this position, TFAM introduces a 180° bend in the DNA (Ngo et al., 2011; Rubio-Cosials et al., 2011). POLRMT and TFB2M have been shown to bind around the transcription start site, covering positions -10 to +10 (Figure 6) (Gaspari et al. (2004) and Paper I).

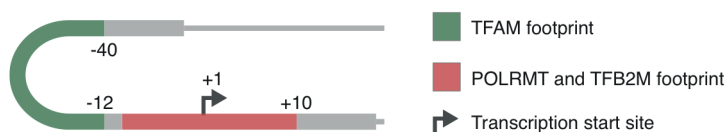


Figure 6. Schematic representation of a mitochondrial promoter. The positions for the TFAM and POLRMT-TFB2M footprints as well as the transcription start site are indicated. Numbering corresponds to the position relative to the transcription start site.

The mitochondrial initiation complex (IC) consists of TFAM, TFB2M and POLRMT (Gaspari et al. (2004), Morozov et al. (2015) and Paper I) whereas the active transcription elongation complex consists of POLRMT and TEFM (Agaronyan et al. (2015) and Paper III) (Figure 7). Polycistronic transcription

initiated from LSP is terminated prior to the HSP transcribed rRNA genes, at the binding site for the transcription termination factor, MTERF1 (Terzioglu et al., 2013; Shi et al., 2016). HSP transcription is terminated either in the TAS region at the 3'-end of the D-loop or, alternatively, in the promoter region, representing full genome transcription (Vijayasathya and Avadhani, 1996; Freyer et al., 2010; Jemt et al., 2015).

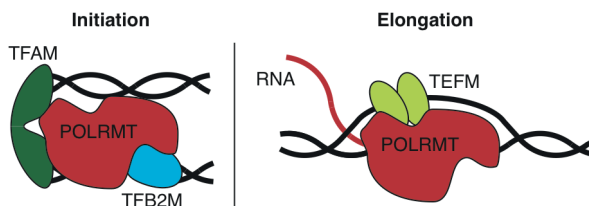


Figure 7. The compositions of the complexes for transcription initiation (left) and transcription elongation (right). Proteins are POLRMT (red), TFAM (dark green), TFB2M (blue) and TEFM (light green). The DNA is shown in black and the RNA in red.

1.3.1. The mitochondrial transcription machinery

Mitochondrial RNA polymerase – POLRMT

The mitochondrial RNA polymerase was first identified in yeast as a factor required for transcription and maintenance of the yeast mitochondrial genome (Greenleaf et al., 1986). Later it was shown that the mitochondrial RNA polymerases from both yeast (Rpo41) and human (POLRMT) are single subunit enzymes with C-terminal polymerase domains related to that of the T7 bacteriophage RNAP (Masters et al., 1987; Tiranti et al., 1997). In 2002, it was shown that POLRMT could initiate promoter specific transcription from mitochondrial promoters *in vitro*, but that this activity, in contrast to the T7 RNAP, called for two additional factors (Falkenberg et al., 2002; Litonin et al., 2010). These transcription factors will be discussed further later in this section.

Human POLRMT is an enzyme of 1230 amino acids with a size of ~140 kDa. The protein contains an N-terminal MTS (amino acids 1-41) that is cleaved off after import, an N-terminal extension (NTE, 42-367), an N-terminal domain (NTD, 368-647) and a C-terminal domain (CTD, 648-1230). During recent years structural information on the mitochondrial transcription machinery has been gathering. The first crystal structure of POLRMT showed that both the more

conserved CTD of the polymerase and the less conserved NTD were structurally similar to the corresponding domains of the T7 RNAP. The CTD contains the core polymerase activity, which is believed to be largely unchanged in evolution. It also contains the promoter recognizing specificity loop, which in POLRMT could still be involved in promoter recognition. The NTD (amino acids 368-647 of POLRMT and 1-271 of T7 RNAP) contains the T7 promoter-binding element of the AT-rich recognition loop as well as the melting inducing intercalating hairpin. These structures, although somewhat changed, seem to also be present in POLRMT (Ringel et al., 2011). Despite this, these structures appear to have, at least to some extent, been functionally replaced or reinforced by the presence of transcription factors. This hypothesis is supported by the fact that TFAM binds to the region corresponding to the element recognized by the AT-rich recognition loop of T7 RNAP and that TFB2M induces promoter melting (Matsunaga and Jaehning (2004), Sologub et al. (2009) and further in Paper II). Despite the requirement for TFAM and TFB2M, POLRMT still critically contributes to promoter recognition (Gaspari et al., 2004). The structures corresponding to the intercalating hairpin and the specificity loop have been shown to interact with the promoter DNA at positions -5 and -10 relative to the transcription start site (Morozov et al., 2015). In addition to these interactions, the polymerase also produces a prominent footprint upstream of the TFAM binding site around position -50 and -60 (Paper I). This interaction has later been mapped at position -49 which was shown to interact with parts of the NTD of the polymerase (Morozov et al., 2015).

In addition to the T7 RNAP-like CTD and NTD, POLRMT also contains the large NTE domain that is absent in the bacteriophage polymerase. With the exception of two pentatricopeptide repeat (PPR) motifs, the structure of the NTE has not been determined (Ringel et al., 2011). PPR motif proteins are mainly found in eukaryotic mitochondria and in chloroplasts of plants. The PPR motifs exhibit RNA binding and many PPR containing proteins are involved in various steps of organelle gene expression such as transcription, RNA processing, RNA stability and translation (Manna, 2015). The function of the PPR motifs found in POLRMT are however still unknown. In yeast, loss of the NTE manifests in genome instability, a phenotype that can be rescued by expression of the NTE as an individual polypeptide *in trans* (Wang and Shadel, 1999). Additionally, deletion of the NTE in yeast causes defects in RNAP-dependent promoter melting but stimulates the transition from transcription initiation to elongation (Paratkar et al., 2011). The NTE of POLRMT seems to have a role in the specificity of the polymerase as removal of the NTE of the mouse enzyme generates a more active polymerase that initiates transcription even in the

absence of TFAM and at non-promoter sites. These features are further covered in Paper I of this thesis.

It is a general belief that POLRMT-dependent transcription initiated from LSP also primes initiation of mtDNA replication at OriH. In support of this idea, loss of POLRMT leads to depletion of 7S DNA and mtDNA. In addition, LSP transcription is favored over HSP at low POLRMT levels, suggesting that replication priming is favored over transcription (Kuhl et al., 2016). A detailed understanding of the OriH-dependent initiation of mtDNA replication and the mechanisms of primer formation at this origin is however still missing.

In addition to its promoter specific transcription activities, POLRMT can also initiate transcription from single-stranded DNA (ssDNA), an activity that does not require the presence of TFAM or TFB2M. In contrast to the T7 RNAP, POLRMT is not processive on such a template, but rather produces shorter stretches of RNA transcripts of 20-100 nucleotides. This property allows POLRMT to function as the primase for OriL dependent DNA replication. OriL consists of a stem-loop structure where the single stranded loop contains a poly(dT) stretch. As POLRMT uses ATP as the initiating nucleotide this is a perfect template for POLRMT dependent priming of replication (Wanrooij et al., 2008; Fuste et al., 2010).

Mitochondrial transcription factor A – TFAM

Aside from being the packaging factor of the mitochondrial nucleoid, TFAM also functions as a transcription factor. TFAM was first identified as an activator of promoter-specific transcription by the mitochondrial RNA polymerase, without affecting the non-promoter-specific RNAP activity of that enzyme (Fisher and Clayton, 1985). TFAM is a 25 kDa protein that specifically recognizes and binds to a region 12 to 40 bp upstream of the transcription start site at mitochondrial promoters (Fisher et al., 1987).

TFAM and its yeast homologue (Abf2) were later identified as members of the high mobility group (HMG) box protein family (Diffley and Stillman, 1991; Diffley and Stillman, 1992; Fisher et al., 1992). In contrast to TFAM, Abf2 was shown not to be needed for transcription initiation (Xu and Clayton, 1992). Members of the HMG-box protein family are involved in many different aspects of DNA and RNA metabolism, one example being the human nuclear factor hUBF1, an RNA polymerase I transcription factor (Parisi and Clayton, 1991). TFAM contains two HMG-boxes that both contribute to DNA binding. The HMG-boxes are connected through the linker region in TFAM. When bound to its LSP promoter binding site, TFAM imposes a U-turn on the DNA with one

HMG box on each side, stabilized by the linker region (Ngo et al., 2011; Rubio-Cosials et al., 2011). The structures of TFAM bound to the HSP binding site and to non-specific DNA, have revealed similar U-turns in the DNA. The TFAM binding site at HSP has been proposed to be oriented in the opposite direction relative to the transcription start site as compared to the binding site at LSP (Ngo et al., 2014). A recent report has however questioned this difference in orientation and suggested a similar setup of initiation complexes at the two promoters (Morozov and Temiakov, 2016).

The transcription factor activity of TFAM has been partly attributed to a C-terminal tail that is absent from the yeast homologue Abf2. Removal of 15 amino acids from this region of TFAM severely impairs promoter-specific transcription, and conversely the addition of the human C-terminal tail to the yeast homologue permits Abf2 to function in the human mitochondrial transcription system (Dairaghi et al., 1995a). The positioning of TFAM at -12 to -40 is crucial for transcription initiation as an increase of this distance completely abolishes promoter activity (Dairaghi et al., 1995b). The presence of TFAM is absolutely required for the promoter binding and transcription initiation activity of POLRMT (Gaspari et al., 2004; Shi et al., 2012), hence TFAM has been suggested to recruit the polymerase to the promoter (Morozov et al. (2014) and Paper I). TFAM-POLRMT interactions have been mapped between the C-terminal parts of the transcription factor and parts of the NTE and NTD of the polymerase (Morozov et al., 2015).

Mitochondrial transcription factor B2 – TFB2M

When mitochondrial transcription was first fully reconstituted *in vitro*, two homologous proteins named TFB1M and TFB2M were reported to function as transcription factors in addition to the previously characterized factor TFAM. TFB2M was identified as the strongest transcription activator, with at least 10-fold stronger activity as compared to TFB1M (Falkenberg et al., 2002). TFB1M and TFB2M are both proteins related to rRNA methyltransferases and have been proposed to derive from such enzymes of the bacterial endosymbiont (Shutt and Gray, 2006b). In later years it has been shown that TFB1M still retains this function in mitochondria as knock out of this protein leads to a complete loss of 12S rRNA adenine dimethylation resulting in impaired mitochondrial translation (Metodieiev et al., 2009). TFB2M however has evolved into a *bona fide* transcription factor, essential for mitochondrial transcription (Litonin et al., 2010). In humans TFB2M is a protein of 396 amino acids and 45 kDa in size (Falkenberg et al., 2002).

