NINA RAJALA

Probing the Mammalian Mitochondrial Nucleoid

ACADEMIC DISSERTATION
To be presented, with the permission of the Board of the BioMediTech of the University of Tampere, for public discussion in the Jarmo Visakorpi Auditorium, of the Arvo Building, Lääkärinkatu 1, Tampere, on May 16th, 2015, at 12 o’clock.

UNIVERSITY OF TAMPERE
NINA RAJALA

Probing the Mammalian Mitochondrial Nucleoid

Acta Universitatis Tamperensis 2052
Tampere University Press
Tampere 2015
No amount of experimentation can ever prove me right; a single experiment can prove me wrong.
Albert Einstein

Ihanalle tyttärelleni Arlesille
Table of Contents

List of original communications .................................................. 7
Abstract ....................................................................................... 8
Lysiennelmä ................................................................................ 10
List of abbreviations ..................................................................... 12
1. Introduction ............................................................................. 15

2. Review of the literature .......................................................... 17
   2.1 Mitochondria ................................................................. 17
   2.2 Mitochondrial genetics .................................................... 19
      2.2.1 Organisation of mtDNA ........................................ 20
   2.3 Mitochondrial DNA maintenance ...................................... 22
      2.3.1 mtDNA transcription ............................................. 23
      2.3.2 mtDNA replication ............................................... 25
      2.3.3 mtDNA replication machinery ............................... 29
      2.3.4 Initiation and termination of replication .................. 37
      2.3.5 mtDNA repair ..................................................... 38
   2.4 MtDNA diseases .............................................................. 40
      2.4.1 Twinkle helicase and mtDNA maintenance diseases .... 41
   2.5 Mammalian mitochondrial nucleoids .................................. 43
      2.5.1 What do we know about nucleoid membrane connection? ........................................ 44
      2.5.2 Organisation of mammalian mitochondrial nucleoids ... 44
      2.5.3 Diversity of mitochondrial mtDNA and nucleoids .......... 47
      2.5.4 Composition of mammalian mitochondrial nucleoids .... 48
      2.5.5 Mass spectrometry based nucleoid identification strategies ........................................ 52

3. Aims of the research ................................................................. 56

4. Materials and methods ............................................................ 57
   4.1 Cell culture systems ......................................................... 57
   4.2 Transfections, confocal microscopy, Edu and BrdU labeling ....... 57
   4.3 siRNA knockdown and ddC treatment ............................... 59
   4.4 Western blot analysis ...................................................... 59
   4.5 Isolation of mitochondria and mitochondrial treatments ........ 60
      4.5.1 Mitochondrial (sub)fractionation ................................ 60
      4.5.2 Mitochondrial carbonate or KCl extraction ............... 61
4.5.3 Treatment of isolated mitochondria by sonication and nucleases .......................................................................................62
4.6 Dot-plot analysis of mtDNA content in digitonin fractions ..........62
4.7 Floatation assay ..........................................................................................................................63
4.8 Formaldehyde cross-linking and immunoprecipitation ...............63
4.9 Mass spectrometry sample preparation .................................................64
4.10 Mass spectrometric measurements ....................................................65
4.11 Mass spectrometric data analysis .........................................................66
4.12 Baculovirus protein expression ...........................................................67
4.13 Twinkle purification for TEV analysis ...............................................68
4.14 TEV protease expression and purification ........................................68
4.15 Further Twinkle purification .................................................................69
4.16 Chromatography based Twinkle purification ....................................70
4.17 Helicase assay ..........................................................................................71
5. Results ............................................................................................................72
5.1 Twinkle structural studies ..........................................................72
  5.1.1 Optimisation of Twinkle expression ........................................72
  5.1.2 Optimisation of Twinkle purification ......................................73
  5.1.3 Structural analysis of Twinkle ......................................................76
5.2 Organisation of nucleoids .................................................................78
  5.2.1 Twinkle shows properties of a mitochondrial inner membrane protein .................................................................78
  5.2.2 Mitochondrial nucleoid proteins show different fractionation properties .................................................................78
  5.2.3 Twinkle is associated to the mitochondrial inner membrane and two pools of mtDNA–protein complexes could be separated on the basis of their solubility .........79
  5.2.4 Modelling of bacteriophage T7 helicase and Twinkle reveals a loop region in Twinkle that seem to have some role in Twinkle anchorage to the inner membrane ..........81
  5.2.5 Mitochondrial nucleoids are heterogeneous and dynamic in nature .................................................................84
  5.2.6 Twinkle and mtSSB are not constitutive nucleoid components .....................................................................................84
  5.2.7 Twinkle organisation is independent of mtDNA .........................84
  5.2.8 mtSSB organisation at mtDNA foci is Twinkle dependent .................................................................................................85
5.2.9 Twinkle and mtSSB are enriched in replicating nucleoids ................................................................. 89
5.2.10 Twinkle determines the distribution of mtDNA and TFAM in membrane fraction ........................................... 90
5.3 Investigation on the composition of mammalian mitochondrial nucleoids .............................................................. 92
5.3.1 Development of whole cell cross linking method for nucleoid associated protein identification ..................... 93
5.4 Identifying potential nucleoid associated proteins using mass spectrometry .......................................................... 94
5.4.1 Validation of whole cell cross linking method by mass spectrometry analysis .................................................... 95
5.4.2 Optimisation of whole cell cross linking method for nucleoid protein identification ........................................... 98
5.4.3 Application of whole cell cross-linking method for identification of proteins in a minimal replication platform in the absence of mtDNA .......................................................... 101
6. Discussion ......................................................................................................................................................... 106
6.1 Nucleoid biology ............................................................................................................................................. 106
6.1.1 Nucleoids are heterogeneous in nature ........................................................................................................ 107
6.1.2 Twinkle shows characteristics of a mitochondrial inner membrane protein .................................................. 108
6.1.3 Organisation of nucleoids at the mitochondrial inner membrane is dynamic and replication dependent ....... 110
6.1.4 Conservation of mechanistic character of nucleoids .................................................................................... 111
6.2 Mitochondrial nucleoid protein identification strategies brought to the 21st century .................................................. 114
6.2.1 The mitochondrial nucleoid “compartment” ................................................................................................. 116
6.2.2 Close liaison between nucleoids and ribosomes ......................................................................................... 117
6.3 Twinkle purification and structural studies ....................................................................................................... 118
6.3.1 Twinkle double hexamer .......................................................................................................................... 119
7. Summary and conclusions ................................................................................................................................... 120
Acknowledgements .............................................................................................................................................. 122
References ............................................................................................................................................................. 124
Appendices .......................................................................................................................................................... 148
Original communications .................................................................................................................................. 154
List of original communications

This thesis is based on the following communications:


*Joint first authorship

Articles I and III, were originally published in Nucleic Acids Research, Oxford University Press. By the Copyright (2014) Oxford University press, authors may use their own material in other publications provided that the Journal is acknowledged as the original place of publication and Oxford University Press as the Publisher. Article II was originally published in PLOSone, PLOS. Authors retain ownership of the copyright of their articles.

Other Publications:
Abstract

Mitochondria are involved in many cellular functions, which the most important is the supply of energy in form of ATP, which is produced by the oxidative phosphorylation system (OXPHOS). In addition, mitochondria are involved in other important cellular functions including providing cellular constituents, calcium storage and apoptosis. Mitochondria have their own DNA (mtDNA) that codes for parts of the OXPHOS and the rest is coded by the nuclear genome. Among the nuclearly encoded proteins are those involved in mtDNA maintenance, which have during evolution moved to the nucleus. The unique feature of having its own DNA means that both nuclear and mtDNA can affect the mitochondrial functions and mutations in either genome can lead to mitochondrial dysfunction. For example, mutations in mitochondrial helicase Twinkle have been implicated in adPEO (autosomal dominant progressive external ophthalmoplegia) and shown to cause mtDNA deletions. This in turn results in a loss of OXPHOS genes and leads to decrease in energy production. In vitro studies have shown that depending on the mutation either the helicase activity, ssDNA binding property or oligomerisation of Twinkle can be affected. In addition, the most severe mutant forms of Twinkle can cause replication stalling.

In this present study, we have investigated the structure of Twinkle and the results show the existence of Twinkle as hexamers and heptamers. Twinkle is a close relative of bacteriophage T7 gp4 primase/helicase protein and belongs to the SF4 family of hexameric replicative DnaB-like helicases. Oligomeric transitions between hexamers and heptamers have also been shown for T7 gp4 protein. Further, we show that the stabilisation of Twinkle oligomeric structure is dependent on the N-terminal portion of the protein and unlike the T7 gp4 protein, the linker region is not solely responsible for the oligomerisation. The functional importance is corroborated by the existence of several Twinkle disease mutants in this domain.

Since the discovery that also mtDNA is found in discrete mtDNA:proteins structures referred to as nucleoids, research on determining the composition of
nucleoids has caught the interest of mitochondrial researchers. Research this far has used nucleoid complexes isolated from mitochondria and has concentrated mostly on non-quantitative methods and on examination of one protein found in the mass-spectrometry-screen. Here we show that whole cell formaldehyde crosslinking combined with affinity purification and tandem mass-spectrometry provides a simple and reproducible method to identify potential nucleoid associated proteins by mass-spectrometry. We also investigated the composition of Twinkle associated nucleoids in cells lacking mtDNA and were able to identify proteins that are reduced or absent when mtDNA is not present.

The fact that nucleoids are membrane bound is a long-standing observation, but the mechanism is still unclear. Here we show that Twinkle is firmly membrane associated even in the absence of mtDNA unlike single-stranded DNA-binding protein (mtSSB) and mitochondrial transcription factor A (TFAM). Further investigation on the association of replication factors with nucleoids show that endogenous Twinkle and mtSSB co-localize only with a subset of nucleoids. Using nucleotide analogs to identify replicating nucleoids we were able to show that nucleoids transiently associate with Twinkle when there is a need for replication while mtSSB is recruited to nucleoids in Twinkle dependent manner.

In conclusion, the work here gives important information on the structure of mitochondrial helicase Twinkle, which will help us to understand Twinkle caused mitochondrial disorders. Furthermore, our results show that Twinkle has an important role during mtDNA replication so that Twinkle recruits or is assembled with mtDNA at the inner membrane to form a replication platform and amount to the first clear demonstration that nucleoids are dynamic in both composition and concurrent activity. The nucleoid isolation work here offers a fast and quantitative method for nucleoid associated protein isolation, which can be applied to screen for example, Twinkle mutant expressing cell lines in search for factors important during diseases. Our data provides a very valuable resource for both basic mitochondrial researchers as well as clinical geneticists working to identify novel disease genes on the basis of exome sequence data.
Lyhennelmä


MtDNA on pakattu nukleoidieiksi kutsuttuihin mtDNA-proteiinkomplekseihin, joiden rakentumista ei ole täysin ymmärretty. Vaikka nukleoidien koostumusta ei

Monet tutkijat ovat olleet kiinnostuneita selvittämään nukleoidien koostumusta ja rakennetta, mutta tähänastinen nukleoiditutkimus on keskittynyt suurimmaksi osaksi ei-kvantitatiiviseen tutkimukseen. Olen kehittänyt uudenlaisen nukleoidiproteiinien eristysmenetelmän, jossa käytetään kokonaisia soluja, jolloin mitokondrioiden eristämiseen ei ole tarvetta. Yhdistettynä immunoaffiniteettipurifikaation ja kvantitatiivisen massaspektometriikan kanssa (käyttäen label free quantification – tekniikkaa), menetelmä tarjoaa yksinkertaisen ja toistettavan tavan nukleoidiproteiinien tunnistamiseen. Tällä menetelmällä olemme tunnistaneet monia jo tiedettyjä nukleoidiproteiineja, sekä listanneet joukon proteiineja, jotka voivat osoittautua tärkeiksi mtDNA:n ylläpidossa.

Väitöskirjatutkimukseni on tuottanut tärkeää tietoa Twinklen rakenteesta ja organisoitumisesta mitokondrion sisäkalvolle. Tutkimukseni auttaa ymmärtämään mtDNA:n ylläpitoa yleisesti ja siten myös paremmin mitokondrion sairauksia. Lisäksi olemme systemaattisesti tutkineet nukleoidien rakennetta massaspektrometrian avulla. Löytyämämme proteiinit ovat myös mielenkiintoinen jatkokutkimuskohde ja tärkeä tiedon lähde muille mitokondriotutkijoille.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>adPEO</td>
<td>autosomal dominant progressive external ophthalmoplegia</td>
</tr>
<tr>
<td>ANT1</td>
<td>adenine nucleotide translocator</td>
</tr>
<tr>
<td>ATAD3</td>
<td>ATPase family AAA domain-containing protein 3</td>
</tr>
<tr>
<td>ADP</td>
<td>adenosine diphosphate</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BER</td>
<td>base excision repair</td>
</tr>
<tr>
<td>CTD</td>
<td>C’-terminal domain</td>
</tr>
<tr>
<td>ddC</td>
<td>dideoxycytidine</td>
</tr>
<tr>
<td>DHX30</td>
<td>DEAH-box helicase</td>
</tr>
<tr>
<td>D-loop</td>
<td>displacement loop</td>
</tr>
<tr>
<td>ds</td>
<td>double stranded</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>EM</td>
<td>electron microscopy</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>ERMES</td>
<td>ER-mitochondria encounter structure</td>
</tr>
<tr>
<td>FA</td>
<td>formaldehyde</td>
</tr>
<tr>
<td>H</td>
<td>heavy</td>
</tr>
<tr>
<td>hMDS</td>
<td>hepatocerebral mtDNA depletion syndrome</td>
</tr>
<tr>
<td>HR</td>
<td>homologous recombination</td>
</tr>
<tr>
<td>HSP</td>
<td>heavy-strand promoter</td>
</tr>
<tr>
<td>ICC</td>
<td>immunocytochemistry</td>
</tr>
<tr>
<td>IF</td>
<td>immunofluorescence</td>
</tr>
<tr>
<td>IAP</td>
<td>immuno affinity purification</td>
</tr>
<tr>
<td>IM</td>
<td>mitochondrial inner membrane</td>
</tr>
<tr>
<td>IMAN</td>
<td>insoluble membrane associated nucleoids</td>
</tr>
<tr>
<td>IMS</td>
<td>mitochondrial intermembrane space</td>
</tr>
<tr>
<td>IOSCA</td>
<td>infantile onset of spinocerebellar ataxia</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>KCl</td>
<td>potassium chloride</td>
</tr>
<tr>
<td>KIF</td>
<td>kinesin motor</td>
</tr>
<tr>
<td>L</td>
<td>light</td>
</tr>
<tr>
<td>LFQ</td>
<td>label-free quantification</td>
</tr>
<tr>
<td>LSP</td>
<td>light-strand promoter</td>
</tr>
<tr>
<td>M</td>
<td>mitochondrial matrix</td>
</tr>
<tr>
<td>MIA</td>
<td>mitochondrial intermembrane space transport and assembly</td>
</tr>
<tr>
<td>MMR</td>
<td>mismatch repair</td>
</tr>
<tr>
<td>MRPL12</td>
<td>mitochondrial ribosomal protein L12</td>
</tr>
<tr>
<td>MRPS22</td>
<td>small mitochondrial ribosomal subunit</td>
</tr>
<tr>
<td>MRPL49</td>
<td>large mitochondrial ribosomal subunit</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>mt</td>
<td>mitochondrial</td>
</tr>
<tr>
<td>mtDNA</td>
<td>mitochondrial DNA</td>
</tr>
<tr>
<td>mtSSB</td>
<td>mitochondrial single stranded binding protein</td>
</tr>
<tr>
<td>MTERF</td>
<td>mitochondrial transcription terminator factor</td>
</tr>
<tr>
<td>mtLuc</td>
<td>mitochondrially targeted luciferase</td>
</tr>
<tr>
<td>mtTOPO1</td>
<td>human mitochondrial topoisomerase I</td>
</tr>
<tr>
<td>mtTFB1</td>
<td>mitochondrial transcription factor B1</td>
</tr>
<tr>
<td>mtTFB2</td>
<td>mitochondrial transcription factor B2</td>
</tr>
<tr>
<td>MTERF</td>
<td>transcription terminator factor</td>
</tr>
<tr>
<td>MYH</td>
<td>non-muscle myosin heavy chain</td>
</tr>
<tr>
<td>NAD</td>
<td>nicotinamide ademine dinucleotide (oxidised form)</td>
</tr>
<tr>
<td>NADH</td>
<td>nicotinamide ademine dinucleotide (reduced form)</td>
</tr>
<tr>
<td>NAP</td>
<td>nucleoid associated protein</td>
</tr>
<tr>
<td>NCR</td>
<td>non-coding regions</td>
</tr>
<tr>
<td>NER</td>
<td>nucleotide excision repair</td>
</tr>
<tr>
<td>NHEJ</td>
<td>non-homologous end-joining</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>O_H</td>
<td>heavy strand origin</td>
</tr>
<tr>
<td>O_L</td>
<td>light strand origin</td>
</tr>
<tr>
<td>OM</td>
<td>mitochondrial outer membrane</td>
</tr>
<tr>
<td>OXPHOS</td>
<td>oxidative phosphorylation system</td>
</tr>
<tr>
<td>PDIP38</td>
<td>polyperase delta interacting protein</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>PFA</td>
<td>paraformaldehyde</td>
</tr>
<tr>
<td>PEO</td>
<td>progressive external ophthalmoplegia</td>
</tr>
<tr>
<td>PH1</td>
<td>heavy strand transcription initiation site 1</td>
</tr>
<tr>
<td>PH2</td>
<td>heavy strand transcription initiation site 2</td>
</tr>
<tr>
<td>PHB</td>
<td>prohibitin</td>
</tr>
<tr>
<td>POLG</td>
<td>DNA polymerase γ</td>
</tr>
<tr>
<td>POLG1</td>
<td>the catalytic subunit of POLGγ</td>
</tr>
<tr>
<td>POLG2</td>
<td>the accessory subunit of POLGγ</td>
</tr>
<tr>
<td>POLRMT</td>
<td>mitochondrial RNA polymerase</td>
</tr>
<tr>
<td>r</td>
<td>ribosomal</td>
</tr>
<tr>
<td>RI</td>
<td>replication intermediates</td>
</tr>
<tr>
<td>RITOLS</td>
<td>RNA incorporation throughout the lagging strand</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNase MRP</td>
<td>mitochondrial RNA processing protein</td>
</tr>
<tr>
<td>RPD</td>
<td>RNA polymerase domain</td>
</tr>
<tr>
<td>SAXS</td>
<td>small angle X-ray scattering</td>
</tr>
<tr>
<td>SMLN</td>
<td>soluble matrix localised nucleoids</td>
</tr>
<tr>
<td>ss</td>
<td>single-stranded</td>
</tr>
<tr>
<td>ssDNA</td>
<td>single stranded DNA</td>
</tr>
<tr>
<td>TAS</td>
<td>termination –association sequence</td>
</tr>
<tr>
<td>TEFM</td>
<td>mitochondrial transcription elongation factor</td>
</tr>
<tr>
<td>TEV</td>
<td>Tobacco Etch Virus</td>
</tr>
<tr>
<td>TFAM</td>
<td>mitochondrial transcription factor A = mTFA</td>
</tr>
<tr>
<td>TIM</td>
<td>transporter of the inner membrane</td>
</tr>
<tr>
<td>TRE</td>
<td>TFAM responsive element</td>
</tr>
<tr>
<td>TMS</td>
<td>two membrane spanning</td>
</tr>
<tr>
<td>TOM</td>
<td>translocase of the outer membrane</td>
</tr>
<tr>
<td>TOPOIIIα</td>
<td>topoisomerase IIIα</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>Twinkle</td>
<td>the mitochondrial DNA helicase</td>
</tr>
<tr>
<td>2DNAGE</td>
<td>neutral two-dimensional agarose gel electrophoresis</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>XL</td>
<td>cross linked</td>
</tr>
<tr>
<td>ZBD</td>
<td>zinc binding domain</td>
</tr>
</tbody>
</table>
1. Introduction

Gottfried Schatz has said: “Mitochondria can tell us who we are and where do we come from - We are an assembly of two different organisms that decided to live together 1.5 billion years ago” (Sagan 1967 and Schatz 2001). Mitochondria are thought to have evolved from symbiosis with oxidative bacteria. The current concept is that mitochondria were once free-living bacteria that were engulfed by larger cells in a unique merger that gave rise to eukaryotic cells. During evolution this symbiosis proved to be beneficial and the bacteria evolved to the “the powerhouse” of the cell – the mitochondria (Lane 2006).

The mitochondrial inner membrane harbours the oxidative phosphorylation system (OXPHOS), which converts the cellular fuels for energy. To paraphrase remarks attributed to Professor Howy Jacobs: “The machinery for cellular respiration is like a car engine: A good engine only uses a little petrol, works efficiently and only produces small amounts of toxic waste. Whereas, a bad engine works opposite.” Because this force providing function is essential to life, and if the mitochondrial engine is not working properly, the by-products can damage the cells and mitochondrial DNA (Prof Howy Jacobs). This can lead to a wide variety of mitochondrial diseases and has been shown to contribute to ageing. In addition, mitochondrial defects have been linked to disorders like Parkinson’s disease, Alzheimer’s disease, epilepsy, diabetes, and neuromuscular diseases. Individual mitochondrial diseases are rare, but because there are so many different mitochondrial diseases, they have a large impact – pathogenic mtDNA mutations affect at least 1 in 5000 people (Chinnery et al 2000 Schaefer et al 2008 and Cree et al 2009).

Words from Gilkerson describe the mitochondria well: “It has been clear that mitochondria play by their own rules”. They have their own DNA; produce some of their own RNA and proteins, including parts of the OXPHOS to produce ATP for life (Gilkerson et al 2013). The molecular mechanism of mtDNA maintenance and the organisation of mtDNA are not fully understood. Research into understanding
how pathogenic mutations in mtDNA maintenance proteins give rise to mitochondrial diseases is important to understand, and to get a cure, for mitochondrial diseases. Furthermore, it is also important to understand which factors and mechanisms are involved in replication, repair, organisation and segregation of mtDNA/nucleoids.

Research for this thesis was carried out to gain more insight into the mammalian mitochondrial nucleoids and the mitochondrial helicase Twinkle, in order to understand the mtDNA maintenance in health and disease. I have investigated the nature and composition of the mammalian mitochondrial nucleoid and its membrane association. A specific aim of this thesis work has been to develop a fast and reproducible method for nucleoid purification and protein identification by mass spectrometry analysis. This method would enable screening of many samples in standardised conditions in search of proteins that are involved in maintenance of mtDNA. Finally, I have studied the structure of mitochondrial helicase Twinkle, which would help us to understand mtDNA maintenance diseases caused by Twinkle mutants.
2. Review of the literature

2.1 Mitochondria

Margulis in 1967 (Sagan 1967) proposed the endosymbiont theory. Although it was heavily based on the ideas first presented by Merezhkovsky and Wallin early in the 20th-century and ignored by the scientific world, the theory by Margulis is the first based on microbiological evidence and is considered the landmark of the endosymbiotic theory (Margulis 1981). There has been debate on the evolution of mitochondria, but the current view is that mitochondria are: “the descendants of a bacterial endosymbiont that became established at an early stage in a nucleus-containing host cell” (Gray et al 1999). It is believed that mitochondria evolved 1.5x109 years ago from prokaryotes that were engulfed by primitive eukaryotic cells and developed a symbiotic relationship (Margulis 1970).

First evidence for the existence of mitochondria came in 1857 when scientist Albert von Kölliker described mitochondria as “granules” in muscle cells. In 1886 Richard Altman called mitochondria as “bioblasts” and described them as basic units of cellular activity, but it was not until 1898 when Carl Benda called these organelles as mitochondria, from Greek thread, mitos and granule chondros (see Ernster & Schatz 1981 for historical review on mitochondria).

As discussed by de Brito & Scorrano (2010), work by Palade (1952) and Sjostrand (1953) showed that mitochondria have two membranes, an outer mitochondrial membrane (OM) and a highly convoluted inner membrane (IM) that is folded into a series of ridges called cristae. Later studies refined the structure to contain compartments of intermembrane space and the matrix that harbours the mitochondrial DNA (mtDNA). Mitochondria reside in the cytoplasm of eukaryotic cell participating in many cellular activities. Most important of those is providing the cell with more than 90% of the energy needed via the oxidative phosphorylation system (OXPHOS). The OXPHOS system consists of five multi protein complexes (I, II, II, IV, V) and two electron carriers, coenzyme Q and cytochrome c embedded
in the inner membrane (figure 2.1). Out of the 85 OXPHOS proteins, 13 are encoded by the mtDNA and rest by the nuclear genome (Smeitink et al 2001). The function of the OXPHOS is to create an electrochemical gradient by coordinated passage of electrons along the complexes and pumping of protons across the inner membrane by the OXPHOS-complexes, which is then used by F1F0-ATPase to synthesise ATP (Saraste 1999 and Nelson & Cox 2000).

Figure 2.1: Schematic representation of the electron transport chain. Electrons are transferred through the complexes I-IV to oxygen and water is formed. Hydrogen ions from the matrix are pumped at the same time from the matrix to the inter membrane space by complexes I, III and IV. The electrochemical gradient formed by this process is the driving force for the ATP synthesis by complex V. (OMM – outer mitochondrial membrane, IMM – inner mitochondrial membrane).

In addition to its role in energy production mitochondria are involved in cellular metabolism including; fuel catabolism; glycolysis, carbohydrate, fatty acid breakdown (beta oxidation) and amino acid metabolism that all produce acetyl-Co-A. The end product of glucose break down in cytosol (glycolysis) is pyruvate that is transported to mitochondria where it is converted to Acetyl-Co-A. Other source of Acetyl-Co-A is the fatty acid break down (beta-oxidation) that occurs inside mitochondria and it is also produced by the breakdown of amino acids. Acetyl-Co-A is the initiator of citric acid cycle, which in turn leads to formation of NADH and
succinate that are used by the OXPHOS system to generate ATP (Nelson & Cox 2000). Mitochondria also have a role in calcium signalling and cell death (Patron et al 2013) and are also involved in urea production, cardiolipin and lipid biosynthesis, providing building blocks for amino acids and in biosynthesis of ubiquinol (Coenzyme Q), (Nelson & Cox 2000). In addition, mitochondria participate in metal metabolism, being involved in heme and Fe-S cluster (Lill & Muhlenhoff, 2008) and steroids (Sewer & Li 2008) synthesis.

2.2 Mitochondrial genetics

The possibility of cytoplasmic inheritance in plants was suggested in the early 20th century by the botanist von Wettstein, who named it the plasmon theory (von Wettstein 1927); later the cytoplasmic inheritance was demonstrated in yeast and the heritable unit was designated as the Rho factor (Ephrussi et al 1949). However, the identity of the cytoplasmic units of inheritance was not known until 1963 when Nass & Nass found that mitochondria have their own DNA and its sequence was completed in 1981 (Anderson et al 1981).

Nuclear and mitochondrial DNA (mtDNA) have several differences; mammalian mtDNA is circular and exists as many copies/cell (Anderson 1981) whereas nuclear DNA is linear and organised in 23 pairs of chromosomes. MtDNA and nuclear DNA have different genetic codes (Barrell et al 1979) and hence the translation machineries residing in cytosol (for nuclear gene expression) and mitochondria can only translate genes encoded in those compartments. Most of the mtDNA is coding DNA and the genes lack introns, but it also contains two non-coding regions that have regulatory functions and take approximately 5% of the genome (Attardi 1985). In contrast, 99% of nuclear DNA is non-coding and 1% codes for genes, which consist of coding and non-coding regions and mRNA processing leads to removal of the non-coding regions before translation takes place. In addition nuclear DNA is much larger, e.g. one chromosome can have approximately 50 000 000 – 250 000 000 base pairs whereas mtDNA is only just over 16 000 base pairs long (Anderson et al 1981). Nuclear DNA is a mixture of paternal and maternal DNA whereas mtDNA is only maternally inherited (Hutchison et al 1974). MtDNA mutation rate is estimated to be 10 fold higher than that of the nuclear DNA (Brown et al 1979),
which is thought to occur due to local environmental factors by being close to the OXPHOS and ROS production, less efficient DNA repair mechanism and also due to the higher replication rate compared to nuclear DNA (von Wurmb-Schwark 2007).

Mammalian mtDNA is inherited maternally (Hutchison et al 1974 and Giles et al 1980) and displays non-Mendelian inheritance (Giles et al 1980). The mechanism of maternal inheritance was enlightened by a study undertaken by Shitara et al (2000) where they showed that during fertilization spermatozoa mitochondria are destroyed soon after it enters the egg. There is only one report of paternally inherited mtDNA associated with disease (Schwartz & Vissing, 2002), which was thought to develop from the recombination of the mutated paternal mtDNA with the maternal mtDNA (Kraytsberg et al 2004). Mitochondria undergo replicative segregation at cell division, have high copy number in a cell (Di Mauro & Davidzon 2005) and can have variable heteroplasmy levels – meaning mtDNA population within the cell can contain both mutant and wild type molecules (Holt et al 1988) and the respective abundance can range from less than 1% to greater than 99% of the mtDNA population. Mutations that result in loss of function are usually tolerated unless they exceed a certain threshold, beyond which they become pathogenic. Most human pathogenic mutations are heteroplasmic, with the threshold for disease manifestation varying according to the specific mutation, individual and tissue (see later).

2.2.1 Organisation of mtDNA

The human mitochondrial genome is a 16,569 base pairs (bp) in length (Anderson et al 1981), a closed, circular double stranded molecule as illustrated in figure 2.2 (Anderson 1981, Chomyn et al 1985 and Andrews et al 1999) and is found in the matrix compartment. Each cell can contain hundreds to thousands of copies of the genome depending on the cell type and each mitochondrion contains several copies of mtDNA (Bogenhagen & Clayton 1974 and Satoh & Kuroiwa 1991). The mtDNA encodes for 37 genes of which two encode for ribosomal (r)RNA and 22 for transfer (t)RNA while 13 encode for essential components of the OXPHOS. The mitochondrial DNA-molecule consists of two strands, the heavy (H) strand that is guanidine rich and the light (L) strand that is cytosine rich named so because of their
different behaviour in denaturing caesium chloride gradient (Kasamatsu & Vinograd 1974). Heavy- and light strand are both transcribed as polycistronic molecules. Heavy-strand encodes for 28 genes and the light-strand for nine genes (Anderson et al 1981). The organisation of mtDNA is economical, so that it contains no introns and some genes overlap or are separated only by few base pairs (Anderson et al 1981, Shadel & Clayton 1997 and Fernandez-Silva et al 2003). MtDNA contains only two short non-coding regions (NCR). The major NCR is a 1.1 kb long region that contains the origin of replication for the heavy-strand (O_H) and promoters for transcription of the light-strand and heavy-strand. A shorter 32 nt NCR is located at origin of replication for the light strand (O_L) (for a review see Shadel & Clayton 1997, Wanrooij et al 2010/2012 and Nicholls & Minczuk 2014).

Figure 2.2: The human mitochondrial DNA gene map. Human mtDNA is a double-stranded, closed, circular molecule of 16.6 kb. Single letters indicate the positions of the corresponding tRNA genes. ND, NADH dehydrogenase genes; cyt b, cytochrome b gene; COX, cytochrome c oxidase genes; A6/8, ATP synthase genes 6 and 8; 12S/16S, ribosomal RNA genes (Reprinted by permission from John Wiley & Sons, Inc, IUBMB Life, Spelbrink 2010).
A region of the 1.1 kb long NCR is often occupied by a 650 nt long third strand forming a displacement loop (D-loop), (Kasamatsu et al 1971 and Nicholls & Minczuk 2014). This 650 nucleotides (nt) long third strand is generated by premature replication termination of de novo heavy strand synthesis at the termination-association sequence (TAS) creating the third strand (Arnberg et al 1971, ter Schegget et al 1971, Doda et al 1981). Kasamatsu et al (1971) referred to the third strand 7S DNA because of its sedimentation coefficient. The functions of the D-loop remain unknown, but it has been hypothesised to be involved in nucleoid organisation (He et al 2007), as an increased access sites for proteins involved in mtDNA transcription and replication (Nicholls & Minczuk 2014) or in control of mitochondrial nucleotide pools (Antes et al 2010). A much shorter non-coding region is located at the \( O_L \), two-thirds downstream of heavy strand origin. According to strand-displacement mtDNA replication model (see below) a stem loop structure at \( O_L \) is required for primer formation for lagging strand synthesis (Wong & Clayton 1985a and Fuste et al 2010) and it has been shown to be essential in \textit{in vivo} studies for mtDNA maintenance and is functionally conserved in evolution (Wanrooij et al 2012).

### 2.3 Mitochondrial DNA maintenance

Because mtDNA codes for 13 polypeptides of the OXPHOS and tRNAs and rRNAs needed for their translation (Anderson et al 1981 and Chomyn et al 1985) correct copying of mtDNA is important for the maintenance of OXPHOS and ATP production. Although mitochondria have their own genome, it’s thought to have lost most of it during evolution and is dependent on nuclear genes (Margulis 1970 and Gray 1999). Why have mitochondria evolved to “keep their DNA” and the cost of maintaining it? Lane & Martin (2010) discuss that the benefit of mtDNA encoding parts of OXPHOS enables individual mitochondria quickly to respond to changes in energy demands and the cost of maintaining mtDNA is outweighed by the benefits. MtDNA maintenance is fully dependent on nuclear genes and many of these proteins are known (see below) but it is clear that the machinery is more complex. For example proteins can have dual functions operating both in nucleus and mitochondrial compartment in DNA maintenance, and to add to the complexity they
may not possess a known mitochondrial targeting sequence like in DNA2 (Duxin et al 2009).

The human mitochondrial proteome consists of approximately 1500 proteins of which 99% is encoded in the nucleus and translated in the cytoplasm (Calvo & Mootha 2010). For mitochondria to be able to use the transferred genes some regulatory elements needed to be acquired, including mitochondrial targeting signal (N’, C’-terminal or internal), which is proteolytically removed after import (Gakh et al 2002) and regulatory elements for nuclear transcription and cytoplasmic translation (Lithgow & Schneider 2010). The targeting sequences can be complex and varied, but a simple N’-terminal sequence with basic, amphipathic, helical sequence is able to translocate proteins from the cytoplasm to mitochondria (Maccecchini et al 1979, Hay et al 1984, Lemire et al 1989 and Lithgow & Schneider 2010). A 15 to 55 amino acids long N’terminal targeting sequence is the most commonly found in proteins designated to mitochondria comprising an amphipathic α-helical segments with a net positive charge (Roise & Schatz 1988 and Vögtle et al 2009. In addition, a system for the transport of the proteins from cytoplasm to mitochondria has been acquired (Dudek et al 2013 and Hewitt et al 2014 for reviews).

2.3.1 mtDNA transcription

MtDNA contains three promoters for transcription, the light-strand promoter (LSP) and the heavy strand-promoter 1 and 2 (HSP), two of them located in the major NCR, while the third promoter HSP2 is located down stream of HSP1 within the sequence of the mitochondrial tRNA for phenylalanine (Montoya et al 1982, Montoya et al 1983 and Chang & Clayton 1984). There has been arguments on the existence of HSP2. The activity of HSP2 was first demonstrated by Martin and co-workers in 2005, but Litonin et al (2010) have questioned the existence of HSP2 as in their in vitro studies using recombinant proteins they were unable to see any HSP2 transcripts. In contrast, two other studies argue against this and show evidence for transcription from HSP2 (Lodeiro et al 2010 and Zollo et al 2012). HSP2 and LSP produce long polycistronic transcripts that are post-transcriptionally processed for final transcripts (Bibb et al 1981, Clayton 1984 and Reichert et al
1998), whereas HSP1 transcription produces a truncated transcript transcribing for two tRNAs and two rRNAs (Montoya et al 1983 and Martin et al 2005). Premature termination of transcription from HSP1 occurs immediately downstream of 16S RNA gene (Montoya et al 1983) and the mitochondrial transcription terminator factor 1 (mTERF1) is partly responsible for the termination (Kruse et al 1989). mTERF1 has also been shown to have a role in stimulation of transcription from the HSP1 promoter, which was dependent on the existence of the mTERF termination site. Authors concluded that the simultaneous binding of mTERF to HSP1 promoter and mTERF termination sequence creates a looping-out rDNA structure important for the reinitiating effect on transcription (Martin et al 2005). In vivo work on mTERF1 knock out mouse carried out in Larsson laboratory contradicts the earlier findings showing no role of mTERF1 in initiation or termination of mtDNA H-strand transcription. Instead they suggest a model where mTERF1 acts as a “traffic light” to prevent L-strand transcription proceeding and hence prevents interference at the LSP promoter (Terzioglu et al 2013).

Mechanism for termination of transcription from HSP2 is not known, but a termination region has been identified upstream of the tRNA$^{\text{Phe}}$ gene in mice (Camasamudram et al 2003). Light-strand transcription and heavy-strand replication is coupled so that transcription from LSP also produces the primer for replication (Walberg & Clayton 1983 and Chang & Clayton 1985).

The machinery for transcription includes mitochondrial RNA polymerase (POLRMT), a 1230 amino acid protein (Tiranti et al 1997), whose activity was first characterised by Shuey & Attardi (1985) and provides RNA polymerisation and promoter recognition (Gaspari et al 2004). In 1988 Fisher & Clayton purified the 24.4 kDa protein, referred as the transcription factor A (TFAM) that is needed for transcription to take place. Binding of TFAM in non-sequence specific manner, upstream of promoter introduces changes to mtDNA structure (Fisher et al 1987 and Fisher et al 1992) to facilitate binding of POLRMT and transcription factor B1 (mtTFB1) or B2 (mtTFB2) to the initiation site (Fisher & Clayton, 1988, Shadel & Clayton, 1997, Falkenberg et al 2002 and Gaspari et al 2004), of which mtTFB2 is thought to be the primary transcription factor (Cotney et al 2007). Role of mtTFB1 is thought to be primarily in 12S rRNA methylation, which in turn is important for ribosome biogenesis and mitochondrial translation (Cotney et al 2007 and Metodiev et al 2009). However, Shut et al (2010) argue that initiation of transcription can also
take place without TFAM. They demonstrated: “that it only takes two to tango” meaning that the basal transcription apparatus can also work as a two-component system, POLRMT and mtTFB2, *in vitro*. They suggest that mammals have evolved an additional regulatory system mediated by TFAM to their two-component transcription mechanism, so that TFAM concentration seems to regulate which promoter is used, LSP or HSP1, (Shutt et al 2010) suggesting that TFAM levels regulate the need for replication and transcription (Shutt et al 2011). Recent structural studies show that the binding of TFAM to the TFAM responsive element (TRE) and the subsequent U-turn formation are important for the transcriptional activation at LSP and HSP1 (Malarkey et al 2012, Ngo et al 2011 and Rubio-Cosials et al 2011). Study by Lodeiro et al (2012) suggests a role for TFAM in regulation of HSP2 transcription by binding to transcription start site hence preventing POLRMT and TFB2 binding to the promoter (Lodeiro et al 2012), but transcription from HSP2 can occur in the absence of TFAM and is only dependent on POLRMT and TFB2 (Lodeiro et al 2010).

Few other proteins were also shown to have some role in mtDNA transcription; mTERF 2 and 3, mitochondrial transcription elongation factor (TEFM), (Minczuk et al 2011) and the mitochondrial ribosomal protein L12 (MRPL12), (Wang et al 2007), but their role in mtDNA transcription begs for further research (reviewed in Shutt & Shadel 2010).

### 2.3.2 mtDNA replication

Replication of mtDNA takes about an hour, is not dependent on the cell cycle like nuclear DNA replication, but instead is continuously amplified (Bogenhagen & Clayton 1977) and is fully dependent on nuclear genes that are translated in the cytoplasm and the proteins imported into mitochondria (Shadel & Clayton 1997).

Clayton 1981, Kang et al 1997 and Bogenhagen & Clayton 2003). In this model mtDNA replication is initiated at different times from two unidirectional origins, one located on the H-strand (O_H) and the other in the L-strand (O_L) and proceeds asynchronously and asymmetrically.

The leading strand replication starts at the heavy strand origin and proceeds unidirectionally until the light strand origin is reached two-thirds around the circular genome. When the light strand origin is exposed replication initiates to the opposite direction. Brown et al (2005) have also suggested a second L-strand origin, but there has been no follow up study to examine the use of such second origin in cell. Holt and co-workers were first to challenge this model in 2000 describing the stand-coupled mtDNA replication model, arising from data using neutral two-dimensional agarose gel electrophoresis (2DNAGE) where double stranded replication intermediates (RI) were found, suggesting a coordinated leading and lagging strand DNA synthesis (Holt et al 2000, Bowmaker et al 2003 and Yasukawa et al 2005). In this model, mtDNA replication is initiated at a broad zone downstream of the major NCR and proceeds bidirectionally until reaching the D-loop region. The lagging stand synthesis is supported by leading strand synthesis early on in replication and the lagging-strand synthesised first as Okazaki fragments.

Work at Holt laboratory later led to modification of the model when they discovered stretches of RNA in the lagging-strand and proposed yet another model termed ribonucleotide incorporation throughout the lagging strand (RITOLS) model (Yang et al 2002 and Yasukawa et al 2006). In this model, the strand-synchronous mtDNA replication proceeds unidirectionally and the second strand is first laid down as RNA that is turned into DNA later. Recently Holt laboratory has refined the RITOLS model and proposed that RITOLS proceeds via “bootlace model” so that RITOLS intermediates are not RNA primers, but preformed RNA hybridises on the displacement strand that are processed further to fully replicated mtDNA (Reyes et al 2013).
Figure 2.3: Schematic representation of mammalian mtDNA replication models. A) strand-displacement B) strand-coupled and C) the RNA incorporated throughout the lagging strand (RITOLS). A) In the asymmetric or strand displacement model replication of the H-strand is initiated at O_H with accompanying displacement of the H-strand thus forming a D-loop. The light strand is single stranded until synthesis of the nascent heavy strand exposes O_L, where synthesis of the L-stand is initiated in the opposite direction. In the strand-coupled model, bidirectional replication is initiated from a zone downs stream of O_H followed by progression of the two forks around the mtDNA circle. In the RITOLS model replication of the leading strand initiates similar to the strand-displacement model but the lagging strand is initially laid down as RNA before being converted to DNA (Kasiviswanathan et al 2012, Yasukawa et al 2006 and Krishnan et al 2008). It is proposed by the bootlace model of mtDNA replication that preformed (L-strand) transcripts, including complementary tRNA and mRNA hybridize to the template lagging strand of mammalian mtDNA as leading strand DNA synthesis proceeds (Kasiviswanathan et al 2012 and Reyes et al 2013). (Picture adapted from Krishnan et al 2008 and Brown et al 2005).
Several studies have investigated the start site for the origin of replication at the H-strand. As discussed in Holt & Reyes (2012) the first identified site for O_H was at nucleotide 191 for the strand-asynchronous mtDNA model of replication (Crews et al 1979), which Attardi later on revised to be located at nucleotide 57 (Fish et al 2004) and Pham et al (2006) have suggested site upstream of nucleotide 191 as the site of O_H. There has been a lot of controversy on the initiation of replication that follows the arguments for mtDNA replication mode. In strand-coupled replication Reyes et al (2005) argued that mtDNA heavy-strand replication is initiated from a broad zone. In the RITOLS model work by the Holt laboratory has suggested two sites for the origin of replication O_H and Ori-b at nucleotides 191 and 16,197 respectively (Yasukawa et al 2005). See later on for more details.

Holt et al (2000) have also postulated that mtDNA may switch from one mode of replication to another according to environmental conditions. The strand-displacement model and RITOLS have in common the fact that there is a delayed synthesis of the lagging-strand. However, instead of mitochondrial single-stranded DNA-binding protein (mtSSB) association with the displaced H-strand, in the strand-displacement model, RNA is deposited in the RITOLS model. Why does RNA incorporation occur? Yasukawa et al (2006) speculated that RNA might block mtDNA transcription until replication has taken place or they may act as stabilizing agents for the displaced DNA. Clayton and co-workers have argued against the RITOLS and strand-coupled models suggesting the 2DNAGE method produces artefacts representing the RIs. They also argue that the RIs are in fact mtDNA transcription intermediates and not incorporation of ribonucleotides (Bogenhagen & Clayton 2003 and Brown & Clayton, 2006). However, Holt & Jacobs (2003) argued that Clayton and co-workers have failed to grasp the central point of their model.

Work by Yang et al (2002) suggests that the displacement replication model can be explained by the degradation of RITOLS replication intermediates during sample preparation, thus creating a single-stranded H-strand, and hence should be considered an artefact. In both strand-displacement model and the RITOLS model O_L is a major site of initiation of second strand DNA synthesis (Yasukawa et al 2006). Furthermore, mutation studies suggest that a stable stem loop structure at O_L is necessary for primer synthesis (Wanrooij et al 2012).
2.3.3 mtDNA replication machinery

During replication the parental dsDNA needs to be unwound to ssDNA intermediates and nascent strands are synthesized using the parental strand as templates. Replication is dependent not only on the replication machinery, but also on the transcription machinery, because transcription from the light-strand provides the RNA primer needed for replication initiation at O\textsubscript{H} for the leading-strand (Xu & Clayton 1996). The replication machinery is related in part to the T7 bacteriophage machinery and is relatively simple (for a review see Shutt & Gray, 2006), (figure 2.4). Most of the replication machinery (POLG, Twinkle and POLRMT) is homologous to the bacteriophage T7 machinery (Ropp & Copeland 1996, Spelbrink et al 2001, Tiranti et al 1997) with one exception of mtSSB that resembles more bacterial SSB (Tiranti et al 1993). A minimal mtDNA replisome in vitro was demonstrated to consist of the helicase Twinkle, mtSSB and the mtDNA polymerase gamma (POLG) (Korhonen et al 2004). Later Wanrooij et al (2008) demonstrated POLRMT to prime the lagging strand mtDNA replication. The proteins known to be involved in mtDNA replication is discussed in next section and the role of these proteins in nucleoids is discussed in later section.

