Mitochondrial DNA replication in mammalian cells: overview of the pathway

Maria Falkenberg
Department of Medical Biochemistry and Cell Biology, University of Gothenburg, P.O. Box 440, 405 30 Gothenburg, Sweden

Correspondence: Maria Falkenberg (maria.falkenberg@medkem.gu.se)

Mammalian mitochondria contain multiple copies of a circular, double-stranded DNA genome and a dedicated DNA replication machinery is required for its maintenance. Many disease-causing mutations affect mitochondrial replication factors and a detailed understanding of the replication process may help to explain the pathogenic mechanisms underlying a number of mitochondrial diseases. We here give a brief overview of DNA replication in mammalian mitochondria, describing our current understanding of this process and some unanswered questions remaining.

Introduction
Mitochondrial DNA (mtDNA) is a double-stranded molecule of 16.6 kb (Figure 1, lower panel). The two strands of mtDNA differ in their base composition, with one being rich in guanines, making it possible to separate a heavy (H) and a light (L) strand by density centrifugation in alkaline CsCl2 gradients [1]. The mtDNA contains one longer noncoding region (NCR) also referred to as the control region. In the NCR, there are promoters for polycistronic transcription, one for each mtDNA strand; the light strand promoter (LSP) and the heavy strand promoter (HSP). The NCR also harbors the origin for H-strand DNA replication (O_H). A second origin for L-strand DNA replication (O_L) is located outside the NCR, within a tRNA cluster approximately 11,000 bp downstream of O_H.

mtDNA replication factors
Mammalian mtDNA is replicated by proteins distinct from those used for nuclear DNA replication and many are related to replication factors identified in bacteriophages [2]. DNA polymerase γ (POLγ) is the replicative polymerase in mitochondria. In human cells, POLγ is a heterotrimer with one catalytic subunit (POLγA) and two accessory subunits (POLγB) [3-5]. Mouse knockouts for POLγA and POLγB have revealed that both factors are essential for embryonic development [6,7]. At least four additional polymerases (PrimPol, DNA polymerase β, DNA polymerase θ, and DNA polymerase ζ) have been reported to play a role mitochondria [8-11]. These polymerases are not essential for mtDNA maintenance and none of them can substitute for POLγ. Most likely they are involved in mtDNA repair, but the exact function of these additional polymerases in mtDNA maintenance needs to be further elucidated (reviewed in [12]).

POLγA belongs to the family A DNA polymerases and contains a 3′-5′ exonuclease domain that acts to proofread the newly synthesized DNA strand [3]. POLγ is a highly accurate DNA polymerase with a frequency of misincorporation lower than 1 x 10⁻⁶ [13]. The accessory POLγB subunit enhances interactions with the DNA template and increases both the catalytic activity and the processivity of POLγA [14-16]. POLγ cannot use double-stranded DNA as a template and a DNA helicase is therefore required at the mitochondrial replication fork [17]. The DNA helicase TWINKLE is homologous to the T7 phage gene 4 protein [18] and during mtDNA replication, TWINKLE travels in front of POLγ, unwinding the double-stranded DNA template. TWINKLE forms a hexamer and requires a fork structure...
Figure 1. Map of human mtDNA
The genome encodes for 13 mRNA (green), 22 tRNA (violet), and 2 rRNA (pale blue) molecules. There is also a major noncoding region (NCR), which is shown enlarged at the top. The major NCR contains the heavy strand promoter (HSP), the light strand promoter (LSP), three conserved sequence boxes (CSB1-3, orange), the H-strand origin of replication (O_H), and the termination-associated sequence (TAS, yellow). The triple-stranded displacement-loop (D-loop) structure is formed by premature termination of nascent H-strand DNA synthesis at TAS. The short H-strand replication product formed in this manner is termed 7S DNA. A minor NCR, located approximately 11,000 bp downstream of O_H, contains the L-strand origin of replication (O_L).

(a single-stranded 5′-DNA loading site and a short 3′-tail) to load and initiate unwinding [18-20]. Mitochondrial single-stranded DNA-binding protein (mtSSB) binds to the formed ssDNA, protects it against nucleases, and prevents secondary structure formation [21,22]. mtSSB enhances mtDNA synthesis by stimulating TWINKLE’s helicase activity as well as increasing the processivity of POLγ [17,19,23].