Both the yeast (Mtf1) and the human (TFB2M) transcription factors have been proposed to be involved in promoter melting (Matsunaga and Jaehning, 2004; Sologub et al., 2009). This property is further elaborated in Paper II of this thesis. TFB2M is situated in close proximity to the priming nucleotide in the transcription initiation complex (Sologub et al., 2009). Interactions between TFB2M and POLRMT have been mapped to the region around the intercalating hairpin of POLRMT and the C-terminal parts of TFB2M. The N-terminal part of TFB2M is responsible for contacts with the priming nucleotide and has also been shown to interact with the promoter at position -5 (Morozov et al., 2015). TFB2M probably only plays a role in the initiation complex, as the yeast Mtf1 dissociates from the polymerase when transcription enters the elongation phase (Mangus et al., 1994).

Transcription elongation factor of mitochondria – TEFM

TEFM was first discovered in RNAi mediated knockdown experiments. Cells depleted of TEFM showed a dramatic decrease of transcript levels with increasing distance from the promoter. Based on these observations and evidence for a direct interaction with POLRMT, TEFM was proposed to be a transcription elongation factor (Minczuk et al., 2011). These *in vivo* observations were later supported by *in vitro* experiments, demonstrating that the addition of purified TEFM has a dramatic stimulatory effect on the processivity of POLRMT (Agaronyan et al. (2015) and Paper III). Furthermore, TEFM was found to stabilize POLRMT interactions with an elongation-like RNA-DNA scaffold (Paper III). A significant portion of POLRMT-dependent transcription events *in vitro* are prematurely terminated at the CSBII sequence downstream of the LSP promoter (Pham et al., 2006). Addition of TEFM to the transcription machinery completely abolishes this termination. As this termination at CSBII has been proposed to be important for primer formation during initiation of mtDNA replication at OriH, TEFM was proposed as a possible regulator of this process (Agaronyan et al. (2015), Paper III and Paper IV). The TEFM protein and its effects on the mitochondrial transcription machinery are further covered in Paper III of this thesis.

1.3.2. Mitochondrial transcription associated proteins

MTERF family proteins

In mammals, the MTERF family contains four proteins, MTERF1-4 (Linder et al., 2005), that are mitochondrial proteins with functions in transcription (MTERF1 and 3), replication (MTERF1), or translation (MTERF3 and 4) (Park et al., 2007; Camara et al., 2011; Spahr et al., 2012; Yakubovskaya et al., 2012; Wredenberg et al., 2013). The function of MTERF2 remains to be elucidated.

MTERF1

MTERF1 is the founding member of the MTERF family of proteins. This protein binds specifically to a site just downstream of the rRNA unit of mtDNA, inside the rRNA^L gene (Figure 4) (Kruse et al., 1989; Shang and Clayton, 1994; Fernandez-Silva et al., 1997). MTERF1 was first identified as a terminator of both LSP and HSP originated transcripts. It was proposed to play a crucial role in regulating HSP-dependent transcripts ensuring that rRNA are produced at a higher frequency than the downstream mRNA on the same strand. Later experiments have demonstrated that MTERF1 exclusively terminates LSP transcription and has no effect on the relative levels of rRNA and mRNA transcripts (Terzioglu et al., 2013; Shi et al., 2016). Recent data suggest that MTERF1 may also have a role in DNA replication. This will be further covered in the mitochondrial DNA replication section of this thesis.

MTERF3

The MTERF3 protein negatively regulates initiation of mitochondrial transcription *in vivo* (Park et al., 2007). The crystal structure of this factor identified the MTERF domain as a nucleic acid binding domain (Spahr et al., 2010), but an exact binding site for MTERF3 in DNA or RNA has not yet been identified. In addition to a positive effect on transcription, loss of MTERF3 also affects regulation of translation (Wredenberg et al., 2013). The molecular mechanisms behind MTERF3 function remain unknown.

MRPL12

MRPL12 is a protein component of the large subunit of the mitochondrial ribosome. This factor has been reported to be a stimulator of POLRMT dependent transcription (Wang et al., 2007), a conclusion that has been refuted by others (Litonin et al., 2010). Later MRPL12 was suggested to increase the

stability of POLRMT (Nouws et al., 2016). The effects of MRPL12 on mitochondrial transcription remain debated.

1.3.3. R-loops in the CSB region

Transcription from the LSP promoter is believed to be responsible for the formation of replicative primers at OriH (Shadel and Clayton, 1997). A feature of the region downstream of LSP that supports this hypothesis is the ability to form RNA-DNA hybrids. These hybrids have been shown to form during *in vitro* transcription across the OriH region from supercoiled DNA templates by retention of the RNA on the DNA strand, forming a triple stranded structure known as an R-loop. These R-loops are resistant to RNaseA and RNaseT1 but sensitive to RNaseH treatment (Xu and Clayton, 1995; Lee and Clayton, 1996; Xu and Clayton, 1996). Later work to determine the mechanism of R-loop formation has demonstrated that *in vitro* transcription frequently terminates downstream of the CSBII sequence and that this termination as well as the R-loop formation, depends upon G-quadruplex (G4) structures formed between CSBII RNA and the corresponding non-template DNA strand (Pham et al., 2006; Wanrooij et al., 2010; Wanrooij et al., 2012a). G4 structures are non-canonical nucleic acid structures involving quartets of guanine bases, bound to each other through Hoogsteen base pairing. The resulting structure is an intermolecular or intramolecular association of four nucleic acid strands in a parallel or anti-parallel manner (Burge et al., 2006). R-loop formation and its importance for DNA replication initiation is further elaborated in Paper IV of this thesis.

1.4. DNA replication

1.4.1. A short introduction to DNA replication

A cell that is dividing has to copy its genome to give each daughter cell one genomic copy. This calls for DNA replication, which is the genome copying process in biology. Organisms from all three domains of life (bacteria, archaea and eukaryotes) replicate their genomes in a semiconservative fashion, meaning that the two daughter molecules each contain one of the original DNA strands

and one newly synthesized strand. DNA polymerases use a DNA template strand to synthesize a new nascent strand in the 5' to 3' direction. Many DNA polymerases also carry a 3' to 5' exonuclease proofreading activity, which can remove mismatched nucleotides from the nascent DNA strand and allow for the correct base to be incorporated (Alberts, 2008; Berg et al., 2012).

In contrast to RNA polymerases, DNA polymerases are unable to initiate nucleic acid synthesis *de novo*. Instead DNA polymerases need a short RNA primer, annealed to the DNA template, as a starting point. The primer is produced by a primase, which in many systems is a specialized RNAP only used for this function. Furthermore, most DNA polymerases display a poor displacement activity and need additional factors that unwind the DNA duplex. A DNA helicase is such a factor that can unwind the double helix as it translocates along one of the DNA strands. Other factors essential for many DNA replication machineries include single-stranded DNA binding (SSB) proteins, topoisomerases and DNA polymerase processivity factors. Together, the proteins involved in the moving replication machinery form the replisome. The front of a moving replisome, i.e. the point of the parental DNA strand separation, is known as the replication fork (Alberts, 2008; Berg et al., 2012).

In most cases, the two strands of DNA are synthesized simultaneously by the replication machinery. Due to the antiparallel nature of duplex DNA, and the fact that DNA synthesis has a directionality (5' to 3'), only one of the two nascent strands can be synthesized in a continuous manner (the leading strand), whereas the other strand (the lagging strand) is synthesized in a non-continuous, piecemeal fashion. During leading strand synthesis, the lagging strand template forms a single stranded loop that can be replicated in fragments by a second polymerase traveling together with the leading strand polymerase. The nascent lagging strand fragments are known as Okazaki fragments. Prior to Okazaki fragment synthesis, the exposed single stranded loop is protected by SSB protein, which prevents re-annealing of the ssDNA (Alberts, 2008; Berg et al., 2012). The T7 bacteriophage has provided an important model for our understanding of DNA replication in general, and mtDNA replication in particular. This replication system will be further discussed below.

1.4.2. The T7 bacteriophage replisome

The T7 replisome can be reconstituted *in vitro* with only four proteins. This bacteriophage replisome constitutes of the T7 DNA polymerase (gene 5 protein, gp5), its processivity factor thioredoxin (trx), the helicase/primase (gp4), and the

T7 SSB protein (gp2.5). Travelling in the front of the T7 replisome is the helicase/primase gp4 (Hamdan and Richardson, 2009). The gp4 protein can load directly onto DNA and forms a hexameric complex surrounding the DNA template (Crampton et al., 2006). The gp4 protein is a 5' to 3' translocating helicase, meaning that it travels along the lagging strand. The translocation process is energy demanding and depends on dTTP hydrolysis. In addition to its helicase activity, gp4 is also the primase of the T7 replisome. As the protein travels along the lagging strand it synthesizes the primers required for initiation of Okazaki fragment synthesis. Protection of the lagging strand template is provided by the gp2.5 protein. The T7 replisome contains two gp5 DNA polymerases, one for each strand. The gp5 enzyme is a DNA polymerase I family protein with a structure similar to that of T7 RNAP. The processivity of gp5 is increased up to 100-fold by the presence of thioredoxin (Hamdan and Richardson, 2009).

1.4.3. Initiation of DNA replication

Bacteria

A well-characterized replication initiation machinery is that of the bacterium *E. coli*. In this organism DNA replication initiates at one unique site in the genome, known as OriC. The first step in the preparation for DNA replication involves the binding of the origin-recognition protein, DnaA. This factor binds as a multimer to specific binding sites, wrapping the DNA around a DnaA protein core. In the next step, the replicative helicase of *E. coli*, DnaB, is loaded onto the DNA aided by its helicase loader DnaC. Regions of OriC are now exposed as ssDNA and bound by SSB protein. At this stage, DNA is accessible to the primase, DnaG, which can synthesize the RNA primer. The replicative DNA polymerase, DNA polymerase III, next loads onto the complex aided by interactions with DnaB, triggering replication initiation and DnaA dissociation (Alberts, 2008; Berg et al., 2012; Costa et al., 2013).

ColE1-like plasmids

Another well characterized system for initiation of DNA replication is that of ColE1-like plasmids, which in some aspects may also resemble mitochondrial DNA replication initiation (Shadel and Clayton, 1997). ColE1-like plasmids carry no replication factor genes, but rely on the host bacterium to supply enzymes for both primer formation and DNA replication. The origins of

replication in these plasmids differ from many other systems, since the normal transcription machinery of *E. coli* produce the primers. *E. coli* RNAP initiates transcription around 550 nucleotides upstream of the replication origin and the transcribed RNA, denoted as RNA II, remains stably associated with the DNA template in an RNA-DNA hybrid R-loop. The R-loop RNA is next processed by RNaseH, followed DNA polymerase I initiation of DNA synthesis from the newly formed primer 3'-ends (Itoh and Tomizawa, 1980). How the RNA II R-loop is anchored to the template DNA is not clear, but mutations in an upstream guanine stretch (-263 to -268 relative to the origin of replication) have a negative effect on RNA-DNA hybrid formation and impair primer utilization (Masukata and Tomizawa, 1984). A second stretch of guanines present just upstream of the replication origin (-16 to -21) is also of importance, as deletion or interference with this sequence impairs replication initiation (Ohmori et al., 1987; Masukata and Tomizawa, 1990). The ColE1 replication origin can function in the absence of RNaseH, but initiation events will then occur downstream of the RNaseH dependent origin. Even if RNaseH is absent, replication still depends on R-loop formation (Dasgupta et al., 1987).

