

UNIVERSITAT AUTONOMA DE BARCELONA
DEPARTAMENT DE BIOQUIMICA I BIOLOGIA MOLECULAR

CONTROL OF THE ACTIVITY OF THE
HUMAN MITOCHONDRIAL
TRANSCRIPTION TERMINATION FACTOR
(mTERF) BY POLYMERIZATION. FIRST
EVIDENCES

TESI PRESENTADA PER JORDI ASIN CAYUELA
BARCELONA, 2003

TESIS DOCTORAL

Programa de Doctorado: Bioquímica y Biología Molecular, opción B.

CONTROL OF THE ACTIVITY OF THE HUMAN MITOCHONDRIAL TRANSCRIPTION TERMINATION FACTOR (mTERF) BY POLYMERIZATION. FIRST EVIDENCES

Memoria presentada por Jordi Asín Cayuela para optar al grado de Doctor.

Esta tesis doctoral ha sido realizada conjuntamente en la División de Biología del California Institute of Technology, Pasadena, Ca. y el Centre de Investigacions en Bioquímica i Biologia Molecular del Hospital Vall d'Hebron de Barcelona, bajo la co-dirección del Professor Giuseppe Attardi y del Dr. Antoni L. Andreu.

Tesis adscrita al Departament de Bioquímica i Biologia Molecular, Universitat
Autònoma de Barcelona.

Tutor: Dr. Simón Schwartz Riera

Acknowledgements

It has been a privilege to spend 5 and a half years in Professor Attardi's laboratory at Caltech. It is difficult to translate into words how much I owe to these years, both in the professional and personal levels. I benefitted not only from Dr. Attardi's expertise and patience towards me, but I also had the opportunity to learn from all the members of his group, and specially from Doctors Anne Chomyn, Petr Hayek, Mark Helm, Yan Wang, Jae Cho and Miguel Martin. Their experience in laboratory work, their expertise in molecular and cellular biology and their will to share them with me in countless conversations have been invaluable to bring this project to a satisfactory end. And for many technical aspects of my work I relied on the skills of Arguer Drew, Loredana Bellavia, Maria del Mar Roldan, Jennifer Fish and Rosie Zedan. It is only fair to say that I would not have been able to finish this project without their help and support. Special thanks need to go to Gwen Murdock, who took care with diligence and effectiveness of all aspects concerning my stay in the U.S. as a foreign national.

From this side of the Atlantic, I received invaluable help from Dr. Andreu. His ability to get things done and to face challenges and deadlines with both resolution and a smile have been and always will be a source of inspiration for me. To Doctors Elena Garcia and Ramon Marti, as well as to Nuri Prim, I thank for their help in the last stages of the composition of the manuscript. In this respect, I need to extend my recognition to Tofol, Ricardo, Gisela and Israel, all of them graduate students at Dr. Andreu's laboratory, who suffered with resignation, generosity and a fair dose of sense of humour my too-frequent incursions in their spaces, desks and computers during the preparation of the manuscript. My most sincere gratitude to all of them.

- **ADP**: Adenosine 5'-diphosphate
- **ATP**: Adenosine 5'-triphosphate
- **AUFS**: Absorbance Units Full Scale
- **bp**: base pair
- **BLASTP**: basic local alignment search tool for proteins
- **bZip**: basic domain-leucine zipper domain
- **COI**: subunit I of Complex IV
- **COII**: subunit II of Complex IV
- **COIII**: subunit III of Complex IV
- **COX**: cytochrome c oxidase
- **CSB**: Conserved Sequence Block
- **CTP**: Citidine 5'-triphosphate
- **Cyt b**: Apocytochrome b
- **Cyt c**: cytochrome c
- **CTP**: Citidine 5'-triphosphate
- **DEPC**: Diethylpyrocarbonate
- **D-loop**: Displacement loop
- **DTT**: Dithiothreitol
- **EDTA**: Ethylenediaminetetraacetic Acid
- **EF**: Elongation factor
- **ELISA**: Enzyme-Linked-Immunosorbent-Assay
- **FAD**: Flavin adenine dinucleotide
- **f-met**: formyl-methionine
- **FMN**: Flavin mononucleotide
- **GFP**: Green fluorescent protein
- **GTP**: Guanidine 5'-triphosphate
- **H₁**: initiation site for the Heavy-strand transcription rDNA unit
- **H₂**: initiation site for the total Heavy-strand transcription unit
- **HEPES**: N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
- **h-mtRPOL**: Human mitochondrial RNA polymerase

- **HMW**: High molecular weight
- **HRP**: Horseradish peroxidase
- **HSE**: Heat shock element
- **HSF**: Heat shock transcription Factor
- **H-strand**: Heavy-strand
- **IF**: Insoluble fraction (also initiation factor)
- **IgG**: Immunoglobulin G
- **IP**: Isoelectric point
- **kDa**: Kilodalton
- **LMW**: Low molecular weight
- **L-strand**: Light-strand
- **Lz**: leucine zipper domain
- **MELAS**: Myoclonic Encephalopathy, Lactic Acidosis and Strokes
- **mRNA**: messenger RNA
- **mtDBP**: Mitochondrial D-loop binding protein
- **mtDNA**: mitochondrial DNA
- **mTERF**: Mitochondrial transcription termination factor
- **MW**: Molecular weight
- **NADH**: Reduced form of nicotinic adenine dinucleotide
- **ND1**: subunit 1 of Complex I
- **ND6**: subunit 6 of Complex I
- **OD**: Optical density
- **O_H**: Origin of replication for the Heavy-strand of mitochondrial DNA
- **O_L**: Origin of replication for the Light-strand of mitochondrial DNA
- **OPD**: Orthophenylenediamine dihydrochloride
- **PAGE**: Polyacrylamide gel electrophoresis
- **PBS**: Phosphate buffer saline
- **PIPES**: 1,4 piperazine bis(2-ethanesulfonic acid)
- **PMSF**: Phenylmethylsulfonyl fluoride
- **Pol γ** : DNA polymerase γ
- **Pol I**: RNA polymerase I

- **PTRF**: Pol I and transcript release factor
- **PVDF**: Polyvinylidenfluorid
- **Q**: Oxidized form of ubiquinone
- **QH₂**: Reduced form of ubiquinone
- **rDNA**: ribosomal DNA (ribosome-coding genes)
- **RI**: Replication intermediate
- **R-loop**: Replication Loop
- **rRNA**: ribosomal RNA
- **SDS**: Sodium dodecyl sulfate
- **SDS-PAGE**: Sodium dodecyl sulfatate - polyacrylamide gel electrophoresis
- **SF**: Soluble fraction
- **SRSEC**: Seryl-tRNA synthetase from *E. coli*
- **TAS**: Termination Associated Sequence
- **TBE**: Tris-Borate-EDTA
- **TE**: Tris-EDTA
- **TEMED**: N,N,N',N'-tetramethylethylenediamine
- **TFAM**: Mitochondrial transcription factor A
- **TFB1M**: Mitochondrial transcription factor 1
- **TFB2M**: Mitochondrial transcription factor 2
- **tRNA**: transfer RNA
- **TTF-I**: Transcription termination factor I
- **UTP**: Uridine 5'-triphosphate
- **UTR**: Untranslated region

	Page
INTRODUCTION	1
1 .The human mitochondrial genome	2
1.1.The mitochondrion: general aspects	2
1.1.1. The electron transport chain (ETC)	5
1.2. Evolutionary origin of mitochondria	8
1.3. The human mitochondrial genome	12
1.3.1. Replication of human mtDNA	15
1.3.1.1. Strand-asymmetric model of replication	15
1.3.1.2. Alternative model of replication	19
1.3.2. Transcription of human mtDNA	20
1.3.3. Translation of human mtDNA	22
2. Leucine zippers	25
3. The human mitochondrial transcription termination factor (mTERF)	29
.	
OBJECTIVES	36
.	
MATERIAL AND METHODS	38
1. Preparation of anti-mTERF antiserum	39
1.1. Purification of His-tagged mTERF	39
1.1.1. Bacterial culture	39
1.1.2. Nickel column chromatography	41
1.1.2.1. Column preparation	41
1.1.2.2. Purification	42
1.1.3. Electroelution of the protein from the gel slice	43
1.2. Immunizations and bleedings of the rabbit	44
1.3. Testing the efficiency of immunization by ELISA	45
2. Purification of mTERF from HeLa cells	46
2.1. Preparation of S-100 from HeLa cell mitochondria	46

2.2. Preparation of DNA affinity column	48
2.3. Heparin chromatography	51
2.4. DNA affinity chromatography	52
3. Gel filtration chromatography	53
4. Study of DNA-binding activity: band-shift analysis	55
4.1. 5'-end labeling of the probe	55
4.2. Band-shift assay	56
4.3. Super-shift assay using the anti-mTERF antiserum	56
5. Transcription termination activity test	57
5.1. Preparation of DNA template (pTER)	57
5.2 In vitro transcription assay	59
5.3. S1 protection assay	60
5.3.1. Preparation of S1 RNA probe	61
5.3.2. S1 protection assay	62
6. SDS-polyacrylamide gel electrophoresis (SDS-PAGE)	63
6.1. General procedure	63
6.2. Coomassie Blue staining	65
6.3. Silver staining	65
7. Electrophoresis on agarose for DNA analysis	65
8. Immunoblotting	66
9. 2D electrophoresis	67
9.1. First dimension: native-PAGE	68
9.2. Second dimension: SDS-PAGE	69
10. Measurement of ³²P incorporation to DNA	70
11. Determination of protein concentration	71
12. Concentration of samples	71
RESULTS	72
1. Preparation of polyclonal anti-mTERF antibody	73

2. Gel filtration chromatography of S-100 from	
HeLa mitochondrial lysate	74
3. Native-PAGE of HMW and LMW pools	77
4. Transcription termination activity assay of GF fractions	77
5. DNA-binding activity assays	80
6. The HMW form of mTERF is a reversible structure	83
7. Homopolymer vs. heteropolymer	85
.	
DISCUSSION	90
.	
CONCLUSIONS	99
.	
REFERENCES	101
.	

