JAAKKO POHJOISMÄKI

Modulation of Mitochondrial DNA Replication and Recombination in Mammalian Tissues and Cultured Cells

ACADEMIC DISSERTATION
To be presented, with the permission of the Faculty of Medicine of the University of Tampere, for public discussion in the Auditorium of Finn-Medi 1, Biokatu 6, Tampere, on February 16th, 2008, at 18 o’clock.
ACADEMIC DISSERTATION
University of Tampere, Institute of Medical Technology
Tampere Graduate School in Biomedicine and Biotechnology (TGSBB)
Finland

Supervised by
Professor Howard T. Jacobs
University of Tampere

Reviewed by
Professor Takehiko Shibata
RIKEN, The Institute of Physical and Chemical Research, Japan
Professor Laurie Kaguni
Michigan State University, USA

Distribution
Bookshop TAJU
P.O. Box 617
33014 University of Tampere
Finland

Cover design by
Juha Siro

Acta Universitatis Tamperensis 1294
ISBN 978-951-44-7234-3 (print)
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 697
ISBN 978-951-44-7235-0 (pdf)
ISSN 1456-954X
http://acta.uta.fi
Hey man of science with your perfect rules of measure -
Can you improve this place with the data that you gather?

- Bad Religion: I want to conquer the world

For my mum and for the memory of my father, Veikko.
# Contents

List of original publications .................................................................................. 7
Abbreviations .......................................................................................................... 8
ABSTRACT ............................................................................................................... 10
1 INTRODUCTION .................................................................................................. 12
2 REVIEW OF THE LITERATURE .......................................................................... 13
   2.1 Mitochondria .................................................................................................. 13
   2.2 Mitochondrial origins .................................................................................. 15
   2.3 The mitochondrial genome .......................................................................... 17
   2.4 The structure and organization of mtDNA .................................................. 17
   2.5 The human mtDNA .................................................................................... 20
   2.6 mtDNA inheritance ..................................................................................... 22
   2.7 Segregation of mtDNA and maintenance of homoplasmy ......................... 25
   2.8 Mitochondrial nucleoids ............................................................................. 29
   2.9 General aspects of DNA replication .......................................................... 30
   2.10 Replication of T7 phage ............................................................................. 32
   2.11 Replication of T4 phage ............................................................................. 34
   2.12 Replication of circular bacterial plasmids .................................................. 37
      2.12.1 The theta mechanism ......................................................................... 37
      2.12.2 Strand displacement replication ......................................................... 40
      2.12.3 Rolling circle replication (RCR) .......................................................... 41
   2.13 Special aspects of replicating linear DNA .................................................. 42
   2.14 Replication of mitochondrial DNA ............................................................ 43
      2.14.1 Theta replication ................................................................................ 44
      2.14.2 Rolling circle replication (RCR) .......................................................... 44
      2.14.3 Recombination-dependent replication .............................................. 45
      2.14.4 Replication of linear mtDNA ............................................................. 46
      2.14.5 Reverse transcription in mitochondria .............................................. 47
      2.14.6 Replication of animal (vertebrate) mtDNA ....................................... 47
      2.14.7 Proteins of the mammalian mitochondrial replisome ....................... 51
   2.15 mtDNA binding proteins in mtDNA replication ......................................... 52
   2.16 mtDNA repair ........................................................................................... 53
   2.17 General aspects of recombination in regard to mitochondrial DNA .......... 54
   2.18 Tissue-specific requirements for energy production and mtDNA copy number control ....... 56
2.19 mtDNA in cardiac function and developmental adaptation .......................................................... 57
2.20 Human mitochondrial disorders and heart pathology ............................................................... 58
2.21 Mitochondria, ischemia, preconditioning and damage response in human heart ....................... 60

3 AIMS OF THE STUDY ...................................................................................................................... 62

4 MATERIALS AND METHODS....................................................................................................... 62
4.1 TFAM constructs .......................................................................................................................... 62
4.2 Twinkle and PolG constructs ......................................................................................................... 64
4.3 mTERF constructs ........................................................................................................................ 65
4.4 Mitochondrially targeted RecA constructs .................................................................................. 65
4.5 Sequencing .................................................................................................................................. 65
4.6 RNA interference studies (RNAi) ................................................................................................. 65
4.7 Cell culture and transfections ....................................................................................................... 66
4.8 Inhibition of replication and transcription .................................................................................... 66
4.9 Creation of doxycyclin-inducible Flp-In™ T-Rex™-293 cell lines .................................................. 67
4.10 Transgenic mice .......................................................................................................................... 67
4.11 Autopsy samples ......................................................................................................................... 68
4.12 RNA and DNA extraction and quantification ............................................................................. 68
4.12.1 DNA from cultured cells ........................................................................................................ 68
4.12.2 DNA from tissues ................................................................................................................... 69
4.12.3 mtDNA copy number estimation ......................................................................................... 69
4.12.4 RNA extraction and quantification from cultured cells ......................................................... 69
4.13 Isolation of mitochondria ............................................................................................................ 70
4.13.1 From cultured cells ............................................................................................................... 70
4.13.2 From human and mouse tissues ............................................................................................. 70
4.14 Isolation of mtDNA .................................................................................................................... 71
4.15 Sub-fractionation of mitochondria ............................................................................................... 71
4.16 SDS PAGE and Western blots ..................................................................................................... 72
4.17 Immunocytochemistry ............................................................................................................... 72
4.18 Enzymatic treatments of DNA .................................................................................................... 72
4.19 One-dimensional agarose gel electrophoresis .......................................................................... 73
4.20 Southern blotting ....................................................................................................................... 73
4.21 Two-dimensional neutral-neutral agarose gel electrophoresis (2DNAGE) ............................... 73
4.21.1 2DNAGE methodology ..................................................................................................... 73
4.21.2 Interpretation of 2DNAGE results ........................................................................................ 74
4.22 Radiolabeled probes and blot hybridization ............................................................................. 77
5 RESULTS ................................................................................................................................................... 79
5.1 Structure of mammalian mtDNA replication and recombination intermediates ................. 79
  5.1.1 Electron microscopy of mtDNA from cultured cells (unpublished data) ...................... 80
  5.1.2 Electron microscopy of mtDNA from human heart (IV) .................................................. 85
  5.1.3 Comparative analysis of mtDNA from various human tissues by 1D- and 2D-AGE (IV,
       unpublished data) ....................................................................................................................... 87
5.2 Effects of mtDNA-binding proteins on mtDNA replication and recombination .......... 91
  5.2.1 Manipulation of TFAM expression in cultured cells (I) .................................................. 92
  5.2.2 Manipulation of Twinkle DNA helicase and PolG expression in cultured cells (II and
       unpublished data) ....................................................................................................................... 95
  5.2.3 Expression of mitochondrially-targeted RecA (unpublished data) .............................. 99
  5.2.4 Manipulation of mTERF expression in cultured cells (III) ........................................... 99
  5.2.5 Overexpression of Twinkle and TFAM in vivo in transgenic mice (IV) ...................... 101
5.3 Physiological variables affecting mtDNA replication and topology ................................. 103
  5.3.1 mtDNA copy number in different tissues (IV) ................................................................. 104
  5.3.2 Oxidative damage of mtDNA and replication stall (unpublished data) .................... 105
  5.3.3 Modified mtDNA RIs in a case of a ischemic heart disease (unpublished data) .......... 106
6 DISCUSSION ........................................................................................................................................... 107
  6.1 Mitochondrial DNA replication in mammalian tissues and cultured cells .................... 107
  6.2 Replication pausing in human mtDNA ............................................................................... 111
  6.3 RDR in mammalian mitochondria ..................................................................................... 112
  6.4 A hypothesis for origin independent COSCOFA replication in cultured cells .......... 117
  6.5 Modulation of mtDNA topology ....................................................................................... 120
  6.6 mtDNA replication and deletion formation ........................................................................ 122
  6.7 Cardiac muscle mtDNA replication – physiology and pathology ............................... 123
7 SUMMARY AND FUTURE PERSPECTIVES ................................................................................. 125
ACKNOWLEDGEMENTS .......................................................................................................................... 127
REFERENCES ........................................................................................................................................... 129
ORIGINAL COMMUNICATIONS .............................................................................................................. 153
List of original publications

This thesis is based on the following original communications. Some unpublished data are included where considered necessary as supplementary data.


† This article has been used in the PhD thesis of Sjoerd Wanrooij

* Joint first authorship
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>2DNAGE</td>
<td>two-dimensional neutral-neutral agarose gel electrophoresis</td>
</tr>
<tr>
<td>8-OHdG</td>
<td>8-oxo-deoxyguanosine</td>
</tr>
<tr>
<td>ad</td>
<td>autosomal dominant</td>
</tr>
<tr>
<td>ADP</td>
<td>adenosine diphosphate</td>
</tr>
<tr>
<td>AGE</td>
<td>agarose gel electrophoresis</td>
</tr>
<tr>
<td>ALT</td>
<td>alternative lengthening of telomeres</td>
</tr>
<tr>
<td>AMP</td>
<td>adenosine monophosphate</td>
</tr>
<tr>
<td>ANT</td>
<td>adenine nucleotide transporter</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>APP</td>
<td>amyloid precursor protein</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CMS</td>
<td>cytoplasmic male sterility</td>
</tr>
<tr>
<td>COSCOFA</td>
<td>conventional strand-coupled Okazaki fragment associated</td>
</tr>
<tr>
<td>CSB</td>
<td>conserved sequence block</td>
</tr>
<tr>
<td>Cyt b</td>
<td>cytochrome b</td>
</tr>
<tr>
<td>DAPI</td>
<td>4',6' diamino-2-phenylindole</td>
</tr>
<tr>
<td>DCM</td>
<td>dilated cardiomyopathy</td>
</tr>
<tr>
<td>ddC</td>
<td>dideoxycytidine</td>
</tr>
<tr>
<td>D-loop</td>
<td>displacement loop</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DSB</td>
<td>double-strand break</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double-stranded DNA</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double-strand origin-sequence</td>
</tr>
<tr>
<td>DUI</td>
<td>doubly uniparental inheritance</td>
</tr>
<tr>
<td>ECR</td>
<td>extensive chromosome replication</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EMSA</td>
<td>electrophoretic mobility shift assay</td>
</tr>
<tr>
<td>EtBr</td>
<td>ethidium bromide</td>
</tr>
<tr>
<td>Fpg</td>
<td>Formamidopyrimidine-DNA-glycosylase</td>
</tr>
<tr>
<td>FRDA</td>
<td>Friedreich ataxia</td>
</tr>
<tr>
<td>HB</td>
<td>homogenization buffer</td>
</tr>
<tr>
<td>HCM</td>
<td>hypertrophic obstructive cardiomyopathy</td>
</tr>
<tr>
<td>HEK</td>
<td>human embryonic kidney</td>
</tr>
<tr>
<td>HMG</td>
<td>high-mobility group</td>
</tr>
<tr>
<td>HSP</td>
<td>heavy-strand promoter</td>
</tr>
<tr>
<td>H-strand</td>
<td>heavy-strand</td>
</tr>
<tr>
<td>IAA</td>
<td>isoamylalcohol</td>
</tr>
<tr>
<td>IFM</td>
<td>infraspiral mitochondria</td>
</tr>
<tr>
<td>IP</td>
<td>ischemic preconditioning</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase (1000 bp)</td>
</tr>
<tr>
<td>KSS</td>
<td>Keams-Sayre syndrome</td>
</tr>
<tr>
<td>LM-PCR</td>
<td>ligation-mediated PCR</td>
</tr>
<tr>
<td>LSP</td>
<td>light-strand promoter</td>
</tr>
<tr>
<td>L-strand</td>
<td>light-strand</td>
</tr>
<tr>
<td>MnSOD</td>
<td>manganese superoxide dismutase</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>mSMY</td>
<td>modified SMY</td>
</tr>
<tr>
<td>mt</td>
<td>mitochondrial</td>
</tr>
<tr>
<td>mtDNA</td>
<td>mitochondrial DNA</td>
</tr>
<tr>
<td>mTERF</td>
<td>mitochondrial transcription termination factor</td>
</tr>
<tr>
<td>MPTP</td>
<td>mitochondrial permeability transition pore</td>
</tr>
<tr>
<td>MTRPOL</td>
<td>mitochondrial RNA polymerase</td>
</tr>
<tr>
<td>mtSSB</td>
<td>mitochondrial single-strand binding protein</td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
</tbody>
</table>
ABSTRACT

Mammalian mitochondrial DNA (mtDNA) is a 16.6 kb circular molecule that encodes parts of the mitochondrial protein synthesis machinery as well as key protein units of the ATP-producing electron transport chain. This feature gives mtDNA a central role in cellular metabolism, and its faithful maintenance is critical for the survival of the organism. Learning to understand how healthy tissues maintain their mitochondrial mtDNA is crucial for understanding the pathological mechanisms of human disorders affecting mtDNA function. In addition to rare mitochondrial disorders, common diseases such as cardiomyopathies and parkinsonism are associated with mtDNA defects.

This work has aimed at increasing our understanding of the similarities and differences of basic molecular mechanisms involved in mammalian mtDNA replication and recombination. To achieve this, I have used a combination of one- and two-dimensional agarose gel electrophoresis and electron microscopy to analyze mtDNA replication intermediates in mammalian cells and tissues in a variety of conditions. In most tissues, mtDNA replicates via a unique strand-symmetric mechanism, which incorporates long stretches of RNA onto the lagging-strand. This RNA is later matured into DNA by a mechanism that is not in relation to the progression of replication fork. These specific molecular features of mtDNA replication were manipulated by altering the expression of the multifunctional DNA binding protein TFAM as well as by expressing catalytic mutants of the mitochondrial DNA polymerase gamma (PolG) and DNA helicase Twinkle in cultured human cells.

However, high energy tissues such as human heart were found to utilize a completely different means of mtDNA replication. This replication mechanism is associated with highly active molecular recombination and most likely involves recombination-dependent initiation, as seen in the mitochondria of plants and fungi. The abundant mtDNA recombination in heart seems to be a human-specific feature. However, a similar pattern of mtDNA recombination could be induced in tissues of transgenic mice that overexpress Twinkle or human TFAM, indicating that the actual molecular mechanisms of mtDNA recombination are conserved and play a physiologically significant role in mtDNA maintenance in mammalian high energy-demand tissues. The TFAM transgene has also been reported to provide protection from ischemia-induced cardiac remodeling.

Recombination is an efficient means of protection against DNA double-strand breaks, which might occur more readily in the highly oxidative environment of active heart muscle mitochondria. Besides protecting from extrinsic damage, mtDNA needs to be protected from
catastrophic errors that might arise from mistakes in replication or transcription. Collisions of transcription and replication machineries can result in double-strand breaks and genome rearrangements. I propose that these can be avoided by tightly regulated passage of the two machineries in mitochondrial replication fork barrier regions (RFBs). As the first RFB-regulating protein from mitochondria a protein called mTERF, which was originally characterized as a mitochondrial transcription termination factor, is reported. These RFBs are also common breakpoints for mtDNA deletion formation. The findings presented in this thesis will help to understand the molecular mechanisms of high energy-demand tissue mtDNA pathologies, myocardial protection from ischemia-induced remodeling and role of RFBs or replication stalling in mtDNA deletion formation.
INTRODUCTION

Life is a phenomenon that distinguishes living organisms from non-living beings and manifests itself through metabolism, reproduction and adaptation to the environment. Metabolism converts energy and raw materials acquired from the surroundings to support an organism’s growth, movement and ability to react to stimuli. The fundamental unit of life is a cell and reproduction at its simplest is the production of new individual cells. The most important feature of this process involves the passage of the genetic material by copying – or in biological terms – replicating it to the offspring.

Mitochondria are eukaryotic intracellular organelles that are central elements of metabolism, being responsible for the transformation of energy to forms that can be utilized by the cell. The unique feature of this energy metabolism is that it exploits atmospheric oxygen, enabling cells - and eventually humans - to respire. The evolutionary ancestors of mitochondria are believed to have been once independent living cells. Mitochondria still retain their own genome that encodes important proteins involved in the respiratory chain as well as tRNAs and rRNAs required for the decoding of the genes resident in mitochondrial DNA into protein products. The mitochondrial genome is necessary for functional cellular respiration. Faithful replication of mtDNA ensures the maintenance of the respiratory chain in succeeding cellular generations. Additionally, changes in environmental or cellular conditions bring about different energy requirements to which the mitochondria need to adapt. These adaptations, which in human examples include tissue differentiation in the embryo or physical exercise in adults, also involve changes in the organization and copy number of the mitochondrial genome.

This thesis will summarize the key aspects of mitochondrial function, common features of mitochondrial DNA (mtDNA) organization and replication in various organisms. The segregation and inheritance of mtDNA will be described in detail, because the appreciation of the underlying evolutionary constraints is crucial to understand the structure of mtDNA, as well as some special aspects of its maintenance. In addition to the peculiarities arising from the extrachromosomal status of mtDNA, mitochondria are also challenged by different physiological requirements. Taking this into account, the work represented in this thesis sheds light on the qualitative and quantitative differences of mtDNA replication and recombination in mammalian tissues and the role of specific proteins involved in these processes. Special emphasis will be on human heart muscle mtDNA and on the possible importance of tissue-specific replication mechanisms in human disease.
2 REVIEW OF THE LITERATURE

2.1 Mitochondria

Mitochondria are double-membraned, pleiomorphic, semi-independent cellular organelles that are found inside the cells of most eukaryotic organisms. Mitochondria are generally stated to be 1-10 μm in size, but in fact they form networks in living cells (Figure 2.1). Mitochondria were discovered by various microscopists in the mid 19th century as grainy entities inside cells (Scheffler 1999). In the end of that century, due to improved staining methods and micro-dissection of cells, researchers were able to conclude that these grains were actually threads – or mitos in Greek - and were defined by a membrane.

Figure 2.1. Mitochondrial network inside a cultured human cell. Nucleus outlined with dashed circle. With kind help from Peter Martinson.

The most obvious function of mitochondria is the conversion of food-derived energy to ATP energy that can be utilized by different cellular processes. In this process the electrons carried by oxidised NADH from the the TCA cycle are transferred to oxygen via the inner membrane-bound electron transport chain coupled to proton pumping (Figure 2.2). The proton pumping creates an electrochemical gradient between the two mitochondrial membranes in a process called
Chemiosmosis. The gradient is released through the F$_1$F$_0$-ATP synthase complex, allowing the synthesis of high energy ATP molecules. This in turn provides energy for all important cellular functions, including the maintenance of membrane potential and muscle contraction. This process of energy conversion from food-derived, highly reduced carbon compounds into ATP with the help of atmospheric oxygen is also known as oxidative phosphorylation (OXPHOS). The understanding of this process has been one of the main achievements in modern biology and has been rewarded by Nobel prices for P. Mitchell in 1978 for describing chemiosmosis, and J. Walker and P. Boyer in 1997 for describing the mechanism of ATP synthesis.

Figure 2.2. Electron transport chain (ETC) and ATP synthesis in the mitochondrial inner membrane. CI-V: respiratory complexes. Electrons carried by NADH enter the ETC via NADH dehydrogenase (CI). CI donates the electrons to the ubiquinone pool (Ubq). Succinate dehydrogenase (CII) is a part in TCA cycle and transports electrons from succinate to UbQ via FAD-coenzyme. Electrons pass through the ETC in subsequent reduction and oxidation reactions and finally end up at oxygen that is reduced to water. CI, CIII and CIV pump protons into the intermembrane space with the energy obtained from the changes in redox state of the complex. The created electrochemical gradient is released through ATP synthase (CV) and the obtained energy is used for driving the energetically unfavourable reaction of ATP synthesis (ADP + P’ → ATP).

Besides energy conversion, mitochondria are important in the biosynthesis of various compounds such as heme and steroids, for the maintenance of Ca$^{2+}$ homeostasis and iron metabolism. Mitochondria are also key players in programmed cell death or apoptosis that is a
crucial process for the regulation of development, tissue homeostasis and immunological responses. They are also important for heat production in homeothermic animals.

2.2 Mitochondrial origins

Already in 1890 R. Altman proposed in his book that the small granules inside the cytoplasm of eukaryotic cells, which he called bioplasts, were autonomous organisms forming bacterial-like colonies (Scheffler 1999). Almost simultaneously investigators working with plants had come up with similar ideas concerning chloroplasts, which were dividing in similar fashion as free-living cyanobacteria (Mereschkowsky 1905). The idea of a bacterial origin of mitochondria and chloroplasts was supported by several following investigators, but the experimental evidence was lacking until the discovery of DNA from these organelles in 1959 (Stocking & Gifford, 1959) and 1963 (Nass & Nass, 1963).

A decade after these discoveries F.J.R Taylor proposed a provocative theory of serial endosymbiosis. This idea was further popularized by Lynn Margulis (1981), who argued that eukaryotic cells originated from communities of interacting prokaryotic entities, including endosymbiotic spirochaetes that developed into eukaryotic flagella and cilia. Regarding the origins of flagellae or ciliae the idea has not received much support, but the evidence for the endosymbiotic origins of mitochondria and plastids is overwhelming.

The endosymbiotic theory of mitochondrial origin proposes that the ancestors of mitochondria were free living α-proteobacteria–like organisms that were capable of using oxygen as an ultimate electron acceptor for aerobic respiration, as a means of energy production. Nowadays we tend to take oxygen for granted, but it was not so for early life that developed in the nearly anaerobic environment of the young Earth. The evolutionary invention of water-splitting photosynthesis by ancestral cyanobacteria brought about the first ecocatastrophe in the form of molecular oxygen and irreducible oxidization of Earths atmosphere some 2.2 - 2.7 billion years ago. As a result many ancient organisms were struggling for survival. Molecular oxygen (O₂) is a strong oxidant and thus extremely harmful for biological macromolecules and for many reduced inorganic compounds that chemosynthetic organisms could utilize as their energy source.

Despite the implied trade-off, the oxidative capacity of molecular oxygen has one important advantage in energy metabolism. All organisms take advantage of a double membrane-coupled electron transport chain in their energy metabolism. The systems vary in detail, but they are
all based on the passage of electrons from one substrate to another while the energy harvested from these reduction-oxidation (redox) reactions is used to pump protons into an intramembrane space. This proton gradient is used in ATP production as described earlier. In order to keep the system working, there needs to be a terminal electron acceptor that is a strong enough oxidant to “pull” the electrons through the electron transport chain. In anaerobic respiration performed by many bacteria the terminal acceptor molecules can be nitrates, sulphates or even carbon dioxide (CO$_2$) that have redox potentials of maximally around 200 mV. Molecular oxygen, however, has a redox potential of 820 mV at physiological pH, meaning that significantly more energy is released when oxygen is reduced to water (O$_2$ $\rightarrow$ H$_2$O). In cellular terms this means that with the same starting material but oxygen as the terminal electron acceptor, many more protons can be pumped and therefore more ATP can be produced. As the energy contained in fuel is transformed to heat by combustion in the presence of oxygen, cells by analogy take advantage of controlled oxidation of food derived substrates.

When engulfing the mitochondrial ancestor and establishing a symbiosis with it, the ancestral eukaryote obtained in exchange extremely efficient means of energy production as well as a means of adaptation to rising oxygen levels. Endosymbiosis is a frequent phenomenon in nature with various examples known of more recent events, mainly involving algae and various animal or fungal partners (McFadden 2001). The evolutionary importance of the mitochondrial endosymbiosis is also seen in phylogenetic analyses, which indicate that all mitochondria have a common ancestor. The mitochondrial symbiosis occurred only once, or at least the ancestral mitochondrial symbiont was the only one to survive. In time the mitochondria were intimately integrated, losing most of their autonomy together with the majority of their genes to the host. It is estimated that more than 99% of mitochondrial proteins are nowadays encoded by nuclear genes (Gray et al. 1999). The process is witnessed in the eukaryotic genome by the presence of many genes of bacterial origin as well as by the existence of different sets of genes in mtDNA in different organisms.

Besides the trend of gradual reduction of the mitochondrial genome and functions through the evolution of eukaryotes, there is also evidence that the complete loss of the mitochondrial genome has also occurred in some lineages (Biagini et al. 1997, Hackstein et al. 1999, Dyall & Johnson 2000). This has happened through adaptation to an anaerobic or micro-aerobic environment, where mitochondrial respiration has no value. Instead, some anaerobes that still compartmentalize the terminal steps of energy metabolism have rudimentary mitochondria or hydrogenosomes (type II anaerobes). These organelles ferment substrates such as malate using H$^+$ as terminal electron acceptor, producing reduced molecular hydrogen H$_2$ and ATP. Also the type I
anaerobes have rudimentary mitochondria even though they are not involved in energy metabolism and many genes of mitochondrial origin can be still found in their genomes (Clark & Roger 1995, Tovar et al. 1999, Mai et al. 1999, Arisue et al. 2002).

2.3 The mitochondrial genome

The discovery of mitochondrial DNA (mtDNA) provided the first strong support for the endosymbiont hypothesis. However, this observation also had more far reaching consequences that comes to our understanding of mitochondrial function. Although in human cells the mtDNA represents only around 1% of the total genetic material it is essential for the function of the respiratory chain (Scheffler 1999). Cultured animal cells without any mtDNA can be created artificially. The cells still have mitochondria, but the OXPHOS system is not working and they use glycolysis as the principal means of energy production. Some yeasts, which can live as facultative anaerobes by fermenting sugars can also survive without any mtDNA. Cells which do not have any mtDNA are called $\rho^0$ ($rho^0$).

Because the mtDNA exists fairly isolated in the mitochondrial compartment of the cell it requires its own transcription and translation machinery, which is distinct from the nuclear / cytosolic one. When researchers eventually learned to interpret the mitochondrial genetic code it was surprising that the genetic code differed from what was thought to be the standard universal code. For instance the codons, UAA, UGA and UAG that function as termination signals in both prokaryotes and the eukaryotic cytoplasm code for amino acids in mitochondria; instead a different set of termination codons is used. Also instead of just one translation initiation codon, AUG, additional initiation codons can be used. There is also significant variance in the codon usage between different evolutionary lineages of mitochondria.

2.4 The structure and organization of mtDNA

All known mitochondrial genomes encode two ribosomal RNAs and most 20 or more tRNAs, with the exception of trypanosomes. Animal mtDNAs encode 13 polypeptides that represent parts of the OXPHOS complexes, but a variety of other genes are found in other organisms. For example the mtDNA from a protozoan, *Reclinomonas americana*, contains 97 genes of which 23 are involved in OXPHOS, 18 encode ribosomal proteins and even one RNA polymerase is known.
Besides the gene content there is also a high variability in the size of the mitochondrial genomes. The mtDNA from *R. americana* contains 69,034 base pairs (Lang *et al.* 1997) whereas the compact metazoan mtDNAs are generally around 17,000 bp, varying from approximately 14,000 bp in the nematode *Caenorhabditis elegans* to over 42,000 bp in the mussel *Placeopecten megallanicus* (Scheffler 1999). Plant mitochondria have the largest genomes known; up to 570,000 base pairs in maize. Curiously, however, they do not have proportionately more coding information; instead the genome expansion is primarily due to the presence of large intergenic regions, repeated segments, introns and intronic open reading frames, as well as incorporation of foreign DNA of plastid, nuclear and plasmid origin (Bullerwell & Gray 2004). The mitochondrial DNAs from apicomplexan parasites are the most reduced in size and gene content, having only five genes in their 6,000 bp genome; two fragmented large subunit rRNAs, *cyt b, cox I* and *cox III* (Wilson & Williamson 1997). These rudimentary mitochondrial genomes exist as variably-sized linear DNAs of tandem repeats and are transcriptionally active. The three polypeptide products are also present and functional in the apicomplexan mitochondrial electron transport chain, indicating that these polypeptides might be the minimal requirement for mtDNA gene content (van Dooren *et al.* 2006).

The compact metazoan mtDNAs are typically circular, double-stranded DNA molecules that in addition to the standard monomeric genome can exist as catenanes of two or several interconnected circles, with a small proportion of circular head-to-tail dimers (Hudson & Vinograd 1967, Clayton & Vinograd 1967, Clayton & Smith 1975, Piko & Matsumoto 1977, Boore 1999). Depending on the tissue, catenated molecules can represent up to 10% of all mtDNA and dimers generally less than 0.2% (Clayton & Smith 1975, Piko & Matsumoto 1977). In the tissues of mice, complex mtDNA molecules larger than pentamer size have not been observed (Piko & Matsumoto 1977).

Curiously, a much higher frequency of both complex molecules as well as dimers can be seen in malignant or pathological tissues, such as tumors, as well as some established cell lines and transformed cells (Clayton & Smith 1975, Clayton & Vinograd 1967, 1969). In human granulocytic leukemia cells up to 50% of mtDNA can exist as dimers and some mouse L-cell lines can have 100% of mtDNA as dimers (Clayton & Smith 1975). Despite the fact that most of these studies were conducted some 40 years ago, the mechanism(s) by which mammalian mtDNA monomers are converted to dimers remain unknown.

In the mesozoan animal *Dicyema* the mtDNA is organized in a population of minicircular DNAs of 1,000-2,000 bp in size that as a whole contains all mitochondrial genes (Watanabe *et al.* 1999). A similarly unusual organization can be found in trypanosomal mitochondria (Shlomai 2004, Lukes *et al.* 2005): their mtDNA or kinetoplastid DNA (kDNA) exists
as a network of thousands of catenated molecules of two molecular types; mini- (500 – 10,000 bp) and maxicircles (20,000 – 40,000 bp). Minicircles come in many sequence classes and in most species there is more than one class of minicircles present in the same cell. The different classes share a strongly conserved region that represents one tenth of the molecule. The sole function of minicircles seems to be to encode guide RNAs, which are involved in the peculiar process of RNA editing of trypanosomal mitochondrial RNAs. This type of RNA editing is unique for these microorganisms and allows the production of several products from one set of genes. Kinetoplastid maxicircles are more homogenous and encode the OXPHOS genes.

The mitochondrial genomes of fungi are generally more compact than those of plants. The molecular geneticists’ favourite yeasts, *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* have circular genomes of 20 kb and 68-85 kb respectively. The size variation in *S. cerevisiae* is due to optional introns that can have open reading frames (ORF). The products of these ORFs appear to function in mRNA splicing, but their absence does not affect any mitochondrial function. *S. cerevisiae* strains can differ also by the presence of an omega (ω+) locus, a 1132 bp group I intron, in the large rRNA gene (Colleaux et al 1988). This intron can move and insert into genomes which lack it (ω-). The intron has an ORF encoding the ω transposase, a site-specific double-strand endonuclease, which is targeted to a sequence within the intron-free allele. Similar introns can be found also in other regions of the genome (Groth et al. 2000). While the ω-factor encoding introns belong to group I, there are also group II introns. These introns encode for retroviral related reverse transcriptase, which is capable of making DNA copies of RNA transcripts of the intron sequence.

Besides a tendency to give rise to ρ⁰ mutant cells, *S. cerevisiae* commonly develops a state called ρ⁻ (Contamine & Picard 2000). These cells are also unable to respire, but whereas ρ⁰ cells lack mtDNA completely, ρ⁻ cells instead have extensive deletion of mtDNA. The remaining fragment of mtDNA consists of repeats of the same ρ⁺-derived sequence. Two major types of repeat arrangements exist: the sequence can be directly repeated with head-to-tail junctions or alternatively, inverted and palindromic repeats, where the repeat unit is an inverted duplication of the retained sequence. The ρ⁻ genomes always have several copies of the so-called ori sequence, which is thought to be involved in the control of replication.

Several filamentous fungi and yeasts have mtDNAs that comprise mostly polydispersed linear tandem arrays of a basic genomic unit, similar to the apicomplexans (Nosek & Tomaska 2003). The mtDNA from the liverwort *Marchantia* is arranged as genome-sized and multimeric linear DNAs, which have branched forms with circular head-to-tail permutations of the sequence.
(Oldenburg & Bendich 1998, 2001, see chapters 2.11 and 2.14 for details). Similar functional forms can be found also in higher plants (Oldenburg & Bendich 1996, 2004). Surprisingly, also the close unicellular relatives of animals, such as choanoflagellate and ichthyosporean protists, have complex mtDNA organization (Burger et al. 2003). For instance Amoebidium parasiticum mtDNA exists as several hundred linear chromosomes totalling 200,000 bp that share elaborate terminal-specific sequence patterns. Other multipartite mtDNAs can be found in the phylum Cnidaria, of which several posses mitochondrial genomes that consists of two to four non-identical linear DNA molecules (Ender & Schierwater 2003). Nothing is known about how segregation of multipartite mtDNA occurs. More organized and uniform linear genomes can be found in the mitochondria of some ciliate protozoans, such as Paramecium and Tetrahymena, as well as in algae like Chlamydomonas and yeasts like Candida (Nosek & Tomaska 2003).

Based on the well known animal examples, it was long thought that mtDNAs would generally be fairly simple circular vestiges of prokaryotic genomes. Anecdotally, reports of linear chromosomes were disregarded as artefactual contamination of nuclear DNA or fragmentation of pre-existing circular mtDNA (as reviewed by Nosek & Tomaska 2003). In fact, it seems that circular mtDNA molecules usually represent only a small proportion of the total and one genome sized circles are extremely rare outside the metazoan lineages.

2.5 The human mtDNA

The human mitochondrial genome is 16,569 bp long and, as in all higher animals, it exists as a circular double-stranded DNA molecule (Scheffler 1999 [Figure 2.3]). The genome is extremely compact with most polypeptide and rRNA genes being separated by one or more tRNA genes and with few, if any, non-coding nucleotides in between. The transcription of the genome occurs in both directions and the open reading frames can be on either strand, although the heavy- or H-strand has many more genes than the light- or L-strand. The nomenclature for the strands originates from their slightly different densities in alkaline CsCl gradient centrifugation due to the different base composition.
Mitochondrial transcription produces polycistronic RNA molecules, which in turn are processed to form mitochondrial mRNAs, rRNAs and tRNAs. A 1122 bp long non-coding region (NCR [Figure 2.4]) can be found between the tRNA⁰ pro and tRNA⁰ phe genes in humans and exists in variable lengths in other animals as well. The NCR contains promoter regions for both strands, the light-strand promoter (LSP) and the heavy-strand promoter (HSP), stretches of evolutionary conserved sequences (conserved sequence blocks, CSBs) that are believed to function in the regulation of DNA replication, a site known as heavy-strand origin O₇ and a Termination Associated Sequence TAS. As discussed later in the chapter on mtDNA replication (2.14) the status of O₇ as a replication origin is currently disputed, however it is likely to be at least the common replication terminus. In the NCR there can be also a peculiar triplex structure, the displacement loop (D-loop), where a 650 bp long single-stranded DNA molecule (7S DNA) is hybridized to the
parental L-strand. The D-loop was postulated to arise from replication initiating at $O_H$ and subsequently terminating at TAS. The function of this enigmatic structure remains unknown.

![Diagram of human non-coding region](image)

Figure 2.4. A more detailed map of the human non-coding region showing the key features. $O_H =$ origin of Heavy-strand replication, $P_L =$ Light-strand promoter, $P_H =$ Heavy-strand promoter. CSBI, II, III = conserved sequence blocks, TAS = termination associated sequence. Assignments based on Mitomap.org: mtDNA function locations.

Due to the reasons described in chapter 2.7, the human mtDNA carries more variation between populations than other regions of our genome. This and its maternal inheritance makes mtDNA a valuable tool in population genetics and anthropology, providing valuable insights into the migrations and origins of human populations (Pakendorf & Stoneking 2005).

### 2.6 mtDNA inheritance

During the evolution of eukaryotes a substantial and unidirectional transfer of genes from mitochondria to the nucleus has occurred. Besides reducing the mitochondrial genetic content, this transfer has also stripped the mitochondria - with some exceptions - of all genes encoding proteins involved in DNA maintenance and expression. As a result mitochondrial genomes have lost their ability to replicate, transcribe and translate their genomes independently. Instead they need to rely on nuclear-coded proteins.

The replication of nuclear genomes is stringent; during each mitotic cycle, each genome is replicated exactly once and each daughter cell gets only one set of chromosomes. In contrast, mitochondrial genomes are thought to replicate and be partitioned rather randomly (Magnusson et al. 2003). Moreover, in one cell there are multiple mitochondria, each with multiple copies of
mtDNA. This in theory exposes mtDNA to extensive intra- and intercellular competition and selection between different mitochondrial genomes (Xu 2005).

Probably, as a result, there is a tendency to maintain the uniformity of mitochondrial genomes – or homoplasmy – in cells. The condition where several mitochondrial genotypes exist in one cell is called heteroplasmy. The trade-offs for the host cell caused by the selfish selection of mitochondrial genomes might represent the most important evolutionary pressure for the reduction of mtDNA, suppression of genetic variability and transfer of genes responsible for its replication under the control of the nucleus. Exceptions, such as invasion of mtDNA by selfish elements and defects in mtDNA segregation and inheritance, highlight this rule.

One important mechanism in maintaining homoplasmy is the separation of different mitochondrial lineages by having uniparental inheritance. So far almost all animal species examined show strict maternal inheritance of mtDNA. The offspring obtains its mitochondria only from the oocyte of the mother. Sperm cells also contain mitochondria, but these are normally not passed on to the offspring. It was long thought that the paternal mitochondria do not even enter the oocyte upon fertilization, but this seems not to be true. Instead, the paternal mitochondria are recognized and tagged for proteolytic destruction by adding ubiquitin peptide marker (Sutovsky et al. 2000). Sometimes in interspecific crosses this mechanism is unable to recognize the foreign mitochondria and paternal leakage of mtDNA can occur (Shitara et al. 1998). Defects of this recognition mechanism might cause some level of biparental mtDNA inheritance, although the phenomenon is rare. Such a case of biparental transmission has even been reported for a human individual (Schwartz & Vissing 2002).

Species of mussels from the families Mytilidae and Unionidae contain two types of mtDNA, F and M. The F type behaves like in other animals, whereas M type is transmitted solely through sperm and exists only in male gonads (Breton et al. 2007). This phenomenon is called doubly uniparental inheritance (DUI), where male and female mtDNAs have separate transmission routes and are kept under strict containment already in the embryonic state. The molecular mechanisms of this selection are unclear, but it is thought to be related to the function of the sex determination genes. Genetic analysis of the M and F mtDNA types has revealed that DUI has persisted over 200 million years and the mtDNA sequences are thus highly diverged even when compared with the interspecific divergence in other animals. The intra-species sequence divergence in these mussels for the \textit{cox1} gene can be >20%, higher than the difference between mouse and rat.

One of the consequences of strictly maternal inheritance of mtDNA is that for selfish mitochondrial genes males are unwanted dead-ends. Some intracellular symbionts or parasites actually can bias the sex ratio of their hosts to favour females thus facilitating their own
transmission. The intracellular bacteria *Wolbachia*, which occurs in many invertebrates, are famous examples of genetic control of host sex determination (see Iturbe-Ormaetxe & O’Neill, 2007 for a recent review). Consequently, mtDNA genotypes that have positive fitness effects in females but potentially deleterious effects in males can have selection advantage with sex-antagonistic effect for the species. It may well be that the mussel DUI evolved to counter this unbalance of sexual selection (Breton *et al.* 2007).

A similar avoidance of mitochondrial competition is obvious in some organisms, such as heterothallic fungi. These fungi are haploid by default and practise sex by exchanging nuclei without cytoplasm (Aanen *et al.* 2004). When two compatible monokaryons meet, they pair and form a dikaryon, with the nuclei homogenously distributed, but being a mosaic of mitochondrial genomes. The dikaryon gives rise to fruiting bodies (mushrooms) containing the haploid basidiospores. The presence of two mitochondrial genomes leads to a genomic conflict, where individual mitochondria can be considered as trying to increase their relative probability of being included in the spores. A monokaryon normally accepts its partner’s nucleus (female role) and donates its own (male role). A mitochondrion that can prevent the male role of the monokaryon while maintaining the female role will have selective advantage over the partner mitochondria. Such a phenomenon is called cytoplasmic male sterility (CMS) and is fairly frequent in fungi and even in higher plants.

Besides the different mitochondrial genomes, the nuclear genome is also involved and has a major influence on the outcome. Nuclear fitness is directly dependent on the dikaryon fitness and any mitochondrial competition through CMS reduces the dikaryon fitness. This leads to selection of nuclear genes that can suppress the selfish behavior of mitochondrial mutants.

In some cases the potential bi-parental inheritance of mitochondria is avoided by nuclear genes giving one mitochondrial population a complete dominance over the dikaryon, as in ascomycetes, some basidiomycetes and slime molds (Mirfakhraei *et al.* 1990, Lee & Taylor 1993, Coenen *et al.* 1996, Yan *et al.* 2007). In slime molds this is achieved by targeted desctruction of the mtDNA from the other mating type (Moriyama & Kawano 2003).

In yeast the zygote inherits the mitochondria from both parents. Also here, numerous studies have shown that the heteroplasmic state of the zygote is transient. In fact, the vegetative multiplication of the zygote is coupled with a rapid segregation of mtDNA molecules, leading to homoplasmic clones (Chen & Butow 2005).

There is no direct evidence of mammalian mitochondria killing or feminizing males. The small size of the mammalian mtDNA already limits the capacity to influence complex phenotypic traits such as sex determination (Zeh & Zeh 2005). As mentioned before, it is likely that this is one
of the benefits of decreased mtDNA content. However, mitochondria can still have tools in their armoury to influence their “host”. There are studies demonstrating that mtDNA variation has an effect on the phenotype from brain development to ageing, resulting from the extensive crosstalk between nuclear and mitochondrial genomes (Kato et al. 2001, Capri et al. 2006). For example, twin and pedigree studies indicate that sexual orientation in males is influenced by maternally inherited factors (Whitham et al. 1993). Despite the publicity surrounding earlier studies that claimed to find a link between homosexuality and the X-chromosomal locus (Turner 1995), subsequent studies were unable to confirm such a link (Rice et al. 1999). An alternative hypothesis suggests that homosexuality might originate from incomplete penetrance of an mtDNA haplotype that is selected for its ability to kill males during embryonal development. This hypothesis gets some support from the observations that maternal aunts significantly outnumber maternal uncles in homosexual men and that such a female bias is not evident in paternal relatives (Turner 1995). The possible mechanisms of this influence is not clear, but a number of steroid receptor influenced effects in mitochondria are known (Gavrilova-Jordan & Price 2007) and especially the effects on transcriptional activation might provide plausible mechanisms for control region polymorphisms to influence physiological outcome.

2.7 Segregation of mtDNA and maintenance of homoplasmy

In contrast to the nuclear genome, mitochondrial genomes segregate both at meiosis and mitosis (Birky 2001). The mechanism in both cases is the same; the mitochondrial genomes are physically divided between daughter cells at every cell division. However, the mitochondrial alleles are not evenly segregated at every cell division. Even if the daughter cells receive a more or less equal number of mitochondrial genomes, some copies of mtDNA can replicate more often than others by chance or in response to selective constraints. Segregation can also be influenced by the spatial position of the genome inside the cell. Thus, mtDNA segregation between daughter cells occurs stochastically.

Most segregation studies have been conducted in yeast, because yeast zygotes inherit mitochondrial genomes from both parents. If the parental genomes differ in their allele composition, the resulting zygote is heteroplasmic and the fate of different alleles through succeeding generations can be followed using molecular markers. In budding yeast, as a result of the stochastic sampling and partitioning of mitochondria to the daughter buds, the majority of progeny achieve homoplasmy in approximately 20 generations. As discussed earlier, assuming that paternal leakage of mtDNA is
rare in humans and other mammals, the only sources of heteroplasmyn are new mutations. Such mutations can be point mutations giving rise to single nucleotide polymorphisms or larger re-arrangements, such as deletions and duplications. Deletions are usually associated with mitochondrial pathologies, as discussed later, but they are also known to exist at a low level in healthy human tissues (Kajander et al. 2000). As default, one mitochondrial genotype is usually prevalent in heteroplasmic cells and the other(s) are present in a very low proportion (Kmiec et al. 2005).

The most efficient way of following the fate of heteroplasmyn in mammalian cells has been to create cytoplasmic hybrids by fusing somatic cells with an enucleated counterpart having a different mitochondrial genotype (Birky 2001). A similar kind of cytoplasmic hybrids – or cybrids – can be also created by microinjection of cytoplasm. Heteroplasmic animals can be created by the same principle of cytoplast fusion.

In cybrid cells the different mtDNA genotypes can have various fates. Usually there is a very gradual segregation of mitochondrial genotypes, so that homoplasmyn is eventually achieved after several generations (Matthews et al. 1995, Shoubridge 1995, Lehtinen et al. 2000). It is noteworthy that in cultured cells and in the absence of selection the overall mitochondrial genotype can be relatively stable, with an estimated segregation number being >1000 (Lehtinen et al. 2000). In the presence of selection, the cells can achieve a level of fairly stable heteroplasmyn, where the different genotypes co-exist in equilibrium. Whole animals, however, behave completely differently and homoplasmyn can be achieved in only a couple of generations (Olivo et al. 1983, Laipis et al. 1988, Ashley et al. 1989, Larsson et al. 1992, Blok et al. 1997). This has been suggested to happen because the primordial germ line cells in female embryos having only a few copies of mtDNA are a genetic bottleneck with strong selection for fittest genomes producing a rapid genotype shift within a generation. However, some data suggests that random drift alone is capable of explaining the purifying effect of primordial germ lines on mtDNA heteroplasmyn (Jenuth et al. 1996). In rapidly dividing somatic cells the changes in heteroplasmyn are driven by clonal expansion and random drift in several generations (Elson et al. 2001). However, in mice heteroplasmic for two different mtDNA haplotypes, a selective bias in different tissues can be observed indicating that selection under some circumstances can play a role in mtDNA segregation (Battersby & Shoubridge 2001).

Recent evidence from mice indicates that the mtDNA copy number in germ line cells is also not small enough to explain the phenomenon on its own (Cao et al. 2007). In fact, the apparent bottleneck might be created by a small number of segregating units, which themselves contain several copies of mtDNA, thus ensuring the maintenance of adaptive potential in the progeny while maintaining homoplasmyn. These units by definition would be mitochondrial nucleoprotein
complexes or nucleoids, which will be discussed in more detail in the next chapter. In fact, based on the work in cultured cells Jacobs and co-workers (2000) had suggested already earlier that the nucleoid would be the basic unit of inheritance in mitochondria, whose faithful replication and segregation – or mitochondrial mitosis – explains the observations of stable heteroplasmy where deleterious mutations complement each other and act as a genetic unit. In contrast to this emergency maintenance of genetic fitness in somatic cells, in germ line a reorganization of nucleoids might occur facilitating the rapid genotypic switch and appearance of new and predominant mtDNA variants within few generations. In other words, the selective constraints in somatic cells and germ line might differ: Somatic cells need to maintain their functional capacity despite accumulating mutations, whereas the priority for germ line cells would be to focus on the purification of fitness reducing mutations by maintaining mtDNA homoplasmy that will be transmitted to the next generations. However, there is no evidence of the mechanisms evaluating mtDNA fitness in primordial germ cells. It is likely that stringent selective constraints on the mtDNA function come to effect during or after embryonic development.

The genetics of mitochondrial DNA has an important evolutionary consequence. A vast majority of newly arising mutations are deleterious. Natural populations are continuously under threat from fitness-reducing new deleterious alleles and their elimination is effected by means of natural selection, creating a standard mutation-selection balance for the level of diversity for fitness. Because the segregating numbers of mtDNA in the primordial germlines are small, there is a tendency that genetic drift influences the outcome more than selection. This in turn leads to a potential accumulation of deleterious mutations in the absence of recombination, in a process called mutational meltdown or Muller’s ratchet (Muller 1964, Felsenstein 1974). Mutational meltdown is known to cause extinction of small asexual populations and is thought to be one of the main driving forces of the evolution of sex (Lynch et al. 1993, Barton & Charlesworth 1998, Gordo & Charlesworth 2001). Additional to the lack of sexual recombination, mammalian mitochondria have high average rates of mutation of around 1.24 substitutions per nucleotide per billion years, 25 times greater than the rate in nuclear genome (Lynch 1996). The mtDNA in other phylogenetic lineages evolves at half the rate of the mammalian. Because the invertebrate nuclear genomes evolve faster than the mammalian ones, the rate of base substitution in mitochondrial genes of invertebrates is only five-fold greater than in the nucleus. Unless there is recombination that can produce progeny molecules with greater fitness than the parental molecules, mutated or not, there is a danger of accumulation of deleterious mutations, unless heteroplasmic lineages of mitochondria are preserved for very long periods of time. As described earlier, the evidence rather points to deliberate elimination of heteroplasmy than maintaining it in the germ line.
In fact, there is evidence that newly arisen mutations in mtDNA have a higher probability of fixation than those in nuclear genes (Lynch 1996, Lynch & Blanchard 1998). Population surveys of mammalian and fruit fly mitochondrial DNA sequences reveal an excess of intra-species non-synonymous polymorphisms relative to expectation from interspecies comparisons (Ballard & Kreitman 1994, Nachman et al. 1994, Ballard et al. 1996, Nachman et al. 1996, Rand & Kann 1996). The same phenomenon is observed also in the mildly deleterious mtDNA mutations in humans, which have a penetration as high as 5% in populations (Wallace 1994).

The germ line mtDNA bottleneck also gives apparent advantage to any selfish genotypes that can influence either their replication or segregation into the germ cells. However, this is only superficially true. As mentioned earlier, selection at the level of the host individual either in embryos or as an adult is far more believable then selectional screening in the germline. One of the consequences of the bottleneck, with stringent selection on the fitness of offspring, is that the mutational meltdown slows as deleterious genomes reach homoplasmy and are selected out at the level of the offspring of the organism (Bergström and Pritchard 1998). Thus while increasing genetic variance among hosts and reducing variance within host, the bottleneck is actually a very effective way to subject the fitness of multiploid mtDNA to natural selection.

As a conclusion, the evolutionary reason for the maintenance of homoplasmy is to make the mitochondrial genetic component of the organism to behave as haploid genome. The maintenance of genetic integrity by means of recombination might still play a role in the somatic cells accumulating mutations and being unable to segregate their mtDNA.

High mutation rate combined with inability to segregate is somewhat analogous in consequences to the starting situation in the yeast zygote, where the initial heteroplasmy level is high. One of the key protein players involved in the segregation of yeast mtDNA is Mhr1p, whose deficiency delays and overexpression accelerates the vegetative segregation of mtDNA (Ling & Shibata 2004). Besides segregation, Mhr1p is also involved in the gene conversion–type homologous recombination of mtDNA and is required for recombination-initiated rolling circle replication in yeast mitochondria (Ling et al. 1995, 2000, Ling & Shibata 2002, 2004). Rolling circle replication produces linear tandem arrays of mtDNA copies – or concatemers – that are selectively partitioned into daughter buds resulting in homoplasmy within a few cell generations. While concatemers seem to be the segregating units of wild-type mtDNA of *S. cerevisiae*, it seems that that the *rho-* mtDNA in the yeast segregating units or nucleoids is held together by recombination junctions and the resolution of these junctions is essential for the partitioning of mtDNA to daughter cells (Lochson et al. 1995, Doe et al. 2000).
Mitochondrial nucleoids

Mitochondrial DNA is organized in protein-DNA complexes that by analogy to those in bacteria are called nucleoids. In cultured human cells there are approximately 2-10 copies of mtDNA per nucleoid (Iborra et al. 2004, Legros et al. 2004) and several studies have identified around 30 proteins to be nucleoid components in various organisms (Chen & Butow 2005). The proteins offer protection of DNA against damage and bring several copies of mtDNA in contact to allow recombination (Holt et al. 2007). Some are involved in replication or transcriptional control. In mammals the latter include the mtDNA helicase Twinkle, DNA polymerase γ (PolG) and the mitochondrial single-strand binding protein (mtSSB) (Spelbrink et al. 2001, Garrido et al. 2003). One of the most prominent nucleoid proteins is TFAM, in yeast its homologue Abf2p, which belong to the family of High Mobility Group (HMG)-box proteins. These proteins show homology to the evolutionarily conserved, positively charged DNA-binding HMG proteins of nuclear chromatin.

Both Abf2p and TFAM can bind and bend DNA, suggesting that they are involved in packaging of mtDNA (Friddle et al. 2004, Kaufman et al. 2007). Abf2p generally binds dsDNA non-specifically, with some exception of being excluded from poly(dA) sequences that might have regulatory significance (Diffley & Stillman 1991, Diffley & Stillman 1992). Abf2p is also capable of promoting supercoiling in relaxed circular DNA.

Whereas Abf2p seems to be abundant enough in yeast mitochondria to coat mtDNA by having one molecule per 15-30 bp, the balance of current data indicates that this is not the case with TFAM (Maniura-Weber et al. 2004, Cotney et al. 2007). However, TFAM is highly abundant and exists mainly complexed with mtDNA in nucleoid structures (Alam et al., 2003, Garrido et al., 2003). Moreover, it shows significant, non sequence-specific DNA binding (Fisher & Clayton, 1988). Mammalian TFAM has also a number of important roles in the transcription and maintenance of mtDNA, which will be discussed in section 2.15.

In the case of Abf2p, it seems that the physiological concentration of this protein might only allow a limited DNA compacting ability (Brewer et al. 2003). This indicates that the packaging of mtDNA requires other proteins to maintain nucleoid morphology, gene expression and genome maintenance (Chen & Butow 2005). The fact that yeast mutants lacking Abf2p retain mitochondrial nucleoids, although with altered morphology, indicates that other packaging factors are present. Besides its proposed structural role in yeast nucleoids, Abf2p is required for recombination and segregation of mtDNA to daughter cells (Zelenaya-Troitskaya et al., 1998), and
genetic evidence also implicates it in non-recombinational mtDNA repair pathways (O’Rourke et al., 2002).

Additionally, in yeast several metabolic enzymes are localized to nucleoids and have an impact on mtDNA maintenance. These include aconitase, two subunits of α-keto glutarate dehydrerogenase and Ilv5p. It is tempting to suggest that these bifunctional factors are working in response to physiological cues by coupling mitochondrial metabolism with the replication and transcription of mtDNA.