Figure 2.4: Proteins involved in mtDNA replication. The Twinkle helicase (black) unwinds mtDNA to 5’to 3´direction. RNA primer for lagging strand synthesis is synthesised ty POLRMT (hexagonal). MtSSB (white) stabilises the ssDNA and mtDNA synthesis is carried out by POLG (dark grey and accessory subunit light grey).
2.3.3.1 Mitochondrial DNA helicase, Twinkle

Helicases are needed during replication to unwind the duplex DNA ahead of the DNA polymerase (for a review Patel & Picha 2000). Mitochondrial helicase Twinkle was discovered in search for a mitochondrial helicase and shown to associate with mtDNA in punctate nucleoid structures and have helicase activity in vitro (Spelbrink et al 2001). It was also realised to be in a critical region on chromosome 10q that is linked to autosomal dominant progressive external ophthalmoplegia (adPEO), (see later for the role of Twinkle in mitochondrial diseases and nucleoids), (Suomalainen et al 1995).

Twinkle is a 77 kDa protein with an N’terminal domain, a linker region and a C’terminal helicase domain, and is structurally similar, mostly in the helicase domain, to bacteriophage T7 gene 4 primase/helicase (T7 gp4) (Spelbrink et al 2001 and Shutt & Gray 2006). Unlike the C’terminal domain, the N’terminal domain of Twinkle has low sequence similarity to T7 gp4, but is classified as a prokaryotic DnaG type primase (Ilyina et al 1992). The N’terminal domain of Twinkle has lost the critical residues thought to be important for primase activity (zinc-binding domain and polymerase domain). Hence, the role of the N’terminal domain function differs from that of T7 gp4, so that the primase activity is considered lost in mammalian Twinkle (Shut & Gray 2006) and instead is thought to be important in ssDNA/dsDNA binding and in DNA unwinding activity (Farge et al 2008). Work carried out in Drosophila has found the N’terminal domain to contain Iron-Sulfur clusters that enhance protein stability and carries out ssDNA and dsDNA binding activity (Stiban et al 2014). The C’terminal domain has helicase activity in 5’to 3’direction (Korhonen et al 2003) and the linker region is needed for multimerisation (Goffart et al 2009), but in contrast to T7gp4 it does not need cofactors for oligomerisation (Ziebarth et al 2010). The C’terminal domain includes Walker A and B motifs that are important in nucleotide hydrolysis, providing the energy for helicase activity (Singleton et al 2007). Sen and co-workers (2012) suggest an additional role of Twinkle as an annealing helicase, but the importance of such function was not investigated. The protein is found in in vitro experiments as hexameric and heptameric form depending on cofactors and salt conditions (Spelbrink et al 2001, Ziebarth et al 2007, Goffart et al 2009 and Ziebarth et al 2010). In T7 gene 4 primase/helicase the heptameric structures cannot bind DNA.
and it is though that the seventh subunit is lost when DNA binding occurs and the heptameric ring opens (Crampton et al 2006). The physiological role of the mammalian heptameric form remains to be elucidated.

Twinkle alone can unwind short stretches of dsDNA (Korhonen et al 2003) and it can form a minimal replisome in vitro together with POLG and mtSSB that can generate a DNA product of 16 kb (Korhonen et al 2004). In contrast to other systems like E. Coli and bacteriophage T7 (Davey & McDonnel 2003) Twinkle can load on the circular template without a loading factor and initiate mtDNA replication together with POLG (Jemt et al 2011) and its activity is stimulated by mtSSB (Korhonen et al 2003). Tyynismaa et al (2004) have shown that mtDNA copy number is increased in mouse over expressing Twinkle. In addition, in cultured cells where Twinkle expression is knocked out by siRNA, mtDNA copy number decreases. Demonstrating that Twinkle has an important role in mtDNA maintenance. Later studies have also shown mtDNA copy number to be directly proportional to Twinkle concentrations suggesting Twinkle to have a regulatory role in mtDNA replication (Milenkovic et al 2013). Twinkle disease mutant expression cause replication stalling (Wanrooij et al 2007 and Goffart et al 2009) by disrupting the oligomerisation, nucleotide hydrolysis or helicase activity depending on the location of the mutation (Goffart et al 2009 and Longley et al 2010). In Twinkle mouse knockout studies Milenkovic and co-workers demonstrated Twinkle to be essential for embryonic development, hence showing it to be the only replicative mtDNA helicase in vivo (Milenkovic et al 2013). Over 40 Twinkle disease mutations have been described resulting in multiple mtDNA deletions and depletion ultimately resulting in neuro-muscular diseases via compromised OXPHOS system (Wanrooij & Falkenberg 2010).

Spelbrink et al (2001) also described a variant polypeptide of Twinkle assigned as Twinky that lacks part of the helicase H4 motif plus the polar C’terminal tail. It does not form hexamers, has no helicase activity and does not localise to nucleoids. The role of Twinky is still not known.
2.3.3.2 Mitochondrial transcription factor A

A mammalian TFAM protein was first discovered for its ability to stimulate transcription by mitochondrial RNA polymerase (Fisher & Clayton 1985). Later work reported it to have functions as a DNA packaging protein because it was able to introduce negative supercoils into mtDNA, bend it and bind abundantly in a cooperative fashion and non-specific manner around the entire mitochondrial genome (Fisher et al 1992 and Alam et al 2003). In vivo work has shown TFAM to exist in nucleoids (Garrido et al 2003), which is discussed further in section 2.5.4.1.

TFAM has a molecular weight of 25 kDa, is a member of the high-mobility-group protein family (Fisher & Clayton 1988 and Parisi & Clayton 1991) and has been reported to function as a homodimer (Kaufman et al 2007). However, there are recent reports that argue that the specific DNA binding also happens in a monomeric manner (Rubio-Cosials & Solà 2013) and monomer sliding and non-specific TFAM patch formation that cause more flexible mtDNA turns, leads to nucleoprotein structures (Farge et al 2012). The crystal structure of TFAM has revealed it to be able to make sharp U-turns in mtDNA, suggesting a mechanism for its function as an mtDNA packaging protein and in transcription (Ngo et al 2011 and Rubio-Cosials et al 2011). Both studies have shown TFAM binding to the same region of LSP and the U-turn at the LSP1 is thought to be a prerequisite for mtDNA transcription (Ngo et al 2011 and Rubio-Cosials et al 2011). TFAM also makes a U-turn at the HSP1, but it is not thought to be important for transcriptional activation. U-turns at unspecific regions of mtDNA and TFAM dimerization are thought to be important for mtDNA packaging, but the dimerization is not required for DNA bending or transcriptional activation by TFAM (Ngo et al 2014). TFAM has high affinity to DNA and binds specifically at LSP and HSP in the NCR (Fisher et al 1987, 1989. Fisher & Clayton 1988) and has a high affinity to other regions of the mtDNA as well (Gadaleta et al 1996, reviewed in D’Errico et al 2005). The concentration of TFAM determines whether LSP or HSP promoter is activated, so that at low TFAM concentrations, LSP is activated and at increasing TFAM concentration, the transcription switches to HSP (Fisher & Clayton 1988 and Shutt et al 2010). Because LSP provides the primer for mtDNA replication and HSP produces two rRNAs it was hypothesises that this could provide an important
regulation between mtDNA replication and gene expression via regulation of TFAM concentrations (Shutt et al 2010).

There are reports on the role of TFAM in the control of mtDNA copy number. In vivo, studies have shown that TFAM over expression mice (expressing the human TFAM) demonstrate increased mtDNA copy number (Ekstrand et al 2004). In vitro cell culture work on a TFAM C’terminal deletion mutant that is not capable of stimulating transcription shows the mtDNA amount to correlate with that of TFAM, independent of transcriptional activity (Kanki et al 2004). Also, embryos from TFAM knockout mouse lack mtDNA (Larsson et al 1998) demonstrating the importance of TFAM in mtDNA maintenance. As human TFAM does not stimulate transcription in mice, it is thought that the mtDNA copy number control is independent of transcription. However, in vitro studies on overexpression of wild type TFAM show no change in mtDNA copy number, but instead result in increased transcription rates (Maniura-Weber et al 2004). Clearly, the role of TFAM in mtDNA copy number control begs for further research.

Later in vitro studies using 2DNAGE show that knock down or over expression of TFAM in cell culture results in change on mtDNA replication intermediates suggesting that TFAM influences the mode of mtDNA replication via its combined effects on transcription and mtDNA organization (Pohjoismäki et al 2006). In Rho^0 cells TFAM levels are reduced (Larsson et al 1994), and it is thought that the relationship between mtDNA and TFAM is dynamic in order to maintain optimal TFAM:mtDNA ratio. RNAi studies on Lon protease in Drosophila Schneider cells show accumulation of TFAM in knockdown cells causing an increase in TFAM:mtDNA ration, an increase in mtDNA copy number and inhibition of mitochondrial transcription. Authors conclude that Lon protease stabilise the TFAM:mtDNA ratio by selective degradation of TFAM and therefore regulates mtDNA transcription (Matsushima et al 2010). Later work has shown that the phosphorylation of TFAM by cyclic-AMP dependent protein kinase A (PKA) selects TFAM for degradation by Lon protease and leads to dissociation of TFAM from DNA and therefore to reduced transcriptional activation (Lu et al 2013).

There are several reports on how many TFAM molecules there are per mtDNA molecule. Fisher et al (1992) estimated the yield of TFAM to be a minimum of ~1 TFAM molecule per 1000 bp of mtDNA. However others have estimated the molar ratio of ~1 TFAM molecule per 10 bp of mtDNA and postulated the whole genome
to be covered with TFAM and hence its main function to be mtDNA packaging (Takamatsu et al 2002, Alam et al 2003 and Ekstrand et al 2004). Furthermore, other studies have found TFAM tightly associated with mtDNA and reported it to be the main component of the nucleoid structure (Alam et al 2003, Garrido et al 2003, Kaufman et al 2007 and Kukat et al 2011).

2.3.3.3 Mitochondrial single stranded DNA binding protein

For DNA maintenance, both copying and repair of dsDNA need to be unwound and manipulated exposing patches of ssDNA. Therefore, mtSSB is an important factor in mtDNA maintenance as it protects the DNA and prevents folding or re-annealing of ssDNA (Pavco & Tuyle 1985). The human mtSSB is a 16 kDa (Curth et al 1994) protein that wraps around ssDNA as a tetramer in a non-sequence specific manner, covering 50-70 nt per tetramer (Mignotte et al 1985 and Yang et al 1997). Several studies have shown the importance of mtSSB in mtDNA maintenance. In vitro studies show mtSSB to stimulate POLG activity (Mignotte et al 1988) as well as to stimulate Twinkle helicase activity (Korhonen et al 2003 and Oliveira & Kaguni 2010). Takamatsu et al (2002) also suggests mtSSB to have a role in stabilising the D-loop region together with TFAM. Studies on mtSSB mutant fly have shown mtSSB to be essential for both mtDNA replication and development as mutant flies lose their mtDNA and die before the larval or pupal stage of development (Maier et al 2001). In addition, in Drosophila siRNA against mtSSB results in depletion of mtDNA (Farr et al 2004). Similarly, siRNA against mammalian mtSSB in Hela cells results in gradual decline in mtDNA (Ruhanen et al 2010) and in yeast deletion of mtSSB results in total loss of mtDNA (Van Dyck et al 1992). Fuste et al (2010) also showed mtSSB to suppress non-specific primer synthesis on ssDNA and stimulate initiation from O_L.

2.3.3.4 Mitochondrial DNA polymerase

A great deal of information about the mtDNA maintenance machinery has first come from yeast, and the yeast POLG MIP1 studies gave the first indication of the important role of POLG in mtDNA maintenance (Genga et al 1986). Later studies
have also shown the importance of POLG in mammalian mtDNA replication (Gray & Wong 1992 and Ropp & Copeland, 1996). Human POLG is an asymmetric holoenzyme forming a heterotrimer containing two 55 kDa accessory subunits, POLG2, and one 140 kDa catalytic subunit, POLG1 (Yakubovskaya et al 2006 and Lee et al 2009). The catalytic subunit polymerises DNA in 5´to 3´ direction and also has a 3´to 5´exonuclease domain for proofreading activity (Gray & Wong, 1992 and Longley et al 1998). POLG2 subunits are involved in enhancing the DNA binding, processivity and the catalytic activity of the enzyme (Carrodeguas et al 1999 and Lim et al 1999). The POLG2 subunit lying close to POLG1, stimulates the holoenzyme binding to DNA. The other, more distant POLG2, accelerates nucleotide incorporation. They are also needed to maintain the holoenzyme structure (Lee et al 2009 and Lee et al 2010). There is also evidence that POLG2 has a role in nucleoid organisation, separate from its accessory function to POLG1 (di Re et al 2009).

2.3.3.5 Other enzyme activities required for mtDNA replication

For DNA replication to take place primer formation and removal and other enzyme activities are needed. These include ligases, RNase H and topoisomerases. It is thought that POLRMT functions as a primase at O_H and O_L. Therefore, in addition to its role in mtDNA transcription POLRMT also has been suggested to have a role in mtDNA replication. Wong and Clayton (1985a/b) were first to report O_L specific primer activity, but did not recognise the enzyme involved. The first indication that POLRMT may be the mitochondrial lagging-strand primase came in 1990 when Tsurumi & Lehman identified primase activity in Vero cell mitochondria after virus infection. They postulated the enzyme responsible might be POLRMT based on the molecular weight and enzymatic activity identified (Tsurumi & Lehman 1990). Later studies have shown POLRMT to have primase activity required for initiation of DNA synthesis from the O_L in vitro (Wanrooij et al 2008). García-Gómez et al (2013) have also identified another enzyme with primase/polymerase activity, PrimPol, which localises to both nucleus and mitochondria. They have demonstrated PRIMPOL silencing in cells to cause impaired mtDNA replication and suggested
that PrimPol reinitiates replication at sites where mtDNA replication is paused due to lesions.

DNA ligases are needed during DNA replication, repair and recombination to seal the nicks that occur during these processes (Ellenberg & Tomkinson 2008). Nuclear DNA synthesis proceeds by coupled leading and lagging strand synthesis where the laggings strand consists of Okazaki fragments (Alberts et al 2002). The maturation of the Okazaki fragments to a continuous DNA occurs via a two-step process in which, the RNA primer that drives the DNA synthesis of Okazaki fragments, is first removed. This occurs in the nucleus via Ribonuclease H2, flap endonuclease 1 and endonuclease/helicase DNA2. In the second step, the Okazaki fragments are sealed together brought about by ligase activity (Alberts et al 2002 and Kao & Bambara 2003).

Evidence for mitochondrial DNA ligase activity was first described in 1976 by Levin & Zimmerman and but it was much later when DNA Ligase III activity was confirmed (Pinz & Bogenhagen 1998) and the protein was shown to translocate into mitochondria (Laksmipathy & Campbell 1999). Decreased levels of Ligase III in cell culture results in decrease in mtDNA levels and increase in ssDNA breaks in mtDNA. The role of Ligase III in mitochondrial DNA repair was suggested and postulated that it also is the ligase responsible for sealing the nicks on mtDNA during replication (Laksmipathy & Campbell 2001). Later in vivo work on conditional knock out mice also show reduction of ligase III to result in loss of mtDNA (Gao et al 2011). However, the role of Ligase III in mtDNA replication was not addressed until Ruhanen and co-workers in 2011 used an in vitro cell culture system to study the role of ligase III in mtDNA replication. Authors first deplete mtDNA and used Ligase III RNA interference during the mtDNA recovery following depletion to show that knock down of ligase III leads to delay on the nick sealing on replicating mtDNA molecules and in the absence of ligase III mtDNA copy number is not restored. They conclude that ligase III is the ligase in mitochondria and functions during the maturation process of Okazaki fragments (Ruhanen et al 2011).

Mice studies have shown RNase H1 to be essential during development for mtDNA amplification (Cerritelli et al 2003). Gaidamakov et al (2005) showed dimerisation to give processivity for RNase H1 and postulated this may have a role in mitochondria in processing long DNA/RNA hybrids that may be present during
mtDNA replication. Furthermore, 2DNAGE experiments have identified blocks of RNA of 200-600 nt in length that anneal to the lagging strand template during DNA replication that are RNase H-sensitive, supporting the important role of RNase H1 in mtDNA replication (Yasukawa et al 2006). Ruhanen et al (2011) show depletion of RNase H1 to lead to halting of mtDNA replication. They postulated that the impaired primer removal of RNase H1-depleted cells arrests mtDNA synthesis.

Human mitochondrial topoisomerase I (mtTOPO1) was discovered in 2001 by Zhang et al who later showed that interference of TOPO1 activity results in reduction of 7S DNA suggesting a possible role in mtDNA replication (Zhang & Pommier 2008). Whereas TOPO1 has a role exclusively in mitochondria (Dalla Rosa et al 2009), topoisomerase IIIα (TOPOIIIα) localisation has been shown both in nucleus and mitochondria (Wang et al 2002), but the specific function of mammalian TOPOIIIα in mitochondria remains unclear. Fly and protozoan studies have suggested it to play a role at the end of replication by removing the last few intertwines of the parental mtDNA strands (Kasiviswanathan et al 2012).

2.3.4 Initiation and termination of replication

Transcription from LSP produces the primer needed for replication initiation at O₁ for the leading strand replication (Crews et al 1979, Gillum & Clayton 1979, Cantatore & Attardi 1980, Chang & Clayton 1985 and Pham et al 2006). There are currently two theories how the primer for heavy-strand replication is produced. One hypothesis is that the primer is produced by site-specific cleavage of the full transcript by RNase mitochondrial RNA processing protein (RNase MRP) (Chang & Clayton 1987). This theory has been questioned because majority of RNase MRP is localised to nucleolus (Kiss & Filipowicz 1992) and the amount of RNase MRP in the mitochondria is insufficient to carry out this function (Kiss & Filipowicz 1992). The other theory suggests that the primer is produced by premature termination of transcription at the conserved sequence blocks located between LSP and O₁ (Pham et al 2006). The D-loop contains three conserved sequence blocks (CSBI, CSBII and CSBIII) (Walberg & Clayton 1983) of which, CSBII is a strong transcription terminator element, and has a role in stabilizing the RNA-DNA hybrid (Xu & Clayton 1996, Pham et al 2006 and Wanrooij et al 2012). Wanrooij et al (2010)
suggest that the termination at CSBII is brought about via G-quadruplex structures in DNA and is a prerequisite for primer formation. Furthermore, this structure is thought to regulate mtDNA synthesis by removing the 3’-end of the RNA primer from the DNA template preventing the replication to take place (Wanrooij et al 2012). The authors speculate that the structure formation is controlled or resolved by yet unknown factor for mtDNA synthesis to take place.

Work by Wanrooij et al (2008) and Fuste et al (2010) have shed light on the initiation of replication from LSP. They demonstrated POLRMT depletion to result in decreased replication initiation in vivo and by further in vitro studies suggested that the stem loop structure formed at O₁ when the site is exposed by leading strand synthesis, allows initiation of primer synthesis by POLRMT. Furthermore, it was demonstrated that POLRMT is able to synthesis short RNA primers that can be used by POLG to initiate DNA synthesis. After 25nt, POLRMT is replaced by POLG and lagging strand synthesis is initiated.

Studies on nuclear DNA in mammals, yeast and bacteria have shown that if transcription and replication are not controlled the two machineries can collide and lead to genomic instability (Prado & Aguiler 2005, Mirkin & Mirkin 2005, Hashizume & Shimizu 2007 and Gottipati et al 2008). Not much is known on the mechanism of replication termination in mitochondria, but it is thought that the two replication forks meet at the NCR (Bowmaker et al 2003). Studies of mTERF, mTERF1 and mTERF3 have suggested they have some role in replication termination (Hyvärinen et al 2007 and Hyvärinen et al 2010), but the mechanism remains to be elucidated. Hyvärinen and co-workers showed that mTERF3 binds preferentially O₁/promoter region at the NCR and that its over-expression results in impaired termination of replication (Hyvärinen et al 2010). Based on these findings, they proposed that mTERF3 might act as a contrahelicase to prevent premature replication termination due to collision with the transcription machinery (Hyvärinen et al 2010).

2.3.5 mtDNA repair

DNA can be damaged by endogenously produced molecules such as reactive oxygen species (ROS) produced by the OXPHOS or by exposure to environmental
agents such as ultraviolet (UV) light, ionizing radiation, heavy metals and air pollutants or by chemotherapeutic drugs or the inflammatory system (Lindahl 1993, Iyama & Wilson 2013 and Alexeyev et al 2013). MtDNA is located at the inner mitochondrial membrane where also most ROS is produced, therefore rendering mtDNA especially susceptible to oxidative damage, and research has shown that mtDNA has a higher mutation rate than nuclear DNA (Khrapko et al 1997 and Stuart & Brown 2006). The discovery that mtDNA is packaged by TFAM and exists as nucleoid like structures, indicated that as a first line of defence mtDNA is protected by this structure to some extent (Albring et al 1977, Alam et al 2003 and Kaufman et al 2007). Yoshida et al (2003) showed tumor suppressor p53 to interact with TFAM and differentially regulate its binding to damaged DNA. Later Canugovi et al (2010) demonstrated TFAM to alter the activity of base excision repair (BER), by modulating the access of BER proteins to mtDNA. They show that TFAM interaction with tumor suppressor p53 modifies TFAM-DNA binding and hence promotes BER in mtDNA.

Eukaryotic organisms have evolved five DNA repair mechanism to fight against detrimental nuclear DNA lesions; non-homologous end-joining (NHEJ) and homologous recombination (HR) for double-strand breaks and for single-strand lesions, and base excision repair (BER), nucleotide excision repair (NER) and mismatch repair (MMR) for single strand lesions (Gredilla 2010). The higher mutation rate of mtDNA was originally thought to occur due to lack of mtDNA repair mechanism in mitochondria, because HeLa and KB cells do not remove UV-induced pyrimidine dimers in mtDNA. This was thought to be the evidence for lack of an excision repair mechanism in the mtDNA of these cells (Clayton et al 1974). In addition, it was thought mtDNA has no histone-like proteins to package and protect it (Gredilla 2010). These early findings were not supported by Myers et al (1988) and Pettepher et al (1991) who showed some repair mechanism to be operable and who suggested that both an alkyltransferase mechanism and excision repair was operating within the mitochondrion.

Like discussed in Martin (2011) it is now evident that several repair mechanisms exist in mitochondria including, most prominently the BER (Stierum et al 1999 and Akbari et al 2008). There is also some evidence for mismatch repair (de Souza-Pinto et al 2009), homologous recombination, and non-homologous end-joining (Bacman et al 2009). In addition repair enzymes that were first identified in nucleus have later
been found also in mitochondria, such as FEN1 and DNA2 (Copeland & Longley 2008, Liu et al 2008, Zheng et al 2008, Duxin et al 2009, Gredilla 2010 and Zheng & Shen 2011). Other factors with a role in mtDNA integrity include the elimination of pre-mutagenic dNTPs (Ichikawa et al 2008), and destruction of heavily damaged mtDNA (Shokolenko et al 2009), (for a review see Liu & Demple 2010).

2.4 MtDNA diseases

The first disease caused by mitochondrial dysfunction was described by Luft in 1959 (Luft 1995 for a review), but to date the molecular basis of this disease is not known (DiMauro 2011 and OMIM 2015). In 1965, some scientist hinted the importance of mtDNA in mitochondrial disorders. To quote Gonatas and Shy 1965 - “If mitochondria are self-replicating organelles as recent chemical and morphological evidence has suggested, these two myopathies [pleoconial and megaconal] may be due to a defective gene” —by implication, a mitochondrial gene” (from review of DiMauro Salvatore 2011, Gonatas & Shy 1965). However the importance of mtDNA to human diseases was not documented until 1988, when mtDNA abnormalities were identified as the genetic cause of two specific human syndromes (Holt et al 1988, Wallace et al 1988 from review of DiMauro 2011).

The mitochondrial proteome consists of 1500 proteins, but mtDNA encodes only 13 of those proteins all of which are parts of the OXPHOS system. In addition, mtDNA provides the components needed for mitochondrial protein synthesis. Hence, abnormalities in mtDNA can have detrimental effects on mitochondrial function. All the other proteins of the proteome are encoded by the nuclear genome and include the rest of the OXPHOS proteins and all of the mtDNA maintenance machinery, such as Twinkle, TFAM, POLG and mtSSB (Calvo & Mootha 2010). Therefore, mtDNA diseases can be caused either by primary mutations of mtDNA or by mutations in nuclear genes encoding for proteins involved in mtDNA maintenance or replication machinery, which then lead to secondary damage or loss of the mtDNA-mtDNA multiple deletions and mtDNA depletion respectively (for a review Russell and Turnbull 2014). In addition to primary mitochondrial diseases, mutations in mtDNA have been associated to other neurological and metabolic disorders, including Parkinson’s and Alzheimer’s disease (Lin & Beal, 2006),
cancer (Kroemer 2006), diabetes (Kelley et al 2002 and Patti & Corvera 2010), and heart disease (Karamanlidis et al 2010). Moreover, mtDNA abnormalities may have a pathogenetic role in severe childhood neurological and muscle diseases (Wallace 2010 and Chinnery 2010) and in ageing (Trifunovic et al 2004, Lombard et al 2005 and Raffaello & Rizzuto, 2011).

Because the mitochondria are under the control of two genomes, mitochondrial diseases can be sporadic, maternally inherited or follow Mendelian laws of inheritance (Craigen 2010). Mitochondrial disorders due to primary mutations in mtDNA are highly heterogeneous, partially due to the heteroplasmy and threshold effects. In most circumstances the mtDNAs of an individual share the same sequence (a condition called homoplasmy); however, some people carry two or more mtDNA variants (heteroplasmy). The wild-type mtDNA genomes are able to compensate for the defective genome until a threshold point is reached, after which the deleterious effects of the mutation become manifest. With one exception to date (Sacconi et al. 2008), heteroplasmic mtDNA mutations are usually ‘recessive’: a high mutant load is required before a clinical phenotype is evident (Schon et al 1997, DiMauro & Schon 1998, Wallace 1999 Greaves et al 2012 and Wallace & Chalkia 2013 for reviews) and the threshold depends on the mutation, tissue and individual.

2.4.1 Twinkle helicase and mtDNA maintenance diseases

In 1989, Zeviani and co-workers were the first to describe a mitochondrial disease due to a faulty nuclear gene. The disorder, inherited as an autosomal dominant trait, was characterized by progressive external ophthalmoplegia (adPEO), with multiple mtDNA deletions in the muscle (Zeviani et al 1989, from DiMauro 2011). Typical adPEO phenotype manifests as ophthalmoplegia, ptosis and exercise intolerance. Associated manifestations include cardiomyopathy, peripheral neuropathy, mild ataxia, psychiatric symptoms, dysphagia, parkinsonism and major depression caused by dominant Twinkle mutations (Suomalainen et al 1997, Suomalainen & Kaukonen 2001 and van Hove et al 2009) with an onset of a disease at age 18 – 40 years-old (Suomalainen & Kaukonen 2001). Since the first description, several faulty genes have been linked to adPEO – making the disease genetically heterogeneous. These
include genes encoding factors of the mtDNA maintenance machinery, such as mtDNA helicase Twinkle (Spelbrink et al 2001), mtDNA polymerase POLG (Van Goethem et al 2001) and its accessory subunit (Longley et al 2006), and enzymes involved in nucleotide metabolism (for a review Copeland 2012 and Suomalainen & Isohanni 2010).

Study by Goffart et al (2009) showed that replication stalling or pausing by mutant Twinkle protein is the common consequence of Twinkle PEO mutations that predisposes to multiple deletion formation. As PEO can follow either dominant (Zeviani et al 1989) or recessive inheritance (Bohlega et al 1996) the primary genetic defect was thought to be in nuclear genes that cause mtDNA instability (from a review of Suomalainen & Kaukonen 2001). PEO can also arise from sporadic mtDNA mutations (Suomalainen & Kaukonen 2001). Recessive Twinkle mutations have been shown to cause hepatocerebral mtDNA depletion syndrome (Hakonen et al 2007), infantile-onset spinocerebellar ataxia (IOSCA) (Nikali et al 2005) and renal tubulopathy (Prasad et al 2013) (figure 2.5). These diseases indirectly interfere with mitochondrial function by affecting the mtDNA that then leads to mitochondrial dysfunction (Zeviani et al 1989 and Wallace 2010).

2.5 Mammalian mitochondrial nucleoids

Electron microscopy work by Nass and van Bruggen and co-workers presented the mtDNA molecule as a circular, non-protein bound, structure (Nass 1966 and van Bruggen et al 1966). Albring and co-workers hypothesised that mtDNA stayed attached to the membrane by specific interaction with unknown protein (Albring et al 1977) giving the first indications that mammalian mtDNA may exist as protein-mtDNA complexes called nucleoids, and believed to play a major role in stabilizing the mtDNA. Later studies in *Xenopus laevis* oocytes confirmed that mtDNA was packaged in compact beaded structure that was membrane associated (Barat et al 1985). Mignotte & Barat (1986) characterised the protein component of the “beads” and found a single molecule of 28 kDa that was able to introduce superhelical turns, later identified as TFAM (Fisher & Clayton 1988). The first *in vivo* evidence of the organisation of specific association of protein with mtDNA was done by co-localisation studies in 2001 with the discovery of mitochondrial helicase Twinkle (Spelbrink et al 2001). However it has taken many years for the acceptance that mtDNA is, as DNA in bacteria, organised in mtDNA:protein complexes termed nucleoids (Spelbrink 2010). The packaging of yeast mtDNA by Abf2p was shown in 1991 and the importance of bacterial histone-like protein HU in bacterial DNA packaging in 1993 (Diffley & Stillman 1991 and Megraw & Chae 1993). However, the mammalian mtDNA was considered naked even 10 years later. This misconception was based on studies by Fisher & Clayton who estimated the amount of TFAM per mtDNA to be only 15 molecules (Fisher & Clayton 1988). It was not until Alam et al (2003) demonstrated that the coverage of mtDNA by TFAM was much greater than could be achieved by 15 molecules and therefore that TFAM might package mtDNA.

We can consider TFAM and mtSSB to be the most characterised and acknowledged nucleoid associated proteins (Alam et al 2003, Garrido et al 2003 and Bogenhagen et al 2008). However, to define other bona fide nucleoid associated proteins is more complicated. Discovery of nucleoids has brought about many attempts for search of nucleoid associated proteins and the results vary greatly for reasons that are discussed below. These researchers have found nucleoid associated proteins to include proteins of the transcription and replication apparatus as well as structural and architectural proteins needed for mtDNA maintenance, believed to
regulate stability, replication, transcription and segregation of mtDNA. In addition also bifunctional proteins are present that are thought to have a signalling purpose to facilitate signals between mtDNA and the rest of the cell (Holt et al 2007, Chen & Butow 2005, Spelbrink et al 2010 and Kasashima et al 2014). It is now well accepted that mtDNA exist as a complex with proteins and is inner-membrane bound. There are some suggestions on the proteins that attach nucleoids to the mitochondria inner membrane, which are discussed in the next chapter. Work of this thesis will further elaborate on the nucleoid membrane association mechanism and nucleoid composition.

2.5.1 What do we know about nucleoid membrane connection?

Already in 1969 Nass suggested that mtDNA could be membrane bound (Nass 1969a) and later studies postulated that mtDNA is attached to the inner membrane involving the major non-coding D-loop region (Albring et al 1977 and Boesch et al 2010). However, the means by which mtDNA is held at the inner membrane have stayed an enigma. Suggestions have been made on the role of OPA1-exon4b in nucleoid membrane attachment. Silencing of OPA1-exon4b isoform leads to mtDNA depletion and altered distribution of nucleoids within the mitochondrial network, and it co-immunopurifies with TFAM (Elachouri et al 2011). ATPase family AAA domain-containing protein 3 (ATAD3) and prohibitin are other proteins postulated to have a role in ribosome and nucleoid membrane attachment. Both proteins co-purify with TFAM and mtSSB and are membrane proteins. Gene silencing leads to impaired mitochondrial protein synthesis, suggesting that these proteins organise nucleoids and mitochondrial translational machinery at the inner membrane (He et al 2012a).

2.5.2 Organisation of mammalian mitochondrial nucleoids

In 1991 Satoh & Kuroiwa demonstrated by DAPI staining the `multinuclear’ nature of mitochondria. Their study showed that nucleoids are organised within the mitochondrial network, as spots, ranging from 1 to more than 10 depending on the size of the mitochondrion and the average mtDNA/nucleoid in that study was found
to be 1.4 (Satoh & Kuroiwa 1991). Since then, there have been several investigations to address the issue of the number of nucleoids per cells and the number of mtDNA molecules per nucleoid. While there are reports which suggest a ratio of 2 to 10 mtDNA molecules per nucleoid (for review see Bogenhagen 2012 or Gilkerson et al 2013), more sophisticated imaging techniques have enabled a more detailed investigation on the mtDNA/nucleoid number giving an average of 1.45 mtDNA/nucleoid in fibroblasts (Kukat et al 2011). However, it is also clear that the number of nucleoids and the mtDNA/nucleoid varies between cell lines (Legros et al 2004, Holt et al 2007, Spelbrink 2010).

2.5.2.1 Higher order structure

It is now emerging that nucleoids may not just be assemblies carrying out mtDNA replication, but are more highly organised structures. The first indication of such higher order structures came from trypanosomes where transmembrane structures have been shown to connect mitochondrial nucleoids to the cytoskeleton (Ogbadoyi et al 2003). In yeast Meeusen & Nunnari (2003) provided evidence for a two membrane spanning (TMS) nucleoid structures – where replicating nucleoids co-localise with the mitochondrial outer membrane protein Mmm1. Later Mmm1 has been shown to be an integral endoplasmic reticulum (ER)-membrane protein and Kornmann et al (2009) proposed an ER-mitochondrial encounter structure (ERMES) that is linked to the TMS structure via Mmm1 and the mitochondrial outer membrane proteins Mdm10, 12 and 34. Furthermore, the actin cytoskeleton is thought to associate with this complex, involving the Arp2/3 (involved in actin polymerisation) and Puf 1/3 (RNA binding protein and ARP binding partner) that are mitochondrial outer membrane proteins (Boldogh & Pon 2007 and as discussed in Spelbrink et al 2010).

Spelbrink (2010) suggests that although the conservation of proteins between yeast and mammalian nucleoids is poor, the principle may persist so that mammalian nucleoids also have higher order structures similar to yeast (figure 2.6). Iborra et al (2004) postulated mammalian nucleoids to have transmembrane linkage as they co-localised with kinesin motor (KIF5). Also KIF5 interacts with the outer membrane protein Miro (Macaskill et al 2009 and as discussed in review of
Spelbrink 2010) further suggesting an existence of a higher order structure for nucleoids in mammals. There is also some evidence for ER – mitochondrial connections in mammalians that is mediated by the outer membrane protein mitofusin 2 (DeBrito & Scorrano 2008).

Figure 2.6: Is there conservation of function between budding yeast and mammalian nucleoid organization and dynamics? This figure illustrates recent advances in our understanding of nucleoid organization in budding yeast and mammals (Reprinted by permission from John Wiley & Sons, Inc, IUBMB Life, Spelbrink 2010).

Iborra and co-workers have suggested that mtDNA is arranged in close proximity to mitochondrial and cytoplasmic translation machineries either side of the mitochondrial membranes, with protein import complexes feeding the products of the cytosolic ribosomes directly into the mitochondria (Iborra et al 2004). This idea pointed towards a more complicated nucleoid structure where the translational and RNA processing machinery are part of the nucleoid structure. Later studies have provided further evidence for the close relationship between nucleoids and ribosomes. Work by He and colleagues (2012b) show that chromosome 4 open reading frame 14 (C4orf14 or Noa1) has a role in mitochondrial 28S ribosome assembly and that it interacts with the nucleoid complex. This lead to a hypothesis that the initial ribosome assembly takes place at nucleoids. The work was further supported by a study where Dalla-Rosa et al (2014) set out to investigate MPV17L2, which had unknown function in mitochondria. Their work demonstrated the protein to localise in nucleoids and that in absence of MPV17L2, proteins of the small
subunit of the mitochondrial ribosome are trapped in enlarged nucleoids, in contrast to a component of the large subunit (Dalla Rosa et al 2014). The close liaison between nucleoid and ribosomes was further corroborated by mass spectrometry research by Bogenhagen et al (2014) where they provide evidence that the initial RNA processing and ribosome assembly takes place in the vicinity of nucleoids. Nucleoid associated proteins may also include proteins involved in metabolic activities that operate a signalling role between mtDNA and the rest of the cell. This has been shown in yeast (Kucej et al 2008) where amino acid starvation leads to the recruitment to the nucleoid of Ilv5, an enzyme involved in branched-chain amino acid synthesis (Kucej et al 2008), and related enzymes have been found associated with the mammalian mitochondrial nucleoid (Wang & Bogenhagen 2006 and Bogenhagen et al 2008).

2.5.3 Diversity of mitochondrial mtDNA and nucleoids

Although mtDNA is small and has restricted genetic information, it is not confined to similar form in all eukaryotes. For example, the size of mtDNA varies between species ranging from 6kb to 200 kb (Burger et al 2003). Trypanosome mtDNA can also exist in a mix of two different sized circular DNA molecules (Lukes et al 2002), while in animals the mtDNA molecules are all the same size (Boore 1999). In contrast, the mtDNA in fungi is mostly found in linear form (Williamson 2002 and Malka et al 2006 for review). In both yeast and mammals, the mitochondrial nucleoids are organized in regular pattern within the mitochondrial network that contains several nucleoids (Miyakawa et al 1987 and Satoh & Kuroiwa 1991). In some organisms e.g. in flagellates, mtDNA is organized in a single structure within the mitochondria, referred to as kinetoplast (Lukes et al 2002 and Malka et al 2006 for review). The information above is based on the review of Malka et al (2006).

When Nass and Nass discovered mitochondrial DNA they described them to resemble the bacterial nucleosomes when studied by electron microscopy (Nass & Nass 1963). Mammalian mitochondria have also other features that resemble bacteria, which is not surprising knowing their evolutionary origin (see introduction). Bogenhagen and co-workers discuss human mitochondrial nucleoids to resemble those of bacteria in many ways. For example, the packing density of
human and bacterial nucleoids is thought to resemble each other (Bogenhagen et al 2008). In addition, both human and bacterial nucleoids are membrane anchored (Nass 1969a and Bogenhagen et al 2008).

2.5.4 Composition of mammalian mitochondrial nucleoids

Previous nucleoid research has clearly pointed towards nucleoids being complex structures that have more functions than only being replication machineries. Bogenhagen et al (2008) discuss the nucleoid structure to be layered and He et al (2012b) point towards an intimate relationship between nucleoids and the protein synthesis machinery, as also previously suggested by Iborra on the basis of fluorescent microscopy analysis (Iborra et al 2004). Further, in a recent paper Bogenhagen and co-workers presents evidence that initial RNA processing and ribosome assembly takes place in the close vicinity of nucleoids (Bogenhagen et al 2014), whereas others have suggested that the entire small subunit of the mitochondrial ribosome is assembled at the nucleoid (He et al 2012b). Hence, it makes the concept of nucleoid proteins difficult. Over the last 10 years, various sets of nucleoid associated proteins were identified, but for various reasons very, few proteins are shared between all these sets (Bogenhagen et al 2003, Wang & Bogenhagen 2006, He et al 2007, Bogenhagen et al 2008 and He et al 2012a). The complete list of proteins found in many different nucleoid purification strategies is vast and cannot be repeated here. In addition, only few have been further characterized and confirmed to be truly nucleoid associated as many found proteins would benefit greatly on further characterisation and establishment of the in vivo localisation of the protein. For example, Wang & Bogenhagen (2006) showed some further characterisation of the protein identified in nucleoid isolation, DEAH-box helicase (DHX30), but failed to convincingly show by other means that it is nucleoid associated. In addition, they did not show DHX30 to have a role in mtDNA maintenance.

Below I discuss some proteins that have been more than once isolated in nucleoid purification research and have been more extensively studied. Many more proteins have been identified, and this highlights the difficulty of identifying the consensus
list of nucleoid proteins. In later section, the challenges encountered during nucleoid research are discussed.

2.5.4.1 TFAM and mtSSB

TFAM and mtSSB were the first nucleoid proteins to be discovered and were already found in the 1980’s in characterisation of mtDNA-protein complexes (van Tuyle & Pavco 1985 and Mignotte & Barat 1986). Both TFAM and mtSSB were later shown to co-localise with mtDNA as punctate foci in immunofluorescence studies (Garrido et al 2003). Bogenhagen and co-workers (2003) also identified TFAM and mtSSB in their nucleoid preparations from frog oocyte mitochondria. Kanki et al (2004) have postulated TFAM to have an architectural role in nucleoids where it was thought to maintain the higher order structure of mtDNA. Later work by Kaufman et al (2007) has suggested a specific role of TFAM in nucleoid assembly. Interaction of mtDNA with nucleoid proteins TFAM and mtSSB is also important, as in the absence of mtDNA, nucleoid integrity is lost (Garrido et al 2003).

2.5.4.2 Twinkle

The discovery and microscopic studies of Twinkle by Spelbrink et al (2001) found Twinkle to co-localise with mtDNA and gave the first hint that it may be a component of mitochondrial nucleoids. Work in the Spelbrink laboratory has also shown TFAM and mtSSB to co-localise with Twinkle in immunofluorescence studies, further suggesting the role of Twinkle as a nucleoid protein (Garrido et al 2003). Above studies were carried out using over-expressed Twinkle protein, but later mass spectrometry based nucleoid research has also identified endogenous Twinkle to purify with nucleoids (Wang & Bogenhagen 2006 and Bogenhagen et al 2008).
2.5.4.3 **POLG1 and POLG2**

Interestingly POLG1 and POLG2 did not always co-localise with nucleoid structures even though it was found in nucleoid purification preps using Western blotting detection. It was suggested that POLG may only co-localise with actively replicating nucleoids (Garrido et al 2003). POLG has also been found in later studies to co-purify with nucleoids in human cultured cells using a mass spectrometry approach for protein identification (Wang & Bogenhagen 2006 and Bogenhagen et al 2008). Gene silencing of POLG accessory subunit POLG2 leads to increased nucleoid number, altered nucleoid structure and a reduction in 7S DNA. It was shown to exclusively bind to the D-loop region leading to hypothesis that it is important in nucleoid organisation and key regulator of nucleoid mtDNA copy number (Di Re et al 2009).

2.5.4.4 **ATAD3 and prohibitin**

Prohibitin (PHB) proteins PHB1 and PHB2 are expressed in multiple cellular compartments, but primarily localise to mitochondria, where they are attached to the inner mitochondrial membrane as oligomeric rings. PHBs are involved in cell proliferation and cristae morphology (Merkwirth & Langer 2009), cell signalling (Mishra et al 2010), apoptosis (Green & Reed 1998 and Theiss & Sitamaran 2011) and have chaperone functions for newly synthesised mitochondrial proteins (Nijtmans et al 2000). PHB1 has co-purified with nucleoid proteins at least in three separate nucleoid isolation experiments (Bogenhagen et al 2003, Wang & Bogenhagen 2006 and He et al 2012a). After the discovery of PHB1 in nucleoid purification experiments Kasashima et al (2008) conducted PHB siRNA experiment that demonstrated PHB1 to decrease mtDNA staining by EtBr and Picogreen. In addition, they showed PHB1 to have some role in mtDNA organisation because PHB1 silencing shifts a significant amount of mtDNA to the soluble fraction during membrane fractionation protocol. Kasashima and co-workers (2008) also conclude that the reduction of TFAM during PHB1 gene silencing shows that PHB1 regulates mtDNA copy number via TFAM.

ATAD3 has been found in nucleoid purification experiments and shown to affect mtDNA supercoiling and to bind to mtDNA D-loop section (He et al 2007,
Bogenhagen et al 2008). ATAD3 silencing leads to decrease in number of nucleoids that also appear larger in contrast to those in control cells. Based on this they hypothesised that ATAD3 may be involved in nucleoid division (He et al 2007 and Holt et al 2007). There has been some debate on whether or not ATAD3 is a nucleoid protein. Bogenhagen et al (2008) argued that ATAD3 is not part of the nucleoid complex as it failed to co-localise with nucleoid components in immunofluorescence studies and did not cross-link to mtDNA in nucleoid purification experiments. In contrast, ATAD3 was found in their earlier nucleoid purification strategy (Wang & Bogenhagen 2006). They argued that cross-linking would identify proteins closely attached to mtDNA and native nucleoid preparations may include contaminant proteins due to less stringent protocol for purification. Moreover, they argued that the N’terminal domain reported to bind D-loop by He et al (2007) is not located in matrix and this was later verified by Hubstenberger et al (2010). In support of He et al (2007) work, they did mention that ATAD3 may only transiently attach to nucleoids, as it was not seen to co-localise 100% with nucleoids in immunofluorescence. Bogenhagen et al (2008) also discusses the possibility that ATAD3 interaction to nucleoids to be indirect. Recent study has found ATAD3 gene silencing to affect mitochondrial translation and together with PHB1 to co-purify with mitochondrial ribosomes, suggesting that these proteins might organise nucleoids and mitochondrial translational machinery at the inner membrane (He et al 2012a).

2.5.4.5 Lon protease, M19 and PDIP38

Lon protease has been isolated in nucleoid purification experiment in formaldehyde cross linked sample (Cheng et al 2005) and co-immunoprecipitates with Twinkle and POLG (Liu et al 2004). Several investigations suggest that Lon protease has an important role in mtDNA maintenance. It has been shown to interact with mtDNA in the D-loop region (Lu et al 2007) and shown to degrade TFAM (Matsushima et al 2010). Research has shown that Lon protease over expression reduced TFAM levels and mtDNA copy number and showed the opposite to occur in Lon protease knockdown, suggesting that Lon protease modulates mtDNA biogenesis by the selective degradation of TFAM (Matsushima et al 2010). Recent finding shows that
phosphorylation of TFAM leads to DNA dissociation and selective degradation by Lon protease, hence providing a mechanism for regulating mtDNA binding and release. The authors suggest that this mechanism is essential for mtDNA maintenance and gene expression (Lu et al 2013).