1.5. Mitochondrial DNA replication

The DNA replication machinery in mitochondria is in many aspects related to that of the T7 bacteriophage (Shutt and Gray, 2006a). For example, the catalytic core of the mammalian mitochondrial DNA polymerase (POL γ), is a single subunit (POL γ A), which is structurally related to the gp5 polymerase. POL γ also contains a processivity factor, the non-T7 related POL γ B dimer. In addition to POL γ A, the mitochondrial replicative DNA helicase, TWINKLE, is also structurally related to its T7 counterpart. An additional replication factor found in mitochondria is the *E. coli* related mitochondrial single-strand binding protein, mtSSB (Tiranti et al., 1993). Together, these three factors, the POL γ holoenzyme, TWINKLE, and mtSSB constitutes a minimal mtDNA replisome (Figure 8), capable of synthesis of products up to 16 kb in length (Korhonen et al., 2004). The characteristics of these individual proteins are further discussed later in this section.

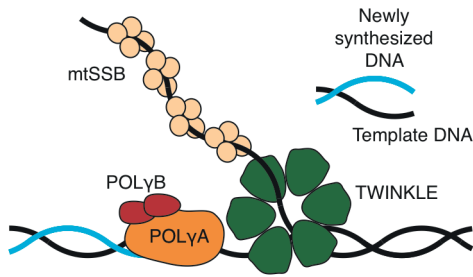


Figure 8. The mitochondrial replisome consists of the replicative helicase TWINKLE (green), the POL γ DNA polymerase (POL γ A in orange and POL γ B in red) and mtSSB (beige). The parental strands are shown in black and the newly synthesized strand in blue. TWINKLE unwinds the double stranded DNA followed by the synthesizing POL γ holoenzyme, while mtSSB covers the displaced DNA strand.

1.5.1. Models for replication of mitochondrial DNA

Different models have been suggested for how the mitochondrial genome is replicated. The first proposed model, known as the strand displacement model (Robberson et al., 1972; Tapper and Clayton, 1981; Clayton, 1982), states that replication is initiated at OriH and as the nascent H-strand is replicated continuously, the parental H-strand is displaced and protected by mtSSB. When the replication machinery reaches approximately two thirds of the genome size, the second origin, OriL, is exposed as single stranded DNA, and thus becomes activated. POLRMT can now perform primer synthesis and POL γ can initiate replication of the L-strand. As the template H-strand is single stranded there is no need for TWINKLE during L-strand DNA synthesis. Both mtDNA strands are continuously synthesized, leading to the formation of two daughter molecules (Figure 9). OriL replication initiation has been reconstituted *in vitro* and has been shown to depend on a hairpin structure composed of a GC-rich stem and a T-stretch containing loop. POLRMT recognizes this structure and initiates primer synthesis from the single stranded T-stretch (Wanrooij et al., 2008; Fuste et al., 2010). OriL is conserved among vertebrates and *in vivo* saturation mutagenesis studies have shown that OriL is required for mtDNA maintenance in mice (Wanrooij et al., 2012b). The strand displacement model was originally based on electron microscopy analysis of replication intermediates (Robberson et al., 1972) and is supported by a wealth of biochemical observations (Clayton, 2003). Recent support for the model comes from genome wide mapping of mtSSB occupancy. The *in vivo* occupancy showed a high degree of protein binding downstream of OriH, which gradually

decreased to OriL. From OriL the protein binding increased again, to drop gradually to OriH. The binding, apart from 7S DNA, was exclusively observed on the heavy strand. The pattern of mtSSB binding corresponds to what would be expected for strand displacement DNA synthesis, providing *in vivo* evidence in favor of this replication model (Miralles Fuste et al., 2014).

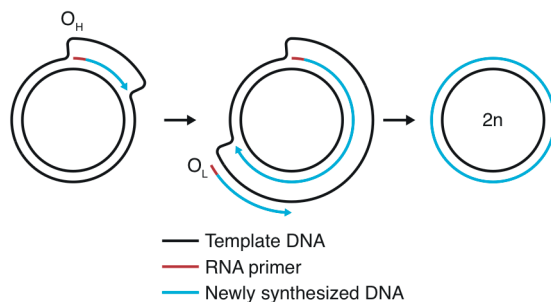


Figure 9. The strand displacement model for mtDNA replication. Replication is initiated at OriH (O_H) and as replication proceeds, the parental H-strand is displaced as a single strand. When replication reaches two thirds of the genome size, OriL (O_L) becomes single stranded and activated. Replication now proceeds in two directions until two fully replicated daughter molecules are formed ($2n$). The parental strands are shown in black, the newly synthesized strands in blue and the replicative primers in red.

Later two competing models for the replication of mtDNA were proposed. These are known as the strand-coupled and the RITOLS models. The strand-coupled model was proposed after the finding of fully double stranded replication intermediates by two-dimensional agarose gel electrophoresis (2D-AGE). These double stranded intermediates suggested a more conventional mode of replication with coupled leading and lagging strand synthesis (Holt et al., 2000). In this model, replication is initiated from a broad region named OriZ near OriH, and OriH instead functions as replication fork pause site (Bowmaker et al., 2003; Reyes et al., 2005). In the identification of the strand-coupled intermediates of replication, molecules containing ssDNA were also identified. Later, these stretches were shown to contain annealed RNA transcripts. A new replication model was suggested with ribonucleotide incorporation throughout the lagging strand or RITOLS (Yang et al., 2002; Yasukawa et al., 2006; Reyes et al., 2013). The RITOLS model also states a strand displacement mechanism of DNA replication but with RNA instead of mtSSB binding to the displaced H-strand.

1.5.2. The mitochondrial DNA replication machinery

DNA polymerase γ – POL γ

POL γ is the replicative polymerase of mammalian mitochondria (Clayton, 1982). It consists of a heterotrimer with one ~140 kDa catalytic subunit, POL γ A, and two ~50 kDa processivity factor subunits, POL γ B (Yakubovskaya et al., 2006). The POL γ A subunit harbors both 5' to 3' polymerase and 3' to 5' exonuclease activities. The exonuclease activity enables proofreading of the incorporated nucleotide, and when lost in mice, a substantial increase in point mutations is observed (Trifunovic et al., 2004). Additionally, the exonuclease deficient POL γ polymerase has an increased strand displacement activity being able to replicate short stretches of double stranded DNA (dsDNA), even in the absence of TWINKLE. This displacement activity impairs subsequent ligation of replication intermediates as the exonuclease deficient POL γ displaces any encountered 5'-ends (Macao et al., 2015). The processivity subunit of POL γ , POL γ B, is structurally related to aminoacyl-tRNA synthetases (Carrodeguas et al., 1999; Carrodeguas et al., 2001). It increases the processivity of POL γ A by increasing the polymerase affinity to the DNA template (Lee et al., 2010).

TWINKLE helicase

The TWINKLE gene was first identified in a search for mutations that caused mitochondrial disease. The TWINKLE protein displayed sequence similarities to the T7 bacteriophage helicase-primase enzyme gp4, immediately suggesting that the protein could be the long-sought replicative helicase in mitochondria (Spelbrink et al., 2001). Similarly to gp4, TWINKLE exhibits a 5' to 3' helicase activity and hydrolyzes nucleotide triphosphates to translocate along the DNA strand (Korhonen et al., 2003). TWINKLE has a molecular weight of about 70 kDa and forms a hexameric complex (Korhonen et al., 2003; Fernandez-Millan et al., 2015). TWINKLE is specifically stimulated by mtSSB (Korhonen et al., 2003) and can load onto closed circular templates (Jemt et al., 2011). TWINKLE appears however to have lost the primase activity present in gp4 (Farge et al., 2008).

Mitochondrial single-strand binding protein – mtSSB

The mitochondrial single-strand binding protein, mtSSB, is a small (~15 kDa), tetramer forming protein. It binds to ssDNA, which is wrapped around the tetramer (Li and Williams, 1997; Yang et al., 1997). The mtSSB protein in *Drosophila* has been shown to stimulate both the polymerase and exonuclease

activity of the mitochondrial DNA polymerase. Additionally, it stimulates primer utilization in replication initiation (Farr et al., 1999). Furthermore, when the human minimal replisome was reconstituted, mtSSB greatly stimulated the replication reaction (Korhonen et al., 2004).

1.5.3. Additional mitochondrial DNA maintenance factors

MGME1

MGME1 was relatively recently identified as a disease associated gene among humans. It has been shown to code for a nuclease that has an apparent role in 7S DNA processing (Kornblum et al., 2013; Nicholls et al., 2014). Furthermore, MGME1 has been suggested to function in flap removal to facilitate successful ligation of newly synthesized mtDNA molecules (Uhler et al., 2016).

RNaseH1

RNaseH1 catalyzes the degradation of RNA hybridized to a strand of DNA. Knockout of RNaseH1 in mice is embryonically lethal and shows evidence of an early decrease of mtDNA (Cerritelli et al., 2003). Additionally, loss of RNaseH1 function leads to retention of primers of mtDNA replication both on the 5'-end of 7S DNA and at OriL (Holmes et al., 2015; Akman et al., 2016), suggesting a role in primer processing.

MTERF1

As discussed under the mitochondrial transcription section of this thesis MTERF1 was first identified as a terminator of transcription. Recent studies have proposed an additional role for this factor, in mtDNA replication, where it functions as a pause site for the heavy strand-synthesizing replication fork (Hyvarinen et al., 2007; Shi et al., 2016). One possible purpose for such a pause site is to create a replication fork barrier to avoid head on collisions of rRNA transcription and the oppositely directed replication fork (Shi et al., 2016).

1.6 Mitochondrial DNA in disease

A large and heterogeneous group of diseases are caused by mutations and deletions in the mitochondrial genome. Usually these diseases affect tissues with high energy demand such as the central nervous system, skeletal muscle or heart. The affected cell types also frequently include the β -cells of the pancreas, cells of the inner ear, and cells of the kidney tubular system. Hence mitochondrial disease can cause many different symptoms, such as progressive external ophthalmoplegia (PEO), fatigue, heart failure, diabetes, deafness, and kidney failure (Greaves et al., 2012). Mutations in mtDNA have also been implicated in the natural process of ageing, a hypothesis that is supported by the premature ageing phenotype of mice expressing the exonuclease deficient POL γ (Trifunovic et al., 2004; Greaves et al., 2012). A disease causing mutation may be found in the mitochondrial genome and thus directly affect the OXPHOS components, but mutations in the nuclear encoded mtDNA maintenance factors may also lead to the accumulation of mtDNA mutations and/or multiple deletions (Greaves et al., 2012). Disease causing mutations have been found in the genes for the DNA polymerase POL γ , the replicative helicase TWINKLE, and the nucleases MGME1 and RNaseH1 (Spelbrink et al., 2001; Kornblum et al., 2013; Copeland, 2014; Reyes et al., 2015).