### 2.9 General aspects of DNA replication

DNA is copied or replicated using the complementary strand as a template (Lodish et al. 2000a). In the replication of double-stranded (duplex) DNA, both parental strands are replicated. When the replication is completed, the new duplexes consisting of one of the original parental strands and its copy are separated. Enzymes that replicate DNA are called DNA polymerases. Unlike RNA polymerases, these cannot initiate chain synthesis de novo; instead they require a short pre-existing RNA or DNA strand – primer – to begin strand synthesis. All nucleic acids are synthesized in 5’ (phosphate) → 3’ (hydroxyl) direction, so that the α-phosphate of the succeeding nucleotide is covalently attached to the 3’ hydroxyl residue of the deoxyribose (or ribose) of the previous one. Both prokaryotes and eukaryotes possess several different DNA polymerases, of which some participate in the synthesis of new DNA and others in the repair or recombination of DNA molecules.

Because duplex DNA consists of two intertwined strands, the replication of each strand requires the unwinding of the original duplex by opening the hydrogen bonds between paired strands. This is accomplished by specific unwinding proteins, DNA helicases with the energy provided by NTP hydrolysis. When DNA is replicated, a DNA helicase opens the duplex ahead of the advancing DNA polymerase, resulting in torsional stress and the formation of supercoils on the DNA template. The supercoiling is removed by topoisomerases. In order to achieve a constant advance of DNA replication, the helicase must unwind the duplex and topoisomerases remove the torsional stress ahead of the DNA polymerase, creating a growing replication fork (Figure 2.5).
Figure 2.5. (A) Directionality of DNA synthesis. (B) A diagrammatic view of the key protein players at the advancing replication fork. Orchestrated cooperation of the different proteins results in growing replication fork.

The directionality of DNA synthesis results in continuous synthesis of the leading-strand, which proceeds in the same direction as the helicase, plus a discontinuous synthesis of the lagging-strand that uses the exposed parental strand, which is transiently protected by the single-strand binding protein SSB, as a template. In most known cases the lagging-strand is primed multiple times with RNA primers as the DNA is being exposed at the replication fork, which in turn are elongated by DNA polymerase to form 1-3 kb long Okazaki fragments (Figure 2.6). As the elongation of the fragment reaches the primer of the previous fragment, the primer is removed and elongation continued until the start of the next DNA fragment is reached. The fragments are then joined by DNA ligase to form a continuous DNA strand. Such strand-synchronous DNA replication is fairly common and exceptions are known only from genetic entities having either ssDNA or RNA intermediates in their replication cycle. These examples represent plasmids, viruses and retroelements.

The proteins mentioned above are the key players of the replication fork; however the functional replicase complexes capable of DNA replication on their own are multicomponent machines. The simplest is that of T7 phage with four replisome components (with many more involved in regulation), dozens in E. coli (DNA polymerase III alone has 10 unique subunits) to perhaps hundreds of replication associated proteins in eukaryotes (Zannis-Hadjopoulos & Price 1999, Benkovic et al. 2001, Frouin et al. 2003, Johnson & O’donnel 2005). A minimal mammalian
mitochondrial replisome capable of pre-primed DNA synthesis can be reconstituted with only three proteins: the PolG, the Twinkle helicase and the single-strand binding protein SSB (Korhonen et al. 2004). Despite the evidence of primase activity in various organisms, no mammalian mitochondrial counterparts are known to-date (Li & Englund 1997, Shutt & Gray 2006b, Falkenberg et al. 2007).

In contrast to mitochondrial genomes chromosomal DNAs are huge molecules, whose sheer size explains most of the complexity involved in their replication (Johnson & O’Donnell 2005). Despite intensive research, most of our knowledge of general trends of DNA replication comes from simple models, such as bacteriophages, eukaryotic viruses and protists (Zannis-Hadjopoulos & Price 1999). It seems that different replication systems use the same fundamental components for DNA synthesis, even though the numbers of individual proteins are different and many lack sequence homology. The greatest challenge of the conventional replication machineries is to synthesize simultaneously two DNA strands with opposite polarities. While the leading-strand synthesis is highly processive, the synthesis of Okazaki fragments is discontinuous and creates a problem of coordination for the polymerases. Additionally, for lagging-strand synthesis, multiple initiation events by primase are required, making the exact understanding of the fork progression system difficult.

There are two mechanistically well characterized systems having significant functional analogy to the mitochondrial counterparts: the T7 and T4 bacteriophages. Besides their genomes being simpler and smaller than prokaryotic or eukaryotic chromosomes, many of the key proteins involved in their replication and transcription seem to share a common origin (Shutt & Gray 2006a). Other genetic entities requiring further attention are extra-chromosomal circular bacterial plasmids, which provide insights into various prokaryotic replication mechanisms.

2.10 Replication of T7 phage

The phage T7 chromosome is a linear duplex DNA of 40,000 bp in size, which replicates as a monomer in the early stage of infection, but forms head-to-tail concatemers in the later stages (Romano et al. 1981, Sugimoto et al. 1987). Replication is initiated by the phage RNA polymerase creating a transcript from the primary origin region. This transcript is used to prime light (L-) strand DNA synthesis. As the L-strand replication proceeds through an adjacent AT–rich region, a gp4 T7 primase/helicase recognition site is exposed allowing the synthesis of an RNA primer on the heavy-(H-) strand. At this point replication proceeds bidirectionally and can be observed as a growing replication bubble on the linear genome.
The T7 replisome consists of four proteins: gp5 DNA polymerase, which forms a tight functional complex with the host *E. coli* thioredoxin protein, a dual function gp4 primase/helicase and gp2.5 single-strand binding protein (Benkovic *et al.* 2001). The gp5 protein needs thioredoxin for processivity and is only capable of incorporating 1-15 nucleotides after a primer without its presence, but catalyzes the polymerization of tens of thousands of nucleotides at the rate of 300 nucleotides per second. T7 gp4 belongs to the family of hexameric helicases including *E. coli* DnaB and RecA, being distantly related to the F1-ATPase. It binds a nucleotide (dTTP, ATP) at the subunit interface that is required for assembly and hydrolysis of ATP fuels the helicase action. The gp2.5 binds specifically to ssDNA, approximately seven nucleotides per monomer. It is required for T7 DNA synthesis and cannot be substituted by *E. coli* SSB.

Gp4 protein interacts with single-stranded parts of the forked DNA structure and when powered by NTP hydrolysis translocates unidirectionally 5’ to 3’ on the lagging-strand template at a rate of 114 bp per second. However, it acts most efficiently in combination with DNA polymerase (Stano *et al.* 2005). The discontinuous replication of the lagging-strand is primed by the primase activity of gp4. This recognizes preferred template DNA sequences C(N)2-3 creating Okazaki fragments of 0.5-6 kb in size. The exposed ssDNA at the replication fork prior to primer synthesis is transiently protected by gp2.5. There is substantial evidence that the coordination of leading- and lagging-strand synthesis is achieved by the replisome complex having two DNA polymerases complexed with the helicase, looping the lagging-strand into the opposite direction allowing the synthesis of both strands in the same orientation (Lee *et al.* 2006 [Figure 2.6]).
Figure 2.6. Coupled leading- and lagging-strand synthesis at T7 replication fork (adapted from Lee et al. 2006). The gp4 primase-helicase unwinds the template duplex and adds short RNA primers on the lagging-strand every 0.5-6 kb. DNA polymerase gp5 catalyzes both the leading- and lagging-strand DNA synthesis. SSB gp2.5 transiently protects the exposed single-stranded parental DNA.

2.11 Replication of T4 phage

The genomes of T4 phages are among the most massive viral DNAs, being up to 170,000 bp in length (Mosig 1998, Kreuzer 2000). Like T7 the T4 genome genome is linear, but is replicated mainly by a recombination-initiated mechanism (recombination-dependent replication or RDR). As in T7, the synthesis of leading and lagging-strands in T4 is closely coordinated (Benkovic et al. 2001). The T4 replisome consists of seven proteins; helicase gp41, primase gp61, SSB gp32, various accessory proteins (gp45, gp44, gp62) that are required for processivity on primed or forked duplex templates and DNA polymerase gp43. The replicative cycle starts with one or few rounds of replication from defined origins in the genome. The initiation for some of these uses an origin R-loop, or stable RNA-DNA hybrid formation, which unwinds the origin and provides the initial primer for leading-strand synthesis as in T7. When the replication fork reaches a chromosomal end, the leading-strand will fill the 5’-terminus, but the lagging-strand leaves a short single stranded region at the 3’-end of the genome.
The single-stranded 3’-end can strand-invade a duplex molecule at a homologous sequence, forming a displacement loop (D-loop) and provide a primer for subsequent replication (Mosig 1998, Kreuzer 2000). The peculiarity of the T4 genome is that its genome contains identical terminal sequences at both ends (terminal redundancy); furthermore these sequences are circularly permutated, meaning that the ends are randomly located, in similar fashion as are the “ends” of a randomly cleaved circular molecule (Figure 2.7).

Because of these features, each single-stranded 3’-end can pair with homologous DNA near the opposite end of the genome; either with the daughter molecule or its sister. Replication initiating from the D-loop proceeds unidirectionally towards the centre of the molecule. Because of the circular permutation, when several genomes are present the end of one DNA will generally be homologous to internal regions of co-infecting molecules, generating DNA molecules that are longer than one genome length. The resulting concatemeric DNA is later required for packaging into the phage head (Kreuzer 2000).

Figure 2.7. A simplified depiction of circularly permutated DNA with terminal redundancy. The terminal redundancy enables successful recombination-mediated replication from any point of the genome by providing more opportunities for finding homology that faces away from the end of the genome. This feature is especially important when the first encountered homology is in the newly replicated daughter molecule.

When a replication fork initiated by RDR reaches the end of the genome, the entire process starts again resulting in a chain reaction-like generation of a very complex branched network of DNA molecules. Before or during packaging the branches are resolved by the cruciform DNA cleaving (or Holliday-junction resolving) endonuclease gp49. The directionality of the replication is not well established. However, the invading strand functions as primer for the leading-
strand replication and the replicative helicase complex gp41 is loaded onto the displaced strand of the D-loop. As for origin-initiated replication in the early phase of infection, the replication can also be bidirectional, so that the first Okazaki fragment becomes the primer for the second direction of replication and recruits its own helicase complex. In situations where the cross-strand structure in the invasion point is not resolved, it leaves Y-shaped fork at the site where the bidirectional replication initiated, resulting in a tridirectional intermediate (Figure 2.8). This fork could initiate another round of replication by direct assembly of a replication complex at the structure.

![D-loop formation diagram](image)

Figure 2.8. T4 replication begins with the strand-invasion that primes the first leading-strand synthesis. The first fork can recruit another replisome and initiate bidirectional replication from the site of the strand-invasion. The second leading-strand is originally primed by the Okazaki fragment of the first replication fork. The invading strand can either be cleaved or left behind to form a special tri-directional structure (adapted from Kreuzer 2000).

One of the advantages of such an extensive chromosome replication (ECR) is its tolerance to double-strand breaks (DSB). There is no need for a special coordination to process and connect broken ends, as the ends can be used to initiate normal semiconservative replication when a break is detected. One broken chromosome together with two intact ones produces four intact replication products when replication is unidirectional, five if bidirectional and six if tridirectional.
2.12 Replication of circular bacterial plasmids

Plasmids are extrachromosomal genetic elements that have fairly constant copy numbers within their host cells (Solar et al. 1998). Plasmids are known ubiquitously from Bacteria and Archea, but also in eukaryotes and even some mitochondria. In some Archea up to 25% of the total genetic material in the cell can be plasmids. Plasmids can deliver genes by recombination or transposition mediating genetic exchanges in bacterial populations. Because they can be introduced to new hosts by a variety of mechanisms, plasmids represent an example of genetic material that can be shared among populations or sometimes between species. They also have ways to communicate with the host replication and transcription machineries, providing some analogies to the mitochondrial genetic system. Besides containing loci involved in the control of replication, plasmids can encode a variety of proteins, including those involved in plasmid transfer, resistance to antibiotics or heavy metals, resistance to radiation as well as metabolic genes or components of the translation machinery.

Despite being apparently small and simple, plasmids replicate autonomously. Amongst their essential loci there are generally origins of replication (or ori), and many plasmids encode a protein involved in the initiation of replication, Rep, as well as other genes involved in the control of replication. Despite their importance for molecular biology and biotechnology, there are still major gaps in our understanding of the mechanistic details of plasmid replication. However, three common replication mechanisms for circular plasmids are known.

2.12.1 The theta mechanism

The theta-type is the most common and intensively studied replication mechanism among circular plasmids of gram-negative bacteria. Theta-replicating plasmids have also been isolated from some gram-positive bacteria (Solar et al. 1998). Their replication, with some exceptions, requires plasmid encoded Rep protein for initiation from ori. Additionally the theta origins may include AT-rich sequence repeats and one ore more dnaA sequence boxes, where the host DnaA initiator protein can bind.

The initiation of replication leads to the assembly of the complete replication machinery on ori, including host DNA polymerase III holoenzyme, DnaB helicase and a primase. Once the DNA replication passes a specific checkpoint for leading-strand synthesis, which controls plasmid copynumber, replication continues to completion. As in the textbook model of DNA replication
described in section 2.9, DNA synthesis is continuous on the leading-strand but discontinuous (Okazaki fragments) on the lagging-strand. The synthesis of the two strands is coupled as in T7, but replication may proceed either bi- or more commonly unidirectionally (Figure 2.9). The coupled synthesis of leading and lagging-strands on a circular molecule results in replication intermediates, which look like the greek theta (Θ).

Figure 2.9. Uni- (A) and bidirectional (B) replication. Bidirectional replication requires always two independently recruited replisomes.

Replication terminates at defined sequences, and the terminus is actively determined by protein interactions. One of the best characterized termini are ter sequences in plasmid R6K, which function as a terminus for unidirectional theta replication. The ter sequence is a binding site for the termination protein Tus and interferes with the helicase progression in an orientation-dependent manner. The replication is sequential; strand-synchronous synthesis first proceeds to the terminus in just one direction, terminates and subsequently the origin initiates to the opposite direction and progresses to completion (Figure 2.10. [Lovett et al. 1975]).
Sequences called terH, which arrest lagging-strand synthesis, are found in the plasmid ColE1, where they induce termination through a stable transcript hybridized at the site. The stalling of the replication fork may result from the presence of an unhybridized run of RNA (Solar et al. 1998).

Catenated molecules containing gaps in the daughter-strands are usually the last step of replication. The catenanes can be resolved by either type I or type II topoisomerases; however it is likely that a class II topoisomerase – TopoIV – has a specialized role for unlinking the daughter replicons in vivo. Maturation of the open-circular gapped forms into supercoiled molecules is further facilitated by DNA polymerase for gap filling and DNA gyrase for introduction of supercoils. With e.g. the plasmid R1 it seems that the maturation of newly replicated DNA is a slow process, preventing the direct utilization of the newly replicated molecules.

In theta replicating plasmids, homologous recombination between daughter molecules can also occur. Replication intermediates provide ideal substrates for recombination and these events may result in the generation of dimeric molecules (Nordström 1983).
2.12.2 Strand displacement replication

Strand displacement replication (SDR) is rare and the best studied examples of such plasmids are the members of the promiscuous IncQ family. These plasmids encode three proteins required for initiation of DNA replication: RepA, RepB and RepC (Scherzinger et al. 1991, Solar et al. 1998).

The minimal ori includes three identical sequence repeats plus other typical palindromic sites and inverted repeats. The identical repeats act as RepC binding sites from which RepC together with RepA helicase can induce partial opening of the sequence, causing the exposed inverted repeats to form hairpin structures, which can be recognized by the plasmid encoded RepB primase. The replication initiation is therefore independent of any host initiation factors, requiring only host DNA Pol III and SSB for replication.

Initiation at either of the inverted repeats can occur independently and result in continuous replication with the RepA helicase facilitating the displacement of the parental strand. Replication initiating from both inverted repeats would result in a theta-shaped intermediate in the overlapping regions and displacement loops beyond these regions (Figure 2.11). The end products of SDR are completely or partially single-stranded displaced circles and double-stranded supercoils. Replication of the displaced strand can also be initiated from the exposed hairpin structure of the inverted repeat.

Figure 2.11. Strand displacement replication from either one (A) or two (B) origins on the different strands.
2.12.3 Rolling circle replication (RCR)

Replication by RCR is unidirectional and strand-asymmetric, the leading and lagging-strand synthesis being uncoupled. Most of the RCR plasmids are smaller than 10 kb and can exist as either ssDNA or dsDNA circles when mature (Solar et al. 1998).

RCR is initiated by a plasmid encoded Rep protein, which introduces a site specific nick on the plus-strand of the plasmid in the region termed double-strand origin \( dso \). The \( dso \) has two loci, \( bind \) and \( nic \), of which the former is involved in the sequence specific binding of the Rep protein and the latter represents a conserved sequence that is nicked by the protein. The two loci can be adjacent or separated by up to 100 bp. The nicking at \( nic \) leaves a free 3’-OH that can prime replication by host replication proteins. As a result, the elongation of the end displaces the parental plus-strand and continues until the \( dso \) is met again. Thus the intermediate products of RCR are a dsDNA molecule constituting parental minus-strand and newly synthesized plus-strand and a single-stranded displaced parental plus-strand (Figure 2.12).

![Figure 2.12. Rolling circle replication](image)

Figure 2.12. Rolling circle replication. The circular template duplex is nicked and the DNA-polymerase initiates synthesis from the free 3’–end of the nick. DNA synthesis replaces the nascent strand from the template. Several rounds of replication can occur producing concatemeric replication products. The joining and the subsequent ligation of the displaced strand is catalyzed by the Rep protein.

Besides initiating RCR, Rep protein is also involved in replication termination. Replication of the leading-strand proceeds more than a full round beyond the nick site, enabling hairpin formation by an inverted repeat region at \( dso \). Rep cleaves the hairpin and facilitates the
ligation reaction of the newly synthesized 3’-OH with the 5’ of the nick and liberating the circular dsDNA product. Continuous replication of the leading-strand is possible, producing head-to-tail concatamers of multiple genome sizes that are later processed into circular monomers by recombination.

The final stage of RCR is the replication of the displaced single-stranded parental (+) strand into dsDNA. The replication is primed by RNA (pRNA) and in most cases evidence suggests that host RNA polymerase (RNAP) is involved. RNAP initiates transcription from the plasmid single-strand origin sso sequence, resulting in a 20 nt RNA primer, which is in turn elongated by the host replication machinery. The newly synthesized dsDNA plasmids are subsequently supercoiled by DNA gyrase into their active form.

2.13 Special aspects of replicating linear DNA

When compared to the circular forms, linear genomes present one disadvantage for replication: because every round of replication needs to be primed and replication can proceed only from 5’ to 3’ direction, the ends of a linear DNA molecule cannot be replicated conventionally. Chromosomal DNA has solved this problem by having structures called telomeres, which avoid this problem by utilizing a different mechanism to synthesize DNA in the chromosome ends, thereby preserving the terminal sequence (Verdun & Karlseder 2007). Telomeres consist of G-rich DNA repeats covered by specialized binding proteins. The actual terminus of a telomere is not blunt-ended, but a single-stranded 3’ overhang of the G-rich strand (or G-strand). This overhang can invade homologous double-stranded telomeric tracts, resulting in a large lasso-like structure, known as a telomeric loop (t-loop), providing an elegant mechanism by which chromosome ends can be protected. At present, it is not clear whether t-loops are present at all chromosome ends, whether they are required for chromosome protection, or whether they have a role in regulating other features of telomeres.

Passage of the replication fork through the telomere is thought to generate a blunt-ended leading-strand product and a lagging-strand product with a short 3’-overhang. The G-rich and repetitive nature of telomeric DNA complicates replication as well allowing the formation of secondary structures, such as G-quartets. To prevent this, telomeric proteins support the progressing replication fork, allowing efficient telomere synthesis. The lagging-strand 3’-overhang results from the removal of the most distal RNA primer used for Okazaki fragment synthesis.

Telomere sequences are extended by telomerases, which belong to a protein subgroup of specialized reverse transcriptase enzymes known as TERT (TElomerase Reverse Transcriptases).
TERTs are involved in synthesis of telomeres in humans and many other, but not all, organisms. They take advantage of an enzyme-associated RNA template to synthesize short telomeric repeat sequences and ligate them to the 3'-ends of the chromosomes. Telomerase-based end maintenance is likely to be an ancient evolutionary invention, because it is found in widely diverged eukaryotic lineages from animals to fungi and plants.

Some organisms as well as some human tumours rely on a distinct telomerase-independent method to maintain their telomeres (Fajkus et al. 2005, Muntoni & Reddel 2005). This mechanism is known as alternative lengthening of telomeres (ALT) and is based on recombination and recombination-mediated replication between telomeres. ALT–like activities exist in other systems as well. It is hypothesized that rolling circle replication of telomeres can take advantage of the t-loop formation. The widespread occurrence of t-loops in eukaryote lineages hint that t-loops replicating via rolling circle mechanism might represent a primordial system of telomere maintenance (Fajkus et al. 2005).

A small fraction of known prokaryotic chromosomes are linear and need specialized mechanisms to maintain them. Some bacteria utilize terminal proteins that can prime replication from molecular ends and some, like Borrelia have genomes with covalently closed hairpin ends (Chen 1996, Casjens 1999). Hairpin ends are also known from fungal plasmid DNAs as well as the poxvirus and parvovirus chromosomes.

### 2.14 Replication of mitochondrial DNA

The diversity in organization of mitochondrial genomes discussed earlier implies an equivalent diversity of replication mechanisms. As we know from present day bacteria, the hypothetical free-living α-proteobacteria-like ancestors of mitochondria most likely had a repertoire of replication, repair and recombination mechanisms to accommodate variable environmental as well as biological challenges. In the course of endosymbiotic evolution most of these functions have become unnecessary or even unwanted and have disappeared either through chance or necessity. Similarly some functions have been more conserved in some evolutionary lineages but not in others.

It is also tempting to suggest, that the existence of various replication mechanisms between and within organisms might be due to the need for redundant mtDNA replication initiation modes as a response to changing environmental conditions, physiological challenges related to development, or to circumvent host defences. Analogies are found in present day free-living prokaryotes (Marians 1992). Unfortunately, the functional details of different replication
mechanisms existing in the mitochondria of various organisms are poorly described. In the following sub-chapters, I will give a more general review of the various mtDNA replication mechanisms in the light of what is known from the better established models mentioned earlier.

2.14.1 Theta replication

The best known examples of conventional, strand-synchronous theta replication of mtDNA are the mini- and maxicircle forms of the trypanosome mitochondrial genome, which as described earlier exist as a complex catenated network resembling the chain armour of medieval knights (Shapiro & Englund 1995, Carpenter & Englund 1995). Before replication, the minicircles must be released from the catenated networks by a topoisomerase II. Newly synthesized minicircles containing nicks or gaps are subsequently attached to the network periphery. Maxicircles remain linked to the network but also replicate via theta intermediates. The replication of mini- and maxicircles results in the growth of the mtDNA network. When the network has doubled in size and all minicircle nicks and gaps are repaired, the network splits in two and is segregated into the daughter cells.

2.14.2 Rolling circle replication (RCR)

Rolling circle replication has been suggested for the yeasts Candida glabrata and Schizosaccharomyces pombe mtDNA based on the fact that linear multimers of genome length molecules can be detected (Maleszka et al. 1991, Han & Stachow 1994). In S. pombe mtDNA replication does not seem to have any preferred directionality and either strand can be used as a leading-strand template (Han & Stachow 1994). RCR seems to be the preferred mechanism for mtDNA replication in many other fungi as well as for numerous mitochondrial plasmids in plants and fungi (Maleszka & Clark-Walker 1992, Maleszka 1992, Backert et al. 1997, Hausner et al. 2006). Unlike common RCR producing ssDNA tails in plasmids (see chapter 2.12.3) or in λ-phage (Kiger & Sinsheimer 1971), the mitochondrial RCR intermediates seem to be fully double-stranded, indicating coupled leading and lagging-strand synthesis (Maleszka et al. 1991, Han & Stachow 1994, Backert 2002). In the plant Chenopodium album the RCR of mitochondrial plasmids is initiated by the formation of a short RNA-DNA hybrid (R-loop), eventually creating a transient theta structure prior to strand-breakage, which generates a RCR intermediate (Backert 2002).
Resulting concatemeric molecules are processed into genome length circles by snap-back recombination (Figure 2.13).

![Diagram of snap-back recombination](image)

**Figure 2.13. Snap-back recombination of linear concatemers to circles**

### 2.14.3 Recombination-dependent replication

Interestingly, RDR seems to be the favoured replication method of the mtDNA in many organisms. Strand-invasion can prime rolling circle as well as theta type replication in mitochondria and confusingly an additional RNA-primed initiation mechanism might exist. In fact, the coexisting R-loop-primed mechanisms show surprising analogy to the T4 system.

For instance, the genetically best studied organism, yeast *Saccharomyces cerevisiae*, appears to be able to initiate mtDNA replication either through an R-loop or through RDR. The R-loop-dependent mechanism is primed by mitochondrial RNA polymerase Rpo41 initiating transcription from a replication-origin promoter in sequences termed *ori* or *rep* (Schaeffler 1999). The transcript remains hybridized and is cleaved by an endoribonuclease to form a primer for DNA synthesis. The yeast mtDNA can replicate in the absence of RNA synthesis, indicating that other priming mechanisms are equally important (Fangman *et al.* 1990).

Because head-to-tail concatemers appear to be the major form of mtDNA replication intermediates in *S. cerevisiae*, it was suggested that yeast mtDNA replicates via rolling circle replication (Maleszka *et al.* 1991). Besides these forms, complex aggregates or networks of replicating and recombining molecules can be seen in electron micrographs. These observations
together with the accumulating data on the importance of molecular recombination for yeast mtDNA maintenance lead eventually to the suggestion that yeast would also employ RDR as its main replication mechanism (Williamson 2002). The recent discovery of recombination initiation in ρ− genomes brings further support for this hypothesis (Ling et al. 2007). Because ori and rep elements occur as reiterated tandem elements in ρ− genomes, and these genomes have a segregation advantage (hypersuppressiveness) over wild type genomes, with their inheritance being independent of RNA polymerase, it is likely that the elements are recombination hotspots and function as origins for RDR. This hypothesis was, in fact, originally proposed in the 1980s (Dieckmann & Gandy 1987).

Notably, the mtDNA replication mechanisms of otherwise well known organisms such as S. cerevisiae are not yet clarified, whereas we have more insight in the RDR of mtDNA from less studied groups of organisms, such as malarial parasites (Apicomplexa) and plants. In the apicomplexan Plasmodium falciparum the main replication mechanism involves massive recombination events between linear 6 kb genome units, creating branched structures highly similar to the ones in T4 phage (Preiser et al. 1996). Similar, but even more complex structures with polydispersed linear concateners, tailed molecules and theta structures involving an invading tail have been reported from the mitochondria of plant Chenopodium album (Backert & Börner 2000).

2.14.4 Replication of linear mtDNA

As discussed earlier, mitochondria of fungal and plant species have mostly linear concateners of their genome-size units. Only a small proportion of the total mass of mtDNA can be detected in circular form. Linear molecules from a number of these species often have defined lengths and terminate in specific mitochondrial telomeres (Tomaska et al. 2004). It seems that these structures are maintained by a specific telomerase-independent mechanism, which in Tetrahymena involves inter-telomeric recombination. In Candida parapsilosis, C. salmansis and Pichia philodendri yeasts the telomeres replicate via peculiar minicircular DNA molecules. These minicircles arise either from intramolecular recombination or from t-loops, involving telomere circularization. The minicircles can replicate independently, re-integrate into the mitochondrial genome by recombination or function as a template for linear genome elongation, where the end of the linear molecule first invades the circle and then uses this for rolling circle replication. Interestingly, in Chenopodium the existence of T4-like replication forms follow the growth cycle in suspension.
cultured cells, indicating that there are underlying physiological stimuli behind the phenomenon that might be significant in the whole organism as well (Backert & Börner 2000).

2.14.5 Reverse transcription in mitochondria

Reverse transcription replication involves an RNA intermediate, which is converted into dsDNA by a RNA-dependent DNA polymerase or reverse transcriptase (RT). The mechanism is widespread in viruses (retroviruses) and also exists as the proliferation mechanism of independent, genome-integrated retroelements in all organisms. Retroelements are thought to be ancient retroviruses and have a significant role in the evolution of genomes (Mourier 2005). Besides the yeast group II introns, some mitochondrial plasmids utilize reverse transcription for proliferation. These unique plasmids (retroplasmids) are known from only six species of filamentous fungi and exist as circular or linear entities (Collins et al. 1981, Akins et al. 1988, Mogen et al. 1991, Katsura et al. 2001, Antal et al. 2002). Each retroplasmid encodes its own RT enzyme and replicates via an RNA intermediate.

Phylogenetic data indicates that the mitochondrial retroplasmids are deeply rooted in the phylogenetic tree and are regarded as ancestral retroelements. The existence of telomere-like repeats in pFOXC plasmids from Fusarium indicates that they might have direct evolutionary relationships with eukaryote telomeres (Eickbush 1997). The Mauriceville retroplasmid encoded RT can initiate cDNA synthesis without a primer, suggesting a functional similarity with RNA polymerases (Wang et al. 1993). The nature of the primer for cDNA synthesis in other retroplasmids remains unknown, but pFOXC RT can utilize weakly associated RNA or DNA primers that can even have 3’-mismatches with the template (Simpson et al. 2004).

2.14.6 Replication of animal (vertebrate) mtDNA

The exact mechanism of mitochondrial DNA replication in animals has caused controversy in recent years. The mtDNA replication mechanism from insects to mammals has been thought to be very similar. However most recent work has concentrated only on mammalian and avian tissues and cultured cells. An intense debate of the exact replication mechanism(s) is still ongoing. Some of the work of this thesis was done to address the issue, including the original articles II and submitted manuscript IV. The submitted manuscript IV will also bring forth evidence of a novel replication-
related mechanism operating in some human cells. The proposed mechanism and their consequences thus will be discussed in more detail in the discussion section of this thesis.

2.14.6.1 Strand-displacement model (SDM)

The long prevailing paradigm in the field describes an unusual mode of strand-displacement replication, where leading- and lagging-strands are synthesized asymmetrically and unidirectionally from separate and well-defined origins, O\textsubscript{H} (heavy- or leading-strand origin) and O\textsubscript{L} (light- or lagging-strand origin) (Robberson \textit{et al.} 1972, Robberson & Clayton 1972, Crews \textit{et al.} 1979, Clayton 1982).

Replication starts from the O\textsubscript{H} origin and is primed by a processed transcript from the light-strand promoter (LSP). This processed RNA strand is highly persistent, forming an R-loop structure in the non-coding region of mtDNA (Shadel & Clayton 1997, Lee & Clayton 1998). The model elegantly links mitochondrial transcription to replication and provides a possible mechanistic way for mtDNA proliferation to respond to various stimuli.

The endonucleolytic processing of the LSP transcript is proposed to be carried out by a site-specific endonuclease, RNase MRP, which is capable of cleaving transcripts sequences spanning from CSB II to CSB III. Interestingly, the substrate specificity of RNase MRP resides in an RNA component encoded by a nuclear gene. Similar ribonucleoproteins are known also from yeast and the frog \textit{Xenopus laevis}. These conclusions have been questioned and instead it has been proposed that CSBII functions as a powerful transcription terminator sequence (Falkenberg \textit{et al.} 2007). Thus the H-strand primer formation could be explained by a protein-modulated transcription termination at CSB II.

The DNA polymerase PolG replicates the nascent H-strand, and with the help of a helicase, displaces the parental H-strand while processing. Consequently, the single-stranded displaced strand is protected by single-strand binding proteins (SSB) from degradation (Figure 2.15). When the replication reaches two-thirds of the genome, O\textsubscript{L} initiates. O\textsubscript{L} is flanked by tRNA clusters and it is suggested that when released in single-stranded form, it tends to form a hairpin structure. A putative primase is proposed to recognise the hairpin and generate an RNA primer for the lagging-strand, after which L-strand replication begins. However, the evidence for such a primase is weak. After replication, the primers are removed by RNase H activity and gaps filled. More recently, additional L-strand origins have been suggested (Brown \textit{et al.} 2005). It is not known
how these origins are primed, but they could enable faster dsDNA transition of lagging-strand replication intermediates.

Figure 2.15. The strand displacement model of mammalian mtDNA replication. The leading-strand replication initiates from $O_h$ and proceeds two-thirds of the genome before the lagging-strand replication is initiated from $O_l$.

2.14.6.2 Strand-synchronous model

Studies involving two-dimensional neutral-neutral agarose gel electrophoresis (2DNAGE) methodology have led to the proposal of a strand-coupled mechanism operating from less discrete origins (Holt et al. 2000, Yang et al. 2002, Bowmaker et al. 2003, Yasukawa et al. 2005, Yasukawa et al. 2006). This mechanism basically involves theta-replication and was originally suggested to function as an alternative to the strand displacement model. Initiation and priming of this replication mode remain unclear. However, there are no mechanistic constraints why the initiation and priming for the leading-strand as described for the SDM would not apply.

One peculiarity of the model is that the lagging-strand DNA is synthesized with delay and the newly replicated lagging-strand is originally laid down as RNA. As a result the RNA rich mode of replication can be detected as RITOLS (Ribonucleotide Incorporation Through-Out the
Lagging-Strand) intermediates (Figure 2.16). RITOLS intermediates are subsequently replaced by or matured to DNA. In this model O_{H} remains the prominent replication origin for the leading-strand, but also functions as the terminus. Replication originating from O_{H} proceeds unidirectionally through the genome. RITOLS replication intermediates (RIs) are prone to RNase H degradation during mtDNA extraction, resulting in single-stranded parental H-strand, explaining the observations that lead to the generation of the strand-asynchronous model (Yang et al. 2002).

Figure 2.16. RITOLS replication of mammalian mtDNA. Strand-coupled, unidirectional replication initiates from O_{H}. Initially the lagging-strand is laid down as RNA, but is subsequently matured to DNA.

Considering lagging-strand synthesis, the mechanism results in superficially similar molecules as the displacement model, although unlike in the SDM, the lagging-strand also terminates at O_{H}. The key challenge in the future will be to show what mechanisms are responsible for persistent RNA formation on the lagging-strand and how the lagging-strand maturation or synthesis is initiated. Based on the same methodology as the previous observations, it seems that a minority of mtDNA molecules take advantage of a more Conventional, Strand-Coupled Okazaki Fragment-Associated replication (COSCOFA) that has no persistent RNA-DNA hybrid formation and is initiated from a broader zone (Holt et al. 2000, Yasukawa 2005). This replication mode seems to be preferred in cultured cells that are amplifying their mtDNA after drug-induced depletion. The replication initiating from this zone starts bidirectionally, but because initiation occurs close to the O_{H} the replication still proceeds through most of the molecule in an apparently unidirectional way (Figure 2.17).
Figure 2.17. COSCOFA replication of mtDNA. Bidirectional and strand-coupled replication initiates from a broader zone (ori-z). Replication proceeding to the Oh direction terminates at Oh, resulting in unidirectional replication of the rest of the molecule. The lagging-strand synthesis does not involve extensive RNA runs.

In rodent and chick liver as well as in untreated cultured human cells RITOLS intermediates are the predominant class (Holt et al. 2000, Yasukawa et al. 2005, Yasukawa et al. 2006).

2.14.7 Proteins of the mammalian mitochondrial replisome

All proteins involved in mammalian mtDNA replication are encoded in the nucleus, translated by cytosolic ribosomes and imported into the mitochondria. However, when compared to bacterial, phage or even nuclear counterparts, only a limited number of mitochondrial replicative proteins have been identified. Among these are a DNA polymerase (POLG1) and its accessory subunit (POLG2 [see Kaguni 2004 for review]), a DNA helicase (Twinkle [Spelbrink et al. 2001, Korhonen et al. 2003]) and the mitochondrial single-stranded DNA-binding protein (mtSSB [Tiranti et al. 1993]).

A minimal mitochondrial replisome capable of genome-length DNA rolling-circle replication on an artificial template can be constructed using PolG holoenzyme consisting of a functional heterotrimer (POLG1 + 2x POLG2), Twinkle and mtSSB (Korhonen et al. 2004). Several components of the mitochondrial replication and transcription machinery (including RNA
polymerase MTRPOL) show similarity to their counterparts in the T-odd bacteriophage family, suggesting that a T-odd phage ancestor made a major contribution to the early endosymbiosis event (Shutt & Gray 2006a). As an example the mtDNA helicase Twinkle has striking similarity with the T7 phage primase/helicase protein gp4 (Spelbrink et al. 2001).

However, whereas the Twinkle homologues in more primitive Eukaryotes have primase function, the metazoan Twinkle, based on sequence comparison, seems to have lost this (Shutt & Gray 2006b). This prediction is backed by the fact that numerous laboratories working on Twinkle have failed to detect any primase activity of the protein. A primase activity has been detected from mammalian mitochondrial extracts, but the data is controversial and no candidate proteins have been suggested (Wong & Clayton 1985, 1986). The most tempting choice would be MTRPOL. This would provide a means for the synthesis of extensive RNA on the lagging-strand in the RITOLS mechanism of mtDNA replication. In fact, the *E. coli* RNA polymerase is capable of synthesizing persistent RNA-DNA hybrids on a single-stranded DNA template (Chamberlin & Berg 1964). However, it is hard to imagine how this might work in the hypothetical Okazaki fragment priming of dsDNA rich COSCOFA replication.

2.15 mtDNA binding proteins in mtDNA replication

Besides the factors involved directly in replisome function, in most genetic systems several additional proteins are required for efficient replication of the genome. Many of these have general roles in DNA maintenance, being involved in the expression of genes as well as protection and packaging, but others have more defined roles in mediating replication initiation as well as replisome movement.

Not much is known about the proteins involved in initiation of mtDNA replication, although certain outlines can be deduced from our understanding of transcription initiation. The mitochondrial transcription factor TFAM is the best studied DNA binding protein in human mitochondria. TFAM is a homologue of yeast Abf2p and likewise an abundant HMG-box protein in mammalian mitochondria (Kang & Hamasaki 2005). It is required for transcriptional initiation *in vitro* at the heavy- or light-strand promoter of the genome together with RNA polymerase and the additional transcription factor TFB1M or TFB2M (Fisher & Clayton 1988, Falkenberg et al. 2002).

Because transcription from the light-strand promoter is required to create the primer for heavy-strand mtDNA synthesis in the strand displacement model, TFAM has also been considered to be an essential protein for mtDNA replication. As mentioned in chapter 2.8, TFAM has also been
proposed to play a structural role in the maintenance of the mitochondrial nucleoids, like Abf2p. However, Abf2p is not involved in transcription of yeast mtDNA. Abf2p can stabilize recombination junctions influencing mtDNA segregation (Zelenaya-Troitskaya et al. 1998). Analogously it has been reported that the mammalian TFAM binds preferentially to Holliday junctions (Ohno et al. 2000), but this feature might also result from its general DNA binding properties. Despite not having a protective role against oxidative damage, TFAM overexpression in rat myoblasts has been reported to speed up the re-establishment of mtDNA levels after peroxide damage (Noack et al. 2006). Similarly, in ischemia, the cardiac muscle surrounding the damaged area seems to be protected from cardiac remodeling in transgenic mice expressing human TFAM (Ikeuchi et al. 2005).

In the replication of bacterial genomes and plasmids, the terminus or pause-binding proteins have important functions. The Tus protein in *E. coli* is capable of binding chromosomal Ter sequences and modulating replication pausing in them by directional contrahelicase activity, stably trapping the replication fork at the terminus coming from one direction but not the other (Neyon et al. 2005, Mulcair et al. 2006). Interestingly Tus can also block transcription in a similar polar fashion and transcription coming from the permitted direction can relieve the replication block (Mohanty et al. 1996). A similar type of transcription-sensitive replication pauses or barriers can be found in yeast and bacterial genomes, often associated with rRNA genes (Mohanty et al. 1996, Krings & Bastia 2005). Similar replication pauses are known in mtDNA, but mitochondrial Tus-like proteins have not been characterized (Mayhook et al. 1992, Holt et al. 2000, Reyes et al. 2005). The only mitochondrial protein known to have contrahelicase activity *in vitro* is mtDBP, a homologue of the mammalian mTERF from sea urchin (Polosa et al. 2005).

## 2.16 mtDNA repair

Different environmental and internal insults can damage DNA. These environmental factors include radiation and some toxic chemicals, but most of the damage is actually due to unprovoked errors in DNA replication or reactive oxygen species (ROS) originating from the electron transport chain. In order to maintain the integrity of the genome the cell needs to have ways to repair such damage. A variety of DNA repair enzyme activities have been detected in mitochondria (Bogenhagen 1999, Bohr & Dianov 1999). Mitochondria seem to be especially efficient in repairing oxidized and alkylated DNA lesions via the base excision repair (BER) pathway (Driggers et al. 1997, Bohr & Dianov 1999). Three uracil glycosylases have been identified, as well as an apurinic endonuclease
activity that is UV-inducible (Anderson & Friedberg 1980, Tomkinson et al. 1990, Kalinowski et al. 1992, Slupphaug et al. 1993, Caradonna et al. 1996). There is also evidence that mitochondria can repair single- and double-strand-breaks as well as intrastrand cross-links, indicating that they might repair DNA via homologous recombination (Thyagarajan et al. 1996, Grishko et al. 1999, LeDoux et al. 1999). Interestingly, different cells vary in their ability to repair mtDNA (LeDoux et al. 1998, LeDoux et al. 1999). However, nothing is known of the tissue-specificity or developmental regulation of mtDNA repair.

2.17 General aspects of recombination in regard to mitochondrial DNA

In homologous recombination two homologous duplex molecules exchange parts and are subsequently resolved so that no tangles or unevenness results (Lodish et al. 2000b). This is achieved by formation of double-strand break, resection of dsDNA to free 3’-ends and subsequent alignment of the two homologous DNA molecules (Cunningham et al. 1980, Cassuto et al. 1980, Szostak et al. 1983). The free 3’-end invades the homologous dsDNA in a process called strand exchange and second end capture results in the generation of four-stranded intermediate, containing two branch points – or Holliday junctions (Schwacha & Kleckner 1995). These Holliday junctions can move along the molecules in a process called branch migration. Homology search and strand-invasion is promoted by a special enzyme called a recombinase, such as RecA in bacteria (Cunningham et al. 1980). The resolution of the recombining molecules is carried out by a special endonuclease enzyme that cleaves branched structures. In DSBR model of recombination, cutting the two Holliday junctions in the same or opposite directions generates non-crossovers or crossovers (Figure 2.18).

In meiosis, genetic recombination mixes the alleles of the parental chromosomes creating more allelic combinations in the offspring. Double-strand breaks (DSBs) in DNA readily initiate recombination, providing an important means of repairing such damage without losing any information (Pâques & Haber 1999). Besides reciprocal crossing over, recombination mediates also non-reciprocal gene conversion, where one allele is replaced by another coded by the sister duplex. This is achieved either by deletion of the allele through double-strand gap formation when ends of the DSB are recessed or by mismatch repair of exchanged strands of the heteroduplex having non-homologous regions (Fogel et al. 1981, Surtees et al. 2004). Crossing over and gene conversion are highly correlated (Fogel et al. 1981, Symington & Petes 1988), however a there is locus specificity, and a number of yeast mutants defective in crossover have normal gene conversion frequencies.
It seems that there are separate pathways for noncrossover and crossover recombination. In yeast *Saccharomyces cerevisiae* the two types of recombination diverge early in the meiosis (Allers & Lichten 2001). It is likely that the noncrossover-type recombination is mediated by synthesis-dependent strand annealing mechanism that is commonly active in the mitotic DSB repair (Figure 2.19).

**Figure 2.18.** Molecular recombination between two homologous DNA duplexes (adapted from Szostak *et al.* 1983).

**Figure 2.19.** of Double-strand break repair via homologous recombination involving synthesis-dependent strand annealing mechanism (after Allers & Lichten 2001).
Besides increasing genetic diversity by combining parental gene sets, molecular recombination is a highly important process for any genetic entity to maintain its genome integrity (Hey 1998, Gessler & Xu 2000, Palsson 2002). It has long been known for long that plants, fungi and some other organisms have active mtDNA recombination (reviewed by Barr et al. 2005) and even some mitochondrial RecA homologues are known (Khazi et al. 2003). In fact, it has been suggested that the early eukaryotes inherited their recombinases from the ancestral endosymbionts (Lin et al. 2006). There is also genetic evidence for some metazoans having recombinogenic mtDNA (Ladoukakis & Zouros 2001, Burzynski et al. 2003). However, the lack of genetic markers and uniparental inheritance has previously made it difficult to detect recombination events in mammalian mitochondria. Molecular recombination activity has been detected in lysates of mammalian mitochondria of cultured cells, and branched molecules with Holliday junction characteristics have been observed in human cardiac muscle mtDNA (Thyagarajan et al. 1996, Tang et al. 2000, Kajander et al. 2001). Many investigators have detected recombinant mtDNA molecules in hybrid crosses where the mechanism responsible for maintaining uniparental inheritance seems to be defective (as reviewed by Ballard & Whitlock 2004). Fairly recently a case of biparental transmission in a human individual with a fraction of recombinant mtDNA between the parental types was reported (Kraytsberg et al. 2004). A follow-up study also demonstrated frequent recombinant mtDNA molecules in individuals with multiple heteroplasmy (Zsurka et al. 2005). Based on this evidence it is safe to say that recombination is part of the natural molecular armoury of mammalian mitochondrial DNA maintenance, although the key enzymatic players and the direct biological consequences of this feature remain to be elucidated.

2.18 Tissue-specific requirements for energy production and mtDNA copy number control

Different tissues have different energy requirements, which is evidenced by the mass and density of mitochondria in the tissue. The control of mitochondrial biogenesis is extremely complex, involving hundreds of genes (Moraes 2001). In skeletal muscles, contractile activity, such as in the case of chronic excercise, results in mitochondrial biogenesis and development of type I muscle fibers (Hood 2001). Similarly, mice feeding on a high fat diet plus a heparin supplement to increase the concentration of free fatty acid in the blood, have induced mitochondrial biogenesis in skeletal muscle (Garcia-Roves et al. 2007). The process is induced by the upregulation of PPARγ (Cresci et
al. 1996), PPARα (Lehman et al. 2000) and PGC-1α–transcription factors, which interact to promote the oxidative capacity of skeletal muscle by stimulating the transcription of genes involved in mitochondrial biogenesis (Hood 2001). These genes act upstream of genes encoding for the transcription factors NRF-1 and NRF2, which also themselves promote mitochondrial biogenesis.

CRE-binding proteins can also induce PGC-1α gene transcription (Wu et al. 2006). They have also been shown to function as a calcium- and cAMP-sensitive mediator, linking calcium in the sarcoplasmic reticulum with the activation of the transcriptional processes that induce mitochondrial biogenesis (Koo et al. 2005). The regulation of mtDNA copy number is unclear (Moraes 2001). For example, genes regulated by NRF-1 and NRF-2 are mainly housekeeping genes, and thus are not directly involved in the copy number regulation process (Moraes 2001). More likely they are a part of the link between mitochondrial biogenesis and external stimuli - such as exercise (Wang et al. 1999) and hormones (Moraes 2001).

Data from cultured cells suggests that cells tend to measure mtDNA molecule amounts rather than the number of gene copies. For instance mouse cells having only circular dimers maintain 65% more genomic units than cells with monomeric mtDNA (Clayton 1982). This is also true in the case of pathological deletions. In tissues, it has been suggested that the increases in energy production might be achieved mainly by increase in mitochondrial biogenesis rather than copy number control and the increase in copy number (e.g. in muscles after intense training) would be just a representation of mitochondrial mass. However, direct measurement of mtDNA copy number per mitochondrion in different mouse tissues clearly shows that the copy number per mitochondrion varies organ-specifically and is related to the energy requirements of the tissue (Veltri et al. 1990). Moreover, the overexpression of human TFAM and mouse Twinkle can specifically increase mtDNA copy number in transgenic mice (Eckstrand et al. 2004, Tyynismaa et al. 2004, Ikeuchi et al. 2005).

2.19 mtDNA in cardiac function and developmental adaptation

The heart is the most energy demanding tissues in the human body consuming 0.1 ml O2/g per minute – an amount that is only surpassed by insect flight muscles (Goffart et al. 2004). Heart has also impressive reserve capacity; human cardiac muscle oxygen consumption can rise four-fold without any effect on the rate of ATP consumption and synthesis. No oxygen debt accumulates as seen in skeletal muscles under prolonged exercise. The continuous circulatory pumping of blood as well as maintenance of ion homeostasis in the cardiomyocytes is dependent on healthy
mitochondrial function (Russel et al. 2005). The importance of oxidative metabolism is highlighted by the high mitochondrial content of heart muscle cells, where up to 40% of the cellular volume can consist of mitochondria. Heart also has two defined subpopulations of mitochondria with slightly different biochemical properties (Judge & Leeuwenburgh 2007). Subsarcolemmal mitochondria (SSM) are located under the plasma membrane and interfibrillar mitochondria (IFM) are arranged in parallel rows between the myofibrils inside the cardiomyocyte. Generally, IFMs oxidize substrates quicker than SSMs and have higher activities of citrate synthase and succinate dehydrogenase.

The high energy requirements manifest also in the mtDNA content. The few studies with adult human cardiac muscle propose a mean copy number of 6970 ± 920 to 9235 ± 5457 with individual numbers ranging from 4 000 to up to 34 000 (Miller et al. 2003, Frahm et al. 2005). For comparison, skeletal muscle cells, depending on the muscle, have copy numbers ranging from 1000 to 4000 (Barthelemy et al. 2001, Miller et al. 2003, Frahm et al. 2005).

Due to the highly active metabolism in adult heart there are substantial changes in the mitochondrial physiology of the heart muscle during development (Lopachuk et al. 1992). The fetal heart mostly utilizes lactate and glucose as the main energy substrates, whereas adult heart relies almost entirely on the β-oxidation of fatty acids. Both mitochondrial mass and mtDNA copy number increase during development in anticipation of the O2-rich environment outside the mother (Marin-Garcia et al. 2000).

### 2.20 Human mitochondrial disorders and heart pathology

Amongst hereditary cardiomyopathies, disorders affecting cardiac energy metabolism are one of the most important (Kelly & Strauss 1994). Many of these affect OXPHOS and β-oxidation, but the underlying molecular mechanisms are diverse and not fully understood (Antozzi & Zeviani 1997). Cardiomyopathies associated with mitochondrial defects can be divided into hypertrophic obstructive cardiomyopathy (HCM) and dilated cardiomyopathy (DCM). The pathological cause can often be pinpointed to mtDNA point mutations and large scale rearrangements or defects in nuclear-coded mitochondrial genes. It should be noted, that also healthy human cardiac muscle has fairly high numbers of rearranged mtDNA molecules compared to other tissues (Kajander et al. 2000).

Hypertrophic cardiomyopathy, where the ventricular wall and especially the septum is thickened resulting in obstructed blood flow, is the most common manifestation of heart

Human mitochondrial syndromes resulting from sporadic mtDNA rearrangements, such as Kearns-Sayre syndrome (KSS) and Pearson’s syndrome also often have cardiac muscle manifestations (Finsterer 2006). In these syndromes actual cardiomyopathy is unusual, but branch blocks, atrioventricular or complete heart blocks in conduction have been reported (Rahman & Leonard 2000, Katsanos et al. 2002, Chawla et al. 2007).

Large scale rearrangements and deletions are often seen in human syndromes affecting mtDNA maintenance (see Copeland 2007 for review). Curiously, a heart phenotype in these syndromes is rarely seen, suggesting that heart mitochondrial DNA maintenance might be more resistant to replication defects than, for example, neurons or skeletal muscle. The few reported cardiac manifestations in DNA maintenance disease are associated with autosomal recessive progressive external ophthalmoplegia (arPEO) (Bohlega et al. 1996). However, some other autosomally inherited diseases involving mitochondria, such as Friedreich’s ataxia (FRDA) can result in mitochondrial iron accumulation and hypertrophic cardiomyopathy together with neurodegenerative disease (Pandolfo 2002, Michael et al. 2006). The pathological mechanism of FRDA is thought to be an increase in oxidative stress in mitochondria. The heart-specific human
pathology can be phenocopied in FRDA transgenic mice and is correlated with increased oxidative stress (Al-Mahdawi et al. 2006). Similarly, knockout of manganese superoxide dismutase (MnSOD), one of the enzymes responsible for elimination of reactive oxygen species in mitochondria, resulted in dilated cardiomyopathy in mice (Li et al. 1995). Studies with Ant1 knockout mice suggest that heart muscle has lower antioxidant defences than other tissues resulting in more severe cellular damage and DNA rearrangements (Esposito et al. 1999). The heart-specific inactivation of the murine TFAM gene also results in mtDNA depletion, dilated cardiomyopathy and conduction block, but is not directly relevant to any human condition (Larsson et al. 1998). Instead, the multiple functions of TFAM might make it fairly difficult to pinpoint the exact molecular mechanism apart from the loss of mtDNA.

Cardiac hypertrophy is caused by an increase of cardiac myocyte size and proliferation of mitochondria together with contractile structures. Despite the proliferation, hypertrophic heart mitochondria are unable to meet the energy requirements of the heart and there is a downregulation of the signalling pathways involved in mitochondrial biogenesis. As a result, there is a transition from fatty acids to glucose-burning metabolism, analogous to a more embryonic-like state. A similar switch has been detected also in the TFAM cardiac knockout mice (Hansson et al. 2004) and the same tendency is seen in aging mouse hearts (Lee et al. 2002). The distinction between cardiac hypertrophy and DCM is not always clear, and often hypertrophy precedes DCM in the clinical evolution of cardiomyopathy.

### 2.21 Mitochondria, ischemia, preconditioning and damage response in human heart

Active energy production, oxygen supply and modulation of physiological signals in heart muscle have important clinical implications in humans. Both a prolonged period of hypoxia or ischaemia as well as the subsequent sudden flow of oxygenated blood (reperfusion) can damage the heart muscle irreversibly (Halestrap et al. 2007). When O₂ is limited, cardiac creatine phosphate is rapidly depleted and fatty acid- as well as pyruvate-oxidation is inhibited resulting in impaired ATP production, accumulation of lactate and subsequent intracellular acidosis, which directly inhibit cardiac contractility. Sustained ischemia eventually results in ATP depletion and necrotic cell death.