INTRODUCTION

1) The human mitochondrial genome

1.1) The mitochondrion: general aspects

Mitochondria are nearly ubiquitous organelles within eukaryotic cells. Only four protozoan phyla, all of them intracellular parasites, are known to lack mitochondria (Cavalier-Smith, 1987). While their principal function is oxidative phosphorylation, they also contribute to the biosynthesis of pyrimidines, aminoacids, phospholipids, nucleotides, folate coenzymes, heme, urea, and many other metabolites. The formation of these organelles involves a major contribution from the nuclear genome and a quantitatively minor, though essential, contribution from the mitochondrial genome. It also involves the operation of sophisticated mechanisms for the import and intramitochondrial sorting of extramitochondrially synthesized proteins, phospholipids, and, in some species, a few RNAs.

Although mitochondria have neither fixed size nor shape they are often represented with a sausage-like shape (that is how they actually look like in hepatocytes and fibroblasts), and with average dimensions of 3-4 μm in length and approximately 1 μm in diameter. The number of mitochondria per cell also varies significantly from cell type to cell type. Estimates from serial sections of cells yield values in the range of a few hundred to a few thousand per cell. For example, hepatocytes contain 800 mitochondria, human oocytes 100,000, while spermatozoa have relatively few, though of larger size. In lower eukaryotes, the number can go as far as 500,000, in the case of the giant amoebae *Chaos chaos*. But the actual number is probably not the most relevant parameter to consider. Of far more relevance may be the ratio between total mitochondrial volume in a cell and cell volume, or even the total surface area of the mitochondrial cristae per cell.

With the perfection of electron microscopy, mitochondria were shown to contain two membranes, an outer membrane and an inner membrane, and it was recognized that the latter was convoluted and folded into *cristae mitochondriales*. This topology compartmentalizes the organelle into an intermembrane space and a matrix. Cristae are

often lamellar in appearance, but they can be tube-like in some cell types, like the amoeba and cells from the adrenal cortex, or even appear as arrays of triangular tubes, like the ones found in mitochondria of cardiac cells (see Scheffler, 1999 and references therein). Considering that the complexes of the respiratory chain are anchored to the cristae, their number and morphology is likely to reflect the response of the mitochondria to the energy demands of the cell. Highly folded, lamellar cristae with a large surface area are typically found in muscle and neurons, where the respiratory rate is the greatest. An interesting example of this tendency can be found in the dramatic changes in the morphology of the inner membrane of *Saccharomyces cerevisiae* in response to different metabolic situations involving the activation or inhibition of the respiratory chain (Johnson and Carlston, 1993; Entian and Barnett, 1992).

Today it is accepted that although cristae are invaginations of the inner mitochondrial membrane, the inner membrane juxtaposed to the outer membrane and the membrane of the cristae are connected by narrow tube-like connections called cristae junctions (Perkins *et al*, 1997). The existence of these junctions raises very relevant questions at the functional level. On the one hand, they may define two compartments in the intermembrane space, intracristal and intermembrane. On the other hand, they may serve to segregate patches of inner membrane with differing protein compositions.

Mitochondria are not distributed randomly in the cytosol. They can either be restrained in preferred positions, or move or be moved to sites where their presence is needed. They are not only not static in their positions in the cytosol, but they are quite changeable in their shape. This intrinsic mitochondrial motility has been observed in animal cells, in fungal hyphae, in algal cells and in plant cells (Bereiter-Hahn and Voth, 1994). It has been suggested that ATP and/or ADP gradients might be responsible for the distribution of mitochondria in the cytosol. That would explain why mitochondria are wrapped around or packed along the axonemes of sperm tails, or why skeletal and cardiac muscle cells require a very uniform distribution of mitochondria in the myofibrils along their entire length. But, although some of the dynamics are due to intrinsic driving forces, other movements are clearly the consequence of the interaction of mitochondria with the cytoskeleton and the possible involvement of molecular motors, like kinesin and dynein.

(Morris and Hollenbeck, 1993; Lee and Hollenbeck, 1995; Morris and Hollenbeck, 1995; Overly *et al*, 1996; Tanaka *et al*, 1998; Hurd and Saxton, 1996).

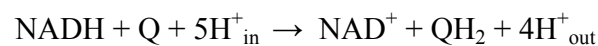
Changes in shape and size can also be the consequence of processes like fusion or fission. Such events have been elegantly studied by time-lapse photography in the phase contrast microscope, corroborating previous observations with electron microscopy observations (Bereiter-Hahn and Voth, 1994). Fission is the postulated mechanism for mitochondrial proliferation. The insertion of new components into the matrix and all membranes will cause mitochondria to grow, and eventually there must be a signal for such a mitochondrion to divide. Fusion has been clearly established in yeast. Serial sections and image reconstructions revealed that in *Saccharomyces cerevisiae* a cell contains from 1 to 10 mitochondrial reticular structures, presumably the product of fusion (Hoffman and Avers, 1973; Stevens, 1981). More recently, the use of modified green fluorescent protein (GFP) for import into mitochondria enabled the use of wide-field fluorescence microscopy for the acquisition of three-dimensional images over long time intervals (Nunnari *et al*, 1997). From these studies, the authors were able to quantitate fusion and fission events. During vegetative growth of wild type cells, 0.39 ± 0.14 fission events per minute and 0.40 ± 0.08 fusion events per minute were reported, indicative of a highly dynamic situation. A similar study with GFP in HeLa mitochondria concluded that mitochondria in these cells also formed a connected, continuous and highly dynamic network (Rizzuto *et al*, 1998). A very interesting aspect regarding mitochondrial fusion is the possibility for fused mitochondria to mix their genetic content. A model has been proposed of human mitochondria functioning as a single dynamic unit in a living cell, with rapid diffusion of mtDNA and/or RNA throughout the organelles, a phenomenon known as complementation (Hayashi *et al*, 1994; Ono *et al*, 2001). This model has been challenged by Attardi and coworkers (Enriquez *et al*, 2000), after finding very low levels of complementation in human cells. These authors argued that evidences of fusion would not necessarily imply a mixing of mtDNA and/or their products, and that the possibility of fusion involving only the external mitochondrial membranes could not be ruled out. In this case, mtDNA and its transcripts would remain compartmentalized in fully fused mitochondria. The differences in the degree of inter-mitochondrial complementation

observed by different groups could also be ascribed to a difference in nuclear background (Attardi *et al*, 2002).

1.1.1) The electron transport chain (ETC)

As stated earlier, the main function of mitochondria is to provide energy to the cell through oxidative phosphorylation. This function is carried out by the electron transport chain (ETC), composed of five protein complexes (I to V), all of them anchored to the inner mitochondrial membrane, plus ubiquinone and cytochrome c acting as mobile electron carriers between complexes. Each complex consists of a number of peptide subunits, 13 of which are encoded by the mitochondrial genome, while the rest is nuclear-encoded.

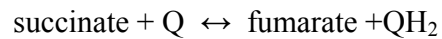
Complex I (NADH-ubiquinone oxidoreductase) catalyzes the following reaction:



where Q and QH₂ refer to the oxidized and reduced forms of ubiquinone, respectively. It is important to signify that the result of this reaction is not only the transfer of one electron from NADH to Q, but also the pumping of 4 protons from the matrix to the intermembrane space. Complex I is composed of at least 41 peptides (~900 kDa) in mammals, of which 7 are mitochondrially encoded (see below). 34 nuclear encoded subunits have been identified in bovine (Walker *et al*, 1995). An overall approach to the structure of the complex has been done by electron microscopy. From these studies it was concluded that complex I is L-shaped, with the long arm integrated in the inner mitochondrial membrane and the short arm extending into the matrix. The short arm contains the co-factor flavin mononucleotide (FMN), at least four Fe-S clusters and the binding site for NADH, and functions as NADH dehydrogenase, while the long arm exerts the ubiquinone hydrogenase activity. The 7 hydrophobic subunits encoded by mtDNA are part of the long arm.

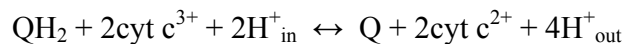
Complex II (succinate:ubiquinone oxidoreductase) consists of four nuclear-encoded peptides and therefore is the simplest of all complexes of the ETC. The two

largest peptides (70 and 27 kDa) constitute the peripheral portion of the complex and function as the enzyme succinate dehydrogenase in the Krebs cycle. They are associated to the inner mitochondrial membrane through two anchor proteins, C_{II-3} (15 kDa) and C_{II-4} (12 kDa). Complex II catalyzes the following reaction:



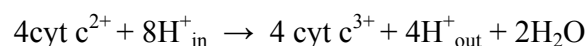
thus linking the Krebs cycle directly to the ETC (Davis and Hatefi, 1972). Flavin adenine dinucleotide (FAD) is linked to the largest peptide (70 kDa), and both form the flavoprotein subunit, or Fp (Kenney *et al*, 1972). The Fp subunit is intimately associated with the iron-protein subunit (Ip), made up of a 27 kDa peptide containing three non-heme Fe-S centers. Fp-Ip subcomplex contains the binding site for succinate, the hydrogen acceptor (FAD) and the Fe-S centers for the transfer of electrons to the anchor proteins, whereas the latter contain the binding site for ubiquinone and a b-type cytochrome with an unclear function (Merli *et al*, 1979).

Complex III (ubiquinone-cytochrome-c oxidoreductase, or bc₁ complex) catalyzes the following reaction:



Similar to the reaction catalyzed by complex I, the oxidation of one of the substrates and the transfer of electrons to the mobile carrier (in this case, cytochrome c) is coupled to the transfer of protons across the inner mitochondrial membrane. Of the 11 peptides that compose this complex, one is encoded by the mitochondrial genome (cytochrome b), while the other 10 are nuclear-encoded. Functionally, the most important subunits are the cytochromes b and c₁ and the Rieske iron-sulfur protein, since they are the only ones participating in the electron transfer and proton pumping. The complete crystal structure of bovine complex III at 2.9 Å resolution was published in 1997 (Xia *et al*, 1997).