Antiviral nucleotide analogues

The mitochondrial transcription and DNA replication machineries are subjects of adverse side effects to therapeutic drug treatments. Both POLRMT and POL γ are sensitive to nucleotide analogues used for treatment of retroviral infections, including HIV and hepatitis C. Nucleotide analogues are developed to inhibit reverse transcriptases, which are members of the DNA polymerase I family and thus similar in structure to the mitochondrial polymerases. These nucleotide analogues have in many cases been shown to exhibit mitochondrial side effects as a consequence of off target inhibition of mitochondrial transcription and replication (Kakuda, 2000; Arnold et al., 2012; Baumgart et al., 2016).

2. Specific Aims

The overall aim of this thesis has been to gain molecular insights into the processes of transcription and DNA replication in human mitochondria. In the four papers of this thesis we have addressed different questions regarding the mechanisms of these processes. In Paper I and II we have investigated the transcription initiation machinery, in paper III the elongation factor TEFM and in Paper IV the initiation of mtDNA replication at OriH. The specific aims for each study are listed below.

Paper I: To investigate the role of the N-terminal extension of POLRMT in transcription initiation.

Paper II: To characterize the importance of TFB2M for DNA unwinding in transcription initiation.

Paper III: To characterize the putative mitochondrial transcription elongation factor TEFM.

Paper IV: To investigate the R-loop in the CSB region and how it can be used as a primer for replication initiation at OriH.

3. Results and Discussion

3.1. Paper I

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

The catalytic domain of POLRMT is structurally similar to that of T7 RNAP. Therefore, the nucleotide addition cycle is believed to be relatively similar for the two enzymes. In addition to the T7-like core of POLRMT, the enzyme also contains a large N-terminal extension (NTE) of unknown function. A crystal structure has been reported for a truncated version of POLRMT, lacking the N-terminal 104 amino acids. With the exception of two PPR motifs, structural information of the NTE was missing in the published data. Whereas the truncated enzyme was fully active, the same study demonstrated that NTE truncations of 200 and 368 amino acids retained activity on a pre-melted template, but failed to initiate transcription from a duplex promoter (Ringel et al., 2011). The NTE of the yeast polymerase Rpo41 has been shown to be important for genome maintenance (Wang and Shadel, 1999). In addition, deletion of the Rpo41 NTE seems to affect the ability of the polymerase to melt the promoter DNA, and possibly lower the affinity for the transcription factor Mtf1. The NTE has also been proposed to inhibit transition from the initiation to the elongation phase (Paratkar et al., 2011). The function of the NTE of mammalian POLRMT and how this domain may contribute to transcription complex assembly at mitochondrial promoters, remain largely unknown.

In this work we generated a number of NTE-truncated recombinant versions of the mitochondrial RNA polymerases from mice (Polrmt) and humans (POLRMT) to assess the function of this domain in transcription. Initial screens showed that only one of these truncations, a version lacking the N-terminal 320 amino acids of mouse Polrmt, generated an active polymerase. This truncated polymerase (hereafter $\Delta 320$ -Polrmt) lacks almost the entire NTE, including the two PPR motifs found in this domain. To study the function of the NTE we employed $\Delta 320$ -Polrmt in a number of biochemical assays, including microscale thermophoreses (MST), *in vitro* transcription assays and DNase I footprinting.

Deletion of the NTE in Rpo41 weakens interactions with Mtf1 and this observation prompted us to use MST to address the affinity between Polrmt (wild type or $\Delta 320$ -Polrmt) and Tfb2m. MST measures the change in migration induced by a temperature gradient of a labeled molecule upon binding of a ligand. Using labeled Tfb2m and titrations of the two Polrmt variants we were able to show that in contrast to the yeast system, a deletion of the NTE increased the affinity of Tfb2m to Polrmt. To investigate how the NTE deletion affected the activity of the polymerase we used a reconstituted *in vitro* transcription system. When compared to wild type Polrmt, $\Delta 320$ -Polrmt showed a clear increase in transcription activity. Furthermore, the truncated polymerase was partially Tfam-independent and could initiate lower levels of transcription from both promoter elements and non-specific DNA sequences in the absence of Tfam. On a template lacking mitochondrial promoters, the non-specific transcription activity of $\Delta 320$ -Polrmt was increased by the presence of Tfam. In contrast, the presence of this transcription factor together with a promoter-containing template completely abolished the non-specific transcription activity.

To further examine the effect of the NTE truncation and the formation of the transcription initiation complex, we used DNase I footprinting. With this technique we were able to show that the overall protection by the transcription initiation complex was unchanged by the absence of the NTE. We could also demonstrate that the initiation complex could be recruited in discrete steps, since Tfam and Polrmt could form a complex on the promoter in the absence of Tfb2m. In addition, the previously characterized Polrmt footprint over the transcription start site were in our assays accompanied with an additional protection upstream of the Tfam binding sequence. Loss of the NTE did not affect the observed protection pattern. To confirm the possibility of stepwise assembly of the mitochondrial transcription machinery, we used a FRET based assay monitoring structural changes at the transcription start site of a human mitochondrial promoter. These results showed that TFAM induces structural changes in the promoter, probably as a consequence of its non-specific DNA binding and bending. These changes were suppressed in the presence of POLRMT indicating binding of the polymerase to the promoter, effectively blocking TFAM from this site. The structural changes reappeared when TFB2M was added to the complex. We believe that these signals arise from TFB2M induced promoter melting (see Paper II below).

The results we obtained in this work differed from studies of the yeast mitochondrial transcription system and also from studies of human POLRMT that had suggested that loss of the NTE resulted in a defective polymerase. In contrast, the $\Delta 320$ -Polrmt studied by us was fully active and even showed

transcription activity on non-promoter templates and in the absence of Tfam. We hypothesized that the NTE functions partly as an inhibitory domain, suppressing transcription initiation in the absence of a promoter-bound Tfam protein. The increased affinity between Polrmt and Tfb2m in the absence of the NTE, together with the possibility of a stepwise assembly of the transcription machinery, led us to suggest a new mechanism for transcription initiation whereby Tfam recruits Polrmt to the promoter, releasing the inhibitory effect of the NTE and successfully increasing the affinity to Tfb2m, leading to recruitment of Tfb2m and transcription initiation.

3.2. Paper II

Human mitochondrial transcription factor B2 is required for promoter melting during initiation of transcription

Previous studies on the yeast mitochondrial transcription system have shown that Rpo41 can initiate transcription in the absence of Mtf1 if provided with a pre-melted bubble template (i.e. a short single-stranded stretch near the transcription start site). In a similar way, negatively supercoiled DNA that can stimulate duplex breathing can also circumvent the requirement of Mtf1 (Matsunaga and Jaehning, 2004). When pre-melted bubble templates were used with the human transcription system, POLRMT was shown to initiate transcription in the absence of both TFAM and TFB2M (Sologub et al., 2009). We found these observations to be in conflict with previous results from our own group and with the existing model for transcription initiation, which had reported that TFAM is needed for successful POLRMT recruitment to mitochondrial promoters (Gaspari et al. (2004), Morozov et al. (2014) and Paper I)

In this paper we used different footprinting techniques as well as *in vitro* transcription to investigate the involvement of TFAM and TFB2M in promoter melting. To monitor melted regions at a mitochondrial promoter we used potassium permanganate (KMnO₄) footprinting. Permanganate is used to detect unpaired thymidine bases found in DNA. In parallel to the permanganate footprinting we performed DNase I footprinting to be able to connect the results with protein binding. Our results clearly showed that the pre-initiation complex of TFAM and POLRMT is a closed complex with no melted regions found near the promoter. In contrast, when TFB2M is present to form the fully assembled

initiation complex a clear melted region appears at -1 to +3 relative to the transcription start site.

To further increase the understanding of promoter melting we performed *in vitro* transcription, using a template with a pre-melted bubble spanning from -4 to +3 relative to the transcription start site. The size of the melted region was identical to what had been used in a previous study (Sologub et al., 2009). In our hands, this template did not result in promoter specific transcription, but rather initiation appeared at multiple sites along the open template strand. This indicates that POLRMT recognizes this template as ssDNA, which explains the observation that this template supports transcription in the absence of both TFAM and TFB2M. The fact that the previous study found promoter specific transcription from this template is probably explained by the use of short primers to direct the initiations to the transcription start site in those experiments. In an attempt to find a promoter specific bubble template we screened pre-melted templates using *in vitro* transcription in the absence of TFB2M. We found that a pre-melted bubble at -3 to +1 was sufficient for TFB2M independent transcription. This template also proved to be TFAM dependent and generated promoter specific transcription products.

As the yeast RNA polymerase could initiate transcription in the absence of its TFB2M homologue from supercoiled templates, we turned to the same experiment using the human transcription machinery. Human POLRMT was unable to initiate transcription from supercoiled DNA templates in the absence of TFB2M, both at standard (64 mM NaCl) and at low (10 mM) salt concentrations.

From this work we concluded that the pre-initiation complex of TFAM and POLRMT is unable to melt a mitochondrial promoter, and that TFB2M is required for open complex formation and transcription initiation. Melting induced by TFB2M occurs at -1 to +3, and does not extend beyond positions -6 to +4. Furthermore, we could conclude that a pre-melted bubble at -3 to +1 is sufficient to obtain TFB2M independent transcription, while still retaining TFAM dependence. These results are in clear agreement with the existing model that states that TFAM is needed for recruitment of POLRMT and that TFB2M may enter the transcription initiation complex in a later step, and upon binding induces promoter melting and transcription initiation.

3.3. Paper III

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

In 2011, TEFM was identified by RNAi mediated knockdown in human cells as a protein of importance for the formation of promoter distal mitochondrial transcripts. Further experiments indicated an interaction between TEFM and POLRMT and a possible stimulation of POLRMT activity in the presence of TEFM (Minczuk et al., 2011).

We found the results of the previous study on TEFM intriguing and decided to investigate this factor further, using our *in vitro* transcription system. To this end we purified a recombinant form of TEFM and employed it in a transcription reaction using an LSP containing plasmid. Interestingly, the presence of TEFM completely abolished premature transcription termination at CSBII and also caused a slight increase in the level of run-off products. As the template produced a relatively short run-off product (~400 nucleotides) in comparison to the transcript length seen *in vivo*, we next turned to using longer templates (>3000 bp). In transcription from these templates it became clear that POLRMT in the absence of TEFM is a non-processive polymerase that generates multiple premature terminations, with only a fraction of the initiation events leading to full-length transcripts. In contrast, the presence of TEFM caused a potent stimulation of processivity, as no prematurely terminated transcripts could be detected and far higher levels of run-off products were generated. Using time course experiments we were able to show that many points of premature termination (especially at CSBII) were instead transcription pause sites in the presence of TEFM.

Using MST we were able to demonstrate a specific interaction between TEFM and POLRMT. In addition we found that TEFM increased the affinity between POLRMT and an RNA-DNA elongation-like scaffold. By using DNase I footprinting we showed that TEFM appears to bind POLRMT at the initiation complex stage, as the protection of the promoter area was altered in the presence of this factor.