Sumitani et al (2009) have also proposed a novel mitochondrial protein M19 to be nucleoid associated as M19 immunoprecipitation led to mtSSB co-purification. M19 also showed punctate labelling in microscopy studies. The authors also showed a link between mtDNA content and M19 concentration speculating the role of M19 in mtDNA organisation and metabolism (Sumitani et al 2009). To elucidate the cellular role of this newly identified protein Cambier and co-workers studied the role of M19 in muscle cells and in pancreatic β-cells. They demonstrated M19 to modulate mitochondrial oxygen consumption and ATP production that led to regulation of major cellular processes such as myogenesis and insulin secretion (Cambier et al 2012).

Polymerase delta interacting protein 38 (PDIP38) was originally identified to bind DNA polymerase delta in the nucleus (Liu et al 2003) and has been shown to have a role in pre-mRNA processing (Wong et al 2013). Cheng and co-workers (2005) discovered PDIP38 to also localise to mitochondria. Furthermore, they demonstrated that PDIP38 associated with mtSSB and proposed that it may be part of the nucleoid structure. However, the interaction with TFAM was not convincingly shown, it did not possess DNA binding properties and no function of PDIP38 in mtDNA maintenance was demonstrated. Nor was co-localisation of the protein with mtDNA or known nucleoid proteins shown. Nevertheless, it has been found in subsequent nucleoid purification experiment (Bogenhagen et al 2008).

2.5.5 Mass spectrometry based nucleoid identification strategies

Uncovering of protein-protein interaction is vital for the understanding of how biological systems function. The current controversies over mtDNA replication could benefit from the identification of proteins that associate with e.g. Twinkle or POLG, which are both involved in replication as well as part of the nucleoid complex (Spelbrink et al 2001, Garrido et al 2003 and Korhonen et al 2004). One approach that has been widely explored in nucleoid purification work, first in yeast
and then in mammals, is mass spectrometry (MS) based protein identification following biochemical isolation such as immunoprecipitation (Kaufman et al 2000, Wang & Bogenhagen 2006, Bogenhagen et al 2008, He et al 2012a). This approach benefits from the fact that protein-protein interactions are investigated in their native environment and hence gives information on the physical relevance of such interaction. Chemical cross-linking has also been used in protein-protein interaction studies to prevent dissociation during sample preparation and hence allowing purification of loosely attached proteins and enables harsher purification strategies to eliminate contaminants (Ethier et al 2006 and Sutherland et al 2008). One such cross-linker is formaldehyde that can produce protein-DNA, protein-RNA and protein-protein cross-links (Möller et al 1977 and Jackson 1978) and has many beneficial features (as discussed in Ethier et al 2006).

1) Proteins must be in close proximity as cross-links are produced at very short distance
2) Formaldehyde enters the cell rapidly and is non-specific
3) Allows snap-shot of interaction at the time of addition (interaction with formaldehyde inactivates enzymes immediately after addition to cells (Hall & Struhl 2002)
4) After cross-linking non-physiological conditions can be applied without interference in structural integrity (Vasilescu et al 2004)
5) Cross-links are reversible, hence allowing analysis by mass-spectrometry (Vasilescu et al 2004)

2.5.5.1 Challenges of nucleoid purification strategies

Mammalian mitochondrial nucleoid identification research thus far has rarely employed quantitative proteomics and is based more on identification of proteins from one or two purification sets with further investigation of individual proteins. Further, the research has been complicated by different approaches to nucleoid purification e.g. by aiming the isolation to either protein of interest or to mtDNA, by using different MS-approaches, purification strategies or starting material (see table 2.1). All these factors are adding to the complexity of finding the comprehensive list of mammalian mitochondrial nucleoid proteins.
Table 2.1: MS-based purification strategies used in mammalian nucleoid research (IP refers to immunoprecipitation and HU to bacterial histone like protein).

<table>
<thead>
<tr>
<th>Source of mitochondria</th>
<th>Nucleoid purification</th>
<th>Mass Spectrometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bogenhagen et al 2003</td>
<td>Xenopus laevis oocytes</td>
<td>Glycerol and metrizamide gradient</td>
</tr>
<tr>
<td>Cheng et al 2005</td>
<td>Jurkat cells</td>
<td>PDIP38 – IP</td>
</tr>
<tr>
<td>Wang &amp; Bogenhagen 2006</td>
<td>HeLa cells</td>
<td>Glycerol gradient combined with TFAM and mtSSB – IP</td>
</tr>
<tr>
<td>He et al 2007</td>
<td>HEK293 TREG cells</td>
<td>HU-coated beads used in IP</td>
</tr>
<tr>
<td>Bogenhagen et al 2008</td>
<td>HeLa cells</td>
<td>Formaldehyde cross linking Glycerol/Nycodenz gradient</td>
</tr>
<tr>
<td>Reyes et al 2011</td>
<td>Rat liver</td>
<td>High salt conditions combined with centrifugation</td>
</tr>
<tr>
<td>He et al 2012a</td>
<td>HEK293 TREG cells</td>
<td>IP against, TFAM, mtSSB or ATAD3</td>
</tr>
</tbody>
</table>

To add to the complexity the nature of the nucleoid is not static as shown by Duxin et al (2009) but proteins may transiently localise to nucleoids. In this study, DNA2 only partially co-localises with wild-type Twinkle, but in cells, expressing mutant variants of Twinkle the co-localisation together with mtDNA was increased to 100%. Therefore, it is impossible to conclude from the data collected this far what is the composition of mammalian nucleoids. Besides, only few of the proteins found have been further characterised and shown to have any role in nucleoid maintenance. Therefore, there would be a need for more systematic approach for nucleoid protein identification before we could come up with a comprehensive list of true nucleoid proteins. In my thesis work, I have investigated possibilities to overcome these problems.
One of the problems with MS-based analysis has been the difficulty of detecting low abundant proteins. For example, Twinkle and POLG are well-characterised nucleoid proteins, but not often identified in MS-analysis, which is especially true for the single band analysis technique (table 2.2). From research presented in table 2.1 Twinkle and POLG are only recognised in work by Wang & Bogenhagen (2006), but only when nucleoids were isolated via TFAM-immunoprecipitation. They also carried out immunoprecipitation based nucleoid isolation using mtSSB, but failed to detect Twinkle or POLG using this bait. The only other research that has successfully detected these replication factors used formaldehyde cross linking and isolated the whole nucleoid complexes (Bogenhagen et al 2008).

Table 2.2 Replication factors identified in different nucleoid purification strategies (Nucleoid = nucleoprotein complexes were isolated, IP = immunoprecipitation, xl = formaldehyde cross linked).

<table>
<thead>
<tr>
<th></th>
<th>Twinkle</th>
<th>POLγ</th>
<th>mtSSB</th>
<th>TFAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single band</td>
<td>Bogenhagen et al 2003 nucleoid</td>
<td></td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Cheng et al 2005 PDIP38-IP</td>
<td></td>
<td>yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>He et al 2007  nucleoid</td>
<td></td>
<td></td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Reyes et al 2011 nucleoid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shotgun</td>
<td>Wang &amp; Bogenhagen 2006 mtSSB-IP</td>
<td></td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TFAM-IP yes</td>
<td>yes</td>
</tr>
<tr>
<td>Bogenhagen et al 2008 nucleoid xl</td>
<td></td>
<td>yes</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>He et al 2012  mtSSB-IP</td>
<td></td>
<td></td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ATAD3-IP</td>
<td>yes</td>
</tr>
</tbody>
</table>
3. Aims of the research

The main purpose of this work was to further characterise the organisation and composition of mammalian mitochondrial nucleoids. Furthermore, the aim of the research was to investigate the structure of mitochondrial helicase Twinkle and study the role of Twinkle in nucleoid organisation. More specifically the aims of the work were as follows:

1. Investigate the known mammalian mitochondrial nucleoid proteins further and elucidate the role of Twinkle in nucleoid membrane connection

2. By developing a more systematic and quantitative approach for nucleoid associated protein isolation and identification by mass spectrometry, to study the composition of mammalian mitochondrial nucleoids

3. Study the structure of mitochondrial helicase Twinkle to gain further knowledge of the mechanistic function of the protein. This would allow us to understand for example the malfunctioning of Twinkle disease mutants
4. Materials and methods

4.1 Cell culture systems

Stable cell lines expressing various mtDNA maintenance proteins upon induction were created as described in (Wanrooij et al., 2007) using the Flp-In™ T-Rex™ 293 host cell line (Invitrogen), a HEK293 variant containing a Flip recombination site at a transcriptionally active locus. The ATAD3-HA expressing cell line was a kind gift of Drs. Ian Holt & Hiroshi Sembongi (Cambridge UK). The transgenic cells were grown in DMEM (Lonza) supplemented with 10% FCS (PAA laboratories), 2 mM l-glutamine, 1 mM Na pyruvate, 50 µg/ml uridine (Sigma), 100 µg/ml Hygromycin and 15 µg/ml Blasticidin (both from Invivogen) in a 37°C incubator at 8.5% CO2. Normal HEK293 cells, U2OS, 143B, 206F (ρº), A549 and B2ρº were grown under similar conditions but without antibiotics. BJ fibroblasts were grown in 4:1 DMEM (Lonza) and M199 (Sigma) containing 15% FCS, 2 mM l-glutamine, 1 mM Na pyruvate. All cell lines were frequently checked for mycoplasma infection and found to be negative. HEK293-Rho0 cells were a kind gift from Prof Ian Holt.

4.2 Transfections, confocal microscopy, Edu and BrdU labeling

For transient transfection of the Twinkle-Myc expression construct (Spelbrink et al 2001) TransIT-LT1 (Mirus, Madison, WI) was used according the manufacturer’s instructions. For immunofluorescent detection cells were grown on coverslips in 6 well plates. Cells were fixed using 3.3% paraformaldehyde (PFA) in cell culture medium for 25 min. This was followed by three washes in PBS and lysis for 15 min with 0.5% Triton X100 in PBS/10% fetal calf serum (FCS). Primary and secondary antibodies were incubated at the following concentrations in PBS/10%FCS for 1 hr-o/n: TFAM rabbit polyclonal, 1:500; Twinkle mouse monoclonal (IgG) or rabbit
polyclonal (Abcam, ab83207) 1:100; anti-DNA mouse monoclonal IgM AC-30-10 (PROGEN), 1:250; mtSSB rabbit polyclonal (kind gift of Drs. M. Zeviani and V. Tiranti), 1:100; POLG1 goat polyclonal, 1:50 (Santa Cruz, sc-5931); myc- rabbit polyclonal (Abcam, ab9106). Secondary antibodies goat-anti-rabbit, goat-anti-mouse IgG or IgM and chicken-anti-goat were AlexaFluor 488, 568, and 647 (Invitrogen) labeled and used in various combinations at a 1:1000 dilution. Slides were mounted using ProLong® Gold antifade with DAPI (Invitrogen). Image acquisition using confocal microscopy was carried out using a Yokogawa spinning disk confocal on a Nikon TiE inverted microscope and equipped with 4 solid-state lasers (405, 488, 561 and 647 nm), AOTF controlled excitation, appropriate narrow band-width emission filters and an Andor iXon DV885 EMCCD camera, all under control of Andor iQ software (Andor, Belfast N. Ireland; see or alternatively using an Olympus FV1000 confocal system of the Microscopic Imaging Centre at the Nijmegen Centre for Molecular Life Sciences (NCMLS; Images were further processed using Photoshop CS2 to adjust brightness/contrast and size).

Mitochondrial DNA labeling using the Click-iTTM EdU (5-ethynyl-2'-deoxyuridine) imaging kits with either AlexaFluor 488 or 568 azide (Invitrogen) was initially done essentially as described by the manufacturer except that we typically used 50 µM EdU in our cell labeling experiments to detect mtDNA label incorporation. For the experiment shown in thesis supplementary figure 7 examining EdU, Twinkle, mtSSB co-localisation we modified the procedure as follows: the Click-iTTM buffer additive was replaced by 50 mM ascorbic acid (final concentration) and the labeling reaction was done twice for 25 min with a freshly prepared labeling mix. Following EdU labeling and detection chemistry we proceeded with antibody incubations as above. MtDNA/Twinkle or EdU/Twinkle positive foci were scored manually by first marking all mtDNA or EdU foci and subsequently overlaying the marks with the Twinkle immunofluorescence. In both cases, experiments were repeated 3 times and in each experiment 10 cells (mtDNA/Twinkle) or 10-20 cells (EdU/Twinkle) were scored to obtain final numbers.

BrdU labeling used the BrdU Labeling and Detection Kit I (Roche) using the manufacturers’ protocol except that we used 50 µM BrdU and the Alexa 568 anti-mouse antibody for BrdU-antibody detection. Fixation using acid-ethanol was also done according to the manufacturer’s protocol but resulted in a significantly reduced
mtSSB antibody staining compared to paraformaldehyde fixation. Both EdU and BrdU labeling, mtDNA/EdU (or mtDNA/BrdU) and Edu/Twinkle (or BrdU/mtSSB) positive foci were scored manually by first marking all mtDNA/Edu or BrdU foci (using Image Pro Plus 6 'create point feature' (Media Cybernetics) or using the 'Event marker' tool using Axiovision 4.8 software, and subsequently overlaying the Twinkle/mtSSB immunofluorescence and counting all double positive and using both numbers to calculate relative percentages. In all cases, experiments were repeated several times as indicated and in each experiment multiple cells were scored to obtain final numbers. Very rare cytoplasmic EdU or BrdU spots that did not appear to co-localise with mtDNA were not considered. MtSSB/BrdU positive foci were only judged positive with clear position overlap and a distinct focal mtSSB signal on the basis of the strong focal mtSSB presence in a subpopulation of mtDNA foci in paraformaldehyde fixed cells. BrdU or EdU foci in the vicinity of the nucleus could often not be unambiguously assigned positive for mtDNA and/or BrdU/Edu on the basis of the often strong nuclear DNA signal and were therefore not used in the quantification.

4.3 siRNA knockdown and ddC treatment

MtDNA depletion in U2OS cells was achieved by a 48-72 hr treatment with 100 µM 2’,3’-dideoxycytidine (ddC). For Twinkle knockdown, U2OS cells and fibroblasts were transfected in 6 well plates with a mixture of three Stealth™ siRNA duplex oligonucleotides (C10Orf2 HSS125596, HSS125597, HSS125598, Invitrogen) against Twinkle, at a concentration of 20 pmol each, using Lipofectamine™ 2000 (Invitrogen) according the manufacturer’s protocol. As a negative control we used Stealth™ Universal negative controls. Cells were fixed and analysed 36-72 hrs post-transfection.

4.4 Western blot analysis

Mitochondrial fractions were analysed for protein expression by immunoblotting after SDS–PAGE (Spelbrink et al 2000). Primary monoclonal c-myc (Roche
Molecular Biochemicals), FLAG (Sigma Aldrich) and HA (BAbCO) antibodies were used for detection of recombinant proteins. Other antibody dilutions were as follows: mtSSB rabbit polyclonal (Sigma) 1:1000; TFAM rabbit polyclonal antibody (kind gift of Dr. R. Wiesner), 1:10000; Twinkle mouse monoclonal (kind gift of Dr. Anu Wartiovaara-Suomalainen) 1:1000; COXII mouse monoclonal (Invitrogen), 1:10000; GDH rabbit polyclonal (kind gift of Prof. RN Lightowlers), 1:2000; POLG1 (Santa Cruz, sc-5931) goat polyclonal 1:1000; TOM20 mouse monoclonal (Santa Cruz, sc-17764), 1:500; FLAG monoclonal (Sigma), 1:4000; and anti-GFP mouse monoclonal (Zymed, 33-2600) was used at 1:10000. Peroxidase-coupled secondary antibody horse-anti-mouse or goat-anti-rabbit was obtained from Vector Laboratories. Donkey-anti-goat was obtained from Santa Cruz. Enhanced Chemiluminescence detection was done essentially as described (Spelbrink et al. 2000) or the Pierce Super Femto Western blot detection kit was used according to the manufacturer’s instructions.

4.5 Isolation of mitochondria and mitochondrial treatments

Cells induced for 24-72 hrs with various concentrations of doxycycline (Sigma) (as indicated) were collected and resuspended to hypotonic homogenization buffer (4 mM Tris-HCl, pH 7.8, 2.5 mM NaCl, 0.5 mM MgCl2 and proteinase inhibitor mixture complete, Roche Molecular Biochemicals) and subjected to homogenisation using a 5ml chilled Dounce homogeniser until 80% of the cells were broken. Alternatively cells were disrupted after short cytochalasin treatment (Yasukawa et al. 2005). With both methods, mitochondria were isolated using differential centrifugation and on occasion further purified using sucrose gradient purification as described (Spelbrink et al. 2000).

4.5.1 Mitochondrial (sub)fractionation

The mitochondrial outer membrane was disrupted by incubation with a digitonin (Sigma Aldrich)/protein ratio ([µg dig] / [µg mitochondria]) = 0.2 (unless otherwise indicated) either in PBS or a buffer containing 225mM Mannitol, 75mM sucrose,
10mM HEPES, pH 7.8, 10mM EDTA, in either case supplemented with a protease inhibitor cocktail (Roche). The mitoplasts were obtained following centrifugation at 8000 xg for 10 min, +4°C. The supernatant was subsequently centrifuged at 100 000 xg for 1 hr: the supernatant containing protein fraction is the intermembrane space (IMSP) fraction, while the pellet contains a fraction of outer mitochondrial membrane proteins. Mitoplasts were suspended in 0.16 mg Brij58 per mg mitoplasts and incubated for 10 min on ice. Membrane (inner + outer) (pellet) and matrix (supernatant) fractions were obtained after centrifugation at 100 000 xg for 1 hour (figure 4.1).

For digitonin based fractionation, crude mitochondria from HEK293e or inducible HEK293 Flp-In™ T-Rex™ wt-Twinkle cells were taken up in 1X PBS (Gibco), the total protein concentration was determined with Bradford assays and lysed by addition of digitonin (Sigma Aldrich) at indicated ratios µg digitonin/µg total mt protein, incubated for 10 min on ice and centrifuged for 5 min at 14 000 xg and 4°C, solubilized supernatant fractions were transferred and insoluble pellet fractions were resuspended in volumes equal to the removed soluble fractions.

4.5.2 Mitochondrial carbonate or KCl extraction

For carbonate extraction, isolated mitochondria were resuspended in a 0.1 M Na₂CO₃ buffer (pH 11.0) and incubated on ice for 30 minutes; the pellet was then recovered by centrifugation (100 000 xg, 60 minutes, 4°C). For salt-wash experiments, mitochondria were diluted tenfold in buffers consisting of either 30
mM KCl or 500 mM KCl in 30 mM Tris-HCl (pH 7.4) and sonicated at 40% power 3mm probe 3 times for 10 seconds per cycle. The pellet was recovered by centrifugation (100 000 xg, 60 minutes, 4°C). Proteins from the resulting supernatants were concentrated by DOC/TCA precipitation as follows: lysates were first treated with 0.02% DOC for 30 min on ice before addition of 10% TCA and incubated at +4°C over night. To collect the precipitate samples were centrifuged at 15 000 xg for 15 min at +4°C.

4.5.3 Treatment of isolated mitochondria by sonication and nucleases

For nuclease treatment mitochondria were resuspended in enzyme-buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl, 3mM CaCl$_2$, 2 mM MgCl$_2$) and sonicated on ice at 40% power for 3 times 20 s before addition of the enzymes as indicated (DNase I (Fermentas) 10U, RNAse A (Fermentas (20µg), Micrococcal nuclease (Fermentas), 50U and Benzonase nuclease (Sigma) 50U), and incubated at +37°C for 30 min. Where appropriate, lysates were further subjected to carbonate extraction as described above.

4.6 Dot-plot analysis of mtDNA content in digitonin fractions

For mtDNA analyses, defined dilutions were taken up in 2xSSC (f.c.) and dot-blotted in triplicate onto positively charged nylon membranes using a manifold device (Biorad). Dot-blots were detected using non-radioactively labeled cyt$b$ probes using a Dig-labeling system (Roche). Hybridisations (at 48°C) and subsequent dig-antibody incubations were carried out using Easy-Hyb (Roche) according the manufacturer’s protocol. ECL detection was performed with CSPD (Roche) and visualized with a ChemiDoc (Biorad). Quantifications of resulting ECL-signals were performed with ImageQuant (Ge Healthcare).
4.7 Floatation assay

Mitochondrial protein yield was determined by Bradford assay and the equivalent of 2mg of total mitochondrial protein was lysed in TN (25mM Tris-HCl, pH 7.5, 150mM NaCl, 1mM DTT, cocktail of protease inhibitors, 10% sucrose) containing either 1% Triton X-100 or digitonin at a ratio of 1.5:1 (w/w) for 30 min on ice. Digitonin lysed samples were centrifuged for 10min and the pellet resuspended in TN containing 1% Triton X-100. Samples were mixed with cold Optiprep™ to a final concentration of 42.5%, transferred into ultracentrifuge tubes and overlaid with 400µl of each 40, 37.5, 35, 32.5, 30, 27.5, 25, 20 and 0% Optiprep™. The gradients were centrifuged at 1000 000g for 14h at 4°C. Fractions were collected from top to bottom and aliquots analysed by Western blotting and dot blotting as described above.

4.8 Formaldehyde cross-linking and immunoprecipitation

Twinkle expression was induced by addition of 3ng/ml doxycycline (Sigma) for 36 hours. From previous experiments we know that this expression level and time is appropriate to preserve nucleoid structures. For cross-linking typically cells from five 145 mm (cross-section) cell culture dishes were harvested and cell number was adjusted to 10×10^6 cells/ml. Cross-linking was carried out in 1% formaldehyde (Sigma) for 10 min at RT with rotation. The reaction was stopped by addition of 125 mM glycine, pH 8.0. Formaldehyde is toxic and was handled in a fume hood. Sample handling after addition of formaldehyde similarly was carried out in a fume hood and formaldehyde disposed appropriately. Cells were transferred on ice and all subsequent centrifugations carried out at +4 °C. Cells were washed four times with ice cold TBS (50 mM Tris-HCl pH 7.4, 150 mM NaCl) and processed further by two different methods. Method A, Triton X-100 method: Cells were lysed in Buffer A (50mM Tris-HCl pH 7.4, 300 mM NaCl, 2mM EDTA 1% Triton X-100). In method B, the X-ChiP method, cells were lysed with RIPA buffer (50mM Tris-HCl pH 8, 150mM NaCl, 1% NP-40 (Igepal), 0.5% sodium deoxycholate, 0.1% SDS). In both methods lysates were sonicated for 1 min at 40 % power (1s on 2s off cooling
on ice), but only with the X-ChiP method sonication was followed by addition of 100µg/ml RNAse A (Sigma), 5U/ml DNAse I (Thermo Scientific) and 50U/ml Benzonase nuclease (Sigma), 2.5mM Mg2+, 1mM CaCl2 and incubated at +37 °C for 30min. With both methods lysates were centrifuged for 10 min at 1200g at +4 °C and the protein content of the lysates was equalised to 2mg/ml in a total volume of 10 ml before addition of 180 µl of FLAG resin (Sigma) and rotation for 2 hours at +4 °C. In method A, FLAG resin was washed once in buffer B, C and D. Buffer B: 50mM Tris-HCl pH 7.4, 800mM NaCl, 0.1% Triton-X 100, Buffer C: 50mM Tris-HCl pH 7.4, 50 mM NaCl, Buffer D: 50mM Tris-HCl pH 7.4, 150 mM NaCl, 0.1% Triton-X100. Nucleoids were eluted with 100 µl 3xFLAG peptide (at 0.25 mg/ml) in 50mM Tris-HCl pH 7.4, 150 mM NaCl. In method B, the FLAG resin was washed three times in RIPA buffer and nucleoids eluted with 100 µl 3xFLAG peptide (at 0.25 mg/ml) in RIPA buffer. All buffers included 1×complete EDTA-free Protease inhibitors (Roche).

4.9 Mass spectrometry sample preparation

Protein samples were incubated with SDS-PAGE sample-buffer for 30 min at 95°C to reverse FA cross-links and fractionated by SDS-PAGE on Any kD™ Mini-PROTEAN® TGX™ Gels (BIO-RAD). Lanes were cut in three equal-sized (approximately 1x2.5 cm) gel slices. No gel-staining was applied following electrophoresis. Each gel slice was subjected to in-gel tryptic digestion and further processed according to standard methods. In short, gel slices were cut into small pieces (~1mm²) and were washed successively at least three times with 50 mM ammonium bicarbonate (ABC) and 100% acetonitrile (ACN). Gel slices were swelled in 10 mM dithiothreitol and incubated for 20 minutes at 56°C to reduce protein disulfide bonds. To remove the reduction buffer, gel pieces were shrunk with ACN. Alkylation of the reduced cysteines was performed by incubation of 50mM chloroacetamide in ABC for 20 minutes at room temperature in the dark. Gel pieces were again washed twice with ACN and ABC before tryptic digestion at 37 °C overnight with 1.25ng/µl sequencing grade modified Trypsin (Promega) in ABC. To recover tryptic peptides from the gel pieces, they were first diluted 1:1 with 2% trifluoric acid (TFA), sonicated for 30 seconds, and incubated at RT for ≥ 15
minutes with gentle agitation. Supernatant was transferred to a fresh tube and the gel pieces were shrunk with 100% ACN at RT at gentle agitation for ≥ 15 minutes to recover remaining peptides from the gel. Supernatant was pooled and subjected to vacuum centrifugation to remove the ACN and concentrate the sample. Thereafter, the peptide sample was desalted and concentrated by “STop And Go“ Extraction (STAGE) tips.

4.10 Mass spectrometric measurements

Measurements were performed by nanoflow reversed-phase C18 liquid chromatography (EASY nLC, Thermo Scientific) coupled online to a 7 Tesla linear ion trap Fourier-Transform ion cyclotron resonance mass spectrometer (LTQ FT Ultra, Thermo Scientific) or by nanoLC 1000 (Thermo Scientific) chromatography coupled online to Q Exactive hybrid quadrupole-Orbitrap mass spectrometer (Thermo Scientific). Chromatography was performed with an Acclaim PepMap 0.3 x 5 mm 5µm 100Å trap column (Thermo scientific) in combination with a 15cm long x 100µm ID fused silica electrospray emitter (New Objective, PicoTip Emitter, FS360-100-8-N-5-C15) packed in-house with ReproSil-Pur C18-AQ 3 µm 140Å resin (Dr. Maisch). Tryptic peptides were loaded onto the trap column using 0.1% formic acid and separated by a linear 60 minutes (LTQ-FT) or 30 minutes (Q Exactive) gradient of 5-35% acetonitril containing 0.1% formic acid at a flow rate of 300 nl/min. For the LTQ-FT; the mass spectrometer was set to positive ion mode and acquired one full MS survey scan in the ICR cell parallel to up to four data dependent collision induced dissociation (CID) fragmentation spectra by the linear ion trap. Full MS precursor scans were performed with a single microscan at 100.000 resolving power (FWHM) at m/z 400 using 1E6 ions or after 2500ms injection time if this came first. Data dependent acquisition of MS/MS spectra by the linear ion trap was performed on 3E4 ions or after 750 ms maximal injection time. Fragmentation of the precursor ion by CID was performed at 30% normalized collision energy for 30 ms and activation Q=0.25. An isolation width of 3 that was set to isolate the precursor ion for MS/MS sequencing events. For the Q Exactive; the mass spectrometer was again set to positive ion mode. Full MS events were performed at 70.000 resolving power (FWHM) at m/z 200 using 1E6 ions or after
20ms of maximal injection time. Data-dependent MS/MS spectra were performed using 1E5 ions at 17.500 resolving power (FWHM) at m/z 200 or after 50ms maximal injection time for the top 10 precursor ions with an isolation width of 4.0 Th and fragmented by higher energy collisional dissociation (HCD) with a normalized collision energy of 30%.

4.11 Mass spectrometric data analysis

Data analysis was performed with the MaxQuant software (version 1.3.0.5) applying default settings with minor modifications. The precursor mass tolerance for Q Exactive measurements was set to 4.5 ppm. For both LTQ-FT and Q Exactive the multiplicity was set to 1 and Trypsin was chosen as the proteolytic enzyme allowing for 2 miscleavages. Default MaxQuant normalizations were applied. Database searches were performed on the human RefSeq database in which, the reversed database is used to calculate the false discovery rate (FDR), which was set to 1% and isoleucine and leucine were forced to be treated equally. Between samples the option “Match between runs” was enabled to detect sequenced peptides, which were not subjected to sequencing event in other samples and Label Free Quantification (LFQ) calculation was applied. Peptide modifications after formaldehyde cross-linking did not occur as tested by the presence of two possible modifications occurring when the cross-linking is not reversed completely. The first modification is the addition of 30 Da considered to be the addition of the whole formaldehyde molecule (O=CH2) as an intermediate step in the cross-linking reaction. The second possible modification is the addition of 12 Da, which equals the addition of formaldehyde followed by the release of a water molecule, and is considered to be the final product. Since neither modification occurred, the reversal of cross-linking seems to be complete. Furthermore, there is an increased possibility of miscleavages since the reactivity of formaldehyde is the highest on those amino acids subjected to tryptic digestion, this did not seem to give any problems since we allowed for maximum of two miscleavages and were not able to detect any miscleavage in combination with peptide modifications. Raw data files provided by MaxQuant were further analysed manually. For the biological replicates LFQ values were used to calculate the ratios between samples per biological sample. For the triplicate
measurements performed on the Q Exactive first the average LFQ values were calculated from the replicates (only proteins identified in all three replica measurements were considered), followed by calculation of the ratios between sample conditions. Whenever the ratio exceeded the value of 2 or was below 0.5, the protein was called to be respectively increased or decreased. Additional protein information such as the Gene Ontology_SLIM_cellular compartment (CC), molecular function (MF), biological process (BP) and the official gene symbol were acquired using ProteinCenter (version 3.12.10015; Thermo Scientific).

4.12 Baculovirus protein expression

For Twinkle expression in Baculovirus system BacMagic system was used (Novagen). For production of Twinkle in SF9 insect cells a gene construct coding for mature Twinkle from aa 30-684 was cloned into vector pTriEx (Novagen) including an N’-terminal or C’-terminal hexahistidine-tag followed by a tobacco etch virus (TEV) proteinase cleavage site. *Spodoptera frugiperda* (Sf9) insect cells (1 × 10^6 cells/35-mm plate) were transfected with the recombinant pTriEx_Twinkle-TEV-6*His construct and virus amplification was carried out until the P3 stage. Protein production was optimized by different concentrations of P3 virus. 2 x 10^6 Sf9 cells/ml in log phase growth were infected with different amounts (50µl, 500µl, 1ml per 50 ml Sf9 culture) of P3 virus and incubated for 24, 48, 72, 96h at +27°C in a shaking incubator and analysed for Twinkle expression by Western blotting. Sf9 cells were harvested by centrifugation at 2500rpm for 5 min and cell pellet lysed in RIPA buffer and incubated on ice for 30 min. After centrifugation at 10000rpm for 30 min at +4°C protein concentration was determined by Bradford assay and 30µg of lysate was run for Western blotting. Twinkle antibody was used for detection as above. To test functionality of TEV site Twinkle protein was incubated at 1h with TEV protease (see below) in protease buffer (50mM Tris-HCl pH 8.0, 0.3mM EDTA, 1mM DTT).
4.13 Twinkle purification for TEV analysis

Twinkle expression was induced in 50 ml of Sf9 cells as described above. Cells were harvested by centrifugation at 2500rpm for 5 min at +4°C. Cell pellet was lysed in lysis buffer (50mM KPO<sub>4</sub> pH 7.0, 1M NaCl, 10% glycerol, 0.5% Triton X-100, 10µM imidazole and 5µM β-ME) and incubated on ice for 30min before sonication for 1 min at 40% power 1s on 1s off. Lysate was centrifuged for 30 min at 10000g at +4°C and Talon resin (Clontech) (equilibrated in lysis buffer) was added to the lysate and rotated at +4°C for 2 hours. Resin was washed once in wash buffer A (50mM Tris-HCl pH 7.4, 1M NaCl, 20mM imidazole, 5µM β-ME), once in wash buffer B (Tris-HCl pH 7.4, 0.8M NaCl, 40mM imidazole, 5µM β-ME). Twinkle was eluted in 500µl of elution buffer (Tris-HCl pH 7.4, 40% glycerol, 50mM L-Arginine, 250mM imidazole, 5µM β-ME).

4.14 TEV protease expression and purification

0.5µl of TEV-plasmid was transformed into E.Coli BL21 cells and plated on agar plates. One colony was incubated in 3 ml of LB medium and incubated over night at +37°C and from that culture 100µl was transferred to 50 ml culture. From the 50 ml culture 8 ml was transferred to 500 ml of LB and bacteria grown at +37°C until Abs<sub>600</sub> of 0.6 was reached. Protein expression was induced with 250µl of 1M IPTG (Sigma) for 500ml of culture and bacteria grown at RT for 8 h. Bacterial cultures were shaken at 150rpm. Bacteria were harvested by centrifugation at 6000g for 20 min at +4°C and kept on ice. Bacteria were lysed in TEV buffer A (10mM Imidazole, 20mM Tris-HCl pH 8, 150mM NaCl), sonicated and centrifuged at 10000g for 30 min at +4°C. For the supernatant 1 ml (0.5ml of actual resin) of Ni-NTA resin (QIAGEN) (equilibrated in TEV buffer A) was added and supernatant rotated at +4°C for 1 h. Resin was washed with 5 ml of TEV buffer A, then with 5ml of TEV buffer B (like buffer A but with 1M NaCl instead) and finally with 5ml of TEV buffer A. TEV was eluted with 0.5ml of TEV buffer C (like TEV buffer A but with 330mM imidazole and 10 % glycerol).
4.15 Further Twinkle purification

40ml of Sf9 culture was induced to express Twinkle and centrifuged (2500rpm 5 min at +4C) followed by lysis in 5ml of lysis buffer (50 mm KH$_2$PO$_4$, pH 7.4, 1M NaCl, 0.1% Triton X-100, 10 mm Imidazole, 5 mm β-mercaptoethanol) incubated on ice for 15min and sonicated 1 min 40% power 1s on 1 s off. Talon resin (Clontech) was equilibrated in lysis buffer and either settled in column or in 15 ml tube for patch purification. For patch purification lysate was incubated with Talon resin for 1h hour in rotation at cold room. For column purification lysate was added to the column and let through by gravity flow. Resin was washed twice in buffer 1 (50mM KPO$_4$ pH 7, 1M NaCl, 30mM Imidazole, 0.5% Triton X-100, 5mM β-ME) and twice in buffer 2 (20mM Tris-HCl pH 7.2, 800mM NaCl, 50mM Imidazole, 10mM β-ME) and eluted in buffer 3 (20mM Tris-HCl pH 7.2, 500mM NaCl, 150mM Imidazole, 10mM β-ME, 50% glycerol, 100mM L-arginine).

To compare different metal affinity based purification resins Twinkle expression was induced as described above. For Ni-NTA (QIAGEN) based purification cells were lysed for 15min at 4°C in buffer C (50 mM Tris-HCl pH 7.4, 1M NaCl, 20mM imidazole, 1% TX-100, 5mM β-MeOH) containing EDTA free protease inhibitor tablets. The lysate was sonicated (30 cycles, 1”on – 3” off) and centrifuged (10000xg) for 15 min at 4°C. The cleared lysate was bound to Ni-NTA beads for 2 h at 4°C on rotation and subsequently pelleted by centrifugation (1500xg) for 3min. The pellet was washed once in 20-30 bed volumes with buffer A (50 mM Tris-HCl pH 7.8, 800mM NaCl, 20mM imidazole, 5mM β-MeOH) and twice in same buffer containing 40 mM imidazole and 0.5M NaCl. The protein was eluted in a single step with buffer A containing 250mM imidazole, 500 mM NaCl and 100mM L-arginine. Talon based purification was carried out as described above. For the high specificity resin (PrepEase, Affymetrix) based purification the cell pellet was lysed in LEW buffer (50mM NaH$_2$PO$_4$, 1M NaCl, pH 8.0, 5mM β-MeOH). The lysate was sonicated (30 cycles, 1”on – 3” off) and centrifuged (10000xg) for 15 min at 4°C. The cleared lysate was bound to high specificity resin for 2 h at 4°C on rotation and subsequently pelleted by centrifugation (1500xg) for 3min. The pellet was washed once in 20-30 bed volumes with LEW buffer containing 20mM imidazole and twice in buffer containing 40mM imidazole and eluted in LEW buffer containing 250mM imidazole and 100mM L-arginine.
For optimization of Twinkle purification using Talon metal affinity resin Sf9 cell pellet was lysed in either 50mM KPO$_4$, 50mM NaPO$_4$ or 50mM Tris-HCl pH 7.4 which were in buffer containing 1M NaCl, 10% glycerol, 0.5% Triton X-100, 10µM imidazole and 5µM β-ME and lysate sonicated and centrifuged as above. The cleared lysate was bound to Talon resin and rotated for 4°C and subsequently pelleted by centrifugation (1500xg) for 3min. The pellet was washed once with the appropriate buffer for each condition (50mM KPO$_4$, 50mM NaPO$_4$ or 50mM Tris-HCl pH 7.4) containing 0,8M NaCl, 20mM imidazole, 5µM β-ME and twice in buffer containing 40mM imidazole, 0,5M NaCl. The protein was eluted in buffer containing 250mM imidazole, 0.5M NaCl and 100mM L-arginine.

For testing different detergents, 50mM KPO$_4$ buffer was used and purification carried out as above with the exception that in all buffers either 0.5% Triton X-100 or Tween 20 was used. For purification carried out in room temperature 0,5% Triton X-100 was used in lysis buffer. As control, purification was carried out where Triton X-100 was only in the lysis buffer as described above.

4.16 Chromatography based Twinkle purification

Twinkle expression was induced in 200 ml of Sf9. Twinkle protein was further purified using ÄKTA-FPLC liquid chromatography (GE Healthcare). For downstream experiments, Twinkle was eluted from Talon resin using 20mM Tris-HCl pH 8 with same buffer conditions described in section 1.16. Before loading to anion exchange column QXL (GE Healthcare) Twinkle eluate from Talon purification was adjusted to NaCl concentration of 200mM with buffer QXL (20mM Tris-HCl pH 8.6, 1mM DTT) and pH adjusted to pH 8.6. QXL was carried using buffers A: 20mM Tris-HCl pH 8.6, 200mM NaCl, 1mM DTT and buffer B: 20mM Tris-HCl pH 8.6, 1200mM NaCl, 1mM DTT, increasing the NaCl concentration using buffer B in a linear gradient. After buffers were cooled in cold room the pH was checked and adjusted to pH 8,6, filtered and DTT was added freshly. For cation exchange chromatography with SP FF (GE Healthcare) column Twinkle was eluted in 50mM KPO$_4$ pH 7.6 buffer. Eluates were adjusted to 200mM NaCl concentration using 50mM KPO$_4$ pH 7.6 buffer with 1mM DTT. SP FF buffer A, 50mM KPO$_4$ pH 7.6, 200mM NaCl, 10% glycerol, Buffer B, 50mM KPO$_4$ pH
7.6, 1M NaCl, 10% glycerol, 1mM DTT using gradient elution with buffer B. For gel filtration, Superdex 20/300G column was used (GE Healthcare).

4.17 Helicase assay

Helicase assay was carried out as described in Wanrooij et al (2007) except 3mM substrate was used. As standard substrate for helicase assays a radioactively end-labeled 60 nt oligonucleotide hybridized to M13 ssDNA was used (5'ACATGATAAGATACATGGATGAGTTTGGACAAACCACAACGTAAAAC GACGGCCAGTGCC 3'), forming a 20 nt double-stranded stretch with a 40 nts 5' overhang. The assay was performed by incubating 1 ng Twinkle protein in 40 μl helicase buffer (25 mM Tris-HCl 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 3 pmol 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 × TBE, dried on a vacuum gel drier and exposed to X-ray film or quantified by phosphoimager.
5. Results

5.1 Twinkle structural studies

Several Twinkle mutations have been associated with mtDNA maintenance diseases that lead to mtDNA deletions or depletion and hence to mitochondrial dysfunction. Therefore, the structural information of the protein would be beneficial for the understanding how the different mutations affect Twinkle function and lead to changes in mtDNA. One aim of this PhD thesis was to elucidate the Twinkle structure and for this, the first task was to purify Twinkle at high enough purity and yield.

5.1.1 Optimisation of Twinkle expression

For production of Twinkle for structural studies in Sf9 cells, the BacMagic Baculovirus expression system was used. After virus amplification the virus stock was tested for optimal protein production. Figure 5.1 a, shows that with 50µl of P3 virus stock in 10 ml Sf9 culture, the production of Twinkle is not optimal as it is expressed at low amounts in all time points. 1ml of P3 stock is also not optimal as Twinkle protein expression was not synchronised. Twinkle was expressed already in day 1 and 2 in lower amounts than in day 4 when 500µl of P3 virus stock was used. Virus volume of 500µl seems to be optimal for Twinkle production. Although a low amount of Twinkle is expressed at day 3, a significantly higher amount of Twinkle is expressed at day 4 giving the most synchronised expression pattern for the protein. To test the TEV proteinase cleavage site functionality Twinkle was incubated with TEV-protease. As can be seen from the Figure 5.1 b the TEV site is functional because with TEV incubation the Twinkle protein migrates faster in SDS-PAGE gel.
Figure 5.1: Optimisation of Twinkle expression A) Twinkle expression was optimised using three different P3 virus stock volumes and Twinkle expression analysed by Western blotting. B) TEV protease site was tested and protein visualised by Coomassie blue staining.

5.1.2 Optimisation of Twinkle purification

The behavior of Twinkle in column and batch purification was first tested. Twinkle is not eluted from the resin in column purification. The elution is more efficient when carried out in batch (Figure 5.2 a and b). Different metal affinity resins were tested for their purification efficiency, Nickel-NTA, Talon and PrepEAsE – high specificity (HS) resin. It should be noted that HS resin eluates were run on a separate SDS-PAGE gel. Talon resin shows best specificity based on Coomassie staining; however, there is a major contaminant that migrates lower than Twinkle that is not seen in HS resin. However, the HS resin seems to have lower specificity than Talon (Figure 5.2 c). Different buffer conditions with Talon purification were next tested to see if the protein could be eluted better and if the purity could be improved. Changing the buffering agent to either KPO4, NaPO4 or Tris-HCl pH 7.4 did not have any effect on the purity of Twinkle (Figure 5.3 a). Twinkle purification was more efficient if carried out at +4°C, but the choice of detergent during lysis did not affect the protein purity or yield, nor did the presence of detergent in all steps during the purification (Figure 5.3 b).
Figure 5.2: Twinkle purification with Talon metal affinity resin in A) column and in B) batch, both eluates (E) and resin (R) after eluation is shown in the figure. C) Testing of different metal affinity resins for Twinkle purification, Coomassie blue staining shows eluates from purification. Protein samples were run in SDS-PAGE gel and visualised by Coomassie staining. (H- PrepEase high specificity resin, T – Talon metal affinity resin, Ni – Ni-NTA metal affinity resin).

Figure 5.3: Optimisation of Talon resin purification. To analyse if Twinkle could be eluated from the resin more efficiently different purification conditions were tested including A) three different buffering agents were tested during Talon purification including KPO₄, NaPO₄ and Tris-HCl (see M&M for full buffer conditions) B) four different purification conditions were tested using KPO₄ as a buffering agent, including purification in room temperature and in +4°C, and in buffers including Triton X-100 or Tween 20 in all buffers, or Triton X-100 in lysis buffer only (TX = Triton X-100, KP=KPO₄, NaP=NaPO₄).
5.1.2.1 Further purification of Twinkle for EM studies using liquid chromatography

To further increase Twinkle purity liquid chromatographic methods were used. Both anion and cation exchange chromatography were tested for their suitability in Twinkle purification. Two different Twinkle purification schemes were employed. (Figure 5.4). After Talon purification cation exchange chromatography was carried out followed by buffer exchange using spin columns to facilitate TEV protease incubation. To remove TEV protease gel filtration was carried out. As can be seen from figure 5.4 a, using cation exchange protocol further Twinkle purity was achieved and by using gel filtration, the sample is further purified while TEV protease is removed. Figure 5.4 b, shows that the major contaminant left after Talon purification is not removed by anion exchange chromatography. After concentrating the peak fractions a TEV protease cut was carried out. To reduce complicated downstream applications Talon resin was added to remove the TEV protease that harbors a His-tag. Twinkle in this case was collected from the flow through fraction. Some of the major contaminant left from Talon purification still remains (figure 5.4 b). However, both of these samples from purifications a and b were sent for EM studies.
Figure 5.4: Further Twinkle purification. A) Twinkle produced in Sf9 cells was first subjected to Talon batch purification, followed by cation exchange chromatography (SP FF) after which Twinkle was subjected to TEV-protease treatment. Gel filtration chromatography was used for further Twinkle purification and to separate TEV from Twinkle. B) Twinkle produced in Sf9 cells was first subjected to Talon batch purification after which, anion exchange chromatography was used to further purify the protein (QXL). After anion exchange chromatography Twinkle was subjected to TEV-protease treatment and TEV removed by binding to Talon resin.

5.1.3 Structural analysis of Twinkle

EM analysis of Twinkle reveals that it is present as hexamers and heptamers (figure 5.5). Initial 3D structural construction also reveals that Twinkle may be present as double hexamers in this initial preparation (data not shown) (see Discussion).

5.1.3.1 Further Twinkle structural analysis

To elucidate the structural organisation of Twinkle, further studies using single-particle negative staining EM studies with Twinkle produced in either E.coli or insect cells were used. Despite the specificity of Talon resin purification, as can be seen above, quite a large amount of Twinkle typically remained bound to the resin following elution. Further testing led to an adapted purification protocol used for
insect cell produced Twinkle that used high salt buffer with NiNTA purification o/n at 4°C. With this protocol the fast majority of Twinkle could be eluted at the first elution step, thus presenting an improvement over the Talon protocol (see article III, and S. Cansız, unpublished observations). For both SF9 insect culture and *E. coli* produced Twinkle the 2D analysis revealed distinct averages with negative staining EM that contained either six or seven radial densities, confirming the initial results in figure 5.5 (see article III figure 2), indicating the coexistence of heptameric and hexameric ring-like structures. The initial structural modelling of samples used in figure 5.5 suggested a head-to-head double hexameric structure for Twinkle. However, later cryo-EM of GraFix Twinkle (see M&M article III, article III figure 3) did not confirm the initial observed double hexameric structure.