The mode of mtDNA replication
A model for mtDNA replication was presented already in 1972 by Vinograd and co-workers (Figure 2) [24]. According their strand displacement model, DNA synthesis is continuous on both the H- and L-strand [25]. There is a dedicated origin for each strand, O_H and O_L. First, replication is initiated at O_H and DNA synthesis then proceeds to produce a new H-strand. During the initial phase, there is no simultaneous L-strand synthesis and mtSSB covers the displaced, parental H-strand [26]. By binding to single-stranded DNA, mtSSB prevents the mitochondrial RNA polymerase (POLRMT) from initiating random RNA synthesis on the displaced strand [27]. When the replication fork has progressed about two-thirds of the genome, it passes the second origin of replication, O_L. When exposed in its single-stranded conformation, the parental H-strand at O_L folds into a stem–loop structure [28]. The stem efficiently blocks mtSSB from binding and a short stretch of single-stranded DNA in the loop region therefore remains accessible, allowing POLRMT to initiate RNA synthesis [26,29]. POLRMT is not processive on a single-stranded DNA templates [27]. Already after about 25 nt, it is replaced by POLγ and L-strand DNA synthesis is initiated [30]. From this point, H- and L-strand synthesis proceeds continuously until the two strands have reached full circle. Replication of the two strands is linked, since H-strand synthesis is required for initiation of L-strand synthesis. The structure and the sequence requirement for mammalian O_L has been studied both in vivo and in vitro, demonstrating that a functional human O_L must include a stable double-stranded stem region with pyrimidine-rich template strand and a single-stranded loop of at least 10 nt [31].
Figure 2. Replication of the human mitochondrial genome

Mitochondrial DNA replication is initiated at O₄ and proceeds unidirectionally to produce the full-length nascent H-strand. mtSSB binds and protects the exposed, parental H-strand. When the replisome passes O₅, a stem–loop structure is formed that blocks mtSSB binding, presenting a single-stranded loop-region from which POLRMT can initiate primer synthesis. The transition to L-strand DNA synthesis takes place after about 25 nt, when POLY replaces POLRMT at the 3'–end of the primer. Synthesis of the two strands proceeds in a continuous manner until two full, double-stranded DNA molecules have been formed.

It should be noted that some aspects of the strand-displacement model have been questioned. For instance, studies have suggested that processed RNA molecules hybridize to the single-stranded H-strand and function as a provisional lagging strand, which is replaced by DNA during later stages of mtDNA replication [32]. This so-called RITOLS model implicates that processed transcripts are successively hybridized to the paternal H-strand as the replication fork advances, but the enzymatic machinery required for this process has not been identified [33]. Arguing against the RITOLS model, single-stranded DNA binding proteins are used to stabilize single-stranded DNA intermediates during DNA replication in all three major branches of life. In mitochondria, there are at least 500 mtSSB tetramers available per mtDNA molecule. Since each tetramer binds 59 nt, the levels of mtSSB are sufficient to cover the entire parental H-strand during mtDNA synthesis [26,34]. In addition, strand-specific chromatin immunoprecipitation has revealed that mtSSB exclusively covers the parental H-strand during mtDNA replication in vivo. The occupancy profile displays a distinct pattern, with the highest levels of mtSSB close to Orih, followed by a gradual decline toward OriL. The pattern is thus as would be predicted if mtSSB functions to stabilize the single-stranded, parental H-strand during strand-displacement DNA replication [26]. Yet another problem for the RITOLS model is the presence of RNASEH1 in mitochondria. This enzyme efficiently removes RNA–DNA hybrids, an activity that is difficult to reconcile with processed RNA molecules stably binding to long stretches of single-stranded DNA during mtDNA replication [35].

Finally, there have been reports suggesting that under certain conditions, strand-coupled replication may function as a backup replication mode in mammalian mitochondria [36,37]. The molecular mechanisms underlying this type of replication have not been elucidated.

The D-loop

Curiously, not all replication events initiated at O₄ continue to full circle. Instead, 95% are terminated already after about 650 nt at the termination associated sequences (TAS) [38,39]. The short DNA fragment formed in this way, 7S DNA, remains bound to the parental L-strand, while the parental H-strand is displaced (Figure 1, top panel). As a result, a triple-stranded displacement loop structure, a D-loop, is formed. The functional importance of the D-loop structure is unclear and how replication is terminated at TAS is also not known [40,41]. It appears however that termination at TAS is a regulated event, providing a switch between abortive and genome length mtDNA replication [42]. In support of this notion, in vivo occupation analysis revealed that POLY under normal conditions stalls at the 3'-end of the D-loop, whereas TWINKLE occupancy is low in this region. When mtDNA is depleted, the situation changes and TWINKLE occupancy increases and at the same time 7S DNA levels are decreased. These data have been interpreted as evidence for TWINKLE reloading in response to increased demand for mtDNA replication [42]. Binding of the helicase to the 3'-end of 7S DNA would allow the stalled POLY to continue replication of 7S DNA to full circle. The model receives support from mouse genetic experiments, demonstrating that TWINKLE is important for mtDNA copy number control. Increased or decreased levels of TWINKLE correlate nicely with mtDNA levels.
[43-46]. It is thus possible that mtDNA replication is regulated at the level of pretermination rather than initiation. The switch may fine-tune the mtDNA copy number in response to cellular demands.