Thus far our experiments had shown that TEFM could aid POLRMT in transcription of G-quadruplex forming structures, such as CSBII, and also at other sites prone to transcription termination. To further investigate the bypass of putative termination points we next used DNA templates containing sites of DNA damage. These templates consisted of the LSP promoter with a

downstream damage point of either an apurinic/aprimidinic site (AP-site) or 8-oxo-2'-deoxyguanosine (8-oxo-dG). We demonstrated that POLRMT is unable to pass an AP-site in the absence and in the presence of TEFM. POLRMT could however to a low extent transcribe past an 8-oxo-dG and the presence of TEFM stimulated this lesion bypass activity.

We could conclude that TEFM is a *bona fide* transcription elongation factor that increases POLRMT affinity for the DNA template making the polymerase less prone to transcription termination. In addition, the presence of TEFM completely abolished transcription termination at CSBII, which indicated that the protein could help to regulate formation of primers required for initiation of mtDNA replication at OriH. Another report published at the same time as ours suggested that TEFM regulates the switch from transcription to replication in human mitochondria (Agaronyan et al., 2015). Although we discussed the same possibility in our report, we also raised a few issues that must be resolved before this statement could be verified. Firstly, if TEFM were the sole regulator of primer formation versus full-length transcription this would also be expected to affect HSP transcription. In our paper we clearly show that the absence of TEFM will give rise to fragmented transcription, terminating all over the transcribed template. This would not be a favorable situation in transcription of the mitochondrial genome. Furthermore, even though the CSBII-terminated transcript has, in multiple reports from our group and others, been proposed to constitute the replicative primer, this has not been conclusively demonstrated *in vitro* or *in vivo*. To address these questions we continued with the work presented in Paper IV of this thesis.

3.4. Paper IV

The molecular mechanism of DNA replication initiation in human mitochondria

Previous work has demonstrated that RNA-DNA hybrids (R-loops) can be formed during *in vitro* transcription across the OriH region (Xu and Clayton, 1995; Lee and Clayton, 1996; Xu and Clayton, 1996). These R-loops have been suggested to constitute the primers for OriH replication initiation. More recent work demonstrated that transcription *in vitro* frequently terminates at the CSBII sequence and that R-loop formation depends upon G-quadruplex structures formed between RNA and the non-template DNA strand (Pham et al., 2006;

Wanrooij et al., 2010; Wanrooij et al., 2012a). Furthermore, the 3'-ends of these prematurely terminated transcripts overlap with sites of RNA to DNA transitions in the nascent heavy strands of mtDNA *in vivo* (Kang et al., 1997; Pham et al., 2006). As demonstrated in Paper III of this thesis, addition of the transcription elongation factor TEFM to *in vitro* transcription reactions abolished transcription termination at CSBII. These observations could suggest that TEFM is involved in regulation of primer formation and the switch between transcription and replication (Agaronyan et al. (2015) and Paper III). Many parts of the mitochondrial DNA replication process have been reconstituted *in vitro*, including the formation of a minimal mtDNA replisome (Korhonen et al., 2004) and replication dependent activation of OriL (Fuste et al., 2010). However, an understanding of the initial step of mtDNA replication, that of replication initiation at OriH, remains missing.

In this work we address the questions of how TEFM activity affects R-loop formation. We have also continued to characterize the CSB region R-loop, its regulatory role and potential to function as a primer for mtDNA replication.

To investigate the effect of TEFM on the formation of CSB region R-loops we used RNaseA cleavage of *in vitro* transcription reaction products, since RNA-DNA hybrids are resistant to this nuclease. With this assay we demonstrated that R-loops only form when POLRMT transcribes a supercoiled LSP template. Premature transcription termination at CSBII is also strongly stimulated by negative supercoils and in contrast to transcription on relaxed templates, nearly no transcription events escaped to full-length transcription. Interestingly TEFM completely abolished R-loop formation. We analyzed the size of the CSB region R-loops and showed that these hybrid structures stretch from close to the transcription start site, all the way to the CSBII termination site. Interestingly, R-loop structures in close proximity to the transcription start site effectively inhibited new rounds of transcription. In a transcription screen of mutant templates in combination with R-loop detection we were able to identify two regions of importance for R-loop formation, CSBII and CSBIII. We also showed that the R-loop anchoring ability of CSBIII was dependent on guanines in the non-template DNA strand, possibly indicating a CSBII-like G4 anchoring mechanism.

As previous attempts to use transcripts prematurely terminated at CSBII as primers for POL γ replication initiation have failed, we hypothesized that the RNA of these hybrids must be processed to create new 3'-ends that can be used by the replication machinery. To address this possibility we included a recombinant version of the mitochondrial RNaseH1 in transcription reactions from supercoiled LSP templates. We found that RNaseH1 can use the CSB

region R-loops as substrates, creating new RNA 3'-ends. To investigate if these processed 3'-ends could be used for priming of mtDNA replication, we set up a transcription reaction with a supercoiled LSP template, including RNaseH1, and added the replicative polymerase POL γ and the four deoxyribonucleotide triphosphates (dATP, dGTP, dTTP spiked with radioactively labeled dTTP and chain terminator ddCTP). By post-treating the reactions with potassium hydroxide (KOH), we ensured that only DNA would be detected. In the absence of RNaseH1 we observed no replication products. However, in the presence of low amounts of RNaseH1, replication products started to appear. In an attempt to increase the efficiency of the reaction we included mtSSB, which is known to stimulate POL γ . The presence of mtSSB greatly enhanced the replication reaction and we could now observe multiple clusters of replication products.

With the help of mutant templates, we could determine where in the CSB region replication was initiated. The main initiation point was observed downstream of CSBIII. A second, less abundant cluster of initiations, was found downstream of CSBII. In order to explore how much of the primer was retained on the replication products we post-treated the reactions either with KOH as before or with KCl, which will not remove ribonucleotides attached to the 5'-ends of the replication products. Although the KCl treatment resulted in a more complex pattern, most of the replication products seemed to have one or two ribonucleotides still attached to their 5'-ends.

We also addressed the effect of TEFM on replication initiation. Interestingly, we failed to observe any prominent replication products in the presence of TEFM, indicating that R-loop formation is required for proper RNaseH1 processing and initiation at OriH.

The results presented in this work represent the first successful initiation of replication from OriH using a defined mitochondrial *in vitro* system. We have shown that the R-loops created by POLRMT at the CSB region of mtDNA are anchored mainly by the CSBII and CSBIII elements, possibly in both locations by G4 structures. Additionally, the R-loop can act as a negative regulator, preventing new rounds of transcription under conditions when mtDNA is replicated. We could further conclude that RNaseH1 activates and mtSSB stimulates the replication initiations seen in our experiments. In contrast, TEFM abolishes R-loop formation and replication initiation. We observed replication initiations both downstream of CSBII and CSBIII, which is somewhat unexpected, as RNA to DNA transitions have previously only been mapped to downstream of CSBII. The reason for this discrepancy is yet unclear. Additional factors that direct the replication initiation events to specific start sites could be missing. Alternatively, CSBIII initiations could be present *in vivo*, but failed to

have been observed due to technical difficulties. One explanation for the latter scenario could be the presence of the strong G4 structure at CSBII. This structure could pose a block to polymerases, which would obscure interpretations of the primer extension experiments that have been used to map RNA to DNA transitions in the past.

4. Conclusions

The presence of a mitochondrial genome was demonstrated over 50 years ago and using recombinant factors *in vitro*, important aspects of transcription and mtDNA replication have over the years been reconstituted. The papers presented in this thesis have continued the molecular mechanistic studies of these processes and our work has established and expanded existing models for transcription initiation and elongation as well as for OriH dependent replication initiation.

Through Paper I and II, and by the support of other reports, we have proposed a sequential model for transcription initiation. In this model TFAM recruits POLRMT to the promoter, relieving the negative regulatory effect of the NTE. In the following step TFB2M enters the transcription initiation complex, which leads to promoter melting over the transcription start site and transcription initiation.

In Paper III we were able to establish TEFM as a part of the core mitochondrial transcription machinery. We found that POLRMT was unable to transcribe the mitochondrial genome by itself, and that it required the presence of this transcription elongation factor. We were additionally able to elucidate aspects of the TEFM-mechanism, as it increases the affinity of the polymerase for the DNA template, thus making POLRMT less prone to transcription termination upon a stall on the template.

In Paper IV of this thesis we could reconstitute replication initiation from an LSP derived primer. We concluded that this process involves an R-loop anchored to the CSB region via G4 structures, but that this R-loop has to be processed by RNaseH1 to work as a primer for replication. The observed initiation events were stimulated by mtSSB, but abolished by TEFM. Aside from its function as a replicative primer we could also demonstrate that the CSB region R-loop could act as a regulatory structure, by shutting down LSP transcription on replicating molecules, and thereby diminishing the risk of a transcribing POLRMT interfering with primer formation or other replication events near OriH.

5. Future Perspectives

We have demonstrated that the NTE of POLRMT is non-essential for the basic transcription reaction, but that it functions as a molecular brake inhibiting non-specific transcription initiations. However, additional roles for this domain likely exist *in vivo*. It would therefore be of great interest to see how an NTE deleted POLRMT behaved in living cells. These types of experiments could also potentially present an indication of the function of the POLRMT PPR motifs. Further, expression of the NTE as an individual polypeptide could show if this domain can function even when not covalently attached to the polymerase.

The emerging technical advances in structural determination of large protein nucleic acid complexes by single-particle cryo-electron microscopy may be used to obtain more structural information for the mitochondrial transcription machinery than the simpler structures that has been determined thus far. For example, a structure of the complete initiation complex could provide important information on transcription factor-polymerase interactions as well as protein-DNA specific contacts that has not yet been identified by present methods. In addition, a possible structure of an elongation complex in the presence of TEFM could further elucidate the mechanism of action of this elongation factor.

The identification of TEFM and the fact that this protein abolishes R-loop formation and replication initiation call for further studies of the regulation of this factor. An interesting possibility would be to identify TEFM-interaction partners, possibly through the use of modern mass spectrophotometric methods. It would also be of interest to investigate the action of TEFM on transcription from individual promoters, especially during replicating conditions. To this end, one could for instance analyze TEFM and POLRMT occupancy on mtDNA during the recovery phase after mtDNA depletion.

The reconstitution of LSP transcription-dependent replication initiation is intriguing. Our continued work will aim to identify which of the start sites are used *in vivo*, and if these differ from those observed *in vitro*, try to uncover how this discrepancy is achieved. For future experiments we will also aim to fully reconstitute mitochondrial replication *in vitro*, which will necessitate the presence of TWINKLE for successful replication of longer stretches of a dsDNA template. If fully reconstituted replication is accomplished, further studies can be undertaken regarding issues such as topology, termination of replication, ligation and decatenation.

Acknowledgements

First of all I would like to express my deepest gratitude to my supervisor **Claes Gustafsson**. I have learnt so much from you regarding science, teaching, writing and the academic world. But most of all, I am thankful for the freedom I have had during my time as a PhD student, the fact that your door has always been open, the enthusiasm you bring whenever something interesting has come up, and the contagious optimism you show even when one cannot find a solution to a scientific problem.