Further modelling of Twinkle (article III figure 4), gives additional information on Twinkle structure and domain orientation important in understanding the Twinkle disease mutants (article III figure 1).

Figure 5.5: Negative staining EM pictures reveal that Twinkle is present both in hexameric and heptameric structures.
5.2 Organisation of nucleoids

5.2.1 Twinkle shows properties of a mitochondrial inner membrane protein

Mammalian nucleoid research thus far has concentrated on finding new nucleoid proteins and there has not really been any studies concentrating on the known nucleoid proteins and how they behave. To investigate the characteristics of known nucleoid-associated proteins (NAP), and also to look into the long lasting enigma of mtDNA membrane association, we first set out to look into the known NAPs distribution in the mitochondria. We used carbonate extraction or high salt (KCl) treatment to investigate the localisation of several known nucleoid proteins. The idea of carbonate extraction is to see if a protein is an integral membrane protein. Sodium carbonate at pH 11.5 makes vesicles converting to sheets and protein-protein interactions are disrupted, but protein-lipid interactions should stay intact (Fujiki et al 1982). High salt treatment was used as an alternative method to confirm carbonate extraction. In addition mitochondrial digitonin fractionation was used to separate the different mitochondrial compartments matrix (M), intermembrane space (IMS), inner membrane (IM) and outer membrane (OM). We also used nuclease treatment (DNAse and/or RNAse) with the different fractionation protocols to see if the membrane connection is dependent on mtDNA. In our laboratory, we have several nucleoid proteins cloned into the HEK293 FlpIn™ TRex™ system that allows an inducible expression of the protein.

5.2.2 Mitochondrial nucleoid proteins show different fractionation properties

To test the localisation, various mitochondrial proteins were stably over-expressed and subjected to digitonin fractionation. From the figure 5.6 and thesis supplementary figure 1, we can see that both over-expressed and endogenous Twinkle is a mitochondrial inner membrane protein. In contrast, mtSSB, TFAM and POLG1 and POLG2 as well as the truncated variant of Twinkle, Twinky, were all found in the matrix fraction demonstrating that they show characteristics of more soluble proteins.
Figure 5.6: Mitochondrial nucleoid proteins show different fractionation properties. Using mitochondrial fractionation protocol Twinkle-Myc and ATAD3-HA fractionated to the IM whereas other nucleoid proteins including endogenous TFAM, POLG1-Myc, POLG2-HA and mtSSB-Myc, as well as the non-nucleoid Twinky-Myc fractionated to both IM and matrix (m refers to mitochondrial lysate, OM to outer mitochondrial membrane, IMSP to inter membrane space and IM to inner mitochondrial membrane).

5.2.3 Twinkle is associated to the mitochondrial inner membrane and two pools of mtDNA–protein complexes could be separated on the basis of their solubility

Because Twinkle does not possess any predictable trans-membrane domain but the above results strongly indicate Twinkle is a membrane protein, we wanted to confirm the Twinkle membrane association. To test this we subjected mitochondria to different mitochondrial fractionation protocols. We also investigated the possible involvement of RNA and/or DNA in the membrane anchorage of Twinkle. Using either carbonate extraction or high salt (0.5M KCl) treatment, we confirmed the Twinkle membrane association for both over-expressed and endogenous protein, indicating a strong association of the protein with the inner membrane. In contrast mtSSB and TFAM are found in both membrane and soluble fractions after KCl treatment indicating dissociation from the membrane (Figure 5.7 a and b).
When mitochondria are treated with nucleases to destroy DNA and/or RNA, Twinkle still remains membrane associated. In contrast, a substantial proportion of TFAM and mtSSB are released to the soluble fraction (figure 5.7 a and b).

Figure 5.7: Twinkle is membrane associated. Endogenous (A) and over-expressed Twinkle (B) Twinkle stays membrane bound, same as the inner membrane protein COXII, using either fractionation method (carbonate extraction or KCl) and is not dependent on DNA or RNA. In contrast, TFAM and mtSSB were found mostly in the soluble fraction and their membrane connection is partly due to DNA and/or RNA (p = pellet, s = supernatant).

To further investigate the nature of nucleoid proteins and their association with each other and mtDNA, a floatation assay was used. In this assay mitochondria were disrupted using either Triton X100 or digitonin and separated to soluble and insoluble fractions by centrifugation. The insoluble fraction of both fractionations were subjected to floatation assay (see M&M). This assay demonstrated that when using mild membrane fractionation by digitonin mtDNA, Twinkle and other nucleoid-associated proteins move up the gradient to a single location. In contrast COXII and MRPL49 (large ribosomal subunit) show different movement up the gradient. Indicating that mtDNA and the associated proteins exists as a single complex. The small ribosomal subunit (MRP S22) is also found in the nucleoid fraction (figure 5.8).
Figure 5.8: Nucleoid proteins form a single fraction in floatation assay. The digitonin lysed mitochondrial pellet fraction show that mtDNA and Twinkle, POLG1, mtSSB, TFAM and ATAD3 moved up the gradient to a single low-density iodixanol concentration. In contrast, inner membrane protein COXII and a marker for large ribosomal subunit MRPL49 remained at relativey high-density fractions.

5.2.4 Modelling of bacteriophage T7 helicase and Twinkle reveals a loop region in Twinkle that seem to have some role in Twinkle anchorage to the inner membrane

Twinkle has amino acid and sequence identity 48% and 15 % respectively to the bacteriophage T7 gene 4 helicase (T7 gp4) (Spelbrink et al 2001). Similar to T7 gp4 helicase Twinkle possess a helicase domain at its C’terminus, but unlike the T7 gp4 the primase function of N’terminal domain is lost (Ziebarth et al 2007), but the binding activity to ssDNA still remains (Farge et al 2007). Due to the homology of T7 gp4 and Twinkle helicase we carried out a tentative molecular fitting of Twinkle on the T7 gp4 structure using Swiss-Model. This analysis suggested a loop-region of amino acids 623-644 that was absent from that of T7 gp4 helicase (figure 5.9). To study if this loop region would be responsible for the membrane anchorage we set out to investigate a Twinkle mutant in which, the loop region was deleted. We also used other Twinkle deletion mutants, Twinkle del 31-49 and del 70-343 and del 29 and del 35 amino acids from the C’terminus (figure 5.10). Mitochondrial fractionation shows that only the TwinkleLoop deletion shows increased solubility,
indicating that this region of Twinkle may have some role in Twinkle membrane anchoring. Twinkle Del31-49 has arginine residues that were thought to have a possible role in membrane anchorage. Based on the mitochondrial fractionation assay these amino acids are not involved in Twinkle membrane association as the proteins stays in the membrane fraction (figure 5.11).

Figure 5.9: Computer modelling of Twinkle reveals a loop region (indicated in red at the top) that is distinct from that of T7 gp4 helicase.
Figure 5.10: Schematic representation of Twinkle mutants used in carbonate extraction and/or high salt treatment to test Twinkle membrane association.

Figure 5.11: Carbonate extraction (a and b) and high salt mitochondrial fractionation (c) reveals that the loop region of Twinkle might be partly responsible for the membrane anchoring. Isolated mitochondria from HEK293 cells transfected with Twinkle mutants were subjected to either carbonate extraction or high salt treatment (see M&M). Twinkle and TFAM were detected by Western blotting. P refers to pellet fraction and s refers to the soluble fraction, WT = wild type.
5.2.5 Mitochondrial nucleoids are heterogeneous and dynamic in nature

The results above clearly show that NAPs have different characteristics. To further test the nature of nucleoid dynamics in mammals we carried out various immunofluorescence based assays. In our studies, we were also able to study the low abundant endogenous protein, which has not been possible earlier due to the lack of a good antibody. The verification of Twinkle antibody results can be found in the thesis supplementary section (Figure 2) and in figure 5.14.

5.2.6 Twinkle and mtSSB are not constitutive nucleoid components

Immunofluorescence studies with endogenous Twinkle and co-staining with TFAM and/or mtDNA show that Twinkle co-localises with a subset of nucleoids. Approximately 48% of the mtDNA positive foci were also Twinkle positive, whereas TFAM always co-localises with mtDNA and vice versa (figure 5.12 a). MtSSB showed lower co-localisation with mtDNA than Twinkle (14%). Also, in addition to focal distribution, mtSSB showed a more uniform distribution within the mitochondrial network (Figure 5.12 b and c). The above results were verified using various cell lines, as well as different fixation and permeabilisation methods to demonstrate that the nucleoid heterogeneity is not cell line or cell fixation specific (see below and in the article supplementary figures).

5.2.7 Twinkle organisation is independent of mtDNA

As the results above show that some of the replication machinery might organise in structures that are independent of mtDNA we set out to investigate this finding further. Also, previous studies in our laboratory have demonstrated that even in the absence of mtDNA in Rho0 cells overexpressed Twinkle-GFP stays as punctate foci while mtSSB is more dispersed as is TFAM that in addition shows much weaker staining (Garrido et al 2003). To further study the organisation of endogenous Twinkle in the absence of mtDNA we used immunofluorescence techniques to visualise cells in which, mtDNA is depleted using dideoxycytidine (ddC) treatment
as well as in cells that mtDNA is absent, Rho0 (206f–ρ0, of the 143B osteosarcoma parental cell line). Results show that even in the absence of mtDNA, either in ddC treated or Rho0 cells, endogenous Twinkle remains as punctate foci (figure 5.13. a and b). In Rho0 cells mtSSB was more uniformly distributed over the mitochondrial network (Figure 5.13.b). The previous results for TFAM also were verified demonstrating much weaker TFAM signal in Rho0 cells when compared to the parental cell line (see thesis supplementary figure 3).

![Figure 5.12](image.jpg)

Figure 5.12: Endogenous Twinkle is found only in a subset of nucleoids. A) All mtDNA foci were TFAM positive, while ~ 50% of mtDNA foci were not positive for Twinkle (some of these foci are indicated with white arrow). In addition, some Twinkle foci were observed also in the absence of mtDNA (foci indicated with green arrows). B) In a second fibroblast line <50% of mtDNA foci were positive to Twinkle, while even fewer mtDNA foci were strongly positive for mtSSB. C) The graph shows the calculated percentages of mtDNA foci positive for Twinkle or mtSSB. In these experiments TFAM foci was never seen without mtDNA.

5.2.8 mtSSB organisation at mtDNA foci is Twinkle dependent

Results thus far indicate the importance of Twinkle in nucleoid membrane connections. Hence, we next wanted to carry out experiments with transient Twinkle knock down to investigate the effect of Twinkle depletion on nucleoids. Transient
Twinkle knock down leads to dramatic loss of mtSSB foci. This is illustrated with the decrease from 70 intense mtSSB foci in wild type cells to only 10 foci remaining in cell treated with Twinkle siRNA (Figure 5.14 a and b1 and 2). The same is illustrated in the larger images where mtSSB foci are completely absent compared to the wild type cells. (Figure 5.14 b1 and 2). Figure 5.14 B2 also shows that upon transient Twinkle knock down mtDNA positive for Twinkle decrease from 30 to 8%. It should be noted that the images also show a non-mitochondrial background fluorescence that is not sensitive for the assay, please see the article supplementary figures for the control experiments. The importance of Twinkle on regulation of mtSSB in nucleoids was further confirmed by immunofluorescence studies carried out using Twinkle replication stalling mutants (Wanrooij et al 2007) or in cells where mtDNA was depleted using ddC resulting in stronger mtDNA loss. Results show that when cells are expressing these Twinkle mutants this results in a loss of mtSSB foci in contrast to cells expressing the wild type protein (thesis supplementary figure 4). When a stronger mtDNA depletion is carried out by ddC more mtSSB foci co-localise with mtDNA when compared to non-treated cells.
Figure 5.13: Twinkle membrane association is not dependent on mtDNA (A1, 2 and 3) After depletion of mtDNA by ddC treatment multiple Twinkle foci still remains (green arrows). B1 and B2) In cells lacking mtDNA (B2p0 or 206f cells) endogenous Twinkle still remain as punctate foci while mtSSB shows more uniform staining when compared to cells containing mtDNA.
Figure 5.14: Knock down of Twinkle cause reduction in mtSSB Twinkle knock down in two different cell lines U2OS (A) and fibroblast (B1) demonstrating that mtDNA foci still remains when Twinkle foci are absent when compared to the control cells (quantified in panel B2). Strong mtSSB/mtDNA or mtSSB/Twinkle or mtDNA/Twinkle foci are absent. Tw = Twinkle, pos = positive (-) Twinkle knock down cells, (+) Twinkle positive cell.
5.2.9 Twinkle and mtSSB are enriched in replicating nucleoids

Yeast research has shown that replicating nucleoids in mitochondria organise at the mitochondrial inner membrane when replication takes place, but this has not been shown in mammals. Above results, strongly suggest a similar type of organisation in mammalian nucleoids. For example, we showed that not all mtDNA foci contain Twinkle or mtSSB. Moreover, as Twinkle is the only known mitochondrial helicase, the Twinkle foci lacking mtDNA suggest that for replication to take place nucleoids must dynamically associate with Twinkle. To investigate the hypothesis if Twinkle and/or mtSSB association with mtDNA is replication dependent we used ClickIt-EdU (Lentz et al 2009) and BrdU labelling to detect de novo mtDNA synthesis. We reasoned that if the association of Twinkle and/or mtSSB with replicating mtDNA is dynamic we could detect this most efficiently by short Edu/BrdU pulses. This hypothesis was confirmed and the results show that with shorter Edu labelling the highest proportion of EdU positive foci were Twinkle positive (73%) and the proportion of positive foci declines over time (thesis supplementary figure 6 shows EdU-Twinkle-mtDNA labelling sample images for 30 min and 90 min EdU labelling time points, quantification is illustrated in figure 5.15 a and b). Similar results were seen in BrdU labelling with mtDNA/mtSSB. In this case after a 30 min pulse, 69% percent of BrdU foci were positive for mtSSB that dropped to 41% after 90 min (Figure 5.15 a, for microscope images see thesis supplementary figure 7). These results show that for EdU or BrdU labelling to occur Twinkle needs to be mtDNA associated, hence demonstrating that the association of mtDNA with Twinkle is dynamic.
Figure 5.15: Twinkle and mtSSB are enriched in replicating nucleoids. A) The graph shows the relative percentage of EdU/Twinkle and BrdU/mtSSB positive foci, with the 30 min time-point set to 100%. In reality, this time-point showed 73% ±8% of all EdU foci to be Twinkle positive and 69% ±8% of all BrdU foci to be mtSSB positive. B) Fibroblasts were labelled for the indicated times for EdU, and Twinkle & mtSSB and co-localisation determined. At the same time, a parallel slide from the same 6 well plate was processed for Twinkle, mtSSB and mtDNA detection to obtain steady-state co-localisation of Twinkle and/or mtSSB with mtDNA.

5.2.10 Twinkle determines the distribution of mtDNA and TFAM in membrane fraction

Subfractionation studies of this thesis work show that nucleoid proteins can be found in both membrane and matrix fractions. In contrast to the over-expressed Twinkle the endogenous protein shows decreased co-localisation with mtDNA. In addition, the immunofluorescence studies demonstrate the dynamic nature of nucleoids; only a subset of nucleoids contain Twinkle and furthermore the nature of Twinkle association with mtDNA is replication dependent. Although Twinkle does not possess a predictable transmembrane domain, above results also strongly point toward Twinkle as the membrane anchor. Hence, we hypothesised that Twinkle could shift the mtDNA and associated proteins membrane association dynamics. To study this hypothesis we overexpressed Twinkle and used mitochondrial fractionation experiments to see if there would be a shift of mtDNA and associated proteins to the membrane fraction.

The results of Western blotting and dot blot analysis showed that Twinkle overexpression resulted in almost complete retention of mtDNA and redistribution
of TFAM to the non-soluble fraction, in particular for the three days induction period (Figure 5.16). Western blot image shows that POLG1 shows similar trend than TFAM and shifts to the non-soluble fraction upon Twinkle over expression.

![Western blot image showing TFAM and POLG1 distribution](image)

**Figure 5.16:** Twinkle over expression resulted in mtDNA and TFAM redistribution. Mitochondria from Twinkle expressing or non-expressing cells were subjected to digitonin lysis and soluble matrix (s) fractions and non-soluble pellet fractions (p) were separated by centrifugation. Upper left panel indicates the amount of mtDNA detected by dot blot analysis and proteins detected in each fraction by Western blotting are shown in the lower left panel. The graphs on the right show the distribution of TFAM and mtDNA as calculated from the Western blot or dot blot images respectively.
To further test the involvement of Twinkle in the mtDNA and associated protein distribution we used Twinkle siRNA treated cells in our mitochondrial fractionation protocol. Results of this work show that Twinkle depletion cause mtDNA to shift more to the soluble fraction, but the shift is not complete and substantial amount of mtDNA still remains in the pellet fraction (Figure 5.17 graph). This result shows that Twinkle is not the only factor responsible for mtDNA distribution at mitochondria inner membrane. MtSSB showed a similar trend with redistribution of the protein to the more soluble fraction after Twinkle depletion (Figure 5.17 Western blot image), which is in agreement with the results presented earlier. The mtSSB presence at the inner membrane is dependent on Twinkle.

Figure 5.17: Twinkle depletion results in a shift of mtDNA and mtSSB to the soluble fraction. Using the same mitochondrial fractionation protocol as in figure 5.7 and using Twinkle knock down cells (3d) show an increase in mtDNA solubility (n = 4) compared with non-treated and non-targeting siRNA.

5.3 Investigation on the composition of mammalian mitochondrial nucleoids

Since previous nucleoid research has been aimed more towards identification and analysis of one or two proteins and not been directed towards quantitative proteomics we still lack the consensus list of nucleoid proteins. We aimed to
develop an easy and fast nucleoid purification method combined with quantitative mass spectrometry analysis, so that we could find a more inclusive list of NAPs and have a method that could be widely used in the identification of mitochondrial protein-protein and protein-nucleic acid interactions. To avoid the laborious mitochondrial isolation and subsequent nucleoid purification the idea was to carry out the nucleoid isolation from the whole cell culture. Formaldehyde (FA) was used to cross link protein-protein and protein-nucleic acid interactions. This has the advantage of preserving weak and dynamic interactions even in a harsher purification environment. One of the advantages of using FA is its capability to rapidly permeate live cells achieving stable covalent but reversible cross-links within minutes. In addition, since FA is very small, cross-linked amino acids and nucleic acid must be in close proximity to each other (Sutherland 2008).

5.3.1 Development of whole cell cross linking method for nucleoid associated protein identification

Previous MS based nucleoid research has concentrated on purifying NAPs from isolated mitochondria, our aim was to test NAP isolation from whole cell cultures using formaldehyde cross linking together with FLAG-immunoaffinity purification (IAP). Twinkle was chosen as the bait as results above and previous research show that it is a core component of nucleoids. Because Twinkle is a low abundant protein we used the inducible cell line system to express TwinkleFLAG to allow for IAP. We have previously shown that low and short over expression of Twinkle does not interfere with mtDNA maintenance (Wanrooij et al 2007 and Goffart et al 2009). To initially test the method we analysed the IAP eluates by Western blot analysis. These experiments show that several known nucleoid proteins TFAM, POLG1 and mtSSB are specifically enriched by FA crosslinking in TwinkleFLAG samples, following FLAG IAP (Figure 5.18), whereas control luciferase FLAG shows no enrichment of TFAM or mtSSB. Twinkle is purified with similar efficiency in samples with or without formaldehyde cross-linking, demonstrating that cross linking does not interfere with immunoprecipitation. (Figure 5.18 a). In the absence of mtDNA there is a marked reduction in TFAM and mtSSB co-purification (Figure 5.18 c).
Figure 5.18: Nucleoid proteins can be enriched using formaldehyde cross linking in whole cell culture in combination with immune affinity purification. Known nucleoid proteins TFAM, mtSSB and POLG1 can be enriched with cross-linking in cells expressing TwinkleFLAG (a and b). In the absence of mtDNA the enrichment of TFAM and mtSSB in substantially reduced.

5.4 Identifying potential nucleoid associated proteins using mass spectrometry

The above results demonstrated the possibility of identifying mitochondrial NAPs from the whole cell culture. To more systematically analyse the samples, address the question of sample complexity and possible non-mitochondrial protein contamination we next used LTQ-FT mass spectrometry analysing the entire protein composition of these samples using shotgun proteomics.
5.4.1 Validation of whole cell cross linking method by mass spectrometry analysis

To validate our whole cell cross-linking method we first analysed several biological repeats from TwinkleFLAG and mtLucFLAG cells. Figure 5.19 illustrates the sample analysis. Full mass spec raw data and analyses tabs can be found in Suppl Table 1 in article II.

Figure 5.19: A schematic representation showing the data analysis for the biological repeats used in the present study.

When comparing the enriched proteins, cross-linking results show enrichment of mitochondrial proteins in both Twinkle and luciferase expressing cells, 70%, when compared to non-cross linked where only 28% of the proteins show mitochondrial annotation. (Figure 5.20). The raw data output and analysis can be found in the article II supplementary data Table 1.

Figure 5.20: Mass spectrometry data analysis shows that we can enrich mitochondrial proteins by formaldehyde cross-linking.
By comparing our findings with previously published nucleoid research data the most interesting proteins are found in TwinkleFLAG +XL compared to TwinkleFLAG –XL and mtLucFLAG XL (marked as red border in Venn diagram (Figure 5.21). If we compare our results to already published nucleoid datasets, we can find many of the known proteins, including TFAM, POLG1, mtSSB and POLRMT, from the TwinkleFLAG XL enriched proteins (168 proteins: an annotated version is presented in Suppl Table 2, found in article II) (table 5.1). The resulting list of 168 is used in later comparisons (see figure 5.22) and is separately given alphabetically by gene name in Supplemental Table 4 (found in article II supplementary figures - first Tab: 'Biol repeats enriched all').

Figure 5.21: Whole cell cross-linking method is applicable for identification of nucleoid associated proteins. Further data analysis of TwinkleFLAG versus LucFLAG both with or without cross-linking shows the potentially interesting proteins enriched in TwinkleFLAG cross-linked samples (168 proteins, marked by red circle).
Table 5.1: Comparison of proteins identified in our nucleoid purification work to previously published nucleoid research. Checkmark = 2 fold increase in the TwinkleFLAG IAP compared to the mtLucFLAG IAP control with XL, cross = no difference, dark boxes = undetected protein. *= proteins, which are increased in TwinkleFLAG vs mtLucFLAG XL but no compared to TwinkleFLAG –XL. For ρ° samples - % of protein, purified in the absence of mtDNA.
5.4.2 Optimisation of whole cell cross linking method for nucleoid protein identification

5.4.2.1 Comparison on QLT FT and QExactive mass spectrometry methods

The above results show that whole cell crosslinking can be used in the identification of mitochondrial nucleoid proteins using shotgun proteomics. Therefore, we next wanted to further optimise the method and to do this we set out to carry the NAP identification method using two different IAP purification conditions. The first method was the one used for above analysis of the biological repeats where IAP is carried out in 300mM NaCl conditions with Triton X-100 lysis and sonication named here the TX-100 method. The second method uses RIPA buffer in all steps during the isolation, named here the X-Chip method (see M&M). After cell lysis and prior to IAP the cell lysates are sonicated in both methods, but in addition to sonication also a nuclease treatment in the X-Chip method was carried out. For this part of the study we used a QExactive mass spectrometer that possesses a higher resolution and sensitivity and also allows comparison of samples using label free quantification (LFQ) values IAP eluates were analysed as triplicate technical repeats.

To first compare the two different MS methods LTQ FT and QExactive we measured one of the biological repeats (IAP carried out by TX-100 method) in triplicate on the QExactive and compared samples shown below (Figure 5.22):

TwinkleXL  mtLuc XL  Twinkle -XL  mtLuc -XL

The 192 interesting proteins found (highlighted in red figure 5.22 A, article supplementary table 4, second tab: ‘TX100 enriched all’ and Supplemental Table 2), in addition is sorted in such a way that the same 192 proteins are the first proteins listed in the LFQ comparison tab (Tab 3) were used in later comparisons. When the interesting protein lists from LFQ-FT and QExactive are compared we can see that that there is substantial overlap of enriched proteins for both sets of experiments.
Overall, however, more proteins were identified with the QExactive instrument. (Figure 5.22, Table 5.1, Suppl Table 3 (can be found in article II). In both experiments the enriched proteins include many of the established nucleoid associated proteins (Table 5.1).

Figure 5.22: Enrichment of proteins in TX100 based NAP purification using Q Exactive mass spectrometry (A) TX100 sample 2 (figure 5.21) was measured as technical repeat and potentially interesting proteins were compared using Venn diagram. B) Comparison of the enriched proteins from the biological repeats using LTQ FT (figure 5.21) to the technical repeats using Q Exactive.

5.4.2.2 **Comparison of two different immuno affinity purification methods using QExactive mass spectrometry method**

To compare the two different IAP methods, the enriched proteins from figure 5.22 were compared to proteins enriched from X-Chip method using Q Exactive mass spectrometry. This comparison shows a considerable number of proteins that were identified with both methods (Figure 5.23 a and b). However, the X-Chip method identified many more NAPs in comparison to the TX100 method. For example a number of proteins with functions in RNA metabolism and translation, mitoribosomal proteins, DDX28, TACO1, MTIF2 and MTRF1. In addition, the X-
Chip method also identified the mitochondrial polymerase accessory subunit, POLG2 and the recently identified nuclease MGME1) (Di Re et al 2009, Kornblum, 2013 and Szczesny, 2013) (Article II supplementary table 2 and 3).

Figure 5.23: The X-Chip method is more stringent in comparison to the TX100 method for NAP identification (A) Q Exactive orbitrap was used to analyse samples from X-Chip NAP protocol. Potentially interesting proteins enriched in TwinkleFLAG XL sample are outlined by red in the Venn diagram. (366 proteins Supplemental Table 4 (see article II), (third Tab: ‘X-ChIP enriched all’) (table 5.2). Supplemental Table 3 (see article II), in addition is sorted in such a way that the same 366 proteins are the first proteins listed in the LFQ comparison tab (Tab 3). (B) Comparison of enriched proteins in TX100 (figure 5.22) to proteins enriched in X-Chip method.

Possibly the X-ChIP protocol, instead of removing all proteins that are indirectly associated with TwinkleFLAG either via DNA or RNA, might result in a less tightly packed complex in turn resulting in better accessibility of the FLAG epitope for TwinkleFLAG IAP. This would explain the approximately 10-fold higher LFQ values for Twinkle with X-ChIP compared to the TX100 Q Exactive measurements, whereas mtLucFLAG LFQ values are comparable between both sets (Supp Tables 2/3). This can then be expected to result also in a much better recovery of cross-linked mitochondrial proteins in the X-ChIP experiment. Not surprisingly, 98% of all proteins enriched with both the TX100 and the X-ChIP method are mitochondrial.
(figure 5.24). Moreover, 88% of proteins that are specifically enriched with the X-ChiP method had a mitochondrial annotation while in contrast, 36% of the proteins that showed specific enrichment only with the TX100 method were mitochondrial, suggesting many of these proteins are contaminants in the preparation (figure 5.24).

![Figure 5.24: Proteins enriched in TX100 and X-Chip using QExactive.](image)

**5.4.3 Application of whole cell cross-linking method for identification of proteins in a minimal replication platform in the absence of mtDNA**

Results of this thesis work show that Twinkle helicase is firmly membrane associated and that even in the absence of mtDNA it forms discrete membrane associated foci within the mitochondrial network. We next wanted to investigate the composition of these Twinkle associated nucleoid structures. For this we analysed NAPs using the X-Chip and QExactive method from normal Twinkle expressing cells and in cells lacking mtDNA ($\rho^0$). In addition this experiment would also allow us to identify proteins that are indirectly associated with nucleoids via DNA/RNA.

For the comparison of proteins identified in cells lacking DNA versus cell containing DNA we used the list of proteins (366 proteins, supplementary table 3 and 4, article II, figure 5.23) enriched with the X-Chip protocol when comparing TwinkleFLAG XL to no-crosslinked control or to LucFLAG XL samples using QExactive mass spectrometry. Out of the 366 proteins 95 are completely lacking in Rho0 cells including proteins with known function in mtDNA metabolism, MGME1, MTERF and POLG2. These proteins also included proteins with function in mitochondrial translation, such as ribosomal proteins, tRNA synthetases,
translation and RNA processing factors. Out of the 366 proteins 163 showed a ≥2 fold decrease including proteins in mtDNA metabolism or RNA processing, such as DHX30 and DDX28, LONP1, GRSF1 and POLRMT (Figure 5.25, table 2). ATAD3, MTERFD2 and ATP synthase subunits were found amongst the proteins that were either ≥2 fold increased or showed no change. These proteins may also be of interest and could represent proteins that are more stably associated with the membrane associated minimal nucleoid structure.

Figure 5.25: The pie-chart shown here illustrates the distribution of the 366 enriched proteins identified with the X-ChIP method in regular HEK293 TwinkleFLAG and measured in HEK293 TwinkleFLAG ρ° in the following classes: not detected (absent), 95 proteins; ≥2 fold decrease, 163 proteins; no change (100) or ≥2 fold increase (8)(see also Suppl Table 3 article II). Light gray boxed text shows abridged lists of proteins in each of the three categories selected from Suppl Table 3 (see article II).

Table 5.2: (see next page) List of the 366, enriched proteins identified with X-Chip method presented in figure 5.23

102
Proteins completely disappeared after mtDNA depletion (95 proteins)
<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Protein Name/Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 kDa heat shock protein, mitochondrial</td>
<td></td>
</tr>
<tr>
<td>AFG3L2</td>
<td>AFG3-like protein 2</td>
</tr>
<tr>
<td>CLPK</td>
<td>ATP-dependent Clp protease ATP-binding subunit clpL-like, mitochondrial precursor</td>
</tr>
<tr>
<td>SUP3L1</td>
<td>ATP-dependent RNA helicase SUP3L1, mitochondrial precursor</td>
</tr>
<tr>
<td>DUAT</td>
<td>dihydroxytyrosine-residue acetyltransferase component of peroxyl dehydrogenase complex, mitochondrial precursor</td>
</tr>
<tr>
<td>POLG</td>
<td>DNA polymerase subunit gamma-1</td>
</tr>
<tr>
<td>POLRMT</td>
<td>DNA-directed RNA polymerase, mitochondrial precursor</td>
</tr>
<tr>
<td>TUFM</td>
<td>elongation Factor Tu, mitochondrial precursor</td>
</tr>
<tr>
<td>LRRPNC</td>
<td>leucine-rich PPR motif-containing protein, mitochondrial precursor</td>
</tr>
<tr>
<td>LONP1</td>
<td>lon protease homolog, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>PTCO3</td>
<td>pentatricopeptide repeat-domain-containing protein 3, mitochondrial precursor</td>
</tr>
<tr>
<td>POLDP2</td>
<td>polymerase delta-interacting protein 2 isoform 1</td>
</tr>
<tr>
<td>IDOX2B</td>
<td>probable ATP-dependent RNA helicase IDOX2B</td>
</tr>
<tr>
<td>DHX30</td>
<td>putative ATP-dependent RNA helicase DHX30 isoform 1</td>
</tr>
<tr>
<td>SHMT2</td>
<td>serine hydroxymethyltransferase, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>SBP1</td>
<td>single-stranded DNA-binding protein, mitochondrial precursor</td>
</tr>
<tr>
<td>TFAM</td>
<td>transcription factor A, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>HADHA</td>
<td>trifunctional enzyme subunit alpha, mitochondrial precursor</td>
</tr>
<tr>
<td>HADHB</td>
<td>trifunctional enzyme subunit beta, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>MRPS10</td>
<td>28S ribosomal protein S10, mitochondrial</td>
</tr>
<tr>
<td>MRPS11</td>
<td>28S ribosomal protein S11, mitochondrial isoform a</td>
</tr>
<tr>
<td>MRPS17</td>
<td>28S ribosomal protein S17, mitochondrial</td>
</tr>
<tr>
<td>MRPS18B</td>
<td>28S ribosomal protein S18b, mitochondrial</td>
</tr>
<tr>
<td>MRPS2</td>
<td>28S ribosomal protein S2, mitochondrial</td>
</tr>
<tr>
<td>MRPS23</td>
<td>28S ribosomal protein S23, mitochondrial</td>
</tr>
<tr>
<td>MRPS25</td>
<td>28S ribosomal protein S25, mitochondrial</td>
</tr>
<tr>
<td>MRPS26</td>
<td>28S ribosomal protein S26, mitochondrial</td>
</tr>
<tr>
<td>MRPS28</td>
<td>28S ribosomal protein S28, mitochondrial</td>
</tr>
<tr>
<td>DAF3</td>
<td>28S ribosomal protein S29, mitochondrial isoform 1</td>
</tr>
<tr>
<td>MRPS31</td>
<td>28S ribosomal protein S31, mitochondrial</td>
</tr>
<tr>
<td>MRPS36</td>
<td>28S ribosomal protein S36, mitochondrial isoform b</td>
</tr>
<tr>
<td>ODGH</td>
<td>2-oxoglutarate dehydrogenase, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>BCKDHA</td>
<td>2-oxoisovalerate dehydrogenase subunit alpha, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>MRPL11</td>
<td>39S ribosomal protein L11, mitochondrial isoform a</td>
</tr>
<tr>
<td>MRPL12</td>
<td>39S ribosomal protein L12, mitochondrial</td>
</tr>
<tr>
<td>MRPL15</td>
<td>39S ribosomal protein L15, mitochondrial</td>
</tr>
<tr>
<td>MRPL22</td>
<td>39S ribosomal protein L22, mitochondrial isoform a</td>
</tr>
<tr>
<td>MRPL23</td>
<td>39S ribosomal protein L23, mitochondrial</td>
</tr>
<tr>
<td>MRPL24</td>
<td>39S ribosomal protein L24, mitochondrial</td>
</tr>
<tr>
<td>MRPL38</td>
<td>39S ribosomal protein L38, mitochondrial</td>
</tr>
<tr>
<td>MRPL39</td>
<td>39S ribosomal protein L39, mitochondrial isoform b</td>
</tr>
<tr>
<td>MRPL40</td>
<td>39S ribosomal protein L40, mitochondrial</td>
</tr>
<tr>
<td>MRPL41</td>
<td>39S ribosomal protein L41, mitochondrial</td>
</tr>
<tr>
<td>MRPL43</td>
<td>39S ribosomal protein L43, mitochondrial isoform a</td>
</tr>
<tr>
<td>MRPL45</td>
<td>39S ribosomal protein L45, mitochondrial isoform 1</td>
</tr>
<tr>
<td>MRPL47</td>
<td>39S ribosomal protein L47, mitochondrial isoform a</td>
</tr>
<tr>
<td>MRPL55</td>
<td>39S ribosomal protein L55, mitochondrial isoform b</td>
</tr>
<tr>
<td>HSDB10</td>
<td>2-oxoacyl-CoA dehydrogenase type-2 isoform 1</td>
</tr>
<tr>
<td>HICBCH</td>
<td>2-oxoacyl-CoA hydrolase, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>ACA2</td>
<td>3-ketoacyl-CoA thiolase, mitochondrial</td>
</tr>
<tr>
<td>NT5DC2</td>
<td>2'-0-methyltransferase domain-containing protein 2 isoform 1</td>
</tr>
<tr>
<td>NT5DC3</td>
<td>2'-0-methyltransferase domain-containing protein 3</td>
</tr>
<tr>
<td>ACAT1</td>
<td>acetyl-CoA acetyltransferase, mitochondrial precursor</td>
</tr>
<tr>
<td>ACS1</td>
<td>acetyl-coenzyme A synthetase 2-like, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>ACSF3</td>
<td>acyl-CoA synthetase family member 3, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>ACOT1</td>
<td>acyl-CoA thioesterase 1</td>
</tr>
<tr>
<td>ACOT13</td>
<td>acyl-CoA thioesterase 13 isoform 1</td>
</tr>
<tr>
<td>ACOT7</td>
<td>acyl-CoA thioesterase 9, mitochondrial isoform a precursor</td>
</tr>
<tr>
<td>FAH1D1</td>
<td>fatty acid hydrolase 1D1, mitochondrial isoform 1</td>
</tr>
<tr>
<td>AARS2</td>
<td>aspartate--tRNA ligase, mitochondrial</td>
</tr>
<tr>
<td>ALDH1B1</td>
<td>aldehyde dehydrogenase X, mitochondrial precursor</td>
</tr>
<tr>
<td>ALDH2</td>
<td>aldehyde dehydrogenase, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>ALDH7A1</td>
<td>alpha-aminoadipic semialdehyde dehydrogenase isoform 1 precursor</td>
</tr>
<tr>
<td>DARS2</td>
<td>dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>BCA7</td>
<td>branched-chain amino-acid aminotransferase, mitochondrial isoform a precursor</td>
</tr>
<tr>
<td>CDS5RAP1</td>
<td>CDS5 regulatory subunit-associated protein 1 isoform a</td>
</tr>
<tr>
<td>ADCX3</td>
<td>acyl-CoA oxidase 3</td>
</tr>
<tr>
<td>LSTE</td>
<td>leucine synthetase, mitochondrial precursor</td>
</tr>
<tr>
<td>MRA8A</td>
<td>co670-tyrosin acid a-c-diamide adenosyltransferase, mitochondrial precursor</td>
</tr>
<tr>
<td>C12BP</td>
<td>complement component 1 C1 subcomponent-binding protein, mitochondrial precursor</td>
</tr>
<tr>
<td>LRYM7</td>
<td>complex III assembly factor LRYM7</td>
</tr>
<tr>
<td>COR07-PAM16</td>
<td>COR07-PAM16 protein</td>
</tr>
<tr>
<td>NF52</td>
<td>cysteine desulfurase, mitochondrial isoform a precursor</td>
</tr>
<tr>
<td>UGCRc1</td>
<td>cysteine-b-c subunit complex L</td>
</tr>
<tr>
<td>UGCRF51</td>
<td>cysteine-b-c complex subunit Rieske, mitochondrial</td>
</tr>
<tr>
<td>COX41R</td>
<td>cysteine-c oxidase subunit 4 isoform 1, mitochondrial precursor</td>
</tr>
<tr>
<td>COX5B</td>
<td>cysteine-c oxidase subunit 5B, mitochondrial precursor</td>
</tr>
<tr>
<td>LAP5</td>
<td>cysteol amidopropylase</td>
</tr>
<tr>
<td>PDS2</td>
<td>2-decaprenylphosphate synthase subunit 2</td>
</tr>
<tr>
<td>ECH1</td>
<td>beta-tetrahydrofolate dehydrogenase, mitochondrial precursor</td>
</tr>
<tr>
<td>DLST</td>
<td>dihydroxyisobutyryl-CoA hydrolase, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>TOP3A</td>
<td>DNA topoisomerase 3-alpha</td>
</tr>
<tr>
<td>ETIA</td>
<td>electron transfer flavoprotein subunit alpha, mitochondrial isoform a</td>
</tr>
<tr>
<td>TFAM</td>
<td>elongation factor T5, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>EC2</td>
<td>enoyl-CoA delta isomerase 2, mitochondrial isoform 2</td>
</tr>
<tr>
<td>FASTD2</td>
<td>FAST kinase domain-containing protein 2</td>
</tr>
<tr>
<td>FASTD5</td>
<td>FAST kinase domain-containing protein 5</td>
</tr>
</tbody>
</table>
### Proteins showing no change after mtDNA depletion (100 proteins)

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAHD2A</td>
<td>fumarate-lyase domain-containing protein 2A</td>
</tr>
<tr>
<td>GLUD1</td>
<td>glutamate dehydrogenase 1, mitochondrial precursor</td>
</tr>
<tr>
<td>GLS</td>
<td>glutaminase kidney isoform, mitochondrial isoform 2</td>
</tr>
<tr>
<td>GARS</td>
<td>glycine--tRNA ligase precursor</td>
</tr>
<tr>
<td>GPRF1L</td>
<td>grpe protein homolog 1, mitochondrial precursor</td>
</tr>
<tr>
<td>GPRF1L2</td>
<td>grpe protein homolog 2, mitochondrial precursor</td>
</tr>
<tr>
<td>AK5</td>
<td>GTP-AMP phosphotransferase AK5, mitochondrial isoform a</td>
</tr>
<tr>
<td>HADH</td>
<td>hydroxyacetyl-coenzyme A dehydrogenase, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>HMGLC</td>
<td>hydroxymethylglutaryl-CoA lyase, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>PPA2</td>
<td>inorganic pyrophosphatase 2, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>ISCA1</td>
<td>iron-sulfur cluster assembly 1 homolog, mitochondrial precursor</td>
</tr>
<tr>
<td>IDH2</td>
<td>isocitrate dehydrogenase (NADP), mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>L2HGDH</td>
<td>L-2-hydroxyglutarate dehydrogenase, mitochondrial precursor</td>
</tr>
<tr>
<td>LTF</td>
<td>factortransferrin isoform 1 precursor</td>
</tr>
<tr>
<td>LETM1</td>
<td>LETM1 and EF-hand domain-containing protein 1, mitochondrial precursor</td>
</tr>
<tr>
<td>LYPAL1</td>
<td>lysophospholipase-like protein 1</td>
</tr>
<tr>
<td>LYZ</td>
<td>lysozyme C precursor</td>
</tr>
<tr>
<td>MDH2</td>
<td>malate dehydrogenase, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>MACAT</td>
<td>malonyl-CoA-acyl carrier protein transacylase, mitochondrial isoform a precursor</td>
</tr>
<tr>
<td>ACADM</td>
<td>medium-chain specific acyl-CoA dehydrogenase, mitochondrial isoform b precursor</td>
</tr>
<tr>
<td>MTFMT</td>
<td>methionyl-tRNA formyltransferase, mitochondrial</td>
</tr>
<tr>
<td>TIMM44</td>
<td>mitochondrial import inner membrane translocase subunit TIM44</td>
</tr>
<tr>
<td>ARHD10</td>
<td>mycophenolic acid acyl-glucuronide esterase, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>NADK2ADK1</td>
<td>NAD kinase 2, mitochondrial isoform 1</td>
</tr>
<tr>
<td>NDUFAF7</td>
<td>NADH dehydrogenase (ubiquinone) complex I, assembly factor 7 isoform 1</td>
</tr>
<tr>
<td>NDUFAF3</td>
<td>NADH dehydrogenase (ubiquinone) flavoprotein 3, mitochondrial isoform a precursor</td>
</tr>
<tr>
<td>MACROD1</td>
<td>O-acetyl-ADP-ribose deacetylase MACROD1</td>
</tr>
<tr>
<td>DAT</td>
<td>dihydrolipoyl dehydrogenase, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>PPAT</td>
<td>peptidyl-prolyl cis-trans isomerase F, mitochondrial precursor</td>
</tr>
<tr>
<td>PKG1</td>
<td>phosphokinase (GTP), mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>NPSR1</td>
<td>probable histidine--tRNA ligase, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>PSM1</td>
<td>proteasome subunit beta type-1</td>
</tr>
<tr>
<td>GBAS</td>
<td>protein Hip3p homolog 2 isoform 1</td>
</tr>
<tr>
<td>NIPSNAP3A</td>
<td>protein Hip3p homolog 3A</td>
</tr>
<tr>
<td>RCC2</td>
<td>protein RCC2</td>
</tr>
<tr>
<td>PYCR2</td>
<td>pyrroline-5-carboxylate reductase 2 isoform 1</td>
</tr>
<tr>
<td>PDHB</td>
<td>pyruvate dehydrogenase E1 component subunit beta, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>PDK2</td>
<td>pyruvate dehydrogenase kinase, isozyme 2 isoform 1 precursor</td>
</tr>
<tr>
<td>PKD3</td>
<td>pyruvate dehydrogenase kinase, isozyme 3 isoform 1 precursor</td>
</tr>
<tr>
<td>RRM1T1</td>
<td>RNA methytransferase-like protein 1</td>
</tr>
<tr>
<td>RPUSD3</td>
<td>RNA pseudouridylate synthase domain-containing protein 3 isoform 1</td>
</tr>
<tr>
<td>SMARCE1</td>
<td>small nuclear ribonucleoprotein E</td>
</tr>
<tr>
<td>SDHA</td>
<td>succinate dehydrogenase (ubiquinone) flavoprotein subunit, mitochondrial</td>
</tr>
<tr>
<td>ALDH5A1</td>
<td>succinate-semialdehyde dehydrogenase, mitochondrial isoform 1 precursor</td>
</tr>
<tr>
<td>SUC2G2</td>
<td>succinyl-CoA ligase [GDP-forming] subunit beta, mitochondrial isoform 2 precursor</td>
</tr>
<tr>
<td>TXN</td>
<td>thioredoxin isoform 1</td>
</tr>
<tr>
<td>PRDX3</td>
<td>thioredoxin-dependent peroxide reductase, mitochondrial isoform a precursor</td>
</tr>
<tr>
<td>MTF2</td>
<td>translation initiation factor IF-2, mitochondrial precursor</td>
</tr>
<tr>
<td>YARS2</td>
<td>tyrosine--tRNA ligase, mitochondrial precursor</td>
</tr>
<tr>
<td>ACO1L</td>
<td>very-long-chain specific acyl-CoA dehydrogenase, mitochondrial isoform 3</td>
</tr>
</tbody>
</table>

### Proteins showing 2 fold increase after mtDNA depletion (8 proteins)

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCK1</td>
<td>pyruvate dehydrogenase (acetyl-transferring) kinase isozyme 1, mitochondrial isoform 2 precursor</td>
</tr>
<tr>
<td>ATP5F1</td>
<td>ATP synthase F(0) complex subunit B1, mitochondrial precursor</td>
</tr>
<tr>
<td>CENPV</td>
<td>centromere protein V</td>
</tr>
<tr>
<td>CLU</td>
<td>clustatin precursor</td>
</tr>
<tr>
<td>ALDH1B1A1</td>
<td>aldehyde-1-pyrroline-5-carboxylate synthase isoform 1</td>
</tr>
<tr>
<td>MTERF2D</td>
<td>mTERF domain-containing protein 2</td>
</tr>
<tr>
<td>PIC1</td>
<td>pyrroline-5-carboxylate reductase 1, mitochondrial isoform 1</td>
</tr>
<tr>
<td>FUS</td>
<td>RNA-binding protein FUS isoform 1</td>
</tr>
</tbody>
</table>
6. Discussion

6.1 Nucleoid biology

Since the discovery of mtDNA in 1960’s, (see introduction) many researchers have tried to understand the organisation and maintenance of mtDNA. Nass suggested mtDNA might be attached to the inner mitochondrial membrane in 1969 (Nass 1969), and when Albring and co-workers in 1977 proposed that this membrane attachment was carried out by an unknown protein, this also led to the hypothesis that also mammalian nucleoids were organised in mtDNA:protein complexes referred to as nucleoids (Albring et al 1977). Yeast research has been at the forefront of much of the mitochondrial nucleoid work, showing, for example, that replicating nucleoids are tethered to the inner mitochondrial membrane (Meeusen & Nunnari 2003). Therefore, a functional conservation of yeast and mammalian nucleoids was suggested (Spelbrink et al 2010). However, both in yeast and mammals the protein(s) carrying out the nucleoid membrane connection remain(s) unresolved, although recent reports have identified candidate proteins (see introduction).