Paradoxically, reperfusion injury can be equally destructive. The activation of the Na⁺/H⁺ antiporter during lactic acidosis to restore pH is inhibited due to declining ATP concentration, leading to an increase in intracellular Na⁺. Concomitantly this leads to the
accumulation of intracellular Ca\(^{2+}\), because the Na\(^{+}\)/Ca\(^{2+}\) antiporter is also inhibited. ATP becomes rapidly degraded into ADP and AMP, AMP into adenosine and other nucleosides, which leak out of the cell causing local vasodilatation. Moreover, the leakage results in depletion of adenine nucleotides inside the cell.

The reperfusion of oxygen is associated with a burst of reactive oxygen species (ROS), the source of which is unclear but might involve complexes I and III of the ETC. The restarted electron transport chain is at least transiently short of adenine nucleotides, resulting in reduced cardiac performance or stunning upon reperfusion. ROS can cause thiol oxidation and inhibition of the ATP-synthase and other disulphides in proteins, further exacerbating the effects of ATP deprivation. ROS also cause peroxidation of unsaturated fatty acids in plasma- and mitochondrial membranes. It is thought that the combined effects of ROS and elevated Ca\(^{2+}\) play the main role in irreversible reperfusion injury, when the mitochondrial permeability transition pore (MPTP) is opened, cytochrome c leaks out and causes apoptotic cell death as well as substantial release of intracellular proteins resulting in necrotic reperfusion injury and inflammation. The opening of MPTP can be prevented by drugs or by ischemic preconditioning (IP), where the heart is exposed to brief periods of ischemia. Preconditioned heart muscle exhibits smaller infarct size and less necrotic cell death while preserving their contractility. IP effect is thought to arise from gradual adaptation of the mitochondrial signalling pathways.

Inflammation resulting from ischemic heart muscle damage may also cause cardiac muscle remodelling in dilated cardiomyopathy (DCM), where the heart’s pumping efficiency diminishes due to the enlargement of the left ventricle (Pankuweit et al. 2004). DCM is progressive and eventually leads to cardiac failure. Concomitant with the disease progression mitochondrial respiratory dysfunction and depletion of mtDNA copy number can be detected (DiMauro & Schon 2003, Marin-Garcia & Goldenthal 2004). Cardiac specific knockout of the TFAM gene or HSP40 results in depletion of mtDNA copy number and eventually DCM and heart failure in mice (Larsson et al. 1998, Hayashi et al. 2006). In agreement with the data, mice overexpressing TFAM do not develop DCM as easily as control littermates after myocardial infarction (Ikeuchi et al. 2005). The cardiac muscle remodelling in DCM is likely due to the retrograde signalling between mitochondria and the nucleus, but the mechanisms remain unknown (Marin-Garcia et al. 2006).

Acquired cardiac hypertrophy is primarily a compensatory reaction resulting from increased workload (Goffart et al. 2004). In the case of mitochondrial diseases, the myocyte function is impaired and a compensatory growth is provoked as a reaction to appropriate body circulation. Prolonged hypertrophy almost inevitably leads to decompensation, resulting finally in cardiac insufficiency and eventually cardiac failure.
3 AIMS OF THE STUDY

This study was designed to address the significance of the abundant recombination in human heart mtDNA (Kajander et al. 2001). The main goals can be defined as:

A. To study the differences and similarities of mtDNA replication mechanisms in different tissues
B. To identify proteins involved in the modulation of replication and recombination in cultured cells and animal models
C. To clarify the physiological significance of recombination or RDR in heart muscle mtDNA

4 MATERIALS AND METHODS

The following sub-chapters describe the source of materials as well as the main methods used in the studies including parts that were not included in the articles. Regarding the articles, the focus is on specific methods employed by myself. For more detailed descriptions of other methods see the original publications I-IV and references therein. Some clarifying explanations are provided, where they are necessary for the more general understanding of the work.

4.1 TFAM constructs

A full-length TFAM cDNA was amplified from total HeLa cell cDNA using the following primers (restriction sites as indicated, underlined, plus start and stop codons in bold italics):

5’-CCGGAATCCGCATGGCGTTTCTCCGAAGC-3’ (EcoRI) and
5’-CGCGGATCCACACTCCTCAGCACCATATTTCG-3’ (BamHI)

The product was verified by cloning into Topo-blunt (Invitrogen) vector and subsequently sequenced to confirm its identity. The TFAM-insert was cut out from the Topo-vector and re-ligated into EcoRI + BamHI-cut pcDNA3.1(-)Myc-HisA (Invitrogen) to create the construct
mtTFA-myc. This construct was then used as template to amplify a full-length variant of TFAM containing its normal stop codon (TFAM-stop), using the following primers (restriction sites as indicated, underlined, plus start and stop codons in bold italics):

5´-CCCAAGCTTGCAGTTGCCTTCTCCGAAGC-3´ (HindIII) and 5´-CGCGGATCCATTACCTCCTCAAGCACCATAATTTTC-3´ (BamHI).

The restriction-digested PCR product was ligated into HindIII + BamHI-cut pcDNA3.1(+) vector (Invitrogen), to create the construct mtTFA-pcDNA3.1 for transient expression. Because TFAM has two putative starting codons (AUGs 133, 155), I also created a pcDNA3.1(+) construct using the following forward primer (mtTFA-155-pcDNA3.1):

5´- CCC AAG CTT AAG GCG TTT CTC CGA AGC–3´

When I analysed the transient expression in HEK293T cells I detected and subsequently cloned also the published splice variant of TFAM (Tominaga et al. 1993). This clone was used to create two constructs using either of the AUGs (mtTFA-sv133, mtTFA-sv155). See Figures 4.1 and 4.2 for details.

Figure 4.1. Alternative AUGs in TFAM mRNA sequence (GenBank accession M62810).
Figure 4.2. (A) Diagrammatic representation of the TFAM gene including the alternative splice site. (B) RT-PCR detecting both splice variants in HEK293T cells. (C) TFAM splice variant sequence (capitals) lacking exon V as cloned from (B). The variant matches the one described by Tominaga et al. (1993).

Both TFAM-stop and TFAM-myc his constructs were used to create inducible Flp-In™ T-Rex™-293 (Invitrogen) cell lines by recloning the inserts into pcDNA5/FrT/TO vector (Invitrogen) by Pmel (Fermentas) digestion and subsequent ligation.

4.2 Twinkle and PolG constructs

The cDNA of POLG1 and Twinkle variants used in paper II were originally cloned in the pcDNA3.1(-)/Myc-His A (Invitrogen), as previously described (Spelbrink et al., 2000; Spelbrink et al., 2001). The corresponding Flp-In™ T-Rex™-293 cell lines containing the relevant insert were created by recloning into pcDNA5/FrT/TO vector.
4.3 mTERF constructs

The mTERF gene was amplified from HeLa cell cDNA (Spelbrink et al., 2000) and cloned into pcDNA3.1(-)/Myc-His A (Invitrogen) vector DNA to create the mTERF and mTERF-MycHis inducible expression constructs, using a similar strategy as before. See paper III for further details.

4.4 Mitochondrially targeted RecA constructs

In order to study effects of a known recombinase on mtDNA maintenance, I used a construct expressing the *E. coli* RecA protein N-terminally fused to the MnSOD mitochondrial localization signal, kindly supplied by Dr. P. Lestienne (Université Victor Segalen, Bordeaux). The creation of the MnSOD-RecA construct was described in Paul et al. (2001).

4.5 Sequencing

The identities of the different transgene constructs and DNA probes used for hybridization (4.22) were confirmed by sequencing using BigDye terminator kit (Applied Biosystems). Primers were universal M13, T7, BGH for vectors, or specific primers for transgene. The sequencing reaction products were analyzed by capillary electrophoresis using ABI 3130 Genetic Analyzer and manufacturers’ software.

4.6 RNA interference studies (RNAi)

TFAM siRNAs were synthesized using the Silencer™ siRNA construction kit (Ambion). The basic of the kit is that siRNAs are created from designed oligonucleotides by *in vitro* transcription using T7 RNA polymerase. Transcripts are hybridized and enzymatically digested to yield the desired dsRNA product.

Six putative TFAM-specific siRNA sequences were selected using the manufacturer’s prediction programme (www.ambion.com/techlib/misc/siRNA_finder.html). After testing by transfection and Western blotting, two were found to be efficient in post transcriptional silencing of TFAM gene expression in HEK293T cells:
Si2: 5´-AAGTTGTCCAAAGAAACCTGTCCTGTCTC-3´ (np 275-293, exon 2)

Si4: 5´-AAGATGCTTATAGGGCGGAGTCCTGTCTC-3´ (np 431-451, exon 4).

The same approach was used for finding suitable siRNAs for mTERF, see (III).

4.7 Cell culture and transfections

HEK293T and Flp-In™ T-Rex™ -293 cells (Invitrogen) were cultured in Dulbecco-modified Eagle’s medium containing 4.5 g/l glucose, 2 mM L-glutamine, 1 mM sodium pyruvate, 50 µg/ml uridine and 10% fetal bovine serum, at 37 °C in a humidified atmosphere with 5.5% CO₂ in air. No antibiotics were added for HEK293T cells, but transgenes in Flp-In™ T-Rex™ -293 cells were maintained under selection with hygromycin and blasticidin according to the manufacturer’s recommendations (see 4.9). All cell lines were routinely detached by pipetting alone, and passaged at 1:10 – 1:5 dilutions every 3-4 days depending on the cell line.

Transfections were carried out using TransFecting™-lipid reagent (Bio-Rad), following the manufacturer’s recommended procedure, with 12 µg of plasmid DNA and 40 µl of reagent, both in 1.5 ml of serum-free medium, per 10 ml plate. For siRNA induced gene knockdown, HEK293T cells were transfected using Lipofectamine 2000 (Invitrogen) and a final concentration of 20 nM of siRNA.

4.8 Inhibition of replication and transcription

To study replication fork behaviour under the arrest mtDNA synthesis by chain termination, cells were treated for various times in medium containing 100 µM dideoxycytidine (Sigma-Aldrich). To suppress both mtDNA replication and mitochondrial transcription, cells were treated with medium containing 50 ng/ml ethidium bromide (EtBr) for 72 h, after which cells were washed and replated in fresh medium, then cultured for a further 48 h. For the analysis of the effect of oxidative damage on mtDNA replication in cultured HEK293T cells, KBrO₃ was used. KBrO₃ is known to induce oxidative DNA damage in lab animals as well as cultured cells (Kawanishi & Murata 2006). From initial trials a concentration of 30 mM for 24 h was chosen.
4.9 Creation of doxycyclin-inducible Flp-In™ T-Rex™ -293 cell lines

The various cell lines used in the studies, enabling inducible expression of stably maintained transgene, were created by Sjoerd Wanrooij and Merja Jokela. The established system made it possible to grow polyclonal but purely transgenic human cells in quantities necessary for successful isolation of purified mitochondrial DNA.

The Flp-In™ T-REx™ 293 host cell-line (Invitrogen) is a variant of HEK293 containing a Flip recombination site at a transcriptionally active locus. The system exploits the bacterial tetracycline-resistance operon that is based on the binding of Tet repressor (encoded by a regulatory plasmid inserted into the host cell), which represses expression of the desired transgene. Tetracyclins, such as doxycyclin, bind to the Tet repressor, in turn inhibiting its binding to the target promoter and allowing transcription of the latter.

Cells were grown supplemented with 100 µg/ml Zeocin (Invivogen) and 15 µg/ml Blasticidin (Invivogen) to maintain Tet-repressor expression. Two days prior to transfection cells were split onto 10 cm plates and grown to about 80% confluence in medium lacking antibiotics. Cells were co-transfected with the appropriate pcDNA5/FrT/TO construct (0.4 µg) and pOG44 (Invitrogen; 3.6 µg), a plasmid encoding the Flp-recombinase necessary for targeted integration into the host genome. Six hours later, the medium was replaced with fresh medium, again without any antibiotics. 24 h after transfection the selective antibiotics hygromycin (150 µg/ml) (Invivogen) and blasticidin (150 µg/ml) were added. Selective medium was replaced every second day for the maintenance of the integrated transgene and Tet-repressor.

To induce expression, doxycyclin (Sigma) was added to the growth medium. The amount of doxycyclin used depends on the application (see original publications I-III), but saturated transgene expression was generally reached with 10 ng/ml. When transgene induction for longer than two days was required, the medium was replaced every second day.

4.10 Transgenic mice

The creation of Twinkle and TFAM mice and estimation of transgene expression and mtDNA copy number in mouse tissues is described in Tyynismaa et al. (2004), Tyynismaa et al. (2005) and Ikeuchi et al. (2005).
4.11 Autopsy samples

Autopsy samples were taken as part of the Tampere Coronary Study, approved by the Ethics Committee of Tampere University Hospital (DNO 1239/32/200/01) and the National Authority for Medicolegal Affairs. The samples were taken during standard forensic investigation for the cause of death and represent individuals from both sexes with no known diseases and of 3-72 years of age.

The study included cardiac muscle samples from 12 individuals of different ages plus, brain, skeletal muscle, kidney and liver samples from two individuals and one multiple tissue survey of heart, brain, skeletal muscle, kidney, liver and lymph node from a single individual. Heart muscle samples were taken from left ventricle and brain samples from unspecified parts of the cerebral cortex. For the isolation of human heart and brain mtDNA I was able to obtain heart, cortex and liver samples from an individual that had been deceased only six hours.

4.12 RNA and DNA extraction and quantification

4.12.1 DNA from cultured cells

For mtDNA copy number analysis, total cellular DNA was extracted using standard techniques. Briefly, the cells were harvested by centrifuging 1200 g<sub>max</sub> for 3 min, washed once with PBS and resuspended in 10 times the pellet volume of DNA extraction buffer (25 mM EDTA pH 8.0, 75 mM NaCl). One tenth volume of 10% SDS and 0.5 mg proteinase K were subsequently added and mixture swirled gently. After overnight incubation at 37 °C, two vol of 25:24:1 phenol:chlorophorm:isoamyl-alcohol pH 8.0 was added and the solution was mixed well by swirling. The mixture was then transferred into Eppendorf 15 ml Phase Lock Gel™ Heavy tubes and centrifuged at 5000 g<sub>max</sub> for 15 min. The aqueous phase was recovered and the extraction step repeated. DNA was precipitated by the addition of 0.2 vol 10 M ammonium acetate and 2 vol of ethanol. The mixture was incubated on ice for 10 min and DNA was spooled out using a glass rod, washed once with 70% ethanol, air dried gently and resuspended in 300-700 μl TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH8.0), depending on the pellet size.
4.12.2 DNA from tissues

Approximately 7 mm x 7 mm 7 mm block of frozen heart, brain, skeletal muscle, kidney tissue, liver and lymph node samples were cut into thin slices with a sterile blade and suspended in 2 ml DNA extraction buffer; one tenth vol of 10% SDS and 0.5 mg proteinase K were added. The crude homogenate was passed several times through a 5 ml pipette tip with a cut tip, to disperse the larger tissue fragments. The homogenate was incubated overnight with gentle swirling at 37 °C and the DNA extraction was continued as above.

4.12.3 mtDNA copy number estimation

Copy number was assessed independently by two different methods, to minimize possible artefacts. For Southern blotting, total DNA was cut by EcoRI and analyzed as described in Lehtinen et al. (2000), with quantitation by phosphorimaging (Storm 840 scanner and ImageQuant 5.1 software, Molecular Dynamics). Copy number was also estimated by real-time quantitative PCR with Taqman probes for mitochondrial cytochrome b and amyloid precursor protein (APP) as a single-copy nuclear DNA standard. Primers and probes were as follows (all 5’ to 3’):

APP Forward: TTTTTGTGTGCTCTCCAGGTCTCT
APP Reverse: TGGTCACTGGTTTGGC
APP Probe (FAM+TAMRA): CCCTGAACTGAGCATCAACAAATGTGGTAG
Cyt-b Forward: GCCTGCCTGATCCTCCAAAT
Cyt-b Reverse: AAGGTAGCGGATGATTCAGCC
Cyt-b Probe (TET+TAMRA): CACCAGACGCCTCAACCGCCTT

The runs were done using ABI Prism 7000 (Applied Biosystems) and analyzes performed using manufacturers´ software.

4.12.4 RNA extraction and quantification from cultured cells

RNA extraction from cells for RT-PCR, agarose or urea-polyacrylamide gel electrophoresis and Northern hybridization was performed with Trizol™ (Invitrogen) reagent using manufacturers’
recommendations. Probes were $^{32}$P end-labelled oligonucleotide as follows (5´to 3´): for ND3 mRNA, GTCACTCATAGGCCAGACTT, for 5S rRNA (loading control), GGTTGGTATGGCCGTAGAC, for tRNA$^{\text{Leu(UUR)}}$ and tRNA$^{\text{Tyr}}$ as described previously (Toompuu et al., 2002). Quantitation was performed by phosphorimaging as for mtDNA copy number.

4.13 Isolation of mitochondria

4.13.1 From cultured cells

To prepare mtDNA (mitochondrial nucleic acids) for analysis of replication intermediates, mitochondria were isolated from cells essentially as described by Spelbrink et al. (2000). Briefly, cells from 10-20 145/20 mm plates were collected by pipetting in PBS, centrifugation at 400 $g_{\text{max}}$ for 3 min at room temperature and transfer to ice. The cell pellet was resuspended by gentle pipetting in 2 volumes of ice-cold 0.1 × homogenization buffer (4 mM Tris-HCl, pH 7.8, 2.5 mM NaCl, 0.5 mM MgCl$_2$), kept on ice for 5 min and homogenized in a glass homogenizer with 20 strokes of a tight-fitting pestle. Disruption of the cells was monitored by microscopy. One-ninth volume of 10 × homogenization buffer was added and nuclei and cell debris were pelleted by sequential centrifugations at 1,200 $g_{\text{max}}$ for 3 min at 4 °C until no pellet was visible. Mitochondria from the post-nuclear supernatants were recovered by centrifugation at 16,000 $g_{\text{max}}$ for 10 min at 4 °C. The mitochondrial pellet was washed once in resuspension buffer (10 mM Tris-HCl pH 7.4, 0.32 M sucrose, 1 mM EDTA, 5 mM MgCl$_2$) and re-centrifuged at 16,000 $g_{\text{max}}$ for 10 min at 4 °C. For preparing mtDNA from highly purified mitochondria, the protocol by Yasukawa et al. (2005) was used, using cytochalasin (Sigma-Aldrich) induced cell breakage, followed by homogenization, differential centrifugations and 1/1.5 M sucrose gradient step.

4.13.2 From human and mouse tissues

Tissue samples were minced finely with sharp scissors in ice-cold HB buffer (225 mM mannitol, 75 mM sucrose, 10 mM Tris-HCl pH 7.6, 1 mM EDTA, 0.1% fat-free BSA) and washed three times with 10 ml ice-cold HB to remove blood, then resuspended in 5 ml HB. The crude tissue
homogenate was further homogenized with 8-10 strokes of a tight-fitting Teflon/glass pestle and transferred into a fresh 15 ml tube, filled up with HB and centrifuged to pellet nuclei and other debris for 5 min 1000 \( g_{\text{max}} \) at 4°C.

The supernatant was transferred to a fresh tube and centrifuged again at 9000 \( g_{\text{max}} \) for 10 min 4°C to pellet the mitochondria. The crude mitochondrial pellet was resuspended in 500 μl HB, layered over a 5 ml sucrose gradient (1/1.5 M) and centrifuged 45,000 \( g_{\text{max}} \) for one hour at 4°C. After the centrifugation the mitochondrial layer was transferred into a fresh 2 ml tube and one volume of HB was added. The purified mitochondria were pelleted by centrifugation 12,000 \( g_{\text{max}} \) for 5 min at 4°C.

4.14 Isolation of mtDNA

Mitochondrial pellets were resuspended in 1 ml of DNA extraction buffer (25 mM EDTA pH 8.0, 75 mM NaCl) and 0.2 vol of 10% SDS and 200 μg of Proteinase K (Fermentas) were added. The mixture was incubated for 30 min at room temperature and subsequently extracted with two volumes of phenol-chloroform-isooamylalcohol (25:24:1), repeating the extractions until no interphase was visible. The mitochondrial nucleic acids were precipitated by the addition of 0.2 vol 10 M ammonium acetate and 2 vol of ethanol, kept at -20°C for 2-15 h and pelleted by centrifugation 16,000 \( g_{\text{max}} \) for 20 min at 4°C. The pellet was washed once with 70% ethanol, air dried and resuspended in TE (10 mM Tris-HCl, 0.1 mM EDTA) buffer pH 8.0.

4.15 Sub-fractionation of mitochondria

For assaying the localization of recombinant TFAM the mitochondrial pellet, prepared as above, was resuspended in 2 vol lysis buffer (0.25 M sucrose, 20 mM Tris-HCl,pH 7.6, 2 mM EDTA, 7 mM β-mercaptoethanol). Mitochondria were lysed by adding 20% NP40 to a final concentration of 0.5% (v/v) with incubation on ice for 1 h. After centrifugation at 16,000 \( g_{\text{max}} \) for 10 min at 4°C the pellet and supernatant fractions were processed for SDS-PAGE.
4.16 SDS PAGE and Western blots

For SDS-PAGE 7.5-12% polyacrylamide gels were used under standard conditions (Laemmli 1970). Sample preparation, Western blotting and immunodetection were carried out as described previously (Spelbrink et al. 2000). Primary antibodies used were: mouse anti-Myc monoclonal 9E10 (Roche Molecular Biochemicals), 1:15.000 dilution of a 5 mg/ml stock, rabbit anti-human TFAM (kind gift of Dr R.J. Wiesner), 1:10.000 dilution, and mouse anti-α-actinin monoclonal AT6/172 (Upstate) 1:5000 of a 1 mg/ml stock. MnSOD-RecA was detected by immunoblotting using mouse monoclonal anti-RecA antibody, MSA-205 lot #B504434, by Stressgen (Victoria, BC, Canada). Signals were quantified using a ChemiDoc XRS chemiluminescence detection instrument and associated QuantityOne software.

4.17 Immunocytochemistry

Immunocytochemistry was performed as in Garrido et al. (2003). The cells were grown on coverglasses and fixed with fixation solution (3.7% formaldehyde, 5% sucrose in PBS) for 15 minutes. Fixation solution was removed, cells washed twice with PBS, lysed with 0.5% Triton X-100 in PBS for 15 min, subsequently washed twice with PBS and blocked for 45 minutes with 5% non-fat milk in PBS.

After removing the blocking solution, the samples were incubated for one hour with 15 μl 1:200 antibody of interest, washed three times with PBS and subsequently incubated with fluorophore-labeled secondary antibody for one hour in the dark. The samples were washed again three times with PBS, mounted on a microscope slide using Vectashield mounting medium with DAPI (Vector Laboratories) and analyzed by either fluorescent (Olympus BX61WI) or confocal microscope (Olympus IX70, with Wallac UltraVIEW confocal and Andor iXON detector). All steps were carried out at room temperature.

4.18 Enzymatic treatments of DNA

One to five μg of total DNA and 100 ng of mtDNA were used in each analysis. Samples were treated with the following modifying enzymes as recommended by the manufacturer: Lambda
exonuclease (New England Biolabs), T7 endonuclease I (New England Biolabs), Fpg (New England Biolabs), topoisomerase I (New England Biolabs) and topoisomerase IV (John Innes Enterprises). If successive enzyme treatments were used, DNA was ethanol precipitated between the steps.

4.19 One-dimensional agarose gel electrophoresis

All gels were run in 1xTBE. 0.4% agarose (Invitrogen 15510-027, Scotland, UK) was used for mtDNA topology analyses and 0.7% for copy number determination.

4.20 Southern blotting

Prior to blotting, the gels were immersed in 0.25M HCl for 20 min, washed briefly with water, immersed in 0.5M NaOH, 1.5M NaCl for 2 x 20 min to denature the DNA. After denaturation gels were neutralized with 1.0 M Tris-HCl pH 7.2, 2.0 M NaCl and blotted on to Hybond N or Hybond N+ nylon membrane (Amersham) 4-15 h by capillary action.

4.21 Two-dimensional neutral-neutral agarose gel electrophoresis (2DNAGE)

4.21.1 2DNAGE methodology

One µg of total mitochondrial nucleic acids or 10 µg (heart and brain) to 20 µg (skeletal muscle, kidney and cultured cells) of total DNA was used per analysis. Restriction digestions and other enzyme treatments were performed following manufacturers’ recommendations. BclI, EcoRI, HincII and PvuII, were supplied by Fermentas, Lithuania. AccI, DraI and BamHI by New England Biolabs (Finnzymes), Finland.

If subsequent treatment with another enzymes or nucleases was used, DNA was first recovered by ethanol precipitation and resuspended in the appropriate reaction buffer, before
treatment with the second enzyme. Reactions were stopped by the addition of an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1, pH 8.0) and immediately extracted.

The following agarose (Invitrogen 15510-027, Scotland, UK) percentages were used for optimal separation:

1 – 2 kb fragments
First dimension (1D): 0.6%, 1.2 V/cm, +4 °C, 20 h
Second dimension (2D): 1.4 %, 9 V/cm +4 °C, 6 h

2 – 5 kb fragments
1D: 0.4%, 1.8 V/cm, 4 °C, 20 h
2D: 1.0% 10 V/cm 4 °C, 6 h

16.6 kb fragments or uncut mtDNA
1D: 0.28%, 1.7 V/cm, room temperature, 24 h
2D: 0.58%, 1.8 V/cm, room temperature, 67h

The first dimension was run without ethidium bromide (EtBr) in TBE buffer. After the run the gel was stained with EtBr (300 ng/ml) in TBE. Individual lanes were cut out, rotated 90° and second dimension agarose containing 300 ng/ml EtBr, precooled to 55 °C, was cast around them. The second dimension was run at 4 °C with constant buffer recirculation. For analysis of high molecular weight fragments (e.g. 16.6 kb mtDNA linears), both dimensions were run at room temperature.

4.21.2 Interpretation of 2DNAGE results

The first dimension separates DNA fragments by size. Separation by size and shape is achieved in the second dimension with addition of EtBr and increasing voltage (Figure 4.4). For example, for a non-origin –containing restriction fragment through which a replication fork proceeds unidirectionally, the growing replication fork starts from the 1n position on the linear arc. As it grows in size, its complexity increases affecting the molecule’s migration in the second dimension. The replication fork is most complex when all the arms are equally long; forks with longer arms begin to be more linear again and start to migrate faster in the second dimension, touching finally
the linear arc at the 2n position. Recombination junctions always start at 2n and increase in complexity, depending on the position of the junction. Replication bubbles follow the same logic: growing in size and complexity and disappearing when the bubble exits the fragment and converts to Y-forms.

Figure 4.4. Interpretation of typical 2DNAGE images. The first dimension separates by size and second by size and shape (Friedman & Brewer 1985, Brewer & Fangman 1987). See text for explanation.

RITOLS intermediates (Yasukawa et al. 2006) seem more complicated to understand. Most restriction enzymes do not cut RNA-DNA hybrids. Therefore, when the lagging-strand is RNA-rich, replication intermediates migrate as higher molecular weight species with varying lengths of asymmetric Y-arms depending on the extent of the RNA incorporation. Essentially this results from failure of restriction enzymes to cut RNA-protected restriction sites (Figure 4.5.). Uncut circular intermediates and bubbles are also observed.
Figure 4.5. RITOLS intermediates on 2DNAGE. Restriction enzymes are not capable of cutting lagging-strand with RNA:DNA hybrid, resulting in molecules that are composites of several restriction fragments. RITOLS Y-arcs (or slow Y-arcs) seldomly start from the linear arc. Note the color-coding for different replication intermediates.

Gels capable of separating mtDNA-sized molecules follow the same logic as the previous (Figure 4.6.). Due to the migration properties of long DNA molecules, some patterns are distorted or more spread out than corresponding forms on regular 2DNAGE gels. These gels can help to visualize the replication intermediates over the whole genome, permitting global, if approximate mapping of replication origins and determining the directionality of replication.
Figure 4.6. Interpretation of long 2DNAGE patterns. (A) A diagrammatic presentation of mtDNA with two different cut sites i and ii. (B) Patterns resulting from both options. (C) Restriction fragments of replication intermediates in cases i and ii. In case i the restriction site is far away from the unidirectional origin, resulting in an almost complete bubble arc (b). When the restriction site is passed, the bubble converts to a double-Y (dy). In RITOLS molecules the RNA-DNA lagging-strand is not cleaved, resulting in molecules with a growing dsDNA tail attached to an uncut circle. These will migrate as a streak from the dimeric 2nc species (uncut circular molecules; either catenanes or termination intermediates). In case ii, most of the replication intermediates are tailed circles (t) starting from uncut monomeric circles (nc) and growing in size until they reach 2nc forming a distinct “eyebrow” arc. See also the Supplementary Figure 1 on original Article IV.

4.22 Radiolabeled probes and blot hybridization

For Southern hybridization, the following probes were created by *Pfu*-PCR, using cloned segments of human mtDNA as template, and subsequently sequenced to confirm their identity: O_H (np 35-611), ND2 (np 4480-4988), A8-6 (np 8460-9107), ND4 (np 11161-11640) and ND5 (np 12992-13670). The probes created from mouse mtDNA by *Pfu*-PCR and gel purified, were the fragments NCR (15,357-136) and CoxI (5383-5910). Pig and rabbit probes were created from total DNA samples by *Pfu*-PCR, gel purified and correspond to the mtDNA fragments ND1 (2821-3422) and
CoxI (2551-2980) respectively. Probes were labelled using Rediprime™ II random prime labelling kit (Amersham) and [α-32P]dCTP (Amersham; 3000 Ci/mmol).

4.23 Transmission electron microscopy (TEM)

For electron microscopy, mtDNA was isolated from gradient-purified mitochondria (see 4.14). Approximately 10 μg of total mitochondrial nucleic acids was treated with 50 u of RNase If (New England Biolabs) at 37 °C for 30 minutes in reaction buffer supplied by the manufacturer, recovered by phenol-chloroform extraction and ethanol precipitation, and redissolved in TEM grade TE-buffer (10 mM Tris-HCl, 0.1 mM ETDA, pH 7.6).

A separate aliquot was treated with 15 u of RNase H (Fermentas) at 37 °C for 30 min in the manufacturer’s recommended buffer, and recovered similarly. Aliquots (0.5-1.0 ng) of RNase-treated mtDNA were prepared and directly mounted on parlodium-coated grids with or without the presence of formamide, following the Kleinschmidt procedure as described by Griffith & Christiansen (1978) and Thresher & Griffith (1992). Imaging and image analysis were performed as described previously (Fouche et al. 2006).
5 RESULTS

In the following sub-chapters, I will briefly summarize the key findings reported in the original publications I-IV. The original articles addressed different aspects of human mtDNA maintenance. In an attempt to combine these findings into a holistic view of mammalian mtDNA replication and recombination, some unpublished supporting data is also presented. To place the findings in context, more background information on mtDNA maintenance is provided.

5.1 Structure of mammalian mtDNA replication and recombination intermediates

The exact mechanism of mtDNA replication in vertebrates has been under intense debate since discovery of strand-coupled mtDNA replication intermediates by 2DNAGE (Holt et al. 2000, Bogenhagen & Clayton 2003a, 2003b, Holt & Jacobs 2003). To complicate the issue further, no revised model of mtDNA replication has crystallized into published form. However, for clarity I will use the term “RITOLS-model” to separate the recent findings of strand-coupled replication from the earlier strand-displacement model (SDM). The SDM and RITOLS replication models differ especially with regard to the outcome of replication under altered physiological conditions. For example, replication stalling due to DNA damage or a defective replisome produces different end products depending on whether replication proceeds strand-symmetrically or not. These in turn, might affect the mechanistic outcomes of mtDNA deletion formation in pathological conditions.

In addition, little is known about tissue-specific features of mtDNA maintenance. There are several published comparative studies from different mouse tissues such as Piko & Matsumoto (1977), but they mainly focused on quantitative differences among different mtDNA forms. The first qualitative tissue-specific difference in mtDNA in humans was reported by Kajander et al. (2001). This study described unprecedented amounts of molecules with the properties of recombination junctions in heart mtDNA, but left open the issue of their origin and physiological significance. In order to understand tissue-specific differences in mtDNA maintenance, it is important to establish the general features of mtDNA replication and recombination. In pursuit of these goals I performed an extensive comparison of both topological as well as replicative forms of mtDNA in various human tissues and cultured cells using transmission electron microscopy (TEM), as well as one- and two-dimensional agarose gel electrophoresis (AGE).
5.1.1 Electron microscopy of mtDNA from cultured cells (unpublished data)

During recent decades vertebrate mtDNAs have been studied extensively using TEM and more recently 2DNAGE. However, most of the work done by electron microscopy in the past was conducted using CsCl gradient-purified mtDNA. CsCl gradient centrifugation separates mtDNA based on its buoyant density and was originally used to remove nuclear DNA (nDNA) contamination from the mitochondrial preparations (Tapper et al. 1983). This was considered especially important when visualizing mtDNA by TEM. However, it is possible that CsCl gradient purification introduces artifacts that result in the observed discrepancies in the replication intermediates compared to the 2DNAGE results, which were obtained analyzing mtDNA extracted from sucrose gradient-purified mitochondria.

To address this issue, I performed TEM on HEK293T cell mitochondrial DNA that was prepared exactly as for the 2DNAGE. The amount of nDNA contamination was very low in these preparations, and less than 10% of total molecular forms of DNA consisted of long heterogeneous linear fragments. However, in order to visualize any DNA forms at all the samples needed to be stripped from large amounts of contaminating single-stranded RNA by pre-treatment with RNase I. As expected, most of the mtDNA in HEK293T cells was found to exist as 16.6 kb monomeric circles. Additionally other molecular forms were seen (Table 5.1). These forms included supercoils, replication intermediates and complexes of two or more molecules. The largest mtDNA forms in cultured cells represent catenanes of 4 to 5 circular molecules. Some examples of these molecules can be seen in Figure 5.2, 5.3 and in the Supplementary Figure 4C of original communication IV.

Table 5.1. The quantities of different molecular forms of mtDNA from human heart and HEK293T cells as counted by TEM. N = number of molecular forms counted, 1n = circular monomers, 2n = circular dimers, pairs = two linked molecules, complexes = more than three linked molecules. RIs = replication intermediates.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>1n</th>
<th>2n</th>
<th>pairs</th>
<th>complexes</th>
<th>RIs</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEK293</td>
<td>1002</td>
<td>85,13 %</td>
<td>0,20 %</td>
<td>7,29 %</td>
<td>2,20 %</td>
<td>5,19 %</td>
</tr>
<tr>
<td>Heart</td>
<td>500</td>
<td>35,60 %</td>
<td>33,40 %</td>
<td>17,60 %</td>
<td>12,20 %</td>
<td>1,20 %</td>
</tr>
</tbody>
</table>
Figure 5.1. Paired mtDNA molecules from HEK293T cells. The region of the paired sequence is short and smooth. However, it is impossible to say whether the molecules are recombining, catenated or perhaps recently replicated and bound by the replication terminus. The upper scale bar represents 1 kb.
The species scored as replicating molecules represented fully double-stranded theta-intermediates of various sizes (Figure 5.3). In some cases short segments of single-strandedness were seen close to one replication fork. Exactly as predicted by the RITOLS model of mtDNA replication, the double-stranded theta-intermediates were transformed into partially single-stranded SDM-like intermediates by RNase H treatment (Figure 5.4). The direct mounting method used for TEM causes ssDNA to coil on itself. This can be avoided by spreading the DNA samples in the presence of formamide that prevents the spontaneous formation of intramolecular hydrogen bonds (Figures 5.3C-D).
Figure 5.3. Theta-type mtDNA replication intermediates from HEK293T cells. Scale bar 1 kb. Short, fully double-stranded replication bubbles seen in (A) and (B) cannot be attributed to the standard strand-displacement model of mtDNA replication. (C) Replication has proceeded half-way through the mtDNA. Also an unrelated linear molecule can be seen entering the field from the bottom right corner. (D) Almost fully replicated mtDNA. Compare to Figure 5.1B, where the replicating molecule is linked to another as a part of a catenane.
Figure 5.4. RNase H treatment converts double-stranded theta-molecules into partially single-stranded intermediates of SDM-type. (A) Untreated mtDNA in formamide spread showing fully double-stranded replication bubble. (B) A closeup of the bubble. (C) Formamide spread of RNase H treated mtDNA showing extensive single-strandedness. Insert showing a closeup of the partially single-stranded fork structure.
5.1.2 Electron microscopy of mtDNA from human heart (IV)

To investigate the peculiar molecular forms reported by Kajander et al. (2001), I also performed TEM analysis of sucrose gradient-purified mtDNA from human heart and cerebral cortex. As seen in Table 5.1, heart has equal amounts of dimeric and monomeric circular molecules whereas only a low frequency of dimers was detected in cultured cells (Supplementary Figure 4a in Article IV). Moreover, significantly more paired and other complex molecules were seen in heart mtDNA preparations. Many of the observed pairs quite likely represent recombining molecules as characterized by Kajander et al. (2001). Examples of these molecules and their connections can be seen in Figure 3c of original communication IV. The most striking complex mtDNA forms in heart form complex network-like aggregates, of which the largest have several connection foci (Figure 5.5). These molecules have defined boundaries and they consist of dozens of circular units (Figure 3a of Article IV). A closer analysis of the nexuses of these networks reveals that the junctions are not more electron dense than the rest of the molecule, indicating that they are possibly not bound by any residual proteins (Figure 3b in IV).

In contrast to HEK293T mtDNA, very few free replicative forms were seen in TEM of heart mtDNA. The few molecules scored as replication intermediates were aberrant circular molecules with double-stranded tails of generally less than one full genome length (Supplementary Figure 4b in IV). However, a careful examination of the large aggregates of heart mtDNA revealed apparent forked molecules inside these structures (Figure 5.6). The partial digestion of these complexes by BamHI prior to the spreading for the EM helped to resolve these junctional molecules into more interpretable form (Figure 5 in IV). These molecules seem to fall into three classes: four-way junctions, usually with each pair of branches of equal length (Fig. 5a-e in IV), three-way junctions representing replication forks (Fig. 5f, g in IV), and more complex junctions of several branches. Some of the four-way junctions appeared to have two closely spaced junctions joined by a short bridge of intertwined duplex strand, equivalent to the classical double Holliday junction intermediate of the double-strand break-repair model of homologous recombination (Szostak et al. 1983). Complex junctions such as seen in Figure 5h-k in IV were in persistent tangles as well as in simpler molecules (Fig. 5k in IV). As a summary, TEM provided support for the RITOLS model in cultured cells, whilst revealing abundant junctional complexes in human heart mtDNA.
Figure 5.5. A complex DNA structure from human heart.

Figure 5.6. Examples of some junctional molecules found inside the complex forms of heart mtDNA. Three-way forks possibly representing replication intermediates are marked with an arrow.
5.1.3 Comparative analysis of mtDNA from various human tissues by 1D- and 2D-AGE (IV, unpublished data)

For comparative analysis, 2D-NAGE of the mtDNA replication intermediates from various human tissues was performed. In order to get a whole mitochondrial genome view, a single-cleavage restriction enzyme approach was applied using gel conditions capable of resolving 16.6 kb fragments. I was able to detect intact mitochondrial theta–like replication intermediates in total DNA preparations of post mortem brain, skeletal muscle (iliopsoas) and kidney that were comparable in quality to the ones from freshly prepared total DNA sample from HEK293T cells (Figure 5.7). In the BamHI digest typical “eyebrow”-forms, indicating RITOLS replication, were also detected (see Supplementary Figure 1 of IV). In the autopsy samples from some tissues, such as liver and lymph node, the mtDNA was generally too degraded to detect any replication intermediates (data not shown).

In heart there were no replication intermediates corresponding to the expected products of theta-replication on single-cleavage restriction enzyme gels, the major replication intermediates seen migrated on a Y-like arc, encompassing the whole mtDNA molecule (Figure 5.7A:i, 5.7B:i). The products of *Pvu*II and *Bam*HI digests were indistinguishable. The majority of other non-linear molecules appeared to be recombination intermediates migrating on a standard X-arc, as already reported in Kajander et al. (2001) for smaller fragments of the genome. Some complex X-like forms of more than twice the genome size and probably corresponding to three or more recombining molecules were also detected (denoted as dx in Figure 5.7). Similar X forms as well as a regular X-arc were also present together with theta-forms in brain samples. Identical results were obtained with sucrose gradient-purified mtDNA from human cardiac muscle and cortex obtained as soon as possible after death (6h), which were later used in TEM (for details see Supplementary Figure 3 in IV).
Figure 5.7. 2DNAGE of total DNA shows theta-type mitochondrial replication intermediates (RIs) in human brain, skeletal muscle, kidney and cultured cells, but not in heart. See chapter 4.20.2 for interpretation. (A) PvuII (nt 2564) digested DNA from various tissues and HEK293T cells. All other panels except heart (A-i) show theta-type replication intermediates. (B) Identical samples digested with BamHI (nt 14259). Note the similarity of (A-i) and (B-i). In heart most of the non-linear forms migrate on X- and double-Y-arcs. Some more than 2n genome length Xs are present (dx). In BamHI digests (bottom row) the dominant RIs are double-Ys (dy) and tailed circles (t) as depicted in Figure 4.6B and 4.6C. Again, the heart RIs migrate as a Y-like arc. Brain mtDNA 2DNAGE patterns show both theta- and heart-like replication intermediates. Note the low amount of any replicative forms detected in skeletal muscle mtDNA.

2DNAGE of smaller fragments revealed further tissue differences in replication intermediates (Figure 2 in IV). As expected, 2DNAGE of purified mtDNA from cultured cells (Figure 2b:vi in IV) shows typical slow-moving-Y (SMY)-like arcs indicative of RITOLS intermediates. These intermediates were preserved in total DNA preparations from the same cells (Figure 2b:v in IV) together with some partially degraded RNA-rich intermediates (mSMY). Traces of these same intermediates could also be seen in total DNA samples obtained from autopsied kidney (Figure 2b:iv in IV). Heart and brain samples however did not show any recognizable RITOLS intermediates, instead the Y-arc was well defined and X-forms were abundant, as already seen in the PvuII and BamHI digests. The high MW X-like structures (dx) were also present in these preparations. Furthermore, no initiation-arcs were detected from O2H-containing or any other fragments of human heart mtDNA.

The X-forms were seen in all 12 healthy human individuals checked so far and seemed to be age-independent although no precise quantifications were done (Supplementary Figure 3 of IV). Interestingly, the abundant X-forms in heart seem to be a human-specific feature, since no
X-forms were detected from cardiac muscle total DNA samples from mouse, rabbit or pig (Supplementary Figure 3 in IV).

On one-dimensional Southern blots of undigested human heart mtDNA, most material migrated in two main positions that are denoted as bands s (for slow) and f (for fast) (Figure 5.8A). The fastest migrating band represents one genome length, topoisomerase I-sensitive supercoils (see also Figure 5.12). The other easily recognizable form is the one genome length (16.6 kb) linear.

A better separation was achieved by applying similar 2DNAGE conditions as for the single-cleavage analysis (Figure 5.8B). This method revealed that band f is actually a heterogeneous mixture of different molecular forms, including 33 kb linear (2n), one genome-length open circles (oc) and topoisomers of larger, most probably two genome length circles. As seen in Figure 5.8B:i the one genome-length supercoils (sc) migrated more slowly in the second dimension, with the same mobility as open circles. Similarly the bulk of band f migrated more slowly in the second dimension, and ran in the same position as the bulk of band s. When topoisomerase I or IV was applied most of the band f disappeared together with the supercoils (Figure 5.8A, B:ii). It should be noted that besides decatenation, topoisomerase IV also relaxes supercoiled molecules.

Figure 5.8. Human cardiac muscle mtDNA is organized in various topological forms consisting of 16.6 and 33 kb molecules. (A) A Southern blot of total DNA from human heart, probed with mitochondrial probes, showing different enzymatic treatments. u – untreated, ti – topoisomerase I, tiv – topoisomerase IV (a decatenase), t7 – T7 endonuclease I, exo – lambda phage exonuclease III. Most of the material in the untreated sample is concentrated in diffuse bands denoted t and f. Band f has a linear mobility of around 30-35 kb, while t migrates slower than a 48.5 kb lambda phage genome fragment. (B) Enzymatic digests as well as 2DNAGE reveal that band f actually consists of multiple forms, which include 33 kb linear molecules, one genome length circles, junctional molecules and topoisomers. (i) Untreated, (ii) topoisomerase IV, (iii) topoisomerase IV and T7 endonuclease I, (iv) T7 endonuclease I. 2n - two genome length, 1n - one genome length, h – heterogeneous large molecules, oc – open circles, sc – supercoils, x - junctional molecules.
Topoisomerase IV treatment revealed an additional species, x, migrating around 33 kb in the first dimension. This species quite likely represents a circle recombining with a linear molecule, since circular molecules would be slower in the first dimension and fully linear partners would form an X-arc. Circular recombining molecules as well as 33 kb circles would likely migrate in the same position as catenated 16.6 kb circles (see Figure 5.12.). In the gels presented here, their presence is masked by band s, which most likely contains all these forms. This notion is supported by the fact that T7 endonuclease I treatment resulted in an increase in 16.6 kb linear and when applied together with topoisomerase IV left a defined residual band (Figure 5.8B:iii-iv). This band has the migration properties of a circle and is resistant to decatenation by topoisomerase IV as well as cruciform cutting endonuclease, thus it is safe to assume that it represents 33 kb circles (2n oc).

Topoisomerase IV and T7 endonuclease resolved the heterogeneous material above band s (denoted as h) producing 33 and 16.6 kb linear, indicating that this material includes catenanes and recombining molecules of one and two genome lengths. Phosphoimager analysis of these residual circles and linear showed that around two-thirds of the total signal corresponds to 33 kb sized molecules. This is in agreement with the EM data, because dimeric genomes logically give twice the signal.

Mitochondrial DNA from other tissues, including brain, differed from heart in appearance on Southern blots of one-dimensional gels of uncut DNA (Figure 5.9). Most of the material in all of these tissues consisted of 16.6 kb open circles. Some low mobility bands possibly corresponding to band s in heart were also seen in samples from these tissues, but were mainly topoisomerase IV-sensitive, two molecule catenanes as seen in cultured cells. High molecular weight forms were generally absent and the amount of supercoiled genomes was relatively high in the skeletal muscle sample. As mentioned before, DNA from autopsy-derived lymph node and liver samples was most degraded (data not shown).
5.2 Effects of mtDNA-binding proteins on mtDNA replication and recombination

In order to learn about the molecular mechanisms involved in the modulation of recombination and replication, several candidate proteins were chosen for detailed study. Unfortunately, no mammalian mtDNA recombinase has been described and rather few proteins involved in mammalian mtDNA maintenance are known. The best candidate for a protein involved in modulation of recombination is the multi-functional DNA binding protein TFAM, whose homologue in yeast is known to stabilize Holliday junctions (Zelenaya-Troitskaya et al. 1998).

Because molecular recombination is catalyzed by specialized recombinase enzymes, some possibilities were considered. The obvious candidate protein is the Twinkle mtDNA-helicase, which belongs to the same superfamily as DnaB–helicases (Spelbrink et al. 2001). Ultimately this makes Twinkle distantly related to the RecA/Rad51/DCM1 family of ATP-dependent recombinases (Leipe et al. 2000). Twinkle is also the only human protein that comes up with BLAST-search using the RecA polypeptide sequence. Therefore, the effects of manipulating the expression of Twinkle protein in cultured human cells were interesting.

The possibility of inducing recombination artificially in vivo in HEK293T cells by allotopic expression of mitochondrially-targeted bacterial RecA was also tested by transiently
transfecting the cells with recombinant *E. coli* RecA containing the MnSOD mitochondrial-localization presequence. The construct was previously reported to be mitochondrially-localized and enzymatically-active in cultured cells (Paul et al. 2001).

I also had the opportunity to test the effects of elevated levels of TFAM and Twinkle at the whole organism level, because transgenic mice overexpressing these proteins existed (Tyynismaa et al. 2004, Ikeuchi et al. 2005) and were made available to me. Both of these mouse lines have elevated mtDNA copy number in many tissues and are also healthy as far as known. Using mouse models in the study was important for addressing the physiological significance of the highly tissue-specific recombination phenomenon seen in humans.

Twinkle and TFAM are known to be involved in mtDNA replication and their manipulation was also expected to yield valuable insight on the general replication mechanisms operating inside mitochondria. As a partner for Twinkle, the involvement of PolG in the modulation of mtDNA replication was studied. Besides overexpressing these proteins, it was considered useful to look at the effects of different catalytically-impaired artificial mutants on mtDNA replication in cultured human cells.

Additionally, some unexpected but extremely important input for the project came from Anne Hyvärinen’s work on another mtDNA binding protein, mTERF. Being originally characterized as a mitochondrial transcription termination factor, mTERF is inferred a novel role in the regulation of mtDNA replication based on 2DNAGE analysis, which I was able to conduct in collaboration with her (III).

### 5.2.1 Manipulation of TFAM expression in cultured cells (I)

In order to manipulate TFAM expression in cultured cells I engineered several variant expression constructs in suitable shuttle vectors (see Materials and Methods). The initial trials were done using transient transfections of the pcDNA3.1 constructs of the different TFAM variants. Only constructs encoding full-length TFAM commencing at the first AUG were chosen for the creation of inducible TFAM-stop and TFAM-myc-his cell lines used in the subsequent experiments. Both AUG variant constructs gave good expression and products were mitochondrially localized (data not shown). No significant effects on transcription or replication were observed with overexpression of the TFAM splice variant constructs (data not shown). Depletion of TFAM protein in HEK293T cells was achieved using an RNAi approach (Supplementary Figure 3d of I).
Both sustained overexpression as well as post-transcriptional silencing of TFAM gene resulted in a gradual copy number decrease (see I, Figure 1). The overexpression of TFAM also depleted the level of an mtDNA-encoded mRNA ND3, whereas TFAM RNAi resulted, surprisingly, in an overall increase of transcription per mtDNA molecule. The effects on mitochondrial RNA as well as DNA levels were strongest in TFAM-myc-his overexpressing cells (Figure 1 in original Article I).

Because of the effects on mtDNA copy number, as well as it being a candidate protein for the modulation of recombination, the effects of TFAM overexpression and RNAi were studied using 2DNAGE to reveal changes in mtDNA replication and recombination intermediates. Cells expressing wild-type TFAM-stop were chosen for a more detailed analysis, because this overexpression had a less severe effect on mtDNA copy number, plus the possibility that the myc-His tag might cause secondary effects could not be ruled out. However, both constructs induced identical changes in mtDNA replication intermediates (Supplementary Figure 5b of I).

TFAM overexpression resulted in a marked decrease in RITOLS intermediates with concomitant enhancement of more conventional COSCOFA-type dsDNA replication intermediates around the genome. Moreover, a specific accumulation of replication intermediates in the rDNA region after a naturally occurring pause site was observed (Figure 5.10).

Figure 5.10. TFAM overexpression-induced changes in the mtDNA replication intermediates in the 4.8 kb AccI ND2-containing fragment of human mtDNA. Top row: mtDNA from uninduced 293T-rex cells. Bottom row: Same cells induced with doxycyclin to express TFAM for 48 h. Note the heterogeneous (h) cloud of RNase I sensitive material in control cells. The slow moving Y arc (SMY) results from RITOLS replication. RITOLS replication intermediates are sensitive to both RNase H and S1 treatment. Notice the significant increase in the S1 resistant standard Y-arc signal in TFAM overexpressing cells. S1 nuclease treatment also reveals a specific accumulation of dsDNA replication intermediates in the rDNA region after the replication pause site (p).
The depletion of TFAM by RNAi resulted did not result in major changes in the representation of different mtDNA RIs. The only observed difference was a concomitant increase in heterogeneous RNA intermediates coinciding with the increased transcription per template (Figure 5.11).

![Figure 5.11](image_url)

**Figure 5.11.** TFAM RNAi increases the heterogeneous RNA intermediates seen on 2DNAGE. Nomenclature as in Figure 5.10.

![Figure 5.12](image_url)

**Figure 5.12.** Southern blots of undigested mtDNA from control and cells induced to express TFAM for 48h. Note the marked increase in topoisomerase I (tI)-sensitive supercoiled forms and the concomitant reduction in the topoisomerase IV (tIV)-sensitive catenated forms in the induced cells. T7 endonuclease I (gp3) resolves junctional molecules. Note the similar residual quantities of open circles and genome length linears in the tIV + gp3 treated samples. (B) TFAM overexpression depletes 7S DNA. Heating the sample briefly in 95°C melts out the 7S DNA strand. Ind = cells induced for TFAM expression, unind = uninduced controls. Ind x 3 = three times longer exposure of the autoradiograph from induced cells.
The mitochondrial genome topology in TFAM overexpressing cells was analyzed in detail using Southern blots of undigested mtDNA treated with various DNA-modifying enzymes. An increase in relative amounts of monomeric supercoils, together with depletion of catenated forms, was observed upon the induction of TFAM-overexpression (Figure 5.12). After topoisomerase IV treatment two additional high molecular weight bands sensitive to the T7 endonuclease I (gp3) were observed in the TFAM overexpressing cells.

5.2.2 Manipulation of Twinkle DNA helicase and PolG expression in cultured cells (II and unpublished data)

While overexpression of wildtype Twinkle or PolG did not produce any significant effects on either mtDNA copynumber or the patterns of RIs (III), catalytically defective mutants of the proteins resulted in distinctive rapid copy number depletion following induction. In severe catalytic mutants of Twinkle, all RITOLS replication intermediates were replaced by enhanced COSCOFA-type dsDNA replication intermediates (Figure 5.13, data by Steffi Goffart). The defective PolG mutants cause a different phenotype, retaining the RITOLS intermediates (see Article II for details). When PolG was inhibited by ddC in cells expressing the severe Twinkle mutant, the RITOLS intermediates were observed to reappear (Figure 5.14, data by Steffi Goffart).