I also want to express my gratitude to **Maria Falkenberg**. You have been involved in most of my projects and I have always felt that discussing scientific problems with you has been rewarding. You have also been my main source of knowledge when it comes to experimental problems.

I would also like to thank my co-supervisor **Per Elias**, especially for making lab-meetings and journal clubs more rewarding by sharing your knowledge and asking sharp questions.

Next, I would like to thank all my co-authors that have contributed to the work presented in this thesis: **Emily Hoberg, Anke Dierckx, Saba Shahzad, Camilla Koolmeister, Nils-Göran Larsson, Marcus Wilhelmsson, Martin Hällberg, Ali Al-Behadili** and **Jay Uhler**.

Further I would like to thank **Tom Nicholls** for helping me with comments and suggestions during the writing of this thesis.

I would also like to thank all the people that have worked in the labs of Claes Gustafsson and Maria Falkenberg during these years. All of you have made this lab into such a good social and scientific environment. There have always been plenty of people to ask for help with problems, to discuss scientific issues with and to do social activities with.

Last but definitely not least, I would like to thank my family and friends for their support, especially my closest family, **Amila, mamma, pappa** and **Emma**. Without the four of you nothing of this would have been possible.

References

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.

Akman, G., Desai, R., Bailey, L.J., Yasukawa, T., Dalla Rosa, I., Durigon, R., Holmes, J.B., Moss, C.F., Mennuni, M., Houlden, H., Crouch, R.J., Hanna, M.G., Pitceathly, R.D., Spinazzola, A. and Holt, I.J. (2016). Pathological ribonuclease H1 causes R-loop depletion and aberrant DNA segregation in mitochondria. *Proc Natl Acad Sci U S A* 113, E4276-4285.

Alberts, B. (2008). *Molecular biology of the cell*. New York, Garland Science.

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., Schreier, P.H., Smith, A.J., Staden, R. and Young, I.G. (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290, 457-465.

Andersson, S.G., Karlberg, O., Canback, 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.

Arnold, J.J., Sharma, S.D., Feng, J.Y., Ray, A.S., Smidansky, E.D., Kireeva, M.L., Cho, A., Perry, J., Vela, J.E., Park, Y., Xu, Y., Tian, Y., Babusis, D., Barauskus, O., Peterson, B.R., Gnatt, A., Kashlev, M., Zhong, W. and Cameron, C.E. (2012). Sensitivity of mitochondrial transcription and resistance of RNA polymerase II dependent nuclear transcription to antiviral ribonucleosides. *PLoS Pathog* 8, e1003030.

Baumgart, B.R., Wang, F., Kwagh, J., Storck, C., Euler, C., Fuller, M., Simic, D., Sharma, S., Arnold, J.J., Cameron, C.E., Van Vleet, T.R., Flint, O., Bunch, R.T., Davies, M.H., Graziano, M.J. and Sanderson, T.P. (2016). Effects of BMS-986094, a Guanosine Nucleotide Analogue, on Mitochondrial DNA Synthesis and Function. *Toxicol Sci* 153, 396-408.

Berg, J.M., Tymoczko, J.L. and Stryer, L. (2012). *Biochemistry*. New York, W. H. Freeman.

Berk, A.J. and Clayton, D.A. (1974). Mechanism of mitochondrial DNA replication in mouse L-cells: asynchronous replication of strands, segregation of circular daughter molecules, aspects of topology and turnover of an initiation sequence. *J Mol Biol* *86*, 801-824.

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. (2012). Mitochondrial DNA nucleoid structure. *Biochim Biophys Acta* *1819*, 914-920.

Bowmaker, M., Yang, M.Y., Yasukawa, T., Reyes, A., Jacobs, H.T., Huberman, J.A. and Holt, I.J. (2003). Mammalian mitochondrial DNA replicates bidirectionally from an initiation zone. *J Biol Chem* *278*, 50961-50969.

Brieba, L.G. and Sousa, R. (2001). T7 promoter release mediated by DNA scrunching. *EMBO J* *20*, 6826-6835.

Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. and Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* *34*, 5402-5415.

Camara, Y., Asin-Cayuela, J., Park, C.B., Metodiev, M.D., Shi, Y., Ruzzenente, B., Kukat, C., Habermann, B., Wibom, R., Hultenby, K., Franz, T., Erdjument-Bromage, H., Tempst, P., Hallberg, B.M., Gustafsson, C.M. and Larsson, N.G. (2011). MTERF4 regulates translation by targeting the methyltransferase NSUN4 to the mammalian mitochondrial ribosome. *Cell Metab* *13*, 527-539.

Carrodeguas, J.A., Kobayashi, R., Lim, S.E., Copeland, W.C. and Bogenhagen, D.F. (1999). The accessory subunit of *Xenopus laevis* mitochondrial DNA polymerase gamma increases processivity of the catalytic subunit of human DNA polymerase gamma and is related to class II aminoacyl-tRNA synthetases. *Mol Cell Biol* *19*, 4039-4046.

Carrodeguas, J.A., Theis, K., Bogenhagen, D.F. and Kisker, C. (2001). Crystal structure and deletion analysis show that the accessory subunit of mammalian DNA polymerase gamma, Pol gamma B, functions as a homodimer. *Mol Cell* *7*, 43-54.

Cerritelli, S.M., Frolova, E.G., Feng, C., Grinberg, A., Love, P.E. and Crouch, R.J. (2003). Failure to produce mitochondrial DNA results in embryonic lethality in *Rnaseh1* null mice. *Mol Cell* *11*, 807-815.

- 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.
- Chang, D.D., Hauswirth, W.W. and Clayton, D.A. (1985). Replication priming and transcription initiate from precisely the same site in mouse mitochondrial DNA. *EMBO J* 4, 1559-1567.
- Cheetham, G.M., Jeruzalmi, D. and Steitz, T.A. (1999). Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* 399, 80-83.
- Cheetham, G.M. and Steitz, T.A. (1999). Structure of a transcribing T7 RNA polymerase initiation complex. *Science* 286, 2305-2309.
- Chen, H. and Chan, D.C. (2004). Mitochondrial dynamics in mammals. *Curr Top Dev Biol* 59, 119-144.
- Chomyn, A., Cleeter, M.W., Ragan, C.I., Riley, M., Doolittle, R.F. and Attardi, G. (1986). URF6, last unidentified reading frame of human mtDNA, codes for an NADH dehydrogenase subunit. *Science* 234, 614-618.
- Chomyn, A., Mariottini, P., Cleeter, M.W., Ragan, C.I., Matsuno-Yagi, A., Hatefi, Y., Doolittle, R.F. and Attardi, G. (1985). Six unidentified reading frames of human mitochondrial DNA encode components of the respiratory-chain NADH dehydrogenase. *Nature* 314, 592-597.
- Clayton, D.A. (1982). Replication of animal mitochondrial DNA. *Cell* 28, 693-705.
- Clayton, D.A. (2003). Mitochondrial DNA replication: What we know. *Iubmb Life* 55, 213-217.
- Copeland, W.C. (2014). Defects of mitochondrial DNA replication. *J Child Neurol* 29, 1216-1224.
- Costa, A., Hood, I.V. and Berger, J.M. (2013). Mechanisms for Initiating Cellular DNA Replication. *Annual Review of Biochemistry*, Vol 82 82, 25-+.
- Crampton, D.J., Ohi, M., Qimron, U., Walz, T. and Richardson, C.C. (2006). Oligomeric states of bacteriophage T7 gene 4 primase/helicase. *Journal of Molecular Biology* 360, 667-677.

Dairaghi, D.J., Shadel, G.S. and Clayton, D.A. (1995a). 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.

Dairaghi, D.J., Shadel, G.S. and Clayton, D.A. (1995b). Human mitochondrial transcription factor A and promoter spacing integrity are required for transcription initiation. *Biochim Biophys Acta* 1271, 127-134.

Dasgupta, S., Masukata, H. and Tomizawa, J. (1987). Multiple mechanisms for initiation of ColE1 DNA replication: DNA synthesis in the presence and absence of ribonuclease H. *Cell* 51, 1113-1122.

Diffley, J.F. and Stillman, B. (1991). A close relative of the nuclear, chromosomal high-mobility group protein HMG1 in yeast mitochondria. *Proc Natl Acad Sci U S A* 88, 7864-7868.

Diffley, J.F. and Stillman, B. (1992). DNA binding properties of an HMG1-related protein from yeast mitochondria. *J Biol Chem* 267, 3368-3374.

Dolezal, P., Likic, V., Tachezy, J. and Lithgow, T. (2006). Evolution of the molecular machines for protein import into mitochondria. *Science* 313, 314-318.

Doublet, S., Tabor, S., Long, A.M., Richardson, C.C. and Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* 391, 251-258.

Dunn, J.J. and Studier, F.W. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J Mol Biol* 166, 477-535.

Durniak, K.J., Bailey, S. and Steitz, T.A. (2008). The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. *Science* 322, 553-557.

Ekstrand, M.I., Falkenberg, M., Rantanen, A., Park, C.B., Gaspari, M., Hultenby, K., Rustin, P., Gustafsson, C.M. and Larsson, N.G. (2004). Mitochondrial transcription factor A regulates mtDNA copy number in mammals. *Hum Mol Genet* 13, 935-944.

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.

Farge, G., Holmlund, T., Khvorostova, J., Rofougaran, R., Hofer, A. and Falkenberg, M. (2008). The N-terminal domain of TWINKLE contributes to single-stranded DNA binding and DNA helicase activities. *Nucleic Acids Res* 36, 393-403.

Farge, G., Laurens, N., Broekmans, O.D., van den Wildenberg, S.M., Dekker, L.C., Gaspari, M., Gustafsson, C.M., Peterman, E.J., Falkenberg, M. and Wuite, G.J. (2012). Protein sliding and DNA denaturation are essential for DNA organization by human mitochondrial transcription factor A. *Nat Commun* 3, 1013.

Farge, G., Mehmedovic, M., Baclayon, M., van den Wildenberg, S.M., Roos, W.H., Gustafsson, C.M., Wuite, G.J. and Falkenberg, M. (2014). In vitro-reconstituted nucleoids can block mitochondrial DNA replication and transcription. *Cell Rep* 8, 66-74.

Farr, C.L., Wang, Y. and Kaguni, L.S. (1999). Functional interactions of mitochondrial DNA polymerase and single-stranded DNA-binding protein. Template-primer DNA binding and initiation and elongation of DNA strand synthesis. *J Biol Chem* 274, 14779-14785.

Fernandez-Millan, P., Lazaro, M., Cansiz-Arda, S., Gerhold, J.M., Rajala, N., Schmitz, C.A., Silva-Espina, C., Gil, D., Bernado, P., Valle, M., Spelbrink, J.N. and Sola, M. (2015). The hexameric structure of the human mitochondrial replicative helicase Twinkle. *Nucleic Acids Res* 43, 4284-4295.

Fernandez-Silva, P., Martinez-Azorin, F., Micol, V. and Attardi, G. (1997). The human mitochondrial transcription termination factor (mTERF) is a multizipper protein but binds to DNA as a monomer, with evidence pointing to intramolecular leucine zipper interactions. *EMBO J* 16, 1066-1079.