Complex IV (cytochrome c oxidase) catalyzes the following reaction:



Molecular oxygen is the terminal electron acceptor, the mobile carrier cytochrome c is reoxidized and four protons are pumped from the matrix to the intermembrane space. Bovine complex IV was the first complex to have its crystal structure determined at high resolution (Tsukihara *et al*, 1995; Tsukihara *et al*, 1996). The mammalian complex contains 13 subunits. The three largest subunits (I, II and III) are encoded by the mitochondrial genome, and the rest (IV, Va, Vb, Via, Vib, Vic, VIIa, VIIb, VIIc and VIII) is nuclear-encoded. Subunit I binds to heme prosthetic groups (a and a₃) and also contains one copper redox center, Cu_B. Subunit II contains another copper redox center, Cu_A. Subunit III is likely to be involved in proton pumping and in modulating the electron transport through the metal centers. Therefore, subunits I-III form the functional core of the complex. The other subunits might perform regulatory functions, or play a role in insulation, stabilization or assembly (Schlerf *et al*, 1988; Schillace *et al*, 1994).

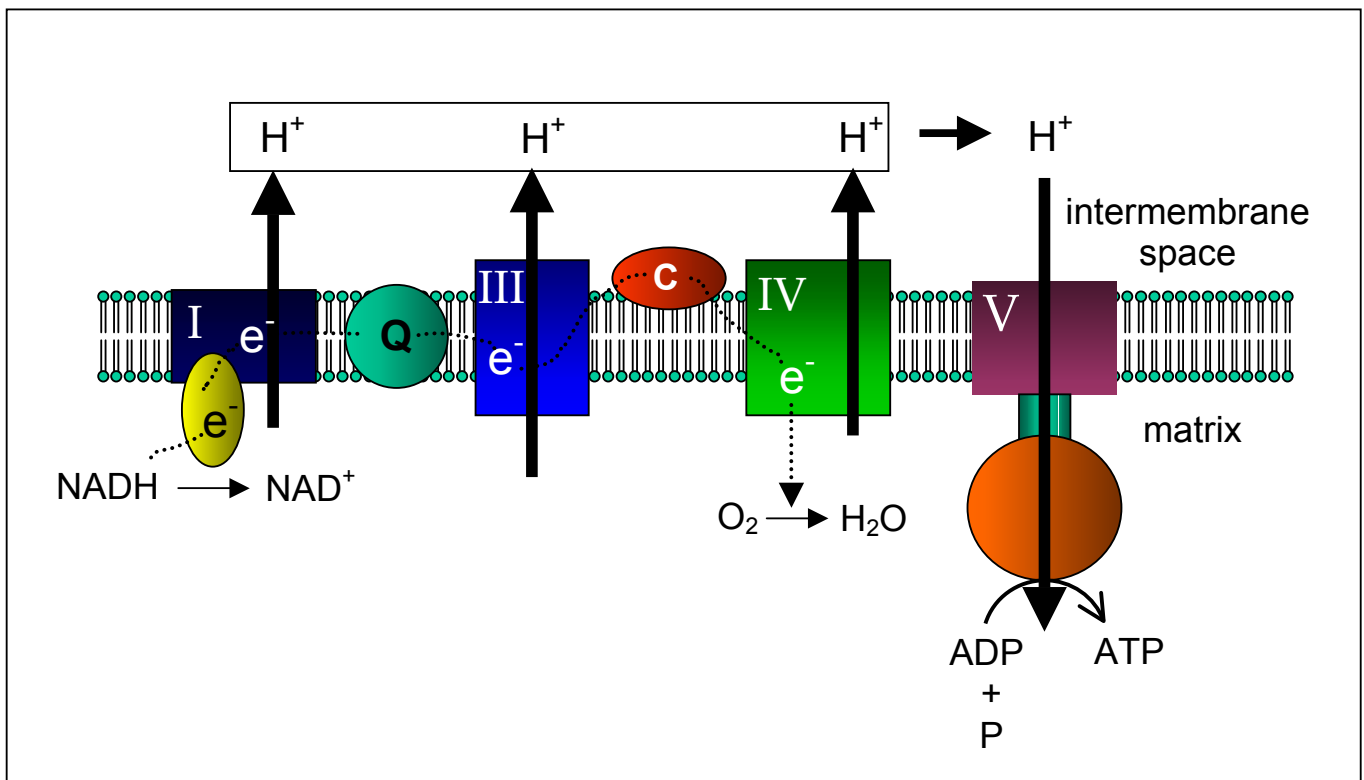
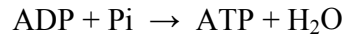


Fig. 1. Schematic illustration of the chemiosmotic hypothesis

Complex V (F₁F₀-ATPase) is responsible for the synthesis of ATP from ADP and inorganic phosphate:



This reaction is driven by the proton gradient generated by the ETC at complexes I, III and IV. Complex V is composed by two subcomplexes, a soluble F₁ ATPase in the mitochondrial matrix, attached to an insoluble membrane complex referred to as F₀ (Hatefi, 1976; Hatefi *et al.*, 1979). F₀ has the ability to translocate protons from the intermembrane space to the matrix, and when coupled to F₁, ATP synthesis can be achieved. The F₁ ATPase has five different peptides, α , β , γ , δ and ϵ , present in the ratio 3:3:1:1:1. The F₀ subcomplex has 11 subunits, of which 3 confer proton translocation activity to this subcomplex, while the rest are accessory proteins necessary for assembly, stabilization and control. Only two subunits, belonging to F₀, are encoded by the mitochondrial genome. The crystal structure of bovine complex V at 2.8 Å resolution was published by Walker's group in Cambridge (Abrahams *et al.*, 1994).

The question of how respiration is linked to ATP synthesis has been the object of debate over the years. A theoretical foundation for the answer to this question was proposed in the early 60s (Mitchell, 1961), but it took 10 to 15 years to convince the scientific community. Nowadays, Mitchell's chemiosmotic hypothesis is fully accepted. According to this hypothesis, the energy required for ATP synthesis by the F₁ subunit of complex V comes from the flow of protons through the F₀ subunit of the same complex. The flow of protons is the result of a proton concentration gradient generated by complexes I, III and IV of the electron transport chain (see fig. 1).

1.2) Evolutionary origin of mitochondria

Soon after mitochondria were first observed in cells of higher organisms, the idea that they were somehow related to bacteria was expressed in a more or less explicit form

by authors like R. Altmann (1890), K.C. Mereschkowsky (1910), P. Portier (1918), J.E. Wallin (1927) and J. Lederberg (1952) (see Margulis, 1981 for a complete bibliography). The hypothesis was not well received, and it was not until the discovery of mitochondrial DNA that it was seriously considered, and nowadays the prokaryotic ancestry of mitochondria is widely accepted. According to this hypothesis, present eukaryotic cells are the consequence of a symbiosis between a protoeukaryote cell and a prokaryote. The first one had its DNA compartmentalized in a nucleus and had capacity for anaerobic glycolysis and fermentation, while the prokaryote did not have a nucleus and contributed the electron transport system. Both must have been capable of DNA replication, transcription, protein synthesis and the biosynthesis of various lipids to form a membrane. Krebs cycle enzymes may also have been present in both for the purpose of interconverting short carbon compounds and aminoacids. Therefore, the initial association between these two cells must have led to a considerable amount of redundancy of genetic information. Some of the redundant genes in the mitochondrion were lost, while others were transferred to the nucleus, leaving only a few remaining in the mitochondrion (Gray and Doolittle, 1982; Gray, 1992).

The link between mitochondria and protozoan endosymbionts was further strengthened with the complete sequencing of the mitochondrial genome of the protozoan *Reclinomonas americana* (Palmer, 1997; Lang *et al*, 1997). It is 69,034 bp long, and contains 97 genes, many more than those present in other known mitochondrial genomes. But the most relevant finding was that, apart from 23 components of the respiratory chain, 18 motoribosomal proteins, 27 distinct tRNAs, 3 rRNAs and a component of the RNA processing enzyme RNase P, this genome encoded 4 components of a multisubunit RNA polymerase resembling a eubacterial-type polymerase. This finding strongly suggests that an eubacterial multisubunit RNA polymerase present in the original endosymbiont was lost during evolution in all mitochondria analyzed so far (except for *R. Americana*) and replaced by a nuclear-encoded single-subunit enzyme with similarities to the T3/T7 RNA polymerase.

Mitochondrial DNA (mtDNA) exhibits an extraordinary diversity in structure, gene organization, gene content, mode of expression and replication in different

organisms (Saccone, 1984). Such diversity probably reflects differences in the evolutionary pathways that have led to the present-day segregation of mitochondrial and nuclear genes. However, it is also possible that this diversity bears witness to a multiplicity of past endocytotic events that may have led to the development of mitochondria in different eukaryotic cells. The genes of the original endosymbionts that have remained sequestered in mitochondria of various present-day organisms are harbored in structures as different as the 46 kilobase linear molecules of *Tetrahymena* (Borst, 1980), the 16-20 kilobase circular molecules of metazoan mtDNA (Attardi, 1985) or the gigantic 570 kilobase circles of maize mtDNA (Lonsdale *et al*, 1984). The biological significance of these differences is not clear. One point of view is that the transfer of genes from the protomitochondria to the nucleus has progressed to different extents in different organisms purely by chance, and that further transfer of genes will occur in the future. In metazoans, the selection appears to have forced the reduction of the mitochondrial genome to its absolute minimum in terms of nucleotides, that is, the highest possible density of genes per unit length of DNA, although it is not clear whether this represents a limit to the genes that can be transferred to the nucleus. Apart from the intergenic sequences, the genetic information in the mitochondrial genome is very similar in the majority of organisms studied. This might imply that the loss of genes from the original symbiont was essentially complete before extensive branching of the phylogenetic tree occurred, or that the present gene organization of different species was achieved independently, leaving inside mitochondria those genes whose products could not be imported at all. This argument cannot be sustained for tRNAs, which are known to be imported in plant mitochondria, in protozoa and in yeasts. And although the proteins encoded by vertebrate mtDNA are highly hydrophobic, and therefore unsuitable for import across the mitochondrial membranes, such rationalization fails to explain why the two integral proteins of complex II are imported in most organisms (Scheffler, 1999). In the same league, knock-out or mutations of mitochondrial genes causing defects in mitochondrial function have been successfully complemented with appropriately engineered versions (codon usage, addition of import signals) in the nucleus (Guy *et al*, 2002; Manfredi *et al*, 2002). Theoretical arguments have been advanced predicting that the mitochondrial genome ultimately needs to retain only two genes encoding peptides, in

addition to the rRNA and tRNA coding genes (Popot and deVitry, 1990; Claros *et al*, 1995; Claros, 1995). The two peptides are apocytochrome b and COX1, due to the high presence of hydrophobic transmembrane domains in their sequences.