Several studies using proteomic, immunocytochemical and fractionation approaches have identified some of the components of the mitochondrial nucleoids, including TFAM, mtSSB (van Tuyle & Pavco 1985, Mignotte & Barat 1986 and Garrido et al 2003) and Twinkle (Spelbrink et al 2001 and Garrido et al 2003). In this part of the PhD work, we have set out to study the organisation of the known nucleoid proteins with the aim to elucidate their role in mtDNA organisation and maintenance. For example, we have investigated how nucleoids are organised inside mitochondria by performing immunocytochemistry (ICC) with antibodies against mtDNA, TFAM, mtSSB and Twinkle and visualised replicating mtDNA by EdU and/or BrdU labelling using confocal microscopy. In addition, we have performed several mitochondrial fractionation protocols to investigate the localisation of the known nucleoid associated proteins.
6.1.1 Nucleoids are heterogeneous in nature

The mitochondrial fractionation work shows that the different known nucleoid proteins and mtDNA are distributed differentially to insoluble membrane associated nucleoid (IMAN) and soluble matrix localised nucleoid (SMLN) fractions, showing that nucleoids are not a single entity with a static nature and demonstrating the existence of nucleoids with varying composition. This is further corroborated by immunocytochemistry (ICC) studies (see below). The nucleoid heterogeneity has been discussed earlier: for example, ICC research carried out in the Spelbrink laboratory identified POLG1 and POLG2 to have a different distribution within the mitochondria when compared to TFAM and mtSSB. TFAM and mtSSB co-localised as punctate foci with Twinkle whereas POLG2 staining showed more uniform appearance with only few intense foci that co-localised with Twinkle. The hypothesis was made that perhaps POLG2 only co-localises with replicating nucleoids. Due to antibody specificity issues the POLG1 co-localisation could not be addressed in that study (Garrido et al 2003). The ICC results of this present study show that not all nucleoids have the same composition. It seems that TFAM is the only constitutive member of the nucleoids (of the proteins studied), whereas Twinkle and mtSSB only localise to a subset of nucleoids. No other mtDNA packaging protein has been found to date and given the function of TFAM as the mtDNA packaging protein (Alam et al 2003), it seems reasonable that it would be always present in nucleoids. These experiments do not address the amount of TFAM/nucleoid and it is likely, based on previous research, that the amount of TFAM in nucleoids is dynamic (Seibel-Rogol & Shadel 2001).

Work by the Bogenhagen laboratory has identified nucleoid fractions with different protein composition when performing sedimentation assays, although their work did not include Twinkle localisation (Wang & Bogenhagen 2006), but nevertheless supporting the nucleoid heterogeneity. The heterogeneity of nucleoids has been discussed in the literature previously and the idea is that the composition of nucleoids changes depending on the needs for e.g. replication or transcription. For example in yeast the nucleoid composition has been shown to change according to metabolic changes (Kucej et al 2008).

Work in this thesis is the first research looking at the distribution of several nucleoid proteins at the same time by mitochondrial fractionation and is the first
research undertaken to investigate the behaviour of the endogenous Twinkle protein. Results show that mtDNA/protein complexes could exist as firmly membrane bound or as either loosely membrane associated or non-membrane bound in mammalian mitochondria, contrary to the current paradigm. This finding is in line with the work of Brown et al (2011) who found that not all nucleoids have close contact with the inner membrane. Next, the organisational role of Twinkle was further elucidated.

6.1.2 Twinkle shows characteristics of a mitochondrial inner membrane protein

Twinkle and ATPase family AAA domain-containing protein 3 (ATAD3) were the only proteins in our study solely found in the IMAN fractions. ATAD3 is a known mitochondrial inner membrane protein (Da Cruz et al 2003, He et al 2007 and Gilquin et al 2010) and the localisation of ATAD3A and B to the IMAN fraction was expected. ATAD3 has a predictable trans-membrane domain and the N’terminal domain has contact sites to the mitochondrial outer membrane (Gilquin et al 2010). However, the appearance of Twinkle in the IMAN fractions was somewhat surprising, because it does not contain any transmembrane domains predicting that it would be an inner membrane associated protein. To investigate the Twinkle membrane association further we set out to use different mitochondrial fractionation protocols.

By using two different fractionation protocols (sodium carbonate and high salt), the data clearly shows that Twinkle is tightly associated with the mitochondrial inner membrane. In all fractionation experiments Twinkle always remained in the membrane fraction. These experiments were carried out both with over-expressed and endogenous Twinkle. As Twinkle has been shown to purify with mtDNA (Wang & Bogenhagen 2003) we wanted to confirm that Twinkle membrane association was not due to RNA and/or DNA interaction. Therefore, fractionations were also carried out in the presence of nucleases. Destroying the RNA and/or DNA did not release Twinkle to the soluble fraction showing that the membrane interaction is independent of nucleic acid. Moreover, Twinkle in contrast to TFAM and mtSSB remains punctate in cells lacking mtDNA, demonstrating the independence of membrane attachment from DNA. The release of TFAM to the soluble fraction after nuclease treatment shows that the presence of TFAM in the
IMAN fraction is partly dependent of DNA or RNA. Furthermore, a floatation assay show that Twinkle migrates in the same fraction as POLG1, mtSSB, TFAM and ATAD3 together with mtDNA demonstrating that these proteins exist as a complex and do not interact due to the experimental condition present in the fractionation assays.

There are three classes of membrane proteins, integral membrane proteins that associate to the lipid bilayer permanently, peripheral membrane proteins that only associate temporarily to the lipid bilayer or with integral membrane proteins. The third class includes lipid-anchored proteins that are bound to the bilayer through lipidated amino acid residues, contact sites of protein alpha-helix or loop regions and by electrostatic interactions, but they do not span the membrane (Blobel 1989). Temporary attachment of the peripheral membrane proteins are via electrostatic, hydrophobic and other non-covalent interactions, which are disturbed by high pH or high salt concentrations. Twinkle did not solubilise with changes in pH or salt and hence is not temporarily attached to the membrane and behaves more like a permanent membrane protein. Twinkle does not possess any predicted transmembrane domains, but we show that it is firmly membrane attached and it can be hypothesised that it behaves more like an integral monotopic protein. Moreover, Garrido et al (2003) showed GFP-tagged Twinkle to have the same velocity and directionality as overall mitochondrial movement, which also substantiates the role of Twinkle as a membrane anchor. The unpublished data of this thesis reveals that a loop region of Twinkle that consists of amino acids 612-644 might be partly responsible for the membrane anchorage, but only a portion of Twinkle missing this loop-region shifts to the soluble fraction, so clearly there are some other factors involved in this process. The cells used for the studies are still expressing the endogenous Twinkle protein and one possibility that cannot be discarded is that the endogenous protein is holding some of the Twinkle structures at the inner membrane. This could lead to only partial release of the protein to the soluble fraction. Although, it is noteworthy that endogenous Twinkle is a low abundant protein. We did not study the importance of Twinkle loop region in the membrane association further in this thesis project. To eliminate the involvement of endogenous Twinkle holding the nucleoids at the inner membrane in TwinkleLoop expressing cells, as discussed above, we could have silenced the endogenous Twinkle specifically.
Research undertaken with over-expressed Twinkle shows good co-localisation with Twinkle and mtDNA (Spelbrink et al 2001 and Garrido et al 2003), but in contrast our study with endogenous Twinkle demonstrates that only a subset of nucleoids contain Twinkle or mtSSB. When mitochondria from Twinkle over-expressing cells were analysed for protein distribution in soluble and insoluble fractions, results show that increasing Twinkle concentration shifts the distribution of TFAM, POLG1 and mtDNA to the Twinkle containing membrane fraction. Similarly the absence of Twinkle release mtDNA and mtSSB to the more soluble fraction. Because Twinkle does not seem to be a constitutive member of nucleoids and is needed for mtDNA replication, nucleoids must be dynamic in nature for replication to take place. Moreover, in this work, we have demonstrated dependency of mtDNA membrane-association on Twinkle. This gives stronger evidence for the idea that there might also be a specific organisational centre for replicating nucleoids, as in yeast. Next, the organisational role of Twinkle was further elucidated.

6.1.3 Organisation of nucleoids at the mitochondrial inner membrane is dynamic and replication dependent

Studies in yeast have suggested that replicating nucleoids are organized at the inner membrane (Meeusen & Nunnari 2003). Mammalian nucleoid work in Bogenhagen’s laboratory has identified slow and fast sedimenting nucleoids during gradient isolation, which he suggested might represent replicating and non-replicating nucleoids. However, the two populations of mtDNA incorporated similar amounts of thymidine label suggesting the differences between them were unrelated to replication (Wang & Bogenhagen 2006). To further study the possibility that replicating mammalian nucleoids might also organise at the inner membrane, we used Edu and BrdU labelling and in situ detection to distinguish replicating and non-replicating mtDNAs.

Helicases are needed for the unwinding of dsDNA during replication and Twinkle alone can unwind short stretches of dsDNA, although its activity is stimulated by mtSSB (Korhonen et al 2003). Moreover, in vitro Twinkle has been shown to form a “replisome” together with mtSSB and POLG1 (Korhonen et al 2004). Edu and BrDu labelling results of this current work shows that Twinkle and
mtSSB are associated with replicating mtDNA molecules. At any given time point only a subset of nucleoids are replicating and results show that for replication to take place mtDNA needs to be associated with Twinkle. In light of our findings on Twinkle membrane association, we propose that mammalian nucleoids are re-organised to form replication centres at the inner mitochondrial membrane, as in fungi. Furthermore, our findings support the dynamic nature of the nucleoids. For example, the presence of mtSSB in nucleoids is Twinkle dependent and it is dynamically recruited to replicating nucleoids. When Twinkle is knocked down and replication cannot take place, the mtSSB foci disappear in ICC and in fractionation studies down regulation of Twinkle shifts the mtSSB to the more soluble fraction. In ddC treated cells, where replication can initiate, but is stalled due to the ddC at some point, mtSSB foci still remain. This makes sense, as mtSSB is needed to stabilise the newly synthesised DNA. Above data is also supported by the finding that when Twinkle replication stalling mutants are expressed in cells mtSSB staining is more uniform compared to wild type cells (Wanrooij et al 2007).

6.1.4 Conservation of mechanistic character of nucleoids

Yeast research has demonstrated that replicating nucleoids are connected to the inner membrane and are faithfully inherited during cell division (Meeusen & Nunnari 2003). It has also been shown that mammalian nucleoids are in close contact with fission protein Drp1 (Ban-Ishihara 2013) and Drp1 is localised at the sites of mitochondria-ER connections (Friedman et al 2011), suggesting mammalian nucleoids have connections to the cytoplasmic sites. Evidence for higher order structure in mammals has been hypothesised earlier by Iborra et al (2004) as a component of a kinesin motor, KIF5B, co-localises often with mtDNA foci. Studies on fungi nucleoids show that the inner membrane localisation of yeast mtDNA polymerase, Mip1, is mtDNA independent and replicating nucleoids appeared dynamic in nature. Same research also showed that, similar to our findings, some of the mtDNA maintenance proteins remained as discrete complexes in the absence of mtDNA (Meeusen & Nunnari 2003). We show that in mammalian Rho0 cells that lack mtDNA, Twinkle retains a punctate pattern, whereas mtSSB is more widely dispersed than in cells containing mtDNA. Yeast does not have a Twinkle
homologue (Shutt & Gray 2010), but it has been shown that yeast mtDNA helicase, Pif1, is membrane associated and the yeast TFAM homologue, Abf2, is partially dissociated by nuclease treatment (Cheng & Ivessa 2010), as is TFAM in this present study.

It could be argued that mtDNA is anchored to the membrane by some other means than Twinkle, which is then recruited to the mtDNA/nucleoid complex when replication takes place. Few proteins have been shown to have a role in nucleoid membrane attachment. For example, prohibitin (PHB) and ATAD3 have been isolated with nucleoids and been postulated to have an architectural role in nucleoids (Wang et al 2006, He et al 2007 and He et al 2012a). He and co-workers showed ATAD3 and PHB to co-sediment and co-purify with nucleoids and the mitochondrial translation machinery, postulating that ATAD3 links mitochondrial ribosomes to nucleoids and that both PHB and ATAD3 link nucleoids to the inner mitochondrial membrane (He et al 2012a). Combined, the data from ATAD3 and prohibitin (He et al 2012 a) suggest that the nucleoid membrane connection is dependent on more than one protein. We can hypothesise that nucleoids have different composition depending on the particular need of mtDNA at any given point (figure 6.1). For instance, DNA2, which is involved in DNA replication and repair, only associates with nucleoids when replication is stalled (Duxin et al 2009). Although in the MS work carried out for this thesis work, to identify nucleoid associated proteins (NAPs), PHB1 and 2 did not pass our selection criteria, because they were also identified in TwinkleFLAG IPs without XL and were not sufficiently enriched in TwinkleFLAG +XL compared to mtLucFLAG +XL. In addition, their levels remained equal in TwinkleFLAG +XL IAP in ρ° cells compared to mtDNA containing cells. ATAD3 also just failed to pass our selection criteria as it showed a <2 fold (1.93) increase comparing TwinkleFLAG +XL and TwinkleFLAG without XL. These results thus maintain the notion that these proteins could be part of a membrane anchor for a minimal mtDNA replication platform that includes Twinkle.
Figure 6.1: Schematic representation of possible nucleoid structures that could exist in the mammalian mitochondria. Nucleoids dynamically interact with Twinkle at the mitochondrial inner membrane when there is a need for replication. MtSSB is also found only in the replicating nucleoids, but in contrast TFAM is a constitutive component of the nucleoids. Other forms of nucleoids are hypothesised to be either membrane associated or matrix located and the shift between the different forms of nucleoids could be based on the metabolic need of the mitochondria at that moment. Membrane association could be via ATAD3 or prohibitin. It has been postulated that ATAD3 also links mitoribosomes to the nucleoids (OM= mitochondrial outer membrane, IM=mitochondrial inner membrane).
The specific organisation of mtDNA in addition to yeast and mammals is seen in other organisms, such as *Trypanosoma brucei* where mtDNA is organised in kinetoplasts at the inner membrane and have connections to the basal body. Bacterial DNA is also organised in nucleoid structures (Lemon & Grossman 1998). Perhaps, the mammalian mtDNA membrane association is reminiscent of its bacterial ancestor and is an evolutionarily conserved mechanism to ensure proper DNA organisation. In this current PhD work, we did not set out to study all the known nucleoid proteins, but we can hypothesise that in mammals a similar organisation has remained to ensure organised replication and inheritance of nucleoids and that the membrane connection is dependent on multiple proteins.

Bogenhagen et al (2008) hypothesised nucleoids to have a layered structure with a catalytic core that includes transcription and replication machineries surrounded by proteins that may have a more architectural role based on proteins found in nucleoid purification experiment, but the method used in that study could not distinguish between different sets of nucleoids. I present here data that agree with the idea of nucleoids having a core that is the synthetic centre for mtDNA maintenance. This does not discard the layered nucleoid theory, but extends the idea of layered nucleoids to different pools of nucleoids or nucleoid compartment (discussed later).

### 6.2 Mitochondrial nucleoid protein identification strategies brought to the 21st century

Considering the number of studies performed on the mitochondrial nucleoid, it remains something of a conundrum. The number of proteins purified with nucleoids is large, but the results vary greatly between different purification strategies and only a few proteins have been further characterised for their role in mtDNA maintenance. Moreover, many strategies did not detect core components like Twinkle and mtSSB (Bogenhagen et al 2003, Cheng et al 2005, He et al 2007, Reyes et al 2011 and He et al 2012a, see review of the literature table 5.2). Problems lie in the usage of different starting materials; purification strategies; buffers used and mass spectrometry methods and the target at which isolation was directed. Also, one important factor is that many of the protein-nucleoid interactions are transient in
nature. The first aim of nucleoid isolation project was to establish a faster, easier, qualitative and quantitative nucleoid purification strategy. Previous nucleoid research has relied on large amounts of starting material and on complicated purification strategies that have always included isolation of mitochondria as a first step, which in itself is labour intensive. Here we have employed a strategy where formaldehyde cross-linking is carried out in intact cells hence avoiding the mitochondrial isolation step.

Sutherland and co-workers (2008) discussed the unsuitability of FLAG tag for formaldehyde cross-linking because it contains several lysine residues. Lysine is one of the amino acids that has been reported to be highly reactive with formaldehyde. However, the short formaldehyde incubation used in this study did not interfere with the efficiency of nucleoid complex isolation, unlike other tags that were tested including the myc- and his-tag (unpublished data). These results demonstrate that the developed approach can have a much wider application in the analysis of mitochondrial protein complexes.

In nucleoid purification strategies, it is important to consider the correct buffer conditions in mitochondrial lysis as well as during subsequent nucleoid purification. For example, Bogenhagen et al (2003) found triton X-100 to be essential during glycerol gradient to exclude additional membrane proteins. We have chosen to use formaldehyde cross-linking, so that we can use stringent buffer conditions during immunoprecipitation. The approach used in this thesis research also shows that by a comparison of lysis conditions and sample handling (TX100 or XChIP), the XChIP method was the most sensitive and inclusive, and, despite the fact that many more proteins were ‘nucleoid’ enriched compared to the TX100 method, the XChIP method nevertheless showed enrichment of the highest percentage of mitochondrial proteins suggesting the method is considerably more stringent than the TX100 method. NAPs that only were identified using the XChIP method include, MGME1, DDX28, MTERF and MTERF2, Topoisomerase 3α, POLG2, TFB2M as well as 50 mito ribosomal proteins.

The data presented here provide an important resource for the discovery not only of potential mtDNA maintenance factors but also for factors involved in transcription, RNA metabolism and protein synthesis. For example amongst the proteins identified, we can find proteins with yet unassigned roles in mitochondrial gene expression, including RNA metabolism and translation.
In the present work cross linked Twinkle-FLAG isolated from cells lacking mtDNA reveals the absence of some of the proteins enriched in the presence of mtDNA and therefore identifies important candidate proteins. For example, four FAST kinase domain-containing proteins were severely reduced or absent. It is noteworthy that these proteins were also found in a recent RNA-binding proteome (Baltz et al 2012). One of the novel proteins identified as nucleoid associated in this same fraction is the methylcrotonoyl-CoA carboxylase (MCC), involved in leucine metabolism. The finding is novel and interesting as another enzyme, enoyl-coenzyme A (CoA) hydratase involved in leucine metabolism has been shown to have a role in mitochondrial protein synthesis (Richman et al 2014). MCC was identified as a NAP from Twinkle FLAG cells, but was not decreased in NAP fractions from cells lacking mtDNA. It may directly interact with Twinkle, but we did not study this further in the present work.

It is also crucial to apply a systematic analysis, comparing several conditions of isolation, using various controls such as a tagged and mitochondrially targeted control protein (in our case Luciferase), measuring both biological and technical repeats and applying stringent selection criteria. However, using the methods chosen we could not pin point a strict list of for example only DNA binding proteins. For instance formaldehyde is not specific and is more efficient in cross-linking protein to protein than protein to DNA. In addition, as shown in this work and given the sensitivity of modern mass spectrometry we can not say that the proteins identified could be considered as all being NAPs until further investigation. However, the approach taken in this study shows that this method can be applied in NAP research and identifies many of the already known nucleoid proteins, such as POLG1, mtSSB and TFAM.

6.2.1 The mitochondrial nucleoid “compartment”

Based on the above, the question arises - can we look at the nucleoids as a tight structure that would have a limited number of proteins associated with it? Can we even expect a consensus list of NAPs if we think how dynamic these complexes are? The evidence suggests that nucleoids cannot be defined as static structures with a fixed composition and that while TFAM is always a nucleoid associated protein
Twinkle and mtSSB are present only with replicating nucleoids. Moreover, if we compare the earlier work of Bogenhagen et al (2008) and Wang & Bogenhagen (2006) to later work from Reyes et al (2011) and He et al (2012a) – we can see that depending on the purification strategy the amount and identity of proteins varies greatly. Here we show that it also depends on whether the mtDNA is replicating or not. If we consider this, which my results discussed above clearly show, during purification any nucleoid can be carrying out any given function at that specific time point.

Instead of thinking nucleoids as small structures containing e.g. just the replication machinery, we can think of them as larger structures that are associated with the proteins involved in for example mitochondrial translation. As suggested by Iborra et al (2004), the translational machinery could well be in close vicinity of nucleoids. Indeed, ribosomal proteins have been found to associate with nucleoids (He et al 2012a). Many of the proteins found to isolate with Twinkle in this present study also fall in the category of RNA metabolism or translation, further corroborating the idea of a larger nucleoid compartment idea. Furthermore, these compartments have NAPs that only transiently associate with membrane associated nucleoids and other NAPs that are permanently membrane associated.

6.2.2 Close liaison between nucleoids and ribosomes

Previous research has pointed towards the close liaison of nucleoids with the translational machinery (see introduction) and the results of this thesis gives more evidence for the close relationship of nucleoids and the mitochondrial protein synthesis apparatus. For example, some of the proteins that are involved in ribosome biogenesis and were suggested by Bogenhagen to interact with nucleoids, were less than 2-fold decreased when comparing TwinkleFLAG IAP enriched proteins from mtDNA containing cells to ρ° protein profiles in this present study (e.g. small ribosomal subunits S7, S9 and S15), suggesting they would be part of the core nucleoid structure that is maintained in the absence of mtDNA and RNA. An organisational role of Twinkle, independent of mtDNA, is further corroborated with the mass spectrometry studies where protein identification in cells lacking mtDNA shows differences in several proteins when compared to mtDNA containing cells.
For example, the levels of TFAM and mtSSB are reduced and several proteins completely disappear in mtDNA-less cells including POLG2, MGME1 and 10 mitochondrial ribosomal proteins. These results agree with those of Iborra et al (2004) and He et al (2012a) strongly supporting the organisation of mitochondrial translational machinery in the vicinity of Twinkle replication organisational centre. Further evidence for the close contact of nucleoids with the translational machinery comes from the mitochondrial floatation assay that shows the small ribosomal protein (MRPS22) partially co-fractionate with Twinkle, TFAM, POLG1 and mtSSB in the gradient, indicating that these proteins exist as a complex or are retained in a specialized nucleoid ‘compartment’.

6.3 Twinkle purification and structural studies

During DNA replication, helicases are needed to unwind the dsDNA ahead of the DNA polymerase to allow replication to take place (Matson & Kaiser-Rogers 1990). For mtDNA maintenance, the correct function of Twinkle helicase is essential for the accurate copying of mtDNA and hence the subsequent health of the cell. Several Twinkle mutations are lined to autosomal dominant progressive external ophthalmoplegia (adPEO) (Spelbrink et al 2001), which affect either the oligomerisation and/or helicase function (Goffart et al 2009, Longley et al 2010 and Ziebarth et al 2010). In adPEO patients, mtDNA deletions occur in post mitotic tissue (Suomalainen et al 1997). Mice expressing a disease-related Twinkle mutation accumulate mtDNA deletions in skeletal muscle, which led the authors to conclude that impaired Twinkle function must be the underlying cause for the deletions (Tyynismaa et al 2005). In vitro studies showed that most severe Twinkle adPEO mutants could cause replication stalling, which was speculated to be the molecular mechanism for the formation of mtDNA deletions (Wanrooij et al 2007 and Goffart et al 2009). Twinkle structural information would be beneficial for the understanding of the mechanistic function of the protein and hence would allow us to understand the malfunctioning of Twinkle disease mutants. There is no molecular structure at atomic resolution for Twinkle DNA helicase. The problems arise from the purification of the protein in adequate quantity and purity. In this thesis work, I was able to purify Twinkle for electron microscopy (EM) studies, however the
quantity of the purified protein did now allow crystallographic studies. The negative staining EM pictures revealed that Twinkle exists as heptamers and hexamers, which was also verified by the later EM pictures. These results are in agreement with previous work where Twinkle has been shown to form hexamers and heptamers (Ziebarth et al 2010). In addition, the close relative of Twinkle, bacteriophage T7 gp4 helicase, is also found in both heptameric and hexameric structures (Crampton et al 2006). In T7 gp4 helicase two hypothesis for the heptameric forms have been put forward. In one model, the heptamer loses one subunit when in contact with ssDNA and this allows the loading of the helicase on DNA (Kato et al 2003). In the other model Toth et al (2003) discusses the role of heptameric structure in dsDNA binding, suggesting that heptamers can translocate along dsDNA whereas hexamers can only translocate along ssDNA. Although, later studies have shown that heptameric form does not bind DNA efficiently and the hexamers are the predominant form of helicases (Crampton et al 2006). The SAXS and EM studies in this present work show a wide central channel in Twinkle heptamers and previous work has shown Twinkle to be able to load on ssDNA and dsDNA (Farge et al 2008) supporting either model suggested for the T7 gp4 protein.

6.3.1 Twinkle double hexamer

The double hexamer structure found in the first EM studies was never observed in later EM work. This can be due to conditions, such as salt, detergent, during purification, storage and/or EM conditions. Twinkle double hexamer has not been found in EM studies of other researchers (Ziebarth et al 2010). Thus, it is likely that the double hexameric structure is due to an artefact. Twinkle, being a DNA binding protein, is sensitive for purifying conditions such as NaCl concentration (Ziebarth et al 2010) and hence the structure may be easily disrupted or the protein may aggregate. The latter is indeed frequently observed in Twinkle purification and storage.
7. Summary and conclusions

Almost 40 years after the first indications of mtDNA membrane association, the work of this thesis points to Twinkle as a membrane anchored replication factor that specifically links replicating nucleoids to the mitochondrial inner membrane. The evidence for Twinkle as a membrane-anchored protein and as the core of replicating nucleoids represents an important advance to knowledge. The work described in this thesis also illustrates the dynamic nature of nucleoids, suggesting that different forms of nucleoids exist to carry out specific functions depending on the requirements of the cells. Therefore, we have confirmed that TFAM and Twinkle are nucleoid components; however, we have shown that TFAM is the only stable component of nucleoids, whereas mtSSB and Twinkle are dynamic partners present in replicating nucleoids. The structural studies of Twinkle are important in understanding the mechanisms of Twinkle function and hence giving valuable information on malfunctioning of the disease mutants of the protein.

Despite the many mass spectrometry based mammalian nucleoid studies, the composition of mammalian nucleoids still remained an enigma and this is not due to the lack of research directed towards nucleoids in the 21st century, but more because very few systematic and quantitative studies have been undertaken. The work of this thesis demonstrate the feasibility of using whole cell formaldehyde cross-linking in combination with immuno-affinity purification in mitochondrial nucleoid identification based on mass spectrometry. Moreover, for the first time different isolation conditions and mass spectrometry approaches have been compared and analysed in search of NAPs, giving valuable information on the experimental conditions that should be considered during NAP identification research. The approach of using whole cell culture as starting material is novel in NAP based research and it can be a useful tool to identify new causes of mitochondrial disease. In addition, previous nucleoid research has concentrated on non-quantitative mass spectrometry, whereas in my research we have used a quantitative approach with the aim to identify a list of proteins that we can refer to as NAPs. We have identified
several interesting novel proteins, although it was not within the scope of this study to test their role in nucleoid organization. However, the list generated represents a good reference for future investigation.

The proteins expressed by mtDNA are vital for life - this highlights the importance of understanding the organisation and maintenance of mtDNA and nucleoids. The work carried out here has “probed the mammalian mitochondrial nucleoid” and given new information on the organisation of mtDNA and its maintenance machinery, which will be helpful in understanding the mammalian mtDNA maintenance in health and in disease.
Acknowledgements

The work presented in this thesis was carried out at the laboratory of molecular biology in the Docent Hans Spelbrink’s mtDNA maintenance group at the institute of BioMediTech, University of Tampere. The research was funded by the Doctoral Programme in Biomedicine and Biotechnology (TGPBB).

I would like to thank Prof Hans Spelbrink for providing the PhD project and for the help he has provided over the years. I will forever be thankful to Prof Markku Kulomaa who I have met first time when applying for MSc studies in molecular biology at the University of Jyväskylä. I am grateful for him that he accepted me as a student when I applied with a BSc (Hons) degree from London University, and he has been a good mentor ever since. Few years later, we met in Tampere University where he supported me as one of the thesis committee members during my PhD. He has been helpful in providing advice many times during my PhD and been a great support throughout the process. I would also like to thank Prof Ian Holt for being one of the thesis committee members, who has provided good advice and good scientific conversations during this process. The official reviewers of this thesis, Dr Antonella Spinazzola and Docent Henna Tyynismaa are both gratefully acknowledged for their valuable insights and comments on this thesis.

I would like to thank the co-authors for their hard work they have carried out. I wish to thank all past and present people in Hans, Howy, Alberto and Kaguni lab for their friendship and scientific discussions and advice. Just to mention few: Eric Dufour has been a great person for scientific discussions and advice. Special thanks to Suvi Vartiainen for nice conversation during lunch breaks and helping me with my “Finglish” at times of need for the Finnish abstracts. It has been great to share the lab space with Giuseppe Cannino. Alberto Sanz has always been helpful on providing scientific help and advice. Especially I would like to mention Peter Martinsson, the former Hans lab member, for good scientific conversations and for being a good friend and a great team member in Rogaining (orienteering)
competitions. I also want to thank all my friends and colleagues at BioMediTech and in the FinMIT groups in Helsinki during these years.

I am very grateful for Merja Jokela for her absolutely magnificent help in the lab. She has grown me countless numbers of cells, mixed hundreds of buffers and more. I could not have asked for a better person to help me. Her knowledge and hardworking attitude are admirable. Genuine thanks are due to Outi Kurronen, who with Merja, keeps the lab going and makes our lives so much easier by keeping everything running smoothly. I would also like to thank both Merja and Outi for being there for me and listening and giving advice in times of need.

Thank you for my friends for keeping me “sane” during this process. The rock climbing trips, hikes, camping trips, great conversations and so on – I have truly enjoyed them. Special thanks to my dear friend Beth Foreman, who I met while working at the University of Liverpool. She has always been a great support and I could not have asked for a better Godmother for my daughter.

I would also like to mention my current group leader Prof Anne Kallioniemi, with whom I have worked as a researcher while finalising my PhD. I would like to thank her for providing me with an interesting job, while finalising my thesis. I have really enjoyed my time in her laboratory. From my new lab, I would also like to thank Dr Emma-Leena Alarmo for her help with some Finnish language issues and her help with other issues relating to constructing and submitting the thesis work. Thank you also for the other Anne lab members for their support.

I especially thank my mom Sirkka who has helped me through this process in many ways. She was always ready to provide babysitting when there was need for late nights in the lab or travel for conferences. She was there in good times and bad; she has picked me up on low times and pushed me forward.

And last but not least, my little girl Arles. She is the power that has kept me going. Her positive attitude towards life and her happy joyful personality keeps me smiling and keeps me going. I love you little one.

Suuri kiitos äidilleni – ilman häntä tämä väitöskirja ei olisi ollut mahdollinen.
References

Akbari M, Visnes T, Krokan HE and Otterlei M (2008), Mitochondrial base excision repair of uracil and AP sites takes place by single-nucleotide insertion and long-patch DNA synthesis, DNA Repair (Amst) 7(4):605-16.


Bogenhagen DF and Clayton DA (1977), Mouse L cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle, Cell, 11:719-727.


Carrodeguas JA, Kobayashi R, Lim SE, Copeland WC and Bogenhagen DF (1999), The accessory subunit of Xenopus laevis mitochondrial DNA polymerase gamma increases processivity of the catalytic subunit of human DNA polymerase gamma and is related to class II aminoacyl-tRNA synthetases, Mol Cell Biol. 19(6):4039-46.


Chen XJ and Butow RA (2005), The organization and inheritance of the mitochondrial genome, Nat Rev Genet. 6(11):815-25.


Gredilla R (2010), DNA damage and base excision repair in mitochondria and their role in aging, J Aging Res.


Hyvärinen AK, Kumanto MK, Marjavaara SK and Jacobs HT (2010), Effects on mitochondrial transcription of manipulating mTERF protein levels in cultured human HEK293 cells, BMC Mol Biol. 16;11:72.


Jackson V. (1978), Studies on histone organization in the nucleosome using formaldehyde as a reversible cross-linking agent, Cell. 15(3):945-54.


Kaufman BA, Durisic N, Mativetsky JM, Costantino S, Hancock MA, Grutter P and Shoubridge EA (2007), The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures, Mol Biol Cell. 18(9):3225-36.


Lee YS, Lee S., Demeler B., Molineux IJ, Johnson KA and Yin Y.W (2010), Each monomer of the dimeric accessory protein for human mitochondrial DNA
polymerase has a distinct role in conferring processivity. J. Biol. Chem. 285:1490-1499.


Martin S.A (2011), Mitochondrial DNA Repair, DNA Repair - On the Pathways to Fixing DNA Damage and Errors, Dr. Francesca Storici (Ed.)


Pavco PA and Van Tuyle GC (1985), Purification and general properties of the DNA-binding protein (P16) from rat liver mitochondria, J Cell Biol. 100(1):258-64.


Schon EA, Bonilla E and DiMauro S (1997), Mitochondrial DNA mutations and pathogenesis, J Bioenerg Biomembr. 29(2):131-49. Review


Sewer MB and Li D (2008), Regulation of steroid hormone biosynthesis by the cytoskeleton, Lipids 43(12):1109-15. Review.


Tsurumi T and Lehman IR (1990), Release of RNA polymerase from vero cell mitochondria after herpes simplex virus type 1 infection, J Virol 64:450-452.

for mtDNA maintenance and regulates mtDNA copy number, Hum Mol Genet.

van Bruggen EF, Borst P, Ruttenberg GJ, Gruber M and Kroon AM (1966), Circular

Van Dyck E, Foury F, Stillman B and Brill SJ (1992), A single-stranded DNA binding
protein required for mitochondrial DNA replication in S. cerevisiae is homologous
to E. coli SSB, EMBO J. 1992 Sep;11(9):3421-30.

Mutation of POLG is associated with progressive external ophthalmoplegia
characterized by mtDNA deletions, Nat Genet. 28(3):211-2.

Van Hove JL, Cunningham V, Rice C, Ringel SP, Zhang Q, Chou PC, Truong CK and
Wong LJ. (2009), Finding twinkle in the eyes of a 71-year-old lady: a case report
and review of the genotypic and phenotypic spectrum of TWINKLE-related

Van Tuyle GC and Pavco PA (1985), The rat liver mitochondrial DNA-protein complex:
displaced single strands of replicative intermediates are protein coated, J Cell

Vasilescu J, Guo X and Kast J (2004), Identification of protein-protein interactions using in
vivo cross-linking and mass spectrometry, Proteomics.4(12):3845-54.

Berlin, (1927), Nachrichten der Gesellschaft der Wissenschaften zu Göttingen,

Von Wurmb-Schwark N (2007), Mitochondrial Mutagenesis in the Brain in forensic and
pathological Diseases, In Oxidative Stress and Neurodegenerative Diseases, p59-82,
Elsevier.

Vögtle FN, Wortelkamp S, Zahedi RP, Becker D, Leidhold C, Gevaert K, Kellermann J,
Voos W, Sickmann A, Pfanner N and Meisinger C (2009), Global analysis of the
mitochondrial N-proteome identifies a processing peptidase critical for protein

Walberg MW and Clayton DA (1983), In vitro transcription of human mitochondrial DNA.
Identification of specific light strand transcripts from the displacement loop region,

Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza AM, Elsas LJ 2nd and
Nikoskelainen EK (1988) Mitochondrial DNA mutation associated with Leber's

Review.

Wallace DC (2010), Mitochondrial DNA mutations in disease and aging, Environ Mol

Wallace DC and Chalkia D (2013), Mitochondrial DNA genetics and the heteroplasmic
conundrum in evolution and disease, Cold Spring Harb Perspect Med.
3(10):a021220. Review.

Wang Y and Bogenhagen DF (2006), Human mitochondrial DNA nucleoids are linked to protein folding machinery and metabolic enzymes at the mitochondrial inner membrane, J Biol Chem. 281(35):25791-802.


Appendices

Thesis supplementary figures
Legends to supplementary figures

Figure 1
To test the solubilisation of different nucleoid proteins mitochondria were subjected to different digitonin concentrations followed by centrifugation. Between 0.5:1 and 1:1 ratio of digitonin/mitochondrial protein (w/w) the inner mitochondrial membrane becomes leaky as the mitochondrial matrix protein glutamate dehydrogenase (GDH) is released to the supernatant (s) fraction. Whereas COXII is still found in the pellet (p) fraction. Both TFAM and POLG1 show a considerable amount of protein present in the s fraction and endogenous Twinkle behaved essentially identical to COXII.

Figure 3
Twinkle monoclonal antibody is specific for Twinkle protein. Twinkle monoclonal antibody detects all overexpressed Twinkle-myc hence demonstrating that Twinkle monoclonal antibody is specific. Twinkle-myc was expressed in U2OS cells via transient transfection and endogenous and overexpressed Twinkle detected using the Twinkle monoclonal antibody while Twinkle-myc was specifically detected using myc-antibody.

Figure 4
TFAM staining in 206 f ρ° cell line compared to the parental 143B cell line. 206f and 143B cells were fixed and stained for Twinkle, TFAM and DNA. Images for both cell lines were acquired and processed using identical setting to illustrate the difference in TFAM signal between control 143B and p cells.
Figure 5
mtSSB foci co-localisation with mtDNA in U2Os cells treated with ddC. U2OS cells were treated for 48 h with 100 µM ddC (panel B and control cells without ddC in panel A) and stained with Twinkle, mtSSB and DNA. Results show that ddC treatment results in a dramatic decrease of mtDNA staining (A1 and B1 larger panels and boxed area). In ddC treated cells many of the mtDNA foci left show intense mtSSB staining when compared to the surrounding mitochondrial network. Quantification of 4 images in both control and ddC treated cells (rightmost panel C) shows that the percentage of mtDNA foci with intense mtSSB staining is higher than in control cells although also there is quite some variability (range 24-47%) compared to control (range 10-22%). C) The quantification of fluorescence however shows that in contrast to the increased percentage, the mtSSB foci are less intense.

Figure 6
Sample images for 30 min (upper panel) and 90 min (lower panel) EdU labelling time points are shown. For each panel the individual mtDNA, EdU and Twinkle images are shown, as well as the EdU-mtDNA and EdU-Twinkle merged images. EdU-mtDNA co-localisation is indicated with pink dots, EdU-Twinkle co-localisation of EdU-mtDNA positive foci is indicated with yellow dots while EdU-only in the EdU-Twinkle merged image is indicated with red dots. The counted numbers for non-co-localisation are also given.

Figure 7
Fibroblasts were labelled for the indicated times with BrdU and processed for BrdU, mtSSB and mtDNA detection. BrdU foci positive for mtDNA are indicated with a red x while distinctly mtSSB-positive BrdU foci are indicated with a yellow x.
Figure 1
Figure 5
Figure 6

Figure 7
Original communications
Replication factors transiently associate with mtDNA at the mitochondrial inner membrane to facilitate replication

Nina Rajala¹, Joachim M. Gerhold², Peter Martinsson¹, Alexey Klymov² and Johannes N. Spelbrink¹,²,*

¹FinMIT Centre of Excellence, Institute of Biomedical Technology & Tampere University Hospital, Pirkanmaa Hospital District, FI-33014 University of Tampere, Finland and ²Department of Pediatrics, Institute for Genetic and Metabolic Disease, Radboud University Medical Centre, Geert Grooteplein 10, P.O. Box 9101, 6500 HB, Nijmegen, The Netherlands

Received June 6, 2013; Revised September 30, 2013; Accepted October 2, 2013

ABSTRACT
Mitochondrial DNA (mtDNA) is organized in discrete protein–DNA complexes, nucleoids, that are usually considered to be mitochondrial-inner-membrane associated. Here we addressed the association of replication factors with nucleoids and show that endogenous mtDNA helicase Twinkle and single-stranded DNA-binding protein, mtSSB, co-localize only with a subset of nucleoids. Using nucleotide analogs to identify replicating mtDNA in situ, the fraction of label-positive nucleoids that is Twinkle/mtSSB positive, is highest with the shortest labeling-pulse. In addition, the recruitment of mtSSB is shown to be Twinkle dependent. These proteins thus transiently associate with mtDNA in an ordered manner to facilitate replication.

INTRODUCTION
Mitochondrial DNA (mtDNA) was first visualized in 1963 (1), and subsequently found associated with the mitochondrial inner membrane (IM) (2). This was confirmed by electron microscopy (3) but to date the nature of mtDNA–membrane association has not been clarified. Microscopic methods have been used to show the organization of mtDNA in distinct structures, nucleoids, within the mitochondrial network (4), but it was not until 2001 when the first specific in situ protein co-localization was shown with the mitochondrial helicase Twinkle and mtDNA (5). The first nucleoid purification method identified two mtDNA binding proteins, mitochondrial single-stranded DNA-binding protein (mtSSB) and transcription factor A (TFAM) (6,7), both shown to co-localize with mtDNA in situ (8–11).

Mammalian mtDNA replication requires the concerted action of several replication factors including the mtDNA polymerase γ (POLG), the mtDNA helicase Twinkle, mtSSB and the transcription and packaging protein TFAM [see e.g. (12) for a review]. A minimal replisome consisting of Twinkle, POLG and mtSSB is capable of synthesizing the equivalent of a full-length mtDNA of 16.5 kb in vitro (13). Although overexpressed Twinkle, as well as endogenous mtSSB and TFAM have been shown in situ to co-localize at least partially with mtDNA, the...
Whole Cell Formaldehyde Cross-Linking Simplifies Purification of Mitochondrial Nucleoids and Associated Proteins Involved in Mitochondrial Gene Expression

Nina Rajala1, Fenna Hensen2, Hans J. C. T. Wessels2,3, Daniel Ives4, Jolein Gloerich2,3, Johannes N. Spelbrink1,2,*

1 Mitochondrial DNA Maintenance Group, BioMediTech, FI-33014 University of Tampere, Tampere, Finland, 2 Department of Pediatrics, Nijmegen Centre for Mitochondrial Disorders, Radboud University Medical Centre, Geert Grootplein 10, P.O. Box 9101, 6500 HB, Nijmegen, The Netherlands, 3 Radboud Proteomics Centre, Department of Laboratory Medicine, Laboratory of Genetic Endocrine and Metabolic Disorders, Radboud University Medical Centre, Geert Grootplein 10, P.O. Box 9101, 6500 HB, Nijmegen, The Netherlands, 4 MRC-National Institute for Medical Research, Mill Hill, London, United Kingdom

* hans.spelbrink@radboudumc.nl

Abstract

Mitochondrial DNA/protein complexes (nucleoids) appear as discrete entities inside the mitochondrial network when observed by live-cell imaging and immunofluorescence. This somewhat trivial observation in recent years has spurred research towards isolation of these complexes and the identification of nucleoid-associated proteins. Here we show that whole cell formaldehyde crosslinking combined with affinity purification and tandem mass-spectrometry provides a simple and reproducible method to identify potential nucleoid associated proteins. The method avoids spurious mitochondrial isolation and subsequent multifarious nucleoid enrichment protocols and can be implemented to allow for label-free quantification (LFQ) by mass-spectrometry. Using expression of a Flag-tagged Twinkle helicase and appropriate controls we show that this method identifies many previously identified nucleoid associated proteins. Using LFQ to compare HEK293 cells with and without mtDNA, but both expressing Twinkle-FLAG, identifies many proteins that are reduced or absent in the absence of mtDNA. This set not only includes established mtDNA maintenance proteins but also many proteins involved in mitochondrial RNA metabolism and translation and therefore represents what can be considered an mtDNA gene expression proteome. Our data provides a very valuable resource for both basic mitochondrial researchers as well as clinical geneticists working to identify novel disease genes on the basis of exome sequence data.
Introduction

Mammalian mitochondrial DNA (mtDNA) was discovered in the 1960's\cite{1,2} and early studies in 1969 by Nass suggested that mtDNA could be membrane bound\cite{3}. Later studies postulated that mtDNA is attached to the inner membrane involving the major non-coding D-loop region \cite{4,5}. The first microscopic observation of mtDNA as discrete structures within mitochondria came from the use of a DNA stain in the yeast \textit{S. cerevisiae}. However, despite this evidence and many additional studies in yeast and many other, often vertebrate, species (see below), mtDNA in mammals was for many years described as naked. This view has changed over the last 15–20 years and mtDNA is now generally considered to be organized in discrete nucleo-protein complexes that are designated nucleoids by analogy to nucleo-protein complexes in bacteria\cite{6,7}. Studies in \textit{Xenopus laevis} oocytes suggested that mtDNA was packaged in a compact beaded structure that was membrane associated\cite{8}. Mignotte & Barat\cite{9} characterised a single 28 kDa protein component of the “beads” that was able to introduce superhelical turns, later identified as mitochondrial transcription factor A (TFAM)\cite{10}. TFAM (Abf2 in yeast) is considered to be the principle mtDNA packaging factor\cite{7}. Mitochondrial single stranded DNA binding protein (mtSSB) was also one of the early identified nucleoid proteins\cite{11}. Twinkle, the mitochondrial DNA helicase, was the first mammalian protein shown to co-localise with mtDNA in immunofluorescence studies\cite{12} and is part of a minimal replisome together with mtSSB and Polymerase gamma (POLG) in \textit{in vitro} studies\cite{13}. TFAM and mtSSB were shown also to co-localise with mtDNA \textit{in situ}\cite{14–16}, the latter showing enrichment in particular with replicating nucleoids\cite{17}.