For comparison with the previous data from HEK293T cells (Chapter 5.1), 100 mtDNA molecules from cells expressing the stalling Twinkle mutant K421A were examined by TEM. Nineteen replicating molecules, all fully duplex intermediates, were observed (Figure 5.15). One replication intermediate had an intact bubble-structure that appeared to have been strand-invaded by a linear molecule (Figure 5.15C-D). The invasion site located exactly in the middle of the replication bubble. Other circular molecules with linear tails, possibly representing broken theta-forms, were also observed. These intermediates might arise either from strand breakage during isolation or by in vivo regression of a stalled replication fork.
Figure 5.13. Examples of 2DNAGE of *Hinc*II digested mtDNA from different cell lines expressing Twinkle variants. Blots are probed for the O\textsubscript{H} containing fragment (nts 13637-1008) that includes the RITOLS replication initiation-bubble (rb). The induction of wild type Twinkle expression does not cause an obvious phenotype apart from the depletion of Holliday junction-like (HJL) intermediates, which represent replication termination at O\textsubscript{H}. The catalytically most impaired mutants like K421A and G575D show the most severe stalling phenotype with the depletion or complete loss of RITOLS intermediates (circled). The dsDNA COSCOFA bubble arc (b) and the conventional Y-arc are significantly enhanced. Note that the dsDNA bubble arc extends over the whole fragment in the catalytically mutants. For further details of the mutants and their relative amounts in cells under the used induction conditions see Article II.

Figure 5.14. Inhibition of PolG restores RITOLS intermediates in mutants with severe stalling phenotype.
Expression of the Twinkle K421A variant resulted in the accumulation of full length, expanded COSCOFA-type initiation bubble arcs in the O_H containing HinlII fragment ([nts 13637-1008] Figure 5.13). One interpretation is that there is random initiation across the fragment and since the Y-arc is still incomplete, most of the replication never reaches the end of the molecule. This observation was supported by TEM of BamHI-digested (nt 14259) mtDNA from the same cells. In unidirectional replication, the shorter fork of the resulting double-Y molecule should normally map at O_H. This was the case in most of the observed molecules (mean position nt 147, N=4 [Figure 5.22A]), but in some cases the fork mapped in other positions (eg. 1070 or 15900 [Figure 5.22B]).
Figure 5.22. (A) A double-Y molecule resulting from BamHI digest of mtDNA from Twinkle K421A expressing cells. Arms of the shorter fork map at O_H (nt 191). (B) Another double-Y molecule from the same cells. Short arms map approximately at nt 15911. Scale 1 kb.
5.2.3 Expression of mitochondrially-targeted RecA (unpublished data)

Because there is no known human mitochondrial RecA homologue that would be responsible for the excessive recombination seen in heart mitochondria, I decided to test whether similar amounts of recombination could be induced in cultured cells by expression of mitochondrially-targeted bacterial RecA. No increase in X-arc intensity was observed (Figure 5.23). Instead, an increase in the Y-arc intensity was evident, suggesting either increased or slowed replication.

![Figure 5.23](image)

Figure 5.23. (A) Allotopic expression of bacterial RecA enzyme in human mitochondria does not induce Holliday-junction formation. Instead an increase in Y-arc intensity can be observed. Dral + S1 digest of total DNA, probed for the ND5 containing fragment (nts 12272-16011). (B) Detection of mnSOD-RecA on Western blot using anti-RecA antibody. Lanes represent mock-transfected and RecA-transfected cells respectively.

5.2.4 Manipulation of mTERF expression in cultured cells (III)

My colleague Anne Hyvärinen detected and characterized numerous novel binding sites for mTERF during her own PhD thesis research project using the combination of SELEX and EMSA (Figure 2 in Article III). Because the apparent occupancy of these sites did not produce any transcriptional effects, as expected from the characterized transcription termination function of the protein, we decided to investigate whether the binding sites had any physiological role in mtDNA replication. 2DNAGE analysis revealed that mTERF overexpressing cells showed enhancement of several naturally occurring replication pause sites that roughly corresponded with the newly identified mTERF binding sites. In order to obtain better resolution of these pause sites, the samples were treated with S1 nuclease, which generally results in better defined conventional replication patterns of mtDNA on 2DNAGE, although the RITOLS-type RIs are preferentially degraded compared to the COSCOFA-type.
Many of the new sites clustered in the region of mtDNA harbouring the well characterized mTERF binding site in the tRNA\textsuperscript{Leu(UUR)} gene. A comparison of different cell lines, as well as human tissues, revealed a number of naturally occurring pause sites in this region, spanning from O\textsubscript{L} to the beginning of the rDNA segment (Figure 5.24). Additional naturally occurring pause sites and mTERF binding sites were detected also in the NCR region. One of the fainter replication pauses, located close to the tRNA\textsuperscript{Leu(UUR)} gene, was greatly enhanced in mTERF overexpressing cells (Figure 5.24). A more abundant, naturally-occurring pause site in the vicinity of IQM tRNA cluster was also significantly enhanced in these cells. Lagging-strand initiation at these sites was later mapped by LM-PCR to locate at nucleotides 3234 in the tRNA\textsuperscript{Leu(UUR)} region and 4320, 4340 and 4476 in the IQM – ND2 regions (Figure 6 in Article III). The most significant effect was seen on the H-strand, where the accumulation of the 5’ ends mapped precisely at O\textsubscript{H}. Similarly, an accumulation of replication intermediates in the O\textsubscript{H} region was observed on 2DNAGE of mtDNA from mTERF-overexpressing cells (Figure 4C in Article III). Depletion of mTERF by RNAi reduced the nt 3234 pause site more than other mTERF dependent pause sites (Figure 5.25). Interestingly, RNAi treatment also depleted the X-forms in the corresponding region in the AccI-PvuII (2564-6287) fragment.

Figure 5.24. Naturally occurring replication pause sites in various human cell lines and tissues visualized with 2DNAGE. (A) PvuII + AccI digest was used in order to resolve the densely spaced pause sites near the distal end of the rDNA region (2564-6287). 2DNAGE panels showing (B) various cell lines and (C) human tissues. Middle panel: S1 treatment to visualize pause sites. (D) An interpretation of the pause sites seen on 2DNAGE. (E) A schematic view of the characterized mTERF binding sites (see III) and the pause sites (denoted a-d) on mtDNA.
Figure 5.25. The rDNA region of human mtDNA shows specific responses to the modulation of mTERF amount in HEK293 derivating cell-line. Overexpression of mTERF (panel ii) enhances the pause site seen in the tRNA$\text{Leu(UUR)}$ gene (a) as well as the IQM pauses (b). Depletion of mTERF (panel iii) results in the decrease of the tRNA$\text{Leu(UUR)}$-pause site, but not of the IQM pause. Notice the difference in the X-form intensity. Equal exposure was ensured by quantification of the radioactive signal by phosphoimager. 4x – four times as long exposure as 1x confirming the depletion of the tRNA$\text{Leu(UUR)}$-pause site as well as the X-forms.

5.2.5 Overexpression of Twinkle and TFAM in vivo in transgenic mice (IV)

In order to determine whether TFAm or Twinkle overexpression induced similar effects on mtDNA replication in vivo to those seen in cultured cells, I analyzed mtDNA from TFAM and Twinkle transgenic mice by 2DNAGE. The mice had been previously shown to have elevated mtDNA copy number in various tissues, such as skeletal muscle and heart. In non-transgenic laboratory mice as well as in the non-transgene expressing littermates of the transgenic mice, mtDNA replication intermediates in heart appeared to be similar to those in liver, although of much lower abundance (Figure 5.26A). The quantities of RIs were comparable to those seen in human skeletal muscle samples and might indicate low levels of mtDNA replication in these tissues. The younger mice, whose organs are still growing, had relatively more replication intermediates than the 10 month-old mice, but no qualitative differences were observed (Figure 6d in Article IV). RITOLS replication-derived slow moving Y-arcs (SMY) were more prominent in liver than in heart (Figure 5.11A:ii).
Figure 5.26. A comparison of 2DNAGE patterns from liver (i,ii) and heart (iii,iv) of control and Twinkle or TFAM overexpressing mice. (i) and (iii) MluI (a single cutter at nt 1771), (ii) and (iv) DraI (probed for the CoxI containing fragment, nts 5275-9820) digest respectively. SMY = slow moving Y-arc of RITOLS replication.

In transgenic mice overexpressing mouse Twinkle or human TFAM protein, a strong increase in the X-arc intensity was observed on 2DNAGE of purified cardiac muscle mtDNA (Figure 5.26B, C). In addition the high molecular weight species denoted dX was observed in these samples (Figure 5.26B:iii, c:iii, iv). In Twinkle-overexpressing mouse heart there was also a slight increase in other replication intermediates when compared to the age matched non-expressing littermates. Concomitant with this increase, slow moving Y-arcs indicative of RITOLS type replication could be seen in the CoxI-containing DraI fragment. In human TFAM-expressing mice the effect was much more pronounced; the X-arc was more intense, the dX-streak appeared (Figure 5.26C:iv), no RITOLS intermediates were detected and the bubble arcs, indicatives of theta-type replication, disappeared. A similar strong effect was observed in Twinkle-overexpressing mouse brain (Figure 6e-f in Article IV). Brain was the only tissue that had relatively high amount of X-form molecules in mtDNA.
There were no obvious differences in the mtDNA organization between control mice liver and heart (Figure 5.27.). Based on Southern blots of undigested mtDNA, Twinkle-overexpressing mice accumulated heterogeneous high molecular weight forms, which were sensitive to both topoisomerase IV and T7 endonuclease I (Figure 7 in Article IV). TFAM overexpression resulted in a similar outcome (Figure 7, IV).

Figure 5.27. Comparison of liver and heart mtDNA forms in control littermate (wt) and Twinkle (tw) overexpressing mice. The majority of molecules represents genome-size open circular (oc) forms, catenated molecules as well as supercoils (sc) can be seen. Twinkle overexpression results in marked increase in high MW mtDNA forms in heart.

5.3 Physiological variables affecting mtDNA replication and topology

Different tissues have different energy requirements. These energy requirements can affect mtDNA copy number as well as the relative amounts of exposure to ROS originating from the respiratory complexes. Both of these variables might impose requirements for mtDNA maintenance. Furthermore, the molecular mechanisms required for the physiologically-meaningful mtDNA maintenance might be affected in various pathological conditions. In order to gather supporting information of the possible involvement of these variables in modulating of mtDNA replication and/or organization, I measured copy number in different human tissues. I also analyzed the effects of deliberately provoking oxidative stress to mtDNA and examined post mortem heart mtDNA replication and recombination intermediates from a person diagnosed with a severe ischemic heart disease. All this data, whilst provocative is very preliminary and more experiments are needed. I would like to present these results only as perspectives for future work.
5.3.1 mtDNA copy number in different tissues (IV)

The mtDNA content of the cell depends on its energy requirements (Moraes 2001). Because mtDNA must be packaged into nucleoids, the copy number must be related either to the size or number of nucleoids. However, nothing is known about how mtDNA is organized in nucleoids. It is plausible that the high MW mtDNA complexes found in human heart (Chapter 5.1) represent the nucleoid organization of mtDNA in these cells. With more mtDNA molecules per nucleoid, the opportunity for inter-molecular recombination would increase, providing an explanation for the high levels of recombination seen in human heart as well as possible resistance to ischemic stress.

In order to see whether mtDNA organization correlated with copy number in different tissues, I measured mtDNA copy number using real time quantitative PCR (qPCR) by comparing the relative quantity of mtDNA to that of a typical diploid nuclear gene. Heart had the highest copy number of around 11,000 and lymph node the lowest of around 500 mtDNA copies per cell. Skeletal muscle was estimated to have 2300 copies and all other tissues around 5000 copies of mtDNA per cell (Figure 5.28).

![mtDNA copy number in different tissues](image)

Figure 5.28. mtDNA copy number per diploid nuclear gene in various human tissues from one individual as measured by qPCR.
Heart is a continuously-working tissue and it is expected that the highly-active OXPHOS system also produces comparatively high amounts of ROS. ROS are a known cause of DNA damage, including double-strand breaks (DSBs [David et al. 2007]). DSBs can result in genomic rearrangements or possibly in a complete loss of a chromosome. As recombination-mediated DNA repair is the most efficient means of repairing DSBs, a highly active molecular recombination system in heart might be a physiologically-regulated response to the fact that heart mtDNA is constantly bombarded by ROS. To evaluate whether recombination or other effects on mtDNA replication could be produced in cell culture by oxidative damage, HEK293T cells were exposed to various amounts of KBrO₃, which is known to be a potent oxidizer of DNA in vivo. The oxidative damage in mtDNA was confirmed by measuring the relative amount of nicks per mtDNA molecule using alkaline AGE after treating the sample with E. coli Fpg, which is an 8-oxo-deoxyguanosine (8-OHdG) DNA glycosylase. Eight hours of exposure to 30 mM KBrO₃ induced changes in the pattern of mtDNA RIs as seen by 2DNAGE, and 24h exposure resulted in the accumulation of replication intermediates indicating severe stalling of replication (Figure 5.29). However, X-arcs were not influenced by this treatment.

Figure 5.29. Oxidative damage by potassium bromate on HEK293T cells after 24h exposure. BclI digest, ND2 probe. Note increase in the standard Y-arc (arrow) resulting from accumulating replication intermediates, indicative of replication stall. RITOLS intermediates are not significantly affected. No increase can be observed in the recombination intermediates.
5.3.3 **Modified mtDNA RIs in a case of a ischemic heart disease**

*(unpublished data)*

There are indications that certain cardiac pathologies either directly or indirectly affect mtDNA copy number or other functions in the diseased heart. However, a survey of somatic mtDNA rearrangements and heart pathology did not reveal any causal relationship with common heart diseases (Kajander *et al.* 2002). The mtDNA rearrangements in heart rather seem to have a loose age-dependent equilibrium with normal mtDNA, indicating that they arise from a non-pathological process. Despite being only a small sample, the heart pathologies checked in the study revealed some interesting outliers. Instead of having elevated numbers of rearrangements as one would expect, some cases had almost undetectable amounts of them. If the rearrangements result from an active recombination mechanism, their absence might indicate a switch to a more conventional DNA replication, which could itself be cited to pathology, if recombination-dependent processes are protective against mtDNA damage. This notion is supported by 2DNAGE, which I carried out on mtDNA from an individual with severe ischemic heart disease. In this case as a distinction from all other individuals that I surveyed, strong RITOLS intermediates can be detected and there are much less X-forms than in control heart samples (Figure 5.26.). This observation is thus far anecdotal, but raises interesting future perspectives (see Discussion).

![Figure 5.26. Unusual *Pvu*II and *Dra*I 2DNAGE patterns of heart mtDNA from a person who suffered from severe ischemic heart muscle disease during his lifetime. Compare with Figure 5.7 above and Figure 2b of original communication IV. Strong theta- and RITOLS intermediates can be detected.](image)
6 DISCUSSION

For nearly the last three decades it was thought that the mechanism of mammalian mitochondrial DNA replication had been elucidated thoroughly and only the identification of the relevant replicative proteins and their characterization remained. While the investigation of plant and yeast mtDNA maintenance has been hampered by the evident complexity of the processes involved, much of the mammalian mtDNA research has been biased by the assumptions of mechanistical simplicity and failure to examine tissue- or species-specific features. Most of the research has been concentrated on cultured cells or on only one model mammal – the mouse. The work by Kajander et al. (2001) was one of the first indications that the extrapolated view is inaccurate, because it described for the first time abundant recombination as a completely new feature of mammalian mtDNA. The observation was obviously not just anecdotal, as the relative quantities of junctional molecules were much higher than those reported in other systems with well-documented active recombination. It was obvious that these molecules represented some central aspects of mtDNA maintenance in healthy human heart, although the finding has been largely ignored by the field.

In this series of studies, I conducted extensive investigations of mammalian mtDNA replication under various conditions in cell culture as well in different tissues of the organism. I found new functions for genes known to be involved in mitochondrial transcription or mtDNA maintenance, characterized replication phenotypes of catalytically-defective replisome proteins, revealed physiologically-significant tissue-specific differences in mtDNA replication, and showed that these features could be manipulated in a transgenic model organism. Finally, by applying several different analytical methods in concert, I revealed evidence for a novel mechanism of mtDNA replication in human heart. Taken together, the work published in original communications I-IV has helped to deepen our understanding of the maintenance of mammalian mtDNA.

6.1 Mitochondrial DNA replication in mammalian tissues and cultured cells

The strand displacement model (SDM) of mammalian mtDNA replication was first proposed in 1972 (Robberson et al. 1972). This model was challenged almost 30 years later by a series of 2DNAGE analyses, which concluded that most mitochondrial replication intermediates (mtRIs) are double-stranded but have RNA incorporation throughout the lagging-strand (RITOLS [Holt et al.
A paper by Brown et al. (2005) questioned these 2DNAGE findings, suggesting that the apparently-duplex mammalian mtDNA replication intermediates most likely arise from artifactual branch migration of strand displacement replication intermediates. Based on AFM evidence, they furthermore proposed that there are additional light strand origins that would also explain some of the 2DNAGE findings. This notion is not new. In fact, Wolstenholme et al. (1973) as well as Pikó and Matsumoto (1977) reported at least two double-stranded regions on the displaced lagging-strand of mouse mtDNA.

I have shown by TEM that mtDNA extracted by our method from sucrose gradient-purified mitochondria contains mostly fully double-stranded theta-type replication intermediates, which can be converted into partially single-stranded species by RNase H digestion, as predicted by the 2DNAGE results (unpublished data of Pohjoismäki et al. 2007; submitted manuscript not included in this thesis). Furthermore, these TEM results are in agreement with the work done by Kirschner et al. (1968) and Wolstenholme et al. (1973), who conducted their studies on partially purified mtDNA without sucrose or CsCl gradient centrifugation steps. All replication intermediates seen by Kirschner et al. (1968) from rat liver were fully duplex, and 10-60% of hepatoma cell line mtDNA replication intermediates were fully duplex theta molecules reaching up to 85% of the genome (Wolstenholme et al. 1973). The partially single-stranded molecules seen by Wolstenholme et al. (1973) could result from degradation of RNA in crude mitochondrial preps, as reported by Yang et al. (2002).

Our experiments show that carefully isolated mtDNA contains RITOLS replication intermediates, although there is, as yet, no evidence as to how they arise. Several possibilities remain open. In the light of the SDM, RNA incorporation would be a good alternative to protect the displaced lagging-strand from degradation instead of having long stretches coated by SSB. The RNA could be synthesized de novo or incorporated from preformed RNA at the replication fork from 3’ to 5’ by an RNA helicase activity. One candidate for such a helicase might be hSUV3, a poorly-characterized mitochondrial RNA-DNA helicase–like protein (Minczuk et al. 2002, 2005). An obvious candidate for de novo RNA synthesis would be the mitochondrial RNA polymerase MTRPOL. If MTRPOL is coupled to the replication fork, the synthesis of the RNA lagging-strand would be discontinuous, just like in the case of Okazaki fragments. These RNA-Okazaki fragments could then function as elongated RNA primers for delayed lagging-strand synthesis. If MTRPOL could be shown to initiate promoter independent-transcription from a single-stranded DNA template without the presence of any transcription factors, it might provide alternative means for RNA-DNA hybrid formation. As mentioned earlier, the E. coli RNA polymerase is capable in synthesizing persistent RNA-DNA hybrid on single-stranded DNA template (Chamberlin & Berg 1964).
If preformed RNA is incorporated by an active mechanism at the replication fork, one would expect to see overhanging RNA molecules that are drawn in at the replication fork like shoelaces (Figure 6.1.), hence the term *bootlace* model (Yasukawa et al. 2006). The 2DNAGE data is very suggestive in the favour of this model. It is possible to clean up a 2D image using the single-strand RNA digesting enzyme RNase If: the heterogeneous high molecular-weight intermediates vanish and better defined RITOLS intermediates appear (Figure 6.1.). For the TEM experiments the samples also had to be treated with RNase If to remove massive amounts of heterogeneous single-stranded RNA in order to visualize any DNA. If any overhanging RNA was present on the DNA molecules, it was lost in the procedure. A careful examination of replication forks seen in TEM show that the single-strandedness at or near the fork is usually very short, indistinguishable from standard strand-coupled replication. If the RNA on the lagging-strand resulted from an artefactual hybridization of RNA on ssDNA during mtDNA purification, some discontinuities would be expected. RNA tends to form secondary structures and these would need to be opened before such an even hybridization on long stretches of ssDNA can be achieved.

![Figure 6.1. The bootlace model intermediates on 2DNAGE. Pre-formed RNA is hybridized on the lagging-strand at the replication fork resulting in heterogeneous replication intermediates with varying lengths of overhanging RNA. RNase If cleaves the overhanging RNA resulting in better defined RITOLS intermediates.](image)

In TFAM-overexpressing cells the increase in the dsDNA RIs coincides with transcriptional depletion (Article I). As TFAM overexpression causes several changes in mtDNA, it is hard to draw conclusions of the exact causal relationships. The loss of heterogeneous RNA intermediates could be due to a lack of available transcripts close to the site of replication. This observation is supported by the fact that the increased steady-state levels of RNA per template
mtDNA in TFAM RNAi cells result in a corresponding increase of heterogeneous RNA intermediates (Figure 5.11). This can be due to the reduced compaction of the DNA template, resulting in better access for RNA polymerase.

However, the effect of TFAM overexpression in cultured cells on mtDNA replication is similar to that caused by ddC and catalytic mutants of PolG and Twinkle, although the effect of TFAM is more specific, being especially pronounced in the rDNA regions. As TFAM overexpression causes compaction of mtDNA (Kaufman et al. 2007), the decreased rate of fork progression is likely to be due to the replisome having difficulties in progressing through the coiled molecule. As in other cases of replication stalling (I, II), TFAM overexpression results in mtDNA depletion over time.

The catalytic mutants of PolG and Twinkle provide some important insight into the nature of lagging-strand synthesis in RITOLS replication. While the severe PolG mutants “freeze” all replication intermediates, the severe Twinkle mutants selectively enhance dsDNA replication intermediates. Coinciding with mtDNA copy number depletion, the accumulation of RIs in both cases is a hallmark of severe replication stalling. In the case of Twinkle variants, this results from the fact that the mutant helicase is unable to unwind the DNA duplex and replication fork progression is halted. Delayed lagging-strand synthesis catches up with the retarded replication fork, resulting in fully dsDNA replication intermediates. More importantly, replication intermediates of all sizes, including those with forks close to O_H, are converted to dsDNA RIs, indicating that lagging-strand synthesis can initiate at any region of the genome (e.g. Figure 5.13). Unless SDM with random initiation of the lagging-strand synthesis is suggested, which in any case would be close to the conventional (COSCOFA) model, the data strongly indicates that the RNA of the lagging-strand can be matured to DNA from any given point. When PolG is inhibited by ddC in addition to the presence of the catalytically-inactive mutant Twinkle, which inhibits fork progression inducing stalling, the RITOLS intermediates reappear (Figure 5.14). This ddC treatment has little further effect on the progression of the leading-strand fork, which is already extremely slowed. Since ddC has a drastic effect on the rate of lagging-strand synthesis, it supports the idea that PolG is responsible for the DNA maturation and the RITOLS intermediates are genuine precursors of the fully-dsDNA species detected in control cells. Their maturation presumably requires processing of the long hybridized RNAs into shorter RNA primers for the DNA polymerase to employ.

As a further objection to the SDM, O_H appears to function as a definite replication terminus for both leading- and lagging-strand synthesis, as demonstrated by single-cleavage 2DNAGEs and TEM, meaning that the first lagging-strand origin is at or close to the O_H. It may
well be that the sister molecules separate before lagging-strand maturation is complete, resulting in nicked molecules seen on 2DNAGE. The end product of replication might be catenated molecules, as proposed earlier (Clayton 1982). Unlike trypanosome minicircles that need to be decatenated prior to replication, it seems that this is not the case with human mtDNA (Figure 5.1B).

In conclusion, RITOLS replication appears to be the main replication mode of mtDNA in mammalian cultured cells, liver, kidney and skeletal muscle. The “bootlace” maturation model is supported by a variety of experiments in cultured cells.

6.2 Replication pausing in human mtDNA

Persistent replication termini, like Ter in bacteria, require protein modulators that trap and stabilize replication forks, preventing their collapse and the formation of double-strand breaks (Neyon et al. 2005). We found binding activity for the mitochondrial transcription termination factor mTERF at the NCR region (Article III), but were unable to obtain direct evidence of mTERF involvement at the mtDNA replication terminus. 2DNAGE analysis of NCR-containing fragments reveals enhanced replication pausing around O$_H$, and LM-PCR detects enhanced stable 5´ ends on the H-strand in cells overexpressing mTERF. This suggests that the resolution of replicated molecules is actually delayed when mTERF is overexpressed.

Other replication pause sites are clearly enhanced when mTERF is overexpressed. These pause sites mainly map immediately beyond the distal end of the rRNA genes, just like replication fork barriers (RFBs) in other systems. We proposed mTERF as the first mammalian mitochondrial protein demonstrated to modulate replication pausing. Lagging-strand initiation sites at the pauses were mapped using LM-PCR and shown to correspond to the earlier known mTERF transcription termination site in the tRNA$^{Leu(UR)}$ gene at nt 3234, as well as nts 4320, 4340 and 4476 in the IQM cluster and the beginning of the ND2 gene. Other, minor 5´ ends were also detected in the vicinity. It is noteworthy that these 5´ ends were mapped on the L-strand. The existence of persistent 5´ ends on the L-strand argues strongly against the SDM model of replication.

The reason for having tightly-controlled RFBs after rRNA genes in many systems is to control the passage of replication forks and transcription complexes in actively transcribed regions. Collisions between the complexes could result both in aborted transcription as well as double-strand breaks triggering genomic instability (Vilette et al. 1995, Takeuchi et al. 2003, Prado & Aguilera 2005). At least in E. coli, head-on collisions of the transcription and replication machineries block the progress of the replication fork, whereas codirectional transcription has no negative effect
In the *E. coli* chromosome almost all essential genes are oriented so that transcription and replication are codirectional, which is proposed to minimize the mutagenic effect of repeated replication stalling and recombinational restart caused by head-on collisions (Rocha & Danchin 2003, Mirkin & Mirkin 2005, Higgins 2007). This notion stresses the importance of the mitochondrial RFBs, as all transcription originating from the HSP is antidirectional compared to replication.

In most described cases the RFB blocks the replication fork until transcription enters the RFB from the permissive direction, removing the pause-binding protein and allowing DNA synthesis to continue (Mohanty *et al.* 1996, 1998). Like mTERF, the *E. coli* Tus protein has a dual function in transcription and replication termination.

Interestingly, in TFAM-overexpressing cells, there is a substantial accumulation of replication intermediates in the rRNA region immediately after the IQM pause site (Figure 5.10). As transcription in TFAM-overexpressing cells is reduced (at least when judged by the short-lived ND3 mRNA levels), this might indicate that, without the transcriptional relief of the pause site, the replication complexes reaching the pause site halt and eventually stutter in slow motion through the rRNA region.

The RNAi depletion of mTERF produced a clear reciprocal effect only on the nt 3234 pause indicating that mTERF might not be the only replication pause binding protein in mitochondria. It is likely that the nt 4476 pause is the main physiologically significant RFB, as it is clearly the dominant replication pause in all tissues examined. Other sites might be more transient pauses or perhaps even initiation sites for lagging-strand maturation.

### 6.3 RDR in mammalian mitochondria

In contrast to most tissues and cultured cells, human cardiac muscle mtDNA has abundant recombination intermediates as well as conventional Y-arc dsDNA replication intermediates. Furthermore, half of the mtDNA molecules seem to be dimeric, which is quite an unusual feature for healthy tissues. When analysed in detail with 2DNAGE and TEM, no standard theta-like replication intermediates can be found. Recombination intermediates are also fairly abundant in human brain, but they coexist with theta-like intermediates and there are many fewer dimeric molecules. As brain is a heterogeneous tissue of several cell types, it is likely that these molecules represent mtRIs from different cells rather than several processes happening in the same
mitochondria. Low amounts of recombination intermediates can also be detected in other mammalian tissues and cultured cells (Figure 5.7, Figure 2 in IV).

Interestingly, the high levels of dimers and recombination intermediates seem to be quite specific for human heart, they cannot be detected in mouse, rabbit or pig heart. The only mouse tissue analysed with abundant recombination intermediates is brain. This fits together with observations of Pikó and Matsumoto (1977) that mouse brain has the highest frequency of circular dimers and complex molecular forms compared to heart, kidney and liver. Recombination of two monomeric circles would be the simplest mechanism of dimer formation.

The reason for such an active recombination in heart mtDNA is hard to interpret in the light of any conventional models of DNA maintenance. Due to homoplasmy, there is no need for genetic recombination in mtDNA. If it is a simple matter of double-strand break repair this would imply that heart mtDNA suffers huge amounts of damage, as inducing similar levels of recombination intermediates involves heavy use of genotoxic drugs or radiation in other systems (eg. Liberi et al. 2005, Donaldson et al. 2006). Moreover, I was unable to induce such forms in cultured cells by inducing DNA damage using KBrO₃. However, there is one abundant and physiologically-significant source of DNA damage in heart: reactive oxygen species (ROS) originating from the highly-active electron transport chain. Although oxidized nucleotides can be repaired effectively by other means, there is evidence that they are a major source of double-strand breaks (Lieber et al. 2003, David et al. 2007). The exact mechanism of DSB induction is unknown. A chain reaction–like oxidation of DNA as well as interference with replication has been suggested.

The absence of standard theta replication intermediates in human cardiac muscle mtDNA, combined with the evidence that mtDNA in this tissue is organized in multimeric networks joined by abundant recombination junctions and by catentation, suggests that mammalian mtDNA might in some cases utilize recombination-dependent initiation for replication. RDR is widely established as the main mode of mtDNA replication in many other organisms as discussed earlier in chapter 2.14.3. This conclusion is supported by the observation that the replication intermediates of PvuII- or BamHI-cut human heart migrate on 2DNAGE more in the position of a double-Y arc instead of bubble- or eyebrow-arcs (Figure 5.7). When PvuII-cut mtDNA from cultured cells is treated with S1 nuclease, the bubble arc is broken and broken bubble intermediates migrate on an arc close to the standard Y-arc form (Supplementary Figure 2e in Manuscript IV). S1 treatment of a similar heart mtDNA digest reveals that no broken bubbles are generated; instead the double-Y becomes more clearly defined. Moreover, the TEM examination of heart mtDNA reveals complex molecular networks, which include forked replication intermediates (Figures 5.5, 5.6). A partial BamHI digest reveals examples of trident structures, where three arms of a four way molecule have
equal lengths (Figure 5K in Article IV). The lengths of these arms would in most cases map one fork to the vicinity of $O_H$. This type of structure could represent invading linear molecules that have initiated replication from the invasion point, as in T4 phage DNA replication (Figure 6.2). If the invading strand is not cut, the resulting molecules have four arms and would migrate on the double-Y arc on 2DNAGE.

![Diagram of replication initiation and trident structure](image)

**Figure 6.2.** Generation of a trident structure at the $O_H$ region of mtDNA based on the T4 phage replication model. Applied from Kreuzer (2000).

Based on the existing data it is impossible to say whether RDR results in the initiation of theta-like replication as in T4 (Mosig 1998, Kreuzer 2000) or rolling circle replication as in yeast (Ling & Shibata 2004). It also cannot be excluded that heart mtDNA replicates as a linear molecule. Although longer linear molecules of more than $2n$ in human heart mtDNA were not detected, linear monomers and dimers would be sufficient to sustain such a mode of replication whilst avoiding the “end attrition” problem mentioned in section 2.13.

Curiously $O_H$ seems to be the definitive replication terminus also in human heart mtDNA; few, if any, replication forks progress past it (Figure 6.3.). This observation has further consequences for the replication of dimeric molecules with two $O_H$ regions. Dimers need to initiate twice in order to become fully replicated. Optionally, rolling circle replication initiating from $O_H$
could replicate half of the dimer and the resulting linear could be integrated into the network, initiate another round of replication or be circularized by snap-back recombination producing monomers. In order to replicate the whole genome-sized molecule RDR must also initiate site-specifically and unidirectionally from \( O_h \). Thus the NCR of human mitochondria could play the same role as the T4 phage terminal repeats. Alternatively, dimers may not be replicated as such at all; but this would require specific molecular mechanism to distinguish DNA sequences associated with dimers from monomers, which is unlikely. It also cannot be excluded that the molecules containing replication intermediates that do not terminate at \( O_h \), contain also an X-junction, making them migrate as a more complex species on 2DNAGE.

Figure 6.3. Persistent replication terminus at \( O_h \). (A) Diagrammatic view of mtDNA showing \textit{HincII} sites flanking the NCR and resulting restriction fragment with paused fork. (B) 2DNAGE pattern of a heart \textit{HincII} digest, probed for the NCR fragment. (C) Interpretation of the result. Note that the Y-arc does not reach the linear arc, indicating that the replication is unidirectional and \( O_h \) is the terminus also in heart mtDNA. It is possible that any readthrough molecules also contain an X-junction and could migrate somewhere else.

It should be noted that, whilst X-forms seen in heart mtDNA are a consequence of highly-active recombination, they do not necessarily represent replication intermediates. X-junctions seem to be fairly evenly distributed around the genome, although the appearance of the X-arc in single cut 2DNAGE might indicate that there are certain hotspot regions for cross-overs.

The distinction between X-forms and replication intermediates is important in the context of TFAM and Twinkle overexpressing mice. In these mice there is a tissue-specific increase in X-forms in heart, brain and skeletal muscle. It is possible that both Twinkle and TFAM function in recombination in mammalian mitochondria. However, they do not influence the abundance of
X-forms in cultured cells, suggesting that underlying tissue-specific molecular mechanisms are involved. The overexpression of either one of these transgenes also results in mtDNA copy number increase in mice, but not in cultured cells. In human heart the low abundance of Twinkle mRNA compared to e.g. skeletal muscle (Spelbrink et al. 2001) makes it an unlikely candidate for modulation of recombination without the involvement of other proteins. However, the possibility of tissue specific, function-affecting post-translational modifications cannot be excluded.

A possible explanation for this observation without invoking completely new activities for these proteins is that a proportion of the mtDNA in the affected tissues already replicates via RDR, but the overall copy number inside nucleoids is relatively low compared to the human heart (Dr. T. Ide, personal communication). It should be noted that wild type mice have few if any high molecular weight mtDNA forms in their mitochondria (Figure 5.27). If these forms are a representation of a more complex, higher copy number nucleoid structure, this might be further evidence for a lower mtDNA copy number per nucleoid in mouse tissues compared to human. Twinkle and TFAM could increase the copy number (maybe in independent ways), resulting in more mtDNA copies per nucleoid and providing more opportunities for strand exchange by an already existing active recombination mechanism. However, TFAM overexpression results in a much stronger effect on the intensities of the X-spike and on the loss of RITOLS intermediates than Twinkle, despite the fact that the increase in mtDNA copy number is comparable. It could be that TFAM is also involved in promoting the formation of recombination junctions, as in the case of Abf2p in yeast (Zelenaya-Troitskaya et al. 1998). The idea of density-induced recombination is further supported by the appearance of high molecular weight mtDNA forms in affected tissues of the overexpressing mice (Figure 5.27). A similar DNA density-related phenomenon is represented in T4, where complex molecular networks are formed only when there are many genomes present (Mosig et al. 1995, Mosig 1998, Kreuzer 2000). In the early phase of the T4 life-cycle, DNA replication is dependent on transcription, but when late-phase proteins such as endonuclease VII (endo VII) are expressed, there is a developmental switch to RDR. It has been suggested that mtDNA replication is also generally dependent on transcriptional activity (e.g. Bonawitz et al. 2006), which fits to the RITOLS model. Expression of a T4 endo VII–like gene in some human tissues could, in theory, be sufficient to trigger a copy number boost via RDR activation. Such a process would be much more efficient than could be achieved by gradual increase in transcription without possible disadvantage of interfering with the transcription regulation. This hypothesis is substantiated by the fact that in the higher plant Chenopodium album mitochondria the formation of complex networks also depends on the cell cycle and is linked to an overall increase in mtDNA synthesis (Backert & Börner 2000). DSB
generation could be also an alternative explanation for the TFAM and Twinkle overexpressor mice mtDNA recombination phenotype. At least in TFAM overexpressing cells there is a marked increase in the Y-forms in 2DNAGE of PvuII digested mtDNA (Figure 2b:iv-vi in I). These intermediates could represent rolling circle–like replication intermediates resulting from strand breakage at O_H. It may well be that mitotic cells recircularize these intermediates via synthesis-dependent strand annealing (Figure 2.19) or by some other means, whereas post-mitotic tissues either initiate RDR or Holliday junction formation. This hypothesis should be easily testable.

6.4 A hypothesis for origin independent COSCOFA replication in cultured cells

A small minority of mtDNA molecules in cultured cells appear to replicate via conventional, bidirectional replication that does not involve long stretches of RNA on the lagging-strand (COSCOFA). This mode of replication has been reported to initiate from a broader initiation zone and is typically characterized by dsDNA replication bubbles initiating in fragments outside O_H. It seems that COSCOFA is the main mode of replication in cultured cells recovering from mtDNA copy number depletion (Holt et al. 2000, Yasukawa et al. 2005), although in this case the bidirectional origin is close to O_H. It might well be that the replication in these reamplifying cells is still under inhibition and the apparent COSCOFA replication is simply a result of slowed replication fork progression, which is constantly being matched by lagging-strand maturation. However, the fact that full-length S1-nuclease resistant bubble arcs can be detected in HEK293T cells from as far as two-thirds of genome length away from O_H, suggests that a low level of conventional strand synchronous replication also occurs in untreated cells and that it can initiate in almost any position on the genome (Figure 6.4). The initiation and priming mechanisms of this apparent origin-independent mode of replication remain unknown.
It is possible that this replication mode could be initiated by transcription terminating elsewhere in the genome than in the NCR and forming an R-loop to initiate replication. Because this initiation does not result in RITOLS replication but in the more conventional strand-coupled (COSCOFA) mode, it is likely that the initiation mechanisms involved are not the same. Based on the previous examples, it is tempting to suggest that the best way to initiate DNA replication in an origin-independent manner would be RDR. Strand-invasion would be carried out by randomly-cut linear molecules resulting from double-strand breaks caused by aborted RITOLS replication or external damage. The fact that origin-independent replication dominates after mtDNA depletion by replication-inhibiting drugs, such as ddC or EtBr, supports the idea: RITOLS replication initiating from O_H is disturbed and the replication forks stall before reaching the terminus. Stalled replication forks need to be resolved, resulting in double-strand breaks (Figure 6.5) or fork regression, generating dsDNA linears. In response, there might more initiation of replication at O_H, but these forks would stall fairly soon resulting in the accumulation of rather short linear molecules originating from O_H. Either end of these molecules could then strand-invade another mtDNA molecule and initiate replication either at O_H or further away. Drug-induced replication stalling would result in randomly terminated linear molecules, the hotspots for the distal end would statistically locate close to but not at O_H, but would be dispersed over a wide area and difficult to map. This would be detected as the so-called origin zone (Ori Z) of COSCOFA replication (Bowmaker et al. 2003). In reamplifying cells, the main origin zone is reported to be quite discrete and located only a short distance downstream from O_H, in TAS-mt5 region ([nt 16197] Yasukawa et al. 2005). As the site was mapped using LM-PCR to detect free 5’-ends of the lagging-strand, it may well be that this “origin” is actually the dominant pause site seen in this region of the HincII digest (13637-1008, see III, Figure 4C:vi, pause site h). Degradation of broken lagging-strand linear molecules of RITOLS replication results in molecules with extensive single-stranded 3’-ends.
capable of strand-invasion and replication priming (Figure 6.5). This invading linear molecule could later be cleaved by an endonuclease as known for T4 phage, replication generating theta-like forms indistinguishable from a standard replication bubble.

Figure 6.5. A hypothetical recombination-initiated mtDNA replication mechanism in cultured cells. A replication stall results in a double-strand break. As the lagging-strand is likely to be cleaved due to the discontinuity directly at the replication fork, resulting linear fragments could be easily degraded exposing a single-stranded 3'-end. The free 3'-end could strand invade homologous sequences in other molecules and initiate replication.

A supporting piece of evidence comes from the investigation of mtDNA from Twinkle K421A-expressing cells. This artificial mutation abolishes helicase activity and results in a strong replication-stalling phenotype with around 20% of the molecules containing stalled replication forks. The situation is somewhat analogous to the drug-induced replication stall. 2DNAGE of mtDNA from these cells shows bubble-arcs that extend all over the HincII O_H containing fragment (13637-1008 [Figure 5.13]), and molecules where the arms of the replication fork map outside O_H (Figure 5.22B) are observed by TEM, both findings indicating random replication initiation. Of nineteen replicating molecules examined from K421A Twinkle-expressing cells, one had a linear molecule protruding out from the middle of a theta replication intermediate (Figure 5.21). Closer examination showed that the linear is genuinely connected by a junction and is not just a separate linear molecule (see insert in Figure 5.21). The distance of the replication forks from the branch point is equal and the distal end of the linear arm would map outside of the bubble region, if the sequences are assumed to be homologous. If the linear fragment stems from an earlier terminated replication
event, the two replication events should have different origins. Moreover, one fork from the second replication event would have passed the branch point without interference, which intuitively seems impossible. Although this is still an anecdotal observation, it is very difficult to postulate any other explanation for such a structure, except that of a linear molecule invading the circle and initiating bidirectional replication from the initiation point, as in T4 phage and illustrated in Figure 6.5.

Also in line with RDR as the primary mechanism of origin-independent replication, mitochondrially-targeted MnSOD-RecA seems to enhance S1 resistant replication intermediates in cultured cells, as expected for an enzyme that facilitates strand-invasion. In contrast, RecA does not alter the X-form abundance bringing further support to the idea of a noncrossover recombination mechanism in cultured cells (Figure 5.25B). Alternatively this might result from a similar DNA replication inhibition effect by a DNA binding protein as seen with TFAM overexpression. If RDR exists in all tissues, it might, most of the time, function as a backup mechanism against short mtDNA fragments that mostly arise from aborted replication replication events, whereas regular DSBs are either repaired by the synthesis-dependent strand annealing or crossover recombination. It should be noted that once the RDR has reached the end of the molecule the result is indistinguishable from synthesis-dependent strand annealing recombination. Clearly further studies are needed in order to resolve this issue.

6.5 Modulation of mtDNA topology

With a few exceptions, in vivo tissues and cultured cells display a similar profile of different topological forms of mtDNA. Human skeletal muscle has the fewest catenated molecules and most monomeric supercoils compared to other tissues. It should be noted that the only skeletal muscle investigated in this study was iliopsoas. Muscles having more red muscle fibers are expected to have both higher mtDNA copy number and perhaps a different organization. Human heart had abundant high molecular weight forms, which based on TEM appear mainly to consist of tangled complexes, as well as dimeric circles and their derivatives. It is highly likely that these high molecular weight forms in heart, as well as the catenanes seen in other tissues and cells, represent the core of functional mitochondrial nucleoids.

Although the correct partitioning and regulation of nucleoid size is essential for mtDNA segregation, nothing is known about the mechanisms involved in these processes in mammalian mitochondria. The loss of catenated mtDNA in favour of monomeric supercoils in TFAM-overexpressing cells may relate to this issue. Decatenation is important for the dispersal of the
recently-replicated mtDNA, and in HEK293T cells the largest aggregates of mtDNA contained maximally four molecules. In yeast, the TFAM homologue Abf2p is involved in the partitioning of mtDNA into daughter cells, although this process involves recombination rather than decatenation (MacAlpine et al. 2000). Recently a novel protein, ATAD3, was suggested to be involved in segregation of nucleoids in mammals, but the functional mechanism is still unknown (He et al. 2007). ATAD3 is reported to have specificity for the D-loop region and it is possible that TFAM-induced depletion of 7S DNA plays a role either in the regulation of the nucleoid or in catenane size.

The mechanism might also be due to the altered kinetics of mtDNA replication – a reduced replication speed cannot keep up with the partitioning of mtDNA in nucleoids. The supercoiling could also be independent from the decatenation and result from excess TFAM compacting the DNA template to which it is bound. This notion is supported by a recent article on in vitro analysis of TFAM and DNA interactions (Kaufman et al. 2007). Similarly, the E. coli HU protein can supercoil circular plasmid DNA (Marians 1987).

Besides its effects on 7S DNA, TFAM overexpression depletes termination intermediates migrating as Holliday junction-like (HJL) species on 2DNAGE of O_{H}-containing fragments (marked as “t” in Figure 2C of I). These structures represent X-like forms comprising replication forks coming from opposite directions and joining at O_{H}. In the TFAM-overexpressing cells their depletion coincides with reduction in mtDNA copy number, suggesting that they are resolved by a post-replicational mechanism. It is possible that Twinkle is involved in their resolution, because HJLs are the only affected mtDNA structure in cells overexpressing wild-type Twinkle (Figure 5.13) that do not show any copy number decrease. The precise molecular structure of the termination-associated HJL is unclear, because they cannot simply comprise two oppositely moving forks that have almost met, such a structure being thermally unstable unless the unreplicated region between them is at least 50 bp (in which case they would not lie precisely on the X-arc on 2DNAGE gels). It is possible that they are held together by hemicatenation after ligation of at least one of the strands. Such junctions have been reported to form in the replication origins of Sulfolobus solfataricus–archaea (Robinson et al. 2007). Hemicatenation can be resolved by the co-operative action of a topoisomerase and helicase, playing a role in the segregation of chromosomes (Harmon et al. 1999).
6.6 mtDNA replication and deletion formation

Mitochondrial DNA molecules with large scale deletions can be maintained in cultured cell cybrids (e.g. Tang et al. 2000), but to my knowledge there are no reports that deletions would accumulate in cell culture as they do in mitochondrial pathologies. It might be that mtDNA deletions are difficult to detect in this system, as they might become lost at mitosis in rapidly-dividing cells. The proposed hypothesis for origin-independent replication provides an alternative explanation: the mitotic cells replicate their mtDNA more actively in order to maintain their copy number in each cell division. As a result, strand-invasion events by linear DNAs could recruit replisomes and initiate replication, resulting in the loss of deleted linear molecules by converting them systematically to fully-replicated mtDNA.

Nevertheless, deleted mtDNA molecules can be detected in various post mitotic human tissues (Ozawa 1999, Kajander et al. 2000, Zeviani & Carelli 2003). Examination of their common break-points in the light of the new data provides potential insight into the relationship between replication and recombination in deletion formation. As seen from the Figure 2.20, many of the breakpoints in human heart mtDNA sublimons coincide with the tRNA\textsubscript{Leu(UUR)} and ND2 replication pauses, just as they do in mice overexpressing PEO-associated mutant of Twinkle (Kajander et al. 2000, Tyynismaa et al. 2005). This is understandable, because RFBs are sites of persistent replication forks that can easily terminate prematurely resulting in a DSB. Furthermore, in mTERF-overexpressing cells the X-forms are enhanced in the tRNA\textsubscript{Leu(UUR)} gene region and conversely depleted by mTERF RNAi (Figure 5.8). A possible explanation is that persistent pausing due to mTERF overexpression results in replication fork collapse and recombinogenic restart of replication. Similar events are known in yeast, where double-strand breaks at RFBs can result in low-level genomic rearrangements (Weitao et al. 2003). This implies that mTERF is a good candidate gene in genetic disorders having mtDNA rearrangements, but whose genetic basis is yet unknown.

However, replication stall at RFBs does not explain the mechanisms of the common 5 kb deletion seen in KSS as well as in many other mtDNA disorders and even in UV-irradiated skin (Schon et al. 1989, Mita et al. 1990 Berneburg et al. 1999). Because one break point is always located at the proximal part of the replicating mtDNA molecule (such as nt 11460 in the 5 kb common deletion), it may be that these deletions arise from replication forks stalling soon after initiation. If strand breakage occurs, the recombination machinery could find the nearest homology in the same molecule (Figure 6.6). This hypothesis is supported by the fact that the majority of mtDNA deletions are flanked by homologous direct repeats (Schon et al. 1989, Mita et al. 1990). It
should be noted that in *S. cerevisiae* the minimal length of homology required for noncrossover recombination is 25-60 nucleotides, whereas crossing over requires 150-200 nucleotides (Hayden & Byers 1992), allowing a synthesis-dependent strand annealing mechanism to initiate illegitimate recombination.

![Diagram of a potential mechanism for the generation of Class I common deletions.](image1)

Figure 6.6. A potential mechanism for the generation of Class I common deletions. Replication stalls soon after initiation resulting in DSB. The recessed 3’-end strand invades a homologous site further away in the same molecule resulting in a deleted molecule when the junctions are resolved.

### 6.7 Cardiac muscle mtDNA replication – physiology and pathology

We still do not know much about the physiological significance of RDR in heart. Fortunately the TFAM and Twinkle mouse models provide us some insight on the issue. Primarily, it seems that recombination is universal in mammalian high energy-demand tissues, although it does not necessarily manifest as stable Holliday-junctions. Recombination provides a means of protection against double-strand breaks as well as against other mutations such as inter- and intra-strand crosslinks. Oxidative damage by KBrO3 induces replication stalling in cultured cells (Figure 5.25A). These stalls can readily result in double-strand breaks. In similar fashion, frequent oxidative damage of heart mtDNA would be highly problematic without an active and frequent recombination system as a backup. In a highly mutagenic environment, recombination could also maintain homoplasmy via a gene conversion–like mechanism. However, it is unlikely that this could discriminate deleterious mutations from the wild type. In their study, Lin and co-workers (2003) report both increased levels of 8-OHdG (an indicator of oxidative DNA damage) and increased amounts of mtDNA deletions in atrial muscle from patients with atrial fibrillation. It may be that in these cases the elevated oxidative stress results in an excess of DSBs for the recombination
machinery to deal with, producing deletions via non-homologous end-joining (NHEJ) or illegitimate recombination. Instead of the strand-invasion step, other factors such as Twinkle might be limiting for the RDR mechanism.

Twinkle and TFAM overexpressors are not only completely healthy mice, but they are also protected from cardiac remodelling after infarction-reperfusion of the heart muscle. It is difficult to think of a role for mtDNA recombination in the event, as the damage to the ischemic tissue itself is not ameliorated. More likely the increased mtDNA copy number could influence an early stage of the remodelling signalling cascade in still healthy cardiomyocytes in the overexpressors. As mentioned in section 2.21, DCM remodelling results in mtDNA copy number decrease and a shift to a more embryonic–like energy metabolism. The increased copy number alone in the tissue surrounding the ischemic damage could be enough to prevent the DCM maladaptation.

This finding has important implications for human heart muscle pathology. It can be assumed that the developing heart at some time point switches from the “typical” RITOLS replication of an embryonic heart into the complex replication and recombination–type mtDNA maintenance of an adult heart. A reciprocal process might be expected when adult cardiac muscle in DCM switches to a more embryonic-like physiology. Indeed, it seems that cardiac muscle mtDNA from a patient with severe ischemic heart disease behaves differently (Figure 5.26.). There are abundant RITOLS–like replication intermediates and theta forms, but very little recombination intermediates. Whether this finding can be generalized to other cases of cardiac diseases and exactly what it means remains to be seen.

If the prediction is true and the change in replication mode reflects also the change in physiological condition of the heart muscle, the observation might have some interesting consequences. The current information on the mechanisms responsible for mitochondrial biogenesis supports an energy-driven stimulus in relation to increased fatty acid availability (Hood 2001, Garcia-Roves et al. 2007). Cardiac muscle utilizes fatty acids for its energy and a developmental change in DCM to glucose burning might be a sufficient stimulus to switch the mtDNA replication mode. Interestingly, overexpression of the cytosolic form of phosphoenolpyruvate carboxykinase (PEPCK-C) in mouse muscle results in a marked increase in the red muscle fibres, mitochondrial content and exercise endurance (Hakimi et al. 2007). These mice preferrably metabolize fatty acids and do not suffer of lactate-induced fatigue during strenuous exercise. Furthermore, they have an elevated mtDNA copy number that is comparable to the levels seen in TFAM and Twinkle overexpressing mice.
SUMMARY AND FUTURE PERSPECTIVES

In this thesis I have presented the results from four different studies on mammalian mtDNA replication under various conditions in cell culture and in mammalian tissues with different energy requirements. There are at least two types of mtDNA replication modes in mammalian tissues. The more commonly operating mechanism is RNA intermediate-rich RITOLS replication. The relative rates of fork progression and lagging-strand maturation in RITOLS replication can be readily manipulated by altering the expression of TFAM or by expressing catalytic mutants of Twinkle and PolG in cultured cells. The results indicate that lagging-strand RNA is maturated with a delay relative to forward fork movement and PolG is involved in the maturation process.

A number of issues remain to be settled before RITOLS replication can be accepted as a new “textbook model” of mtDNA replication. While it is perfectly plausible that the MTRPOL primes leading-strand synthesis as in the SDM, the mechanism by which lagging-strand RNA is incorporated remains to be elucidated. Transient inhibition of MTRPOL by RNAi might prove a useful approach, but as an ultimate test of the bootlace model more TEM studies are needed. With careful preparation of mtDNA combined, for example, with gel filtration, it might be possible to purify mtDNA replication intermediates that have bootlace RNA overhangs from the abundant single-stranded RNA seen otherwise in all preparations not subjected to RNase treatment. As for the origin-independent mechanism, it would be informative to study reamplifying cell mtDNA using TEM in order to detect clear-cut cases of recombination-mediated initiation.

While it is likely that the recombination-dependent mechanism operates in high energy-demand tissues, it remains to be seen, whether the origin-independent dsDNA-rich replication mode in cultured cells relates to RDR. The highly active RDR mechanism in human heart could represent a physiological adaptation against a highly oxidative environment, protecting from catastrophic replication errors and effects of double-strand breaks while maintaining the maximum fidelity of the gene products. It is likely that the same mechanism operates in all mammals, but in human heart the mtDNA copy number per nucleoid could be high enough to result in complex DNA networks as seen in other phylogenetically more diverged organisms. Moreover, the importance of the heart and its dependency on OXPHOS during a very long adult lifetime argues in favour of special protective mechanism, of which this may be a manifestation.