On the other hand, there seems to be convincing evidence that the complete loss of mitochondrial genome has in fact occurred in evolution (Palmer, 1997). The best studied example is the parasitic flagellate protist *Trichomonas vaginalis*, an air-tolerating anaerobe that has an energy-producing organelle known as hydrogenosome. This organelle is believed to be a highly modified mitochondria. The criteria for such an assertion are the existence of a double membrane, the fact that they import proteins posttranslationally and that they divide by fission, and above all that they contain Hsp10, Hsp60 and Hsp70 (Bui *et al*, 1996). These proteins are considered the most reliable tracers of eubacterial ancestry of both mitochondrion and chloroplast (Palmer, 1997). Thus, the evidence supports a common origin for hydrogenosomes and mitochondria.

The size range of mtDNAs found in multicellular animals is relatively narrow (~17 kb), with some exceptions varying from 14 kb in the nematode *Caenorhabditis elegans* (Okimoto *et al*, 1992) to 42 kb in the scallop *Placopecten magallanicus* (LaRoche *et al*, 1990), and all are single, circular DNAs. An exception to this trend is the cnidarian *Hydra attenuata*. Its mtDNA consists of two unique, linear DNA molecules of 8 kb (Warrior and Gall, 1985). Complete sequences are now available from various mammals, chicken (*Gallus domesticus*), toad (*Xenopus laevis*), sea urchin (*Strongylocentrotus purpuratus*), the fruit fly (*Drosophila yokuba*) the nematode *Caenorhabditis elegans* and the sea anemone (*Metridium senile*), and it's striking that deviations in length from the human sequence are less than 1 kb (Saccone, 1994).

Although most organisms studied have circular mtDNA, an increasing number of organisms have been shown to present linear mitochondrial genomes, from among the ciliata (*Paramecium Aurelia*), the algae (*Chlamydomonas reinhardtii*), the fungi (*Candida*, *Pichia* and *Williopsis* species) and oomycetes (Nosek *et al*, 1998). The size of these linear mtDNAs is in the 30-60 kb range, and although it has been argued that its existence might indicate a different evolutionary origin, this form of DNA might have arisen by

accident and become stabilized by the fortuitous existence of a replication machinery capable of dealing with linear ends (Nosek *et al*, 1998).

1.3) The human mitochondrial genome

The human mtDNA (see fig. 2) was the first complete mitochondrial genome to be sequenced from any organism (Anderson *et al*, 1981). It is a double stranded circular molecule of 16,569 nucleotides. Both strands have different sedimentation coefficients in CsCl alkaline gradients due to their different content in G + T. Therefore, they are known as heavy (H) and light (L) strands. A third strand (known as 7S DNA) is present between the origins of replication for the H strand and nucleotide 16106* (with minor 3'-ends at nucleotides 16105 and 16104, according to Doda *et al*, 1981), conferring a triple stranded configuration known as the D-loop. 7S DNA turnover is very high, and though its function has not been clearly established, it is widely accepted that it might play a crucial role in priming mtDNA for replication. The identification of a conserved 15 nucleotide sequence located a short distance upstream from the 3' D-loop ends in human and mouse cells led to the proposal that arrest of 7S DNA synthesis is a template-directed event (Doda *et al*, 1981). Consistent with this suggestion, closely related termination-associated sequences (TASs) have been identified at similar positions relative to mapped D-loop DNA 3' ends in *Xenopus laevis* (Dunon-Bluteau and Brun, 1987), cow, pig (Mackay *et al*, 1986) and several primates (Foran *et al*, 1988). It is therefore accepted that TAS elements are candidate *cis* elements for the regulation of 7S DNA synthesis arrest, and by extension, of H-strand DNA synthesis. Consistent with this hypothesis, protein binding to TAS elements in bovine mtDNA (Madsen *et al*, 1993) and rat (Kumar *et al*, 1993) mtDNA have been reported.

* Human mtDNA was sequenced at the MRC Laboratory of Molecular Biology at Cambridge, and the numeration adopted then, known as 'Cambridge sequence', has been universally adopted to number human mtDNA.

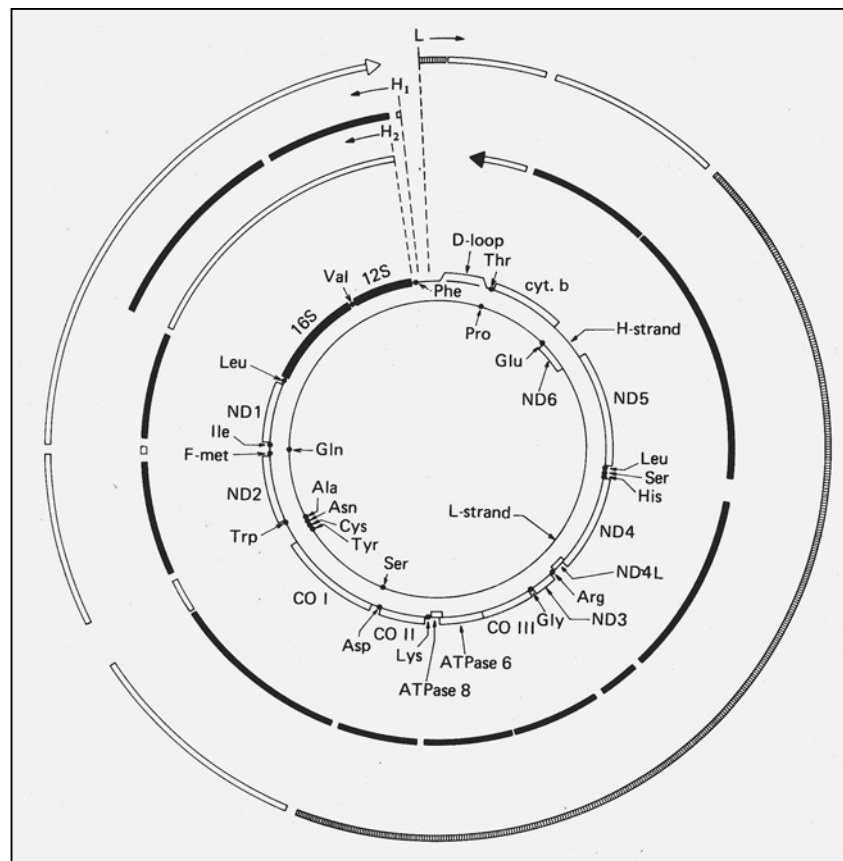


Fig. 2. Genetic and transcription maps of the human mitochondrial genome. The two inner circles show the two strands of mitochondrial DNA, H and L, with the positions of the 2 rRNA genes (12S and 16S, in black), 22 tRNA genes (black dots) and 13 protein coding genes (in white). In the outer portion of the diagram, curved black bars represent the identified functional RNA species other than the tRNAs resulting from processing of the two polycistronic primary transcripts of the H-strand starting at H_1 (rDNA transcription unit) and H_2 (total H-strand transcription unit). Cross-hatched bars represent the identified RNA species resulting from processing of the polycistronic primary transcript of the L-strand. The white bars represent unstable, presumably non-functional by-products. COI, COII and COIII: subunits I, II and III of complex IV; cytb: apocytochrome b; ATPase 6 and 8: subunits 6 and 8 of complex V. ND1-ND6: subunits 1-6 of complex I (adapted from Attardi, 1986).

Human mtDNA encodes two ribosomal RNAs (12S and 16S), 22 tRNAs and 13 peptides, all of them components of the respiratory chain, namely:

- Subunits 1, 2, 3, 4, 4L, 5 and 6 of complex I (NADH-coenzyme Q oxidoreductase).
- Subunit b (cytochrome b) of complex III (cytochrome c reductase).
- Subunits I, II and III of complex IV (cytochrome c oxidase).
- Subunits 6 and 8 of complex V (ATPase)

The genetic information on human mtDNA is extremely compact, which has no parallel in the living world except in viral genomes (Anderson *et al*, 1981, 1982; Bibb *et al*, 1981). Apart from a short segment around the origin of replication, mammalian mtDNA is completely saturated with genes, all of which lack introns. Most of the genes are transcribed from the H-strand, including the 2 rRNA genes, 14 tRNA genes and 12 protein coding genes (Ojala *et al*, 1980). These genes are in most cases butt jointed to each other or separated by only a few nucleotides. Almost all reading frames lack significant nontranslated flanking regions (Montoya *et al*, 1981), and in some cases genes overlap. That is the case for subunits 6 and 8 of ATPase, and also for ND4 and ND4L, which overlap in 46 and 7 nucleotides, respectively. Most genes also lack a complete termination codon. In these cases, completion of the termination codon TAA occurs postranscriptionally by polyadenylation of the mRNAs (Ojala *et al*, 1981). Another distinctive feature of the mammalian mitochondrial genome is the scattered distribution of the tRNA genes, which separate with nearly absolute regularity the rRNA and protein-coding genes (see fig. 2). These tRNA structures appear to function as signals for the RNA processing enzymes that generate the mature RNA species.