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

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., Lisowsky, T., Parisi, M.A. and Clayton, D.A. (1992). DNA wrapping and bending by a mitochondrial high mobility group-like transcriptional activator protein. *J Biol Chem* 267, 3358-3367.

- 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.
- Freyer, C., Park, C.B., Ekstrand, M.I., Shi, Y., Khvorostova, J., Wibom, R., Falkenberg, M., Gustafsson, C.M. and Larsson, N.G. (2010). Maintenance of respiratory chain function in mouse hearts with severely impaired mtDNA transcription. *Nucleic Acids Research* 38, 6577-6588.
- Fuste, 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.
- 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.
- Giles, R.E., Blanc, H., Cann, H.M. and Wallace, D.C. (1980). Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A* 77, 6715-6719.
- Gray, M.W., Burger, G. and Lang, B.F. (1999). Mitochondrial evolution. *Science* 283, 1476-1481.
- Greaves, L.C., Reeve, A.K., Taylor, R.W. and Turnbull, D.M. (2012). Mitochondrial DNA and disease. *J Pathol* 226, 274-286.
- 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.
- Gustafsson, C.M., Falkenberg, M. and Larsson, N.G. (2016). Maintenance and Expression of Mammalian Mitochondrial DNA. *Annual Review of Biochemistry*, Vol 85 85, 133-160.
- Hallberg, B.M. and Larsson, N.G. (2014). Making proteins in the powerhouse. *Cell Metab* 20, 226-240.
- 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. *EMBO J* 15, 4767-4774.

He, B., Kukarin, A., Temiakov, D., Chin-Bow, S.T., Lyakhov, D.L., Rong, M., Durbin, R.K. and McAllister, W.T. (1998). Characterization of an unusual, sequence-specific termination signal for T7 RNA polymerase. *J Biol Chem* 273, 18802-18811.

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. *Proc Natl Acad Sci U S A* 112, 9334-9339.

Holt, I.J., Lorimer, H.E. and Jacobs, H.T. (2000). Coupled leading- and lagging-strand synthesis of mammalian mitochondrial DNA. *Cell* 100, 515-524.

Hyvarinen, A.K., Pohjoismaki, 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.

Itoh, T. and Tomizawa, J. (1980). Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. *Proc Natl Acad Sci U S A* 77, 2450-2454.

Jemt, E., Farge, G., Backstrom, S., Holmlund, T., Gustafsson, C.M. and Falkenberg, M. (2011). The mitochondrial DNA helicase TWINKLE can assemble on a closed circular template and support initiation of DNA synthesis. *Nucleic Acids Res* 39, 9238-9249.

Jemt, E., Persson, O., Shi, Y., Mehmedovic, M., Uhler, J.P., Davila Lopez, M., Freyer, C., Gustafsson, C.M., Samuelsson, T. and Falkenberg, M. (2015). Regulation of DNA replication at the end of the mitochondrial D-loop involves the helicase TWINKLE and a conserved sequence element. *Nucleic Acids Res* 43, 9262-9275.

Kakuda, T.N. (2000). Pharmacology of nucleoside and nucleotide reverse transcriptase inhibitor-induced mitochondrial toxicity. *Clin Ther* 22, 685-708.

Kaneda, H., Hayashi, J., Takahama, S., Taya, C., Lindahl, K.F. and Yonekawa, H. (1995). Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. *Proc Natl Acad Sci U S A* 92, 4542-4546.

Kang, D.C., Miyako, K., Kai, Y., Irie, T. and Takeshige, K. (1997). In vivo determination of replication origins of human mitochondrial DNA by ligation-mediated polymerase chain reaction. *Journal of Biological Chemistry* 272, 15275-15279.

- Korhonen, J.A., Gaspari, M. and Falkenberg, M. (2003). TWINKLE Has 5' → 3' DNA helicase activity and is specifically stimulated by mitochondrial single-stranded DNA-binding protein. *J Biol Chem* 278, 48627-48632.
- 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.
- Kornblum, C., Nicholls, T.J., Haack, T.B., Scholer, S., Peeva, V., Danhauser, K., Hallmann, K., Zsurka, G., Rorbach, J., Iuso, A., Wieland, T., Sciacco, M., Ronchi, D., Comi, G.P., Moggio, M., Quinzii, C.M., DiMauro, S., Calvo, S.E., Mootha, V.K., Klopstock, T., Strom, T.M., Meitinger, T., Minczuk, M., Kunz, W.S. and Prokisch, H. (2013). Loss-of-function mutations in MGME1 impair mtDNA replication and cause multisystemic mitochondrial disease. *Nat Genet* 45, 214-219.
- 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.
- Kuhl, I., Miranda, M., Posse, V., Milenkovic, D., Mourier, A., Siira, S.J., Bonekamp, N.A., Neumann, U., Filipovska, A., Polosa, P.L., Gustafsson, C.M. and Larsson, N.G. (2016). POLRMT regulates the switch between replication primer formation and gene expression of mammalian mtDNA. *Sci Adv* 2, e1600963.
- Kukat, C., Davies, K.M., Wurm, C.A., Spahr, H., Bonekamp, N.A., Kuhl, I., Joos, F., Polosa, P.L., Park, C.B., Posse, V., Falkenberg, M., Jakobs, S., Kuhlbrandt, W. and Larsson, N.G. (2015). Cross-strand binding of TFAM to a single mtDNA molecule forms the mitochondrial nucleoid. *Proc Natl Acad Sci U S A* 112, 11288-11293.
- 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.
- Larsson, N.G., Wang, J., Wilhelmsson, H., Oldfors, A., Rustin, P., Lewandoski, M., Barsh, G.S. and Clayton, D.A. (1998). Mitochondrial transcription factor A is necessary for mtDNA maintenance and embryogenesis in mice. *Nat Genet* 18, 231-236.
- Lee, D.Y. and Clayton, D.A. (1996). Properties of a primer RNA-DNA hybrid at the mouse mitochondrial DNA leading-strand origin of replication. *J Biol Chem* 271, 24262-24269.

- Lee, Y.S., Lee, S., Demeler, B., Molineux, I.J., Johnson, K.A. and Yin, Y.W. (2010). Each monomer of the dimeric accessory protein for human mitochondrial DNA polymerase has a distinct role in conferring processivity. *J Biol Chem* *285*, 1490-1499.
- Li, K. and Williams, R.S. (1997). Tetramerization and single-stranded DNA binding properties of native and mutated forms of murine mitochondrial single-stranded DNA-binding proteins. *J Biol Chem* *272*, 8686-8694.
- Linder, T., Park, C.B., Asin-Cayuela, J., Pellegrini, M., Larsson, N.G., Falkenberg, M., Samuelsson, T. and Gustafsson, C.M. (2005). A family of putative transcription termination factors shared amongst metazoans and plants. *Curr Genet* *48*, 265-269.
- 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.
- 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. *J Mol Biol* *280*, 201-213.
- Macao, B., Uhler, J.P., Siibak, T., Zhu, X., Shi, Y., Sheng, W., Olsson, M., Stewart, J.B., Gustafsson, C.M. and Falkenberg, M. (2015). The exonuclease activity of DNA polymerase gamma is required for ligation during mitochondrial DNA replication. *Nat Commun* *6*, 7303.
- Macreadie, I.G., Novitski, C.E., Maxwell, R.J., John, U., Ooi, B.G., McMullen, G.L., Lukins, H.B., Linnane, A.W. and Nagley, P. (1983). Biogenesis of mitochondria: the mitochondrial gene (*aap1*) coding for mitochondrial ATPase subunit 8 in *Saccharomyces cerevisiae*. *Nucleic Acids Res* *11*, 4435-4451.
- 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.
- Manna, S. (2015). An overview of pentatricopeptide repeat proteins and their applications. *Biochimie* *113*, 93-99.
- Martin, C.T., Muller, D.K. and Coleman, J.E. (1988). Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry* *27*, 3966-3974.

- 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.
- Masukata, H. and Tomizawa, J. (1984). Effects of point mutations on formation and structure of the RNA primer for ColE1 DNA replication. *Cell* *36*, 513-522.
- Masukata, H. and Tomizawa, J. (1990). A mechanism of formation of a persistent hybrid between elongating RNA and template DNA. *Cell* *62*, 331-338.
- Matsunaga, M. and Jaehning, J.A. (2004). Intrinsic promoter recognition by a "core" RNA polymerase. *J Biol Chem* *279*, 44239-44242.
- Meisinger, C., Sickmann, A. and Pfanner, N. (2008). The mitochondrial proteome: from inventory to function. *Cell* *134*, 22-24.
- Metodiev, M.D., Lesko, N., Park, C.B., Camara, 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.
- Minczuk, M., He, J., Duch, A.M., Etema, T.J., Chlebowski, 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 Fuste, J., Shi, Y., Wanrooij, S., Zhu, X., Jemt, E., Persson, O., Sabouri, N., Gustafsson, C.M. and Falkenberg, M. (2014). In vivo occupancy of mitochondrial single-stranded DNA binding protein supports the strand displacement mode of DNA replication. *PLoS Genet* *10*, e1004832.
- 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., Elmlund, H., Kalisman, N., Bushnell, D.A., Adams, C.M., Azubel, M., Elmlund, D., Levi-Kalisman, Y., Liu, X., Gibbons, B.J., Levitt, M. and Kornberg, R.D. (2013). Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* *342*, 1238724.

Nass, M.M. and Nass, S. (1963a). Intramitochondrial Fibers with DNA Characteristics. I. Fixation and Electron Staining Reactions. *J Cell Biol* *19*, 593-611.

Nass, S. and Nass, M.M. (1963b). Intramitochondrial Fibers with DNA Characteristics. II. Enzymatic and Other Hydrolytic Treatments. *J Cell Biol* *19*, 613-629.

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. *Nature Communications* *5*.

Nicholls, T.J., Zsurka, G., Peeva, V., Scholer, S., Szczesny, R.J., Cysewski, D., Reyes, A., Kornblum, C., Sciacco, M., Moggio, M., Dziembowski, A., Kunz, W.S. and Minczuk, M. (2014). Linear mtDNA fragments and unusual mtDNA rearrangements associated with pathological deficiency of MGME1 exonuclease. *Hum Mol Genet* *23*, 6147-6162.

Nouws, J., Goswami, A.V., Bestwick, M., McCann, B.J., Surovtseva, Y.V. and Shadel, G.S. (2016). Mitochondrial Ribosomal Protein L12 Is Required for POLRMT Stability and Exists as Two Forms Generated by Alternative Proteolysis during Import. *J Biol Chem* *291*, 989-997.

Oakley, J.L., Strothkamp, R.E., Sarris, A.H. and Coleman, J.E. (1979). T7 RNA polymerase: promoter structure and polymerase binding. *Biochemistry* *18*, 528-537.

Ohmori, H., Murakami, Y. and Nagata, T. (1987). Nucleotide sequences required for a ColE1-type plasmid to replicate in *Escherichia coli* cells with or without RNase H. *J Mol Biol* *198*, 223-234.