The significance of recombination in other tissues is also worth of further investigation. Low amounts of recombination intermediates can be detected in all tissues and cultured cells analyzed in this study. As mitochondria are highly-dynamic organelles capable of fission and fusion,
it is likely that they also exchange genetic material. In fact, indirect evidence of this exists, as discussed earlier. On one hand, the high segregation number in cultured cells indicates that, at least in these systems, there is no significant recombination and gene conversion. On the other hand, sudden heteroplasmy changes or “nucleoid reorganizations” occur (Jacobs et al. 2000). It is plausible that these are due to the mixing of different mtDNA alleles into the same nucleoids. With enough time recombination would eventually create mismatching recombinant strands, resulting in gene conversion. If some somatic cell types were more prone to mitochondrial mixing than others, this might be an explanation for tissue specific differences and developmentally programmed changes in heteroplasmy levels, distinct from the random genotypic drift seen in rapidly dividing cell lines. To investigate the extent of RDR, highly active red muscle-fibre skeletal muscles such as diaphragm might be interesting to analyze in the future.

Replication fork barriers (RFBs) play an important role in the maintenance of the fidelity of mtDNA transcription and replication. The mitochondrial transcription termination factor mTERF is the first such protein known from human cells and it is likely that mTERF-induced replication fork pausing has an important physiological function. Labile replication pausing can result in replication stalling and double-strand breaks with undesirable results, such as highly recombinogenic ends promoting pathological rearrangements of mtDNA. As further evidence, the replication pause sites characterized in this study represent also break points for many common mtDNA deletions seen in human tissues.

While NHEJ and recombination play an important part in deletion formation in post mitotic human tissues, mitotic cells could eliminate deleted molecules by initiating RDR from any DSB. The circularization of linear molecules resulting from DSBs might be a slow process and linear molecules not able to amplify, thus becoming lost in rapidly dividing cells. In the few mouse models of mitochondrial diseases two mechanisms of deletion formation can be observed: Twinkle linker-duplication and mitochondrially targeted-PstI mice show circular mtDNA deletions very similar to those seen in human heart or in mtDNA rearrangement disorders, whereas the “deletions” in the PolG mutator mouse are linear (Trifunovic et al. 2004, Tyynismaa et al. 2005, Srivastava & Moraes 2005). Further molecular comparison of these mice might provide interesting insights into the mechanisms of deletion formation in different cases of genetic stress.

The developmental regulation of mtDNA replication in the PEPCK-C mice could prove to be informative. The comparison of preferred energy source, mtDNA copy number and exercise tolerance should also be done with TFAM and Twinkle mice in order to test the dependence of mtDNA maintenance on energy metabolism in the tissue.
ACKNOWLEDGEMENTS

This will be more of an epilogue, but I feel that the past five plus years deserve more than just a couple of lines. A lot of things have happened and changed for me since April 2002. I never thought that doing a PhD would be this difficult... Although, I remember that a senior researcher back in my undergraduate days mentioned that “a PhD thesis should not be easy”. I can’t remember anymore if it was meant as an anecdote or was he commenting the incidence of a graduate student named Mika ending up in ER because of overstraining. Fortunately I was always a bit lazier than Mika.

The thesis work was carried out in the laboratory of Molecular Biology, Institute of Medical Technology at the University of Tampere. Some of the experiments were performed at the Lineberger Cancer Institute in UNC, Chapel Hill. Many of the experiments would have been impossible without the collaboration of the Department of Forensic Science in Tampere University hospital as well as Dr. Ian Holt’s group in the MRC Dunn Research Institute in Cambridge. I also like to thank Anu Wartiovaara’s group in the Department of Neurology in Helsinki, Dr. Tomomi Ide and Professor Dongchong Kang in the Kyushu University Graduate School of Medical Sciences in Fukuoka for the important collaboration regarding the transgenic mice. The research was made possible by the Tampere Graduate School in Biomedicine and Biotechnology, the European Union, the Academy of Finland, the Juselius Foundation, the Tampere University Hospital Medical Research Fund and the Tampere University Support Fund.

I am grateful to Prof. Takehiko Shibata and Prof. Laurie Kaguni for reviewing my thesis and for the excellent suggestions to improve both the content and the language. I would especially like to thank Prof. Eric Schon for being my opponent, something I feel quite privileged about.

It has been extremely nice to get to know Jack Griffith and Smaranda Willcox. Thank you as well as the rest of the Griffith-lab for the nice stay in Chapel Hill and a great new method. I think it will prove to be a very fruitful collaboration. I would like to thank Professor Pekka Karhunen for all the conversations, encouragement, samples and science. I am also thankful for the scientific input and fruitful discussions with Dr. Ian Holt, Prof. Anu Wartiovaara, Prof. Deepak Bastia, Dr. Tom Schneider, Prof. Arnold Bendich and Dr. Patric Lestienne.

Despite of – or should I say thanks to - the more or less anarchistic supervision by Howy in the early stages of the project, combined with a “find-out-what’s-happening” research outline, I think I have developed more as a scientist as I would have with any other career plan that I had back in 2002. I didn’t need to worry about funding, there were plenty of opportunities to travel, to interact with people in the field and to exchange ideas. The original scientific communications that were (and some that will be) published in the highly appreciated journals in the field are completely thanks to the brain power and superb writing skills of Howy. He taught me how to make the best scientific representation of data, something that is equally valuable as his inventions and original ideas. Thanks Howy for the good example and for providing the scientific environment and resources. I think I found my way of conducting research, which really satisfies, and self confidence for my work. Moreover, I think I actually managed to contribute something that has significance for the field.

I would especially like to thank Steffi for all the things she knows anyway (♥). I appreciate sincerely Anne, Sjoerd and Henna for their involvement in the project as well as their friendship. Hans deserves special thanks for guidance and ideas as well as involving me in the Hans-lab. Thanks for all present and former members of H&H-labs for being such a nice and social but somewhat peculiar collection of people. Sorry for being too occupied lately to participate in
anything. I predict more social events for 2008! Special thanks for Anja, Outi and Merja for running the labs! Iso Kiitos myös Kaisu Pekoselle kaikenlaisten sekalalisten asioiden selvittämisestä, paketeista ja yökaluista!

There are plenty of people who have improved my quality of life during these years. Especially the Joensuu-originated crew, who shares the same passion for life (biodiversity s. lat.), bugs, hunting, fishing and good fun: Jussi, Pekka, Henkka, Perttu, Miikka, Vesa et al. Thanks for the past trips and evenings together. Lisätään intensiteettiä ja jatketaan samaa!

Big appreciation for all of “them good old boys” in Tampere, some of who have been my best friends for 20 years: Antti, Ville, Teemu and Kalle. Love you guys (in laconic – or was it platonic way, of course)! Kiitos reissuista, koheltamisista, pitkistä illoista ja aina hyvistä puheenaiheista.

An important balance for the white-coat scientist work has been the ability to pursue my holistic naturalist vocation in the field of entomology and ecology. Special thanks for integrating me into the circles and for the interesting discussions on ecology and determination of all sorts of creepy-crawlers: Juha Salokannel, Keijo Mattila, Tero Piirainen, Matti Koivikko, Jukka Salmela, Jere Kahanpää et al. Many thanks also to Markus Rantala and Ilari Sääksjärvi for involving me in their research interests as well as for being such nice guys.


Tampere, January 2008

Jaakko Pohjoismäki
REFERENCES


Jacobs HT, Lehtinen SK & Spelbrink JN. 2000: No sex please, we’re mitochondria: a hypothesis on the somatic unit of inheritance of mammalian mtDNA. *BioEssays* **22:**564-72.


Maniura-Weber, K., Goffart, S., Garstka, H.L., Montoya J. & Wiesner, R.J. 2004: Transient overexpression of mitochondrial transcription factor A (TFAM) is sufficient to stimulate mitochondrial DNA transcription, but not sufficient to increase mtDNA copy number in cultured cells. *Nucleic Acids Res.* **32:**6015-27.


Matthews, P.M., Brown, R.M., Morten, K., Marchington, D., Poulton, J. & Brown, G. 1995: Intracellular heteroplasmy for disease-associated point mutations in mtDNA:
implications for disease expression and evidence for mitotic segregation of heteroplasmic units of mtDNA. *Hum Genet.* 96:261-8


Stocking, C. & Gifford, E. 1959: Incorporation of thymidine into chloroplasts of Spirogyra. *Biochem Biophys Res Comm.* 1:159-64


148


ORIGINAL COMMUNICATIONS
Alterations to the expression level of mitochondrial transcription factor A, TFAM, modify the mode of mitochondrial DNA replication in cultured human cells

Jaakko L. O. Pohjoismäki1, Sjoerd Wanrooij1, Anne K. Hyvärinen1, Steffi Goffart1, Ian J. Holt2, Johannes N. Spelbrink1 and Howard T. Jacobs1,3,*

1Institute of Medical Technology and Tampere University Hospital, FI-33014 University of Tampere, Finland, 2MRC-Dunn Human Nutrition Unit, Hill Road, Cambridge CB2 2XY, England, UK and 3IBLS Division of Molecular Genetics, University of Glasgow, Glasgow G12 8QQ, Scotland, UK

Received July 17, 2006; Revised September 8, 2006; Accepted September 11, 2006

ABSTRACT
Mitochondrial transcription factor A (TFAM) is an abundant mitochondrial protein of the HMG superfamily, with various putative roles in mitochondrial DNA (mtDNA) metabolism. In this study we have investigated the effects on mtDNA replication of manipulating TFAM expression in cultured human cells. Mammalian mtDNA replication intermediates (RIs) fall into two classes, whose mechanistic relationship is not properly understood. One class is characterized by extensive RNA incorporation on the lagging strand, whereas the other has the structure of products of conventional, strand-coupled replication. TFAM overexpression increased the overall abundance of RIs and shifted them substantially towards those of the conventional, strand-coupled type. The shift was most pronounced in the rDNA region and at various replication pause sites and was accompanied by a drop in the relative amount of replication-termination intermediates, a substantial reduction in mitochondrial transcripts, mtDNA decatenation and progressive copy number depletion. TFAM overexpression could be partially phenocopied by treatment of cells with dideoxycytidine, suggesting that its effects are partially attributable to a decreased rate of fork progression. TFAM knockdown also resulted in mtDNA depletion, but RIs remained mainly of the ribosubstituted type, although termination intermediates were enhanced. We propose that TFAM influences the mode of mtDNA replication via its combined effects on different aspects of mtDNA metabolism.

INTRODUCTION
In mammalian cells, mitochondrial DNA (mtDNA) was long believed to replicate by an unusual, strand-asymmetric mechanism (1). However, recent studies, using two-dimensional neutral agarose gel electrophoresis (2DNAGE), have revealed the presence, both in vertebrate tissues and cultured cells, of two classes of mtDNA replication intermediates (RIs) whose structures are not consistent with the strand-asymmetric model. Both classes are essentially duplex throughout their length, but differ in their ribonucleotide content (2,3). One class shows extensive RNA incorporation on the lagging strand [ERIOLS, Ref. (2)], whereas the other has structures fully consistent with conventional, strand-coupled DNA replication (3–5). ERIOLS intermediates are generally nicked or gapped on the RNA strand (2) and are hence labile to partial degradation during extraction.

The mechanistic relationship between RIs of the ERIOLS and strand-coupled types, as well as how they relate to the ‘orthodox’, strand-asymmetric replication model, are not properly understood. ERIOLS intermediates have been suggested to be processed to resemble those of the strand-coupled type via a maturation step (2). Different replication modes may also operate simultaneously in the same cell. In solid tissues, strand-coupled replication appears to initiate bidirectionally in a broad origin zone, spanning at least several kilobases downstream of the major non-coding region (NCR) of the genome (5,6). In cultured cells recovering from drug-induced mtDNA depletion, such initiation is confined to a much narrower region of the NCR (3). The initiation mechanism which gives rise to RIs of the ERIOLS type remains unclear. Initiation within the NCR can also give rise to the synthesis of 7S DNA (1), which establishes the characteristic D-loop form of mtDNA, although its relationship with productive replication of the genome remains enigmatic.

*To whom correspondence should be addressed. Tel: +358 33 55 17 731; Fax: +358 32 15 77 10; Email: howard.t.jacobs@uta.fi

© 2006 The Author(s).
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
The mitochondrial transcription factor TFAM, an abundant HMG-box protein of mitochondria, appears to have multiple functions in mtDNA metabolism (7). It was originally characterized by the absolute requirement for the protein for transcriptional initiation in vitro at either the heavy- or light-strand promoter of the genome (8). More recently, these findings were confirmed using a fully reconstituted system, containing mtDNA-derived templates, RNA polymerase and the additional transcription factor TFB1M or TFB2M (9). Transcription from the light-strand promoter is required to create the primer for heavy-strand mtDNA synthesis according to the orthodox, strand-asymmetric replication model. Therefore TFAM has been considered to be an essential protein also for mtDNA replication. Consistent with this view, abolition of TFAM expression using a conditional knockout strategy in the mouse showed clearly that TFAM is required for mtDNA maintenance as well as cellular function and survival (10). However, this finding is also consistent with TFAM protein having other essential roles in mtDNA metabolism.

TFAM has been proposed to play a structural role in the maintenance of the mitochondrial chromosome, independent of its transcriptional activity. It is highly abundant, is mainly (or entirely) complexed with mtDNA in nucleoid structures (11,12), shows significant, non-sequence-specific DNA-binding (8) and promotes DNA compaction (13), leading to the suggestion that it coats the entire DNA in a manner similar to histones in the eukaryotic nucleus or the HU protein in bacteria (14). Its homologue in yeast, Abf2p, has been shown to induce compaction by introducing sharp bends into the DNA backbone (15) and is required for the stable maintenance of wild-type mtDNA (16).

Mammalian TFAM has preference for binding to branched DNA structures such as Holliday junctions (17) and to cisplatin-damaged or oxidized DNA (18). In vitro, TFAM promotes the resolution of D-loop forms (19). It also interacts physically with p53 [Ref. (20)], suggesting a possible function in DNA repair or other recombinational processes. Although it does not have a directly protective role, TFAM overexpression in rat myoblasts has been reported to accelerate the recovery of mtDNA levels after peroxide damage (21) and transgenic expression of human TFAM in mice mitigates mtDNA loss and other mitochondrial defects after cardiac ischemia (22). In yeast, Abf2p is required for recombination and segregation of mtDNA to daughter cells (23) and genetic evidence also implicates it in non-recombinational mtDNA repair pathways (24). All of these pieces of evidence point to TFAM and its homologues being key regulators of DNA transactions in mitochondria.

In organello, TFAM imported into rat liver mitochondria stimulates the synthesis both of mitochondrial RNAs (25) and 7S DNA (26). Transient overexpression of TFAM in cultured HEK cells also results in increased transcription, but with no change in mtDNA copy number (27). However, high levels of TFAM added exogenously in vitro (9,28,29), as well as prolonged over expression in HEK cultured cells (27), bring about a paradoxical decrease in transcription. This effect may be attributable to an over-condensed state in the template DNA. The transient increase in mitochondrial transcription brought about by TFAM over-expression in HEK cells is accompanied by an increased level of RNase H-sensitive mtDNA species (27), which may correspond with RIs of the ERIOLS type.

Several lines of evidence suggest that TFAM regulates mtDNA copy number independently of its role(s) in transcription. Heterozygosity for TFAM knockout produces copy number depletion of ~40% in mice (10) and 50% in chicken cells (30), but with minimal effects on RNA levels. Although human TFAM has only a weak transcription-stimulatory effect on mouse mtDNA promoters in vitro, transgenic expression of human TFAM in mice produces a stoichiometric increase in mtDNA levels (31). Overexpression of a transcriptionally inert variant of TFAM in human cells also results in a proportionate copy number increase (14), whilst TFAM knockdown by RNA interference (RNAi) causes copy number depletion with only minor effects on transcription per template molecule (14). In chicken cells in which the endogenous c-TFAM gene has been disrupted, transgenic c-TFAM lacking the C-terminal tail region required for transcription-promoting activity is nevertheless able to support maintenance of mtDNA at 50% of wild-type levels (30).

The idea that TFAM regulates mtDNA copy number by a simple titration model is, however, contradicted by the observation that, following transient, ethidium bromide (EtBr)-induced mtDNA depletion in cultured cells, TFAM levels were observed to recover more slowly than mtDNA. This suggests that the packing ratio of TFAM on mtDNA can vary and may influence the rate of mtDNA replication (32).

In order to investigate further the effects and mode of action of TFAM on mtDNA replication and copy number modulation, we analysed mitochondrial RIs from cells overexpressing TFAM and from cells in which TFAM expression was knocked down by RNAi. We report here that overexpression of TFAM brings about a dramatic change in the relative abundance of strand-coupled versus ribosubstituted (ERIOLS-type) RIs, accompanied by systematic alterations in copy number, transcript levels and mtDNA topology. In contrast, TFAM knockdown results in copy number depletion, but with only minimal effects on mtDNA RIs. The findings are consistent with the idea that the synthesis of mtDNA is dependent on at least two different TFAM-influenced processes, one of which is transcription-associated, the other related to mtDNA organization.

**MATERIALS AND METHODS**

**TFAM constructs**

Full-length TFAM cDNA lacking its usual stop codon was amplified from a cDNA clone (27, kind gift of Dr R. Wiesner), using the following primers (restriction sites as indicated, underlined, start codon in bold italics): 5'-CCGGAATCCGGCATGCGTTTCTCCGAAAGC-3' (EcoRI) and 5'-CGGGATCCACACTCTCCAGCACCACATATTTCGC-3' (BamHI). The restriction-digested PCR product was ligated into EcoRI + BamHI-cut pcDNA3.1(−)Myc-HisA (Invitrogen) to create the construct mtTFA-myc, capable of directing the expression of C-terminally Myc-His epitope-tagged TFAM. The full-length TFAM cDNA, including the natural stop codon, was amplified using the following primers...
(restriction sites as indicated, underlined, plus start and stop codons in bold italics):

5'-CCCAAGTTGGCGTCTCTCCGGAAGC-3' (HindIII) and 5'-CCGCGATCCATTACACTCTCCGACCATATTTTCC-3' (BamHI).

The restriction-digested PCR product was ligated into HindIII + BamHI-cut pcDNA3.1(+) vector, to create the construct mTFAPcDNA3.1 for transient expression.

In order to create cell lines inducibly expressing either natural TFAM or C-terminally Myc epitope-tagged TFAM, the two plasmids described above were digested with PmeI and the liberated inserts recloned into the vector pcDNA5/FRT/TO (Invitrogen), which was then transfected into the Flp-InTm T-RexTm-293 host cell line (Invitrogen) according to the manufacturer’s recommendations. Full details of the use of this system to create cell lines inducibly expressing proteins involved in mtDNA metabolism will be published elsewhere (S. Wanrooij et al., manuscript in preparation).

siRNAs for TFAM knockdown
TFAM siRNAs were synthesized using the SilencerTm siRNA construction kit (Ambion). Six putative TFAM-specific siRNA sequences were selected using the manufacturer’s prediction programme (www.ambion.com/techlib/misc/siRNA_finder.html). After testing by transfection and western blotting (see below), two were found to be efficient. The sequences of the relevant mRNA targets were as follows, Si2: 5′-AAATGGTGCTCAAGAAACCTGT-3′ (np 273–293 of the TFAM mRNA sequence, Genbank NM_003201) and Si4: 5′-AAGATGCTTATAGGGCGGAGT-3′ (np 431–451, exon 4). See legend to Supplementary Figure 3 for details of other siRNAs tested in trial experiments.

Cell culture and transfection
HEK293T and Flp-InTm T-RexTm-293 cells (Invitrogen) were cultured in DMEM containing 4.5 g/l glucose, 2 mM L-glutamine, 1 mM sodium pyruvate, 50 μg/ml uridine and 10% fetal bovine serum, at 37°C in a humidified atmosphere with 5% CO2 in air. No antibiotics were added for HEK293T cells, but transgenes in Flp-InTm T-RexTm-293 cells were maintained under selection with hygromycin and blasticidin according to the manufacturer’s recommendations. All cell lines were routinely detached by pipetting alone and passaged at 1:10 dilution every 3–4 days. TFAM transgene expression was induced by adding 10 ng/ml doxycyclin (Sigma-Aldrich) to the culture medium for the times indicated in the Figures and legends. Transfections were carried out using TransFectinTm-lipid reagent (Bio-Rad), following the manufacturer’s recommended procedure, with 12 μg of plasmid DNA and 40 μl of reagent, both in 1.5 ml of serum-free medium, per 10 ml plate.

For TFAM knockdown, HEK293T cells were transfected using LipofectamineTm 2000 (Invitrogen) and a final concentration of 20 nM of siRNA. To arrest mtDNA synthesis by chain termination, cells were treated for various times in medium containing 100 μM dideoxycytidine (Sigma-Aldrich). To suppress both mtDNA copy number and mitochondrial transcription, cells were treated with medium containing EtBr (50 ng/ml) for 72, after which cells were washed and replated in fresh medium, then cultured for a further 48 h.

DNA and RNA extraction and quantitation
For mtDNA copy number analysis, total cellular DNA was extracted using standard techniques (33). Copy number was assessed independently by two different methods, to minimize possible artefacts. For Southern blotting, total DNA was cut by EcoRI and analyzed as described in Ref. (34), with quantitation by phosphorimaging (Storm 840 scanner and ImageQuant 5.1 software, all from Molecular Dynamics). Copy number was also estimated by real-time quantitative PCR (35) with Taqman probes for mitochondrial cytochrome b and for amyloid precursor protein (APP), used as a single-copy nuclear DNA standard. Primers and probes were as follows (all 5′-3′): APP Forward: TTGTGTGTGGTCTCC-CAGGTCT, APP Reverse: TGGCTACGTGGTGTTGGC, APP Probe (FAM+BHQ): CCGTGAACTCGACGACTCCA-ATGTGAG, Cyt-b Forward: GCCTGCTGATCCCT-AAAT, Cyt-b Reverse: AAGGTAGCGATGTATCCAGCC, Cyt-b Probe (TET+BHQ): CACCAGACGCTCTACCC-OCCTT.

RNA extraction from cells, agarose or urea–PAGE and Northern hybridization were as described previously (34,36). Probes were 32P end-labelled oligonucleotide as follows (5′–3′): for ND3 mRNA, GCTACTCATAGGCCAGCATT, for 5S rRNA (loading control), GGGTGGTATGGCCTAGAC, for tRNA1out(Ujk) and tRNA1yr as described previously (37). Quantitation was by phosphorimaging as for mtDNA copy number.

For the preparation of mtDNA (mitochondrial nucleic acids) for analysis of Rls, mitochondria were isolated from cells essentially as described by Spelbrink et al. (38). Briefly, cells from 10 to 20, 14 cm plates were collected by pipetting in PBS, centrifugation at 400 g max for 3 min at room temperature and transfer to ice. The cell pellet was resuspended by gentle pipetting in two volumes of ice-cold 0.1x homogenization buffer (4 mM Tris–HCl, pH 7.8, 2.5 mM NaCl, 0.5 mM MgCl2), kept on ice for 5 min and homogenized in a glass homogenizer with 20 strokes of a tight-fitting pestle. Disruption of the cells was monitored by microscopy. One-ninth volume of 10x homogenization buffer was added and nuclei and cell debris were pelleted by sequential centrifugations at 1200 g max for 3 min at 4°C until no pellet was visible. Mitochondria from the post-nuclear supernatants were recovered by centrifugation at 16000 g max for 10 min at 4°C. The mitochondrial pellet was washed once in resuspension buffer (10 mM Tris–HCl pH 7.4, 0.32 M sucrose, 1 mM EDTA, 5 mM MgCl2) and re-centrifuged at 16000 g max for 10 min at 4°C. The pellet was placed immediately on ice and resuspended thoroughly in 500 μl of DNA extraction buffer (25 mM EDTA pH 8.0, 75 mM NaCl) followed by the addition of 50 μl 10% SDS with gentle mixing and incubation on ice for a further 10 min. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1, pH 8.0) was added and the tube was shaken gently overnight at 4°C on a rotatory shaker, then centrifuged at 5000 g max for 15 min at 4°C. The aqueous phase was transferred to a fresh tube and phenol extraction repeated several times until the interface was clear, after which 0.2 vol. 10 M ammonium acetate and 2 vol. 80% EtOH were added to the final aqueous phase. The solution was gently mixed, incubated on ice for 15 min and centrifuged at 5000 g max for 15 min at 4°C. The precipitated nucleic acids were washed once with 80%
EtOH and centrifuged at 5000 g_{\text{max}} for 5 min at 4°C min. The pellet was air-dried, dissolved in 80 µl of TE buffer and stored at 4°C.

**Enzymatic treatment of DNA**

MtDNA samples were treated with the following DNA-modifying enzymes under conditions recommended by the manufacturers: T7 gp3 endonuclease (New England Biolabs), topoisomerase I (New England Biolabs) and topoisomerase IV (John Innes Enterprises).

**Two-dimensional neutral agarose gel electrophoresis**

One microgram of total mitochondrial nucleic acids was used per analysis. Restriction digestions were performed following manufacturers’ recommendations, except for BclI which was carried out at 37°C for double the usual reaction time. If subsequent nuclease treatments were used, DNA was first recovered by ethanol precipitation and resuspended in the appropriate reaction buffer, before treatment with 50 U of RNaseI (New England Biolabs), 2 U of RNaseH (Promega), each for 1 h at 37°C or 50 U S1 Nuclease (Promega) for 30 s. Reactions were stopped by the addition of an equal volume of phenol–chloroform–isoamyl alcohol (25:24:1, pH 8.0) and immediately extracted. 2DNAGE was performed essentially as described in Kajander et al. (39). The first-dimension was run without EtBr in a 0.4% agarose gel in TBE buffer, 1.2 V/cm for 24 h at 4°C. The gel was stained with EtBr (300 ng/ml) in TBE. Individual lanes were cut out, rotated 90° and 1.0% agarose containing 300 ng/ml EtBr, precooled to 55°C, was cast around them. The second dimension was run at 6 V/cm for 5 h at 4°C with constant buffer recirculation. For analysis of high molecular weight fragments (e.g. 16.6 kb mtDNA linears), the first-dimension gels were 0.28% agarose, run at 1.4 V/cm for 24 h at room temperature, with the second dimension in 0.58% agarose, 300 ng/ml EtBr, run at 2.6 V/cm for 67 h at room temperature with constant buffer recirculation. Gels were processed for Southern blotting using standard procedures.

**Radiolabelled probes and blot hybridization**

For Southern hybridization, the following probes were created by Pfu-PCR, using cloned segments of human mtDNA as template and subsequently sequenced to confirm their identity: O H (np 35–611, Anderson et al., 1981), ND2 (np 4480–4988), A8-6 (np 8460–9107), ND4 (np 11161–11640) and ND5 (np 12992–13670). Probes were labelled using Rediprime II random prime labelling kit (Amersham) and [α-32P]dCTP (Amersham; 3000 Ci/mmol).

**Sub-fractionation of mitochondria**

For assaying the localization of recombinant TFAM the mitochondrial pellet, prepared as above, was resuspended in 2 vol. of lysis buffer (0.25 M sucrose, 20 mM Tris–HCl, pH 7.6, 2 mM EDTA, 7 mM β-mercaptoethanol). Mitochondria were lysed by adding 20% NP40 to a final concentration of 0.5% (v/v) with incubation on ice for 1 h. After centrifugation at 16,000 g_{\text{max}} for 10 min at 4°C the pellet and supernatant fractions were processed for SDS–PAGE.

**SDS–PAGE and western blotting**

SDS–PAGE used 7.5–12% polyacrylamide (Laemmli) gels under standard conditions. Sample preparation, western blotting and immunodetection were carried out as described previously (38). Primary antibodies used were: mouse anti-Myc monoclonal 9E10 (Roche Molecular Biochemicals), 1:15 000 dilution of a 5 mg/ml stock, rabbit anti-human TFAM (kind gift of Dr R. J. Wiesner), 1:10 000 dilution and mouse anti-α-actinin monoclonal AT6/172 (Upstate) 1:5000 of a 1 mg/ml stock. Signals were quantified using a ChemiDoc XRS chemiluminescence detection instrument and associated QuantityOne software.

**RESULTS**

**Effects on mitochondrial nucleic acids of modulating TFAM expression levels in vivo**

We manipulated TFAM expression levels in cultured human cells using transient or inducible expression of epitope-tagged or untagged TFAM, as well as TFAM knockdown by RNAi (Figure 1, see also Supplementary Figures 1–3). Transient expression (data not shown) produced essentially the same effects on all parameters studied as inducible over-expression using the Flp-In™ T-Rex system, although with less quantitative reproducibility. Inducible over-expression of TFAM carrying a C-terminal MycHis tag produced qualitatively similar but quantitatively more dramatic effects, namely a small but transient increase in mtDNA copy number, followed by progressive mtDNA depletion. After 10 days of full induction (10 ng/ml doxycyclin, full details on the use of the induction system to be published elsewhere, Wanrooij et al., manuscript in preparation), TFAM-MycHis expression resulted in the reduction of mtDNA levels to ~20% of control levels. Over-expression of TFAM with its normal stop codon (TFAM-stop) caused a decrease in mtDNA copy number of 40–60%. Copy number depletion was verified by two independent methods, real-time PCR and phosphorimaging of Southern blots (Supplementary Figures 1 and 2).

Although the absolute amount of TFAM protein did not increase markedly during induction, measured relative to a loading control (Supplementary Figures 1 and 2), the progressive drop in mtDNA copy number means that the amount of TFAM protein per mtDNA molecule did increase substantially and consistently during induction: in the case of TFAM-stop to over twice the endogenous level, in the case of TFAM-MycHis by an order of magnitude. Crude fractionation of mitochondrial protein lysates by sucrose density-gradient centrifugation (Supplementary Figure 2) showed that over-expressed TFAM-MycHis partitioned, like endogenous TFAM, mainly into the pellet fraction, forming high-molecular weight complexes in the same proportion as endogenous TFAM, indicating that it is likely complexed with mtDNA.

Previous experiments in vitro and in organello have suggested that a large excess of TFAM can suppress rather than activate transcription. We measured the steady-state levels of various mitochondrial transcripts using northern blot analysis (Supplementary Figures 1 and 2 and other
data not shown), under conditions where over-expression of TFAM-stop or TFAM-MycHis was induced in vivo. The level of short-lived mRNAs such as ND3 showed a marked decline even relative to the decreased amount of mtDNA. Although modulation of post-transcriptional processing and RNA stability contribute to changes in the steady-state level even of short-lived transcripts, the drop in ND3 mRNA levels is consistent with a substantial drop in transcription per template molecule, under conditions of TFAM over-expression, which was especially marked for TFAM carrying the C-terminal MycHis tag. As a more rigorous and direct test of its effects on transcriptional activity, we also analysed the consequences of TFAM-MycHis expression on the rate of recovery in mitochondrial tRNA levels in cells treated for 72 h with EtBr, following removal of the drug (Supplementary Figure 4).

In conformity with the published literature, RNAi knockdown of TFAM produced a progressive reduction of TFAM protein levels and mtDNA copy number. A combination of two siRNAs was selected, based on preliminary trials, which decreased TFAM protein to low levels (<10%) during 7 or 14 days continuous culture (Supplementary Figure 3). Over shorter-time periods, both TFAM protein and mtDNA depletion by RNAi were rather modest and the levels of TFAM protein or ND3 mRNA per template mtDNA molecule were almost unchanged from control cells (Figure 1).

**Altered TFAM expression leads to systematic effects on mtDNA replication intermediates**

The copy number depletion of mtDNA produced either by RNAi knockdown or by over-expression prompted us to investigate further the effects of these treatments on mtDNA replication, using 2DNAGE. Transient expression of TFAM-stop or inducible overexpression of either TFAM-stop or TFAM-MycHis, produced dramatic and systematic effects on the patterns of mtDNA RIs, which were essentially the same in all three cases (Supplementary Figure 5b). The most consistent effects on RIs were seen after 48 h of induced over-expression (or 48 h after transient transfection), when effects on mitochondrial transcripts were clearly evident and quite similar in all cases.

TFAM overexpression resulted in a substantially increased abundance of RIs relative to the unit-length restriction fragment (Figure 2; see Supplementary Figure 5 for further explanations and interpretations of the gel data). For example,
using a restriction enzyme which cuts only once in the genome (PvuII, np 2560, Figure 2b), TFAM over-expression gave a clear enhancement of the clubheaded bubble arc indicative of initiation far upstream of the restriction site, as well as revealing a prominent simple Y-arc. In uninduced cells these RIs were either much fainter or absent. The main features were the clubheaded bubble arc and corresponding double-Y arc, as well as a number of species that were sensitive to or modified by, RNaseI treatment. The RNaseI-sensitive species included circular molecules, a short arc which corresponds with dimeric circles and/or broken theta forms (T. Yasukawa, personal communication), plus a diffuse cloud of material migrating in the region between circular molecules and the bubble arc. A putative termination arc was also visible only in uninduced cells, although this was revealed more clearly when other species were modified by RNaseI (Supplementary Figure 5).

In other digests, the exact transformations of RIs resulting from TFAM overexpression varied according to which region of the genome was being analysed (Figure 2c–e). In general, ERIOLS-type RIs, i.e. slow-moving arcs and ‘clouds’ of heterogeneous, nuclease-sensitive material were diminished and migrated as more discrete entities. There was general enhancement of nuclease-resistant arcs, in particular those resistant to S1 nuclease. In addition, rather specific subtypes of RIs were seen to accumulate along the arcs, which differed from those seen in uninduced or untransfected cells. Changes were most dramatic in the rDNA region, extending to O\(_L\) (Figure 2d). In uninduced cells, in contrast to the patterns of RIs from this region of the genome seen in solid tissues or in cultured cells recovering from drug-induced mtDNA depletion (2–4), complete Y-arcs of the strand-coupled type were not detectable, even on long exposure. Instead, a heterogeneous ‘cloud’ of complex, high-molecular weight material was seen, plus the ascending portion of a Y-like arc, ending in a faintly detected replication pause site within the ND1 gene. These forms were partially sensitive to or modified by nucleases. In TFAM overexpressing cells, a complete Y-arc was easily detected, even at relatively low exposure. The descending portion of this arc was now the most prominent, although this segment was relatively sensitive to nucleases. The cloud of heterogeneous material was replaced by at least two discrete, slow-moving Y-like arcs, which were also nuclease-sensitive. The replication pause

Figure 2. 2DNAGE analysis of mtDNA replication intermediates (RIs) in cells induced to overexpress TFAM-stop. (a) Diagrammatic map of human mtDNA, showing the origins of heavy- (O\(_H\)) and light-strand (O\(_L\)) replication according to the orthodox model, relevant restriction sites and probes for the three regions of the genome analysed (approximate location of probes indicated by asterisks). NCR shown as dark grey bar, rDNA as pale grey bar. (b–e) 2DNAGE of mtDNA from uninduced cells and from cells induced to express TFAM-stop for 48 h, analysed using the restriction digests and probes indicated, with or without additional enzymatic treatments as shown. In each panel, the various arcs and other salient features are denoted as follows: Y, standard Y arcs, dY, standard doubleY arcs, c, ‘cloud’ of RNase-sensitive material, o, circular molecules, b, standard bubble arcs, s, slow-moving Y-like arcs, sensitive to various nucleases, t, termination intermediates lying on a portion of a standard X arc, p, prominent pause sites. See Supplementary Figure 5 for diagrammatic interpretations of the various arcs. Panels i and iv of part (b) and panels iii and iv of part (c) are equivalent exposures, for comparison. Other panels of uninduced cell mtDNA are 5- to 10-fold more exposed than induced cell material, in order to reveal the main features of the arcs. Note the general enhancement of RIs, relative loss of nuclease-sensitive species, of termination intermediates and of bubble arcs, following TFAM induction. The appearance of a complete or almost complete, Y arc in Figure 2b, panels iv–vi, is consistent with frequent strand breakage at O\(_H\) or with recombinational strand-switching (also generating a free end at O\(_H\)) or with frequent initiation far distant from the NCR.
in ND1 was prominently detected after S1 nuclease treatment.

The major NCR (Figure 2c), containing the principal sites of replication initiation and termination, showed more subtle changes, affecting most obviously the termination region. In uninduced cells, the most prominent species were termination intermediates lying on or close to the apex of the X-arc, plus a heterogeneous cloud of material migrating at high molecular weight which, after nuclease treatment, was either modified (RNaseI or H) or abolished (S1). The ascending portion was the most prominent portion of the partial Y arc and this was also sensitive to ribonucleases, as was the bubble arc. In TFAM overexpressing cells the bubble arc was barely visible even on long exposure and the descending portion of the partial Y arc, leading to the termination site, was strongly enhanced. However, termination intermediates lying on the X-arc were less prominent, especially after S1 nuclease treatment. Heterogeneous, high molecular weight, S1-sensitive material was less dispersed and its migration less affected by ribonucleases.

We next analysed the nature of fully double-stranded (i.e. S1 nuclease-resistant) RIs around the genome in further detail, under conditions of TFAM overexpression (Figure 3). Different restriction digests (Figure 3b) confirmed that TFAM overexpression led to an accumulation of material on the standard Y arc in the region approaching the terminus at O_H, with a corresponding loss of termination intermediates in which the fork had entered the fragment from the other end and stalled at the terminus prior to resolution.

A number of pause sites or regions were strongly enhanced, in addition to the strong pause in ND1 (Figure 3d). These included the pause at O_L (Figure 3d), several discrete sites in the regions of ND4, ND3, COXIII, A6 and A8, most of the COXII gene (Figure 3c) and a broad region of the ND5 gene (Figure 2c). Conversely, the abundance of 7S DNA was diminished by TFAM overexpression (Figure 6).

RNAi knockdown of TFAM expression produced more subtle effects on RIs (Figure 4). RIs resistant to S1 nuclease were little altered (Figure 4c), but ribonuclease-sensitive ‘clouds’ of heterogeneous, nuclelease-sensitive material were somewhat enhanced, especially in the rDNA region (Figure 4c) or showed altered mobility. The clearest transformation was seen in the origin/termination region encompassing O_H, where there was an increase in the abundance of double-Y termination intermediates lying on the X-arc, relative to the bubble and partial Y arcs (Figure 4b), an opposite result to that produced from TFAM overexpression, which diminished the abundance of termination intermediates (Figures 2 and 3).

Figure 3. 2DNAGE analysis of S1 nuclease-resistant mtDNA RIs in cells induced to overexpress TFAM-stop. (a) Diagrammatic map of human mtDNA, nomenclature as for Figure 2a. (b–d) 2DNAGE of mtDNA from uninduced cells and from cells induced to express TFAM-stop for 48 h, analysed using the restriction digests and probes indicated. All samples were treated with S1 nuclease before electrophoresis. Nomenclature as for Figure 2, plus pr, pause region (replication slow-zone). See Supplementary Figure 5 for diagrammatic interpretations. Comparable exposures are shown, to illustrate the general enhancement of S1-resistant RIs, the strengthening of pause sites and regions and the decrease in termination intermediates. Note that bubble arcs are not visible at these exposures following S1 nuclease treatment.
TFAM over-expression shows similarities with ddC treatment

Arcs of fully double-stranded, ribonuclease-insensitive RIs should be generated by true strand-coupled replication. However, they could also arise from maturation of ERIOLS type intermediates and, if so, should be enhanced where leading-strand synthesis has stalled randomly or has been drastically slowed (i.e. if the rate of lagging-strand maturation is now comparable with that of fork progression). The effects of TFAM overexpression could thus be interpreted either as a switch to strand-coupled-type replication or as a general slowing in progression of the replication fork, such that maturation of the lagging strand now occurred as fast as the fork progressed.

To address this issue we compared the effects on mtDNA RIs of TFAM over-expression with treatment of cells with the replication inhibitor dideoxycytidine, ddC (Figure 5). After conversion to ddCTP this drug produces repeated chain termination events during mtDNA replication, requiring removal of the incorporated dideoxynucleotide by exonuclease action or recombination, thus greatly slowing down the overall rate of fork progression and leading to mtDNA depletion. As shown in Figure 5, ddC treatment induced rather similar transformations in the pattern of mtDNA RIs as those brought about by TFAM overexpression: progressive and dramatic reduction in ERIOLS-type RIs, a corresponding increase in fully double-stranded RIs of the strand-coupled type (Figure 5d) and the disappearance of termination intermediates (Figure 5b, panels i–iii). Removal of the drug rapidly induced a burst of mtDNA replication but with intermediates remaining initially of the strand-coupled type and with a delayed re-appearance of termination intermediates (Figure 5b, panel vii). However, both during and following ddCTP treatment, the enhancement of RIs of the strand-coupled type was more general than that resulting from enhanced TFAM expression, rather than being concentrated in the rDNA region and at replication pause sites (e.g. compare Figure 5e, panels ii and iii).

Alterations to TFAM expression modify mtDNA topology

The effects of TFAM overexpression on mtDNA replication might reflect TFAM-induced changes in transcription or in the overall organization of mtDNA. To test whether modulation of TFAM expression affects mtDNA topology or organization, we analysed uncut mtDNA from TFAM-induced and uninduced cells, both before and after treatment with various DNA-modifying enzymes (Figure 6). In uninduced cells, most of the mtDNA migrated either as relaxed circles or in high molecular weight catenated forms that were sensitive to topoisomerase IV but not topoisomerase I. A pronounced smear of material was also visible in the high molecular weight region of the gel. Induced TFAM over-expression resulted in a pronounced shift towards monomeric supercoils, with much less catenated mtDNA, including the high molecular weight smear, as well as creating novel functional forms that were resolved by phage T7 gp3 endonuclease. Despite these differences, the residual products from combined treatment with topoisomerase IV and T7 gp3 were strikingly similar, when comparing TFAM-induced and uninduced cells. ddC treatment also resulted in a shift away from catedated forms in favour of monomeric circles, both relaxed and supercoiled (Supplementary Figure 6). TFAM knockdown by RNAi produced more subtle changes in mtDNA topology (Figure 6), with an increased level of one particular high molecular weight species (arrowed in Figure 6) and of linear molecules. The arrowed species corresponded in mobility with a catenated form which could be enhanced in control cell mtDNA by treatment with T7 gp3 endonuclease and thus might represent an abortive termination product.

Finally, we analysed the effects of various treatments on the steady-state level of 7S DNA. Both ddC treatment [(40); Supplementary Figure 6] and TFAM overexpression (Figure 6b) resulted in a substantial drop in the amount of 7S DNA relative to other forms of mtDNA, whereas TFAM knockdown resulted in a small increase in 7S DNA (Figure 6b).

DISCUSSION

In this study we investigated the effects on mtDNA replication of modulating TFAM expression in cultured human cells. Sustained over-expression, resulting in a >2-fold increase in the ratio of TFAM to mtDNA, greatly enhanced the steady-state levels of RIs of the strand-coupled type, with corresponding depletion of ribonucleotide-rich, ERIOLS-type RIs. This was accompanied by decreased mitochondrial transcription, depletion of 7S DNA and of replication-termination intermediates, reduced copy number and decatenation of mtDNA. TFAM knockdown in the same cell background also produced copy number depletion,
but with different alterations to mtDNA topology, no systematic effects on transcription and, apart from a strong enhancement of termination intermediates, virtually no change in the pattern of RIs. The findings suggest that TFAM can influence mtDNA replication and copy number in several different ways.

TFAM as a structural protein of the mitochondrial chromosome

Previous views of TFAM as a copy number regulator have assumed that the ratio of TFAM:mtDNA is invariant, so that changes in the rate of TFAM synthesis determine the amount of mtDNA present in the cell (10,14,30,31).
However, the observation of a reduced ratio of TFAM protein to mtDNA during mtDNA reamplification following EtBr treatment (32) suggests that the TFAM:mtDNA ratio is not invariant and may indeed be subject to regulation, affecting mtDNA copy number in ways distinct from a simple titration model.

The present study offers support to the latter view, by confirming that an increased ratio of TFAM protein to mtDNA can be sustained in human cells and has systematic, but opposite effects on mtDNA copy number than would be predicted by the simple titration model. Moreover, whereas most experiments hitherto conducted on TFAM have employed highly non-physiological tools, such as cells treated with EtBr, heterologous expression of a human protein in mice or of a truncated TFAM variant never seen in nature, we observed alterations in mtDNA replication and copy number resulting from increased expression of a protein identical to endogenous TFAM. The fact that we were able to achieve and sustain an increased ratio of TFAM to mtDNA contradicts the titration model, but is consistent with the results of studies using in organello footprinting (41–44), which indicate that protein binding to mtDNA is not uniform and thus that there are sites ordinarily unoccupied by protein where over-expressed TFAM can potentially bind.

The effects of TFAM over-expression on RIs were qualitatively very similar to those brought about by treatment with the drug ddC which, following conversion to ddCTP, is assumed to be a potent inducer of repeated replication stalling via premature DNA chain termination. ddC treatment also resulted in a large increase in the steady-state levels of nucleose-resistant RIs of the strand-coupled type from all around the mitochondrial genome, with concomitant loss of termination intermediates. Although the effects of these treatments were not absolutely identical, with TFAM over-expression generating a subtly different pattern of such RIs, especially in the rDNA region and in the vicinity of pause sites, their overall similarity strongly suggests that the main effect of TFAM is, like ddC, to provoke a substantial decrease in the net rate of DNA synthesis. This may reflect an increased compaction of mtDNA when the ratio of TFAM protein to mtDNA is increased, such that decondensation of the nucleoid becomes rate-limiting for fork progression. The TFAM-specific transformations of RIs may reflect different degrees of compaction and inhibition of fork progression in different regions of the mitochondrial genome, depending on their affinity for TFAM and other proteins.

The increased abundance of strand-coupled RIs may thus be due to maturation of ERIOLS-type intermediates, which we assume normally to be a slow step compared with the rate of fork progression. If the latter is slowed by increased compaction, the rate of maturation may become comparable with it. Alternatively, if the two classes of RI represent entirely different modes of DNA replication, ERIOLS type replication might be unable to use a highly compacted mtDNA template, leaving only strand-coupled replication (with a less discrete origin, as implied by Figure 2b). Initiation of bidirectional, strand-coupled replication in many bacterial plasmids is stimulated by or dependent on, DNA-bending proteins functionally related to the HMG superfamily (45). Excess TFAM may thus lead to copy number depletion by enforcing a switch to an inherently slower replication mode or simply by suppressing the alternative mode.

Exactly the same transformations of RIs were produced by transient expression of TFAM as were produced by inducible expression of either natural or C-terminally MycHis-tagged TFAM. However, the two inducible variants brought about copy number depletion with markedly different kinetics (Figure 1). Copy number depletion may thus not be due entirely to a change in the rate of fork progression, but involve also another TFAM-related process with which the C-terminal tag may interfere. Such an effect could, however, be indirect. The amount of TFAM-MycHis protein rose much more steeply than TFAM-stop during induction (Figure 1, Supplementary Figures 1 and 2), suggesting that it may escape physiological turnover mechanisms. This is supported by the observation (Supplementary Figure 2) that TFAM-MycHis almost completely replaced endogenous TFAM during induction. The fact that sustained overexpression of natural TFAM is comparatively difficult to achieve may reflect a natural homeostatic mechanism, whereby TFAM levels can modulate copy number only within certain limits.

Another possibility is that, as for other HMG proteins (46), the C-terminal tail of TFAM recruits or interacts with other nucleoid proteins. It is already implicated in interactions with specific components of the transcriptional apparatus (47). The epitope tag may therefore disturb interactions with other nucleoid proteins involved in copy number homeostasis. Copy number derangement produced in cells or mice by C-terminally truncated or heterologous TFAM variants (14,30,31) may prove to be due to loss of such regulation, rather than by the titration model inferred previously.

The fact that RNAi knockdown of TFAM results in copy number depletion without substantial changes in the patterns of RIs is further evidence that at least one other TFAM-dependent process is critical for mtDNA maintenance. One obvious possibility already mentioned is the compaction of nascent mtDNA into TFAM-containing nucleoid structures, in the absence of which the newly replicated mtDNA may simply be unstable. Other possibilities are discussed below.

TFAM as a regulator of mitochondrial transcription

Although TFAM was originally identified and named on the basis of its being essential for mtDNA transcription, previous studies have shown that excess TFAM, supplied either to a reconstituted in vitro system (9) or produced by sustained overexpression in HEK cultured cells (27), results in a paradoxical suppression of transcription. Our own observations are consistent with this (Figure 1, Supplementary Figures 1 and 2).

A drop in transcriptional activity could, conceivably, underlie the switch in DNA replication mode that favours the generation of strand-coupled RIs. Although the decrease in transcriptional activity which accompanies TFAM overexpression appears modest, according to the data of Figure 1 (Supplementary Figures 1 and 2), the steady-state level of ND3 mRNA is a relatively insensitive measure of the actual transcription rate. The mechanism by which ERIOLS-type RIs are generated remains unknown. However, since on 2DNAGE they include heterogeneous ‘clouds’ of material sensitive to both RNaseI and RNaseH which are suppressed
by TFAM overexpression (Figure 2), it is possible that they could arise by a mechanism involving either preformed RNA or nascent transcripts. Transcriptional suppression by excess TFAM, rather than over-compaction of the nucleoid, may therefore be the mechanism driving replication towards the slower, strand-coupled mode. In support of this, the segment of the genome most affected by TFAM overexpression is also the most heavily transcribed region (rDNA), in which discrete RIs are ordinarily hard to detect (Figure 2d), but became prominent when TFAM was overexpressed.

Whereas copy number depletion brought about by TFAM overexpression was accompanied by a clear drop in transcription, transcription per template molecule appeared unchanged when copy number was depleted by TFAM knockdown. Copy number regulation may therefore be independent of transcription. However, transcription was more severely affected by overexpression of epitope-tagged TFAM than natural TFAM and also provoked a more rapid drop in mtDNA copy number. The issue of a relationship between the transcriptional activity of TFAM and copy number control thus remains open.

TFAM overexpression leads to reduced levels of 7S DNA, which might reflect decreased transcriptional activity at the light-strand promoter and/or enhanced resolution of D-loop forms in vivo, another known property of TFAM [19], see also following section. If D-loops represent a precursor step in DNA replication, as proposed by the orthodox model, their depletion by TFAM overexpression may be crucial in bringing about copy number reduction or in inducing a switch to strand-coupled replication. On the other hand, the increased level of 7S DNA resulting from TFAM knockdown shows clearly that the D-loop form is not sufficient to maintain a high copy number.

TFAM as an enhancer of replication pausing

Replication pauses are well documented in both bacteria and eukaryotes, as well as in plasmids and mtDNA [48]. However, their roles in DNA homeostasis are unclear, except where they function as definitive terminators. In human mtDNA, prominent pauses occur at the so-called termination-associated site delimiting 7S DNA, at O1, and in the region immediately downstream of rDNA, within the ND1 gene (Figures 2 and 3). One possibility is that, as in yeast rDNA, these pauses are the signatures of proteins which bind at specific sites to facilitate the passage of oppositely moving replication and transcription complexes [49] or simply markers of collision sites, as in bacteria [50], at which replication finally resumes after dissociation of the transcriptional machinery [51]. Although TFAM has only a low sequence-specificity for DNA-binding, it may enhance such pauses by promoting DNA-bending, as proposed for the protein Sap1p at one of the replication pause sites in Schizosaccharomyces pombe rDNA [52]. Such bending may also facilitate the binding of other, more sequence-specific DNA-binding proteins, which directly function in fork arrest. Both copy number depletion and transcriptional inhibition following TFAM overexpression may therefore be due, at least in part, to its effect as a strong enhancer of replication pausing.

TFAM as a cofactor in junctional resolution

TFAM overexpression and TFAM knockdown produced opposite effects on the abundance of termination intermediates. Whereas these were depleted by TFAM overexpression (or dDC treatment), they were strongly enhanced by TFAM knockdown (Figures 3–5). A simple interpretation is that resolution of these forms is a late and slow step in mtDNA replication, which is in some way dependent on the supply of TFAM. Under conditions of TFAM overexpression, either that step is facilitated by increased loading of TFAM onto the nascent DNA or the rate of fork progression is slowed down so much that resolution of termination intermediates is no longer rate-limiting. Under conditions of TFAM ‘starvation’ this resolution step would conversely be inhibited and this may be the primary reason for copy number depletion following TFAM knockdown.

TFAM is already known to have binding preference for junctional structures [17], to facilitate the resolution of D-loops [19] and to bend DNA in a manner analogous with bacterial proteins such as HU and IHF. The exact molecular structure of mtDNA replication-termination intermediates is unclear, although it cannot simply comprise two oppositely moving forks that have almost met, since such a structure would be thermally unstable unless the unreplicated region between them were at least 50 bp, in which case they would not lie precisely on the X-arc on 2DNAGE gels. One possibility is that they are held together by hemicatenation after ligation of at least one strand, although such forms should be S1 nuclease-sensitive [53]. Another is that they are converted to true Holliday junctions. A third option is that they contain a single or even double chicken-foot structure.

Whatever their precise nature, a complex enzymatic machinery should be required both to form and to resolve them, in which TFAM is a plausible player, both because of its DNA-binding preferences and bending properties and by its putative ability to recruit other proteins to the DNA. DNA-bending is required for the protein-based partition systems of low copy-number bacterial plasmids such as P1 [54] and binding sites for the DNA-bending proteins H-NS and FIS appear to be clustered around the replication terminus in Escherichia coli [55]. HMG and related proteins are widely implicated as cofactors in recombination. For example, mammalian HMGB1 facilitates V(D)J recombination by the RAG proteins [56]. In bacteria, HU is essential for efficient homologous recombination [57] and IHF is required for the action of lambda integrase [58]. HMG proteins are able to facilitate recombination even in heterologous systems [see Ref. [59]]. In yeast mitochondria, overexpression of Abf2p promotes the formation of recombination intermediates [60]. DNA-bending by TFAM may facilitate analogous processes in human mtDNA.

A rather different interpretation would also be consistent with the 2DNAGE data, namely that TFAM overexpression leads to a high frequency of strand breakage at O1, which would account both for the relative paucity of bubble arcs in O1-containing fragments (Figure 2b and c), for the corresponding increase in simple Y arcs and for the loss of termination intermediates (Figures 2e, and 3b). Such strand-breakage is very unlikely to be an extraction artefact, since these differences from control cells were seen in all of the
many DNA preparations analysed. If this interpretation is correct, it may imply a need for true recombination reaction to resolve daughter molecules, which could be much slower than the usual termination/resolution step. Copy number depletion might therefore be due to interference with termination when TFAM is either overexpressed or downregulated.