In addition to the highly compacted genes described above, mammalian mtDNA also contains a small non-coding region, which in the case of humans is 1122 nucleotides long. This region is located between tRNA^{pro} and tRNA^{phe} genes and contains the replication origin for the H-strand, the promoters for transcription of both strands, as well as the D-Loop, and is therefore referred to as the control region (some authors also refer to it as D-loop region; see fig. 3). This region also contains three highly conserved sequence blocks (CSBs I, II and III), believed to play a relevant role in the control of priming of mtDNA replication (Chang and Clayton, 1985, Chang *et al*, 1985, Xu and Clayton, 1996).

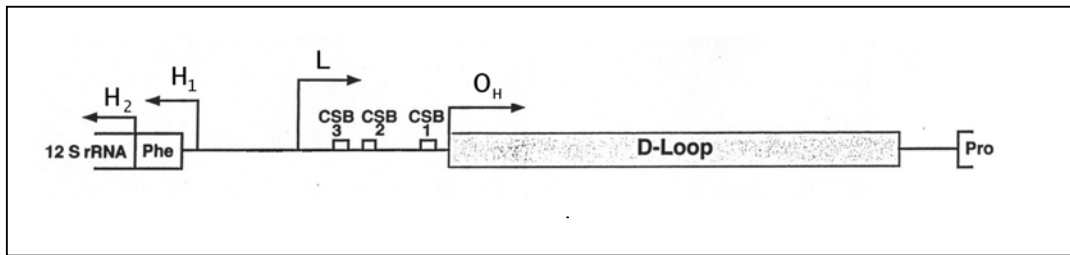


Fig. 3. Schematic representation of the non-coding region of the human mitochondrial DNA. It contains the heavy strand promoters (H_1 and H_2), the light strand promoter (L), the initiation sites for the replication of the heavy strand (O_H represents the main initiation site, at position 191), the D-loop and the three conserved sequence blocks, CSB I, CSB II and CSB III.

1.3.1) Replication of human mtDNA

For the past 20 years, the strand-asymmetric model of replication of mtDNA has been widely accepted (Clayton, 2000). However, recent studies have raised a controversy concerning the validity of this model (Holt *et al*, 2000; Yang *et al*, 2002). In this chapter we will describe the strand-asymmetric model in detail, and will include a comment about the works by Holt and co-workers and Yang and co-workers.

1.3.1.1) Strand-asymmetric model of replication

According to this model (see fig. 4), a commitment to mtDNA replication begins by the initiation of H-strand synthesis that results in strand elongation for the entire length of the genome. Initiation of the L-strand synthesis only occurs after the initiation site for replication of the L-strand (O_L) is exposed as a single-stranded template by displacement. Thus, in this model, the H-strand (leading-strand) origin is the dominant element for initiation of replication, while the L-strand (lagging-strand) origin plays an essential but secondary role.

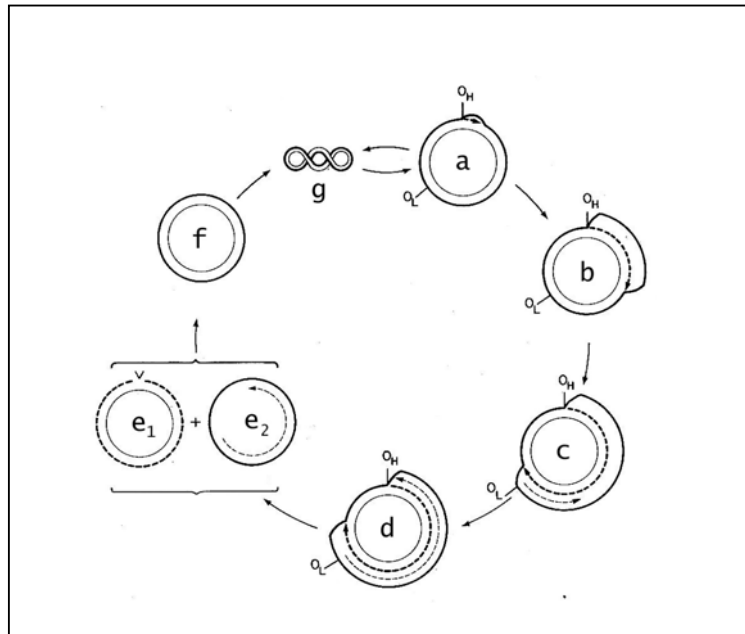


Fig. 4. Asymmetric model of replication. Thick solid lines: parental H-strands. Thin solid lines: parental L-strands. Thick dashed lines: daughter H-strands. Thin dashed lines: daughter L-strands. O_H and O_L : origins of H- and L-strand synthesis, respectively. (a): D-loop containing close circle mtDNA. (b): replication intermediate in which H-strand synthesis has proceeded past the D-loop. (c) and (d): replicative intermediates in which O_L has been exposed, allowing L-strand synthesis to proceed. (e_1) and (e_2): daughter molecules, both of which are converted into closed circles (f), which in turn becomes superhelical mtDNA (g) by the addition of about 100 negative superhelical turns. (Adapted from Clayton, 1982).

Initiation of replication of the H-strand requires the synthesis of an RNA primer (Chang and Clayton, 1985; Chang *et al*, 1985) by the same transcription machinery described below. This primer is synthesized from the L-strand promoter, starting at the transcription initiation site at position 407.

The next step is the formation of an extremely stable RNA-DNA hybrid, called R-loop (Lee and Clayton, 1996; Lee and Clayton, 1997). This structure (see fig. 5) consists of the two parental DNA strands and the RNA primer, and seems to be of a unique type, involving interactions of all three strands, rather than a simple displacement-loop (Clayton, 2000). R-loop formation requires the presence of intact CSB II and CSB III regions (Scheffler, 1999).

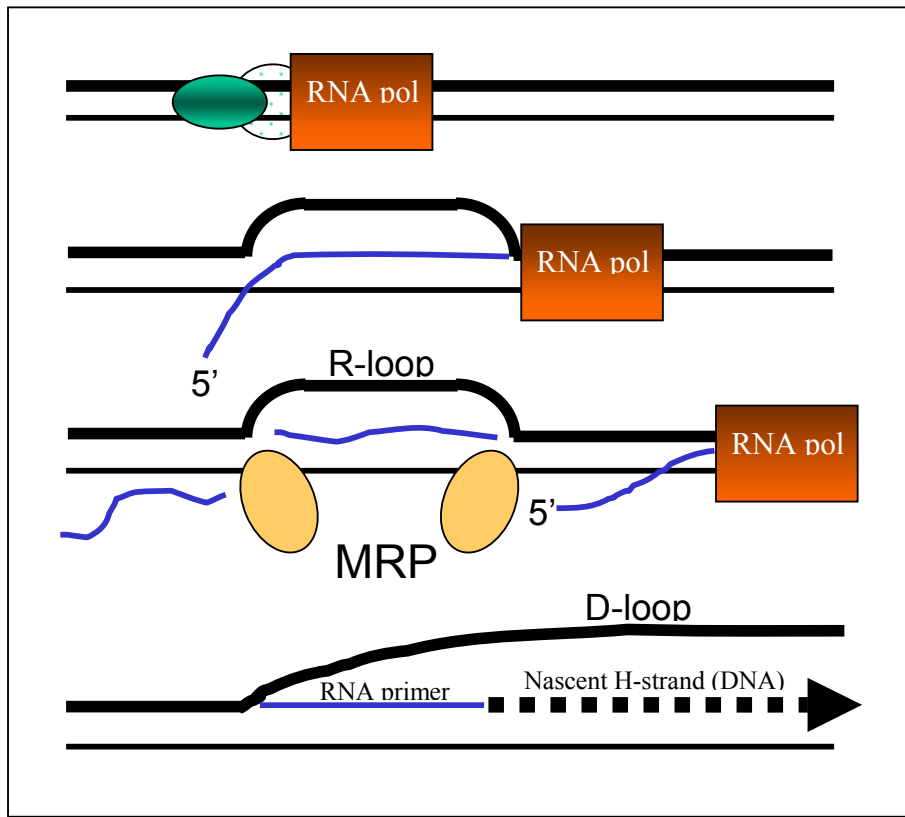


Fig. 5. Schematic representation of the generation of the RNA primer for DNA replication. RNA polymerase and factors generate an RNA transcript. This transcript is cleaved by RNase MRP, leaving a short RNA tightly annealed to the DNA and creating an R-loop. This RNA serves as primer for DNA synthesis. (Modified after Scheffler, 1999).

For an RNA molecule to serve as a primer for replication it must have a 3'-hydroxyl group available for extension by DNA polymerase. In humans, the transition RNA-DNA occurs at different points, between positions 100 and 320 of the mtDNA, but most DNA nascent chains start at position 191 (Chang and Clayton, 1985; Kang *et al*, 1997). To provide such primer, the RNA transcript initiated at the L-strand promoter is processed by a site-specific endoribonuclease, RNase MRP (see fig. 5). This enzyme was first identified in mouse and human cells, where it was shown to be able to process origin-containing RNA substrates at sites that match some of the DNA replication sites. The enzyme contains, in addition to protein components, an RNA essential for its activity (Chang and Clayton, 1987a; 1989; Topper and Clayton, 1990). This RNA component has

been characterized in a variety of organisms (Schmitt *et al*, 1993). Interestingly, cleavage by RNase MRP requires a triple stranded structure, like the R-loop described above. Cleavage by this enzyme does not occur when single stranded RNA is used as a substrate (Clayton, 2000). Testing RNase MRP with an appropriate R-loop substrate has given a cleavage pattern matching precisely the majority of RNA to DNA transition sites mapped in mouse mitochondrial DNA (Chang and Clayton, 1987b). These cleavages were completely dependent on the presence of CSB I. Similar results were obtained using human RNase MRP on a human R-loop substrate (Clayton, 2000).

Once the RNA-DNA transition has occurred, the first thousand basepairs of nascent H-strand remain associated with the circular parental molecule, forming the D-loop, until replication is permitted to progress further around the circle. Whether true replication requires new initiation or just elongation of the preexisting D-loop strands is not known.