- Ojala, D. and Attardi, G. (1974). Identification of Discrete Polyadenylate-Containing Rna Components Transcribed from Hela-Cell Mitochondrial-DNA. *Proceedings of the National Academy of Sciences of the United States of America* *71*, 563-567.
- Ojala, D., Montoya, J. and Attardi, G. (1981). Transfer-Rna Punctuation Model of Rna Processing in Human Mitochondria. *Nature* *290*, 470-474.
- Palade, G.E. (1952). The fine structure of mitochondria. *Anat Rec* *114*, 427-451.
- Palade, G.E. (1953). An electron microscope study of the mitochondrial structure. *J Histochem Cytochem* *1*, 188-211.
- 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.
- Parisi, M.A. and Clayton, D.A. (1991). Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. *Science* *252*, 965-969.
- Park, C.B., Asin-Cayuela, J., Camara, Y., Shi, Y., Pellegrini, M., Gaspari, M., Wibom, R., Hultenby, K., Erdjument-Bromage, H., Tempst, P., Falkenberg, M., Gustafsson, C.M. and Larsson, N.G. (2007). MTERF3 is a negative regulator of mammalian mtDNA transcription. *Cell* *130*, 273-285.
- 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.
- Piko, L. and Taylor, K.D. (1987). Amounts of mitochondrial DNA and abundance of some mitochondrial gene transcripts in early mouse embryos. *Dev Biol* *123*, 364-374.
- Reyes, A., Kazak, L., Wood, S.R., Yasukawa, T., Jacobs, H.T. and Holt, I.J. (2013). Mitochondrial DNA replication proceeds via a 'bootlace' mechanism involving the incorporation of processed transcripts. *Nucleic Acids Res* *41*, 5837-5850.

- Reyes, A., Melchionda, L., Nasca, A., Carrara, F., Lamantea, E., Zanolini, A., Lamperti, C., Fang, M., Zhang, J., Ronchi, D., Bonato, S., Fagiolari, G., Moggio, M., Ghezzi, D. and Zeviani, M. (2015). RNASEH1 Mutations Impair mtDNA Replication and Cause Adult-Onset Mitochondrial Encephalomyopathy. *Am J Hum Genet* *97*, 186-193.
- Reyes, A., Yang, M.Y., Bowmaker, M. and Holt, I.J. (2005). Bidirectional replication initiates at sites throughout the mitochondrial genome of birds. *J Biol Chem* *280*, 3242-3250.
- 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.
- 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.
- Robinson, P.J., Trnka, M.J., Bushnell, D.A., Davis, R.E., Mattei, P.J., Burlingame, A.L. and Kornberg, R.D. (2016). Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* *166*, 1411-1422 e1416.
- Rubio-Cosials, A., Sidow, J.F., Jimenez-Menendez, N., Fernandez-Millan, P., Montoya, J., Jacobs, H.T., Coll, M., Bernado, P. and Sola, M. (2011). Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter. *Nat Struct Mol Biol* *18*, 1281-1289.
- Saraste, M. (1999). Oxidative phosphorylation at the fin de siecle. *Science* *283*, 1488-1493.
- Shadel, G.S. and Clayton, D.A. (1997). Mitochondrial DNA maintenance in vertebrates. *Annual Review of Biochemistry* *66*, 409-435.
- 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.
- Sheftel, A.D. and Lill, R. (2009). The power plant of the cell is also a smithy: the emerging role of mitochondria in cellular iron homeostasis. *Ann Med* *41*, 82-99.

Shi, Y., Dierckx, A., Wanrooij, P.H., Wanrooij, S., Larsson, N.G., Wilhelmsson, L.M., Falkenberg, M. and Gustafsson, C.M. (2012). Mammalian transcription factor A is a core component of the mitochondrial transcription machinery. *Proc Natl Acad Sci U S A* *109*, 16510-16515.

Shi, Y., Posse, V., Zhu, X., Hyvarinen, 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.

Shmookler Reis, R.J. and Goldstein, S. (1983). Mitochondrial DNA in mortal and immortal human cells. Genome number, integrity, and methylation. *J Biol Chem* *258*, 9078-9085.

Shutt, T.E. and Gray, M.W. (2006a). Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends in Genetics* *22*, 90-95.

Shutt, T.E. and Gray, M.W. (2006b). Homologs of mitochondrial transcription factor B, sparsely distributed within the eukaryotic radiation, are likely derived from the dimethyladenosine methyltransferase of the mitochondrial endosymbiont. *Mol Biol Evol* *23*, 1169-1179.

Sjostrand, F.S. (1953). Electron microscopy of mitochondria and cytoplasmic double membranes. *Nature* *171*, 30-32.

Sologub, M., Litonin, D., Anikin, M., Mustaev, A. and Temiakov, D. (2009). TFB2 is a transient component of the catalytic site of the human mitochondrial RNA polymerase. *Cell* *139*, 934-944.

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*, 593-599.

Spahr, H., Habermann, B., Gustafsson, C.M., Larsson, N.G. and Hallberg, B.M. (2012). Structure of the human MTERF4-NSUN4 protein complex that regulates mitochondrial ribosome biogenesis. *Proc Natl Acad Sci U S A* *109*, 15253-15258.

Spahr, H., Samuelsson, T., Hallberg, B.M. and Gustafsson, C.M. (2010). Structure of mitochondrial transcription termination factor 3 reveals a novel nucleic acid-binding domain. *Biochem Biophys Res Commun* *397*, 386-390.

Spelbrink, J.N., Li, F.Y., Tiranti, V., Nikali, K., Yuan, Q.P., Tariq, M., Wanrooij, S., Garrido, N., Comi, G., Morandi, L., Santoro, L., Toscano, A., Fabrizi, G.M., Somer, H., Croxen, R., Beeson, D., Poulton, J., Suomalainen, A., Jacobs, H.T., Zeviani, M. and Larsson, C. (2001). Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria. *Nat Genet* 28, 223-231.

Steitz, T.A., Smerdon, S.J., Jager, J. and Joyce, C.M. (1994). A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science* 266, 2022-2025.

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, 43-50.

Tapper, D.P. and Clayton, D.A. (1981). Mechanism of replication of human mitochondrial DNA. Localization of the 5' ends of nascent daughter strands. *J Biol Chem* 256, 5109-5115.

Teixeira, P.F. and Glaser, E. (2013). Processing peptidases in mitochondria and chloroplasts. *Biochim Biophys Acta* 1833, 360-370.

Terzioglu, M., Ruzzenente, B., Harmel, J., Mourier, A., Jemt, E., Lopez, M.D., Kukat, C., Stewart, J.B., Wibom, R., Meharg, C., Habermann, B., Falkenberg, M., Gustafsson, C.M., Park, C.B. and Larsson, N.G. (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.

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.

Trifunovic, A., Wredenberg, A., Falkenberg, M., Spelbrink, J.N., Rovio, A.T., Bruder, C.E., Bohlooly, Y.M., Gidlof, S., Oldfors, A., Wibom, R., Tornell, J., Jacobs, H.T. and Larsson, N.G. (2004). Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature* 429, 417-423.

Twig, G., Hyde, B. and Shirihai, O.S. (2008). Mitochondrial fusion, fission and autophagy as a quality control axis: the bioenergetic view. *Biochim Biophys Acta* *1777*, 1092-1097.

Uhler, J.P., Thorn, C., Nicholls, T.J., Matic, S., Milenkovic, D., Gustafsson, C.M. and Falkenberg, M. (2016). MGME1 processes flaps into ligatable nicks in concert with DNA polymerase gamma during mtDNA replication. *Nucleic Acids Res* *44*, 5861-5871.

Vannini, A. and Cramer, P. (2012). Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell* *45*, 439-446.

Vijayarathy, C. and Avadhani, N.G. (1996). Site specific termination of the mouse mitochondrial H strand transcription. *Faseb Journal* *10*, 2566-2566.

von Heijne, G. (1986). Why mitochondria need a genome. *FEBS Lett* *198*, 1-4.

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, P.H., Uhler, J.P., Shi, Y., Westerlund, F., Falkenberg, M. and Gustafsson, C.M. (2012a). A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. *Nucleic Acids Res* *40*, 10334-10344.

Wanrooij, P.H., Uhler, J.P., Simonsson, T., Falkenberg, M. and Gustafsson, C.M. (2010). G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. *Proc Natl Acad Sci U S A* *107*, 16072-16077.

Wanrooij, S., Fuste, 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.

Wanrooij, S., Miralles Fuste, J., Stewart, J.B., Wanrooij, P.H., Samuelsson, T., Larsson, N.G., Gustafsson, C.M. and Falkenberg, M. (2012b). In vivo mutagenesis reveals that OriL is essential for mitochondrial DNA replication. *EMBO Rep* *13*, 1130-1137.

Wredenberg, A., Lagouge, M., Bratic, A., Metodiev, M.D., Spahr, H., Mourier, A., Freyer, C., Ruzzenente, B., Tain, L., Gronke, S., Baggio, F., Kukat, C., Kremmer, E., Wibom, R., Polosa, P.L., Habermann, B., Partridge, L., Park, C.B. and Larsson, N.G. (2013). MTERF3 regulates mitochondrial ribosome biogenesis in invertebrates and mammals. *PLoS Genet* 9, e1003178.

Xu, B. and Clayton, D.A. (1995). A persistent RNA-DNA hybrid is formed during transcription at a phylogenetically conserved mitochondrial DNA sequence. *Mol Cell Biol* 15, 580-589.

Xu, B. and Clayton, D.A. (1996). RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers. *EMBO J* 15, 3135-3143.

Xu, B.J. and Clayton, D.A. (1992). Assignment of a Yeast Protein Necessary for Mitochondrial Transcription Initiation. *Nucleic Acids Research* 20, 1053-1059.

Yakubovskaya, E., Chen, Z., Carrodeguas, J.A., Kisker, C. and Bogenhagen, D.F. (2006). Functional human mitochondrial DNA polymerase gamma forms a heterotrimer. *J Biol Chem* 281, 374-382.

Yakubovskaya, E., Guja, K.E., Mejia, E., Castano, S., Hambardjjeva, E., Choi, W.S. and Garcia-Diaz, M. (2012). Structure of the essential MTERF4:NSUN4 protein complex reveals how an MTERF protein collaborates to facilitate rRNA modification. *Structure* 20, 1940-1947.

Yang, C., Curth, U., Urbanke, C. and Kang, C. (1997). Crystal structure of human mitochondrial single-stranded DNA binding protein at 2.4 Å resolution. *Nat Struct Biol* 4, 153-157.

Yang, M.Y., Bowmaker, M., Reyes, A., Vergani, L., Angeli, P., Gringeri, E., Jacobs, H.T. and Holt, I.J. (2002). Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. *Cell* 111, 495-505.

Yasukawa, T., Reyes, A., Cluett, T.J., Yang, M.Y., Bowmaker, M., Jacobs, H.T. and Holt, I.J. (2006). Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. *EMBO J* 25, 5358-5371.

Youle, R.J. and van der Bliek, A.M. (2012). Mitochondrial fission, fusion, and stress. *Science* 337, 1062-1065.