In addition, TFAM overexpression resulted in apparent decatenation of mtDNA, whereas TFAM knockdown increased the level of linear molecules and of one decatenated species that could also be generated in control cell mtDNA by treatment with T7 gp3 endonuclease, a promiscuous junctional resolvase. It is tempting to suggest that these species may be aberrant breakdown products resulting from failure of the normal termination step under conditions of TFAM deficiency. The loss of catenated species in favour of the linear forms may be due to interference with the usual termination/resolution step. Copy number reductions in these experiments were not as pronounced as in the experiments in which TFAM expression was decreased, but mtDNA was analyzed after a longer period of time in cell culture, suggesting that the loss of catenated species is a consequence of decreased TFAM expression and not a result of DNA replication. The loss of catenated species in favour of linear forms by treatment with T7 gp3 endonuclease, a promiscuous junctional resolvase, is consistent with the idea that decatenation is an early event in mtDNA replication, whereas TFAM knockdown increa-

 Physiological role of TFAM as a mtDNA copy number regulator

Our findings indicate that, at least in this particular proliferating cell background, TFAM expression is finely poised. Any marked deviations from what appears to be an optimal expression level provoke alterations to mtDNA replication that result in decreased mtDNA levels.

Previous studies of the transcriptional regulation of the TFAM gene indicate that it is sensitive to signals connected with metabolite supply (63) and redox stress, via the phosphorylation of NRF-1 (64) and to proliferative, differentiation-linked and environmental signals via the co-activators PGC1 and PRC (65–67) and the transcription factor Myc (68). This raises the issue of whether there is an additional homeostatic, mechanism to fine-tune TFAM expression to the physical state of mtDNA, which can be circumvented by the manipulations we carried out. A retrograde signalling pathway of this type might serve to ensure that disturbances in mtDNA replication or segregation do not lead to a copy number catastrophe. Alternatively, the level of TFAM expression and possibly also its post-translational modification (69), may function in vivo to accelerate or decelerate mtDNA replication, thus constituting a key determinant of copy number.

ACKNOWLEDGEMENTS

We thank Outi Kurronen and Merja Jokela for technical assistance, Takehiro Yasukawa, Aurelio Reyes, Gertjan Hakkaart, Anu Suomalainen (Wartiovaara) and colleagues for constant advice and critical input and Rudi Wiesner for supplying the anti-TFAM antibody. For financial support of our research we are grateful to the Academy of Finland, Juselius Foundation, Tampere University Hospital Medical Research Fund, the University of Tampere, the EU (EUMITOCOMBAT and MitEURO projects) and the UK Medical Research Council. Funding to pay the Open Access publication charges for this article was provided by the foregoing organizations.

Conflict of interest statement. None declared.

REFERENCES


Alterations to the expression level of mitochondrial transcription factor A, TFAM, modify the mode of mitochondrial DNA replication in cultured human cells

Jaakko L O Pohjoismäki¹, Sjoerd Wanrooij¹, Anne K Hyvärinen¹,
Steffi Goffart¹, Ian J Holt², Johannes N Spelbrink¹, & Howard T Jacobs¹,³

¹Institute of Medical Technology & Tampere University Hospital, FI-33014 University of Tampere, Finland;
²MRC-Dunn Human Nutrition Unit, Hill Road, Cambridge CB2 2XY, England, UK; ³IBLS Division of Molecular Genetics, University of Glasgow, Glasgow G12 8QQ, Scotland, UK.

SUPPLEMENTARY DATA
Legends to Supplementary Figures

Supplementary Figure 1
Effects of induced expression of TFAM-stop. Mitochondrial proteins, DNA and RNA were analysed from Flp-In™ T-Rex™ -293 cells stably transfected with the TFAM-stop construct, induced over the times indicated. In each case, error bars indicate means ± SEs from at least three independent experiments. (a) Western blot probed for TFAM protein, stripped and reprobed for α-actinin as loading control. Bar chart shows TFAM chemiluminescence signals normalized first against those of α-actinin, then against that the value for uninduced cells (t=0). (b) Southern blot probed for nuclear and mitochondrial DNA, using ND4 and 18S rDNA probes, respectively. Bar chart shows the ratio of hybridization signals, normalized against the mean value for uninduced cells. (c) Q-PCR analysis of mtDNA copy number on the same samples, again normalized to the mean value for uninduced cells. Note that this assay measures all mtDNA, not just the full-length mtDNA detected in Southern blots, i.e. it includes also replication intermediates. Conversely, mtDNA which is amplified poorly, e.g. due to ribosubstitution, may be under-quantified by this method. Southern blotting, on the other hand, measures only mature mtDNA, but is potentially subject to artefacts arising from minor variations in sample quality. The two methods in fact gave slightly different results. The transient increase in copy number in the first hours of TFAM over-expression seen by Q-PCR, may, indeed, be due to an increased level of replication intermediates. To minimize any artefacts introduced by one or other method, the mean of the two values was used to calculate the data plotted in Fig. 1. (d) Northern blot probed for mitochondrial ND3 mRNA, stripped and reprobed for 5S rRNA. Bar chart shows the ratio of hybridization signals, normalized against the mean value for uninduced cells.

Supplementary Figure 2
Effects of induced expression of TFAM-MycHis. Mitochondrial proteins, DNA and RNA were analysed from Flp-In™ T-Rex™ -293 cells stably transfected with the TFAM-MycHis construct, induced over the times indicated. In each case, error bars indicate means ± SEs from at least three independent experiments. (a) Western blot probed for TFAM protein, stripped and reprobed for α-actinin as loading control. The identity of the band denoted as the TFAM-MycHis protein (TFAM-mh) was confirmed by probing with anti-
Myc monoclonal antibody. Bar chart shows total TFAM chemiluminescence signals (i.e. endogenous TFAM plus transgenic TFAM-MycHis) normalized first against those of α-actinin, then against the value for uninduced cells (t=0). Note that endogenous TFAM declines as overexpressed TFAM-MycHis increases, indicating that TFAM levels are also regulated translationally or post-translationally, most likely at the level of incorporation into protein-DNA complexes. (b) Southern blot probed for nuclear and mitochondrial DNA, using ND4 and 18S rDNA probes, respectively. Bar chart shows the ratio of hybridization signals, normalized to the mean value for uninduced cells. (c) Q-PCR analysis of mtDNA copy number on the same samples, again normalized to the mean value for uninduced cells. See legend to Supplementary Fig. 1 for discussion of the issues relating to the two methods used to estimate mtDNA copy number changes. (d) Northern blot probed for mitochondrial ND3 mRNA, stripped and reprobed for 5S rRNA. Bar chart shows the ratio of hybridization signals, normalized against the mean value for uninduced cells. (e) Western blot of mitochondrial protein extracts from TFAM-MycHis after 24 h of induction, sub-fractionated into pellet (Pel) and supernatant (Sup) fractions, and probed for TFAM. The adjacent tracks are equally loaded in regard to the starting material. Based on chemiluminescence, approximately 90% of both the endogenous and the transgenic TFAM fractionate in the pellet, along with mtDNA.

Supplementary Figure 3

Effects of RNAi knockdown of TFAM expression. Mitochondrial proteins, DNA and RNA were analysed from HEK293T cells, following transfection with siRNAs Si2 and Si5, at the times indicated. In each case, error bars indicate means ± SEs from at least three independent experiments. (a) Western blot probed for TFAM protein, stripped and reprobed for α-actinin as loading control. Bar chart shows TFAM chemiluminescence signals normalized first against those of α-actinin, then against the value for untreated cells (t=0). (b) Q-PCR analysis of mtDNA copy number, normalized to the mean value for untreated cells. (c) Northern blot probed for mitochondrial ND3 mRNA, stripped and reprobed for 5S rRNA. Bar chart shows the ratio of hybridization signals, normalized against the mean value for untreated cells. (d) Western blot of mitochondrial protein extracts from cells treated for 5 d in trial experiments with various combinations of putative TFAM-specific siRNA oligonucleotides. The sequences of the relevant mRNA targets were as follows. Si1: 5´-AAAGAAACCTGTAAGTTCTTA -3´ (np 282-302 of the TFAM mRNA
sequence), Si3: 5´-AAAGAAAAAAATATATCAAGA-3´ (np 414-434), Si5 5´-AAAGAAGAGATAAGCAGATTT-3´ (np 463-483), Si6: 5´-AAGCAGATTTAAAGAACAGCT-3´ (np 474-494). The final lane shows protein from cells treated with an siRNA directed against mTERF mRNA, as a negative control.

**Supplementary Figure 4**

Suppression of *de novo* mtDNA transcription in cells induced to express TFAM-MycHis. Cells were cultured in EtBr-containing medium for 72 h, with or without concomitant induction of TFAM-MycHis expression by doxycyclin, followed by replating in fresh medium containing no EtBr. Mitochondrial tRNA levels at different time-points were measured by Northern hybridization and phosphorimaging, with normalization first to 5S rRNA as a loading control, and then to the starting level of the relevant tRNA in uninduced cells on the first day of measurement (24 h after the start of the experiment). Over 5 days of induction, TFAM-MycHis expression reduces mtDNA copy number only by 30-50% (Supplementary fig. 2), therefore the almost complete suppression of *de novo* transcription revealed by this experiment is not due to the absence of template.

**Supplementary Figure 5**

Explanations and supplementary data for 2DNAGE analyses. (a) Various panels from Figs. 2, 3 and 4 are reproduced, alongside interpretative illustrations of the various classes of replication intermediate assumed to be represented by the arcs indicated. Restriction sites remaining uncut due to ribosubstitution on the lagging strand are indicated by red bars. The clouds of RNAse-sensitive material associated with slow-moving arcs are suggested to have attached tails of nascent RNA, as shown in red. Replication pause regions are shown as filled circles: the one at O_k in yellow, a second one within the ND1 gene in blue. The short ‘eyebrow’ arc, migrating in the region of the termination intermediates from Fig. 2b, panel ii, is proposed to derive from broken theta molecules forms as shown, although it may, alternatively, comprise dimeric circles. A comprehensive explanation of 2DNAGE methodology and gel interpretation is given in Refs. 70 and 71. (b) Comparison of 2DNAGE analyses of the ND2 region (AccI digest, ND2 probe, see Fig. 2a) in cells overexpressing TFAM-stop or TFAM-MycHis (TFAM-mh) by induction in stably transfected Flp-In™ T-
Rex™-293 cells, or TFAM-stop by transient transfection of HEK293T cells, with or without additional enzymatic treatments as shown. Panels v and vi are scaled versions of panels i and ii of Fig. 3d. The patterns of mtDNA RIs are virtually indistinguishable in the three cases of TFAM overexpression, but quite different from those seen in uninduced cells.

**Supplementary Figure 6**

Effects of ddC treatment on mtDNA topology. DNA samples were analysed as in Fig. 6. ddC treatment was for the times indicated (h), 72 + 2 meaning 72 h of treatment followed by 2 h of recovery in fresh medium etc. The panel showing 7S DNA is a longer exposure of the bottom of the same gel blot.
Additional References for Supplementary Figures


Supplementary Figure 1: Pohjoismäki et al
Supplementary Figure 2: Pohjoismäki et al.
Supplementary Figure 3: Pohjoismäki et al
Supplementary Figure 4: Pohjoismäki et al

![Graph showing tRNA levels over time](image)

- tRNA\textsuperscript{Leu(UUR)}
- tRNA\textsuperscript{Tyr}

- **●** induced
- **△** not induced

**tRNA levels (a.u.)**

**Time of EtBr treatment + recovery (h):**

- 0
- 24
- 48
- 72
- 72 +24
- 72 +48
Supplementary Figure 5: Pohjoismäki et al.

(a) Figure 2b, panel ii

(b) Induced TFAM-mh, TFAM-stop

Fig. 3b, panel iii

Fig. 5b, panel vi

Fig. 2c, panel i

Fig. 3d, panels iv, vi

Fig. 2d, panel viii

Fig. 4c, panel ii

Fig. 3b, panel ii

Fig. 2c, panel i

Fig. 3d, panels iv, vi

Fig. 2d, panel viii

Fig. 5b, panel vii

Fig. 4c, panel ii

Fig. 3d, panels iv, vi

Fig. 2d, panel viii

Fig. 5b, panel vii

Untreated

+RNase I

+S1

Induced TFAM-mh

Induced TFAM-stop

+RNase I

+S1

Acl, ND2, 4.8 kb
Supplementary Figure 6: Pohjoismäki et al

ddC treatment

U 24 48 72 2 24 48

catenanes etc
relaxed circles
linears
monomeric supercoils

7S DNA
Expression of catalytic mutants of the mtDNA helicase Twinkle and polymerase POLG causes distinct replication stalling phenotypes

Sjoerd Wanrooij¹, Steffi Goffart¹, Jaakko L.O. Pohjoismäki¹, Takehiro Yasukawa² and Johannes N. Spelbrink¹,*

¹Institute of Medical Technology and Tampere University Hospital, Tampere, Finland and ²MRC-Dunn Human Nutrition Unit, Wellcome Trust-MRC Building, Cambridge, UK

ABSTRACT

The mechanism of mitochondrial DNA replication is a subject of intense debate. One model proposes a strand-asynchronous replication in which both strands of the circular genome are replicated semi-independently while the other model proposes both a bidirectional coupled leading- and lagging-strand synthesis mode and a unidirectional mode in which the lagging-strand is initially laid-down as RNA by an unknown mechanism (RITOLS mode). Both the strand-asynchronous and RITOLS model have in common a delayed synthesis of the DNA-lagging strand. Mitochondrial DNA is replicated by a limited set of proteins including DNA polymerase gamma (POLG) and the helicase Twinkle. Here, we report the effects of expression of various catalytically deficient mutants of POLG1 and Twinkle in human cell culture. Both groups of mutants reduced mitochondrial DNA copy number by severe replication stalling. However, the analysis showed that while induction of POLG1 mutants still displayed delayed lagging-strand synthesis, Twinkle-induced stalling resulted in maturated, essentially fully double-stranded DNA intermediates. In the latter case, limited inhibition of POLG with dideoxycytidine restored the delay between leading- and lagging-strand synthesis. The observed cause-effect relationship suggests that Twinkle-induced stalling increases lagging-strand initiation events and/or maturation mimicking conventional strand-coupled replication.

INTRODUCTION

Human mitochondrial DNA (mtDNA) is a closed circular molecule of ~16.5 kb and was sequenced 25 years ago (1,2). The two strands of mtDNA are denoted as the Heavy-(H)-strand and the Light-(L)-strand on the basis of their mobility in a denaturing caesium chloride gradient. The strand-asynchronous or strand-displacement model for mammalian mitochondrial DNA replication was first proposed in the early 70s (3). In this model, synthesis of the nascent H-strand starts at a fixed point in the major non-coding region (NCR) of mtDNA denoted O_{H}. O_{H} was originally defined by mapping the 5' ends of the so called D-loop and is located on the L-strand upstream of three conserved sequence blocks. Leading-strand (nascent H-strand) synthesis proceeds two-thirds of the way around the molecule, displacing the parental H-strand in the process with mitochondrial single-stranded DNA-binding protein (mtSSB) suggested to provide protection against the action of nucleases and other insults such as reactive oxygen species. Following exposure of the lagging-strand initiation site (O_{L}) synthesis of the nascent L-strand begins (4,5).

More recently, Holt and co-workers proposed two models of mtDNA replication, one a more conventional strand-synchronous theta mode (6–9) where mtDNA replication initiates bidirectionally at various sites across an initiation zone (OriZ). In this case termination occurs at or near O_{H}. The other mode of replication is similar to the strand-asynchronous mode of replication so that the nascent L-strand DNA was suggested also to be synthesized with a considerable delay. Initiation is essentially unidirectional and occurs in the NCR, importantly however RNA is deposited on the displaced H-strand rather than mtSSB, thus forming ribonucleotide incorporation throughout the lagging strand (RITOLS) intermediates, which is a crucial difference from the strand-asynchronous model (10). Although the high levels

*To whom correspondence should be addressed: Tel: +358 3 35518598; Fax: +358 3 35517710; Email: hans.spelbrink@uta.fi

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

© 2007 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
of mtSSB (11) could be seen as supporting the strand displacement model, also for example Escherichia coli is estimated to have several thousands of molecules of SSB (12) even though it contains a single copy genome and replicates via conventional theta replication. SSB is nevertheless essential, as it would be in mammalian mitochondria, not only at the replication fork but also in repair, recombination and other DNA maintenance processes. Given the various essential functions of SSB, the high levels might simply reflect a cell’s precaution to ensure it is readily available.

The RITOLS model requires that the RNA is replaced by DNA to produce a dsDNA lagging-strand. It was shown that the RITOLS replication intermediates (RIs) are prone to RNaseH degradation during mtDNA purification leaving a single-stranded parental H-strand (7), thus producing RIs originally predicted by the strand-asynchronous model. Strand-asynchronous RIs are therefore considered purification/degradation artefacts. In rodent and chick liver and cultured human cells under normal culture conditions RITOLS intermediates are the predominant class (6,9,10). However, in cultured human cells recovering from mtDNA depletion, the majority of the replication intermediates are essentially double-stranded DNA suggesting a switch from the RITOLS replication mode to more conventional theta replication (9). Alternatively, initiation of lagging-strand DNA synthesis occurs more frequently resulting in an increased rate of conversion of RITOLS RIs to dsDNA RIs.

All proteins responsible for mammalian mtDNA maintenance are encoded in the nucleus, translated by cytosolic ribosomes and imported into the mitochondrial compartment. So far, a limited number of proteins has been identified. These include the mitochondrial DNA polymerase gamma (POLG1) and its accessory subunit (POLG2) [see, e.g. (13)], the mitochondrial DNA helicase Twinkle (14,15), mitochondrial single-stranded DNA-binding protein (mtSSB) (16) and various proteins with a more general role in mtDNA maintenance. The POLG holoenzyme, Twinkle and mtSSB can form a minimal mitochondrial replisome capable of genome length DNA synthesis on an artificial template (17). Some of the components of the mitochondrial replisome and transcription machinery show similarity to their counterparts in T-odd bacteriophages suggesting that a T-odd phage ancestor contributed to the early ‘mitochondrial’ endosymbiosis event (18). For example, Twinkle shows striking similarity to the T7 phage primase/helicase protein gp4 (T7 gp4) (14). The Metazoan primase domain of Twinkle has diverged from the ones of more primitive Eukaryotes and T-odd phages suggesting it has lost its primase function (19).

The genes for Twinkle, POLG1 and more recently POLG2 (20) have been implicated in human diseases. Autosomal dominant (ad) mutations in Twinkle are associated with Progressive External Ophthalmoplegia (adPEO) (14), while a single recessive mutation is associated with infantile onset spinocerebellar ataxia or IOSCA (21). Mutations in POLG1 are associated with a variety of disorders, including dominant and recessive PEO, various types of ataxia, Parkinsonism and the severe mtDNA depletion Alpers syndrome (see, e.g. (22) and references therein, and http://dir-apps.niehs.nih.gov/polg/).

The catalytic subunit of polymerase gamma, POLG1, is well-characterized biochemically (see (13), bearing similarity with prokaryotic A-type DNA polymerases such as E. coli DNA polymerase I and T7 DNA polymerase. Conserved regions include a C-terminal domain responsible for polymerase activity and an N-terminal 3'-S' exonuclease domain involved in proofreading. Several disease associated POLG1 mutations have been studied using purified recombinant enzyme. These include the common autosomal dominant Y955C and other mutations, which result in a moderate to severe decrease in polymerase activity, reduced nucleotide selectivity or reduced processivity (23).

In vivo, the properties of POLG1 have also been partly characterized in yeast and in cultured human cells (24–27). In both cases, expression of a mutant form of the protein deficient in 3'-S' exonuclease activity results in the accumulation of mtDNA mutations. An exonuclease deficient variant in mouse also results in a mutator phenotype and shows a whole-organism phenotype of reduced lifespan with a variety of tissue specific ageing associated defects (28,29). Expression of an adPEO associated Twinkle mutation in transgenic mice has shown a late onset phenotype with striking similarities to late onset PEO (30).

Although there is a need for further biochemical characterization of POLG1 and Twinkle mutants, and animal models can provide a wealth of information on disease aetiology and pathogenesis, both approaches have their limitations. We therefore chose an alternative approach of inducible expression of wild-type and mutant POLG1 and Twinkle in cultured human cells, allowing us to study protein function and mtDNA replication dynamics in vivo. Using this inducible system in combination with two-dimensional neutral/neutral agarose gel-electrophoresis (2DNAGE), we show here that the induced expression of either Twinkle or POLG1 mutants results in distinct replication stalling phenotypes suggesting defined roles for both proteins in mtDNA replication and in particular the frequency of initiation of lagging-strand maturation/synthesis.

MATERIALS AND METHODS

Cloning of expression constructs

The full-length cDNA of POLG1 and Twinkle variants were originally cloned in the pcDNA3.1(−)/Myc-His A (Invitrogen, Carlsbad, CA, USA), as previously described (14,27). All constructs were re-cloned in the pcDNA5/FrT/TO vector (Invitrogen) taking advantage of two PmeI restriction sites flanking the multiple cloning sites of the original pcDNA3 vectors and target vector. The resulting fusion proteins contained the sequence of the respective proteins followed by the Myc-His. All resulting plasmid constructs were confirmed by DNA sequencing.
Creation and maintenance of stable transfected inducible cell lines

The Flp-In™ T-REx™ 293 host cell-line (Invitrogen), a HEK293 variant containing a Flip recombination site at a transcriptionally active locus, was grown in DMEM medium (Cambrex Bioscience, Walkersville, MD, USA) with 2 mM L-glutamine (Cambrex Bioscience), 10% FCS (Euroclone, Milan, Italy) and 50 µg/ml uridine (Sigma, St. Louis, MO, USA) supplemented with 100 µg/ml Zeocin (Invivogen) and 15 µg/ml Blasticidin (Invivogen) in a 37°C incubator at 8.5% CO₂. Two-day prior to transfection cells were split to 10 cm plates and grown to ~80% confluence in medium lacking antibiotics. Cells were co-transfected with TransFectin (Bio-Rad, Hercules, CA, USA) according the manufacturer’s protocol with the appropriate pcDNA5/Frt/TO construct (0.4µg) and pOG44 (Invitrogen; 3.6µg), a plasmid encoding the Flp-recombinase necessary for targeted stable integration. Six hour later, transfection medium was replaced with regular fresh medium lacking antibiotics. Twenty-four hours after transfection the selective antibiotics Hygromycin (150 µg/ml) (Invivogen) and Blasticidin (15 µg/ml) were added. Selective medium was replaced every 2 days for cell maintenance. All inducible cell lines were created according this method. To induce expression the indicated medium was added. With longer than 2 days induction medium was refreshed every 2 days.

Western blot analysis

Cell lysates were prepared and analyzed for protein expression by immunoblotting after SDS-PAGE (27). A primary monoclonal c-myc (Roche Molecular Biochemicals, Nutley, NJ, USA) antibody was used for detection of recombinant proteins. Peroxidase-coupled secondary antibody horse-anti-mouse was obtained from Vector Laboratories, Burlingame, KS, USA. Enhanced Chemiluminescence detection was done essentially as described (27).

Quantitative PCR

The copy number of mitochondrial DNA per cell was determined by real time PCR of cytochrome b using the gene for amyloid precursor protein APP as a nuclear standard as described (31). Briefly, crude nucleic acid extracts were obtained from cells by lysis, proteinase K digest and subsequent isopropanol precipitation, and copy numbers of cytochrome b and APP were determined in a duplex Taqman PCR on an Abiprism 7000 (Applied Biosciences, Foster City, CA, USA) using pCR 2.1-TOPO (Invitrogen) vectors containing the cytochrome b and APP ampiclon as standards.

Immunocytochemistry

Immunofluorescent detection was done essentially as described previously (32). For the detection of mtDNA we used a monoclonal anti-DNA antibody AC-30-10 (PROGEN, Shingle Springs, CA, USA) as described previously (33). Secondary antibodies were anti-mouse IgG-Alexa Fluor®488 (Invitrogen; Myc) and anti-mouse IgM-Alexa Fluor®568 (DNA). Image acquisition using confocal microscopy was done as described (32).

Protein isolation and helicase assays

In vitro assays for determination of helicase activities were performed with highly enriched Twinkle preparations derived from 293 Flp-In™ T-REx™ cells. The cells were induced with 50 ng/ml doxycycline (Sigma) for 2 days, harvested and mitochondria isolated by hypotonic lysis and differential centrifugation (32). The mitochondrial pellet obtained was lysed in high salt buffer (50 mM KH₂PO₄ pH 7.0, 1 M NaCl, 1% Triton X-100, 1 x complete Protease inhibitors EDTA-free, Roche) and sonicated on ice (Sonics Vibra-cell, 1 min 40% amplitude, 1 s pulses with 2 s break). The insoluble DNA fraction was pelletted for 10 min at 12,000 g and 4°C. Supernatant was incubated with Talon metal-affinity resin (Clontech, Palo Alto, CA, USA) for 1–2 h at 4°C to allow binding of Histagged proteins. Resin was washed twice with high salt buffer and twice with low salt buffer (25 mM Tris–HCl pH 7.6, 40 mM NaCl, 4.5 mM MgCl₂, 10% glycerol, 100 mM L-Arginine) containing 20 mM Imidazole. Elution was carried out with low salt buffer containing 500 mM Imidazole. The supernatant of this step was shock-frozen in liquid nitrogen and stored at −80°C.

As standard substrate for helicase assays a radioactively end-labeled 60 nt oligonucleotide hybridized to M13 ssDNA was used (5’ACATGATAAGATACATGGATG AGTTTTGACAAAAACACTGTAACACGAGGCC AGTGCC 3’), forming a 20 nt double-stranded stretch with a 40nts 5’ overhang.

The assay was performed by incubating 1 ng Twinkle protein in 40 µl helicase buffer (25 mM Tris pH 7.6, 40 mM NaCl, 4.5 mM MgCl₂, 100 mM L-Arginine-HCl pH 7.6, 10% glycerol, 3 mM UTP, 1 mM DTT, 5 µM unspecific oligonucleotide) with 2 amol substrate for 30 min at 37°C. The reaction was stopped by adding 10 µl loading buffer (90 mM EDTA, 6% SDS, 30% glycerol, 0.25% bromophenol blue). Twenty microliter reaction mixes were separated on a 15% acrylamide gel in 1 x TBE, dried on a vacuum gel drier and exposed to X-ray film or quantified by phosphoimager.

Mutation sequencing

Point mutation levels in the NCR and cytochrome b region of mtDNA of POLG1 cell lines were measured as previously described (34).

Brewer-Fangman 2D neutral/neutral Agarose electrophoresis

Mitochondrial nucleic acids were extracted using cytochala-sine (Sigma-Aldrich) as described (9). Purified mtDNA was digested with HincII and where mentioned further treated with RNase H or S1 nuclease (Fermentas, Hanover, MD, USA) with the indicated amounts and time. The fragments were separated by 2DNAGE as described (35,36) and the gels were blotted and hybridized with a 32P-labeled DNA probe for human mtDNA nts 14846–15357 (9).
Table 1. Expresssed proteins with their predicted sizes

<table>
<thead>
<tr>
<th>Variant</th>
<th>Size (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twinkle&lt;sup&gt;a&lt;/sup&gt; wt</td>
<td>74</td>
</tr>
<tr>
<td>K421A, G575D</td>
<td></td>
</tr>
<tr>
<td>Twinkle&lt;sup&gt;a&lt;/sup&gt; ΔAA346–376 (Δ346–376)</td>
<td>71</td>
</tr>
<tr>
<td>Twinky&lt;sup&gt;b&lt;/sup&gt;</td>
<td>64</td>
</tr>
<tr>
<td>Twinkle&lt;sup&gt;a&lt;/sup&gt; ΔAA70–343 (Δ70–343)</td>
<td>41</td>
</tr>
<tr>
<td>POLG1 wt</td>
<td>142</td>
</tr>
<tr>
<td>D198A, D890N, D1135A</td>
<td></td>
</tr>
<tr>
<td>POLG1 ΔCAG</td>
<td>141</td>
</tr>
</tbody>
</table>

All expressed proteins had the MycHis epitope tag and this tag was included in the calculation of size in kiloDaltons (kDa).

<sup>a</sup>Size based on predicted processing at AA42.

RESULTS

Establishing stable inducible cell lines expressing mtDNA replication factors

In order to study the mtDNA maintenance machinery in cultured human cells, inducible cell lines expressing wild-type and mutant variants of Twinkle, the Twinkle splice variant Twinky (14) or POLG1 were established using HEK293 Flp-In™ T-REx™ cells (see Table 1 for a list of all variants). Twinkle mutants included a lysine mutation (K421A) in the highly conserved Walker A motif implicated in ATP binding and hydrolysis; a mutation (G575D) in helicase motif H4 and implicated in DNA binding (37); a deletion of 31 amino acids (Δ346–376) of the region that shows similarity with the T7 gp4 linker region that has been implicated in multimer formation [see e.g. (38)]; a large deletion (Δ70–343) in the region of the protein that shows similarity with the T7 gp4 primase domain (19). POLG1 mutants included two polymerase deficient mutants (D890N and D1135A), one exonuclease deficient mutant D198A and a non-deleterious deletion mutant (ΔCAG) of 10 consecutive glutamines in the N-terminus, all as previously described (27). All cell lines and >99% of all cells expressed the recombinant proteins upon doxycycline (DC) induction and expressed proteins were all targeted to the mitochondrial compartment (Figure 3 and not shown).

Since DC is a mitochondrial protein synthesis inhibitor at μg/ml concentrations, we first determined the lowest possible levels of DC to achieve full induction. Figure 1A shows an increase of protein expression of wild-type (wt) Twinkle in cells with increasing DC concentrations (0–1000 ng/ml). Both after one or three days, maximum induction levels were reached at low ng/ml concentrations, but at slightly lower concentrations after three days induction.

Figure 1B shows the expression of all proteins used in this study confirmed by immunoblotting, following 72 h of treatment with 0, 3 and 10 or 20 ng/ml DC. All proteins were detected using their respective epitope tag and gave bands of the expected size (Table 1) upon induction. In the absence of DC we could detect leaky expression of most Twinkle variants, but only when films were overexposed considerably (not shown).

However, immunofluorescence in the absence of DC induction did not result in any mitochondrial signal above background fluorescence suggesting the expression levels were very low (see also below and Figure 3A). Some of the analyzed Twinkle variants, such as Twinky, showed reproducibly lower protein levels with full induction of expression, indicating differences in protein (or mRNA)
stability of these variants. Similarly, the POLG1 D1135A mutant showed lower expression levels than the other POLG1 variants, suggesting that the mutant protein is less stable.

As a final test for the inducible expression system, we created a cell-line expressing a POLG1 variant (D198A) in which exonuclease activity is abolished. We have shown previously that constitutive expression of D198A in cultured human cells results in the accumulation of point mutations in mtDNA (27). To validate the obtained inducible cell-line we determined the point mutation levels in two regions of mtDNA (Figure 1C). After 60 days of induction, both the cytochrome b and control region showed elevated mutation levels in the D198A cell-line while non-induced D198A cells showed low mutation levels similar to cells expressing POLG1 wt.

**Expression of several Twinkle and POLG1 mutants results in mtDNA copy-number depletion**

The relative mtDNA copy number in the various inducible cell lines was compared by quantitative real-time PCR (QPCR) using the nuclear amyloid precursor protein (APP) gene as a standard (31) (Figure 2). The absolute copy number determined by us for the HEK293 Flp-InT<sup>TM</sup> T-REx<sup>TM</sup> and the majority of non-induced transgenic cells was ~3000 copies/cell (2798 ± 450 (n = 4) for the non-induced parental cell line). Induced overexpression of POLG1 wt or Twinkle wt did not significantly change mtDNA copy number per cell, indicating that abundance of these proteins is not rate-determining for mtDNA replication at least in cell culture. More importantly it also showed that overexpression per se did not otherwise interfere with mtDNA replication. Similarly, overexpression of Twinky, Twinkle Δ346–376, POLG1-D198A and POLG1-ΔACAG did not influence steady-state mtDNA levels. In contrast even low-level expression of Twinkle mutants K421A or G575D and POLG1-D890N or D1135A lead to a dramatic decrease of mtDNA levels within a few days. The Twinkle K421A and G575D cell lines showed a significant steady-state reduction in mtDNA copy number of ~60% even prior to induction, presumably caused by the slightly leaky expression of the Twinkle variants. This suggests these mutants are strongly dominant in nature. In contrast, the POLG1 D890N and D1135A mutants did not show a significant copy-number reduction without induction (not shown). Notwithstanding the minor leakiness, depletion upon induction was dose-dependent, as higher expression levels lead to a faster depletion than low-level expression (data not shown). For the Twinkle K421A and G575D as well as the POLG1 D890N and D1135A mutants, the mtDNA levels after three days of full DC induction were ~20–30% compared to non-induced cells, indicating a complete abolishment of successful replication and dilution of mtDNA by cell division.

![Figure 2](image-url)

**Figure 2.** POLG1 and Twinkle mutants can cause mitochondrial DNA copy-number depletion. MtDNA copy number was determined using duplex Taqman PCR. Shown are the copy numbers for each cell line relative to their copy number in non-induced cells following either 3 ng/ml or full induction (10 ng/ml for Twinkle variants; 20 ng/ml for POLG1 variants). Note that although Twinkle K421A and G575D cell lines had a reduced steady-state copy number in non-induced cells, induction resulted in a further strong drop in copy number (see also main text). The same general trends as observed here were obtained by Southern analysis (not shown). Copy number decline was highly significant in Twinkle K421A, G575D and Δ70-343 lines as well as in the POLG1 D890N and D1135A mutant lines.
Several Twinkle mutants show altered nucleoid localization

The localization of Twinkle variants was analyzed by immunofluorescence using the myc-tag of the recombinant proteins (Figure 3A). Overexpressed Twinkle wt showed the typical punctuated pattern within mitochondria, indicating the normal localization in mtDNA nucleoids. Twinky and Δ346–376 both showed diffuse mitochondrial staining. The N-terminal deletion variant Δ70–343 showed almost normal punctuate nucleoid-like localization. The variant G575D could be detected in punctate foci, but in addition showed enhanced diffuse staining in mitochondria. The K421A variant having a mutation in the WalkerA motif was detected in few enlarged spots, indicating either abnormal nucleoid segregation or protein aggregation. To differentiate between these possibilities we used an anti-DNA antibody to see to what extent this and other Twinkle mutants co-localized with mtDNA. As previously shown (32), Twinkle wt showed excellent co-localization with DNA as did Δ70–343 (not shown). The K421A variant, despite its abnormal appearance of enlarged foci, did co-localize with mtDNA (Figure 3B). The number of DNA foci, however, was severely reduced compared to non-expressing or Twinkle wt expressing cells. For Twinkle G575D, the number of nucleoids was again severely reduced (Figure 3B) but most of the protein-foci co-localized with mtDNA. Both Twinky and the Δ346–376 variant showed normal nucleoid numbers judging from detection with the anti-DNA antibody.

Helicase activity of Twinkle mutants is reduced or absent

The helicase activity of Twinkle variants was compared using His-affinity purified protein in an in vitro helicase assay. Twinkle wt showed the expected helicase activity and was able to unwind a DNA substrate with a 5' overhang of >20 bases, as long as the double-stranded part was less than 25 base pairs (Figure 4 and data not shown). No such helicase activity could be detected with Δ346–376 or Twinky. Similarly, proteins bearing the
The effect of overexpression of the various mutant proteins on replication was studied using two-dimensional neutral/neutral agarose gel electrophoresis (2DNAGE) and Southern blotting, as first established by Brewer and Fangman (36,39). This method allows visualization of RIs, as DNA fragments are separated both by size and shape.

When applied to the analysis of mtDNA isolated from cultured cells, Holt and co-workers showed that various types of RIs can be detected [see e.g. (9)]. Figure 5 shows a schematic figure of human mtDNA, indicating the appropriate restriction fragments and probes used for detection of these fragments on 2DNAGE gel blots. The first 2 panels of Figure 6A show the example of a HincII digest of purified HEK293 mtDNA run on a 2DNAGE gel and probed for the 3.9 kb mtDNA fragment (nucleotide number [nt] 13 636–1006) that includes the whole NCR (for a detailed explanation of the various RIs see Supplementary Data). All the analyses shown in the subsequent figures consider this HincII fragment. Analysis of a second region of the genome was also performed (Supplementary Figure 1).

We applied the 2DNAGE methodology to analyze the effects of overexpression of POLG1 and Twinkle variants on mtDNA RIs and in particular to test the hypothesis that the variants depleting mtDNA do so by causing non-site specific, generic replication stalling or pausing. First we examined the effect of overexpression of wild-type Twinkle by comparing mtDNA RIs with and without induction (Figure 6A, right two panels). Very little change in quality or quantity of RIs was observed, with the exception of a presumed resolution intermediate which markedly decreased after induction with >3 ng/ml DC. Similar results were observed with induced expression of untagged wild-type Twinkle (not shown). Increased expression of Δ346–376 or Twinky showed no effect on overall replication (Figure 6B) or RIs.

When overexpressing the Twinkle mutant K421A or G575D, a considerable increase in γ- and bubble RIs was observed (Figure 6C). Since at the same time the total amount of mtDNA decreased strongly, these results indicate a severe, non-site specific, slowing down of replication fork movement. In parallel there was a concomitant decrease in RNA containing RIs, RITOLs or partially single-stranded RIs. Thus, the bubble arc was not only sharper and longer than normal (Figure 6C) but also substantially resistant to RNase H treatment (Supplementary Figure 2). In K421A and especially G575D expressing cells the majority of mtDNA molecules were found on the bubble arc, indicating stalling occurred in the early phase of replication as replicating molecules forming bubble arcs on 2DNAGE harbor initiation site(s) in the fragment. The fraction of molecules in a replicative...
Figure 6. Some Twinkle mutants cause replication stalling. 2DNAGE samples for all panels consisted of purified mtDNA digested with HincII and probed with a radiolabeled cytochrome b gene fragment (nt 14,846–15,357). The detected fragment includes the non-coding region of mtDNA also including the cytochrome b, ND6, part of the ND5 gene and intervening tRNA genes (nt 13,636–1006). (A) The first two panels on the left show the RIs of HEK293 cells and the interpretation based primarily on earlier 2DNAGE analysis of mtDNA RIs (7,9). Abbreviations: 1n, 3.9 kb non-replicating HincII fragment. (b) bubble arcs. MtDNA bubble arcs are usually very sensitive to RNase H due to the presence of patches of RNA–DNA especially on the lagging strand; these therefore also fall in the category of RITOLS as do various other RIs. Here, y and y' indicate ascending and descending parts of the y arc and (dy) indicates double-Y structures. These will eventually form resolution intermediates resembling Holliday junctions (HJL-Holliday junction like molecules). The two panels on the right show a comparison of RIs of non-induced and fully induced cells expressing Twinkle wt. The only notable difference in this case is a reproducible reduction in one of the HJL RIs as indicated. (B) A normal pattern of RIs similar to non-expressing cells was observed in cells expressing Twinky or the iLinker variant. (C) K421A, G575D and i70–343 show similar patterns of replication stalling, with increased bubble (b) and descending Y-arc (y') intensities, a sharpening and lengthening of the bubble arc and loss of RITOLS (ovals). The right-most panel shows the same exposure 2D gel pattern of the non-induced i70–343 line showing the typical HincII fragment pattern, including abundant RITOLS. (D) A limited S1 digestion illustrates that stalled RIs observed in panel C are S1 insensitive (right two panels). The effectiveness of the S1 treatment is illustrated by the left two panels, showing the effect on Twinkle non-induced cells. Similar to the S1 treatment RNase H treatment shows that the stalling RIs observed with Twinkle mutants are largely insensitive to this enzyme (Supplementary Figure 2), showing that the observed stalled RIs in panel C are essentially dsDNA. Although the intensities in subfigures A–D cannot be directly compared due to differences in exposure time, each panel contains appropriate controls of similar exposure. For example, the exposures of the left two panels in C have been chosen to be similar in comparison to the right-most panels to properly illustrate the severity of the stalling phenotype.
state was ~30% in the 3.9 kb HincII fragment, compared to <5% in control cells. In contrast, the Δ70–343 mutant showed a somewhat milder stalling phenotypic (Figure 6C), with the majority of RIs in the upper part of the bubble arc and on the descending area of the y-arc suggesting that replication is not aborted so frequently in the very early stages as was the case for the K421A and G575D mutants. Nevertheless, in this case also RITOLS and single-stranded RIs were considerably reduced. Finally, S1 nuclease treatment for various lengths of time with a fixed amount of enzyme resulted only in very minor changes in the abundance of the stalled RIs in Twinkle stalling mutants (Figure 6D shows the treatment of Δ70–343 isolated mtDNA). Only with the longest treatment was there some reduction in the bubble arc intensity with a slight concomitant increase in the ascending part of the y-arc, suggesting some single-strandedness in single-strandedness in the bubbles resulting in broken bubbles upon S1 nuclease treatment. This is not unexpected as some single-strandedness is always expected close to the junction point of the bubble structure of replication intermediates. The Δ70–343 stalling in this case still showed some RITOLS intermediates and these were efficiently removed by the S1 treatment (compare regions marked by arrows in the right two panels of Figure 6D). Nevertheless, most RIs were insensitive to the S1 nuclease showing that they are essentially dsDNA (this was further supported by the analyses shown in Supplementary Figures 5 and 6). In contrast, simultaneous S1 treatment in samples with abundant RITOLS showed a strong reduction in RITOLS RIs, illustrating the effectiveness of the S1 treatment (compare regions marked by arrows in the left two panels in Figure 6D).

Overexpression of POLG1 wt (Figure 7A) and POLG1 ΔCAG (not shown) did not result in any obvious effect on the RIs whereas even modest expression of the mutant variants D890N and D1135A lead to a clear increase in y- and bubble arcs, suggesting these POLG1 mutants also caused replication stalling (Figure 7B). However, unlike the stalling Twinkle variants, RITOLS and single-stranded RIs persisted in the case of D890N and D1135A POLG1 mutants (Figure 7B), even after full induction of the mutant proteins for three days (not shown).

The proofreading deficient D198A POLG1 showed little change in the appearance of RIs. Three ng/ml DC slightly enhanced both the Y- and bubble arc (Figure 7C). Only after full induction did we observe a phenotype suggestive of stalling, similar to but clearly less severe than that observed with D890N and D1135A overexpression at low induction levels.

Comparison of Twinkle induced replication stalling with POLG induced replication stalling showed a considerable quantitative difference in RITOLS and single-stranded RIs (see also Supplementary Figures 3 and 5). We hypothesized (see also Discussion) that this difference could be explained by involvement of POLG1 in initiation of lagging-strand DNA synthesis or maturation. In this scenario, expression of POLG1 mutants would not only result in stalling of the leading-strand but also in delayed lagging-strand DNA synthesis. To test whether inhibition of POLG1 function could delay lagging-strand DNA synthesis in cells overexpressing either Twinkle wt or the stalling mutations Δ70–343, G575D and K421A, cells were treated with the cytidine analogue deoxycytidine (ddC), a competitive inhibitor of POLG. Incorporation instead of deoxycytidine can result in chain termination. Surprisingly, already a short 3–4 h treatment with a high dose of ddC (200 μM) in cells showing modest stalling due to Twinkle mutants like G575D, resulted in the reappearance of considerable levels of RITOLS, while accumulated y- and bubble-arc RIs were as prominent as in untreated cells (Figure 8). The appearance of RIs under the applied conditions was highly similar to the appearance observed with POLG1 stalling (compare lower panels of Figure 8 with those in Figure 7B). Due to the slightly leaky nature of the Flp-In™ T-REx™ system, the strongest Twinkle stalling mutants G575D and K421A already showed signs of stalling without induction. In particular, the G575D mutant had already lost most RITOLS in the absence of DC. Treatment of these cells with ddC showed a similar reappearance of RITOLS (Figure 8). The same regime of ddC treatment did not result in obvious changes in RIs in cells expressing Twinkle wt or in non-induced Twinkle wt cells (not shown). Finally, ddC treatment of cells showing a strong stalling phenotype caused by higher level expression of Twinkle mutants did not result in a clear increase in RITOLS (not shown), suggesting that replication was completely abolished under these conditions (see Discussion section).

DISCUSSION

Because mtDNA encodes some of the central components required for cellular energy metabolism, its maintenance is essential for development and overall cell function. While human and mouse mtDNA were sequenced 25 years ago, for a long time the knowledge of the replication and repair machinery was lacking. This made it very difficult to test predictions of replication models for example by reconstituting the various components needed for replication or by manipulating the individual enzymes.

Using inducible overexpression of wild type or mutant variants of Twinkle and POLG1 in cell culture, we show here that replication stalling results in changed patterns of RIs that can best be understood by considering the proposed mechanisms of lagging-strand synthesis. Most notably our results show that stalling induced by deficient Twinkle results in Rs that mimic conventional strand-coupled RIs and suggest that initiation of lagging-strand DNA synthesis or maturation occurs at multiple sites across the genome. We furthermore propose that this maturation involves POLG1.

Catalytically deficient POLG1 and Twinkle mutants

In this study we have used the HEK293 Flp-In™ T-Rex™ inducible expression system to study mutants of POLG1 and Twinkle expected to result in severe catalytic defects of these enzymes.
As the Flp-In™ T-Rex™ system is a transgenic system, endogenous copies of the genes under study are still being expressed. Thus, in order to discuss here the severity of the mutations studied and to better value the use of the inducible Flp-In™ T-Rex™ system we tried to estimate relative levels of endogenous and recombinant proteins. The results (presented in Supplementary Figures 7 and 8) showed that at 3 ng/ml DC following 2 days induction, POLG1 recombinant and endogenous protein levels were comparable, suggesting this regime recreates the effects of expression of a ‘heterozygous’ mutant. The situation was different for Twinkle. Both northern blot analysis and western blot analysis suggested that at 1 ng/ml DC for 2 days, the wt recombinant Twinkle protein level is 4–5-fold the level of endogenous Twinkle, whereas at 3 ng/ml for 2 days recombinant Twinkle is ~8–10-fold the endogenous level.

We have previously shown that the polymerase mutants D890N and D1135A of POLG1 are deficient in a commonly used reverse transcriptase assay. Transient expression in cell culture showed modest mtDNA depletion (27). The corresponding aspartate 890 and 1135 residues in *E. coli* polymerase I (Asp705 and Asp882) have both been shown to be essential for catalysis (40,41). A model of POLG1 based on the family A polymerase structures including *E. coli* polymerase I places both these residues at critical sites for catalysis (23), which is corroborated by the fact that cell lines for stable
constitutive expression of these variants could not be established. More recently, HEK293 T-Rex cells were established expressing the D1135A mutation (42). The authors showed, as we repeated here, that induced expression of this variant resulted in fast mtDNA copy-number depletion. Similarly we now show that the D890N mutation resulted in similar rapid mtDNA depletion upon induction. For both variants we observed a reduction in copy number that suggests that upon full induction mtDNA is diluted by half with each cell division and thus can no longer be replicated. The generally low expression levels of recombinant POLG1 and the 1:1 ratio to endogenous PolG1 at 3 ng/ml DC imply that both the D890N and D1135A mutation behave as dominant and furthermore suggest that the results are unlikely due to the capture of the entire pool of the POLG1 accessory subunit POLG2 by excessive expression of catalytically inactive mutants.

Like the POLG1 polymerase domain mutations, the Twinkle K421A mutation is predicted to abolish catalytic activity. This mutation is an invariant residue of the Walker A motif that based on the T7 helicase crystal structures is essential for nucleotide binding and hydrolysis (43,44). Indeed, using partially purified Twinkle K421A, we showed that this mutation abolished helicase activity of Twinkle whereas wild-type purified protein showed robust helicase activity and similar specificity (not shown) as published (15). The splice variant Twinky showed no evidence of being able to unwind the helicase substrate, even though most of the core helicase motifs except for the last few amino acids of motif H4 are left intact. The absence of helicase activity is nevertheless not unexpected as we showed previously that this variant is mostly monomeric (14), whereas the helicase activity is expected to require hexamer formation. In contrast to full-length Twinkle, overexpressed Twinky does not co-localize with mtDNA in nucleoids within the mitochondrial network. Likewise, the 31 amino acid deletion of the region, that by analogy with the region in the T7 primase/helicase we tentatively call the linker region, did not show nucleoid localization and was deficient in the unwinding assay. However, using glutaraldehyde cross-linking (Supplementary Figure 4) this mutant was still able to form multimers, unlike Twinky. The analogous mutation to G575D has previously been studied in the T7 primase/helicase (G488D) and was shown to be severely defective in its unwinding as well as its primase activity (37), although the remaining primase activity was still sufficient to allow for phage growth. It was furthermore shown to have weak DNA-binding capacity, in part explaining its weak enzymatic activities. In our hands, the Twinkle G575D mutant did not show helicase activity and its nucleoid localization was partially compromised consistent with weak DNA binding. Of the variants tested here, G575D and K421A behaved as dominant mutations since they reduced mtDNA copy number even at very low levels of expression, which occurred without induction by DC caused by leakiness of protein expression. Based on the comparison above and the difficulties to detect the protein without induction we are confident that under these conditions all recombinant Twinkle mutants are present at levels well below that of the endogenous protein. Twinky and Δ346–376 did not reduce copy number and are thus suggested not to compromise the activity of the endogenous Twinkle still expressed in these cell lines. Based on recent mechanistic insight in the T7 primase/helicase (45) and the fact that the purified Δ346–376 enzymes is inactive, this mutant is unlikely to form hetero-multimers with the endogenous enzyme. Finally, the Δ70–343 mutant reduced mtDNA copy number and helicase activity. Although this showed the importance of this region for Twinkle function, these results and the analysis by 2DNAGE have not yet established a specific function for this part of the protein such as a primase activity.

Replication stalling phenotypes are compatible with delayed lagging-strand synthesis

One of the powers of 2DNAGE is that it allows accurate predictions of results on the basis of a given replication model. The method is therefore very useful to generate and test hypotheses as demonstrated by the adaptation of the method in a computer model (46). One of the unusual features of the strand-displacement model (4) is the prediction of extensive single-stranded RIs. Although these RIs have been observed by electron microscopy and more recently by AFM after mtDNA coating by SSB (47), the RITOLS replication model considers them
artefacts of degradation of the RNA patches of the initial lagging-strand. Indeed, single-stranded RIs predicted depending on the origin of replication and the restriction enzymes used (7,48) are not readily observed by 2DNAGE. It has been suggested however that the failure to observe these RIs is due to extensive branch-migration (47), thus arguing that the 2DNAGE technique is flawed and that the majority of RIs predicted by the strand-displacement model are not detected as such by this technique. This however does not provide an explanation for the observation of abundant RITOLS RIs.

Our observations of mtDNA copy-number depletion and a concomitant overall increase in RIs observed by 2DNAGE are evidence of extensive replication stalling or slowing down of replication fork movement in a non-site-specific manner. Replication stalling induced by the Twinkle K421A and G575D mutations has direct bearing on the question of delayed lagging-strand DNA synthesis as proposed by both the strand-displacement and RITOLS models. Using 2DNAGE we have shown here that accumulated RIs were visible using 3 different restriction enzymes (HincII in Figures 6–8, BclI in Supplementary Figure 1 and DraI in Supplementary Figure 5A) and probes for fragments that cover half of the genome. The results show that reduced fork progression occurs throughout mtDNA and not at a few specific sites since we did not observe appearance of discrete spots on replication arcs but a general increase of the intensities of arcs of RIs. Probing for the 3.9 kb HincII fragment from mtDNA nt 13 636–1006, which includes O$_H$ as well as most of the initiation zone (OriZ) covering the cytochrome b, ND6, part of the ND5 gene region (8), showed both an intense bubble and descending y-arc typical of a fragment containing multiple initiation sites within that fragment. However, since the ascending y-arc signal in this fragment was generally weak even when the bubble arc signal was very strong, the results nevertheless suggest that most initiation occurred in the NCR region, possibly at or near O$_H$ and not in OriZ, which would result in a more uniform y-arc (8). 2DNAGE analysis of a DraI fragment spanning most or all of OriZ but not including O$_H$ further supports this hypothesis as it showed a strong y-arc signal in Twinkle stalling mutants but only a very weak bubble arc (Supplementary Figure 5A). These results thus suggest that despite the apparent lack of RITOLS on 2DNAGE gels in the Twinkle stalling mutants, the predominant initiation occurs at O$_H$ as was proposed for both the RITOLS (10) and the strand-displacement model (4). Since the RIs observed with the Twinkle stalling mutants were RNaseH and S1 nuclease insensitive they were most likely double-stranded DNA and thus resemble strand-coupled RIs. The double-stranded nature of the RIs in question was further confirmed by comparing RIs obtained by digestion with DraI and analysis of a fragment close to but not including the NCR origin (Supplementary Figure 5A and B). As explained in detail in Supplementary Figure 5C the prediction is that double-stranded RIs would yield a conventional Y-arc, whereas RITOLS and single-stranded intermediates being non-digestible on the nascent lagging-strand would yield a retarded arc resembling a Y-arc (SMY or slow-moving Y). Also by comparison with Supplementary Figure 5A, the results show that the vast majority of RIs resulting from Twinkle stalling mutants are on a conventional Y-arc whereas by comparison RIs from cells expressing Twinkle wt or expressing the POLG1 D1135A mutant are on the predicted SMY (7,10). As an alternative we purified replication intermediates from preparative Twinkle G575D 2DNAGE gels and analyzed the purified RIs by restriction digestion using two dsDNA-specific enzymes (XhoI and DraI) and AccI that is capable of digesting ds and ssDNA. As demonstrated and illustrated in Supplementary Figure 6, these digests also fully support the conclusion that RIs observed with the Twinkle stalling mutants are essentially double-stranded DNA. As it is highly unlikely that leading- and lagging-strand synthesis have become coupled due to fork stalling, the observed double-stranded RIs can only be reasonably explained by an increased rate of lagging-strand initiation and/or maturation relative to the rate of fork movement. The results furthermore imply this to occur at sites other than the proposed lagging-strand initiation site(s) because the 3.9 kb HincII is unlikely to contain any, based on the mapping of the mouse second preferred maturation start-site proposed to be between nt 12966 and 12671 (10). Although not excluded at the moment, in the strand displacement model this would require novel priming events by unknown proteins and mechanisms. In the case of RITOLS intermediates, processing of incorporated RNA can be more easily envisaged to yield fully dsDNA intermediates.