In vertebrates there is only one DNA polymerase devoted to mtDNA synthesis, called DNA polymerase γ (pol γ). Pol γ is distinguished from other cellular DNA polymerases by certain chemical criteria, including high activity with synthetic RNA templates *in vitro*, inhibition by both *N*-ethylmaleimide and dideoxynucleoside triphosphates, resistance to aphidicolin and stimulation by salt. Pol γ from all sources studied co-purifies with a 3' – 5' exonuclease domain, probably responsible for the very high fidelity reported for this enzyme (Kunkel, 1985; Wernette *et al*, 1988), which contrasts with the high mutation rate observed in mtDNA. Pol γ is a heterodimer composed of a larger catalytic subunit (~125 – 140 kDa) that confers both DNA polymerase and exonuclease activities and a smaller subunit (35 – 40 kDa) of unknown function (Shadel and Clayton, 1977). This smaller subunit presents structural similarities to aminoacyl-tRNA synthetases, and it has been suggested that its function might be related to the recognition of tRNA-like primer structures within the mtDNA (Fan *et al* 1999).

1.3.1.2) Alternative model of replication

The first work challenging the strand-asymmetric model of replication appeared in the year 2000. In this work, using two-dimensional agarose gel electrophoresis (Brewer and Fangman, 1987;1988;1991) the authors described the existence of two types of replication intermediates (RIs) in mtDNA from human and mouse cells (Holt *et al*, 2000). In addition to S1 nuclease-sensitive RIs, compatible with the classic strand-asymmetric model, they found strong evidence of the presence of double-stranded RIs, suggestive of a strand-coupled mechanism of replication, analogous to the one observed in nuclear DNA. The authors proposed that both mechanisms coexist, and that the strand-asymmetric model operates mainly on basal conditions, while the strand-coupled model is activated in those stress situations in which the mtDNA needs to replicate more than once per cell cycle, for example, after depletion of mtDNA by exposure to ethidium bromide or dideoxycytidine.

Recently, the same group moved a step further, concluding that the partially single-stranded RIs observed previously were attributable to an artifact of extraction due to the presence of RNase H in the preparations of mitochondria (Yang *et al*, 2002). According to these authors, RNase H degrades the RNA component of the RNA:DNA hybrid regions known to exist in mtDNA (Grossman, 1973), rendering artifactual single-stranded areas that were misinterpreted as strand-asymmetric RIs. When mtDNA replication intermediates prepared from sucrose gradient-purified mitochondria were analyzed, they were in fact almost entirely duplex. This led the authors to conclude that mammalian mtDNA replication proceeds mainly, or exclusively, by a strand-coupled mechanism.

In view of the evidences presented by Holt and co-workers, it is perhaps too early to abandon the asymmetric model proposed by Clayton. Nevertheless, the work by Holt and co-workers has supposed a fascinating re-visit to a topic that seemed well established, like mtDNA replication. Furthermore, they have proved that mtDNA replicates, at least partially, or in some circumstances like mtDNA depletion, by a strand-coupled mechanism.

1.3.2) Transcription of human mtDNA

Transcription of mtDNA also differs from transcription of nuclear genes. The first reports on the subject showed that, once transcription is initiated, both strands are transcribed completely (Aloni and Attardi, 1971a; Aloni and Attardi, 1971b; Murphy *et al*, 1975). the L-strand is transcribed 2-3 times faster than the H-strand (Cantatore and Attardi, 1980), although most of its transcripts have a much shorter half-life than those from the later, and 29 out of 37 mitochondrial genes are located in the H-strand.

Oligo (dT)-cellulose chromatography separates mtRNAs in two fractions: polyadenylated and non-polyadenylated. The non-polyadenylated mtRNAs correspond to the two ribosomal RNAs (12S and 16S) and the 22 tRNAs. CH₃HgOH-denaturing electrophoresis of the polyadenylated fraction identifies 18 species of mtRNAs. These contain a poly-A tail about 55 nucleotides long at its 3'-end (Amalric *et al*, 1978; Attardi, 1984; Hirsch and Penman, 1974; Montoya *et al*, 1981; Ojala and Attardi, 1974). This poly-A tail is shorter than the one observed in cytoplasmic mRNAs, is not DNA-encoded and it is added to the mRNA after transcription. By convention, each species is assigned a number according to its molecular weight. The three biggest ones (1,2 and 3) and the smallest (18) are encoded by the L-strand. The rest are encoded by the H-strand.

There are two overlapping H-strand transcription units, each of them starting at a different nucleotide (Montoya *et al*, 1982; Chomyn and Attardi, 1992). One is located 19 nucleotides upstream of the tRNA^{Phe} gene (H₁), and the other is very close to the 5'-end of 12S rRNA (H₂). The same authors also described the only initiation site for transcription of the L-strand, located at position 407, in the control region. Both H-strand transcription units overlap in the region of the rRNAs (Montoya *et al*, 1983). Transcription initiated at H₁ terminates at the 3' end of 16S rRNA and is responsible for the synthesis of both rRNAs. The primary transcript (u4) is processed by cleavage of tRNA^{Phe} and a 25 nucleotide leading sequence to render RNA u4a, which is further processed to produce the mature forms of rRNAs 12S and 16S. The other H-strand transcription unit, starting from H₂, is synthesized at a 20 to 50-fold lower rate (Gelfand

and Attardi, 1981) and produces a giant polycistronic species which covers almost the entire length of the H-strand. Transcription of the L-strand follows a similar pattern. In this case, a single polycistronic species covering the entire length of the L-strand originates from a single initiation site. Since the primary transcripts contain tRNA sequences interspersed between and contiguous to rRNA and/or mRNA sequences, the existence of a RNA processing apparatus recognizing tRNA sequences as processing signals was postulated (Ojala *et al*, 1981). Accordingly, a RNase P activity responsible for the endonucleolytic cleavage of tRNAs at its 5'-end was characterized in HeLa cells (Doersen *et al*, 1985, Puranam and Attardi, 2001), and a 5'- and a 3'- tRNA processing activity was identified and characterized in rat liver mitochondria (Manam and Van Tuyle, 1987). The HeLa cell enzyme has been shown to contain an RNA and a protein moiety that are both necessary for activity.

The differential expression of the two H-strand transcription units involves an attenuation phenomenon at the border between the 16S RNA and tRNA^{Leu(UUR)} genes (Christianson and Clayton, 1988; Kruse *et al*, 1989). A central role in this attenuation is played by the mitochondrial transcription termination factor (mTERF), a DNA-binding protein that protects a 28 bp region within the tRNA^{Leu(UUR)} gene at a position immediately adjacent to and downstream of the 16S rRNA gene (Kruse *et al*, 1989). This protein will be the object of a detailed discussion in the last chapter of this introduction.

Recently, there have been interesting developments in the identification of the components of the human mitochondrial transcription machinery. A partially purified fraction of mitochondrial RNA polymerase, known as h-mtRPOL (Tiranti *et al*, 1997), and recombinant TFAM protein (also known as mtTFA. See Fisher and Clayton, 1985; Parisi and Clayton, 1991) have been known to be sufficient for activating transcription from light and heavy strand promoters *in vitro* (Dairaghi *et al*, 1995). However, attempts to reconstitute human mtDNA transcription with recombinant h-mtRPOL and TFAM proteins proved unsuccessful, leaving the field open for the identification of other transcription factor(s) necessary to support transcription. Falkenberg and co-workers identified two transcription factors, B1 (TFB1M) and B2 (TFB2M) necessary for basal transcription of mammalian mtDNA (Falkenberg *et al*, 2002; Shoubbridge, 2002). Each

factor alone can support promoter-specific mtDNA transcription in a pure recombinant *in vitro* system containing h-mtRPOL and TFAM. The fact that TFB2M is at least one order of magnitude more active than TFB1M might allow flexible regulation of mtDNA gene expression (Falkenberg *et al*, 2002). TFB1M is identical to mtTFB, a protein previously cloned and characterized by Shadel and colleagues (McCulloch *et al*, 2002).

1.3.3) Translation of human mtDNA

The main function of the mitochondrial translation machinery is to provide a few components of the oxidative phosphorylation system, 13 in the case of humans (see above). This machinery is composed of components encoded by the nuclear and mitochondrial genomes. Interestingly, this system is not essential for cell survival, as long as glucose is abundantly available for glycolysis. Unlike replication and transcription, many aspects of mitochondrial DNA translation are still obscure, and all attempts to translate mitochondrial mRNAs with a mitochondrial set of tRNAs, ribosomes and translation factors *in vitro* have failed so far.

The mitochondrial genetic code presents some differences in codon recognition, when compared to the universal code, as depicted in the following table:

	<u>Universal</u>	<u>Human mtDNA</u>
UGA	STOP	Tryptophan
AUA	Isoleucine	Methionine
AUU	Isoleucine	Methionine
AGG, AGA	Arginine	STOP

The human mitochondrial genome presents an unusual codon recognition system which allows proper reading of all the genetic code with only 22 tRNAs, far fewer than the number available in the cytosol. This system is based on the recognition of the two

first bases of the codon (Lagerkvist, 1978). For 8 aminoacids, the third base is not discriminative. Thus, CUX is translated as leucine, GUX as valine, UCX as serine, CCX as proline, ACX as threonine, GCX as alanine, CGX as arginine and GGX as glycine. For each one of these aminoacids there is only one tRNA. For the other aminoacids, the first two bases are also discriminative, and depending on the third base being A-G or U-C, the aminoacid will vary. For example, AAU-AAC are both translated as asparagine, and AAA-AAG as lysine. Also in this case, there is one tRNA for each aminoacid.

Human mtDNA encode only one tRNA^{f-Met} (having the anticodon CAU), and it is the only tRNA for methionine. Therefore, both internal (AUG, AUA and AUN) and start codons (UUG, GUG and GUU) must be recognized by this tRNA, if it is assumed that all peptides are initiated with formylmethionine.

