Stony Brook University

Alterations to the unique DNA binding mode mediated by MTERF1 have implications of pathogenesis in mitochondrial disease.

A Dissertation Presented

by

James Byrnes

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

August 2014

Stony Brook University

The Graduate School

James Byrnes

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Miguel Garcia-Diaz – Dissertation Advisor
Associate Professor, Department of Pharmacology

Markus Seeliger - Chairperson of Defense
Assistant Professor, Department of Pharmacology

Mark Bowen – Committee Member
Assistant Professor, Department of Physiology and Biophysics

Daniel Bogenhagen – Outside Member
Professor, Department of Pharmacology

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Alterations to the unique DNA binding mode mediated by MTERF1 have implications of pathogenesis in mitochondrial disease.

by

James Byrnes

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

2014

Human mitochondria are found in all eukaryotic cells. They are dynamic double membrane organelles that can vary in shape, size and amount depending on cell types. In humans, mitochondria contain a 16.5kb genome that encodes for 22 tRNAs, two rRNAs and 13 proteins that make up the mitochondrial contribution of the oxidative phosphorylation system (OXPHOS). This system produces the majority of energy used by the cell in the form of ATP and is critical for cell viability. In order for the OXPHOS system to produce ATP, it is dependent on the proper expression of the mitochondrial genome. Thus, errors in gene expression can lead to defects in OXPHOS and ultimately pathogenesis of mitochondrial disease. Therefore, it is important to understand the mechanisms of mitochondrial gene expression and how alteration of these mechanisms can result in pathogenesis. The initial stage in gene expression is transcription, a process that involves several events including initiation, elongation and termination, all of which are regulated by proteins that are produced in the nucleus and transported to the mitochondria. Interestingly, termination is one of the least understood aspects of transcription. It has been established that termination is regulated by the mitochondrial transcription termination factor MTERF1. Our lab has solved the crystal structure of MTERF1 bound to its canonical DNA sequence located within the $tRNA^{Leu}$ gene and made great progress in understanding the mechanisms critical for MTERF1 mediated transcription termination. Most strikingly, MTERF1 exhibits a unique DNA binding mode that involves the formation of sequence specific interactions, helix unwinding and the eversion of three nucleotides stabilized by stacking interactions outside the double helix. Further structural and biochemical characterization of the base flipping mechanics has revealed a stepwise order to the base-flipping event that is important for function. In addition, we show that pathogenic mutations present within the DNA binding site perturb sequence recognition and base flipping. These result in termination defects that have implication for pathogenesis of mitochondrial disease.

Dedication Page

I would like to dedicate this work to my lovely wife, Lauren. I will always treasure your unwavering love, support and understanding. I hope we never keep it simple.