The conclusions above are further strengthened by the ddC experiments. In our study overexpression of POLG1 mutants induced replication stalling but maintained RITOLS RIs. In contrast, Twinkle stalling mutants invariably resulted in an extensive loss of RITOLS RIs on 2DNAGE and in an associated increase in dsDNA RIs. The suggested cause-effect that dsDNA RIs increase as RITOLS decrease is further reinforced by the demonstration that a short inhibition of POLG by ddC led to a re-emergence of the RITOLS RIs under mild stalling conditions induced by Twinkle mutants and demonstrates that RITOLS intermediates precede dsDNA replication intermediates. Importantly, higher expression levels of, e.g. G575D for two or three days followed by ddC treatment did not result in the reappearance of RITOLS. We interpret this as a sign of a complete halt in ongoing replication and initiation whilst visible stalled RIs have already been matured. Because under these conditions no replication (re)initiation occurs, also no new RITOLS intermediates are observed upon inhibition of POLG1 by ddC. In contrast, low G575D expression levels allow for a continued fork progression albeit at a reduced rate. Accordingly, very low (leaky) G575D expression resulted in a reduced mtDNA copy number, but a steady-state level was maintained over many cell generations (not shown). As the only plausible target for ddC treatment is POLG, the results seem to imply that the lagging-strand POLG is more sensitive to ddC treatment than the leading-strand polymerase, suggesting a different composition of leading and lagging-strand polymerase holoenzymes. Although the proposal of RITOLS replication (10)
leaves open many questions regarding the synthesis of the initial RNA-rich lagging-strand, our data suggest that POLG is involved in lagging-strand maturation. Furthermore, as proposed previously (7) our data confirm that this is one of the slower steps in mitochondrial replication. Although various RNA-associated activities of POLG1 have recently been characterized (49), to our knowledge a displacement synthesis where POLG1 would displace annealed RNA from a template strand has not yet been demonstrated. Alternatively, POLG1 exonuclease activity might be involved in removing nascent RNA before initiating DNA synthesis. Our data however seem to exclude this idea, as overexpression of the exonuclease deficient POLG1 D198A did not seriously affect mtDNA copy number or the abundance of RNA-rich versus RNA-poor RIs. Nevertheless, moderate changes in replication intermediates at high D198A expression levels were observed by 2DNAGE and warrant further investigation.

The system we have established here provides an excellent tool to study the function of other proteins in mtDNA maintenance, to establish their enzymatic or structural roles in, e.g. replication or repair and to address questions concerned with the mechanisms of DNA replication. It will also be invaluable to determine the in vivo effects of mutations in Twinkle and POLG1 associated with human disorders of mtDNA maintenance. Low-level expression of the severe Twinkle mutations K421A and G575D due to the leakiness of the expression system also in the absence of DC caused reduced mtDNA steady state levels of ~60%. Medium to high induction of the same mutations ultimately lead to the loss of protein expression, probably due to growth arrest or death of cells that lost their mtDNA completely and outgrowth of a few remaining cells that either had lost protein expression or did not have expression to start with, very similar to our earlier observations with constitutive expression of the POLG1 D198A mutant in 293T cells (27). In contrast expression of the null mutations did not have any effect on mtDNA levels also after 7 days and no negative selection occurred. These considerations also illustrate that the system is best used for transient expression of mutant proteins, that care must be taken when long-term effects are studied and that long-term maintenance of master plates without DC is not without hazard.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We like to acknowledge Outi Kurronen and Merja Jokela for their skilful technical assistance, and Ian Holt and Howy Jacobs for many fruitful discussions. This research was financially supported by the Academy of Finland (Grants 110689, 103213 and CoE funding), the Juselius Foundation, Tampere University Hospital Medical Research Fund, the University of Tampere and the EU sixth Framework Programme for Research, Priority 1 ‘Life sciences, genomics and biotechnology for health, contract number LSHM-CT-2004-503116’ supporting EUMITOCOMBAT. T.Y. was supported by the UK Medical Research Council and was the recipient of a Postdoctoral Fellowship for Research Abroad from the Japan Society for the Promotion of Science (April 2003–April 2005) and is currently funded by the UK Muscular Dystrophy Campaign. We gratefully acknowledge Drs Massimo Zeviani and Valeria Tiranti for providing us with a polyclonal antibody against Twinkle. Funding to pay the Open Access publication charges for this article was provided by EUMITOCOMBAT.

Conflict of interest statement. None declared.

REFERENCES


Supporting data and Figure Legends

2DNAGE analysis of mtDNA RIs: a detailed explanation of 2DNAGE patterns of highly purified human mtDNA isolated from cultured cells

In 2DNAGE, all fragments that are not in the process of being replicated will run as any normal dsDNA restriction, on the so-called 1n linear. When probed for a single fragment, the most dominant species will usually be a regular non-replicating restriction fragment situated at the so-called 1n-spot. Secondly, probes for most fragments of mtDNA will show a y-arc containing a subpopulation of molecules where the replication machinery passes through the fragment from one end. In a fragment such as the 3.9 kb HincII fragment, containing the region downstream of O_H, bubble arcs are visible, indicative of replication initiation. In addition, partial Y-arcs will form as bubbles reach one end of the fragment before they reach the other end, i.e. the y-arc continues from the point where the bubble leaves the fragment (theoretically, when a single unidirectional, precisely positioned origin is used, the bubble will burst giving arise to a defined single spot on the Y-arc that will continue as double-Y arc). In fragments containing O_H additional double-Y structures can be found. These structures are relatively S1 and RNaseH resistant, indicating a double-stranded DNA structure and they most likely represent replication resolution intermediates, where two replication forks moving in opposite orientation meet in the D-loop region. Holliday junctions will also appear at a similar position, but will usually start from the 1n-linear. Finally, double-Y or X shaped-like molecules resulting from replication stalling and fork reversal will also be visible at this region on 2DNAGE. In the case of mtDNA, several unusual species arise from the presence of RNA:DNA hybrid molecules. Since most restriction enzymes cannot cut RNA:DNA hybrids or single-stranded DNA that can arise from even minor contamination with RNase H, larger than expected fragments are found. Linear fragments containing one or more protected restriction sites are found as spots on the 1n linear arc, while replication intermediates of larger size accumulate in arcs of higher molecular weight than the resolution intermediates. These intermediates are indicated as RITOLS in Figure 6 (see also Yasukawa et al., 2006). As already indicated, ribonucleotide incorporation makes replication intermediates more susceptible to degrading enzymes that can also result in single-stranded patches on more conventional structures. This is for example demonstrated by the bifurcation of the bubble arc.

Supporting Figure S1 Stalling of mtDNA replication caused by Twinkle or POLG1 mutants is apparent also in other regions of the genome. Purified mtDNA from the various indicated cell-lines was digested o/n with BclI at 37°C. The Southern blots were probed with an mtDNA probe nt 4,480–4,984 that detects the restriction fragment between nt 3,658–7,657. The results show an accumulation of replication intermediates in cell-lines expressing Twinkle variants K421A, G575D and Δ70-343 as well as in POLG1 D1135A compared to wt Twinkle or POLG1 expressing cells, indicating that in all the tested cell-lines stalling occurs throughout the mitochondrial genome. Note that exposures of either the Twinkle variants or the POLG1 variants are identical as samples were run on the same gels.

Supporting Figure S2 Twinkle K412A and G575D RIs are relatively insensitive to RNase H (RH). A comparison of RH sensitivity of non-induced wt Twinkle cell RIs and K421A and G575D induced cells treated with -/+ 1 U RH for 30 min at 37°C. While non-induced Twinkle wt cells show strong sensitivity of RIs to RH, K421A and G575D show only a minor general sensitivity.

Supporting Figure S3 A direct comparison of stalling phenotypes. HincII digest of wt Twinkle, Twinkle K421A and G575D and POLG1 D1135A probed with a cytochrome b fragment (nt 14,846–15,357) (see Fig. 6). All four samples were treated and run in parallel on the same
2DNAGE gels and blotted on the same membrane. The figure demonstrates the near complete absence of RITOLS for Twinkle stalling mutants whereas POLG1 D1135A induced stalling shows abundant RITOLS. Note that due to unintended differences in loading (compare 1n spots) of D1135A, K421A and G575D, the stalling in D1135A looks much stronger compared to K421A and G575D than it is in reality. Also note that due to incomplete removal of protein in the Twinkle wt sample, RITOLS appear less intense then they normally are. A smear of protein associated RITOLS was clearly visible on the top left (indicated by the black oval). Gray ovals indicate the position of RITOLS, (b) indicates bubble arcs, (y’) indicates the descending part of the y arc.

**Supporting Figure S4** Twinkle multimerization. Glutaraldehyde crosslinking was used to examine the multimerization properties of all Twinkle variants. For this 1 ng Twinkle and Twinkle variants were incubated in 25 mM Tris pH 7.6, 40 mM NaCl, 4.5 mM MgCl2, 100 mM L-Arginine-HCl pH 7.6, 10% glycerol, 1mM DTT for 10 min on ice to allow oligomerization. After addition of 0.05% Glutaraldehyde the reaction was incubated for further 10 min at RT. The crosslinking was stopped by addition of 1 Vol SDS-PAGE sample buffer and heat denaturation. Samples were run on denaturing 3-8% Tris-acetate gels and analysed by Western blot analysis. Lane 1 shows wt Twinkle without (w/o) Glutaraldehyde (GA) crosslinking. All other lanes show crosslinked protein. Monomeric Twinkle in this and other lanes is indicated with an asterisk where visible. Note that the amount of loaded protein in lane 1 and 2 is identical, suggesting that the monomeric band is not detected efficiently. The results show non-crosslinked, but denatured Twinkle running as a single band of the expected molecular weight. After crosslinking Twinkle, as well as the K421A and G575D, showed multiple forms corresponding in molecular weight with monomer, dimer, trimer, tetramer and higher order forms that were more difficult to assign due to limited resolution of the gel. Twinky was mostly present in its monomeric form although a dimer was clearly visible as well. Δ70-343 showed the most convincing multimerization with the used conditions, with a very abundant species that would agree with the expected size for hexa- and/or heptameric species. Δ346-376 somewhat surprisingly showed an abundant multimeric form that could correspond with a hexa- and/or heptamer.

**Supporting Figure S5** (A) Twinkle stalling mutants initiate replication primarily in the NCR. Frequent replication initiation in OriZ just downstream of the NCR should show a prominent bubble arc in digests that include this region but not the NCR. To see the site of initiation mtDNA from the cell-lines for the Twinkle stalling mutants K421A, G575D and Δ70-343 was digested with DraI and probed with the cytochrome b fragment (nt 14,846–15,357), detecting the mtDNA fragment between nts 12,271-16,010. The results show that despite a very prominent y-arc in the stalling Twinkle mutants a bubble arc was only very faintly visible. Combined with the HincII digests (Figure 6), this indicates that in all studied cells initiation of replication occurs mainly in the NCR and does not involve OriZ.

(B, C) 2DNAGE blots probed for the DraI fragment further demonstrate that mutant-Twinkle stalling RIs are dsDNA. A comparison of the DraI digest analysed by 2DNAGE as shown in (A) with various controls is shown in B. The controls are mtDNA from Twinkle wt and POLG1 D1135A cells both induced for 2 days at 3 ng/ml. The right-most panel in B shows again a DraI digest for Twinkle K421A mutant which was digested and run along the control samples simultaneously. (C) The DraI 12,271-16,010 fragment that we looked at in this Figure overlaps largely with the HincII fragment and ends just outside the NCR. As DraI only cuts dsDNA, the bubble arc replication intermediated achieved by a HincII digest can only be cut with DraI (at position 16,010) if they are dsDNA. If they are partially single-stranded or contain ribonucleotides they should not be cut on both strands and will be retarded and the next DraI site would be used at nt 2,051. The prediction is, as illustrated in this panel, that most RIs that are essentially dsDNA should be on a conventional Y-arc when a DraI digest is probed for the 16,010-12,271 fragment,
whereas RITOLS intermediates should be mostly on a retarded Y-arc. Thus, the results from panels A and B are in agreement with this prediction and fully confirm that stalling RIs induced by Twinkle mutants are mostly dsDNA.

**Supporting Figure S6** Gel-purified *HincII* 13,636–1,006 fragment RIs resulting from Twinkle mutant induced stalling can be digested by dsDNA-specific restriction enzymes. MtDNA from Twinkle G575D cells induced for 2d with 3ng/ml was run on a preparative 2DNAGE gel. By overlay with an exposure of a Southern blot run in parallel, the 1n spot, the descending Y and the upper ~30% of the bubble arc were cut out, DNA was eluted by electroelution and precipitated. The resulting fragments were digested with restriction enzymes that can either only cut dsDNA once within the original *HincII* fragment from 13,636–1,006 (*XhoI* at 14,955 and *DraI* at 16,010) or can cut both ds and ssDNA once within the *HincII* fragment (*AccI* at 15,255). Undigested and digested samples were then run on a regular 0.4% agarose gel without ethidium bromide, blotted and probed with a *cytochrome b* fragment (nt 14,846–15,357). The results confirm that the purified RIs are almost completely digestible by two commonly used restriction enzymes confirming they are indeed dsDNA. The results also show that the purified replication intermediates are of predicted size of partially replicated dsDNA molecules as illustrated and calculated in the Figure. *AccI* gave essentially the same results and did not give shorter fragments that are predicted assuming the lagging strand was not synthesized or partially degraded. Single asterisks next to band sizes indicate that these are the predicted sizes based on a single major origin at OIH at nt position 191. We speculate that the slightly anomalous size of these bands is caused by ethidium bromide still present in the DNA following the purification from the second-dimension 2DNAGE gel. Double asterisks indicate bands resulting from melting out of more complex RIs. In case of the upper bubble arc, the majority of each fragments appeared as such which could be expected as the bubble-junction is close to the end of the fragment. Nevertheless, especially in the DraI digest a smear is clearly visible (indicated by a thick line on the autoradiograph; a longer exposure is shown on the right) from approximately 4.8 kb down to 3.5 kb. Both the smearing and size is exactly what would be predicted for a digested growing bubble arc (from 16,010 to 13,636). The fact that we here isolated only the upper part of the bubble limits the smear to approximately 3.5 kb. As can be seen in the autoradiograph also some melting occurred for the descending Y fragments. This accounted for less than 25% of the total amount and based on various tests was due in large part to the electroelution as it increased dramatically with longer electroelution times, even though the electroelution was done at +4°C.

**Supporting Figure S7** Recombinant POLG1 is expressed at levels similar to endogenous POLG1. (A) To test whether we would be able to easily detect endogenous POLG1 from 293 cells we prepared a partially purified protein by using the accessory protein POLG2, a known interaction partner to POLG1. For this purpose we used and HA-affinity matrix (EZview™ Red Anti-HA affinity gel, Sigma) to precipitate POLG2.HA from stable 293T cell lines (J.N. Spelbrink, unpublished data) expressing either POLG2.HA alone or POLG2.HA and POLG1.MycHis. as a negative control we performed the same purification step on cells expressing only POLG1.MycHis. The purified samples were run on a 7.5% Laemmli gel, blotted and probed first with an anti-Myc monoclonal antibody (Roche), and subsequently with a anti-POLG1 specific antibody (R-18 sc-5930, Santa Cruz Biotechnology). Results show that endogenous POLG1 could be precipitated by overexpressed POLG2.HA. The equivalent of approximately 3 x 10^6 was loaded indicating that POLG1 should be easily detectable in a mitochondrial lysate of one 10 cm-diameter plate. (B, C) Mitochondrial lysates prepared from the indicated cell lines with the indicated treatments were analysed by Western blot analysis using the anti-POLG1 antibody. Results indicate that at 3ng/ml DC recombinant POLG1 wt or D890N, are expressed at levels similar to the endogenous POLG1.
For better visibility a section of the gel in B has been enlarged. The single asterisk marks the endogenous POLG1, whereas the double asterisks marks POLG1.MycHis. The identity of POLG1.MycHis was confirmed by stripping and reprobing with the anti-Myc antibody (not shown). For reference, purified endogenous POLG1 and/or POLG1.MycHis were run alongside the samples shown in B and C. Note that POLG1 in the purified samples appeared slightly retarded compared to POLG1 in the mitochondrial lysates, very likely due to differences in buffer and detergent composition of the samples. Finally as an additional control for the assignment of bands, samples in C were run on duplicate gels and while the first gel was probed only with the anti-POLG1 antibody, the second gel was incubated with antibody and blocking peptide (sc-5930p, Santa Cruz Biotechnology) at 10-fold excess. The assigned bands could no longer be detected except for a still weakly visible purified POLG1 showing they were specifically detected by the antibody. Reprobing with antibody alone again visualized the assigned POLG1 bands.

Supporting Figure S8 Recombinant Twinkle is expressed at moderately increased levels compared to the endogenous protein. As with the existing antibodies endogenous Twinkle cannot be detected in mitochondrial lysates two alternative approaches were used. (A) shows the comparison by Northern blot analysis of the transgene derived mRNA with the endogenous mRNA. For this total RNA from the indicated cell lines and treatments (DC treatment was for 2 days) was isolated using Trizol (Invitrogen). PolyA+ RNA was isolated from total RNA using the Oligotex kit (Qiagen) and run on a 1.2% agarose MOPS-formaldehyde gel. The blotted gel was probed first for Twinkle using a Twinkle cDNA probe (nts 1149-2057 of the cDNA sequence). Following Phosphorimager analysis the blot was reprobed with a reference PCR probe detecting glutamate dehydrogenase (GDH, nts 445-699 of the cDNA sequence). Below the blot images we indicated the amount of transgene derived Twinkle mRNA relative to the endogenous Twinkle mRNA in control 293 Flp-In™ TREx™ cells, after correction for the GDH derived signal. Note that we did not compare Twinkle transgene to the endogenous signal from the same sample in the DC treated Twinkle wt cells because the signal for the endogenous Twinkle was too blurred in these lanes to allow for accurate Phosphor-image analysis. (B) Because Twinkle is able to form multimers we reasoned that if we purified Twinkle.MycHis, endogenous Twinkle should also be pulled down. Talon-pulldown of His-tagged Twinkle from Twinkle wt cells induced with low levels of DC (1 and 3 ng/ml) showed the purification of an additional product by Western blot analysis, that by comparison with overexpressed untagged Twinkle was of the correct size (not shown). This product, in contrast to the upper band was not detected by the anti-Myc antibody, but was recognized by a polyclonal anti-twinkle antibody. In addition, with both 1 and 3 ng/ml DC similar amounts co-purified suggesting that this is not a breakdown product. Western blot analyses to detect endogenous Twinkle used a polyclonal antibody kindly provided by Drs. Massimo Zeviani and Valeria Tiranti. Quantification by enhanced chemiluminescence and the Chemidoc gelimaging system and ImageQuant software (BioRad) showed that at 1 ng/ml DC a similar 4-fold overexpression of recombinant Twinkle over endogenous Twinkle was observed.

Detailed Materials and Methods for all Supporting Figures can be obtained upon request.
Supporting Fig. 1 Wanrooij, Goffart et al.
Supporting Fig. 2 Wanrooij, Goffart et al.
Supporting Fig. 3 Wanrooij, Goffart et al.
Supporting Fig. 4 Wanrooij, Goffart et al.
HincII bubble arc intermediates, dsDNA (Twinkle stalling mutants)

Same RIs following digestion with Dral instead of HincII

Ritols intermediates predict a blocked nascent L-strand for Dral

Supporting Fig. 5 Wanrooij, Goffart et al.
### G575D purified stalling replication intermediates

<table>
<thead>
<tr>
<th>Eluted 1n-spot</th>
<th>descending Y</th>
<th>upper bubble-arc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncut</td>
<td>Uncut</td>
<td>Uncut</td>
</tr>
<tr>
<td>XhoI</td>
<td>XhoI</td>
<td>XhoI</td>
</tr>
<tr>
<td>Dral</td>
<td>Dral</td>
<td>Dral</td>
</tr>
<tr>
<td>AccI</td>
<td>AccI</td>
<td>AccI</td>
</tr>
</tbody>
</table>

**Probe:**

- Hincl 13,636
- XhoI +Xho: 1.3 +2.6
- Dral +Dral: 2.4 + (1.5)
- AccI +Acc: 1.6 + 2.4

**1n:**

- 3.9 kb

**Example XhoI:**

- 2 x 1.3 kb fragment
- + 1x ((1x 0.8) + 2 x 1.8) = 1x 4.4 kb

**Descending Y**

- Hincl 13,636
- XhoI
- Dral
- Hincl 1006

- 816 bp
- 1803 bp

**Bubble**

- 13,636
- XhoI 14,955
- Accl 15,255
- Dral 16,010
- 16,568/0

Supporting Fig. 6 Wanrooij, Goffart et al.
A

**Endogenous POLG1**

B

**293 Flp-In™-T-REx™**

C

Supporting Fig. 7 Wanrooij, Goffart et al.
Supporting Fig. 8 Wanrooij, Goffart et al.
The mitochondrial transcription termination factor mTERF modulates replication pausing in human mitochondrial DNA

Anne K. Hyvärinen, Jaakko L. O. Pohjoismäki, Aurelio Reyes, Sjoerd Wanrooj, Takehiro Yasukawa, Pekka J. Karhunen, Johannes N. Spelbrink, Ian J. Holt and Howard T. Jacobs

1Institute of Medical Technology and Tampere University Hospital, FI-33014, University of Tampere, Finland, 2MRC-Dunn Human Nutrition Unit, Cambridge, UK, 3Department of Forensic Medicine and Tampere University Hospital, FI-33014, University of Tampere, Finland and 4Institute of Biomedical and Life Sciences, University of Glasgow, Scotland, UK

Received May 10, 2007; Revised August 8, 2007; Accepted August 19, 2007

ABSTRACT

The mammalian mitochondrial transcription termination factor mTERF binds with high affinity to a site within the tRNA\(^{\text{Leu(UUR)}}\) gene and regulates the amount of read through transcription from the ribosomal DNA into the remaining genes of the major coding strand of mitochondrial DNA (mtDNA). Electrophoretic mobility shift assays (EMSA) and SELEX, using mitochondrial protein extracts from cells induced to overexpress mTERF, revealed novel, weaker mTERF-binding sites, clustered in several regions of mtDNA, notably in the major non-coding region (NCR). Such binding \textit{in vivo} was supported by mtDNA immunoprecipitation. Two-dimensional neutral agarose gel electrophoresis (2DNAGE) and 5’ end mapping by ligation-mediated PCR (LM-PCR) identified the region of the canonical mTERF-binding site as a replication pause site. The strength of pausing was modulated by the expression level of mTERF. mTERF overexpression also affected replication pausing in other regions of the genome in which mTERF binding was found. These results indicate a role for TERF in mtDNA replication, in addition to its role in transcription. We suggest that mTERF could provide a system for coordinating the passage of replication and transcription complexes, analogous with replication pause-region binding proteins in other systems, whose main role is to safeguard the integrity of the genome whilst facilitating its efficient expression.

INTRODUCTION

The mitochondrial genome of animals is organized in a highly compact manner, with virtually no non-coding information between or within its 37 genes. The circular genome is transcribed by a phage-type RNA polymerase into polycistronic transcripts which, in mammals, encompass the entire genome on both strands (1,2). Production of these transcripts depends upon a set of closely spaced promoters located in the major non-coding region (NCR). The primary transcripts are then processed to mature mRNAs, rRNAs and tRNAs via a series of enzymatic steps requiring the tRNA-processing endonucleases RNase P and tRNAse Z, as well as other enzymes. The major coding strand (informationally the L-strand, but for the purposes of transcription conventionally referred to by the name of the template, H-strand) is transcribed from two distinct initiation sites at the heavy-strand promoter (HSP), \(P_{H1}\) and \(P_{H2}\), separated by \(\sim 100\) bp. The \(P_{H2}\)-derived precursor transcript covers virtually the entire genome and can give rise to all of the transcription products of the heavy-strand except tRNA\(^{Phe}\), whose coding sequence overlaps the \(P_{H2}\) initiation site. The \(P_{H1}\) initiation site gives rise to a truncated transcript encompassing just the rRNAs (plus two tRNAs) and thus defines a distinct mitochondrial rDNA transcription unit. Termination at the 3’ end of the rDNA is brought about by a transcription termination factor, mTERF (3–6), which has also been proposed to interact with the RNA polymerase in initiation site selection (2,7). Recent data suggest that this involves formation of a DNA loop in which RNA polymerase complexes are recycled around the rDNA segment of the genome after terminating (7). mTERF binds sequence specifically with high affinity to a

*To whom correspondence should be addressed. Tel: +358 3 3551 7731; Fax: +358 3 3551 7710; E-mail: howard.t.jacobs@uta.fi

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

© 2007 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
sequence element within the coding sequence of tRNA^{Leu(UUR)}, located immediately downstream of the rDNA (4). Current evidence indicates that mTERF interacts with its asymmetric-binding site as a monomer (8), although the tertiary structure of the protein and the structural basis of its interaction with DNA are unknown.

mTERF belongs to a recently identified superfamily of proteins whose functions are largely unknown (9–11). Homologues in Drosophila and in sea urchins have variously been implicated in transcriptional regulation (12–14), regulation of DNA replication (15) and even mitochondrial protein synthesis (11). The sea urchin mTERF homologue mtDBP (D-loop-binding protein) has recently been shown to terminate transcription in a polar manner (14,16), analogous with the activity of mTERF (5). However, mtDBP is also a contrahelicase (15), and has been proposed to play a role in regulating the expansion of the short D-loop of sea urchin mtDNA and thus the initiation of productive replication of the genome.

Transcription and replication of mtDNA have long been regarded as interlinked processes. The primer for initiation of DNA replication has been assumed to be a product of transcription by the mitochondrial RNA polymerase. However, there is no consensus concerning the mechanism by which 3' ends are generated for extension by DNA polymerase, variously proposed to be RNA processing by endonuclease MRP (17) or protein-independent termination at one of the conserved sequence blocks of the NCR (18). The exact site of replication initiation is also unclear, and may vary between cell-types. A prominent cluster of 5' ends in H-strand DNA, designated as O_H, is generally regarded as the major origin of (unidirectional) replication. However, there is no direct experimental evidence that it functions thus, and bidirectional initiation clearly occurs in some molecules at sites downstream of O_H, both in cultured cells (19), especially when recovering from drug-induced mtDNA depletion (20), and in solid tissues (21,22). In a minority of molecules (21) these initiation sites can encompass the entire genome (in birds) or almost the entire genome (in mammals).

Mitochondrial DNA (mtDNA) was for over 25 years assumed to replicate by a unique, strand-asynchronous mechanism (23). However, more recent analysis of mtDNA replication intermediates (RIs) by two-dimensional neutral agarose gel electrophoresis (2DNAGE) failed to detect the extensively single-stranded products of such a replication mechanism (19–22,24–26) and instead revealed two classes of double-stranded RIs. One class consists of the predicted products of conventional strand-coupled replication (19,21,24,25); the other contains extended RNA segments (26) encompassing the entire lagging strand (RITOLS, ‘RNA incorporation throughout the lagging strand’, 19). Maturation of the lagging strand to DNA appears to occur with different kinetics and distinct sites of initiation in different organisms (19), and some RIs of the first class could be interpreted as molecules in which lagging-strand DNA synthesis has effectively caught up with the advancing fork, as a result of replication pausing. The mechanism by which the RNA lagging strand is created is unknown.

mtDNA replication also depends on the HMG-box protein TFAM, named for its essential role as a cofactor for efficient and specific transcriptional initiation. TFAM is required for mtDNA maintenance (27) and appears to have several distinct roles in mtDNA metabolism. It is a major structural protein of the mitochondrial chromosome, but also influences mtDNA replication in ways connected with transcription. Overexpression of TFAM leads to a drop in mitochondrial transcript levels and a pronounced shift toward conventional, strand-coupled RIs (28). This could represent either a general slowing or stuttering of fork advance, attributable to a decreased availability of RNA to form the lagging strand (i.e. in which lagging-strand maturation frequently catches up with fork advance) or else a programmed switch to standard DNA synthesis.

As a factor affecting the outcome of mitochondrial transcription, mTERF might be expected also to have some influence over mtDNA replication if, as suggested, replication is intimately connected with transcription. We therefore embarked on a series of experiments to document the effects on mtDNA replication of modulating the expression of mTERF in cultured human cells. To this end, we set out initially to characterize better the binding specificity of mTERF, especially given recent reports of possible additional binding sites for mTERF in vivo (6,29). Electrophoretic mobility shift assays (EMSA) using mitochondrial protein extracts from cells induced to overexpress mTERF revealed additional, though weaker mTERF-binding sites clustered in strategically important regions of the mitochondrial genome. 2DNAGE and lagging strand 5' end mapping by ligation-mediated PCR (LM-PCR) identified the canonical mTERF-binding site as a replication pause site, with the frequency of pausing subject to the expression level of mTERF. Replication pausing in other regions of the genome, notably the NCR, was also affected by mTERF overexpression. These results indicate a role for mTERF in mtDNA replication, in addition to its role in transcription.

MATERIALS AND METHODS

Cells and cell culture

Human embryonic kidney-derived HEK293T cells, Flp-In<TM> T-Rex<TM>-293 cells (Invitrogen), 143B osteosarcoma, Jurkat and HeLa cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM, Sigma) containing 4.5 g/l of D-glucose, 10% foetal calf serum (Sigma), 50 μg/ml uridine (Sigma) and 2 mM L-glutamine (BioWhittaker/Cambrex) at 37°C in an incubator with 5% CO₂ in air. Flp-In<TM> T-Rex<TM>-293-derived cell-lines were cultured under selection with blasticidin and hygromycin according to the manufacturer’s protocol. Expression of mTERF or mTERF-MycHis (C-terminally tagged) was induced in transfected Flp-In<TM> T-Rex<TM>-293 cells with 10 ng/ml doxycyclin (Sigma-Aldrich), which was replenished every 48 h. Cells were passaged routinely.
every 3–4 days at 1:10 or 1:20 dilution. Adherent cells were detached either by pipetting alone or, for HeLa cells, by treatment with Trypsin-EDTA (Bio-Whittaker/Cambrex). Suspension cells (Jurkat) were passaged by centrifugation and resuspension in fresh medium.

**Oligonucleotides and plasmids**

Oligonucleotides used to create EMSA or hybridization probes by PCR from purified human mtDNA or cloned segments thereof are listed in Supplementary Table I. The mTERF coding region, including the mitochondrial targeting signal and 24 nt of the 5’ untranslated region, a total of 1221 bp, was amplified from HeLa cell cDNA (30) using the following chimeric primer pairs (all sequences shown 5’ to 3’, restriction sites used for cloning underlined): BamHI-mTERF F1, CGCGATCCTGCCTTCGACGCTTTCTGG plus HindIII-mTERF R1; CCCAAAGCTTCCAGAATCTGCATTAACTTTTT to create an in-frame C-terminal fusion to the Myc epitope tag; BamHI-mTERF F1 plus HindIII-mTERF R STOP, CCCAAGCTTCCAGAATCTGCATTAACTTTTT to create an mTERF expression construct containing the stop codon at the natural position (shown in italics, underlined, complementary strand). After sequence verification PCR products were digested with BamHI and HindIII (Fermentas, manufacturer’s recommendations) and ligated to similarly digested pcDNA3.1(-)/MycoHis A (Invitrogen) vector DNA to create the mTERF and mTERF-MycHis expression constructs. For induction expression using the Flp-In™ T-REx™ system these plasmids were digested with PmeI (New England Biolabs), which cuts on either side of the insert, ligated into PmeI-digested DNA of the vector pcDNA5/FRT/TO (Invitrogen), and stably transfected into the recipient cells as previously (31).

**DNA and siRNA transfections**

HEK293T cells were transfected either with 3 μg of plasmid DNA and 30 μl of LipofectAMINE (Invitrogen) or 10 μg of DNA and 40 μl of TransFectin™ Lipid Reagent (Bio-Rad), according to manufacturers’ protocols. Transiently transfected cells were subsequently harvested for different assays, or placed under selection using 2 mg/ml G418 Sulfate (Calbiochem), in order to select clones of stably transfected cells expressing mTERF-MycHis. mTERF-specific siRNAs were synthesized using the Silencer™ siRNA construction kit (Ambion). Candidate target sites for specific mTERF silencing were chosen using a prediction programme provided by Ambion (http://www.ambion.com/techlib/misc/siRNA_finder.html). One out of five tested siRNAs was found to be efficient in mTERF silencing (see Results section), the relevant target site in mTERF mRNA being nt 585–605 (5’-AAGCAGGGUGAAACGCUAAAU-3’). To knockdown mTERF expression, HEK293T cells (with or without prior stable transfection with the mTERF-MycHis expression construct) were transfected with 10 nM (final concentration) of mTERF-specific siRNA molecules using Lipofectamine™ 2000 transfection reagent (Invitrogen), as per manufacturer’s recommendations. An siRNA reagent targeted on 5’-GGAGAAGGGUACAGGGGC AU-3’ (siRNA Control) was used as a negative control.

**Immunocytochemistry**

For immunocytochemistry cells were grown on coverslips, seeded at low density. Twenty-four hours after transfection or induction with doxycyclin, cells were washed with DMEM and then incubated in fresh medium containing 100 mM MitoTracker® Red CMXRos (Molecular Probes) at 37°C for 10 min, then washed twice with PBS. After incubation in fresh medium at 37°C for 2 h, cells were again washed twice with PBS and fixed in 4% formaldehyde/5% sucrose in PBS at 37°C for 15 min. After three further PBS washes cells were permeabilized by incubation in 0.5% Triton X-100 in PBS at 37°C for 15 min, washed twice with PBS, incubated in blocking solution (5% w/v non-fat milk powder in PBS) at room temperature for 45 min, then again washed three times with PBS. After incubation in primary antibody solution, mouse anti-Myc monoclonal 9E10 (Roche Molecular Biochemicals, stock 5 mg/ml) 1:1000 in PBS for 1 h at room temperature, cells were washed three times with PBS, then incubated for 1 h at room temperature in a 1:200 dilution of secondary antibody, fluorescein-conjugated horse anti-mouse IgG (Vector Technologies, stock 1.5 mg/ml). After three final PBS washes the coverslips were mounted on slides using Vectashield with DAPI (Vector Technologies). Cells were visualized and photographed using an Olympus IX70 inverted confocal microscope at 100× magnification, with excitation at 568 nm (emission 607/45) for MitoTracker Red and 488 nm (emission 525/50) for fluorescein, using an Andor iXon DV885 front-illuminated CCD camera.

**Western blotting**

SDS–PAGE used 12% polyacrylamide gels under standard conditions (32). Protein extraction and western blotting were carried out essentially as described previously (30). Primary antibodies used were mouse anti-Myc monoclonal 9E10 (as above, diluted 1:15000) and rabbit anti-human mTERF antibody, custom-supplied (Invitrogen) as an anti-peptide (KLH-conjugated CSNDYARRSYANIKE) antibody, 1 mg/ml, diluted 1:5000. Kodak Biomax™ ML X-ray film was exposed to the filter membrane for between 5 s and 5 min.

**Preparation of mitochondrial lysates**

Cells were harvested without trypsinization, resuspended in 1 ml (per 10 cm plate of cells) of resuspension buffer (0.133 M NaCl, 5 mM KCl, 0.7 mM Na 2HPO 4, 25 mM Tris–HCl pH 7.5) and centrifuged at 1200g max for 2 min at 4°C. The pellet was resuspended in 0.5 ml of swelling solution (10 mM NaCl, 1.5 mM CaCl 2, 10 mM Tris–HCl pH 7.5) and incubated on ice for 15 min. After swelling, the cells were dounce-homogenized (20–25 strokes, tight-fitting pestle) on ice and breakage of the cells was checked microscopically. An equal volume of sterile filtered sucrose/EDTA buffer (0.68 M sucrose, 2 mM EDTA, 20 mM Tris–HCl pH 7.5) was added immediately after
breaking the cells. Nuclei and debris were pelleted by centrifugation at 1200 g_max for 10 min at 4°C. The supernatant was transferred to a fresh tube and centrifugation was repeated. The supernatant was collected and recentrifuged at 16 000 g_max for 30 min at 4°C. The mitochondrial pellet was washed once with 200 µl of PBS and frozen at −80°C or lysed immediately. For processing large quantities of cells the volumes were scaled up. Mitochondrial lysates were prepared essentially as described by Fernandez-Silva et al. (33), except using ‘Complete, Mini protease inhibitor cocktail’ (Roche) instead of PMSF.

**EMSA**

DNA fragments for EMSA were PCR amplified using mtDNA as template and primer pairs shown in Supplementary Table 1, followed by sequence verification of the product. dsDNA oligonucleotide probes for EMSA (Supplementary Table 1) were prepared by mixing equal amounts of complementary oligonucleotide pairs in 500 µl of H_2O to a final concentration of 2 mM, followed by incubation for 5 min at 100°C and cooling to room temperature on the bench. Total of 300 ng of each PCR fragment or 20 pmol of each dsDNA oligonucleotide were labelled using 8 U of T4 polynucleotide kinase (Fermentas) and 15 µCi of [γ-^32P] ATP (Amersham Pharmacia Biotech, 3000 Ci/mmol) in 15 µl final volume of PNK buffer (MBI Fermentas). Reactions were stopped on ice and diluted to 100 µl with H_2O. EMSA was carried out in 20 µl binding reactions according to Fernandez-Silva et al. (33) with minor modifications. Reactions contained at least 10 µl of the binding buffer (25 mM HEPES-KOH, pH 7.5, 12.5 mM MgCl_2, 20% glycerol, 0.1% Tween-20, 1 mM DTT), 0.2 pmol of labelled dsDNA oligonucleotide or 3 ng of labelled PCR product as probe, 5 µg of mitochondrial lysate, 100 mM KCl, 5 µg BSA and 5 µg of non-specific competitor DNA poly(dI-dC)-(dI-dC) (Amersham Pharmacia Biotech). Reactions were incubated at room temperature for 20 min and terminated on ice with addition of 0.25 volumes of 30% glycerol. Competition EMSA reactions contained also up to 100-fold excess of the non-labelled competing probe. Supershift EMSA reactions contained 0.5 µg of anti-Myc antibody (as above), or 1 µg of anti-FLAG® M2 antibody (Sigma), which was added 30 min prior to the labelled probe. Depending on the length of the fragment, reaction products were analysed on 5–10% non-denaturing polyacrylamide TBE gels, pre-run at 4°C in 2.2× TBE at 100 V for 1 h at 4°C, then run at 100 V for 30 min and 175 V for 3–5 h depending on the size of the probe fragment. Gels were dried and autoradiographed using KODAK BioMax™ MS film.

**SELEX**

Creation of a randomized DNA ligand library was carried out essentially as described by Blackwell (34). The 46 nt long oligonucleotide template contained 14 internal random nucleotides, flanked on either side by 16 nt fixed ends corresponding with standard primers, containing recognition sites for BamHI and EcoRI, respectively. Second-strand synthesis was carried out in a reaction volume of 20 µl containing 1.6 µg of template, 500 µmol of primer, 2 mM dNTPs and 5 U of Klenow fragment (Fermentas) in Klenow fragment buffer at 46°C for 1 min, followed by 37°C for 7.5 min. The reaction was stopped by heating at 75°C for 10 min and the dsDNA ligand library was gel-purified from an EtBr-stained 14% native polyacrylamide gel using the QIAEX kit (QIAGEN) according to the manufacturer’s protocol. Ligand selection was carried out in 25 µl reactions under essentially the same conditions as EMSA, using 10 µg of mitochondrial protein lysate from mTERF-MycHis expressing Flp-In™ T-Rex™-293-cells, 0.8 µg of the ligand DNA and 6.25 µg of non-specific competitor DNA poly(dI-dC)-(dI-dC) incubated for 20 min at room temperature. Pre-swollen anti-myc-Sepharose beads (Amersham Biosciences) were suspended in EMSA buffer (25 mM HEPES-KOH, pH 7.5, 12.5 mM MgCl_2, 20% glycerol, 0.1% Tween-20, 1 mM DTT, 100 mM KCl, 0.2 µg/µl BSA), washed once in the same buffer and resuspended in 1.5 volumes of the same buffer containing 0.25 µg/µl poly(dI-dC)-(dI-dC). To each binding reaction was added 100 µl of the bead suspension, followed by gentle rotation for 2 h at 4°C. Beads were then washed in EMSA buffer containing 0.25 µg/µl poly(dI-dC)-(dI-dC), followed by a further seven times in the buffer without poly(dI-dC)-(dI-dC) and gentle rotation overnight in 100 µl of K buffer (10 mM Tris–HCl, 0.5 mM EDTA, 50 mM NaCl, pH 8.0) containing 100 µg/ml of freshly dissolved proteinase K (Fermentas). DNA was recovered from the beads by extraction with phenol–chloroform–isoamyl alcohol (25:24:1) and ethanol precipitation, washed once with 70% ethanol and resuspended in a minimal volume of H_2O (~7.5 µl). PCR was then carried out using 2 µl of this template in a 50 µl reaction volume containing 0.2 µM of each SELEX primer (GTTGAAT TCGCTCACG and GAACGGATCCTTCTCG, both shown 5’ to 3’, with restriction sites for cloning underlined) and 2.5 U of Pfu DNA polymerase (Promega). Thirty amplification cycles were carried out using a 15s extension step, after which the enriched ligand DNA was gel-purified from an EtBr-stained 12% native polyacrylamide gel as above. After seven such enrichment cycles, the ligand DNA was cloned into pCR®4Blunt-TOPO® vector (Invitrogen) and individual clones were sequenced using standard primers on an ABI 3100 sequencer using the BigDye® Terminator v3.1 cycle sequencing kit (Applied Biosystems).

**DNA extraction and mtDNA copy number estimation**

For the preparation of mtDNA (mitochondrial nucleic acids) for analysis of RIs from cultured cells, mitochondria were isolated and processed as described by Pohjoismäki et al. (28). Total DNA for analysis of mtDNA copy number was extracted from cells by standard methods (35), and copy number was determined using quantitative PCR, as described previously (28), with amyloid precursor protein (APP) as a single-copy nuclear DNA standard. Human placental mtDNA was prepared as previously (24). Total DNA for analysis of mtDNA RIs
was extracted from frozen human tissue blocks ~7 × 7 × 7 mm³ obtained via forensic autopsies. The samples were taken as part of the Tampere Coronary Study, approved by the Ethics Committee of Tampere University Hospital (DNO 1239/32/200/01) and the National Authority for Medicolegal Affairs. Heart, brain, skeletal muscle and kidney tissue samples were cut into thin slices with a sterile blade and suspended in 2 ml DNA extraction buffer (28). One-tenth volume of 10% SDS and 0.5 mg proteinase K were added. The crude homogenate was passed several times through a 5 ml pipette tip with a sawn-off end, to disperse the larger tissue fragments. The homogenate was incubated overnight with gentle swirling at 37°C. After incubation, 2 volumes of phenol–chloroform–isoamyl alcohol (25:24:1) were added, and the mixture was shaken gently for 1 h. The mixture was then transferred to Eppendorf 15 ml Phase Lock Gel™ Heavy tubes and centrifuged at 5000 g for 15 min. The aqueous phase was recovered and the extraction step repeated. DNA was precipitated by the addition of 0.2 volumes of 10 M ammonium acetate and 2 volumes of ethanol. The mixture was incubated on ice for 10 min and DNA was spooled out using a glass rod, washed once with 70% ethanol, air dried gently and resuspended in 300–700 µl TE buffer (10 mM Tris–HCl, 1 mM EDTA, pH 8.0), depending on the pellet size. 2DNAGE analysis used 10 µg aliquots of heart and brain DNA and 20 µg aliquots of kidney and skeletal muscle DNA.

**Two-dimensional neutral agarose gel electrophoresis**

One microgram of total mitochondrial nucleic acids was used per analysis. Restriction digestions were performed following manufacturers’ recommendations, except for BglI which was carried out at 37°C for double the usual reaction time. If subsequent treatment with S1 nuclease was used, DNA was first recovered by ethanol precipitation and resuspended in the appropriate reaction buffer, before treatment with 50 U S1 Nuclease (Promega) for 30 s. Reactions were stopped by the addition of an equal volume of phenol–chloroform–isoamyl alcohol (25:24:1, pH 8.0) and immediately extracted. 2DNAGE was performed as described previously [(28), note different gel conditions for fragments in different size classes].

**Radiolabelled probes and blot hybridization**

For Southern hybridization, probes were created by Pfu-PCR, using cloned segments of human mtDNA as template (see Supplementary Table 1), and subsequently sequenced to confirm their identity. Probes were labelled using Rediprime™ II random prime labelling kit (Amersham) and [α-³²P] dCTP (Amersham; 3000 Ci/mmol).

**LM-PCR**

LM-PCR was carried out as described by Yasukawa et al. (20), using oligonucleotide primer sets as indicated in figure legends and as detailed in Supplementary Table 2.

Mitochondrial DNA immunoprecipitation

Cells were processed for mitochondrial DNA immunoprecipitation (mIP) essentially as described by Lu et al. (36). The mtDNA was sheared to fragments of average size 500–600 bp using a Sonics Vibra-Cell sonicator, 3 mm tip at 25% power for 3 × 20 s (1 s on, 1 s off) with incubation on ice for 30 s between. Complete, Mini protease inhibitor cocktail (Roche) was included in the lysis buffer. Lysates were pre-cleared with pre-swollen Protein A Sepharose (Amersham Biosciences) and immunoprecipitations were carried out with 5 µg mouse anti-Myc monoclonal 9E10 (Roche Molecular Biochemicals) or anti-FLAG® M2 antibody (Sigma) overnight at 4°C. Final PCR reactions used primers listed in Supplementary Table 1 and the minimum number of amplification cycles required to generate substantial product bands from the input DNAs (generally 25–26 cycles, depending on the fragment, based on preliminary tests), thus avoiding saturation.

**RESULTS**

**mTERF has multiple binding sites in the human mitochondrial genome**

In order to study the effects of mTERF on mtDNA replication, we established HEK293-derived cells expressing both natural mTERF and C-terminally Myc epitope-tagged mTERF. Mitochondrial targeting was verified by immunocytochemistry of transiently transfected HEK293T cells expressing mTERF-MycHis (Figure 1a). Induction of protein expression in Flp-In™ T-Rex™-293 cells stably transfected with the mTERF or mTERF-MycHis constructs was verified by western blotting (Figure 1b). Protein levels were the same after 24 or 48 h of induction. Prolonged overexpression of mTERF (6 days) had no significant effect on mtDNA copy number as estimated by Southern blotting (data not shown) or by quantitative PCR (Supplementary Figure 1a).

Since the main aim of the study was to determine the effects of altered mTERF expression on mtDNA replication *in vivo*, we first tested the effects of mTERF expression on protein binding to mitochondrial DNA, using EMSA with mitochondrial protein extracts from cells overexpressing mTERF. In contrast to earlier studies using purified, bacterially expressed mTERF, this tests the effects of altered mTERF expression level on protein–DNA interactions in the mitochondrial milieu, in which other mitochondrial proteins, including TFAM, are present and may influence binding.

Using EMSA with probes covering the previously identified, canonical mTERF-binding site in the tRNA⁴⁰Leu(UUR) gene, we confirmed that overexpression of natural mTERF, whether by transient or stable transfection (data not shown), or under tetracycline induction (Figure 1c), leads to a large increase in sequence-specific binding activity. The protein complex formed from the Myc epitope-tagged protein migrated slightly slower than the complex formed by endogenous
or overexpressed natural mTERF, and was supershifted by an anti-Myc monoclonal antibody (Figure 1c), but not by other antibodies (e.g. anti-FLAG, Supplementary Figure 1b). The anti-Myc antibody did not supershift the complex formed by endogenous or overexpressed natural mTERF (Supplementary Figure 1b).

These properties next allowed us to test other regions of the mitochondrial genome for specific binding of mTERF to DNA, using EMSA. Using overlapping fragments of ~150 bp, we scanned the major NCR and its flanking sequences, the minor NCR (O_L), its surrounding tRNA gene cluster, the region extending from O_L to the canonical mTERF-binding site in the tRNA_{Leu(UUR)} gene, the ATPase 6 gene and its junction with the COIII gene and several other segments of the genome (Figure 2, Supplementary Figures 2 and 3). We estimated relative binding affinities using competition EMSA against the tRNA_{Leu(UUR)} gene fragment and vice versa. As shown in Figure 2 and Supplementary Figure 2, we identified a cluster of four moderately strong mTERF-binding sites within the ND1 coding sequence and the adjacent IQM tRNA gene cluster (see Figure 2e for summary). Competition EMSA indicated that the binding to fragment ND1.1 (Figure 2b) was between one and two orders of magnitude weaker than to the canonical binding site in tRNA_{Leu(UUR)}. Binding to the ND1.1 fragment was tested further, using shorter, overlapping fragments (Supplementary Figure 2d). The results suggest that fragment ND1.1 contains two distinct binding sites. We also identified a binding site adjacent to O_H (fragment OH1) at least two orders of magnitude weaker than the canonical binding site, based on competition EMSA data (data not shown), as well as four other binding sites in the D-loop portion of the NCR and one at O_L, plus a possible site at the HSP (fragment OH5, see Supplementary Figure 2e).

Alignment of the sequences of these binding sites suggested a consensus which was verified by SELEX (Table 1). Most of the SELEX output clones analysed (82/109) contained at least one match to the consensus TGGT or TYGGT, and 43 clones showed an identical or almost identical (8/9) match to the extended consensus TGGT(N8)TYGGT (or its complement). Of 28 control clones analysed, subjected to the same number of amplification cycles but without antibody selection, none matched this consensus. Comparing the SELEX consensus with the canonical mTERF-binding site in the tRNA_{Leu(UUR)} gene, and with the findings of an earlier application of PCR-based selection on a smaller scale using only EMSA (37), the invariant features of the binding site would appear to be two pairs of G residues on the same strand, separated by eight nucleotides (see also Supplementary Figure 2e).

In order to verify that mTERF is able to bind to at least some of its non-canonical binding sites in vivo, we carried out semi-quantitative mIP, using a minor adaptation of the method recently published by Lu et al. (36). For this assay we used cells inducibly expressing mTERF-MycHis, and carried out immunoprecipitation using anti-Myc antibody, as well as a control antibody (anti-FLAG) or no antibody.

**Figure 1.** Overexpression of mTERF in cultured cells. (A) Immunocytochemistry of HEK293T cells transiently or stably transfected with mTERF-MycHis, using anti-Myc monoclonal antibody, counterstained with Mitotracker Red. (B) Western blots of mitochondrial protein extracts from Flp-In™ T-REx™-293 cells transfected with the mTERF or mTERF-MycHis constructs and induced for expression as indicated (0, 24, 48 h) or from transiently transfected (t) HEK293T cells, probed with anti-Myc or anti-mTERF antibodies, as indicated. The endogenous mTERF protein detected by the anti-mTERF antibody is singly arrowed. The mTERF-MycHis fusion protein detected by the same antibody is indicated by a double arrow. (C) EMSA using Leu-short dsDNA oligonucleotide probe and mitochondrial protein extracts from Flp-In™ T-REx™-293 cells transfected with the mTERF or mTERF-MycHis constructs and induced for expression as indicated. EMISA was carried out with or without anti-Myc antibody as shown (left-hand panel), or (right-hand panel) in the presence of an increasing amount of cold Leu-short dsDNA oligonucleotide competitor (1-, 10-, 100- and 1000-fold mass excess) or without competitor (−). The free probe (F), complexes formed by natural mTERF (B_N) or the mTERF-MycHis fusion protein (B_F), and the antibody-supershifted complex (S) are indicated. See also Supplementary Figure 1.
Figure 2. EMSA and mIP analysis of alternate mTERF-binding sites in human mtDNA. (A) Schematic diagram of regions of the mitochondrial genome in which binding was detected, showing NCR (white box), 16S and 12S rRNA genes (pale grey boxes), protein-coding genes ND1, ND2, COI and cyt b (darker grey boxes), tRNA genes (cross-hatched boxes), O₁, O₂ and the promoters/transcriptional initiation sites of the two strands (P₁, P₂ and P₃). Genes transcribed to the right shown above the centre line, genes transcribed to the left shown below. Nucleotide coordinates are as Ref. (82). Black bars indicate the positions of the 150 bp probe fragments which were found by EMSA to contain strong or moderate binding sites for mTERF, as shown in panels b and c. (B) Competition EMSA using the probes and competitors as shown, plus mitochondrial protein extract from cells induced to express mTERF-MycHis. The amounts of cold competitor represent 1-, 10- and 100-fold mass excess over the probe. Similar results were obtained using extracts from cells overexpressing natural mTERF (data not shown). (C) EMSA analyses of binding to 150 bp probe fragments as indicated, using mitochondrial protein extracts from Flp-In T-REx cells transfected either with natural mTERF or with mTERF-MycHis (mTERF-mh) and induced for expression (or not) as indicated. Supershifting with the anti-Myc monoclonal antibody was performed for the lanes indicated. Supershifted complexes are denoted by arrows. Although the supershifted complex is minor in some cases, the main complex is always efficiently removed by the antibody, confirming the presence of mTERF-MycHis. Other antibodies tested (e.g. anti-FLAG) gave no supershifting and did not inhibit the formation of these complexes. For further experiments confirming specificity of binding and negative/weak findings using other fragments, see Supplementary Figures 1 and 2. (D) mIP analysis of mTERF-MycHis binding in vivo. Immunoprecipitation used anti-Myc (M), anti-FLAG (F) or no antibody (–). Amplification of immunoprecipitates alongside corresponding input DNAs used the same primer pairs as were employed to generate the corresponding fragments for EMSA (see Supplementary Table 1). Samples were from Flp-In T-REx T-REx-293 cells induced for mTERF-MycHis expression, except for fragment Leu, where extracts from uninduced cells were also tested.
Induction of mTERF-MycHis expression enabled immunoprecipitation of several key fragments of the mitochondrial genome in which binding was found in vitro (Figure 2d). The fragment containing the canonical mTERF-binding site in the tRNA_{Leu(UUR)} gene was routinely detected in the anti-Myc immunoprecipitate from induced cells, but was not immunoprecipitated by control antibody (anti-FLAG) or no antibody. Immunoprecipitates from uninduced cells were negative under comparable conditions, but using excess anti-Myc antibody we sometimes observed weak amplification of this fragment (data not shown), consistent with a low level of leaky expression of the mTERF-MycHis transgene and the high affinity of the protein for the canonical binding site. Consistent positive signals were also seen in the anti-Myc immunoprecipitate from induced cells, but not control immunoprecipitates, for the HSP-containing fragment OH5 and for the three D-loop fragments (NCR1, NCR5 and OH1) which gave the strongest EMSA signals in vitro (Figure 2c). The ND1.1 fragment internal to the ND1 coding sequence was also weakly amplified from anti-Myc immunoprecipitates from induced cells (Figure 2d). Fragments from the ND3 gene (e.g. ND3.4), or others which were negative for binding in vitro using EMSA, gave either very faint signals or no signal at all after immunoprecipitation. Overall, these findings are consistent with the proposition that mTERF, when overexpressed, can bind in vivo to specific, non-canonical binding sites, which correspond with binding sites detected in vitro.