When mitochondrial ribosomes were discovered, their similarity to bacterial ribosomes, in terms of sensitivity to chloramphenicol and insensitivity to cycloheximide, was used as an early argument in favour of the endosymbiont hypothesis. But a comparison of their sedimentation coefficients (70S for bacterial and 55S for mitochondrial) already gave an indication that the similarity was limited. And when the size of the ribosomal RNAs was compared, the difference became even more evident, as depicted in the following table:

	Prokaryotes	Mammalian (cytosol)	Mitochondria
Large rRNA	2900 nt / 23S	4800 nt / 28S	~ 1600 nt / 16S
Small rRNA	1540 nt / 16S	1900 nt / 18S	~ 950 nt / 12S
5.8S	-	160 nt / 5.8S	-
5S	120 nt / 5S	120 nt / 5S	120 nt / 5S*

* The 5S is not present in vertebrate mitochondria, but is common in plants.

Comparing the buoyant density in CsCl gradients of mitochondrial ribosomes (1.43 g/cm³) with that of cytoplasmic robosomes (1.58 g/cm³) it can be deduced that the former have a higher protein/nucleic acid ratio (O'Brien and Denslow, 1996). It is estimated that there are ~80 ribosomal proteins, all nuclear-encoded, and while some are

homologous to their prokaryotic counterparts, many are unique to mitochondrial ribosomes (O'Brien and Denslow, 1996).

The extreme economy of spacing genes on mammalian mitochondrial genes has already been discussed. mRNAs do not possess 5' or 3' untranslated regions (UTRs) nor 5' caps, and some terminal stop codons are completed by the posttranscriptional addition of a poly-A tail. By contrast, almost all cytoplasmic mRNAs are capped and have 5' and 3' UTRs of varying size. These elements are crucial in determining the intracellular localization of the mRNA, its stability and the efficiency with which an mRNA is translated (Belasco and Brawerman, 1993; Sachs *et al*, 1997). If similar mechanisms are applicable to mammalian mitochondrial mRNAs, these regulatory elements must be contained within the coding sequences. The same dilemma is not posed in yeast. The mitochondrial genome is much larger, and contains long intergenic sequences. In this case, the importance of the mitochondrial 5'UTRs has been convincingly demonstrated. They are the target for imported proteins controlling the turnover and translation of these mRNAs (Scheffler, 1999).

So far, a single initiation factor (IF-2_{mt}) and three elongation factors (EF-Tu_{mt}, EF-Ts_{mt} and EF-G_{mt}) have been purified from animal mitochondria (Schwartzbach *et al*, 1996). EF-Tu_{mt} and EF-Ts_{mt} have been shown to be able to replace their corresponding factors in *E. coli* in an *in vitro* translation system using bacterial ribosomes, a clear indication that elongation in mammalian mitochondria might function like in prokaryotes. On the other hand, the mechanisms of initiation and termination remain virtually unknown.

2) Leucine zippers

As discussed in the next chapter, mTERF contains three leucine zippers. Due to the importance of this structural feature as a support for the main hypothesis behind this project, a detailed description of this motif has been included in this introduction.

The leucine zipper was first proposed as a hypothetical structure in 1988 (Landschulz *et al.*, 1988). This proposal derived from studies on C/EBP, a transcription factor with a distinct DNA-binding specificity. A cDNA clone encoding C/EBP had been isolated and the minimal DNA-binding motif identified was a region with a predicted α -helical structure. In fact, a 35 amino acid region was proposed to form an amphipathic helix with a leucine residue every seventh position, corresponding to two helix turns. Furthermore, computer-assisted searches showed that a related structure was found within the C-terminal domains of other DNA-binding proteins, such as the yeast transcription factor GCN4, *c-fos* and *c-jun*. However, the only conserved amino acids between all these proteins were the leucine residues. As the resulting helix is amphipathic, to be stable in solution a matching surface had to be provided, and Landschulz *et al.* suggested that this might be the same structure in a second protein monomer. The leucine zipper was therefore proposed to be a dimerization motif allowing the association of dimers of a single protein or potentially different proteins.

Leucine zippers are often represented in the form of a helical wheel diagram with the seven amino acids of each repeat referred to by the letters **a** to **g**, with the leucine residues at position **d** (fig. 6). Another characteristic of this motif is the high frequency of hydrophobic β -branched amino acids (valine, threonine, isoleucine) at position **a**. This gives the classic 4-3 repeat of hydrophobic residues that characterizes all leucine zippers (Alber, 1992; Hurst, 1994). Apart from these two common features, there are protein families sharing other characteristics. The most studied is the bZip protein family (for reviews, see Hurst, 1994; Alber, 1992). GCN4 (O'Shea *et al.*, 1991; ElleMBERG *et al.*, 1992), *c-fos*, *c-jun* (Neuberg *et al.*, 1989; van Dam and Castellazzi, 2001) and C/EBP (Calkhoven *et al.*, 1992; Landschulz *et al.*, 1988), among others, are members of this family. All of them are transcription factors, therefore they all are DNA-binding proteins,

and their leucine zippers are essential for their DNA-binding activity. bZip proteins present a basic domain adjacent to the NH₂-terminus of the leucine zipper. The monomeric form of these proteins is inactive, and in order to interact with DNA they require the formation of homo- and in some cases heterodimers (such is the case for *c-fos* and *c-jun*). Leucine zippers act as protein-protein interaction domains, forming a parallel coiled-coil structure^{*}, and thus bringing the two basic domains in close contact, allowing interaction with the DNA target sequence. Consequently, it has been shown that the DNA binding sites for all these factors show dyad symmetry. Thus, each binding site is constituted by two 'half sites', each of which is contacted by the basic region of one of the polypeptides in the dimer. To further stabilize the interaction between the two leucine zippers, these proteins show an accumulation of charged residues at positions e and g (see fig. 7).

However, not all transcription factors with leucine zippers belong to the bZip family. Such is the case for the heat shock transcription factor, HSF. Genes encoding HSF have been isolated from yeast, *Drosophyla*, tomato, human, mouse and chicken (Clos *et al*, 1990; Nakai and Morimoto, 1993; Rabindran *et al*, 1991; Sarge *et al*, 1991; Scharf *et al*, 1990; Schuetz *et al*, 1991; Wiederrecht *et al*, 1988). All these homologue proteins share two main conserved regions, the DNA binding domain and the oligomerization domain. The DNA binding domain is located close to the NH₂-terminus of the protein and does not resemble any known DNA-binding motif. It contains a three-helix bundle that is capped by a four-stranded antiparallel β sheet (Harrison *et al*, 1994). The oligomerization domain, located in the middle of the polypeptide, is composed of two leucine zippers, and is responsible for the formation of homotrimers through a parallel triple coiled-coil structure (Peteranderl *et al*, 1999). Oligomerization confers DNA-binding activity (Westwood and Wu, 1993; Zuo *et al*, 1994), though transcription activation activity requires hyperphosphorylation (Xia and Voellmy, 1997 and references therein). The inactive monomeric polypeptide is in a dynamic complex with Hsp90 or,

* Coiled coils are formed by two or three α helices in parallel and in register that cross at an angle of 20 degrees, are strongly amphipatic, and display a pattern of hydrophilic and hydrophobic residues that is repeated every seven residues (Lupas *et al*, 1991 and references therein).

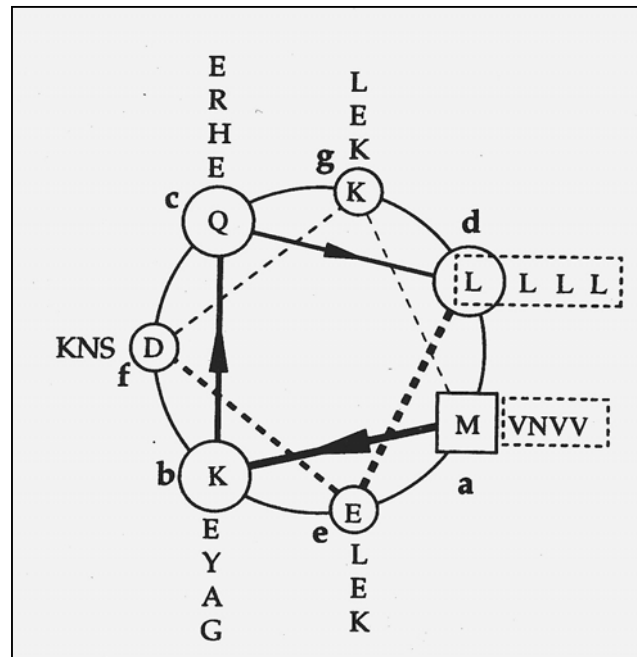


Fig. 6. Organization of a leucine zipper. This example corresponds to residues Met²⁵⁰ to Glu²⁸¹ of the GCN4 protein sequence. View is from the N-terminus, and the residues in the first two helical turns are boxed or circled. Heptad positions are labeled **a** through **g**. (From Harbury *et al*, 1993)

possibly, an Hsp90-containing multichaperone complex (Ali *et al*, 1998; Zou *et al*, 1998, Guo *et al*, 2001). HSFs from higher eukaryotes contain an extra leucine zipper region near the C-terminus (Rabindran *et al*, 1991; Schuetz *et al*, 1991), and a role of this domain in the stabilization of the monomeric form has been proposed (Zuo *et al*, 1994).

The DNA-binding site of HSF, called heat shock element (HSE) is found in the promoters of stress-responsive genes and is composed of multiple inverted arrays of the pentameric consensus sequence 5'-nGAAn-3' (Amin *et al*, 1988; Perisic *et al*, 1989) and, despite the fact that HSF is not a bZip protein, HSE has dyad symmetry.

Finally, there is a functionally heterogeneous group of leucine zipper proteins, all of them lacking DNA-binding activity. Members of this group are: SRSEC, haemagglutinin and spectrin. SRSEC is a seryl-tRNA synthetase from *E. coli*, formed by two identical subunits interacting through their leucine zipper domains (Cusack *et al*, 1990).

With regard to the structure of mTERF, the most evident feature of its primary structure is the presence of three leucine zippers (Fernandez-Silva *et al*, 1997, see figs. 10 and 11a). The most typical leucine zipper is Lz3, situated near the COOH-terminus, between residues 292 and 326, in which the characteristic heptad, with an abundance of hydrophobic residues at positions a and d is repeated five times. Another potential leucine zipper motif (Lz2), with five heptad repeats, occurs between residues 185 and 219. Finally, in the segment between positions 116 and 171, a bipartite leucine zipper is located. It consists of a two-heptad repeat between residues 116 and 129 (Lz1a) and a three heptad repeat between residues 151 and 171 (Lz1b), separated by a stretch of 21 residues.