The above proteins (and where conserved, their yeast counterparts) are all considered \textit{bona fide} nucleoid associated proteins (NAPs) and have a clear function in mtDNA packaging, replication and transcription. However, it has become clear that many additional factors associate with mtDNA to facilitate mtDNA maintenance as well as gene expression\cite{18–26}. In particular in yeast, these factors have been show to associate both transiently and under specific metabolic conditions\cite{27}. By comparison of yeast and vertebrate nucleoid proteomes it has also become clear that there appears to be little conservation of associated additional factors. This is considered a consequence of divergent protein-mtDNA co-evolution\cite{28}.

In order to fully understand mammalian mtDNA maintenance and gene expression, and solve conflicting models for example for mtDNA replication, the identification and functional study of the full set of proteins involved in mtDNA metabolism is important. One approach to identify NAPs is via biochemical isolation and mass spectrometric identification. Over the last 10 years various sets of NAPs were identified, but as we discussed recently, few proteins are shared between all these sets\cite{29}, a consequence of the various methods and starting materials employed, the stringency of isolation, the target at which isolation was directed and the fact that many protein-nucleoid interactions are transient in nature. On the basis of this comparison we also concluded that the most inclusive method, identifying most factors known to interact with mtDNA involved a formaldehyde cross-linking step. However, very few of the studies published so far have used quantitative proteomics and typically have presented the data of just one or two purifications (see \cite{29}). This low replicate number is probably due to the complexity of some of the isolation procedures involved that require large quantities of starting material.

Here we present the shotgun proteomics results using a greatly simplified mtDNA nucleoid proteomics analysis using whole cell formaldehyde cross-linking followed by cell lysis and affinity purification. Here, induced overexpression of a FLAG-tagged mtDNA helicase Twinkle was used because overexpressed Twinkle specifically co-localizes with mtDNA nucleoids\cite{17} and short, low level induction was previously shown to minimally impact on nucleoid structure, mtDNA levels and transcription\cite{30,31}. Because we carry out the cross-linking in whole
cells, this eliminates the need to isolate mitochondria and use subsequent nucleoid purification steps. By comparing non-cross-linked with cross-linked Twinkle-FLAG samples versus non-cross-linked and cross-linked control cells that express a mitochondrially targeted and FLAG-tagged Luciferase we show that many previously identified nucleoid proteins were specifically enriched in cross-linked Twinkle-FLAG purification. Here, the comprehensive use of both non-cross-linked and cross-linked samples and controls in combination with multiple biological and technical repeats by accurate label free quantification (LFQ) provides a firmer basis for the consideration of many putative NAPs and identifies an inclusive list of proteins not just for mtDNA maintenance but also for mitochondrial RNA metabolism and translation. In addition, we identified several potential new NAPs. Finally, in a comparison of Twinkle-FLAG expressing cells either or not containing mtDNA we identify those proteins that co-purify with Twinkle-FLAG because of the presence of mtDNA/RNA, which suggests that many of these proteins interact with mtDNA/RNA but not directly with Twinkle. The ease of our method and application of LFQ is expected to find much wider application in the study of dynamic mitochondrial protein-protein and protein-nucleic acid interactions.

Materials & Methods

Routine cell culture and creation and maintenance of stable transfected inducible expression cell lines

Stable cell lines expressing various mtDNA maintenance proteins upon induction were created as described[30] using the Flp-In T-Rex 293 host cell line (Invitrogen), a HEK293 variant containing a Flip recombination site at a transcriptionally active locus, or Flp-In T-Rex 293 ρ0 cells (see below). The resulting cells were grown in DMEM medium (Sigma) supplemented with 10% FCS (Sigma), 2 mM L-glutamine, 1 mM Na-pyruvate, and with the addition of 50 μg/ml uridine (Sigma) in ρ0 cells, 100 μg/ml Hygromycin and 15 μg/ml Blasticidin (Invivogen) in a 37°C incubator at 8.5% CO2. Flp-In T-Rex 293 expressing a mitochondrially targeted and FLAG tagged Luciferase (mtLucFLAG) were a kind gift of Profs. Robert Lightowlers and Zosia Chrzanowska-Lightowlers (see also[32]).

To isolate a ρ0 variant of the HEK293 Flp-In T-Rex cell line, cells were grown for an extended period of time in standard medium supplemented with 50 ng/ml Ethidium Bromide (EB) and 50 μg/ml uridine. EB treated HEK293 Flp-In T-Rex cells were tested for mtDNA depletion by growth on galactose medium. Galactose medium contained glucose-free DMEM, 1 mM (0.5 mg/l) pyruvate and 5 mM (0.9 mg/ml) filter-sterilised D-(+)-galactose (Sigma). Cells were further tested for total mtDNA depletion by southern blot of total DNA with D-loop (H1) probe (S1 Fig.). This result suggested that prolonged EB treatment had successfully depleted HEK293 Flp-In T-Rex cells of their mtDNA. This was confirmed when the putative ρ0 cells were grown in the absence of EB for a period of several months and still found to lack any detectable mtDNA by dot-blot analysis (unpublished data Ş. Cansiz-Arda and J.M. Gerhold, Spelbrink lab). Prior to southern blot, total DNA was extracted by isopropanol precipitation, digested overnight with PvuII at 37°C, heat denatured at 70°C for 10 minutes and separated on a 1.0% TBE agarose gel at room temperature for 3 hours at 100 volts. D-loop (H1) probe (16241–141) primers: Forward – TTACAGTCAAATCCCTTCTCGT, Reverse – GGATGAGCGATGGAATCAAAGACG.

Western blot analysis

Immunoprecipitation eluates were analysed for proteins by immunoblotting after SDS–PAGE [33]. Antibody dilutions were as follows: primary FLAG monoclonal (Sigma), 1:4000, TFAM
rabbit polyclonal antibody (kind gift of Dr. R. Wiesner), 1:10000; Twinkle mouse monoclonal (kind gift of Anu Wartiovaara-Suomalainen) 1:1000; mtSSB rabbit polyclonal (Sigma, HPA002866), 1:2000; POLG1 goat polyclonal (Santa Cruz, sc-5931), 1:1000. Peroxidase-coupled secondary antibody horse-anti-mouse or goat-anti-rabbit (Vector Laboratories) 1:5000.

**Formaldehyde cross-linking and immunoprecipitation**

Twinkle expression was induced by addition of 3ng/ml doxycycline (Sigma) for 36 hours. From previous experiments we know that this expression level and time is appropriate to preserve nucleoid structures[31]. For cross-linking typically cells from five 145 mm (cross-section) cell culture dishes were harvested and cell number was adjusted to 10×10^6 cells/ml. Cross-linking was carried out in 1% formaldehyde (Sigma) for 10 min at RT with rotation. The reaction was stopped by addition of 125 mM glycine, pH 8.0. Formaldehyde is toxic and was handled in a fume hood. Sample handling after addition of formaldehyde similarly was carried out in a fume hood and formaldehyde disposed appropriately. Cells were transferred on ice and all subsequent centrifugations carried out at +4°C. Cells were washed four times with ice cold TBS (50 mM Tris-HCl pH 7.4, 150 mM NaCl) and processed further by two different methods. Method A, Triton X-100 method: Cells were lysed in Buffer A (50mM Tris-HCl pH 7.4, 300 mM NaCl, 2mM EDTA 1% Triton X-100). In method B, the X-ChiP method, cells were lysed with RIPA buffer (50mM Tris-HCl pH 8, 150mM NaCl, 1% NP-40 (Igepal), 0.5% sodium deoxycholate, 0.1% SDS). In both methods lysates were sonicated for 1 min at 40% power (1s on 2s off cooling on ice), but only with the X-ChiP method sonication was followed by addition of 100μg/ml RNase A (Sigma), 5U/ml DNase I (Thermo Scientific) and 50U/ml Benzonase nuclease (Sigma), 2.5M Mg2+, 1mM CaCl2 and incubated at +37°C for 30min. With both methods lysates were centrifuged for 10 min at 1200g at +4°C and the protein content of the lysates was equalised to 2mg/ml in a total volume of 10 ml before addition of 180 μl of FLAG resin (Sigma) and rotation for 2 hours at +4°C. In method A, FLAG resin was washed once in buffer B, C and D. Buffer B: 50mM Tris-HCl pH 7.4, 800mM NaCl, 0.1% Triton-X 100, Buffer C: 50mM Tris-HCl pH 7.4, 50 mM NaCl, Buffer D: 50mM Tris-HCl pH 7.4, 150 mM NaCl, 0.1% Triton-X100. Nucleoids were eluted with 100 μl 3xFLAG peptide (at 0.25 mg/ml) in 50mM Tris-HCl pH 7.4, 150 mM NaCl. In method B, the FLAG resin was washed three times in RIPA buffer and nucleoids eluted with 100 μl 3xFLAG peptide (at 0.25 mg/ml) in RIPA buffer. All buffers included 1×complete EDTA-free Protease inhibitors (Roche).

**Mass spectrometry sample preparation**

Protein samples were incubated with SDS-PAGE sample-buffer for 30 min at 95°C to reverse FA cross-links and fractionated by SDS-PAGE on Any kD Mini-PROTEAN TGX Gels (BIO-RAD). Lanes were cut in 3 equal-sized (approximately 1x2.5 cm) gel slices. No gel-staining was applied following electrophoresis. Each gel slice was subjected to in-gel tryptic digestion and further processed according to standard methods[34]. In short, gel slices were cut into small pieces (~1mm^2) and were washed successively at least three times with 50 mM ammonium bicarbonate (ABC) and 100% acetonitrile (ACN). Gel slices were swelled in 10 mM dithiothreitol and incubated for 20 minutes at 56°C to reduce protein disulfide bonds. To remove the reduction buffer, gelpieces were shrunk with ACN. Alkylation of the reduced cysteines was performed by incubation of 50mM chloroacetamide in ABC for 20 minutes at room temperature in the dark. Gel pieces were again washed twice with ACN and ABC before tryptic digestion at 37°C overnight with 1.25ng/μl sequencing grade modified Trypsin (Promega) in ABC. To recover tryptic peptides from the gel pieces, they were first diluted 1:1 with 2% trifluoractic acid (TFA), sonicated for 30 seconds, and incubated at RT for ≥ 15 minutes with gentle agitation.
Supernatant was transferred to a fresh tube and the gel pieces were shrunk with 100% ACN at RT at gentle agitation for ≥ 15 minutes to recover remaining peptides from the gel. Supernatant was pooled and subjected to vacuum centrifugation to remove the ACN and concentrate the sample. Thereafter, the peptide sample was desalted and concentrated by “STop And Go Extraction (STAGE) tips”[35].

Mass spectrometric measurements

Measurements were performed by nanoflow reversed-phase C18 liquid chromatography (EASY nLC, Thermo Scientific) coupled online to a 7 Tesla linear ion trap Fourier-Transform ion cyclotron resonance mass spectrometer (LTQ FT Ultra, Thermo Scientific) or by nanoLC 1000 (Thermo Scientific) chromatography coupled online to Q Exactive hybrid quadrupole-Orbitrap mass spectrometer (Thermo Scientific). Chromatography was performed with an Acclaim PepMap 0.3 x 5 mm 5μm 100Å trap column (Thermo scientific) in combination with a 15cm long x 100μm ID fused silica electrospray emitter (New Objective, PicoTip Emitter, FS360-100-8-N-5-C15) packed in-house with ReproSil-Pur C18-AQ 3 μm 140Å resin (Dr. Maischi)[36]. Tryptic peptides were loaded onto the trap column using 0.1% formic acid and separated by a linear 60 minutes (LTQ-FT) or 30 minutes (Q Exactive) gradient of 5–35% acetonitril containing 0.1% formic acid at a flow rate of 300 nl/min. For the LTQ-FT; the mass spectrometer was set to positive ion mode and acquired one full MS survey scan in the ICR cell parallel to up to four data dependent collision induced dissociation (CID) fragmentation spectra by the linear ion trap. Full MS precursor scans were performed with a single microscan at 100.000 resolving power (FWHM) at m/z 400 using 1E6 ions or after 2500ms injection time if this came first. Data dependent acquisition of MS/MS spectra by the linear ion trap was performed on 3E4 ions or after 750 ms maximal injection time. Fragmentation of the precursor ion by CID was performed at 30% normalized collision energy for 30 ms and activation Q = 0.25. An isolation width of 3 Th was set to isolate the precursor ion for MS/MS sequencing events. For the Q Exactive; the mass spectrometer was again set to positive ion mode. Full MS events were performed at 70.000 resolving power (FWHM) at m/z 200 using 1E6 ions or after 20ms of maximal injection time. Data-dependent MS/MS spectra were performed using 1E5 ions at 17.500 resolving power (FWHM) at m/z 200 or after 50ms maximal injection time for the top 10 precursor ions with an isolation width of 4.0 Th and fragmented by higher energy collisional dissociation (HCD) with a normalized collision energy of 30%.

Mass spectrometric data analysis

Data analysis was performed with the MaxQuant software (version 1.3.0.5)[37] applying default settings with minor modifications. The precursor mass tolerance for Q Exactive measurements was set to 4.5 ppm. For both LTQ-FT and Q Exactive the multiplicity was set to 1 and Trypsin was chosen as the proteolytic enzyme allowing for 2 miscleavages. Default MaxQuant normalizations were applied. Database searches were performed on the human RefSeq database in which the reversed database is used to calculate the false discovery rate (FDR) which was set to 1% and isoleucine and leucine were forced to be treated equally. Between samples the option “Match between runs” was enabled to detect sequenced peptides which were not subjected to sequencing event in other samples and Label Free Quantification (LFQ) calculation was applied. Peptide modifications after formaldehyde cross-linking did not occur as tested by the presence of two possible modification occurring when the cross-linking is not reversed completely. The first modification is the addition of 30 Da considered to be the addition of the whole formaldehyde molecule (O = CH2) as an intermediate step in the cross-linking reaction. The second possible modification is the addition of 12 Da which equals the...
addition of formaldehyde followed by the release of a water molecule and is considered to be the final product [38]. Since neither modification occurred, the reversal of cross-linking seems to be complete. Furthermore, there is an increased possibility of miscleavages since the reactivity of formaldehyde is the highest on those amino acids subjected to tryptic digestion, this did not seem to give any problems since we allowed for maximum of two miscleavages and were not able to detect any miscleavage in combination with peptide modifications. Raw data files provided by MaxQuant were further analyzed manually. For the biological replicates LFQ values were used to calculate the ratios between samples per biological sample. For the triplicate measurements performed on the Q Exactive first the average LFQ values were calculated from the replicates (only proteins identified in all three replica measurements were considered), followed by calculation of the ratios between sample conditions. Whenever the ratio exceeded the value of 2 or was below 0.5, the protein was called to be respectively increased or decreased. Additional protein information such as the Gene Ontology SLIM cellular compartment (CC), molecular function (MF), biological process (BP) and the official gene symbol were acquired using ProteinCenter (version 3.12.10015; Thermo Scientific).

Results

Mitochondrial nucleoid proteins can be isolated following whole cell cross-linking

In order to test the applicability of whole cell XL in the analysis of mtDNA-protein nucleoid complexes we first set out to establish that we can enrich for some of the proteins associated with nucleoids using Western blot analysis. We used the inducible HEK293 FlpIn TREx system to inducibly express the mtDNA helicase Twinkle (as previously described[30,31]) with a FLAG tag at its C-terminus. Twinkle was selected as target protein since all available evidence suggests it to function as a core component of the mtDNA replisome (e.g.[13,17]). Here, the use of tagged Twinkle overexpression over immunopurification of endogenous Twinkle was preferred because of the very low abundant endogenous expression of the protein. In addition, short Twinkle induction with a low concentration of doxycycline does not interfere with mtDNA maintenance or gene expression[30,31]. As controls we not only used parallel cultures in which no FA was applied (-XL), but also parallel cultures expressing FLAG-tagged and mitochondrially targeted Luciferase (mtLucFLAG)[32] without and with FA. Western blot analysis of these samples showed that several proteins implicated in mtDNA maintenance such as TFAM, POLG1 and mtSSB are specifically enriched by FA crosslinking in TwinkleFLAG samples, following FLAG immuno affinity purification (IAP) (Fig. 1). The results also showed that following XL, TwinkleFLAG can be affinity-purified almost as efficiently as without XL and that in principle whole cell XL in combination with IAP can be used to enrich for nucleoid associated proteins (this is validated by our mass-spectrometry analysis below). This result also shows that the FLAG tag is suitable for FA applications despite the presence of several lysine residues. Please note that these Western-blot analyses do not assess sample complexity or the percentage of mitochondrial proteins in the preparation which require mass spectrometry based methods.

Identifying potential nucleoid associated proteins using mass spectrometry

To more systematically analyse samples we next applied LTQ-FT mass spectrometry on IAP eluates, analysing the protein composition of these samples by shotgun proteomics. To optimize the procedure and establish the robustness of the crosslinking and IAP method we first
measured several completely independent biological repeats over an extensive period of more than 1 year using various batches of TwinkleFLAG and mtLucFLAG cells with and without XL. Following individual sample analysis at the time of sample preparation, raw mass spectrometry data files of all samples were analysed in one batch using MaxQuant[37]. This allows for the post-hoc comparison of signal intensities of peptides between samples to provide a relative abundance measure for identified proteins. Based on this analysis we initially compared biological repeats by taking LFQ ratio’s for the identified proteins between the 4 conditions tested (being mtLucFLAG -XL or +XL, and TwinkleFLAG -XL or +XL), compiling lists of proteins with a least a 2-fold increase compared to its control and comparing these lists between the biological repeats. From this we extracted ‘≥2 fold increase’ lists based on the further condition that this was observed in at least 2 out of 3 samples. To finally extract meaningful protein sets we generated Venn diagrams simultaneously comparing the four generated protein lists.

Fig 1. Validation of TwinkleFLAG IAP following whole cell cross-linking. HEK293 Flp-In T-Rex cells expressing either TwinkleFLAG or a mitochondrially targeted Luciferase FLAG (mtLucFLAG) were induced for 36 hrs with 3 ng/ml doxycycline, harvested, samples equalized by protein content and incubated for 10 min with 1% formaldehyde (FA) for whole cell crosslinking. Following cross-linking, cells were lysed and FLAG-tagged protein purified using FLAG immunoaffinity resin. Precipitated complexes were analysed using Western blot analysis (see M&M and main text for full details). Results (A, B) show that proteins of the mtDNA maintenance machinery are enriched with cross-linking in TwinkleFLAG expressing cells. (C) HEK293 Flp-in T-Rex cells expressing TwinkleFLAG were established and crosslinked samples of TwinkleFLAG expressing cells were compared with their mtDNA-containing parental cells also expressing TwinkleFLAG. Results show a very substantial decline in levels of co-purifying TFAM and mtSSB, in the absence of mtDNA.

doi:10.1371/journal.pone.0116726.g001
A comparison of the enriched protein sets shows that both with Twinkle and mitochondrially targeted Luciferase, cross-linking results in a marked enrichment of mitochondrial proteins: cross-linking increased the percentage of mitochondrial proteins in both TwinkleFLAG +XL versus -XL and mtLucFLAG +XL versus -XL from 28 to 70% based on Gene-Ontology (GO)-SLIM annotation (Fig. 2). This analysis illustrates a considerable enrichment of mitochondrial proteins with cross-linking, suggesting the fixation of specific direct and indirect interactions with the respective baits.

By applying a stringent comparison between experiments the most interesting lists from the perspective of identifying potential NAPs and comparing identified proteins with previously published datasets are those proteins that are enriched in TwinkleFLAG +XL compared to respective controls. The resulting list of 168, used for later comparison (see Fig. 3) is separately given alphabetically by gene name in S4 Table (first sheet: ‘Biol repeats enriched all’). S1 Table, in addition is sorted in such a way that the same 168 proteins are the first 168 proteins listed in the LFQ comparison sheet (sheet 3).

Fig 2. Whole cell cross-linking followed by IAP enriches for mitochondrial and nucleoid associated proteins. Protein complexes purified using FLAG-tag targeted isolation from 3 independent biological repeats using various batches of TwinkleFLAG (Twinkle) and mtLucFLAG (Luc) cells, treated either with or without FA and further processed as described in Fig. 1, were analysed by shotgun mass spectrometry. Using MaxQuant, LFQ values were derived and ratios calculated comparing TwinkleFLAG versus mtLucFLAG without cross-linking (-XL) with crosslinking (+XL) as well as TwinkleFLAG +XL versus -XL and mtLucFLAG +XL versus -XL. Protein lists were compiled based on a ≥2 fold increase in LFQ values in at least 2 out of 3 experiments (see S1 Table). (A) Gene Ontology (GO)_SLIM_Cellular Compartment (CC) (see also M&M) annotation was used to calculate percentages of mitochondrial proteins in each set. This analysis illustrates that all crosslinked sets (being either with TwinkleFLAG or mtLucFLAG) showed approximately 70% mitochondrial annotation whereas the TwinkleFLAG versus mtLucFLAG -XL showed only 28% mitochondrial annotation. (B) To identify potentially interesting proteins we compared all 4 generated lists simultaneously using Venny (http://bioinfogp.cnb.csic.es/tools/venny/index.html), that generates a 4-way Venn diagram and separate lists for all intersecting and non-intersecting parts of the diagram. The region for potentially interesting proteins, being enriched with TwinkleFLAG +XL compared to respective controls is further outlined in red. The resulting list of 168, used for later comparison (see Fig. 3) is separately given alphabetically by gene name in S4 Table (first sheet: ‘Biol repeats enriched all’). S1 Table, in addition is sorted in such a way that the same 168 proteins are the first 168 proteins listed in the LFQ comparison sheet (sheet 3).

(Fig. 2) using Venny (http://bioinfogp.cnb.csic.es/tools/venny/index.html). MaxQuant raw data output and analyses sheets can be found in S1 Table. A comparison of the enriched protein sets shows that both with Twinkle and mitochondrially targeted Luciferase, cross-linking results in a marked enrichment of mitochondrial proteins: cross-linking increased the percentage of mitochondrial proteins in both TwinkleFLAG and mtLucFLAG samples from 28 to 70% based on Gene-Ontology (GO)-SLIM annotation (Fig. 2). This analysis illustrates a considerable enrichment of mitochondrial proteins with cross-linking, suggesting the fixation of specific direct and indirect interactions with the respective baits.
machineries, such as TFAM, POLG1, mtSSB and POLRMT. Combined our analysis of several biological repeats and comparison with published datasets of NAPs (see also below), including a dataset of enriched proteins purified with the mitoribosomal associated protein ICT1 that also included many potential NAPs[29,38], shows that whole cell crosslinking in combination with IAP is a valid method to identify these proteins.

To further optimize our method to be able to more confidently identify potential novel NAPs we compared two different isolation conditions, considering that different isolation methods should yield at least a similar core set of proteins. The first is the condition used for the biological repeats above, which involves a relatively high-salt (300 mM NaCl) Triton-X100 lysis with sonication followed by IAP and washing with buffers both with high and low salt (see Materials & Methods). The second involves a representative protocol used for cross-linking chromatin immune precipitaton (X-ChIP) using sonication in RIPA buffer (see Materials & Methods) with the addition also of RNAse A, DNAse I and Benzonase since our interest is in protein analysis and not DNA analysis typical for X-ChIP. With the addition of nucleases we also hoped to more specifically identify proteins that are in close association with Twinkle and not proteins that co-purify via indirect DNA and/or RNA association (see Discussion & below).

In addition, to give our analysis a more solid basis we measured samples as triplicate technical repeats on a Q Exactive mass spectrometer that possesses a greater sensitivity and faster MS/MS duty cycle, and again allows comparison of samples using LFQ values generated by MaxQuant.

By first measuring one of the biological repeats (sample 2) comparing TwinkleFLAG +XL, mtLucFLAG +XL versus TwinkleFLAG -XL and mtLucFLAG -XL with high-salt Triton X100 (TX100) lysis in triplicate on the Q Exactive allowed us to compare this measurement with the three biological repeats measured on an LTQ FT mass spectrometer. This showed that there is substantial overlap of enriched proteins for both sets of experiments (Fig. 3 & S3 Table) as expected. Overall, however, more proteins were identified on the Q Exactive instrument due to its greater sensitivity and faster MS/MS duty cycle. This set of measurements was now compared to a duplicate Q Exactive measurement of samples purified with the X-ChIP purification method. This comparison shows a considerable number of proteins that were identified with both methods (Fig. 4A and B), despite the presence of nucleases in the X-ChIP based purification. In fact many identified nucleoid associated proteins were detected using the X-ChIP method that were not identified using TX100 lysis. In particular a large number of mitoribosomal proteins and proteins with possible or established roles in RNA metabolism and translation, such as DDX28, TACO1, MTIF2 and MTRF1, were found. In addition, proteins that are considered nucleoid associated proteins by their demonstrated molecular function, such as POLG2 and the recently described nuclease MGME1[39–41] were specifically identified with the X-ChIP method. Possibly the X-ChIP protocol, instead of removing all proteins that are indirectly associated with TwinkleFLAG either via DNA or RNA, might result in a less tightly packed complex in turn resulting in better accessibility of the FLAG epitope for TwinkleFLAG IAP. This would explain the approximately 10-fold higher LFQ values for Twinkle with X-ChIP compared to the TX100 Q Exactive measurements, whereas mtLucFLAG LFQ values are comparable between both sets (S2 and S3 Tables). This can then be expected to result also in a much better recovery of cross-linked mitochondrial proteins in the X-ChIP experiment. Not surprisingly, 98% of all proteins enriched with both the TX100 and the X-ChIP method are mitochondrial (Fig. 4C). Moreover, 88% of proteins that are specifically enriched with the X-ChIP method had a mitochondrial annotation while in contrast, 36% of the proteins that showed specific enrichment only with the TX100 method were mitochondrial, suggesting many of these proteins are contaminants in the preparation.
We recently have shown that Twinkle helicase is firmly membrane associated and that even in the absence of mtDNA it forms discrete membrane associated foci within the mitochondrial network [17]. Based on these findings we suggested also by analogy with baker’s yeast[42] that a subset of nucleoid associated proteins might organize in a replication platform even in the absence of mtDNA. These observations could thus allow us to identify proteins associated with Twinkle in a minimal replication platform, but in addition tackle the question of indirect association via DNA/RNA binding, by purifying TwinkleFLAG following FA cross-linking using cells without mtDNA (hereafter ρ°). HEK293 FlpIn-TREx ρ° cells were established (S1 Fig.) and stable inducible TwinkleFLAG ρ° cells were subsequently generated. As ρ° cells lack mtDNA they also lack mitochondrial tRNAs as well as the two mitoribosomal RNAs and thus functional mitoribosomes cannot be assembled.

Having generated lists of proteins that are enriched in TwinkleFLAG +XL compared to both TwinkleFLAG -XL and mtLucFlag +XL we now considered only those 366 proteins enriched with the X-ChIP protocol (S3 and S4 Tables) in a direct comparison of TwinkleFLAG +XL in regular HEK293 FlpIn-TREx or HEK293 FlpIn-TREx ρ° cells, each measured using the X-ChIP protocol. This revealed that 258 of 366 proteins showed a ≥2 fold decrease in ρ° cells IAP while 95 of those 258 proteins were completely absent (Fig. 4D, S3 Table). The 95 proteins that were absent in this particular ρ° TwinkleFLAG IAP included several nucleoid associated proteins on the basis of earlier demonstration of nucleoid association or a clear function in mtDNA metabolism and expression. Examples hereof include MGME1, MTERF and POLG2, while many other
**Fig. 4. X-ChIP based affinity purification provides the most inclusive analysis of nucleoid associated proteins.** (A) Protein complexes using FLAG-tag targeted isolation using TwinkleFLAG (Twinkle) and mtLucFLAG (Luc) cells, from cells treated either with or without FA were isolated using an X-ChIP based isolation buffer. Samples were analysed (in duplicate for TwinkleFLAG + XL, otherwise in triplicate) by shotgun mass spectrometry using a Q ExactiveOrbitrap. To again identify potentially interesting proteins we compared all 4 generated lists simultaneously using Venny, similar as in Figs. 2/3. The region for potentially interesting proteins, being enriched with TwinkleFLAG +XL compared to respective controls again is further outlined in red. The resulting list of 366 proteins, used for later comparisons (see 4B/C/D) is separately given alphabetically by gene name in S4 Table (third sheet: ‘X-ChIP enriched all’). S3 Table, in addition is sorted in such a way that the same 366 proteins are the first proteins listed in the LFQ comparison sheet (sheet 3). (B) An area-proportional Venn diagram shows the comparison of the enriched set obtained using TX100 lysis compared to the enriched set obtained using the X-ChIP method. An analysis of the proteins identified as enriched in both sets shows that of these 111, 109 proteins (98%) have a Gene Ontology (GO)_SLIM_Cellular Compartment (CC) annotation (C).
proteins were proteins involved in mitochondrial gene expression such as ribosomal proteins, tRNA synthetases, translation and RNA processing factors. The more inclusive list of all proteins that were reduced ≥2 fold included many additional proteins in the same categories, including for example DHX30 and DDX28, LONP1, GRSF1, POLRMT and so on. The ρ° TwinkleFLAG IAP results point to proteins that co-purify with TwinkleFLAG in regular HEK293 FlpIn-TREx cells by means of association with DNA/RNA or possibly other higher order structures that are modified or absent in ρ° cells. A comparison of commonly identified proteins associated with purified cross-linked nucleoids and mitochondrial ribosomes [29], proteins purified using our two isolation methods and cell lines either with or without DNA is given in Fig. 5. This figure again illustrates not only that with the X-ChIP protocol we identify the majority of previously identified proteins but also how these proteins change in ρ° cells.

As pointed out above, those proteins that are not changed more than 2-fold or that are increased might also be of interest. Similar protein levels might indicate nucleoid associated proteins that directly interact with TwinkleFLAG or with a membrane platform and/or a minimal nucleoid that is still present in ρ° cells and that Twinkle is part of [17]. These proteins (see Fig. 4D, S3 Table) include for example ATAD3, MTERFD2 and ATP synthase subunits (see Discussion).

Discussion
In this paper we demonstrate the feasibility of using whole cell formaldehyde cross-linking in combination with immuno-affinity purification and tandem mass spectrometric analysis in the identification of a mitochondrial protein complex, in this case the nucleoid mtDNA-protein complex. We show that this method identifies many of the same proteins as previously published mitochondrial FA cross-linking experiments combined with several subsequent more laborious purification steps. The method in addition identified several novel proteins that should be considered prime candidate nucleoid associated proteins. By the application of label free quantification we could analyse the effects of isolation buffers and the effect of isolating Twinkle in the absence of mtDNA and consequently all mitochondrially encoded RNAs. The latter experiment was very revealing in that it identified many proteins that were considerably reduced or absent in TwinkleFLAG IAP from ρ° cells pointing to their association with the nucleoid in mtDNA containing cells on the basis of DNA/RNA association and notwithstanding the possibility that a number of these proteins might also be less stable in the absence of mtDNA/RNA. Many of these proteins should thus be considered as nucleoid associated. In addition it pointed to at least some proteins that are found in close vicinity or directly interact with Twinkle and could form part of a minimal membrane associated platform. Combined our results position the mitochondrial gene expression machinery including proteins involved in RNA processing and translation in close vicinity to nucleoids while at the same time providing an
<table>
<thead>
<tr>
<th>ID</th>
<th>Protein</th>
<th>Gene</th>
<th>ICT1-FLAG IP</th>
<th>TX100</th>
<th>TX100 triplicate</th>
<th>X-ChIP method triplicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;g4507401</td>
<td>TFAM</td>
<td>TFAM</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g374821775</td>
<td>mTSSB</td>
<td>SSBP1</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g34147030</td>
<td>IFITu</td>
<td>TUFM</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g20382924</td>
<td>DEAH box polypeptide 30-2</td>
<td>DHX30</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g20127496</td>
<td>Trifunctional enzyme subunit alpha, mitochondrial</td>
<td>HAOA</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g11092853</td>
<td>mTrNA polymerase</td>
<td>POLRMT</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g19835919</td>
<td>Serine hydroxymethyltransferase</td>
<td>SHMT2</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g31675937</td>
<td>LRRP130</td>
<td>LRPRC</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g283439222</td>
<td>AAA-ATPase</td>
<td>ATAD3A</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g40525494</td>
<td>LRRC59</td>
<td>LRRC59</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g45289773</td>
<td>Prohibitin 1</td>
<td>PHB1</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g22139704</td>
<td>Prohibitin 2</td>
<td>PHB2</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g15717127</td>
<td>mtDNA polymerase g A</td>
<td>POLG</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g70867730</td>
<td>mtDNA polymerase g B</td>
<td>POLG</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g15907459</td>
<td>ANT2</td>
<td>SLC25A5</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g15870482</td>
<td>ANT3</td>
<td>SLC25A6</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g72422016</td>
<td>CGI caseinolytic protease X homolog</td>
<td>CLPX</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g39725943</td>
<td>Twinkle</td>
<td>PED1</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g45034272</td>
<td>Trifunctional enzyme subunit beta</td>
<td>HADHB</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g19424872</td>
<td>HSP70</td>
<td>HSPA1</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g21398489</td>
<td>Lon protease</td>
<td>LONP1</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g31543967</td>
<td>SU(V)-like helicase</td>
<td>SUPV1L1</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g15945992</td>
<td>TFB1M</td>
<td>TFB1M</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g11641289</td>
<td>TFB2M</td>
<td>TFB2M</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g382849407</td>
<td>Branched chain ketoadehyrogenase subunit E2</td>
<td>DBT</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g203569338</td>
<td>(TID1)</td>
<td>DNAJA3</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g281804130</td>
<td>Serine/threonine-protein phosphatase PGAMS5</td>
<td>MSG5302 (PGAMS5)</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g7561872</td>
<td>PDP38</td>
<td>PDP38</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g11225820</td>
<td>DNA topoisomerase 1, mitochondrial</td>
<td>TOP1MT</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g5502010</td>
<td>Mert</td>
<td>MERF</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g25772277</td>
<td>DEAD box 26</td>
<td>DDXXB6</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g30019203</td>
<td>AF03 ATPase family gene 3-like 2</td>
<td>AF03L2</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g34147076</td>
<td>MTERF domain-containing 1</td>
<td>MTERFD1</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g21372685</td>
<td>YME1-like isoform 1</td>
<td>YME1L1</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g75677573</td>
<td>ATPase family, AAA domain containing 3B</td>
<td>ATAD3B</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g14500778</td>
<td>Coatr4</td>
<td>NOA1</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g4501859</td>
<td>Ict2</td>
<td>ACTP1</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g38893895</td>
<td>Pentatricopectin repeat-containing protein 3</td>
<td>PTCD3</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g4518697</td>
<td>Polybi-1-DNA helicase ICT1, mitochondrial</td>
<td>ICT1</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g15343515</td>
<td>GADD45G1P1</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>&gt;g24384988</td>
<td>MTHSF75</td>
<td>HSP90</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g3171196</td>
<td>FRC-E2</td>
<td>DAP</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g12976788</td>
<td>Myosin, heavy chain 9, non-muscle</td>
<td>MYH9</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g14150128</td>
<td>Mitochondrial nucletic factor 1 (M19)</td>
<td>MNF1</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g22072317</td>
<td>Mitochondrial ribosome recycling factor isoform 1</td>
<td>MNRF</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g19321616</td>
<td>NIP51NAP</td>
<td>NIP51NAP1</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g4505283</td>
<td>heat shock 10kDa protein 1 (chaperonin 10)</td>
<td>HSP10</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g3952089</td>
<td>36S ribosomal proteins (~50 total)</td>
<td>36S</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>&gt;g3952089</td>
<td>36S ribosomal proteins (~50 total)</td>
<td>36S</td>
<td></td>
<td>26</td>
<td>9</td>
<td>25</td>
</tr>
</tbody>
</table>

Fig. 5. Comparing whole cell cross-linking TwinkleFLAG immune affinity purification with previous nucleoid isolations. Comparison with most commonly identified potential mtNAPs as published in[29] with their enrichment in the TwinkleFLAG +XL IAP. The data here is reduced to compare previously published mitochondrial formaldehyde cross-linking followed by nucleoid purification as performed by[22], in which for simplicity reasons both published protein list are combined to one list and the data from ICT1-FLAG IP as performed by[38]. For the full table see Hensen et al[29]. Shown are the comparison of the three biological repeats on the LTQ-FT Ultra, Q Exactive TX100 and X-ChIP method datasets. Green checkmark indicates an > 2 fold increase in the TwinkleFLAG IAP compared to the mtLucFLAG IAP control with cross-linking. A light red cross indicates no difference while a dark red cross indicates undetected protein. Green checkmark indicated with an asterix represent proteins which are increased in TwinkleFLAG compared to mtLucFLAG with cross linking but not compared to non cross- linked TwinkleFLAG control (Twinkle itself is a logical representative of this class). For the ρ° samples we indicate the percentage of protein, based on LFQ ratios, co-purified in the absence of mtDNA.

doi:10.1371/journal.pone.0116726.g005
advantage of formaldehyde is that the cross-link is reversible. Potential formaldehyde-induced protein modifications were not observed by us following heat-induced reversal of the cross-link, allowing for efficient mass spectrometry based analysis of protein samples. Formaldehyde cross-linking is not considered to be specific in literature, which might result in many false positives. By applying stringent analysis criteria, the use of various controls and a combination of cross-linking and IAP we show here that we nevertheless most consistently identified mitochondrial proteins that are furthermore considerably enriched when we compare cross-linked with non cross-linked samples. This was especially the case using the X-ChIP protocol. The use of a FLAG epitope tag poses another potential problem [43] as the FLAG tag contains several lysines that are substrates for FA cross-linking, but we have shown here, both by Western blot analysis and by LFQ-based quantitative mass spectrometry, that in our hands the combination of a short formaldehyde exposure in whole cell crosslinking did not result in dramatic adverse effects on the efficiency of FLAG IAP. This is very important as it shows that we can directly compare -XL with +XL conditions. Likewise comparison of LFQ values for TwinkleFLAG between regular HEK293 FlpIn- TREx or HEK293 FlpIn- TREx ρ° show only a 24% lower level in the ρ° cell IAP showing the validity of the comparison of LFQ values of co-precipitated proteins. The analysis presented here thus shows that our approach can have a much wider application in the analysis of mitochondrial protein complexes.

Can we define a consensus list of nucleoid associated proteins based on formaldehyde cross-linking?

Formaldehyde can cross-link proteins to nucleic-acid but more efficiently cross-links proteins to proteins. Combine this with a high mitochondrial protein density and the tremendous sensitivity and speed of modern mass spectrometers, which is also illustrated here by the considerable increase of identified proteins by the use of a Q Exactive Orbitrap compared to a LTQ-FT mass spectrometer, and the answer to the above question clearly is no. What we do show here however, similar to what was recently discussed [29], is that formaldehyde cross-linking in combination with an appropriate isolation method yields an inclusive list of proteins, proteins that in addition might be found in close vicinity to the nucleoid in what could be considered a mitochondrial nucleoid ‘compartment’. This compartment, similar to earlier suggestions [22,24,26,44], would contain not only mtDNA and associated factors but also the many proteins involved in mitochondrial RNA metabolism and translation. In fact a large fraction of proteins we have identified as ‘nucleoid’-enriched fall in this last category, as also found by He et al[24]. This nevertheless does not discredit our method to identify potential nucleoid associated proteins if the translation and RNA processing machinery is nucleoid associated, as recent papers indeed have suggested [24,26]. However, as we have shown here, it is important to apply a systematic analysis, optimizing the condition of isolation, using various controls such as a tagged and mitochondrially targeted Luciferase, measuring both biological and technical repeats and applying stringent selection criteria in a comparative proteomics approach. This approach has for example shown that a number of proteins can be categorized as consistently enriched (Fig. 5), including many proteins that based on other research has pinpointed them as nucleoid-associated. Our approach has also shown that by a comparison of lysis conditions and sample handling (TX100 or X-ChIP), the X-ChIP method was the most sensitive and inclusive. Despite the fact that many more proteins were ‘nucleoid’ enriched compared to the TX100 method, the X-ChIP method showed enrichment of the highest percentage of mitochondrial proteins suggesting the method nevertheless is considerably more stringent than the TX100 method. Nucleoid associated proteins that only were identified using the X-ChIP method include, MGME1, DDX28, MTERF and MTERF2, Topoisomerase 3α, POLG2, TFB2M as
well as 50 mitoribosomal proteins and a considerable number of other proteins of mitochondrial gene expression. To immediately assign novel candidates that are likely core nucleoid proteins with a function in mtDNA metabolism is difficult on the basis of our results, but based on the fact that many mtDNA maintenance proteins are among the proteins identified suggests that various candidates with no current assigned role in mtDNA metabolism are present among the remaining proteins. Other isolation methods that more directly probe the interaction of proteins with mtDNA could in the future more specifically identify those proteins. If we examine the data from a more holistic point of view, we can expect that a number of proteins with an as yet unassigned role in mitochondrial gene expression, including RNA metabolism and translation, are amongst the enriched proteins. The analysis of proteins that are ≥2 fold reduced or completely absent in cross-linked TwinkleFLAG IAP from ρ° cells further identifies some of these candidates. These include 4 FAST kinase domain-containing proteins (S3 Table) that were recently also identified in a published RNA-binding proteome[45]. A recent analysis of 107 proteins with a possible function in mitochondrial RNA processing also identified FASTKD4 as being involved in mRNA stability[46]. Of the 107 proteins analyzed in this paper 47 are identified in our set of 366 proteins enriched in TwinkleFLAG cross-linked samples, while 34 of these 47 proteins are ≥ 2-fold reduced in TwinkleFLAG IAP from ρ° cells. Our data provide a valuable additional resource for identification of further mitochondrial RNA metabolism proteins. One possible example is methylcrotonoyl-CoA carboxylase, an enzyme involved in leucine breakdown and to our knowledge not previously identified as nucleoid associated. Interestingly, a second enzyme in the leucine breakdown pathway, enoyl-coenzyme A (CoA) hydratase with AUUU RNA binding activity (AUH), was recently shown to reside in the mitochondrial inner-membrane and matrix and possess a function in mitochondrial protein synthesis[47] and according Wolf and Mootha also has an RNA processing phenotype [46]. AUH in our dataset was specifically enriched in TwinkleFLAG IAPs but was equally enriched without or with cross-linking. Furthermore, it was not substantially decreased in TwinkleFLAG IAP from ρ° compared to IAP from mtDNA-containing TwinkleFLAG expressing cells, suggesting this protein might be one of several proteins that more specifically interacts directly with Twinkle or is part of a Twinkle-containing membrane platform. Apart from proteins with known functions that might have adopted additional functions, such as AUH, our dataset also contains several proteins of unknown function that might be worth investigating including von Willebrand factor A domain-containing protein 8 (VWA8) and Williams-Beuren syndrome chromosomal region 16 protein (WBSCR16), both of which have a very high mitochondrial localization prediction. WBSCR16 was, similar to the FASTKD proteins, also identified in recently published RNA binding proteomes as were many other known mitochondrial RNA binding proteins [45,48].

Few proteins have been shown to have a role in nucleoid membrane attachment. We showed recently that Twinkle organises replicating nucleoids to the inner mitochondrial membrane compartment and that Twinkle remains associated to the membrane in discrete foci in ρ° cells[17]. In other work Prohibitin (PHB) and ATAD3 have been isolated with nucleoids and been postulated to have an architectural role in nucleoids[21,24,49]. He and co-workers showed ATAD3 and PHB to co-sediment and co-purify with nucleoids and the mitochondrial translation machinery, postulating that ATAD3 links mitochondrial ribosomes to nucleoids and that both Prohibitin and ATAD3 link nucleoids to the inner mitochondrial membrane. This was recently further corroborated using complexome profiling, showing that a substantial number of proteins of the small ribosome subunit, ATAD3A and PHB1/2 co-migrate in Blue-native gels[34]. Although PHB1 and 2 did not pass our selection criteria, because they were also identified in TwinkleFLAG IPs without XL and were not sufficiently enriched in TwinkleFLAG +XL compared to mtLucFLAG +XL, their levels remained equal in TwinkleFLAG +XL.
IAP in \( \rho^- \) cells compared to mtDNA containing cells. ATAD3 also just failed to pass our selection criteria as it showed a <2 fold (1.93) increase comparing TwinkleFLAG +XL and TwinkleFLAG without XL. These results thus maintain the notion that these proteins could be part of a membrane anchor for a minimal mtDNA replication platform that includes Twinkle. Interestingly the X-ChIP method also identifies a number of ATP synthase subunits being enriched in TwinkleFLAG IAP while remaining constant or increasing in TwinkleFLAG IAP from \( \rho^- \) cells, in contrast to subunits of for example Complex I that were mostly \( \geq 2 \) fold reduced or absent.

A recent RNAi screen for proteins with a possible role in nucleoid organization and mtDNA maintenance in *Drosophila* identified most of the nuclear ATP synthase subunits[50]. Given the involvement of ATP synthase in mitochondrial membrane organization[51–53], the combined results suggest that ATP synthase could also be involved in the membrane organization of Twinkle containing complexes.

Previous nucleoid research has clearly pointed towards nucleoids being complex dynamic structures that have more functions than only being replication machineries. Bogenhagen et al [22] discusses the nucleoid structure to be layered and He et al[24] points towards an intimate relationship between nucleoids and the protein synthesis machinery, as also previously suggested by Iborra on the basis of fluorescent microscopy analysis[44]. In addition, in a recent paper Bogenhagen et al present evidence that initial RNA processing and ribosome assembly takes place in the close vicinity of nucleoids[26], whereas others have suggested that the entire small subunit of the mitochondrial ribosome is assembled at the nucleoid (see[25] and above [34]). This was further substantiated by a recent study that showed that failure to form the monosome prolongs the association of the 28S subunit with the nucleoid leading also to mtDNA aggregation[54]. Our comparison of proteins purified with TwinkleFLAG in HEK cells and their \( \rho^- \) counterparts indicates that some of the proteins suggested by Bogenhagen (in particular of the small ribosomal subunit; S7, S9 and S15) to associate with nucleoids to facilitate the early steps in ribosome biogenesis, to be less than 2-fold decreased suggesting these proteins might maintain a stable association with a minimal nucleoid structure also in the absence of mtDNA and RNA. A less than 2-fold decrease was also observed for some proteins that might facilitate ribosome biogenesis such as RNMTL1[55,56] and early steps in translation such as MTIF2.