Frontispiece

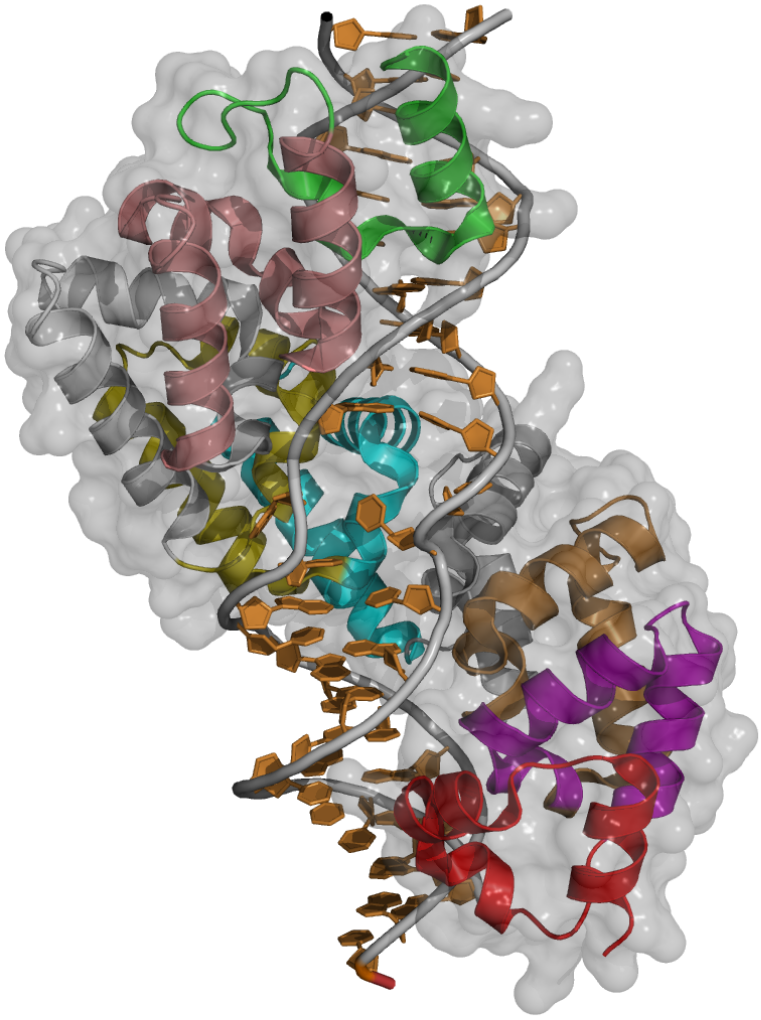

Table of Contents

List of Figures

List of Tables

List of Abbreviations

mtDNA- mitochondrial deoxyribonucleic acid

HS- Heavy Strand

LS- Light Strand

HSP- Heavy Strand Promoter

LSP- Light Strand Promoter

CsCl- Cesium Chloride

kb- kilobase

bp- base pair

OXPHOS- Oxidative Phosphorylation System

MTERF- Mitochondrial Transcription Termination Factor

POLRMT- Human mitochondrial RNA Polymerase

TFAM- Transcription Factor A of Mitochondria

TFB2M- Transcription Factor B2 of Mitochondria

MBP- Maltose Binding Protein

SNP- Single Nucleotide Polymorphism

NCBI- National Center for Biotechnology Information

NADH- Nicotinamide adenine dinucleotide

$FADH_2$- Flavin adenine dinucleotide

ETC- Electron transport chain

ATP- Adenosine Tri-phosphate

ADP- Adenosine Di-phosphate

P_i- Inorganic phosphate

tRNA- Transfer ribonucleic acid

rRNA- Ribosomal ribonucleic acid

mtRNA- Mitochondrial ribonucleic acid

mRNA- Messenger RNA

Kd- Kilodalton

mtDBP- mitochondrial DNA binding protein (of sea urchin)

KO- knockout

RNAP- RNA polymerase

ITC- isothermal calorimetry

dsDNA- double stranded DNA

ssDNA- single stranded DNA

MELAS- Mitochondrial encephalomyopathy, lactic acidosis and stroke-like episode

TEV- Tobacco Etch Virus

HMG- High Mobility Group

Acknowledgments

My scientific journey through graduate school was not taken alone. I am truly indebted to everyone who has played a part in the development of my scientific career. Whether we shared a story or two over lunch or a long scientific discussion over coffee, you have all had an impact on my scientific progress.

First I would like to thank Miguel Garcia-Diaz, my advisor. I truly appreciate your kindness and patience as well as your intelligence that you have shared with me throughout the years.

I would like to thank Kip Guja, Matt Burak, Elena Yakubovskaya, Woo Suk Choi and Edison Mejia for their insightful scientific discussions and friendship.

I would like to thank all of my students and particularly Leah Norona who began to characterize the mechanism of base flipping and Jessica Yamanda who helped me start the allelic variant project.

I would like to thank Kevin Hauser and the Carlos Simmerling for great discussions (and lunches!) and their MD work in analyzing the base-flipping mechanism of MTERF1. I truly appreciate our friendship.

Finally, I would like to thank all the members of my committee: Markus Seeliger, Mark Bowen and Dan Bogenhagen.