Interestingly, two closely related 15 nt and evolutionary conserved palindromic sequence motifs (ATGN₃CAT) are located on each side of the D-loop region. One motif is located just upstream of the 5’-end of the 7S DNA, where it forms a part of conserved sequence box 1, CSB1. The second motif, core-TAS, is located within the TAS region, just downstream of the 3’-end of 7S DNA (Figure 1, top panel) [42]. The physiological role of these motifs remains unclear, but sequence-specific DNA binding proteins often recognize and bind palindromic sequences. In support of this notion, there are published in organello footprints located to the TAS region [47], but despite substantial efforts in different laboratories, a TAS-binding protein has so far not been identified. It is possible that the proteins binding to CSB1, core-TAS, and other regions within TAS are difficult to purify by traditional methods. Perhaps the missing protein is membrane bound and difficult to retain in solution during chromatography. Alternatively, binding could be a regulated event, requiring precise redox conditions or nucleotide concentrations. Finally, it cannot be excluded that secondary structures in mtDNA may play a role, e.g. stem–loops or G-quadruplexes, which could also contribute to the observed in vivo DNA footprints.

**Initiation of mtDNA replication at Oₕ**

We know that POLRMT forms the primers necessary to initiate H-strand synthesis Oₕ [48-51]. Transcripts initiated at LSP provide RNA 3’-ends from which POLγ can initiate DNA synthesis. In human mitochondria, there are multiple transitions points (RNA-to-DNA transitions) located downstream of LSP, clustering around two conserved sequence motifs, CSB3 and CSB2 (Figure 1, top panel) [52-54]. These conserved sequence elements are guanine-rich and during transcription, a G-quadruplex structure can form between nascent RNA and the nontemplate DNA strand at CSB2. In this manner, the nascent transcript is anchored to mtDNA, forming an R-loop structure [30,55]. The G-quadruplex structure also causes premature transcription termination at sites roughly corresponding to RNA to DNA transition sites mapped in the CSB2-region [56]. Based on these observations, it was hypothesized that sequence-dependent transcription termination may be responsible for primer formation at Oₕ [54-56]. The transcription elongation factor TEFM, strongly reduces transcription termination and R-loop formation at CSB2, leading to the suggestion that active TEFM may influence the ratio between primer formation and full-length, productive transcription [57,58]. Arguing against this idea, knockdown of TEFM in cells only has very limited effects on mtDNA copy number and mitochondrial replication intermediates [59]. In addition, there are no direct experimental evidence demonstrating that R-loop-forming, prematurely terminated transcripts can be directly used by POLγ to initiate DNA synthesis. Further experiments are clearly needed to define the precise role of R-loops and TEFM in replication initiation.

The mechanisms of DNA replication initiation at Oₕ may resemble those previously described for initiation of DNA replication in the *E. coli* plasmid CoIE1. In the plasmid, a transcript denoted RNAII associates with the template strand, forming an R-loop that is used to prime DNA synthesis [60,61]. Furthermore, the CoIE1 origin of replication is situated downstream of a guanine-rich stretch that is essential for both replication initiation and R-loop formation [61,62]. In CoIE1, the R-loop is cleaved by RNase H, before it is used to prime DNA synthesis. If the mitochondrial RNASEH1 plays a similar role in mammalian cells remains to be determined [63,64].

**Termination of mtDNA replication**

When POLγ has completed synthesis, the newly formed DNA strands are ligated by DNA ligase III [65,66]. To allow for efficient ligation, the 5’- and 3’-ends of the nascent DNA ends must be juxtaposed, which means that the RNA primers used to initiate mtDNA synthesis must first be removed [67]. A likely candidate for primer removal is RNASEH1, inasmuch as RNA primers are retained in the mitochondrial origin regions in mouse embryonic fibroblasts lacking Rnaseh1 and there is a loss of mtDNA in *Rnaseh1* knockout mice [35,68].