To summarize, we here show that whole cell cross-linking in combination with IAP and appropriate lysis conditions enriches for mitochondrial nucleoids and associated proteins. This method is much less elaborate and complicated compared to previously published isolation protocols that include a formaldehyde cross-linking step. Whole cell cross-linking followed by IAP results in an inclusive list of enriched proteins that we show by the use of appropriate controls and cells lacking mtDNA to contain known and candidate mtDNA maintenance proteins and factors that are involved in mitochondrial gene expression. Our method and data therefore provide a valuable tool and resource for mitochondrial researchers. Our results add further weight to the idea that mtDNA nucleoids are an important organizing centre for mitochondrial biogenesis that might even include a local and specialized membrane organization in a ‘micro-compartment’, as recently suggested[57].

**Supporting Information**

**S1 Fig. Total depletion of mtDNA in \( \rho^- \) HEK293 Flp-In T-Rex cells.** DNA was extracted from cells, digested with *Pvu*II and imaged by exposure to Ultra-Violet (UV) light, or blotted and probed for mtDNA and exposed to a phosphor screen for two hours (2 hour) or 16 hours (16 hour) respectively. (A) 1 kb DNA Ladder. (B) Un-treated HEK293 Flp-In T-Rex cells total DNA. (C) Ethidium bromide-treated HEK293 Flp-In T-Rex cells total DNA at 95days. (D)
Ethidium bromide treated HEK293 Flp-In T-Rex total DNA at 116 days.

S1 Table. TX100 Biological repeats. Data file of the comparison of the three biological repeats measured on the LTQ-FT generated by MaxQuant. In sheet 1 ("RAW") the raw data MaxQuant analysis output is shown with two separate sheets showing the corresponding peptide count per protein (sheet 2; "Peptides") and the LFQ values with their calculated ratios across samples (sheet 3; "LFQ ratio"). Per experiment the ratios were calculated and shown with arrows if there was a change observed (green arrow up, ≥ 2 fold increase; yellow arrow horizontal, no change; red arrow down, ≥ 2 fold decrease). Whenever an increased was observed in at least two out of the three experiments, this was indicated with a green checkmark (instead of a red cross when this was not observed. Please note that all LFQ values of 0 have been replaced by 1E-12 to avoid division by 0.

S2 Table. TX100 Q Exactive triplicate. Data file of the comparison of the three technical repeats of the samples prepared with the TX100 method measured on the Q Exactive generated by MaxQuant. In sheet 1 ("RAW") the raw data MaxQuant analysis output is shown with two separate sheets showing the corresponding peptide count per protein (sheet 2; "Peptides") and the LFQ values with their calculated ratios across samples (sheet 3; "LFQ ratio"). To calculate the LFQ ratio, first the average is calculated from the three technical repeats. Whenever a protein was not identified in every single repeat, it was not considered (shown separately sorted on Twinkle-FLAG occurrence). Next to the ratio it is indicated if a change was observed (green arrow up, ≥ 2 fold increase; yellow arrow horizontal, no change; red arrow down, ≥ 2 fold decrease). Please note that all LFQ values of 0 have been replaced by 1E-12 to avoid division by 0.

S3 Table. X-ChIP method Q Exactive triplicate. Data file of the comparison of the three technical repeats of the samples prepared with the X-ChIP method measured on the Q Exactive generated by MaxQuant. In sheet 1 ("RAW") the raw data MaxQuant analysis output is shown with two separate sheets showing the corresponding peptide count per protein (sheet 2; "Peptides") and the LFQ values with their calculated ratios across samples (sheet 3; "LFQ ratio"). To calculate the LFQ ratio, first the average is calculated from the three technical repeats (for technical reasons TwinkleFLAG + XL is only represented by 2 repeated measurements). Whenever a protein was not identified in every single repeat, it was not considered (shown separately sorted on Twinkle-FLAG occurrence). Next to the ratio it is indicated if a change was observed (green arrow up, ≥ 2 fold increase; yellow arrow horizontal, no change; red arrow down, ≥ 2 fold decrease). The table is further sorted so that the enriched TwinkleFLAG + XL set of 366 proteins as indicated in Fig. 4 are listed first, further sorted by their level detected in Twinkle-FLAG + XL IAP from ρ° cells, as follows from top to bottom: 95 proteins not detected in ρ° cells, 163 proteins with a ≥ 2-fold decrease, 8 proteins with a ≥ 2-fold increase, and 100 proteins with no change (< than 2 fold change). Please note that for calculation purposes all LFQ values of 0 have been replaced by 1E-12 to avoid division by 0.

S4 Table. Datasets of enriched proteins including annotations. Enriched proteins from 4-sample Venn diagrams depicted in Figs. 2–4 and demarked by a red circumference are listed here in alphabetical order by Gene Symbol (first 3 sheets). These datasets were used to generate the area-proportional Venn diagrams in Figs. 3B and 4B and associated protein lists, comparing i) the 3 biological (Biol) repeats measured on and LTQ-FT (FT) with a triplicate measurement of sample 2 (Biol 2) of the three biological repeats measured on a Q Exactive (QE) mass
spectrometer (sheets: Biol FT & Biol 2 (TX100) QE, 97 proteins; Biol FT NOT QE, 71 proteins; Biol 2 FT NOT QE, 95 proteins) and ii) the Biological repeat 2 Triton X100 based method with the X-ChIP based purification method both measured on a Q Exactive instrument (sheets: TX100 & X-ChIP, 111 proteins; TX100 NOT X-ChIP, 81 proteins; X-ChIP NOT TX100, 255 proteins).

Acknowledgments

We thank Ian Holt for his critical reading of a draft of this manuscript, Prof. Ulrich Brandt for providing us measurement time on the Q Exactive Orbitrap mass spectrometer, Marijke Beenes for her expert assistance in the operation of this instrument and Merja Jokela for excellent technical help during the project.

Author Contributions

Conceived and designed the experiments: NR FH HJCTW JNS. Performed the experiments: NR FH DI. Analyzed the data: NR FH HJCTW DI JG JNS. Contributed reagents/materials/analysis tools: HJCTW JG. Wrote the paper: NR FH HJCTW JNS.

References


Whole Cell Cross-Linking Nucleoid Isolation


possible temporal nature of interactions of endogenous mtDNA replication factors has never been demonstrated. Although mtDNA–nucleoids in recent years have been presented as rather static, one might expect many nucleoid-associated proteins such as transcription, replication and repair factors to interact transiently with mtDNA depending on their requirement. This would be reminiscent of many factors that interact with, for example, nuclear DNA in both a spatial and temporal manner. We here set out to ask whether the same applies to mtDNA by examining mtDNA co-localization of two mtDNA replication factors with distinct function, namely Twinkle and mtSSB, and show that their association with mtDNA is indicative of active replication. We previously showed that Twinkle–GFP was present in discrete foci within the mitochondrial network even in the absence of mtDNA in p0 cells (5), which we here confirm for endogenous Twinkle. This observation provided us with a handle on the spatial organization of mtDNA replication within the mitochondrial network. We here provide evidence that Twinkle is firmly membrane associated, is one of the proteins of a membrane-associated replication factory and is at least partially involved in mtDNA membrane association.

MATERIALS AND METHODS

Cell culture

Stable cell lines expressing mtDNA maintenance proteins on induction were created as described (14) using the Flp-In™ T-Rex™ 293 host cell line (Invitrogen). The ATAD3-HA expressing cell line was a kind gift of Drs Ian Holt and Hiroshi Sembongi (Cambridge UK). Transgenic cells were grown in Dulbecco’s modified Eagle’s medium (DMEM; Lonza) supplemented with 10% FCS (PAAS laboratories), 2 mM l-glutamine, 1 mM Na pyruvate, 50 μg/ml uridine (Sigma), 100 μg/ml Hygromycin and 15 μg/ml Blasticidin (Invivogen) in a 37°C incubator at 8.5% CO₂. Normal HEK293E, U2OS, 143B, 206f and B2p cells were grown under similar conditions but without antibiotics. BJ (ATCC® CRL-2522™) human foreskin derived primary fibroblasts, and other primary human skin fibroblast lines were grown in 4:1 DMEM (Lonza) and M199 (Sigma) containing 15% FCS, 2 mM l-glutamine and 1 mM Na pyruvate. BJ fibroblast lines were used on the basis of availability and because these can be cultured to relatively high passage number without showing senescence, resulting also in no or only a relatively weak autofluorescence at 488 nm excitation. Other fibroblast lines were used on the basis of availability from our diagnostics service and were derived from healthy anonymous donors. These were not used with a passage number higher than 20. All cell lines were frequently checked for mycoplasma infection and found to be negative.

Western blot analysis

Mitochondrial fractions were analyzed by immunoblotting after sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) [(15) & Supplemental Experimental procedures].

Isolation of mitochondria

Cells were collected, resuspended in hypotonic buffer (4 mM Tris–HCl, pH 7.8, 2.5 mM NaCl, 0.5 mM MgCl₂ and protease inhibitor complete, Roche Molecular Biochemicals) and subjected to homogenization using a 5-ml chilled Dounce homogeniser until 80% cells were broken. During the testing phase of mitochondrial subfractionations (see below), cells were also disrupted after short cytochalasin treatment (16) and on occasion further purified using sucrose gradient purification as described (15) without noticeable differences in the final results (not shown). With both methods, mitochondria were isolated using differential centrifugation.

Mitochondrial (sub)fractionation

The mitochondrial outer membrane was disrupted by incubation with a digitonin (Sigma Aldrich)/protein ratio ([μg digitonin]/[μg mitochondrial]) = 0.2 (unless otherwise indicated) in phosphate buffered saline (PBS) or a buffer containing 225 mM Mannitol, 75 mM sucrose, 10 mM HEPES, pH 7.8, 10 mM EDTA, in either case supplemented with a protease inhibitor. The mitoplasts were obtained by centrifugation at 8000g for 10 min, +4°C. The supernatant was centrifuged at 100 000g for 1 h to obtain intermembrane space supernatant and pellet containing a fraction of outer mitochondrial membrane proteins (see Supplementary Figure S3 and Results). Mitoplasts were suspended in 0.16 mg of Brij58/mg mitoplasts and incubated for 10 min on ice. Membrane (inner + outer) (pellet) and matrix (supernatant) fractions were obtained after centrifugation at 100 000g for 1 h. Proteins from intermembrane space and matrix were precipitated by deoxycholate (DOC)/trichloroacetic acid (TCA) (see below). Equivalent protein concentrations were run on gel for western blot analysis of the various fractions (Supplementary Figure S6).

For digitonin-based fractionation, crude mitochondria from HEK293E or inducible HEK293 Flp-In™ T-Rex™ wt-Twinkle cells were taken up in 1 × PBS (Gibco). The total protein concentration determined with Bradford assays and lysed by addition of digitonin at indicated ratios μg digitonin/μg total mt protein, incubated for 10 min on ice and centrifuged for 5 min at 14000g and 4°C. Solubilized supernatant fractions were kept separately while pellet fractions were resuspended in volumes equal to the removed soluble fractions. Both supernatant and pellet fractions were brought to a final concentration of 1% SDS.

Crude mitochondria for flotation were further purified over 30% Percoll gradients (30% Percoll, 225 mM sorbitol, 25 mM Tris, pH 7.4, 1 mM EGTA). Purified mitochondria were washed once in 5 volumes of 225 mM sorbitol, 25 mM Tris, pH 7.4, 1 mM EGTA and taken up in 1× PBS. Total mitochondrial protein yield was determined by Bradford assays and the equivalent of 2 mg of total mitochondrial protein was lysed in TN (25 mM Tris–HCl, pH 7.5, 150 mM NaCl, 1 mM DTT,
cocktail of protease inhibitors, 10% sucrose) containing either 1% Triton X-100 or digitonin at a ratio of 2.5:1 (w/w) for 30 min on ice. Digitonin-lysed samples were centrifuged for 10 min at 14000g, the supernatant discarded and the pellet resuspended in TN containing 1% Triton X-100. Samples were mixed with cold Optiprep™ to a final concentration of 42.5%, transferred into MLS-55 centrifuge tubes and overlaid with 400 μl of each 40, 37.5, 35, 32.5, 30, 27.5, 25, 20, and 0% Optiprep™ in TN containing 1% Triton X-100. The gradients were centrifuged at 100000g for 14 h at 4°C. Fractions were collected from top to bottom and aliquots analyzed by western blotting or dot blotting, respectively, as described before.

**Treatment of isolated mitochondria with carbonate or KCl**

For carbonate extraction, isolated mitochondria were resuspended in a 0.1 M Na2CO3 buffer (pH 11.0) and incubated on ice for 30 min; the pellet was recovered by centrifugation (100000g, 1 h, 4°C). For salt-wash experiments, mitochondria were diluted 10-fold in buffers consisting of either 30 mM KCl or 500 mM KCl in 30 mM Tris–HCl (pH 7.4) and sonicated at 40% power 3 mm probe 3×10 s per cycle. The pellet was recovered by centrifugation (100000g, 1 h, 4°C). Proteins from the resulting supernatants were concentrated by DOC/TCA precipitation: lysates were treated with 0.02% DOC for 30 min on ice before addition of 10% TCA, incubated at +4°C over night and precipitated samples were centrifuged at 15000g for 15 min at +4°C. Pellet and precipitated supernatant were finally re-solubilized in equal volumes and the same volume loaded on gel for SDS-PAGE and western blot analysis.

**Treatment of isolated mitochondria by sonication and nucleases**

Mitochondria were resuspended in enzyme-buffer (50 mM Tris–HCl, pH 7.5, 50 mM NaCl, 3 mM CaCl2, 2 mM MgCl2), sonicated on ice at 40% power for three times 20 s before addition of the enzymes as indicated DNase I (Fermentas) 10U, RNase A (Fermentas) 20 μg, Micrococcal nuclease (Fermentas) 50 U and Benzonase nuclease (Sigma) 50 U, and incubated at +37°C for 30 min. Where appropriate, lysates were further subjected to carbonate extraction as described above.

**Dot-blot analysis of mtDNA content**

For mtDNA analyses, samples of supernatant and pellet lysates [see Mitochondrial (sub)fractionation] were suspended in 2× SSC, boiled for 15 min at 95°C and dot blotted in triplicates onto positively charged nylon membranes. Dot blots were detected using nonradioactively labeled cyt b probes using Dig-labeling (Roche). Hybridizations at 48°C and dig-antibody incubations were carried out using Easy-Hyb (Roche) according to the manufacturer’s protocol. ECL detection was performed with CSPD (Roche) and visualized with a ChemiDoc (Biorad). Quantifications of resulting ECL signals were performed with ImageQuant (GE Healthcare).

**Transfections, fluorescence microscopy, ddC treatment, EdU and BrdU labeling**

Immunofluorescence (IF) detection of proteins was done as described previously (8) with minor modifications (for detailed procedures, see Supplement). MtDNA depletion in U2OS cells used 100 μM 2′,3′-dideoxycytidine ddC for 48 h. For Twinkle knockdown, cells were transfected in six-well plates (for IF) or 10-cm cell culture dishes (for biochemical fractionation experiments) with a mixture of three Stealth™ siRNA duplex oligonucleotides (C10Orf2 HSS125596, HSS125597, HSS125598, Invitrogen) against Twinkle, at a concentration of 20 pmol each, using Lipofectamine™2000. As controls we used Stealth™ Universal negative controls. Cells were fixed and analyzed 36–72 h after transfection. Transient transfection of a Twinkle–Myc expression construct (5) used TransIT-LT1 (Mirus, Madison, WI) according to the manufacturer’s instruction. Twinkle knockdown followed by biochemical fractionation involved a short exposure to Lipofectamine™2000 (4 h) after which medium was replaced with regular cell culture medium and replaced again 24 h prior to cell isolation.

MtDNA labeling using Click-iTTM EdU (5-ethyl-2′-deoxyuridine) imaging kits with either AlexaFluor 488 or 568 azide (Invitrogen) was initially done as described by the manufacturer except that we used 50 μM EdU to detect mtDNA label incorporation. Only for the experiment shown in Figure 4C examining EdU, Twinkle, mtSSB co-localization we modified the procedure: the Click-iTTM buffer additive was replaced by 50 mM ascorbic acid and the reaction was done twice for 25 min with a freshly prepared labeling mix. This increased the signal and signal-to-noise ratio (17). Following EdU labeling and detection, we proceeded with IF as above. 5-bromo-2′-deoxyuridine (BrdU) labeling used the BrdU Labeling and Detection Kit I (Roche) using manufacturers protocols except that we used 50 μM BrdU and Alexa 568 anti-mouse for BrdU-antibody detection. Fixation used acid-ethanol resulting in significantly reduced mtSSB antibody staining compared to paraformaldehyde fixation. Both for EdU and BrdU labeling, mtDNA/EdU (or mtDNA/BrdU) and EdU/Twinkle (or BrdU/mtSSB) positive foci were scored manually by first marking all mtDNA/EdU or BrdU foci [using Image Pro Plus 6 ’create point feature‘ (Media Cybernetics) or using the ‘Event Marker’ tool using Axiovision 4.8 software] and overlaying the Twinkle/mtSSB IF, counting all double positives and using both numbers to calculate relative percentages. Twinkle–mtDNA positive foci were similarly scored using Image Pro Plus 6. In all cases, experiments were repeated several times as indicated and in each experiment multiple cells were scored to obtain final numbers. Very rare cytoplasmic EdU or BrdU spots that did not appear to co-localize with mtDNA were not considered. MtSSB/ BrdU positive foci were only judged positive with clear position overlap and a distinct focal mtSSB signal on the basis of the strong focal mtSSB presence in a subpopulation of mtDNA foci in paraformaldehyde fixed cells. BrdU or EdU foci in the vicinity of the
nucleus could not be assigned positive for mtDNA and/or BrdU/Edu on the basis of the often strong nuclear signal and were therefore not used in the quantification. Intensity line scans were made with the ‘Profile tool’ (Axiovision 4.8 software).

RESULTS

Endogenous Twinkle and mtSSB at steady state are found in a subset of mtDNA–nucleoids and are enriched in replicating mtDNA foci

The mtDNA helicase Twinkle is a low abundant protein [see Supplement of (14)]. For this reason and the lack of a good antibody, the analysis of the cellular functions of Twinkle so far has used overexpression of Twinkle variants that contain C-terminal epitope tags. These analyses have shown a high degree of in situ co-localization of Twinkle with mtDNA and mtDNA-associated proteins such as TFAM (8). However, overexpression of tagged Twinkle might not accurately copy the properties of the endogenous protein, while at the same time the total mitochondrial pool of the protein is considerably increased. Since Twinkle and, for example, also the mtSSB are mitochondrial proteins considered to be essential for mtDNA replication, but each mtDNA molecule is not continuously replicated, we asked whether either of these proteins dynamically associates with mtDNA. To this end we tested antibodies for Twinkle in immunofluorescence (IF) using immortal cell lines and primary fibroblasts. This analysis identified two monoclonal antibodies both recognizing the C-terminus of Twinkle based on peptide mapping (Anu Suomalainen-Wartiovaara & Milla Lampinen, personal communication) able to detect endogenous (Figure 1 and 2 and Supplementary Figure S1A and C) as well as overexpressed Twinkle (Supplementary Figure S1B), while siRNA mediated depletion of Twinkle showed loss of mitochondrial antibody signal (see below). Overexpression of Myc-tagged Twinkle showed that the antibody recognized all Twinkle–Myc/mtDNA foci as the Twinkle monoclonal and a Myc polyclonal antibody showed perfect overlap of signals (Supplementary Figure S1B).

Analysis of IF images revealed that endogenous Twinkle co-localized only with a subset of mtDNA–nucleoids based on co-staining for mtDNA and/or TFAM (Figure 1A and B) and Figure 1C for osteosarcoma 206f cells, a result that was also observed using a different fixation and permeabilization method (Supplementary Figure S1A). Similar observations were made in HEK293E cells (Supplementary Figure S1C). An analysis in primary fibroblasts of between 9 and 12 cells in three independent experiments showed the percentage of mtDNA foci that were Twinkle positive to be 48 ± 7% (Figure 1B1). In all, 1.2 ± 0.25 of every 10 Twinkle foci did not appear to co-localize either with mtDNA or TFAM while all mtDNA foci were TFAM positive and vice versa, no TFAM foci were observed that were not also mtDNA positive. These results, apart from showing that less than half of all mtDNA/TFAM foci contained Twinkle, suggested that Twinkle might organize in discrete structures independent of the presence of mtDNA. Earlier we showed that Twinkle–GFP expressed in mtDNA-less (ρ−) cells also shows a punctate mitochondrial fluorescence (5). This was corroborated here for endogenous Twinkle in cells partially depleted of mtDNA using dideoxycytidine (ddC) still showing discrete Twinkle foci both, with and without, mtDNA co-localization (Figure 2A). In 206f–ρ− cells, derived from the 143B osteosarcoma cell line, Twinkle foci are again discrete in an otherwise connected mitochondrial network (Figure 2B1 and B2). In 206f cells (Figure 2B2), endogenous mtSSB was more uniformly distributed over the mitochondrial network as also shown previously (8). This is illustrated by an intensity line profile of a small section of the mitochondrial network showing fluctuating mtSSB intensity at levels well above the baseline, whereas Twinkle foci appear as sharper peaks from the baseline (Figure 2B2). TFAM staining in 206f cells (Supplementary Figure S2A2) showed a much weaker and more uniform signal in comparison to the signal detected in the parental 143B cells (Supplementary Figure S2A1), as previously observed (18).

In fibroblasts (Figure 1B1, details in Figure 1B2 and B3) and U2OS cells (Figure 1C1, details in Figure C2 and C3), mtSSB showed both a relatively uniform mitochondrial distribution as well as a focal accumulation of presumably higher concentrations of mtSSB. These intense mtSSB foci co-localized only with a subset of mtDNA foci at a substantially lower percentage (14.8% in fibroblasts) compared with Twinkle. This was observed with two polyclonal mtSSB antibodies that we used in the course of this study. Cells in which Twinkle was depleted using transient (48 h) siRNA transfection only showed a modest decline of detectable mtDNA foci but showed almost complete loss of intense mtSSB foci (Figure 3). This was observed both in U2OS (panels A) and primary skin fibroblasts (panels B1). Detailed images (panels B2) of a small section of control versus knockdown Twinkle fibroblasts show the loss of most mitochondrial Twinkle foci (8% of mtDNA foci are still Twinkle positive compared with 30% in the control cell) while background antibody staining appearing in the cytosol remains similar to that seen in the control. In the whole-cell image (panel B1, upper right), about 10 intense mtSSB foci remain visible with Twinkle knockdown (compared with on average ~70 in the control), while in the detailed section, 0 remain, compared with 11% in a similar section shown for the control. Similar to U2OS (panels A), various transient knockdown experiments using multiple control fibroblasts lines showed an on-slide correlation between the effectiveness of knockdown in individual cells based on Twinkle IF and the level of loss of intense mtSSB foci (not shown). In further support of these findings, transient expression of strong Twinkle stalling mutants K421A and G575D (14) result in the disappearance of mtSSB foci while these foci remain with expression of wild-type Twinkle or in cells on the same slide in which mutant Twinkle is not expressed (Supplementary Figure S3). In contrast to these findings, a short (48 h) but stronger mtDNA depletion using ddC (Figure 2A and Supplementary Figure S4B) showed a...
Figure 1. Endogenous Twinkle is not a constitutive nucleoid protein. (A1) BJ fibroblasts were stained with a mouse IgG monoclonal antibody for Twinkle (green), a mouse IgM monoclonal for mtDNA (white) and a rabbit TFAM antibody (red) and imaged by confocal microscopy. Detailed (A2) and merged (A3) images show that all mtDNA foci were TFAM positive, while \( \sim50\% \) (see also ‘Results’ section) of mtDNA foci were not positive for Twinkle (some of these foci are indicated with white arrows at the inset). In addition, some Twinkle foci were observed also in the absence of mtDNA (foci indicated with green arrows). (B1) A second primary skin fibroblast line was stained with antibodies for Twinkle (green), mtDNA (white) (same as above) and a rabbit mtSSB polyclonal (red) and imaged in this case using a Zeiss apotome. Detailed (B2) and merged (B3) images show that, similar to BJ fibroblasts, \(<50\%\) of mtDNA foci were positive for Twinkle, while even fewer mtDNA foci were strongly positive for mtSSB against a weaker more uniform mtSSB staining that nevertheless appears to show some preferential localization with mtDNA. The percentage of mtDNA foci positive for Twinkle or mtSSB was determined in three independent experiments in primary skin fibroblasts, showing only a partial co-localization (see Main text). This is here presented as a small graph in panel B1 (right). The percentage of mtDNA foci positive for TFAM is here set at 100\% as we have never seen evidence of any mtDNA foci not also showing a positive TFAM signal. (C1) Similarly, U2OS cells were stained with antibodies (same as above) for Twinkle (green), mtDNA (white) and mtSSB (red) showing multiple mtDNA foci not containing Twinkle and/or high concentrations of mtSSB (detailed in C2 and C3). Scale bars in figures are 10\( \mu \)m. For additional control experiments see Supplementary Figure S1.
Figure 2. Twinkle is found in punctate foci within the mitochondrial network in cells depleted of mtDNA and p0 cells. (A1) Treatment of U2OS cells with ddC resulted in a severe drop of mtDNA positive foci and revealed multiple Twinkle foci in the absence of mtDNA (detailed in A2 and A3). (B) In mtDNA-less 206f cells, endogenous Twinkle also showed a focal staining, while staining for mtSSB was comparatively uniform, in contrast to these proteins in cells containing mtDNA (8). 206f cells are here also stained for DNA showing a clearly positive nuclear DNA signal but absence of any mtDNA signal confirming the mtDNA-less character of these cells. A detailed section (indicated with a white box) of the merged Twinkle–mtSSB images clearly shows the rather uniform character of the mtSSB staining while Twinkle staining is punctate. This is further illustrated by a profile line scan (region indicated in the ‘Detail merge’ image by a white line) that shows fluctuating mtSSB intensity at levels well above the baseline, whereas Twinkle foci appear as sharper peaks from the baseline.
Figure 3. Twinkle depletion results in a modest decline of mtDNA but a dramatic loss of focal mtSSB. (A) Depletion of Twinkle in U2OS cells using transient (48 h) siRNA transfection. In this image view, several cells with Twinkle knockdown (−) are shown alongside two cells that show unsuccessful knockdown (+) to illustrate the considerable reduction of Twinkle immunofluorescence in knockdown cells and, therefore, showing the specificity of the Twinkle antibody used in this study. Cells with loss of Twinkle immunofluorescence, also show loss of mtSSB at mtDNA foci but only a modest reduction in mtDNA signal. (B) In fibroblasts (fibro), similarly, knockdown of Twinkle after 48 h still showed substantial amounts of mtDNA foci while Twinkle was dramatically reduced compared with a control knockdown experiment using nontargeting siRNA (for a quantification of the indicated section in the periphery of the cell, numbers are indicated on the right in panel (B2) Tw = Twinkle; pos = positive), while only a few intense mtSSB foci remained (10 in the whole cell, none in the detailed section). For clarity we here show detailed sections (regions indicated with a white square in the larger mtSSB image) for the three different antibodies (smaller panels B1) as well as enlarged merged images for both control and Twinkle knockdown (B2). These illustrate the presence of mtDNA foci in Twinkle knockdown cells, the absence of strong mtSSB foci co-localizing with mtDNA and Twinkle, and the loss of mitochondrial Twinkle foci co-localizing with mtDNA, which was quantified as 8% compared with 30% in the section analyzed in the control. The enlarged merged images clearly also illustrate that the knockdown looks deceptive as remaining fluorescence is still clearly visible. Most of this fluorescence, however, is nonmitochondrial background, which is also visible in the control, but as shown here, is not sensitive to Twinkle knockdown. For additional control experiments see Supplementary Figure S3.
The absence of endogenous Twinkle from many mtDNA–nucleoids in primary fibroblasts and the presence of Twinkle foci in the absence of mtDNA suggested that nucleoids might dynamically associate with Twinkle foci (or vice versa) dependent on signals that would indicate the need to replicate mtDNA. Even more dramatically, only a small fraction of mtDNA foci in fibroblasts showed a strong accumulation of mtSSB. To address whether Twinkle and/or mtSSB association with mtDNA showed a positive correlation with ongoing mtDNA replication, we made use of ClickIt–EdU (19) and BrdU labeling to detect de novo mtDNA synthesis. Since mtDNA can incorporate EdU/BrdU at any point during ongoing mtDNA replication, we reasoned that if Twinkle and/or mtSSB temporarily associate with nucleoids to enable replication and would subsequently dissociate or disassemble, short EdU/BrdU pulses would show a relatively larger proportion of labeled nucleoids positive for these proteins, than longer pulses. In initial experiments, we tested Twinkle and mtSSB separately with EdU and BrdU labeling, respectively (Figure 4A and B). The data (Figure 4C and Supplementary Figure S5) indicate that with the shortest possible EdU pulses for unambiguous visualization (30 min), the highest proportion of EdU positive foci were Twinkle positive (73%), while after 60 and 90 min and at steady state (see above and Figure 4C) this percentage was significantly lower, indicating that for EdU incorporation to occur Twinkle needs to be mtDNA associated. Similarly, following BrdU labeling and mtSSB/mtDNA, IF showed 69% strong mtSSB positive BrdU foci after a 30-min pulse that dropped to 41% after 90 min and declined even further to 27% after a 90-min pulse followed by a 60-min chase. These data thus show that core components of the replication machinery and mtDNA dynamically associate with one another to enable replication.

To analyze the dynamics of mtDNA association of both Twinkle and mtSSB in more detail, we examined Twinkle/mtSSB co-localization with EdU-labeled mtDNA. At the same time, parallel slides were processed for mtDNA, Twinkle and mtSSB detection under otherwise identical conditions in order to determine overall levels of mtDNA occupancy by Twinkle and mtSSB. As mtDNA replication occurs throughout the cell cycle, these values represent steady-state co-localization values. Since EdU foci were essentially all mtDNA positive (see Supplementary Figure S5), we chose not to perform EdU labeling with quadruple staining including mtDNA staining because axial chromatic aberration in the ultraviolet range could not be corrected for on the microscope used in these experiments.

The results of the EdU–Twinkle–mtSSB detection (Figure 4C) show the same general kinetics as observed with EdU–Twinkle–mtDNA (Supplementary Figure S5) and BrdU–mtSSB–mtDNA detection as depicted in Figure 4B. In this particular experiment, steady-state Twinkle–mtDNA and mtSSB–mtDNA co-localization were 39 and 14%, respectively (Figure 4C). In contrast, at a short 30-min EdU pulse labeling, Twinkle–EdU and mtSSB–EdU co-localization were 64 and 66%, respectively, and the percentage of EdU–mtSSB foci that were positive for Twinkle at this short pulse was 72%. With longer EdU pulses, co-localization percentages declined to more closely reflect steady-state Twinkle and mtSSB, mtDNA co-localization. At a 90-min pulse plus a 60-min chase, Twinkle–EdU co-localization was somewhat below the steady-state Twinkle–mtDNA co-localization while the mtSSB–EdU co-localization was still somewhat above. The percentage of mtSSB–EdU positive foci that were also Twinkle positive showed slightly different kinetics such that at a 30-min pulse it closely reflected the steady-state level of 73%, but at all later time points, it settled at a relatively stable lower-than-steady-state percentage of 52–55% (see ‘Discussion’ section).

Twinkle is firmly membrane associated, enhances mtDNA tethering to the membrane when overexpressed and reduces mtDNA tethering on Twinkle knockdown

Although the first suggestions of mammalian mtDNA membrane association stem from the late 1960s and 1970s (see ‘Introduction’ section), little is known about the nature of this association. The observation that Twinkle forms discrete foci even in the absence of mtDNA raised the possibility that Twinkle is not a matrix-soluble protein, as this would show a uniform staining. In contrast, TFAM and mtSSB lose their localization in discrete foci in ρ− cells and thus require mtDNA for their localization in a discrete complex (see above). We used classical biochemical fractionation to examine the localization and solubility of nucleoid-associated proteins (see legend to Supplementary Figure S6 and ‘Materials and Methods’ section for details). We first used inducible overexpression of tagged proteins, as this required much less material and allowed their detection with tag antibodies. This showed that Twinkle was exclusively present in the mitochondrial membrane fraction as was ATAD3, an IM protein with functions reported in lipid shuttling and mtDNA binding (20,21). In contrast TFAM, mtSSB and POLG1 and 2 were distributed over the membrane and matrix fractions (Supplementary Figure S6). Although Twinkle does not contain any predicted transmembrane helices, these data suggested a strong membrane association. To confirm this, we used two alternative methods to confirm these results for endogenous Twinkle, based either on sodium carbonate fractionation (22) or on 0.5 M KCl extraction (23) (Figure 5A). Both of these methods showed tight association of endogenous Twinkle with mitochondrial membranes. As expected, the same was also observed with overexpressed Twinkle (Figure 5B and C and Supplementary Figure S6). As a control, a combination
Figure 4. Twinkle and mtSSB are enriched in mtDNA foci showing de novo mtDNA synthesis. (A) Fibroblasts were labeled for the indicated times with BrdU and processed for BrdU, mtSSB and mtDNA detection. (B) Fibroblasts were labeled for the indicated times with EdU (also see Supplementary Figure S5) or BrdU and processed for EdU/BrdU, Twinkle/mtSSB and mtDNA detection. The graph shows the relative percentage of Edu/Twinkle and BrdU/mtSSB positive foci, with the 30' time point set to 100%. In reality, this time point showed 73 ± 8% of all EdU foci to be Twinkle positive and 69 ± 8% of all BrdU foci to be mtSSB positive, both significantly higher not only compared with the number of Twinkle positive EdU foci at 60' and 90' or mtSSB positive BrdU foci at 90' or at 90' pulse (p)+60' chase (ch) (paired t-test) but also to the steady-state percentage of Twinkle-positive or mtSSB-positive mtDNA foci (see ‘Results’ section and panel C). Error bars show SD. (C) Fibroblasts were labeled for the indicated times for EdU and slides processed for EdU detection using Alexa Fluor 555, Twinkle detection using Alexa Fluor 488 and mtSSB detection using Alexa Fluor 647 and co-localization determined. At the same time, a parallel slide from the same six-well plate was processed for Twinkle, mtSSB and mtDNA detection to obtain steady-state co-localization of Twinkle and/or mtSSB with mtDNA. For each slide, 10 images were taken and co-localization percentages determined, the bars indicate the upper and lower limits of these percentages, i.e. the range, for each experiment. Numbers in each bar show the average percentage of each experiment.
Figure 5. Twinkle is membrane associated. (A) Isolated mitochondria of HEK293E cells were subjected to either KCl or sodium carbonate extraction (Na₂CO₃) as described in the main text. Endogenous Twinkle was detected using a monoclonal antibody and blots were re-probed with antibodies for TFAM and mtSSB. Results show that endogenous Twinkle fractionates mostly to the pellet fraction using both methods illustrating its strong membrane association, similar to overexpressed Twinkle (Supplementary Figure S6C), whereas TFAM and mtSSB mostly became soluble, in particular, in combination with 0.5M KCl and sonication. Please note that to detect Twinkle with confidence, more protein was used for these western blots sometimes resulting in overloading of TFAM. Sonication in combination with DNAseI (D) or DNaseI/RNase A (R, RNase)/Benzonase (B) released more TFAM and mtSSB than sonication alone, showing that a proportion of TFAM and mtSSB can be found in the insoluble fractions of the various experiments by means of their interaction with mtDNA (rightmost blot panel A). (B) Na₂CO₃ fractionation shows that overexpressed Twinkle-Myc is almost exclusively in the pellet (p) fraction again indicative of tight membrane association, whereas TFAM is
of TritonX-100 lysis and sodium carbonate extraction showed that the insolubility is not a peculiarity of the Twinkle protein or its overexpression (Figure 5B), as Twinkle is mostly found in the supernatant under these conditions. In contrast to Twinkle, and depending to some extent on the cell line used, substantial proportions of endogenous TFAM and mtSSB could be dissociated using sodium carbonate or 0.5 M KCl extraction. This suggested that the fraction of TFAM or mtSSB that was membrane associated on the basis of the more classical mitochondrial supercomplexes (24).

On the basis of our digitonin titration experiments (Supplementary Figure S6) it was clear that also the IM could be disrupted by high digitonin concentrations. We reasoned that if many nucleoid-associated proteins are found in both membrane and matrix fractions based on classical mitochondrial subfractionation and only a subset of nucleoids contain Twinkle based on IF, perhaps two pools of mtDNA–protein complexes could also be separated on the basis of their solubility. To test this, we examined the solubility of several mitochondrial marker proteins by titrating the w/w ratio of digitonin/total mt protein, but this time using only one centrifugation step to separate the solubilized components (supernatant) from the digitonin-insoluble (pellet) fraction. This analysis (Supplementary Figure S6D) showed that the IM became somewhat permeable to glutamate dehydrogenase (a matrix localized enzyme) at a 0.5:1 w/w digitonin/protein ratio and was maximally permeable at a 2.5:1 ratio. At this ratio, COX II was still mostly in the pellet fraction, but at a 3:5:1 ratio it also became more soluble. This agrees with supercomplex blue-native PAGE analysis protocols, where a 4:1 ratio is used to solubilize supercomplexes (24).

TFAM and POLG1 behaved similar to glutamate dehydrogenase to the extent that a sizeable proportion became soluble at the lower range of digitonin concentrations. However, a substantial pool was resistant to digitonin solubilization at a 2:5:1 and even a 3:5:1 ratio, as it remained in the pellet. As expected, endogenous Twinkle behaved essentially the same as the transmembrane COX II protein in this assay. Using a 2:5:1 digitonin ratio, we subjected pellet fractions to a combined flotation/fractionation on an iodixanol gradient having re-solubilized the pellet with Triton-X100 and compared this with a total mitochondrial Triton-X100 lysate (Figure 5D). This analysis showed that Twinkle, mtDNA, TFAM, mtSSB and POLG1 all migrated in a single fraction high up the gradient, while, in contrast, for example, COXII or a marker for the large mitoribosomal subunit, MRPL49, migrated only a small distance up the gradient. The small mitoribosomal subunit marker MRPS22 showed the presence of this subunit in the total lysate but it was essentially absent from the digitonin pellet fraction. These results thus showed again that the Twinkle-containing fraction is not an insoluble aggregate of proteins. More importantly, it indicated various proteins that are involved in mtDNA replication and that were insoluble at a 2:5:1 digitonin ratio to co-migrate in a single fraction at low density in the gradient substantiating the idea that they form a single discrete membrane-associated complex (see ‘Discussion’ section).

Having established the relatively simple fractionation procedure based on digitonin lysis, dot-blot analysis was performed for mtDNA and showed that ~35% of the mtDNA pool was soluble at a 2:5:1 digitonin ratio (see below, Figure 6). This suggested that on mild mitochondrial lysis with digitonin, two pools of mtDNA and associated proteins exist, one that remains in the pellet and contains Twinkle and one that is more soluble and contains little Twinkle.

Overexpressed Twinkle in cultured cells typically shows a good co-localization with mtDNA (8) and (Supplementary Figure S1B). Given this and the results presented so far we hypothesized that overexpression of Twinkle might increase the fraction of mtDNA in the insoluble pellet fraction. To test this hypothesis, we used inducible expression of wt Twinkle without epitope-tag and tested the distribution of mtDNA and associated proteins using digitonin fractionation. Previously we have shown that the used level of Twinkle induction has little effect on nucleoid structure and has no effect on mtDNA levels, replication, mitochondrial transcript levels or cell growth (25). The results of western- and dot-blot DNA analysis showed that, in contrast to
Figure 6. MtDNA membrane association varies with overexpression or knockdown of Twinkle. (A and B) Mitochondria were isolated from noninduced HEK293 FlpIn™ TRex™, or induced as indicated to express Twinkle. Isolated mitochondria were subsequently lysed under mild conditions (see Supplementary Figure S6 panel D and ‘Results’ section) to release matrix constituents (s) but not membrane components (such as cytochrome c oxidase subunit II—COXII) found in the pellet (p) on centrifugation. Each fraction, including nonlysed mitochondria and their wash
noninduced cells, Twinkle overexpression resulted in almost complete retention of mtDNA and a reproducible redistribution of TFAM to the nonsoluble fraction, in particular, for the 3-day induction period (Figure 6B). The analysis of POLG showed a similar redistribution as TFAM on the blot shown here but was not always as clearly detectable owing to its low abundance, precluding a statistical analysis.

Finally, we performed the same analysis using Twinkle siRNA (Figure 6C). This showed an increase in the amount of soluble mtDNA. Nevertheless, a substantial amount still remained insoluble showing that mtDNA membrane association is not solely dependent on Twinkle (see ‘Discussion’ section). In agreement with the suggested requirement for Twinkle, mtSSB showed a redistribution to the more soluble fraction, corroborating the IF analysis following Twinkle knockdown.

**DISCUSSION**

In this article we show, using IF and biochemical fractionation, the presence of at least two pools of mtDNA in human mitochondria, one (or more) that is isolated in an insoluble fraction and is likely membrane associated, and one that is more soluble. Similarly, we show that Twinkle behaves as a membrane protein independent of DNA association, whereas mtSSB and TFAM are present with mtDNA and Twinkle in membrane fractions (by biochemical isolation) or with nucleoids (by IF) on the basis of their DNA association and ongoing replication. This is corroborated by IF in cells not containing mtDNA, showing Twinkle in punctate foci, whereas mtSSB and TFAM are more uniformly labeling the mitochondrial network in contrast to their partial punctate co-localization in cells that contain mtDNA. We show, using EdU/BrdU pulse-chase labeling, a clear causative relationship between the presence of Twinkle and mtSSB at mtDNA foci and concurrent mtDNA replication. Finally, using overexpression or knockdown we show a partial dependency of mtDNA membrane association on Twinkle. These findings and the presence in a single fraction of Twinkle, mtDNA and various mtDNA replication factors on flotation of digitonin purified mitochondrial membranes suggest that Twinkle is a core component of a membrane-associated mtDNA replication factory. We thus provide at least a partial explanation for the long-standing observation of an mtDNA membrane connection (see ‘Introduction’ section). Supplementary Figure S7 shows a model that incorporates the major findings of this article.

**Twinkle and mtSSB preferentially co-localize with replicating mtDNA; mtSSB is recruited to the replisome in a Twinkle-dependent manner**

The steady-state percentages of Twinkle and mtSSB co-localization with mtDNA and the comparison of their co-localization with replicating mtDNA provides insight both into the mechanism of mtDNA replication as well as its dynamics and that of its associated factors. It is clear from the data that Twinkle at steady state co-localizes with a higher number of mtDNA molecules than the number that is being replicated at any given time, as the number of mtDNA foci positive for Twinkle (~40–50%) is much higher than the number of mtDNA foci that are positive for EdU/BrdU (4–10%) at the shortest labeling time (Supplementary Figure S8). This suggests that mtDNA–Twinkle association in itself is not enough to initiate replication or alternatively that replication following this association is frequently aborted or prematurely terminated. This hypothesis is supported by the observation that at steady state and following a short 30-min EdU pulse, the percentage of mtDNA molecules being positive for mtSSB and also positive for Twinkle was highest (~70%), whereas at all the other time points of EdU labeling it seemed to have settled at ~55%. MtDNA copying also involves the frequent synthesis of 7S DNA in the so-called noncoding region. Although the function of this relatively short DNA fragment is currently still unclear, it has a higher synthesis and turnover rate compared with the full-length genome [see e.g. (26) and references herein], which depends on the recently identified mitochondrial genome maintenance exonuclease 1 (MGME1) (27). 7S DNA synthesis might thus provide an explanation for the observed difference, and it was recently shown that 7S DNA synthesis is dependent on the Twinkle protein (28). Alternatively, results could indicate, for example, DNA repair processes, with limited EdU incorporation that would fall below the detection limit but require Twinkle mtSSB association with mtDNA.

The much lower steady-state mtSSB–mtDNA co-localization and the concomitant low percentage of mtSSB that does not co-localize with EdU/BrdU at the shortest labeling pulses shows that mtSSB is a better in situ marker for ongoing full-genome replication than Twinkle.
This is supported by siRNA-mediated Twinkle depletion: a 48-h Twinkle knockdown in fibroblasts (as shown in Figure 3) shows a modest reduction in visible mtDNA foci, which agrees with previous siRNA experiments showing a slow decline in mtDNA copy number (29). At the same time, few intense mtSSB positive foci, as seen in untreated cells, remain. Biochemical fractionation shows similarly that mtSSB becomes more soluble on Twinkle siRNA. In contrast, cells that are more strongly depleted of mtDNA following ddC treatment show mtDNA foci containing Twinkle and strongly positive for mtSSB by IF. The use of Twinkle siRNA allows us to conclude that there is a correlation between the level of Twinkle expression and the strong accumulation of mtSSB in distinct foci. As Twinkle is depleted by siRNA treatment, mtDNA replication can no longer initiate and mtSSB will no longer be recruited in large quantities, as single-stranded mtDNA will no longer be generated by the unwinding action of Twinkle. In contrast, in case of ddC mediated depletion, replication can still initiate but will stall or stutter and both Twinkle and mtSSB will be trapped on partially replicated DNA molecules (Supplementary Figure S7). It is noteworthy that it was recently shown in vitro that Twinkle is capable of loading on a circular DNA molecule without the assistance of a helicase loader (30). Combined, the data suggest that also in vivo Twinkle–mtDNA association is an important step in the initiation of mtDNA replication and show that Twinkle mtDNA unwinding is required for the recruitment of larger quantities of mtSSB during ongoing replication.

**Does twinkle mark specialized replication factories at the mitochondrial IM?**

Since Twinkle does not appear to be a constitutive mtDNA–nucleoid component, nucleoids must dynamically associate with Twinkle at, as we suggest here, specialized foci at the mitochondrial IM. Alternatively, mtDNA might dynamically associate with preexisting platforms at the IM and subsequently recruit Twinkle. The presence of distinct Twinkle foci in rho-zero cells and cells depleted of mtDNA by ddC argues that Twinkle is present at preexisting membrane foci with which mtDNA dynamically interacts. This is substantiated by a considerable decrease of mtDNA membrane association on Twinkle depletion using siRNA. Twinkle–mtDNA membrane association is also substantiated by the observation that GFP-tagged Twinkle was previously observed to show the same velocity and directionality as overall mitochondrial movement (8).