Replication pause sites map close to sites of mTERF binding in human mtDNA

In previous studies using 2DNAGE we noted the occurrence of a number of stereotypic pause sites in mitochondrial DNA of both sea urchins (38) and vertebrates (21,24). In sea urchins, pause sites correspond with sites of specific protein binding (39–41). We therefore considered the hypothesis that some of the replication pause sites in human mtDNA may map to locations of mTERF binding.

We initially analysed the region of the genome in which the canonical mTERF-binding site in the tRNA_{Leu(UUR)} gene is located. 2DNAGE analysis of the PvuII–AccI fragment covering this site, extending from OH1 into the rDNA, in several different cell lines and tissues (Figure 3), revealed a number of pause sites of varying prominence. To visualize their positions more clearly we treated parallel samples with S1 nuclease, thus digesting partially degraded RITOLS intermediates, including any attached RNA tails. The two epithelia-derived cell-lines, HEK293T and HeLa, gave very similar patterns, with a clear, though relatively weak pause site signal in the region of the tRNA_{Leu(UUR)} gene (designated ‘a’ in Figure 3b), a second, more prominent pause located in the 3’ part of the ND1 gene or in the adjacent IQM tRNA cluster (designated ‘b’), a third, near O1 (designated ‘d’), and accumulated material in a broad region of ND2 (designated ‘c’).

In 143B (osteosarcoma) and Jurkat (T-cell leukaemia) cells the steady-state abundance of all mtDNA RIs was quantitatively less, though the patterns were qualitatively similar to those seen in HEK293T or HeLa cells. In S1-untreated material the pause sites were poorly resolved, and the descending segment of the Y-arc was very weak. Region ‘c’ was not seen as a discrete species, even after S1 treatment. Following S1 treatment, the ratio of the other pauses differed between cell-types: for example, pause ‘b’ was much more prominent than pause ‘d’ in Jurkat cells, whereas in 143B cells they were at similar abundance. The tRNA_{Leu(UUR)} gene pause ‘a’ was seen clearly in all cell-lines tested.

In tissue samples (Figure 3c), pause ‘a’, near to the canonical mTERF-binding site, was most prominent in the brain, but weak in other tissues tested. Pause ‘d’ was more prominent than pause ‘b’ in heart and brain, but weaker than pause ‘b’ in skeletal muscle and in kidney. Pause ‘c’, was seen only in kidney, where both it (and pause ‘b’) appeared to be even more diffuse than in other tissues or cell-lines. An additional pause site was seen in brain, between ‘c’ and ‘d’ (denoted ‘d’). Pause ‘d’, near O1, was also detected as an extended pause region in human placenta ([24], Supplementary Figure 3).

Pasing near two other sites at which mTERF binding to DNA was seen both in vitro and in vivo (Figure 2), namely OH1 and the TAS region, is already well documented from previous studies, and further examples are seen in Figure 4 (see also Supplementary Figures 3 and 4). Although originally proposed as a unique unidirectional origin, recent data indicate that O1 also functions as a site of fork arrest when bidirectional replication initiates elsewhere, and may thus also be considered as the terminus of replication ([20–22,24], see also Supplementary Figure 3). The TAS region is, by definition, adjacent to the termination site for the synthesis of D-loop 7S DNA.

mTERF overexpression enhances replication pausing in human mtDNA

In order to test whether the level of mTERF expression influences replication pausing we carried out 2DNAGE
Table 1. SELEX analysis of the mTERF-binding site

<table>
<thead>
<tr>
<th>Position</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occurrences&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>30</td>
<td>17</td>
<td>23</td>
<td>21</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>8</td>
<td>43</td>
<td>43</td>
<td>2</td>
<td>4</td>
<td>16</td>
<td>6</td>
<td>16</td>
<td>31</td>
<td>0</td>
<td>43</td>
<td>43</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>5</td>
<td>4</td>
<td>13</td>
<td>3</td>
<td>11</td>
<td>43</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>41</td>
</tr>
<tr>
<td>Consensus&lt;sup&gt;b&lt;/sup&gt;</td>
<td>t</td>
<td>G</td>
<td>G</td>
<td>t</td>
<td>a</td>
<td>r</td>
<td>a</td>
<td>r</td>
<td>g</td>
<td>T</td>
<td>Y</td>
<td>G</td>
<td>G</td>
<td>t</td>
</tr>
</tbody>
</table>

<sup>a</sup>Out of 43 clones analysed which matched a clear consensus (see text).

<sup>b</sup>Nucleotides found in 43/43 clones shown in upper case, others in lower case, Y = pyrimidine, R = purine.

analysis of RIs in mtDNA extracted from cells induced to overexpress mTERF, compared with uninduced cells (Figure 4, Supplementary Figure 4). Within the 3.6 kb PvuII–AccI ND2-containing fragment (Figure 4b), pause site ‘a’ (tRNA<sub>Leu(UUR)</sub>), mapping near the canonical site of mTERF binding, was strongly enhanced by mTERF overexpression, compared with the unit-length restriction fragment (denoted ln in Figure 3d). Pause sites ‘b’ (ND1/ IQM tRNA cluster) and ‘d’ (O<sub>1</sub>) were also enhanced, as was the more diffuse pause region ‘c’. In mTERF-overexpressing cells we also detected a more prominent X-form intermediate (designated ‘x’ in Figure 4b, Supplementary Figure 4a) in restriction fragments (e.g. HincII or AccI) containing the tRNA<sub>Leu(UUR)</sub> gene at a central location.

Within the NCR mTERF overexpression enhanced the abundance of a paused intermediate migrating near or beneath the bubble arc (designated ‘t’ in the HincII fragment and ‘n’ in the AccI fragment, O<sub>1</sub> probe), as well as the arc leading to it from the unit-length fragment (Figure 4c, Supplementary Figure 4b). These forms were sensitive to S1 nuclease (Supplementary Figure 4b) and are probably equivalent to the classical D-loop. mTERF overexpression also appeared to increase the relative amount of 7S DNA as well as introducing subtle alterations to the various forms of mtDNA resolved on 1D gels (Supplementary Figure 4f). mTERF overexpression also diminished the relative abundance of termination intermediates (designated ‘t’ in Figure 4c and d) and increased that of Y-form intermediates in which a single fork appears to have paused when approaching O<sub>1</sub> (designated ‘g’). The distribution of material on the termination arc are also appeared to be subtly different from that of uninduced cells.

In other regions, a prominent pause site (‘h’, Figure 4c), located near to the ND5/ND6 gene boundary, was unaffected by mTERF overexpression, whereas a novel pause was induced in the coding region of ND3 (Supplementary Figure 4c). Note, however, that strong mTERF binding was not found <em>in vitro</em> in either region (Supplementary Figure 2 and other data not shown).

Digestion of mtDNA from mTERF-overexpressing cells with restriction enzymes having only a single recognition sequence in the genome generated 2DNAGE patterns consistent with enhanced pausing in the region of ND1/tRNA<sub>Leu(UUR)</sub> and consequent delayed resolution in the NCR (Figure 4d, Supplementary Figure 4d and e). Note that mTERF overexpression produced subtle, site-specific effects, rather than a general slowing of replication e.g. as would be attributable due to non-specific stalling.

mTERF knockdown diminishes replication pausing in the ND1/tRNA<sub>Leu(UUR)</sub> region

To test whether the modulation of replication pausing resulting from mTERF overexpression represents the signature of a finely tuned physiological process rather than just an overexpression artefact, we downregulated the expression of mTERF by RNA interference. This produced a reciprocal effect on mtDNA replication pausing at the canonical mTERF-binding site. We first tested several different mTERF-directed siRNAs in transient transfection assays, using cells stably transfected with the mTERF-MycHis expression construct, enabling us to evaluate knockdown at the protein level by western blotting (Figure 5a). One particular siRNA (mTERF.1) gave consistently strong knockdown, as judged also by immunocytochemistry on mTERF-MycHis-expressing cells (Figure 5b) and EMSA (Figure 5c). Based crudely on the autoradiographic EMSA signals, functional knockdown of >90% was routinely achieved 48 h after transfection with siRNA mTERF.1. The effects of mTERF knockdown on replication pausing in the ND1/tRNA<sub>Leu(UUR)</sub> region were then studied using 2DNAGE (Figure 5d). Pause site ‘a’ (tRNA<sub>Leu(UUR)</sub>) was no longer detectable, even on long autoradiographic exposure, and the prominence of X-forms was also diminished by mTERF knockdown. The abundance of other pauses was altered less substantially, though the relative amount of species ‘b’ compared with ‘c’ appeared to be decreased.

mTERF overexpression enhances lagging strand 5’ ends near to specific replication pause sites

One signature of increased replication pausing during strand-coupled DNA replication should be the enhancement of persistent, lagging strand 5’ ends adjacent to pause sites (Supplementary Figure 5). We used LM-PCR to map such ends in the vicinity of the major pauses regulated by mTERF, and to determine the effects upon them of mTERF overexpression. Comparing mtDNA from cells overexpressing mTERF with that from uninduced cells, we analysed 5’ ends on the L-strand near the canonical tRNA<sub>Leu(UUR)</sub> binding site, as well as in the whole of ND1 and the adjacent tRNA genes, and also on the H-strand in the NCR. A cluster of L-strand 5’ ends in or adjacent to the tRNA<sub>Leu(UUR)</sub> gene, notably at np 3211,
3234 and 3310, were strongly enhanced during 72 h of induction of mTERF overexpression (Figure 6a, Supplementary Figure 5b). In the vicinity of the IQM tRNA gene cluster, LM-PCR revealed L-strand 5′ ends enhanced by mTERF overexpression at np 4476 (close to the 5′ end of ND2 mRNA) and np 4434 (within the tRNAMet gene), against a background of heterogeneous 5′ ends that were generally unaffected by mTERF overexpression (Figure 6b and c). The significance of this heterogeneous background of abundant 5′ ends is unclear, although the two sites enhanced by mTERF overexpression lie in the vicinity of pause ‘b’. Within the remainder of the ND1 and ND2 coding regions (Supplementary Figure 5c–h) we detected only weak LM-PCR signals which were not affected by mTERF overexpression. The prominent 5′ ends in the tRNA Cys gene adjacent to O_L were also unaffected by mTERF overexpression.

In the NCR, H-strand 5′ ends at O_H, as well as those clustered in the distal region of the D-loop (np 16331, 16337, 16370, 16411 and more weakly at np 16197) were strongly induced by mTERF overexpression. This is consistent with delayed resolution, arising from more frequent pausing at the canonical mTERF-binding site. H-strand 5′ ends in the NCR are on the lagging strand for initiation events outside of the NCR, and thus may also be enhanced by pausing of replication forks entering the NCR from the ‘cytochrome b side’.

**DISCUSSION**

mTERF is a modulator of replication as well as transcription

In this study, we showed that mTERF binding at its canonical binding site in the tRNA^{Leu(UUR)} gene
influences replication pausing near to this site. Overexpression of mTERF enhanced this pausing and increased the steady-state abundance of lagging strand 5' ends adjacent to the binding site, whereas mTERF knockdown by RNA interferences decreased pausing in the tRNA\textsubscript{Leu(UUR)} gene region. In addition, based on studies \textit{in vitro} (EMSA, SELEX) and supported by findings \textit{in vivo} (mIP), we identified novel sites of mTERF binding, elsewhere in the genome. Binding at these sites was weaker than in the tRNA\textsubscript{Leu(UUR)} gene, but replication pausing in these regions was nevertheless influenced by mTERF overexpression. The data support a role for mTERF as a modulator of replication, especially at its canonical binding site. Close parallels with the properties of replication pause-region binding proteins in bacteria and eukaryotic nuclei provide some intriguing hypotheses which we now discuss.

mTERF binding to additional sites in human mtDNA

The present study revealed novel sites of mTERF binding in the NCR and ND1 gene. mIP assays were consistent with binding at least at some of these sites by over-expressed mTERF \textit{in vivo}. mTERF homologues in invertebrates, such as DmTTF in \textit{Drosophila} (12) or DBP in sea urchins (39,41), also have diverse and multiple binding sites, typically demarcating the 3' ends of oppositely transcribed gene clusters. Although mTERF binding to the novel binding sites appeared weaker than at the canonical binding site, many of them are clustered, suggesting that cooperative binding might promote site occupancy \textit{in vivo}, consistent with the results of mIP. The binding we observed \textit{in vivo} might also depend on other mitochondrial nucleoid proteins, including TFAM, as well as possible post-translational modifications.

The effects of mTERF overexpression on mtDNA RIs from the NCR (Figure 4c and d), suggest that mTERF may interact with this region \textit{in vivo} to promote fork arrest at the termination locus. The termination zone for mtDNA replication appears not to be a single point (O\textsubscript{H}), but an extended region of the NCR (Figure 4c). mTERF overexpression resulted in increased stalling of replication forks as they approach O\textsubscript{H} from the ‘cytochrome b side’. It also appeared to elevate the abundance of 7S DNA (Supplementary Figure 4f) and of S1 nuclease-sensitive species probably equivalent to the classical D-loop (Figure 4c), consistent with increased pausing in the TAS region. Protein-binding sites within the TAS region were previously mapped by \textit{in vivo} footprinting (42) and by EMSA (43), and mTERF might be one of the
proteins involved. The D-loop remains enigmatic. DNA synthesis arrest at TAS might be a switching mechanism relevant to copy number control, or may have other purposes, such as mitochondrial nucleoid organization (44).

The *Escherichia coli* Tus protein, which regulates the termination of chromosomal DNA replication, may represent a useful paradigm for mTERF. Tus binds multiple copies of the Ter sequence flanking the terminator region, and acts directionally as a contrahelicase (45) to trap replication forks in this region (46). The sea urchin mTERF homologue DBP has also been shown to function as a contrahelicase *in vitro* (15). Like Tus, mTERF binding to sites on both sides of the replication terminus region might regulate the entry of oppositely moving replication forks into the region, facilitating their orderly synopsis. Increased mTERF expression resulted in elevated levels of persistent H-strand 5'0 ends in the NCR (Figure 6d), an expected signature of delayed resolution if fork passage through rDNA is more restricted. The orientation of potential mTERF-binding sites in the genome appears highly non-random. Taking the simplified sequence GG(N 8)GG as the minimal binding site, its 12 occurrences in the NCR all bear the same orientation. The same applies to the cluster of seven such sites in the 3'0 portion of ND1 and the adjacent tRNA gene cluster. In contrast, the canonical binding site shows the opposite orientation, although is flanked on each side by two oppositely oriented copies of the minimal binding site (Supplementary Figure 6).

![Figure 5. Downregulation of mTERF expression by RNA interference. (A) Western blot assay of mTERF knockdown by siRNA mTERF.1 (directed against mTERF mRNA) and siRNA Control. HEK293T cells were either untransfected (−), transiently transfected (t) or stably (+) transfected with an mTERF-MycHis expression construct. Cells were then assayed 24, 48 and 72 h following siRNA transfection or else without such transfection (−). The arrowed band is the mTERF-MycHis fusion protein, migrating between two background bands which appear in all westerns and thus provide an internal loading control. Note that the sample from untransfected, non-siRNA-treated cells in the upper panel (pentultimate lane) is approximately 3-fold overloaded. (B) Immunocytochemistry of HEK293T cells stably transfected with mTERF-MycHis expression construct and then either mock transfected or transiently transfected with siRNA mTERF.1. Immunocytochemistry used the anti-Myc monoclonal antibody, and counterstaining with Mitotracker Red. (C) EMSA using Leu-short dsDNA oligonucleotide probe and mitochondrial protein extracts from HEK293T cells with or without stable transfection of mTERF-MycHis expression construct, followed by transient transfection for 48 h with or without siRNA mTERF.1. Despite the apparent difference in signal, the experimental conditions are the same as in Figure 1c: only the exposure time is different, and the amount of background signal in the gel. (D) 2DNAGE of mtDNA from untransfected HEK293T cells or cells transfected with siRNA mTERF.1 for 48 h. *Pvu*II + *Acc*I digest (S1 treated) probed for the 3.6 kb fragment using ND2 probe. Panels iii and iv are longer exposures of panels i and ii, respectively. Note the down-regulation of the X-spike (‘x’) and pause site ‘a’, as well as of pause site ‘b’ relative to pause region ‘c’ (see Figure 3). (E) Phosphorimager-calibrated exposures of 2DNAGE blots from siRNA-treated cells (panel iii) alongside the corresponding images (panels i and ii) from uninduced and induced mTERF over-expressing cells, reproduced from Figure 4b. A longer exposure (panel iv) confirms the absence of pause ‘a’.
findings suggest that the site is efficiently bound \textit{in vivo} by overexpressed mTERF. The latter is consistent with previous findings that recombinant mTERF (7,29) binds only weakly to this site, whereas partially purified, endogenous mTERF binds more strongly and establishes a DNA loop required for efficient rDNA transcription (7). This may require a post-translational modification or limiting accessory factor found only \textit{in vivo} (48). The initiator fragment also contains binding sites for TFAM, which may promote binding but may also interfere with the interpretation of the EMSA assay (see Supplementary Figure 2e).

\section*{Regulated passage of replication and transcription complexes}

The role of mTERF as a transcriptional terminator is well established from \textit{in vitro} studies, and DmTTF also functions thus \textit{in vivo} in \textit{Drosophila} (13). However, there is no compelling evidence that mTERF regulates mitochondrial RNA levels physiologically. The disparity in relative abundance between mRNAs and rRNAs in mammalian mitochondria can largely be accounted for by post-transcriptional regulation, notably differences in half-life (49) and RNA processing efficiency (50). Despite causing reduced mTERF-binding affinity and terminator activity \textit{in vitro}, the 3243A>G MELAS mutation has almost no effect on mitochondrial RNA levels \textit{in vivo} (5). Moreover, manipulation of mTERF levels \textit{in vivo} by overexpression or RNA interference has remarkably little effect on steady-state mitochondrial RNA levels (Hyvärinen \textit{et al}., manuscript in preparation). The observation that mTERF also modulates mtDNA replication pausing suggests a different physiological meaning for its action as a transcriptional terminator, i.e. it coregulates replication and transcription.
The unregulated collision of oppositely moving transcription and replication complexes drastically inhibits DNA replication and provokes genomic instability in both bacteria (51,52) and yeast (53,54). In *E. coli*, head-on collision of the transcription and replication machineries severely impedes the progress of the replication fork (51) whereas codirectional transcription has no effect. Within the *E. coli* chromosome almost all essential genes are oriented such that transcription and replication are codirectional (55), which is proposed to minimize the mutagenic effect of repeated replication stalling and recombinational restart, following head-on collisions (51,56). In bacterial or yeast plasmids, or yeast rDNA, such head-on collisions can trigger genomic instability (53,57,58), e.g. due to knotting of daughter duplexes (59). In mammalian nuclei, head-on collisions can trigger the formation of HSRs (homogeneously staining regions of chromosomes), the signature of massive gene amplification events (60).

Proteins with dual roles in replication and transcriptional arrest are well documented. The *E. coli* Tus protein, described in the preceding section, preferentially blocks transcription with a similar polarity as DNA replication (61,62). Passage of a transcription complex from the permissive direction relieves the block on DNA synthesis (61) by provoking the dissociation of bound Tus (63). The mouse TTF-I protein binds at the 3' end of the rDNA transcription unit, where it terminates transcription by RNA polymerase I (64) and arrests replication forks arriving from the other direction (65,66), via its polar contrahelicase activity (67). This organization of the rDNA locus is relatively conserved throughout eukaryotes, although in some species the TTF-I homologue co-operates with or depends upon other proteins to maintain the replication fork barrier (RFB), including Sap1p and Reb1p in *Schizosaccharomyces pombe* (68–70), with involvement of Swi1p and Swi3p to stabilize the stalled forks (71), or Fob1p in *Saccharomyces cerevisiae* (72,73), with Sir2p regulating recombination at the stalled forks (74).

The entry of replication forks into the mtDNA termination zone around *OH* requires traversal of the heavily transcribed rDNA region in the antisense direction, with potentially catastrophic consequences if a transcription complex is encountered. By binding at the rDNA boundary, mTERF may thus serve a function related to those of both Tus and TTF-I, facilitating the regulated passage of oppositely moving transcription and replication machineries, and regulating fork access to the termination zone.

The passage of a transcriptional complex in the permissive direction may also serve a regulatory role, such as hypothesized for mtDBP in relieving the block on D-loop expansion in sea urchins (15).

mTERF overexpression enhanced X-like species in fragments where the canonical mTERF-binding site was centrally located (Figure 4b, Supplementary Figure 4a), whereas mTERF knockdown by RNA interference depleted such species (Figure 5d). Although these may be recombination intermediates (see below), they might also comprise termination complexes centred on the mTERF-binding site. If increased mTERF activity enhances rDNA transcription (6), it may also restrict the entry of replication forks into rDNA in the antisense direction and perhaps even shift the resolution site in some molecules from the NCR to the tRNA<sub>Leu(UUR)</sub> gene. However, this must be a very minor fraction of molecules, since we did not see a complete Y-arc in *OH*-containing fragments.

**Transcription termination and the bootlace model**

Previous studies of vertebrate mtDNA replication indicated that, in the majority of molecules, the lagging strand is initially laid down in the form of extended RNA segments which are subsequently converted to DNA via a maturation step (19). We hypothesized that the RNA lagging-strand may arise via either of two highly unorthodox mechanisms: either via a primase capable of synthesizing extended RNA primers, or by the hybridization of preformed L-strand RNA with the displaced H-strand in a 3' to 5' direction as the replication fork advances, the so-called bootlace model.

mTERF-dependent replication pausing may be construed as circumstantial evidence supporting the bootlace model. Transcriptional termination by mTERF adjacent to a paused replication complex would provide a 3' end capable of priming lagging-strand DNA synthesis, at the same time as delivering a fresh RNA bootlace to enable the replication fork to proceed in the forward direction. The mTERF-dependent enhancement of lagging strand 5' DNA ends near the canonical (and some other) mTERF-binding sites, (Figure 6), supports this idea. In sea urchin mtDNA, the major replication pause-region, which interacts with at least two DNA-binding proteins (38,40,41), also appears to be a major lagging-strand origin, as well as a site of transcriptional termination and/or RNA processing.

The effect of overexpression of mTERF is much more site-specific than that produced by overexpression of TFAM (28), by treatment with mtDNA replication inhibitors such as dideoxycytidine (28), or by expression of dominant-negative versions of the mtDNA helicase Twinkle (31). Unlike these treatments, mTERF overexpression did not cause a general slowing of replication, and did not alter globally the ratio of strand-coupled and RITOLS type RIs. By facilitating lagging-strand maturation, mTERF may serve merely to minimize the extent of the region of mtDNA maintained in the more vulnerable RNA-DNA hybrid form, thus contributing to genome stability.

**Recombination at the mTERF-binding site**

An alternate interpretation of the X-like molecules centred on the tRNA<sub>Leu(UUR)</sub> gene region, which were enhanced by mTERF overexpression and depleted by mTERF knockdown, is that they represent true recombination intermediates. Such forms would be expected to arise if persistent mTERF binding and consequent prolonged pausing entrain fork collapse, requiring either a double-strand break or fork regression to generate a recombination genomic end for restart of replication. We previously
noted that the canonical mTERF-binding site was a frequent break-point in rearranged mtDNA molecules (‘sublimons’) detectable at a low level in all cell-types, but especially prominent in human heart (75), and also in mice expressing a disease-equivalent version of the Twinkle helicase(76) associated with autosomal dominant external ophthalmoplegia (PEO). It is tempting to ascribe such molecules to aberrant recombination following pausing and fork collapse at the rRNA^Leu(UUR) gene.

In yeast, double-strand breaks at the site of the RFB in rDNA, giving rise to low-level genomic rearrangements implicated in ageing, are evident even in wild-type strains (54). However, in strains defective for the DNA helicase Rrm3p, which is required for the processing of paused replication forks at sites of protein binding (77,78), the frequency of such events is greatly increased. If the balance between the pause-inducing and pause-processing machineries is disturbed, recombinational mechanisms must be employed to restart replication, with the concomitant risk of genomic instability. Thus, even though replication pausing systems such as mTERF may have evolved to limit genomic instability by preventing collisions of the replication and transcription machineries, their dysregulation, including by overexpression, could itself lead to instability. It follows that mTERF is a candidate gene for involvement in those cases of genetic disorders mediated by mtDNA rearrangements (e.g. PEO), whose genetic basis has not yet been elucidated.

In some cases replication pausing is merely a signature of defective replication (79,80). In others it is clearly a programmed event which facilitates other processes and preserves genome stability (81). As indicated by our findings (Figure 3), the phenomenon of pausing in human mtDNA is not confined to just one cell-type, nor is it an in vitro artefact seen only in cultured cells. The fact that it exhibits differences between cell-types and tissues strengthens the proposition that it is of physiological significance. The involvement of mTERF in modulating mtDNA replication pausing in human mtDNA, and the analogies with programmed replication pausing in other systems, support the idea that mTERF represents a system for safeguarding the integrity of the mitochondrial genome, whilst facilitating its efficient expression.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

This work was supported by funding from the Academy of Finland, Tampere University Hospital Medical Research Fund, the UK Medical Research Council, the UK Muscular Dystrophy Campaign, the Juselius Foundation and the European Union EUIMITOCOMBAT and MitEURO projects. We thank Steffi Goffart for advice and assistance with mtDNA copy number determination and interpretation of 2DNAGE, Peter Martinsson for help with microscopy, and Gertjan Hakkaart, Merja Jokela, Outi Kurronen, Anja Rovio, Tea Tuomela, Paivi Turunen and Ernesto Zanotto for advice and technical assistance. We also thank Claes Gustafsson, Heather Lorimer, Sanna Marjavaara, Maria Sola Villarubias and Anu Wartiovaara, as well as many colleagues in the MitEURO network, for useful discussions. Funding to pay the Open Access publication charges for this article was provided by University of Tampere and the Academy of Finland.

Conflict of interest statement. None declared.

**REFERENCES**


The mitochondrial transcription termination factor mTERF modulates replication pausing in human mitochondrial DNA

Anne K. Hyvärinen1,*, Jaakko L. O. Pohjoismäki1,*, Aurelio Reyes2, Sjoerd Wanrooij1, Takehiro Yasukawa2, Pekka J. Karhunen3, Johannes N. Spelbrink1, Ian J. Holt2, & Howard T. Jacobs1,4

1Institute of Medical Technology and Tampere University Hospital, University of Tampere, Finland
2MRC-Dunn Human Nutrition Unit, Cambridge, UK
3Department of Forensic Medicine and Tampere University Hospital, University of Tampere, Finland
4Institute of Biomedical and Life Sciences, University of Glasgow, Scotland, UK

* these authors contributed equally to the work

SUPPLEMENTARY DATA
LEGENDS TO SUPPLEMENTARY FIGURES

Supplementary Figure 1
Effects of overexpression of mTERF on mtDNA copy number and sequence-specific protein binding activity. (a) Q-PCR analysis of mtDNA copy number of Flp-In™ T-REx™-293 cells transfected with the mTERF construct and induced for the times indicated, normalized to the mean value for uninduced cells (means ± S.D. of 3 independently isolated DNA samples). The analysis compared the level of mtDNA (cytochrome b probe) with that of amyloid precursor protein (APP), as a single copy nuclear DNA standard, as previously (28). (b) EMSA supershift assay of protein binding to the tRNA_{Leu(UUR)} gene-containing fragment Leu (see Fig. 2 and Supplementary Table 1) using mitochondrial protein extract from Flp-In™ T-REx™-293 cells, with (+) or without (−) induced overexpression of mTERF or the mTERF-MycHis fusion protein, as indicated. Supershift was carried out using anti-Myc (m) or anti-FLAG (f) antibodies as shown (see Materials and Methods). The supershifted complex formed using the anti-Myc antibody, but not the anti-FLAG antibody, is arrowed. This supershifting property makes it possible to distinguish complexes actually containing the over-expressed mTERF-MycHis. Note also the low mobility bands visible at this exposure in the left-hand panel, which are probably due to binding by TFAM. Similar complexes are revealed in the gel shown in the right-hand panel (mTERF-MycHis expressing cells), at longer (more comparable) exposures. Their appearance is dependent on the fragment tested and the autoradiographic exposure used, not on the expression of any transgene.

Supplementary Figure 2
EMSA analysis of mTERF binding. See Figure 2 of the main paper for other data. (a) Schematic diagram of regions of the mitochondrial genome in which binding was tested using 150 bp fragments, nomenclature as in Fig. 2a, with additional segments of protein-coding genes A8, A6, COIII, ND3 and ND4L as shown. The overlap between A8 and A6 is shown in black. [The additional regions analysed for binding were selected, based on the data of the main paper, as regions showing enhanced replication pausing in mTERF-overexpressing cells]. (b) Map showing overlapping, shorter fragments
used in EMSA to localize the binding sites within fragment ND1.1 (see below, part d). (c) Further detail of binding to 150 bp fragments shown in Fig. 2. Mitochondrial protein extracts were from cells induced (or not) to over-express natural mTERF, with or without 100-fold mass excess of cold competitor. (d) Localization of mTERF binding sites within fragment ND1.1 using shorter, overlapping fragments (see part b), and supershifting. Cold competitor was present, where indicated, in several-fold mass excess. The simplest interpretation is that there are two mTERF binding sites within ND1.1, one located in fragments ND1.12 and ND1.13 but absent from ND1.11 and ND1.16, i.e. lying between np 3830-3855, the second located within ND1.14 but absent from ND1.13, i.e. lying between np 3880-3907. Within each of these short segments a good match to the SELEX consensus is found (see Table 1). (e) EMSA analyses of binding to 150 bp probe fragments in addition to those shown in Fig. 2. Nomenclature as for Fig. 2. A very high molecular weight mTERF-dependent complex (arrowed) was formed by fragment OH5. Since fragment OH5 also contains an identified binding site for TFAM, this complex may contain TFAM as well as mTERF, and could correspond with the large complex visualized by Martin et al (6) at the initiator site for rDNA transcription. Other fragments gave no reproducible EMSA signals indicative of mTERF binding. The ND5-6 dsDNA oligonucleotide covering the ND5-ND6 gene junction (Supplementary Table 1, data not shown) also gave no signal.

Supplementary Figure 3

Examples of replication pausing near mTERF binding sites. (a) Schematic map of human mtDNA showing relevant restriction sites, O_H, O_L, the approximate locations of the probes used (O_H and COI, see Materials and Methods), denoted by asterisks, the NCR (bold, dark grey) and rDNA (bold, pale grey). (b) 2DNAGE analysis of O_H and O_L-containing fragments from human placental mtDNA. Panels i and ii were treated, after digestion, with S1 nuclease and are images of gels already published in Ref. 24. Pause zones near O_H and O_L arrowed.
Supplementary Figure 4

Further 2DNAGE data showing effects of mTERF overexpression. (a) Comparable exposures of ND2-containing AccI fragment, treated with S1 nuclease, in which pause site ‘a’ (tRNA\textsubscript{Leu(UUR)}) is centrally located, showing enhancement of the species at the tip of the X-arc (‘x’), as well as of the double-Y arc leading to it, in mtDNA from cells in which mTERF overexpression was induced for 48 h. (b) Longer exposure of panels i and ii from Fig. 4c, plus similar gels of material treated with S1 nuclease. Nomenclature as Fig. 4c. (c) BclI digest (treated with S1 nuclease), probed for the 3.3 kb ND3-containing fragment (ND4 probe). Novel pause in mTERF-overexpressing cells arrowed. (d) BamHI digest probed as for the gels shown in Fig. 4d, using similar nomenclature. The steeply descending trajectory of the ‘eyebrow’ (broken-theta) arc, denoted ‘e’, is due to compression. i – initiation (bubble) arc, o – uncut circles, dY – double-Y molecules. Pausing in the ND1/tRNA\textsubscript{Leu(UUR)} region generates complex double-Y or broken-theta species in this digest (region of the gel designated ‘l’), whose migration and interconversion depends on the timing of lagging-strand maturation to dsDNA at the BamHI site. The enhanced pausing when mTERF is overexpressed is proposed to facilitate the more complete maturation of the lagging strand, such that the termination forms in the BamHI digest are resolved mainly as double-Y molecules (in which the site is cut on both daughter duplexes), denoted ‘u’, as opposed to the more electrophoretically retarded broken-theta molecules, denoted ‘s’, in which the site is cut only on one daughter branch, as seen in uninduced cells. (e) Illustration of the prominent intermediates generated by pausing in the ND1/tRNA\textsubscript{Leu(UUR)} region, which were detected in single-hitter digests (Fig. 4d and part (d) of this figure), as indicated. Filled and open ovals denote O\textsubscript{II} and the ND1/tRNA\textsubscript{Leu(UUR)} pause region, respectively. (f) One-dimensional agarose gel-blots of undigested mtDNA, with or without heating to 95 °C for 30 s to release 7S DNA (arrowed) by branch migration, as indicated. DNA was from cells induced for 24 h (ind) or not induced (unind) to overexpress mTERF. Lower panel indicates 7S DNA region of the gel (longer exposure). The upper panel indicates approximately equal loading and reveals subtle differences resulting from mTERF overexpression, in the molecular forms of mtDNA resolved on 1-D gels. Brackets indicate areas of the gel in which the migration of various forms of mtDNA from mTERF-overexpressing cells differed from those seen in uninduced cells.
**Supplementary Figure 5**

LM-PCR analysis of additional 5´ ends. (a) Schematic diagram indicating that sites of replication pausing, such as may result from a bound protein (filled oval), correspond also with persistent 5´ ends on the lagging strand. Nascent strands are shown with arrowheads to denote the 5´ to 3´ direction of synthesis. (b)-(i) LM-PCR reactions using the following primer sets: (b) TL1/TL2/TL7, (c) TL1/TL2/TL6, (d) TL1/TL2/TL5, (e) TL1/TL2/TL4, (f) TL1/TL2/TL3, (g) TL8/TL9/TL12, (h) TL8/TL9/TL10 and (i) L1/L2/L3, alongside sequencing ladders of the corresponding segments. Samples were as in Fig. 6. Amount of input DNA for uninduced cells in part (b) was 50% more than for other samples analysed in parallel. Note that the 5´ end at np 3211 does not increase uniformly during the time-course of mTERF over-expression, a curious but consistent anomaly (see also Fig. 6a) for which do not have any explanation. The minor bands, which differ in prominence between the genomic regions analysed, are unlikely to be due to nonspecific DNA nicking or polymerase stalling since there is no obvious reason why this would affect some portions of the mitochondrial genome more than others. It may be that they represent genuine DNA ends on the nascent lagging strand.

**Supplementary Figure 6**

Proposed minimal mTERF binding sites in the 16S/tRNA<sub>Leu(UUR)</sub>/ND1 boundary region. The canonical binding site (highlighted in orange) is the only one in which the pairs of GG residues are on the top (L-) strand. This site is flanked on each side by pairs of binding sites in the opposite orientation (highlighted in yellow). The tRNA<sub>Leu(UUR)</sub> gene sequence is boxed. Numbering as Ref. 82. The gap represents the original nt 3107, which was later deleted from the sequence upon revision, the gap having been inserted to preserve the original numbering.
FOOTNOTE

The authors’ contributions were as follows: AKH, JLOP, AR, SJW, TY, PJK and JNS performed the experimental work: JNS prepared cDNA, SJW recloned mTERF for inducible expression and established inducible cell-lines, JLOP extracted mtDNA and carried out 2DNAGE, partly together with AKH, AR performed LM-PCR, AKH carried out molecular cloning, transfections including RNAi, immunocytochemistry, Western blotting, Q-PCR for copy number determination, EMSA, SELEX and mIP, PJK prepared tissue samples from forensic autopsy and TY performed the initial 2DNAGE analyses to map pause sites. HTJ, IJH and JNS co-supervised the experimental work. HTJ drafted the figures and manuscript.
### SUPPLEMENTARY TABLE 1: EMSA<sup>a</sup> and hybridization probes<sup>b</sup>

<table>
<thead>
<tr>
<th>Name of probe</th>
<th>Coordinates</th>
<th>Length (bp)</th>
<th>Use</th>
<th>Genes or sequence features covered&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>15829-15978</td>
<td>150</td>
<td>EMSA</td>
<td>Cyt b, tRNA&lt;sup&gt;16S&lt;/sup&gt;, tRNA&lt;sup&gt;Pro&lt;/sup&gt;</td>
</tr>
<tr>
<td>NCR1</td>
<td>15959-16108</td>
<td>150</td>
<td>EMSA</td>
<td>NCR, tRNA&lt;sup&gt;Pro&lt;/sup&gt;, NCR, ETAS1</td>
</tr>
<tr>
<td>NCR2</td>
<td>16089-16238</td>
<td>150</td>
<td>EMSA</td>
<td>NCR, ETAS1, TAS, mt5 element</td>
</tr>
<tr>
<td>NCR3</td>
<td>16219-16368</td>
<td>150</td>
<td>EMSA</td>
<td>NCR, ETAS2</td>
</tr>
<tr>
<td>NCR4</td>
<td>16349-16498</td>
<td>150</td>
<td>EMSA</td>
<td>NCR</td>
</tr>
<tr>
<td>NCR5</td>
<td>16479-58</td>
<td>150</td>
<td>EMSA</td>
<td>NCR</td>
</tr>
<tr>
<td>NCR</td>
<td>35-611</td>
<td>576</td>
<td>Hybridization</td>
<td>NCR</td>
</tr>
<tr>
<td>OH1</td>
<td>41-190</td>
<td>150</td>
<td>EMSA</td>
<td>NCR, O&lt;sub&gt;16&lt;/sub&gt;, CSB1, CSB2, CSB3, P&lt;sub&gt;L&lt;/sub&gt;, P&lt;sub&gt;H1&lt;/sub&gt;, P&lt;sub&gt;H2&lt;/sub&gt;, 12 S rRNA</td>
</tr>
<tr>
<td>OH2</td>
<td>171-310</td>
<td>140</td>
<td>EMSA</td>
<td>NCR, O&lt;sub&gt;16&lt;/sub&gt;, CSB1, TFAM binding sites</td>
</tr>
<tr>
<td>OH3</td>
<td>291-440</td>
<td>150</td>
<td>EMSA</td>
<td>NCR, CSB2, CSB3, P&lt;sub&gt;L&lt;/sub&gt;</td>
</tr>
<tr>
<td>OH4</td>
<td>421-570</td>
<td>150</td>
<td>EMSA</td>
<td>NCR, P&lt;sub&gt;L&lt;/sub&gt;, TFAM binding sites</td>
</tr>
<tr>
<td>OH5</td>
<td>520-669</td>
<td>150</td>
<td>EMSA</td>
<td>NCR, TFAM binding sites, P&lt;sub&gt;H1&lt;/sub&gt;, P&lt;sub&gt;H2&lt;/sub&gt;, 12S rRNA</td>
</tr>
<tr>
<td>Leu</td>
<td>3193-3342</td>
<td>150</td>
<td>EMSA</td>
<td>16S rRNA, tRNA&lt;sup&gt;16S(r)(LCR)&lt;/sup&gt;, ND1</td>
</tr>
<tr>
<td>Leu-short</td>
<td>3230-3256</td>
<td>27</td>
<td>EMSA (ds oligo)</td>
<td>tRNA&lt;sup&gt;16S(r)(LCR)&lt;/sup&gt;</td>
</tr>
<tr>
<td>ND1.4</td>
<td>3312-3490</td>
<td>179</td>
<td>EMSA</td>
<td>ND1</td>
</tr>
<tr>
<td>ND1.5</td>
<td>3451-3600</td>
<td>150</td>
<td>EMSA</td>
<td>ND1</td>
</tr>
<tr>
<td>ND1.6</td>
<td>3581-3730</td>
<td>150</td>
<td>EMSA</td>
<td>ND1</td>
</tr>
<tr>
<td>ND1.7</td>
<td>3601-4080</td>
<td>481</td>
<td>Hybridization</td>
<td>ND1</td>
</tr>
<tr>
<td>ND1.11</td>
<td>3711-3860</td>
<td>150</td>
<td>EMSA</td>
<td>ND1</td>
</tr>
<tr>
<td>ND1.12</td>
<td>3800-3950</td>
<td>150</td>
<td>EMSA</td>
<td>ND1</td>
</tr>
<tr>
<td>ND1.13</td>
<td>3931-4080</td>
<td>150</td>
<td>EMSA</td>
<td>ND1</td>
</tr>
<tr>
<td>ND1.14</td>
<td>3800-3839</td>
<td>40</td>
<td>EMSA</td>
<td>ND1</td>
</tr>
<tr>
<td>ND1.15</td>
<td>3800-3873</td>
<td>74</td>
<td>EMSA</td>
<td>ND1</td>
</tr>
<tr>
<td>ND1.16</td>
<td>3813-3893</td>
<td>81</td>
<td>EMSA</td>
<td>ND1</td>
</tr>
<tr>
<td>ND1.17</td>
<td>3831-3907</td>
<td>95</td>
<td>EMSA</td>
<td>ND1</td>
</tr>
<tr>
<td>ND1.18</td>
<td>3813-3927</td>
<td>115</td>
<td>EMSA</td>
<td>ND1</td>
</tr>
<tr>
<td>ND1.19</td>
<td>3847-3950</td>
<td>104</td>
<td>EMSA</td>
<td>ND1</td>
</tr>
<tr>
<td>ND1.20</td>
<td>4061-4210</td>
<td>150</td>
<td>EMSA</td>
<td>ND1</td>
</tr>
<tr>
<td>Gene</td>
<td>Start-End</td>
<td>Length</td>
<td>Assay</td>
<td>Description</td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>--------</td>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>IQM1</td>
<td>4201-4350</td>
<td>150</td>
<td>EMSA</td>
<td>ND1, tRNA&lt;sup&gt;ile&lt;/sup&gt;, tRNA&lt;sup&gt;leu&lt;/sup&gt;</td>
</tr>
<tr>
<td>IQM2</td>
<td>4331-4480</td>
<td>150</td>
<td>EMSA</td>
<td>tRNA&lt;sup&gt;ile&lt;/sup&gt;, tRNA&lt;sup&gt;can&lt;/sup&gt;, tRNA&lt;sup&gt;met&lt;/sup&gt;, ND2</td>
</tr>
<tr>
<td>ND2</td>
<td>4480-4988</td>
<td>589</td>
<td>Hybridization</td>
<td>ND2</td>
</tr>
<tr>
<td>ND2.1</td>
<td>5161-5310</td>
<td>150</td>
<td>EMSA</td>
<td>ND2</td>
</tr>
<tr>
<td>ND2.2</td>
<td>5221-5370</td>
<td>150</td>
<td>EMSA</td>
<td>ND2</td>
</tr>
<tr>
<td>OL</td>
<td>5685-5834</td>
<td>150</td>
<td>EMSA</td>
<td>tRNA&lt;sup&gt;asp&lt;/sup&gt;, OL, tRNA&lt;sup&gt;cys&lt;/sup&gt;, tRNA&lt;sup&gt; tyr&lt;/sup&gt;</td>
</tr>
<tr>
<td>COI</td>
<td>5831-6244</td>
<td>414</td>
<td>Hybridization</td>
<td>tRNA&lt;sup&gt; tyr&lt;/sup&gt;, COI</td>
</tr>
<tr>
<td>A8A6</td>
<td>8475-8625</td>
<td>150</td>
<td>EMSA</td>
<td>A8, A6</td>
</tr>
<tr>
<td>A6.1</td>
<td>8501-8650</td>
<td>150</td>
<td>EMSA</td>
<td>A6</td>
</tr>
<tr>
<td>A6.2</td>
<td>8631-8780</td>
<td>150</td>
<td>EMSA</td>
<td>A6</td>
</tr>
<tr>
<td>A6.3</td>
<td>8761-8910</td>
<td>150</td>
<td>EMSA</td>
<td>A6</td>
</tr>
<tr>
<td>A6.4</td>
<td>8891-9040</td>
<td>150</td>
<td>EMSA</td>
<td>A6</td>
</tr>
<tr>
<td>A6.5</td>
<td>9021-9170</td>
<td>150</td>
<td>EMSA</td>
<td>A6</td>
</tr>
<tr>
<td>A6.6</td>
<td>9151-9300</td>
<td>150</td>
<td>EMSA</td>
<td>A6</td>
</tr>
<tr>
<td>ND3.1</td>
<td>9961-10110</td>
<td>150</td>
<td>EMSA</td>
<td>ND3</td>
</tr>
<tr>
<td>ND3.2</td>
<td>10091-10240</td>
<td>150</td>
<td>EMSA</td>
<td>ND3</td>
</tr>
<tr>
<td>ND3.3</td>
<td>10221-10370</td>
<td>150</td>
<td>EMSA</td>
<td>ND3</td>
</tr>
<tr>
<td>ND3.4</td>
<td>10351-10500</td>
<td>150</td>
<td>EMSA</td>
<td>ND3</td>
</tr>
<tr>
<td>ND3.5</td>
<td>10481-10630</td>
<td>150</td>
<td>EMSA</td>
<td>ND3</td>
</tr>
<tr>
<td>ND4</td>
<td>11161-11640</td>
<td>480</td>
<td>Hybridization</td>
<td>ND4</td>
</tr>
<tr>
<td>ND5</td>
<td>14134-14160</td>
<td>27</td>
<td>EMSA (ds oligo)</td>
<td>ND5-ND6 gene junction</td>
</tr>
</tbody>
</table>

<sup>a</sup>Leu-short and ND5-6 were dsDNA oligonucleotides (see text). All other EMSA probes were created by PCR using two 20 nt oligonucleotide primers commencing at the coordinates indicated (82), except for the H-strand primers for ND1.5, ND3.5, A6.1 and IQM1 (17 nt), NCR3, A6.3, A6.4, A6.6 and A8A6 (18 nt), ND3.1, Leu and OL (19 nt), ND3.2 (26 nt) and ND3.4 (27 nt), the L-strand primers for ND1.6 (16 nt), NCR4, A6.5 and OH5 (18 nt), ND3.3 (24 nt) and ND3.5 (27 nt), and both primers for probes ND1.11-16 (all 27 nt).

<sup>b</sup>Hybridization probes were synthesized by PCR using two 20 nt oligonucleotide primers commencing at the coordinates indicated (82).
**SUPPLEMENTARY TABLE 2: Oligonucleotides used in LM-PCR**

<table>
<thead>
<tr>
<th>Name of oligonucleotide</th>
<th>Oligonucleotide sequence (5’ to 3’)</th>
<th>Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>TL1</td>
<td>TTAGCGCTGTGATGAGTGTG</td>
<td>4536-4517</td>
</tr>
<tr>
<td>TL2</td>
<td>AGTGTGCCTGCAAAAGATGTAG</td>
<td>4522-4501</td>
</tr>
<tr>
<td>TL3</td>
<td>TGCTAGGGTGAGTGAGGGAAGTTTTT</td>
<td>4203-4177</td>
</tr>
<tr>
<td>TL4</td>
<td>CGGCTATGAAGAATAGGGCGGAAG</td>
<td>3985-3963</td>
</tr>
<tr>
<td>TL5</td>
<td>AGTGTGCCTGCAAAAGATGTAG</td>
<td>3812-3788</td>
</tr>
<tr>
<td>TL6</td>
<td>AAATAGGAGGCGCTAGGTTGAGGTTGA</td>
<td>3618-3593</td>
</tr>
<tr>
<td>TL7</td>
<td>GGGCCTTTTCGTAGTTGTATATAGCC</td>
<td>3416-3391</td>
</tr>
<tr>
<td>TL8</td>
<td>TCATGTATAGCTTGGTTCAGGTG</td>
<td>5186-5165</td>
</tr>
<tr>
<td>TL9</td>
<td>GCTTGTTCAGGTGCGAGAT</td>
<td>5178-5159</td>
</tr>
<tr>
<td>TL10</td>
<td>GGGGTGCTTTGGTAACCTC</td>
<td>4837-4818</td>
</tr>
<tr>
<td>TL11</td>
<td>TGCCAGCTTCTGGGAACGA</td>
<td>4638-4619</td>
</tr>
<tr>
<td>TL12</td>
<td>GGTATGGGCCTAATAGCCTATT</td>
<td>4438-4416</td>
</tr>
<tr>
<td>TL18</td>
<td>GGTTGATTGCTGTACCTTG</td>
<td>16223-16204</td>
</tr>
<tr>
<td>TL19</td>
<td>AGGGGGTTTTGAGTGTGGATT</td>
<td>16189-16170</td>
</tr>
<tr>
<td>TL21</td>
<td>ATGAGGATGGTAGTAATAGGGCAAGG</td>
<td>15645-15619</td>
</tr>
<tr>
<td>L1</td>
<td>TAGAGCCTGTGCTTAGACCT</td>
<td>6002-5983</td>
</tr>
<tr>
<td>L2</td>
<td>CACGCATCCTCGCCGCAATAAGGG</td>
<td>5982-5960</td>
</tr>
<tr>
<td>L3</td>
<td>GGTATAGTGTTCCAATGTCTTTGTGGTTTAG</td>
<td>5961-5929</td>
</tr>
<tr>
<td>H1</td>
<td>CCTCCTAGGGCAACCA</td>
<td>15490-15505</td>
</tr>
<tr>
<td>H2</td>
<td>AGACACATATACCCCTTAGCCA</td>
<td>15505-15524</td>
</tr>
<tr>
<td>H3</td>
<td>CACACGTTCCTCTTAAATAGACATACAGATG</td>
<td>16537-16569</td>
</tr>
<tr>
<td>H5</td>
<td>TCCACCATTAGCACCACAAAGCTAAGATTCT</td>
<td>15976-16005</td>
</tr>
</tbody>
</table>

*Coordinates as in Ref. 82.*
**SUPPLEMENTARY TABLE 3**

**SELEX analysis of the mTERF binding site and possible binding sites in mtDNA**

<table>
<thead>
<tr>
<th>Position</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELEX Consensus&lt;sup&gt;a&lt;/sup&gt;</td>
<td>t</td>
<td>G</td>
<td>G</td>
<td>t</td>
<td>a</td>
<td>r</td>
<td>a</td>
<td>r</td>
<td>g</td>
<td>T</td>
<td>Y</td>
<td>G</td>
<td>G</td>
<td>t</td>
</tr>
<tr>
<td>Actual&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu&lt;sup&gt;c&lt;/sup&gt;</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>ц</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>ND1.12&lt;sup&gt;d&lt;/sup&gt;</td>
<td>a</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>a</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>ц</td>
</tr>
<tr>
<td>ND1.14&lt;sup&gt;e&lt;/sup&gt;</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>OH1&lt;sup&gt;f&lt;/sup&gt;</td>
<td>a</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>ц</td>
<td>a</td>
<td>G</td>
<td>G</td>
<td>A</td>
</tr>
</tbody>
</table>

<sup>a</sup>See Table 1 of main paper

<sup>b</sup>Best matches to consensus within several of the inferred binding-site regions of human mtDNA, based on EMSA

<sup>c</sup>Canonical binding site for mTERF within tRNA<sup>Leu(UUR)</sup> gene
Proposed binding site in ND1 gene between np 3830-3855; see Supplementary Fig. 2d.

Proposed binding site in ND1 gene between np 3880-307; see Supplementary Fig. 2d.

Possible binding site within fragment OH1; see Fig. 2
Hyvärinen et al, Supplementary Figure 1

(a) mtDNA copy number (a.u.)

(b) Induction - + + - + +
Antibody - - m - m f

mTERF mTERF-MycHis
Hyvärinen et al, Supplementary Figure 2, page 1 of 2
Hyvärinen et al, Supplementary Figure 3

(a) Diagram showing restriction enzyme sites (HinfII, AccI, BclI) and bands (16569/1, OH, OL, COI) with probe locations.

(b) Gel images showing bands sized as follows:
- HinfII, 3.9 kb
- AccI, 2.8 kb
- BclI, 4.0 kb
Hyvärinen et al, Supplementary Figure 4, page 1 of 2

(a) uninduced induced

uninduced induced

Accl 4.8 kb, ND2 probe, +S1

(b) uninduced induced

untreated +S1

Accl 2.8 kb, O_H probe

(c) uninduced induced

BclI 3.3 kb, ND4 probe, +S1

(d) uninduced induced

BamHI 16.6 kb, ND2 + ND4 probes
3001 ggacatcccg atggtgcagc cgctattaaa ggttcgtttg ttcaacgatt aaagtcctac

3061 gtgatctgag ttcaga ccgg agtaatccag gtcggtttct atctac-ttc aaatctctcc

3121 ctgtacgaaa ggaagaga aataagggct acctcacaaa gcgccttccc ccgtaaatga

3181 tatcatctca acttagttt ataco cccaco ccacccaaga acagggtttg ttaagatggc

3241 agagccgggt aatcgcataa aacctaaaac tttaacgtca gaggttcaat tctctctctt

3301 aacaacatac ccatggccaa cctccctactc ctcattgtac ccattctaat cgaatggca

3361 ttctcaatgc ttaccgaacg aaaaattcta ggctatatac aactacgcaaa aggcccccaac