After completing a full circle-replication, POLγ encounters the 5’-end of the nascent full-length mtDNA strand it has just produced. At this point, POLγ initiates successive cycles of polymerization and 3’-5’ exonuclease degradation at the nick [67,69]. This process, idling, is required for proper ligation. POLγ lacking exonuclease activity is unable to idle and instead continues DNA synthesis into the dsDNA region past the 5’-end, thereby creating a flap-structure that cannot be ligated. Failure to create ligatable DNA ends may explain why mice with exonuclease-deficient POLγ display strand-specific nicks at Oₕ [67,70].

Interestingly, there is a major 5’-end of nascent DNA located approximately 100 bp further downstream of the identified RNA-to-DNA transitions sites. Historically, this site (position 191 in human mtDNA) has been seen as a part of Oₕ, but how this 5’-end is generated is not clear [25]. Although initially identified as a start site for mtDNA replication, the free 5’-end at position 191 may be generated in other ways. For instance, the nascent H-strand may
undergo considerable 5′-end processing during primer removal, removing not only the RNA primer, but also ~100 nt of downstream DNA [42]. In this way, the site for RNA-to-DNA transition would be separated from the site of nascent H-strand ligation at the end of replication. A possible candidate for this effect is the mitochondrial genome maintenance exonuclease 1 (MGME1), a mitochondrial RecB-type exonuclease belonging to the PD-(D/E)KK nuclease superfamily [71,72]. In vitro analysis revealed that MGME1 cuts both ssDNA and DNA flap substrates. Human cells lacking active MGME1 display impaired ligation at O_H and the formation of linear deleted mtDNA molecules spanning O_H and O_L [72]. In addition, there are increased levels of 7S DNA. The 5′-ends of these 7S DNA products are located further upstream (i.e. closer to CSB2) than what is observed in normal cells, suggesting that MGME1 is involved in processing the 5′-end of the nascent H-strand. Further studies of MGME1 and how it works together with RNASEH1 to process nascent replication products will be important.

**Separation mtDNA**

During DNA replication, the parental molecule remains intact, which poses a steric problem for the moving replication machinery. Topoisomerases belonging to the type 1 family can relieve torsional strain formed in this way, by allowing one of the strands to pass through the other. In mammalian mitochondria, TOP1MT a type IB enzyme can act as a DNA "swivel", working together with the mitochondrial replisome [73]. Knockout of the *Top1mt* gene in mouse generates viable offspring that shows altered mtDNA supercoiling [74,75].

In other systems, replication of intact, circular DNA generates daughter molecules linked together as catenanes, i.e. mechanically interlocked, but not yet completely finished DNA circles. Therefore, replication of circular genomes requires decatenation to generate complete daughter molecules separation (Figure 3). The existence of catenanes in mitochondria was reported already in 1967 by Vinograd and co-workers, who identified mtDNA molecules linked together by X-type branches, which they suggested were formed during completion of mtDNA replication [76]. Recently, it was demonstrated that these X-type structures are hemicatenanes, i.e. double-stranded DNA molecules linked together via a single-stranded linkage [77]. A mitochondrial isoform of Topoisomerase 3α (Top3α) is required to resolve the hemicatenane structure (Figure 3), inasmuch as loss of Top3α causes decrease in mtDNA and the formation of large, catenated mtDNA networks. Patient mutations that decrease Top3α activity cause symptoms similar to those

Figure 3. Separation and segregation of the human mitochondrial genome

After mtDNA replication, the new daughter molecules are mechanically linked via a hemicatenane structure, which requires Top3α to be resolved.
caused by mutations in other mitochondrial replication factors, including muscle-restricted mtDNA deletions and chronic progressive external ophthalmoplegia [77]. Interestingly, the hemicatenanes holding these mtDNA networks together are located to the O1t-region, suggesting that these structures are formed during the completion of mtDNA replication. Even if the exact mechanisms remain unclear, previous theoretical work that has postulated ways by which hemicatenanes can be formed during replication of circular DNA molecules [78]. Further work is required to understand how mtDNA replication is terminated and hemicatenanes are formed.

Even if Top3α is required for separation of newly replicated mtDNA, additional proteins are probably also needed. There is a nuclear isoform of Top3α that functions together with three other proteins; the helicase BLM and the OB-fold proteins RM11 and RM12. Together, these proteins form the BTR complex, which acts to dissolve double Holliday junctions, and Top3α requires the other subunits to exert full topoisomerase activity [79]. However, since neither BLM, RM11, or RM12 have mitochondrial isoforms, other proteins may partner with Top3α in mitochondria to regulate and/or stimulate its activity [77].