Although siRNA-mediated Twinkle depletion is not 100%, the observation that a considerable fraction of mtDNA remains membrane associated suggest other means by which it associates with the membrane. Other proposed proteins for involvement in mtDNA membrane association include a processed splice variant of OPA1 (31), ATAD3 (32) and prohibitin (33). Although we have not tested all possible candidates in the analysis of mtDNA fractionation, all available data combined suggest that mtDNA membrane association will likely depend on multiple proteins as well as physiological conditions and signals, and is a dynamic process. Current data suggests the existence of a soluble and possibly distinct membrane bound fractions. This explains published nucleoid proteome analyses in which proteins involved in mtDNA replication and repair, transcription, translation and biogenesis have all been identified (33,34). It is also substantiated by high-resolution immunofluorescence that shows partial but noncomplete overlap of signals or closely adjacent signals of mtDNA/TFAM and proteins with functions in protein synthesis, import and biogenesis (35). This has also been observed, for example, for ATAD3 colocalization (20) while both prohibitin and ATAD3 have recently been implicated in mitochondrial protein synthesis and mtDNA membrane association (34). We thus not only support the original suggestions by Iborra and colleagues that nucleoids are found adjacent to sites of mitochondrial biogenesis, but show that distinct and dynamic populations of mtDNA must exist in association with the IM, possibly dedicated to distinct functions such as translation, replication or repair. In contrast to the above, TFAM generally shows a perfect mtDNA co-localization pattern. It has been argued on the basis of super-resolution microscopy and theoretical considerations that perhaps the only permanently mtDNA-associated factor is TFAM (36,37).

Contrary to the current paradigm, we have shown that a substantial nonmembrane-bound or loosely membrane-associated mtDNA–protein fraction also exists in mammalian mitochondria. A second recent super-resolution microscopy study has similarly suggested that not all nucleoids are in direct close contact with the inner mitochondrial membrane (38) and that consequently mtDNA–membrane interactions could be transient in nature. We can agree with these findings using a different, and in this case biochemical, approach, and show that transient interactions include association with Twinkle to facilitate mtDNA replication.

Finally, in yeast, mtDNA replication showed a similar dynamic association of some of its nucleoid proteins including the POLG1 homologue, Mip1p, with mtDNA (39). In the same study, it was demonstrated that some of the core components of the mtDNA maintenance machinery are present as discrete complexes also in the absence of mtDNA. Although yeast does not have a Twinkle homologue (40), it was shown that one of the yeast mtDNA helicases, Pif1p, is membrane associated. Abf2p, the yeast TFAM homologue, could be partly dissociated by nuclease treatment (41), similar to our findings here for Twinkle and TFAM. Our data thus show a strong mechanistic similarity between yeast and mammalian mtDNA organization and replication, and suggest that mtDNA replication factories in close association with the IM exist throughout the Eukaryote lineage.

To conclude, in this article, we present the first clear evidence that human mtDNA replication factors dynamically associate with mtDNA to facilitate replication. Our type of analysis provides a direct handle on the functional characterization of putative mtDNA replication factors as we would predict these factors to co-localize only partially...
with mtDNA and to be enriched at foci that contain Twinkle and/or mtSSB. In particular, mtSSB seems to provide a good marker for mtDNA replication because it will be positive for fewer foci that are not in the process of replication, and because of its higher abundance, it gives a clearer immunofluorescent read-out when used with high-resolution imaging. On the basis of our results, we can anticipate that similar dynamic protein–mtDNA associations will exist for transcription and mtDNA repair factors. Evidence for the transient interaction of repair proteins includes the organization of base-excision repair proteins in discrete structures distinct from nucleoids (42); the increased co-localization of Cockayne syndrome group B protein with TFAM on menadione treatment (43); and the increased nucleoid co-localization of DNA2 on replication stalling (44). The dynamic nature of protein–mtDNA interactions raises the important question how these interactions are regulated at the molecular level.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
We thank Ian Holt, Howy Jacobs, Anu Suomalainen-Wartiovaara and Brendan Battersby for their interest and discussions.

FUNDING
Academy of Finland [CoE funding to J.N.S., 110463 and 108380 to P.M., currently Photonics grant 134893]; Tampere University Hospital Medical Research Fund [9J119, 9K126 and 9L097 to J.N.S.]; Netherlands Organization for Scientific Research [NWO: VICI grant 016.10.004]; European Molecular Biology Organization (EMBO LTF co-funded by Marie Curie Actions, EMBO LTF 1066_2011 (including MCA-EMBOCOFUND2010, GA-2010-267146) to J.M.G.]; Funding for open access charge: Netherlands Organization for Scientific Research [NWO].

Conflict of interest statement. None declared.

REFERENCES
The hexameric structure of the human mitochondrial replicative helicase Twinkle

Pablo Fernández-Millán1,†, Melisa Lázaro2,†, Şirin Cansız-Arda3,†, Joachim M. Gerhold3,†, Nina Rajala4, Claus-A. Schmitz1, Cristina Silva-Espiña1, David Gil2, Pau Bernadó5, Mikel Valle2,*, Johannes N. Spelbrink3,4,* and Maria Solà1,*

1Structural MitoLab; Department of Structural Biology, Molecular Biology Institute Barcelona (IBMB-CSIC), Barcelona, E-08028, Spain, 2Structural Biology Unit, Centre for Cooperative Research in Biosciences, CICbioGUNE, Derio, E-48160, Spain, 3Department of Pediatrics, Nijmegen Centre for Mitochondrial Disorders, Radboud University Medical Centre, Nijmegen, 6525 GA, The Netherlands, 4Mitochondrial DNA Maintenance Group, BioMediTech, University of Tampere, Tampere, FI-33014, Finland and 5Centre de Biochimie Structurale, INSERM-U1054, CNRS UMR-5048, Université de Montpellier I&II. Montpellier, F-34090, France

ABSTRACT

The mitochondrial replicative helicase Twinkle is involved in strand separation at the replication fork of mitochondrial DNA (mtDNA). Twinkle malfunction is associated with rare diseases that include late onset mitochondrial myopathies, neuromuscular disorders and fatal infantile mtDNA depletion syndrome. We examined its 3D structure by electron microscopy (EM) and small angle X-ray scattering (SAXS) and built the corresponding atomic models, which gave insight into the first molecular architecture of a full-length SF4 helicase that includes an N-terminal zinc-binding domain (ZBD), an intermediate RNA polymerase domain (RPD) and a RecA-like hexamerization C-terminal domain (CTD). The EM model of Twinkle reveals a hexameric two-layered ring comprising the ZBDs and RPDs in one layer and the CTDs in another. In the hexamer, contacts in trans with adjacent subunits occur between ZBDs and RPDs, and between RPDs and CTDs. The ZBDs show important structural heterogeneity. In solution, the scattering data are compatible with a mixture of extended hexa- and heptamer models in variable conformations. Overall, our structural data show a complex network of dynamic interactions that reconciles with the structural flexibility required for helicase activity.

INTRODUCTION

During DNA replication, dedicated replicative helicases unwind double-stranded (ds) DNA while cognate primases generate short DNA/RNA heteroduplexes on the lagging strand. These heteroduplexes prime 5′→3′ DNA synthesis by a DNA polymerase. Characterization of the gene linked to late-onset autosomal dominant progressive external ophthalmoplegia (adPEO) indicated that its product, Twinkle, was the mitochondrial DNA (mtDNA) helicase (1). The C-terminal region of Twinkle shares high similarity with the helicase C-terminal domain (CTD) of the bacteriophage T7 gene 4 (T7 gp4), a bifunctional primase-helicase protein (1–3). This CTD comprises a catalytic hexamerization domain reminiscent of the archetypal helicase RecA, and includes the Walker A and Walker B signatures (also termed H1 and H2, respectively) (4). In helicases and translocases, these motifs provide the catalytic residues for the hydrolysis of nucleoside triphosphates (NTPs), thus supplying the energy required for protein activity on DNA (4). Additional specific signatures within the helicase (H) domain—H1a, H3 and H4—assign Twinkle to the SF4 family of replicative DNA-like helicases (3,5), whose members have a ring-shaped hexameric structure (4). In general, a cleft at the interface between two neighboring RecA-like domains of the...
hexamer binds the NTP, the hydrolysis of which triggers a movement between subunits. This, in turn, leads to 5′→3′ DNA unwinding (6). Further studies confirmed the assignment of Twinkle to the SF4 helicase family: functional studies with recombinant protein in vitro revealed a nucleotide hydrolysis-dependent 5′→3′ DNA unwinding activity (7–9). In addition, Twinkle, the mitochondrial DNA polymerase γ (Polγ), and the mitochondrial single-stranded (ss) DNA-binding protein (mtSSB) form a minimal replisome in vitro (10).

Previous biophysical and EM analyses showed that Twinkle forms hexamers that convert to heptamers at low salt concentrations and in the presence of Mg2+ and the ATP-analog ATPγS (5,11). Oligomeric transitions between heptamers and hexamers have also been described for T7 gp4 (6,12), as well as for SF6 helicases, which include the mini chromosome maintenance (MCM) protein complex (present in archaea and eukarya) and bacterial HB8 RuvB, involved in branch migration in Holliday junctions (4).

Furthermore, studies in vitro demonstrated that Twinkle is capable of loading itself onto circular ssDNA (7,9) or onto dsDNA bubbles (7). However, Twinkle also differs from T7 gp4. It binds to dsDNA in the absence of nucleotides, and with higher affinity than when binding to ssDNA (9,11,13,14). It also displays two binding sites for ssDNA, which have been related to its annealing ability and suggests regulatory functions beyond replication (9).

The N-terminal domain (NTD) of Twinkle shows only limited sequence similarity with that of T7 gp4, but displays a similar predicted overall fold (1,3) and is classified as a prokaryotic DnaG-type primase (15). Primases of this type contain an N-terminal zinc-binding domain (ZBD), which binds ssDNA and detects trinucleotide patterns from priming-sites, and an RNA polymerase domain (RPD), which receives the priming site from the ZBD and initiates RNA primer synthesis (16). Bacterial and phage primase ZBDs typically contain a four- or five-stranded antiparallel β-sheet with two loops at one edge, which provide four cysteines (sometimes replaced by histidines) that coordinate a zinc ion (17–19). RPDs vary among primases: in E. coli DnaG (but not in T7 gp4) the RPD includes an N-terminal subdomain that binds ssDNA, orienting it toward an additional catalytic TOPRIM sub-domain (20). The TOPRIM subdomain is present in DnaG, T7 gp4 as well as in topoisomerases, nuclease and DNA repair proteins (21), and it contains an active site with metal-coordinating acidic residues that catalyze primer chain elongation (17,18,22).

In addition, functional studies with either T7 gp4 (23) or DnaG (24) showed that the ZBD can interact with the RPD from its own subunit (in cis) or with that from another subunit (in trans), in either case efficiently regulating primer synthesis.

Currently, the three-dimensional (3D) organization of a bifunctional SF4 primase-helicase is only exemplified by the crystal structure of a T7 gp4 natural fragment containing only the RPD and the CTD, thus lacking the ZBD (12) (Figure 1A). In the T7gp4 hexamer, the RPD and RPD-CTD linker of one subunit interacts with the CTD of the adjacent subunit, suggesting that the interaction of each subunit with its neighbor coordinates the helicase and primase activities thus the mechanism of translocation along ss-DNA (6). However, despite the predicted structural similarity between T7 gp4 and Twinkle, the bifunctional helicase-primase mechanism described for the former cannot be extrapolated to the latter because most of the activity-related metal-coordinating residues in both ZBD and RPD are absent in human Twinkle (1,3), thus precluding primase activity (13). However, there is a functional requirement for ZBD and RPD domains in Twinkle, since their partial or complete ablation results in a dramatic decrease in helicase activity (13,25). At the physiological level, Twinkle is essential for mtDNA copy number maintenance (25,26), which depends on DNA replication. Furthermore, around 50 mutations (and one in-frame 39-nucleotide duplication) within the Twinkle gene are related with the aforementioned rare diseases adPEO (1), infantile-onset spinocerebellar ataxia (IOSCA) (27), infantile onset mtDNA depletion syndromes (MDS) (28) and epileptic seizures (29), among others (see Figure 1B for domain localization of disease-related mutations and Supplementary Table S1 for references).

In order to examine the structural basis of Twinkle’s function and gain further insight into its dysfunction in disease, we developed an efficient recombinant protein expression procedure and performed electron microscopy (EM) and small angle X-ray scattering (SAXS) studies. Our results reveal that the hexameric organization of Twinkle is clearly asymmetric at the level of ZBDs and RPDs, which allows for a network of diverse interactions between domains along the oligomer, and provides the structural framework for the polar and asymmetric processing of nucleic acids.

MATERIALS AND METHODS

Cloning, E. coli heterologous expression and purification of recombinant Twinkle

A gene construct coding for mature Twinkle (Twinkle full-length, aa 43–684, UniProt Q96RR1.1) was cloned into vector pHART1201 with a C-terminal 6 histidines tag (30). Bases encoding the 5′ and 3′ SfiI restriction sites were added to the genetic construct by PCR with oligonucleotides (oligo) complementary to Twinkle 5′ end TACCATGCGCCACCTCCTGGCC-GAAGGAGATATACATATGGAGACTCTCAAGCGCTTG and 3′ end ACTTAGTGGCCGAGGCG-GCCGCTTGTGAACGCTTGGAGGTGTC (SIGMA). Large-scale cultures of E. coli BL21-DE3 containing a pG-KJE8 plasmid (TaKaRa) were induced with 1 mg/ml L-arabinose and 6 ng/ml tetracycline, and grown at 37°C until the culture reached an A595 of 0.6. Twinkle protein expression was induced with 0.25 mM IPTG overnight at 18°C. Cells were lysed by sonication in buffer A (50 mM sodium citrate pH 6.5, 1.5M NaCl, 50 mM L-arginine, 50 mM L-glutamate) containing 50 mM imidazole and EDTA-free protease inhibitor cocktail (Roche), and centrifuged at 20 000 rpm (JA20 Beckman rotor) for 20 min at 4°C. The supernatant was loaded onto a Ni-NTA-affinity chromatography column (HisTrap HP, GE HealthCare) connected to an AKTA FPLC system at room temperature and washed in buffer A containing 50 mM imidazole. The protein eluted at 35% of a linear gradient of buffer B (buffer A plus 0.5M imidazole). The elution fractions were
subsequently analyzed by 10% SDS-PAGE gels: the purest fractions were pooled and concentrated in a 30KDa cutoff ultrafiltrator (VIVAspin), and loaded onto a size exclusion Superose6 10/300 column (GE-HealthCare) previously equilibrated with buffer A. Eluting fractions were analyzed by 10% SDS-PAGE.

### Recombinant production of Twinkle in insect cells and purification

To produce Twinkle in SF9 insect cells, a gene construct coding for mature Twinkle was cloned into vector pHAR1201 as described above, except that it spanned from aa 30 to 684. BacMagic generated Twinkle cDNA-containing virus was first tested for protein production by testing various volumes of virus stock in 10 ml SF9 cultures containing virus was first tested for protein production by

## Helicase activity assays

The helicase activity was assayed by radiolabeling with $^{32}$P the 5' end of a 60nt oligo of sequence 5'ACATGATAAGATACATGATGAGTTTGGACAAACCAAGTTAAACGACGCAGCCGTCGCT-3', whose 3' end was complementary over 20 bases to plasmid M13(+), thus leaving a free 5' end of 40nt. The helicase reaction was performed at 37°C for 1h at increasing concentrations of Twinkle (10, 20, 30, 60 and 100ng), and 5 μM of plasmid M13(+), in buffer G (Tris-HCl 25 mM pH 7.5, MgCl2 4.5 mM, UTP 3 mM, NaCl 27 mM, 50 mM L-Glutamate), final pH adjusted to 8.0. Peak fractions were analyzed by 7.5% SDS-PAGE and stored in 10% glycerol.

### Gradient fixation (GraFix)

Twinkle oligomers were stabilized by using glutaraldehyde cross-linking combined with density gradient centrifugation following the GraFix method (31). A 4.5 ml gradient was formed with buffers I (HEPS 50 mM pH 7.5, NaCl 1 M, 10% glycerol) and J (same as I but with 30% glycerol and 0.15% glutaraldehyde) by using a gradient mixer and allowed to settle for 1h at 4°C. Thereafter, 300 μg of protein (200–300 μl) was applied on top of the gradient and centrifuged at 95 000x g in an MLS-50 swing-out rotor (Beckman Coulter) for 16h at 4°C. Subsequently, 300 μl fractions were collected and 80 mM glycine pH 8.0 was added to quench the cross-linking.
EM sample preparation and image processing

For cross-linked and non-crosslinked Twinkle samples, negative stained grids were prepared with 2% uranyl acetate and visualized on a JEM-1230 transmission electron microscope (JEOL Europe) at an acceleration voltage of 80 KV. The images were taken in low dose conditions at a magnification of 30 000X in a Gatan CCD camera, resulting in 2.3 A/pixel sampling. For cryo-EM data collection, the cross-linked Twinkle oligomer was dialyzed with Millipore membrane filters to remove the glycerol and decrease the salt concentration to 100 mM. Vitrification was performed using Quantifoil holey grids (with a thin carbon film floated on) rapidly plunged into liquid ethane in a FEI Vitrobot. The images were taken in a JEM-2200FS/CR electron microscope working at 200 KV at a magnification of 50 000X, and recorded at low-dose conditions on Kodak SO-163 films. Micrographs were scanned with a Z/I Photoscan scanner (Zeiss) with a step size of 7 μm, resulting in a final pixel size of 1.4 Å.

The single particles were extracted using EMAN (32) and Spire-SPIDER package (33,34), and 2D classifications were performed in a reference-free manner based on maximum-likelihood methods implemented in XMIPP (35). The initial models were built based on the top (6-fold rotational symmetry) and side (two-fold symmetry) views of the class averages. The references for the two techniques, negative staining and cryo-EM, were calculated independently and converged on similar 3D density maps. The particles were subjected to iterative refinements following the projection-matching scheme in Spire-SPIDER. During image processing several rotational symmetries were used, from C6 up to C1, starting with 6-fold symmetry and relaxing it in a sequential manner. Given that one of the two domains lining the axial channel departed from strict six-fold symmetry, during the reconstruction step of each iteration the generated asymmetric 3D volume was sliced along the z axis and a C1 symmetry pattern was imposed in layers that construct the N-terminal region whereas a 6-fold symmetry was imposed in layers that construct the CTD region, using XMIPP tools. This approach allowed us to maximize the signal-to-noise ratio in the final reconstructed map at the level of the CTD region and to prevent the smearing-out of ZBD and RPD domains. The resolution of the final cryoEM maps was estimated at 0.15 threshold in the Fourier Shell Correlation (FSC) calculated between maps reconstructed from two halves of the image data sets (36).

Monomeric and oligomeric homology model building

A sequence similarity search with BLAST against the entire Protein Data Bank (PDB) yielded the 3D structures from T7 gp4 helicase as closest structural relatives (ZBD-RPD primase domain, PDB code 1NU1; RPD and helicase domain, 1Q57; helicase domain, 1E0J and 1E0K, see Results for references). However, none of these structures covers the longest form of T7 gp4 protein. In addition, the natural NTD of T7 gp4, especially the ZBD, is shorter than that of Twinkle. Additional searches with the Twinkle NTD (aa 43–346) alone pointed to E. coli DnaG primase (PDB code 2AU3) as closest structural relative. Thus, we built a continuous homology model of Twinkle by superposing the RPD domains of coordinates 2AU3 and 1Q57, thus connecting the Twinkle-like ZBD from DnaG with T7 gp4 helicase domain via the common RPDs. Twinkle sequence alignment to the 2AU3 and 1Q57 templates was guided by secondary structure prediction of the former, with PSIPRED (37). With these structures and the alignment, a Twinkle monomer was generated with MODELLER (version 9.13, (38)), and this was fitted six times into the cryo-EM density. The Twinkle heptamer for SAXS was generated by “fusing” seven structures of DnaG (2AU3) to the seven chains of T7 gp4 1Q57, threading the sequence with MODELLER. The Twinkle hexamer was likewise generated by using T7 gp4 1E0J CTD hexamer.

Molecular dynamics flexible fitting

The hexamer was initially fitted manually into the cryo-EM map with Chimera (39) keeping the orientation for the monomers as in the crystal structure (6,12). This was appropriate for the CTD, whereas for the NTDs we carried out molecular dynamics simulations with NAMD 2.9 (40) through the MDFF plug-in (41). The protonation state of histidine residues in the initial model were predicted with PROPKA software from the PDB2PQR package (42). Simulations of 9ns were run with the CHARMM27 force field with CMAP corrections in generalized Born implicit solvent (43,44). We performed intrinsic solvent simulations with dielectric constant 1; cutoff 16 Å; pairlistdist 18 Å; GIBS on; ion concentration 0.3M; alpha cutoff 14 Å; sasa on and surface tension 0.006Kcal/ml/A2. The first minimization step was performed with a grid scaling of 0 to stabilize the initial model. Non-hydrogen atoms were coupled to the U_EM potential derived from the corresponding cryo-EM density maps with a grid scaling of 0.3 kcal/mol. Simulations used restraints for secondary structure, chirality and cis-peptides derived from the initially assembled atomic model. During the initial steps of the simulation, symmetrical restraints was applied between monomers, a symmetry restraint force constant k was applied, linearly increasing from 0 to 10 (Kcal/mol)/Å2. Essentially, the overall structure of globular domains (their secondary structure elements and the relationship between them) was conserved to avoid overfitting, the motions were only allowed in linker regions, and by applying tmdk 500, a constant which scales the harmonic force applied by the restraint on a domain.

Small angle x-ray scattering (SAXS)

SAXS data of Twinkle in size exclusion chromatography (SEC) buffer were recorded at 20°C at 0.3, 0.6 and 0.9 mg ml−1 in buffer A at beamline BM29 at ESRF (Grenoble, France), covering a momentum transfer range of 0.0282 < q < 4.525 Å−1. Ten 1s frames were collected for each sample. Frames with radiation damage were discarded automatically, and buffer scattering profiles measured before and after the sample were averaged and subtracted from the protein scattering profiles by using standard protocols (45). The forward scattering, I(0), and the radius of gyration, Rg, were calculated with the Guinier approximation assuming that, at very small angles (s < 1.3/Rg), intensity is represented...
as \( I(s) = I(0)\exp(-sR_g^2/3) \). The maximum particle dimension, \( D_{\text{max}} \), and the distance distribution were calculated from the scattering pattern with program GNOM (46). The molecular weight of the particle was calculated by comparison of the forward scattering \( I(0) \) with that from a reference solution of BSA (at 3.6 mg ml\(^{-1}\)). In order to characterize the conformational variability of Twinkle detected in our SAXS studies (see Results), we first generated a model of a monomer with program MODELLER (38), with which we generated hexamers and heptamers (see above). Subsequently, by applying the structural model generation module of the ensemble optimization method (EOM v2.0 (47)) we defined the ZBD, RPD and the helicase domain as rigid bodies, and the connecting linkers and N- and C-terminal segments as flexible regions, and generated 5000 models with different conformations for the hexamers and heptamers. For all these models the theoretical SAXS profile was computed with the program CRYSOL (48) using standard parameters. From all these models/crimes EOM selected a sub-ensemble of five conformations that optimally described the experimental SAXS curve. In order to have a better description of the conformational flexibility of Twinkle, the optimization was performed 100 times. The same procedure was repeated for sub-ensembles of 2, 10, 15, 20 and 50 conformations. Importantly, equivalent \( R_g \) and \( D_{\text{max}} \) distributions were obtained independently of the size of the sub-ensemble, suggesting no overfitting (see Results).

RESULTS

In order to find optimal conditions to stabilize Twinkle in a homogeneous population with no nucleic acid bound, we screened several chemical solutions by differential scanning fluorimetry (49). Twinkle was consistently more stable at high concentrations of NaCl (typically \( \sim 1\)M). Therefore, by using high salt we ensured that the sample did not bind DNA, as assessed by monitoring the ratio between the ODs at 260 and 280 nm during all steps of purification, while the particle maintained its oligomeric structure, as monitored by gel filtration (Figure 1C). Activity analysis of Twinkle in *E. coli* and in insect cells showed that the DNA binding and helicase activities were indistinguishable from those of Twinkle expressed by mammalian cells (Figure 2A).

**Twinkle forms flexible hexameric and heptameric rings**

To elucidate the structural organization of Twinkle, we first carried out single-particle negative staining EM studies with Twinkle produced in either insect cells or in *E. coli*. In both cases the two-dimensional analysis of recombinant Twinkle revealed distinct averages that display either six or seven radial densities (Figure 2B and Supplementary Figure S1), indicating the coexistence of heptameric and hexameric ring-like species. Each monomer showed one region connected to neighboring subunits and contributing to a ring with a central channel, while another region was exposed to the solvent with variable orientations in radial arrangement (Figure 2B), thus suggesting that, in the conditions tested, Twinkle oligomers have flexible regions. This flexibility is reminiscent of the T7 gp4 crystal structure, in which the C-terminal segment of the RPD-CTD linker performs strong contacts with the CTD of the neighboring subunit, whereas the N-terminal linker segment allows variable orientations of the RPDs over adjacent CTDs. This results in a loose arrangement of all N-terminal domains with respect to the helicase ring (12).

**Cryo-EM studies show asymmetry in the N-terminal ring**

Despite the initial successful classification by negative staining EM, further analysis by 3D averaging techniques such as cryo-electron microscopy (cryo-EM) required a more homogeneous sample. SEC revealed species with different hydrodynamic radius, and thus heterogeneous. Fractions collected at the end of the SEC peak (Figure 1C), which were expected to contain mostly hexamers, were selected for stabilization by the GraFix method (31). This technique rigidifies the macromolecules without inducing major structural rearrangements, by fractionating different conformations or oligomerization states in a density gradient of both glycerol and cross-linker (31,50). An initial analysis by negative staining of distinct GraFix-generated fractions showed that only few contained homogeneous molecular populations.

![Figure 2](http://nar.oxfordjournals.org/ Downloaded from http://nar.oxfordjournals.org/)
These were mostly hexamers, and suitable for cryo-EM. No heptamers were detected. The initial two-dimensional characterization of cross-linked Twinkle showed a more uniform diameter of the oligomers (Figure 3A) when compared with the negatively stained free sample (Figure 2B). After processing 12,790 particles, two types of electron density map (at 11.6 Å, cut-off of 0.15 in the FSC, Supplementary Figure S2) were calculated. One map was obtained by applying six-fold rotational symmetry to the entire particles (map C6-C6), which, consistent with the negative-staining EM, showed two stacked rings of different overall diameter encircling a central channel that widened when going from the upper narrow ring to the lower open lobules (Figure 3B). The upper ring was assigned to the helicase CTDs (‘TOP’ in Figure 3C; ~130 Å in diameter in panel B) and the lower to the NTDs (‘BOTTOM’; ~160 Å in diameter). The NTDs appeared in the form of six symmetric densities with no visible connections between them (Figure 3C). In addition, fostered by the T7 gp4 crystal structure, which showed variable orientations of the N-terminal domains (12), we further broke the six-fold symmetry of the widest NTDs ring while keeping the six-fold symmetry for the CTDs. This resulted in a 3D map (map C1-C6, at estimated resolution of 12 Å, Supplementary Figure S2) that showed the smaller ring with continuous density in a six-tip star shape, which was similar to the CTDs ring of map C6-C6 (compare Figure 3C and 3D, left panel). In contrast, in the ring attributed to the NTDs, four extra density blobs arose connecting five N-terminal lobules, while one NTD remained isolated (Figure 3D, central and right panels; compare with Figure 3C, same panels). Fully asymmetric image processing for both NTD and CTD yielded map C1 (Figure 3E) with features similar to map C1-C6, but in the last the signal-to-noise ratios were better. The sequence of slices from the N-terminal to the C-terminal ring of the reconstructed volume further evidenced the asymmetry of the particle (Figure 3E).

Cryo-EM-based homology modeling

In order to analyze the molecular architecture of Twinkle oligomers within the cryo-EM maps, a homology-model of each of the constituting domains was built. Among the SF4 family of helicases, only partial crystal structures of T7 gp4 have been solved by X-ray crystallography to date: the CTD in complex with NTPs, which forms a helical filament (PDB codes 1CR0 to 1CR4) (51); the CTD and the linker in complex with a non-hydrolysable ATP analog forming a hexameric ring (1E0J and 1E0K) (6); a 56-kDa natural short form with excised ZBD, which crystallized as a heptameric ring (1Q57, Figure 1A) (12); and the isolated primase domain showing ZBD-RPD domain swapping between two crystallographic partners (1NU1) (17). Separately, these structures did not enable us to build a full-length model of Twinkle because the NTD of the latter contains insertions, in particular at the ZBD (see Supplementary Figure S2). Searches with BLAST against the Protein Data Bank (PDB; www.pdb.org) using only the sequence of the Twinkle N-terminal region pointed to the crystal structure of DnaG primase from A. aeolicus (PDB code 2AU3) (24), which includes both a ZBD and an RPD, as a potential structural relative. Therefore, taking into account the secondary structure prediction of Twinkle, its sequence was threaded into a chimera of the DnaG (PDB 2AU3) and T7 gp4 (PDB 1Q57) models, which were superimposed on their respective RPDs, using MODELLER (38) (Figure 1A, see Experimental Procedures).

Rigid-body fitting was performed against both C6-C6 and C1-C6 maps. In T7 gp4 oligomers, the NTDs are placed on top of the CTDs from the neighboring subunit by virtue of an extended interconnecting linker, and the same arrangement was compatible with our cryo-EM maps of Twinkle (Figure 4A). Since in both C6-C6 and C1-C6 maps a six-fold symmetry had been applied to the C-terminal ring, the CTD fit was identical in both cases and, indeed, similar to the hexameric and heptameric structures of T7 gp4, which show similar CTD interfaces (12). However, for the NTDs, fitting into the C6-C6 map was hampered by the small size of the averaged reconstructed volumes. In contrast, NTD fitting into the C1-C6 map was better but required the introduction of inter-domain flexibility by molecular dynamics (see Experimental Procedures and Supplementary Figure S4). In particular, this latter fitting required rotation of the NTDs to place four ZBDs of consecutive subunits (ZBD1 to ZBD4) into the four extra densities. Following this arrangement, the four ZBDs contact the surface facing the channel of a neighboring RPD (contact in trans) while also contacting the RPD of their own subunit (contact in cis), thus showing a head-to-tail interaction between RPDs and ZBDs (Figure 4B and C). However, whereas the RPDs fit into well-defined density and are arranged similarly within the helicase ring, the intervening ZBDs show variable orientations (Figure 4B, left panel) and protrude differently from the particle (Figure 4B, central panel): ZBD2 is closest to the RPD ring, while ZBD1, ZBD4 and ZBD3 are moved outward by, respectively, 8 Å, 9 Å and 12 Å along the vertical axis. Therefore, the RPD-ZBD contacts are not constant. No suitable density was observed for ZBD5 or ZBD6 (Figure 4B and C, left panel), thus indicating that in two neighboring subunits the ZBDs are highly flexible. Overall, the NTD has a stable part provided by the RPDs and a highly flexible part afforded by the ZBDs.

SAXS studies reveal high flexibility at the inter-domain connectors

To characterize the conformational dynamics of Twinkle in solution, we conducted SAXS studies with Twinkle produced in E. coli, whose gel filtration elution profile and negative-staining analysis were equivalent to those of protein produced in insect cells. The molecular mass estimated by SAXS was 454 KDa, which suggested the presence of a mixture of hexamers (442 KDa) and heptamers (515 KDa). The Rg of Twinkle in solution was 69.8 Å and the pairwise distance distribution function of the curve, ρ(r) (Supplementary Figure S5), which reflects the distribution of the intra-molecular distances, showed a smooth decrease toward a maximum distance (Dmax) of 240 Å. In addition, the Kratky plot displayed a pronounced peak that did not drop to zero at high s values, which is consistent with a globular protein with flexible regions (Supplementary Figure S5). In order to explain the SAXS curve at the molecular level, hex-
Figure 3. Mitochondrial helicase 3D reconstruction. (A) Cryo-EM 2D classes of a crosslinked GraFix fraction (see Results) showed hexamers. (B) The 3D reconstruction applying C6 symmetry to the whole particle (C6-C6 map) shows two stacked rings, one with continuous density and the other with no density between lobules. (C) Three views of the electron density map after applying 6-fold symmetry to the whole particle. (D) Relaxation of the symmetry at the discontinuous ring (C1-C6 map) resulted in extra density between lobules, indicated by arrows. (E) Rendering of the 3D map without imposed symmetry (C1) is displayed together with slices of the reconstructed volume. The sequence of slices reveals 6-fold symmetry in the C-terminal ring but open asymmetric rings at the N-terminal region.
amers (based on T7 gp4 PDB 1E0J) and heptamers (PDB 1Q57) of the aforementioned homology model were constructed, but they did not fit the experimental SAXS curves (the associated $\chi^2$ values were $>10$ in all cases). Considering the Kratky plot and previous reports suggesting flexibility at the interdomain loops (5), we attributed this disagreement to the absence of flexibility in our atomic models. Therefore, we applied the ensemble optimization method (EOM, (47)) to generate hexamers and heptamers with flexible inter-domain arrangements. Two pools of 5000 hexamers (based on the aforementioned C1-C6 map homology model) and heptamers (based on a chimera model of T7 gp4 heptamer 1Q57 and DnaG primase 2AU3) were generated using RanCh (47). In these models the ZBD-RPD linker (residues 157–161) and the N-terminal segment of the RPD-CTD linker (residues 361–365), as well as the N-terminal and the C-terminal tails, were defined as flexible regions with complete conformational freedom. The EOM genetic algorithm selected a sub-ensemble of five conformations with a hexamer/heptamer ratio of 3/2, which, collectively, was in agreement ($\chi^2 = 0.71$) with the SAXS curve in the complete momentum transfer range (Figure 5A). These results substantiate equilibrium between hexamers and heptamers as observed in EM. The wide $R_g$ distribution of the EOM-selected models reflected the heterogeneity of conformations present in the sample (Supplementary Figure S5). In addition, calculation of EOM sub-ensembles using a variable number of conformations (from 2 to 50) presented similar quality of description of the experimental SAXS curve and, importantly, showed similar $R_g$ and $D_{max}$ distributions, suggesting no overfitting of the data (Supplementary Figure S5). The representation of the five conformations sub-ensemble reveals that the helicase ring domains are connected to highly flexible linkers that orientate the NTDs mostly on one side of the Twinkle ring (Figure 5B). In addition, ZBDs and RPDs also showed multiple relative orientations toward each other. Both Twinkle models, that from the T7 gp4 heptamer ring and that obtained by EM, have a similar overall diameter (127 and 124 Å, respectively, Figure 5C). Both have a channel of more than 40 Å wide, so it could only thread an ssDNA strand (Figure 5C). Considering the ability of EOM to detect and quantify the flexibility of macromolecules in solution (52), this result strongly suggested that our particles in solution displayed a wide central channel. Overall, these results indicate compatibility of the flexible structure derived from the cryo-EM hexamer with a fraction of the population in solution, the rest being compatible with flexible T7 gp4-based heptamers. In addition, unbound Twinkle oligomer dimen-
Figure 5. Analysis of Twinkle in solution. (A) Left panel: the experimental scattering-intensity curve (in black) is represented on a logarithmic scale as a function of the momentum transfer, $s = 4\pi\sin(\theta)/\lambda^{-1}$ (28, scattering angle; $\lambda = 0.9919$ Å, X-ray wavelength). The fitted EOM (see Experimental Procedures) curve for the mixture of flexible hexamers and heptamers (in red, $\chi^2 = 0.71$) and for the flexible gp4-like hexamer (in green, $\chi^2 = 1.80$) is shown. Below the panel, residuals show respective quality of fit. (B) Top and lateral molecular representation of the sub-ensemble of two models of T7 gp4 structure-based (1Q57, gp4 like) heptamers (right column) and three of EM-based hexamers (EM like, left) that collectively fit the data, superimposed by their CTD. ZBDs and RPDs are represented in ribbons while CTD rings are represented as a surface; each subunit has a different color. Red dots represent the position of C atoms of the flexible linkers and N- and C-terminal tails. (C) Surface representations of Twinkle CTD rings based on the T7 gp4 heptamer (PDB 1Q57) and hexamer (1E0J) and the cryo EM structure described here. Relative internal and external diameters values are represented by circles.

sions are consistent with binding of either ss- or dsDNA through the central channel.

DISCUSSION

Our structural studies show that Twinkle forms hexamers and heptamers of variable conformation in the presence of high salt but in the absence of NTPs, Mg\textsuperscript{2+} or DNA. These findings are consistent with previous reports showing that both stability and flexibility of Twinkle increased with salt concentration (11). In general, DNA-binding proteins are stabilized by solutions containing high concentrations of NaCl that compensate the protein electropositive surfaces involved in contacts with the negatively charged phosphate backbone of nucleic acids. On the other hand, high ionic strength may disrupt electrostatic interactions between domains, resulting in greater flexibility. This has been illustrated for *A. aeolicus* DnaG primase, in which the ZBD and RPD, connected by a flexible linker, crystallized at low salt conditions in a compact conformation with the two domains docked through a hydrophilic interface (24) (PDB 2AU3). The same authors used FRET at physiological salt concentrations to reveal a compact form that, at high salt, extended thanks to the flexible linker. Similarly, the crystal structure of the T7 gp4 primase domain (PDB...
1NU1) showed two protomers that swap their ZBDs, which perform electrostatic contacts with the symmetric RPDs by virtue of a flexible extended inter-domain linker (17). For Twinkle, flexibility of the ZBD-RPD linker was previously suggested (5) and is hereby confirmed. On the other hand, our combination of SEC with GraFix sub-fractionation allowed isolation of stabilized hexamers with a compact arrangement. This contrasts with the description of Twinkle in solution as an ensemble of more or less extended conformations. We hypothesize that a subpopulation of hexamers was locked in their compact state by the GraFix method. Overall, these results suggest that the inter-domain contacts between ZBD and RPD, but also those between RPD and CTD, are predominantly electrostatic and putatively tunable during the activity cycle of Twinkle, potentially undergoing structural rearrangements triggered by flexible linkers between domains.

**Structural variability in ZBDs**

For an SF4 protein displaying a structural organization homologous to a bifunctional primase-helicase, the hexameric model we present here includes a complete protomer comprising ZBD, RPD and CTD. Interestingly, the ZBDs were visible only in the absence of symmetry restraints and in four out of six protomers, thus indicating different positions and orientations of ZBDs in the stable hexamer. Since partial ZBD excision (construct Δ1–121) particularly reduces ssDNA binding (13), our cryo-EM model suggests that during translocation the ZBDs may bind ssDNA from different positions, two of them highly flexible. On the other hand, the clustering of delocalized ZBDs at one region of the NTDs ring or the open rings detected by negative staining suggest particles broken on one side. Notably, ring opening is triggered by the NTD in T7 gp4 (53) and is required for DNA loading (54).

Within the cryo-EM Twinkle hexamer, RPDs and ZBDs contact in a head-to-tail manner. Contacts between RPDs and ZBDs from different subunits are part of the transfer mechanism of the ssDNA priming site from the ZBD to the RPD in active primases (16). *E. coli* and *A. aeolicus* DnaGs and phage T7 gp4 perform efficient primer synthesis by cis and trans crosstalk between ZBDs and RPDs from different subunits (23,24). Cis and trans ZBD-RPD contacts have been shown by crystal structure analysis (17,24), yet they do not involve the active sites. In our cryo-EM model, the ZBDs contact neighboring subunit RPDs in a region facing the channel. At the back of this region, exposed to the solvent, are the amino acids that would be involved in primer synthesis in an active primase (compare Figure 4C and Supplementary Figure S6). If this contact occurred in an active SF4 primase, ZBD and RPD active sites could interact only by means of a small linker extension and a slight RPDs and ZBDs reorientation.

**Contacts between N-terminal and helicase domains**

Progressively longer deletion of the NTD—i.e. removal of ZBD, or ZBD and part of RPD (truncation Δ1–314)—increasingly affects ATPase and helicase activities of Twinkle, suggesting a functional or structural role of the NTD that influences the CTD folding and function (13). In our hands, a Twinkle construct that contained the CTD and the preceding inter-domain linker (Δ1–357) showed high heterogeneity both in size-exclusion chromatography and in fluorescence-based thermal shift assays (data not shown). This contrasted with the homogeneity of the full-length protein and suggested a weak contribution of the linker to the hexamer assembly. It also contrasts with T7 gp4, in which the linker alone stabilizes multimerization of the CTDs (6,51). Instead, the N-terminal truncation Δ1–314, which keeps the C-terminal region of the RPD, can form stable hexamers, indicating that the RPD fragment contains residues that contribute to oligomer stability (13). This is consistent with the Twinkle EM-based oligomeric model, which shows all RPD C-terminal regions facing the CTDs in rather similar orientations. In particular, vicinal RPD/CTD surfaces include motif V and VI from the RPD and motif Ia and following residues to motif 2 from the CTD (Supplementary Figure S6), which, together with the linker, are hot-spots for disease-linked mutations (Supplementary Table 1). According to previous reports (14), most of these mutants show structural instability that could have deleterious consequences in the long term. However, due to the limitation of our resolution, these detailed structural observations should be taken with caution. Nonetheless, in T7 gp4, the contact between RPD and CTD, despite being far from the active site, has shown to modulate the activity of the phage helicase (55), pin-pointing its functional importance. The functions of Twinkle include DNA translocation, which may involve structurally relevant subunit-specific CTD rotations to configure the central channel for ssDNA binding and pulling. This is the case in T7 gp4 (6), DnaB (56) and in papilloma-virus E1 helicase (from SF3) (57). Based on the Twinkle EM structure, a rotation of a CTD would affect the contact with the adjacent subunit RPD. Whether the RPD/CTD interface is disrupted or, by means of a flexible RPD-CTD linker remains intact during translocation, needs to be addressed by further research.

**Heptamers, hexamers and variable central channel diameter**

Studies describing hexamers and heptamers show disparate results, possibly attributable to differences in the experimental conditions (9,11). We found both hexamers and heptamers in the absence of ligands at high salt. T7 gp4 forms hexamers when bound to nucleoside di- or triphosphate and ssDNA, while a mixture of hexa- and heptamers appears in the absence of DNA. An ssDNA loading mechanism was proposed, according to which a heptamer ejects one subunit when it contacts ssDNA. This creates a gap in the ring through which the hexamer loads the ssDNA to the internal channel (54). However, because T7 gp4 binds dsDNA, other authors have proposed that heptamers translocate on it (12). Our unfigurated heptamer selected by SAXS shows a wide central channel compatible with either functional model, i.e. subunit ejection upon ssDNA detection or dsDNA translocation. The EM hexamer also shows a wide channel, and could thus allocate dsDNA. However, structural studies of hexameric helicases systematically show that, upon ssDNA binding, the ring converts to a right-
handed spiral with a dramatic narrowing of the central channel, which harbors ssDNA (54,56). Unliganded Twinkle hexamers could also undergo similar reshaping upon binding of cofactors or ssDNA.

Twinkle is particular in binding both dsDNA and ssDNA but, in contrast to T7 gp4, it has higher affinity for the former (9,11,13,14). In addition, a study analyzing the DNA re-annealing ability of Twinkle detected a total of three DNA-binding sites, two for ssDNA (putatively in the central channel and on the external ring surface) and one for dsDNA (9). Accordingly, sequence alignment (Supplementary Figure S3) suggests that Twinkle’s central channel may also contain the CTD ssDNA-binding loops found in other SF4 helicases, and our EM model shows the RPDs oriented so that the regions involved in ssDNA binding in active primases (e.g. DnaG (20)) are on the outer surface (Supplementary Figure S6). Therefore, the cryo-EM model of Twinkle described here is also compatible with internal and external binding sites for ssDNA reannealing.

A number of mutations in Twinkle have been found to be associated with various pathologies, and insights into the molecular mechanisms underlying its function in health and disease are provided by examination of its 3D structure. In summary, the hexamer structure revealed by EM shows a compact arrangement, in which a ZBD and an RPD from one subunit contact, respectively, the RPD and CTD of a neighboring subunit. RPDs and CTDs show broadly similar arrangements, suggesting a stable, functionally relevant contact. In contrast, the ZBDS show independent orientations, which is compatible with the asymmetric, differential ZBD positioning required for DNA pulling. Interdomain flexible loops allow for multiple conformations, and in solution ZBD/RPD/CTD contacts are disrupted while the hexameric ring remains intact. Taken together, our results show that Twinkle’s 3D structure supports the high plasticity needed for DNA loading and helicase activity.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the European Synchrotron Radiation Facility (ESRF) in Grenoble (France) for access to the BioSAXS beamline and for excellent support.

FUNDING

MINECO [BFU2009-07134, BFU2012-33516 to M.S., BFU2012-34873 to M.V.]; Generalitat de Catalunya [SGR2009-1366, 2014 SGR 997 to M.S.]; EU [FP7-HEALTH-2010-261460, FP7-PEOPLE-2011-290246, FP7-HEALTH-2012-306029-2 to M.S.]; P.F.-M. held an FPI fellowship from MINECO [BES-2008-05384REF]. Academy of Finland [CoE funding to J.N.S.]; Tampere University Hospital Medical Research Fund [91119, 9K126, 9L097 to J.N.S.]; Netherlands Organization for Scientific Research [NWO: VICI grant 865.10.004 to J.N.S.]; European Molecular Biology Organization [EMBO LTF co-funded by Marie Curie Actions, EMBO LTF 10662011 (including MCA-EMBOCOFUND2010, GA-2010-267146) to J.M.G.]; Agence Nationale de la Recherche SPIN-HD [CHEX-2011]; ATIP-Avenir. Funding for open access charge: MINECO [BFU2012-33516 to M.S.]. Conflict of interest statement. None declared.

REFERENCES