Another relevant structural feature of mTERF is the presence of two basic domains (Fernandez-Silva *et al*, 1997, see figs. 10 and 11a). One (B1) is a 22 amino acid segment between residues 70 and 91, in the NH₂-terminus-proximal third of the protein, and contains seven basic and two acidic residues. The second one (B2) is a 16 amino acid stretch adjacent to the COOH-end of Lz3. It contains seven basic residues, and only one that is acidic. Deletion and site-directed mutagenesis experiments demonstrated that these two basic regions and the three leucine zipper domains are necessary for the DNA-binding activity of mTERF.

Another relevant contribution by Fernandez-Silva and co-workers was to show that, unlike other transcription factors with leucine zipper motifs, mTERF binds DNA as a monomer. This fact was clearly demonstrated by direct measurement of the molar ratio of mTERF to DNA in band-shift experiments in which mTERF and the probe had been previously labeled with ³⁵S and ³²P, respectively, as well as band-shift experiments with mixtures of wild type and truncated forms of mTERF and after treatment with cross-linking agents. The first experiments showed a molar ratio of 0.9, and the band-shift experiments failed to find any evidence of homo- or heterodimers binding to DNA. This conclusion is consistent with the fact that the mTERF binding site lacks a dyad symmetry (Kruse *et al*, 1989), as is expected for binding sites of typical bZip transcription factors (Hurst, 1994). This fact led the authors to propose that mTERF leucine zippers establish intramolecular interactions (possibly a trimeric coiled-coil structure) that bring the basic

which folding of the protein is assumed to bring two separate DNA-binding domains adjacent to each other, thus conferring DNA-binding activity (Morrow *et al*, 1993).

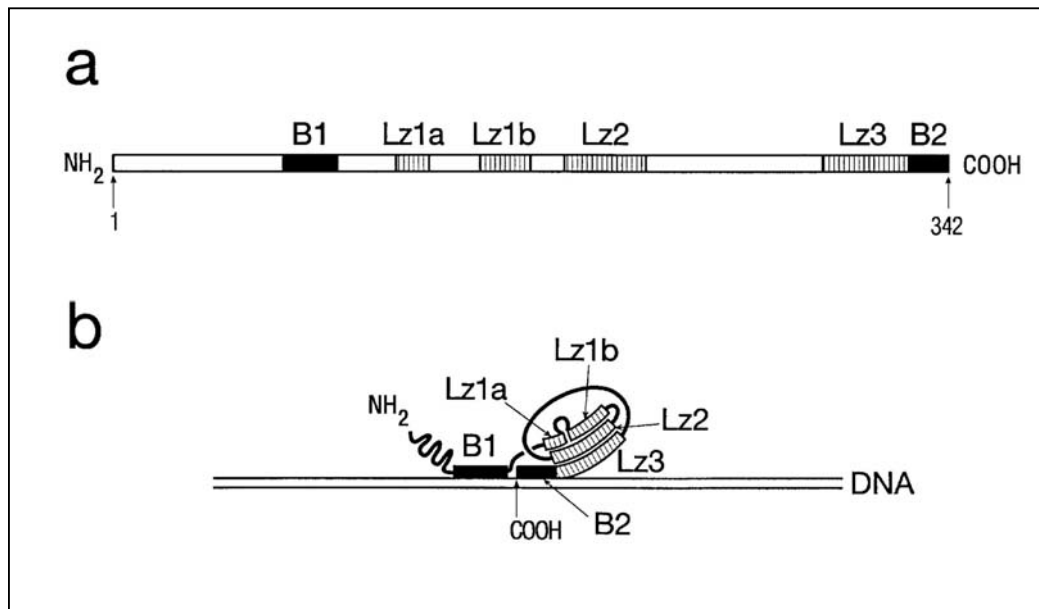


Fig. 11. Proposed model of the role of the leucine zippers in DNA binding of mTERF. (a) Linear representation of mTERF structural features. (b) proposed mechanism of DNA binding to mTERF. The model assumes that intramitochondrial interactions of the leucine zipper domains of the protein are required to bring the basic domains in close register with the mTERF target DNA sequence. (From Fernandez-Silva *et al*, 1997).

Many of the structural and functional characteristics of mTERF have been further corroborated in mtDBP, a sea urchin protein identified in 1999, with high homology with mTERF (Loguercio-Polosa *et al*, 1999). A BLASTP analysis yields 22% amino acid identity and 61% amino acid similarity between these two proteins. MtDBP is a 40 kDa protein which binds two regions of sea urchin mtDNA (one at the 3'-end of the D-loop region and the other one in the boundary between the oppositely transcribed ND5 and ND6 genes). It contains two leucine zipper domains, one of which is bipartite, and two small N- and C- terminal basic domains, all of them, except the basic C-terminal region, are necessary for binding to its target sequence. MtDBP binds DNA as a monomer, and is able to promote transcription termination in a heterologous *in vitro* system, using human

transcriptional machinery and DNA templates containing human promoter regions and sea urchin mtDBP binding sites (Fernandez-Silva *et al*, 2001).

It is interesting to compare the mitochondrial transcription termination system with that involved in nuclear transcription termination of rRNA genes. Transcription of rRNA genes in the nucleus is carried out by RNA polymerase I (Pol I). These genes are arranged in tandem in mammalian cells, and their ~ 14 kb coding region is separated by ~ 30 kb intergenic sequences. Eukaryotic ribosomal transcription units are flanked both at their 5'- and 3' – side by one or more terminator elements. In the mouse, a repeated 18 bp sequence motif (AGGTCGACCAGA/TT/ANTCCG), termed 'Sal box', functions as the transcription terminator (Grummt *et al*, 1985; Bartsch *et al*, 1987). In addition, a T-rich sequence upstream of the terminator has been shown to be involved in both release of terminated transcripts and 3'-terminal processing of pre-rRNA (Kuhn *et al*, 1990; Mason *et al*, 1997). The 'Sal box' is recognized by TTF-I (transcription termination factor), a specific DNA binding protein that stops elongating Pol I when bound to the terminator sequence (Evers *et al*, 1995; Grummt *et al*, 1986). Cloning of mouse and human TTF-I revealed homology between the DNA binding domains of TTF-I and Reb1p, the Pol I termination factor in *Saccharomyces cerevisiae* (Morrow *et al*, 1993). DNA-bound TTF-I on its own is not sufficient for transcript release. In mammals, dissociation of the ternary transcription complex at the termination site requires another factor, PTRF (Pol I and transcript release factor), a 44 kDa protein that interacts with Pol I, TTF-I and the 3'-end of pre-rRNA (Mason *et al*, 1997; Jansa *et al*, 1998). PTRF releasing activity requires the presence of the T-rich sequence mentioned above. Thus, termination of Pol I-dependent transcription of nuclear rRNA genes in mammals requires two DNA domains, the 'Sal box' and the T-rich element, and two proteins, TTF-I that stops elongating Pol I and PTRF that dissociates TTF-I-paused transcription complexes.

The fact that binding sites for TTF-I are present both upstream and downstream of the rDNA transcription unit suggests a functional linkage between transcription initiation and termination. A model has been proposed in which each rDNA transcription unit forms a protein-mediated loop that connects the promoter and termination regions (Kempers-Veenstra *et al*, 1986; Kulkens *et al*, 1992). According to this model, Pol I molecules that

have terminated nascent pre-rRNA chains could be transferred directly to the gene promoter without being released from the template. TTF-I would be a perfect candidate for mediating intramolecular interactions between the 5' and 3' ends of rDNA. In support of this view, TTF-I was shown to interact simultaneously with two separate DNA fragments bearing 'Sal box' elements (Sander and Grummt, 1997) and, therefore, to be potentially capable of linking the promoter-proximal 'Sal box' with the distal terminator site. Further testing of this model, comparing transcription on templates that contained the upstream and downstream 'Sal box' elements, or either one of them, showed that, although transcription was more efficient when termination was allowed to occur, TTF-I-mediated stimulation of Pol I transcription did not require the presence of the promoter-proximal 'Sal box', indicating that the increase in the amount of transcripts was not due to 'handover' of Pol I via a DNA loop, and that the increase in transcription initiation was rather caused by an increase of Pol I available after its PTRF-mediated release from the termination site (Jansa *et al*, 2001)

Understanding of the functional role of mTERF acquired special significance after the demonstration that an A→G transition within the mTERF DNA-binding site is associated with several human diseases, like MELAS encephalopathy (Goto *et al*, 1990, Kobayashi *et al*, 1990), progressive external ophthalmoplegia (Hurko, 1991) and some forms of adult onset diabetes (van der Ouweland *et al*, 1992), and by the fact that this mutation reduces the binding affinity of mTERF for its target sequence (Hess *et al*, 1991; Chomyn *et al*, 1992). Hess and co-workers reported a severe impairment of 16S rRNA transcription termination associated with the mutation using *in vitro* transcription assays with templates containing the mutation, but these results were not confirmed by Chomyn and co-workers, who found in transformant cells lines containing virtually pure mutant mtDNA that, although the MELAS mutation causes defects in protein synthesis and in respiration, the steady-state amounts of mitochondrial rRNAs, mRNAs and tRNA^{Leu(UUR)} are not significantly affected when compared to the parental cell line (Chomyn *et al*, 1991).

OBJECTIVES

- The main objective of this project is to study the capacity of the human mitochondrial transcription-termination factor (mTERF) to interact with other peptides, or with other mTERF molecules, to form a protein complex.

- In order to achieve the main objective, a preliminary aim is to obtain polyclonal antibodies against mTERF.

- If it is proven that mTERF is a component of a protein complex, secondary aims will be:
 - Determination of DNA-binding and transcription-termination activities of the protein complex.
 - Identification of the components of the protein complex.