### Nucleoid replication

mtDNA is not a naked molecule, but packaged into large nucleoprotein complexes, nucleoids [80], which can be visualized by various fluorescent microscopy approaches [81]. Studies have revealed that nucleoids have an average size of ~100 nm in diameter [82,83] and in most cases there is a single mtDNA molecule per nucleoid [82]. The major structural protein component of the nucleoid is TFA, which is present at a ratio of 1 subunit per 16–17 bp of mtDNA. TFA is a member of the high mobility group (HMG) box domain family and it binds DNA without sequence specificity [84]. TFAM is also an essential component of the mitochondrial transcription machinery [85]. During transcription initiation, the protein binds upstream of the transcription start site and induces a sharp bend into DNA [86-88]. Nucleoid-like particles can be reconstituted by simply mixing TFAM and mtDNA, implying that TFAM on its own can fully compact mtDNA [89-91]. TFAM has two DNA binding sites, and appears to compact mtDNA by cross-strand binding and loop formation. In addition, TFAM binds DNA in a cooperative manner, forming protein-patches on mtDNA [90-92].

That TFAM acts as an epigenetic regulator of mtDNA replication is a tantalizing possibility. Super-resolution microscopy has revealed different forms of nucleoids [91]. Perhaps, the more compact nucleoids represent a mtDNA storage form, whereas the larger forms are involved in active replication and/or transcription. Nucleoids involved in active DNA replication have been localized to contact points between the endoplasmic reticulum (ER) and mitochondria. At these sites, mitochondrial division takes place, leading to the idea that contacts between the endoplasmic reticulum and mitochondria can coordinate mtDNA synthesis with division to ensure even distribution of newly replicated nucleoids within the mitochondrial network [93].

**In vitro**, small changes in the TFAM to DNA ratio can have dramatic consequences. At physiological ratios, there are large variations in mtDNA compaction, fully compacted nucleoids and naked DNA can be observed simultaneously. Under these conditions, a small increase in TFAM concentrations can dramatically increase the number of full compacted mtDNA molecules. [89,90]. This model may explain why TFAM levels in vivo remain roughly proportional to mtDNA levels, and suggests that relatively small changes in TFAM concentrations can have strong effects on both gene expression and mtDNA replication. In support of this notion, longer patches of TFAM prevents DNA unwinding and as a consequence, the mtDNA replication and transcription machineries cannot progress [90]. TFAM may therefore function as an epigenetic regulator, which controls the number of mtDNA molecules available for active transcription and/or mtDNA replication.

### Concluding remarks

A detailed understanding of mtDNA replication is not only important from a basic science point of view, but may also explain the formation of deletions and point mutations associated with human disease and aging. A detailed knowledge of these process may pave the way for development of new therapeutic strategies that can be of use in mitochondrial medicine.

### Summary

- Mammalian mtDNA is replicated by proteins distinct from those used for nuclear DNA replication.
According to the strand displacement model, replication is initiated from two distinct origins, OH and OL.

Transcripts initiated at LSP provide the primer from which POLγ can initiate DNA synthesis at OH.

OL forms a stem–loop structure and POLRMT initiates primer synthesis from the single-stranded loop region.

The role of the mitochondrial D-loop is not understood.

RNASEH1 and MGME1 play important roles in primer removal, but the details of this process are not fully understood.

Top3α is required to resolve hemicatenane structures formed between new mtDNA molecules at the end of replication.

mtDNA is not a naked molecule, but packaged into nucleoprotein complexes, nucleoids.

Mitochondrial division is linked to active mtDNA synthesis.

Acknowledgment
I am grateful to Dr Jay P. Uhler who prepared the illustrations.

Competing Interests
The author declares that there are no competing interests associated with the manuscript.

Funding
This work was supported by Swedish Research Council (2013-3621); Swedish Cancer Foundation (CAN2016/816); European Research Council (DELMIT); the IngaBritt and Arne Lundberg Foundation; and the Knut and Alice Wallenberg Foundation.

Abbreviations
HSP, heavy strand promoter; LSP, light strand promoter; MGME1, mitochondrial genome maintenance exonuclease 1; mtSSB, mitochondrial single-stranded DNA-binding protein; NCR, noncoding region; TAS, termination-associated sequence.

References

© 2018 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY-NC-ND).