Chapter 1- Introduction

Mitochondria are organelles found in all eukaryotic cells and are sometimes referred to as the "powerhouse" of the cell (1). This title highlights the major role of mitochondria as an energy generator, which is critical for normal cellular function and viability (2). In humans, mitochondrial defects that result in alterations to energy production can lead to a wide range of clinical symptoms in any organ and at any age, making the diagnosis of mitochondrial disorders difficult (3-5). Antithetically, mitochondria, the same organelle that is responsible for cellular viability, can also carry out programmed cell death, known as apoptosis (6-8). In fact, these two functions, among others, are so important that certain types of cancers are able to manipulate mitochondria and utilize them to the advantage of the cancer cell (9,10). Cleverly, some cancers can turn off the apoptotic pathway in order to keep the cancer cell alive (11,12). In addition, to feed a growing cancer cell quickly, they can increase the rate at which energy is produced (Warburg effect) (13). Moreover, mitochondria have implications in aging and age related neurodegenerative diseases such as Parkinson's and Alzheimer's Disease (14-16). Thus, mitochondria are "powerful," multifunctional, complex organelles and play a critical role in the life cycle of a eukaryotic cell.

The double membrane structure of mitochondria is suited for high-energy output. The inner membrane, which is highly impermeable to small molecules, consists of many folds called cristae that increase the surface area of the inner membrane to the matrix and the intermembrane space (Figure 1.1) (17). In humans, embedded within the inner membrane is a series of five protein complexes (Figure 1.2) (18). Complexes I-IV catalyze a series of oxidative reduction reactions that remove electrons from

electron donors and pass them down to electron acceptors (19). Complexes I & II obtain their electrons from donors such as NADH and $FADH_2$ that are produced mainly from the glycolytic pathway. The electrons reduce ubiquinone to ubiquinol and shuttle the electrons to Complex III. Then, the electrons are then reduced to cytochrome c, which shuttles electrons to Complex IV. After the electrons enter into complex IV, the final step is the reduction of molecular oxygen (O_2) to water. This stepwise process is known as the electron transport chain (ETC). As the electrons move through the ETC, H^+ ions are generated due to the oxidation of NADH and $FADH_2$ to NAD^+ and FADH. With the exception of Complex II and V, the H^+ ions are then pumped from the mitochondrial matrix to the intermembrane space (Figure 1.2). Due to the impermeability of the inner membrane, the H^+ ions accumulate in the intermembrane space and generate a powerful electrochemical gradient. Under normal circumstances, the H^+ ions are able to cross the inner membrane and move down their concentration gradient in a controlled manner. This occurs at Complex V where the movement of H^+ ions drives a molecular motor located within the complex and catalyzes the formation of ATP from ADP and P_i. Thus, the energy stored in the electrochemical gradient is used to form ATP, which contains high-energy bonds utilized by the cell for various functions. This entire process constitutes the oxidative phosphorylation system (OXPHOS). In humans, OXPHOS occurs in the presence of oxygen and produces the majority of energy for the cell. Perturbations of this process can result in decreased production of ATP and has serious implications for cell viability (20,21).

Structure and organization of the human mitochondrial genome.

The structural components of the OXPHOS system requires the assembly of ~80-100 proteins (22). The majority of the proteins are encoded for and expressed in the nucleus. After expression in the nucleus, they are transported to the mitochondria where they are inserted into the inner membrane. Interestingly, core proteins that are part of Complexes I, III, IV and V are encoded by a separate genome that is located within the mitochondrial matrix. In humans, the mitochondrial genome contributes 13 proteins in total to OXPHOS (23). Perturbations to the expression of the genes encoding for the 13 proteins have been shown to result in decreased cellular levels of ATP (24-26). Therefore, the mitochondrial genome and its products play an integral role in maintaining levels of ATP (27).

In humans, depending on cell type and the number of mitochondria present, the total number of genome copies per cell can range from 10's to 1000's. The structure of the human mitochondrial genome resembles that of a bacterial plasmid and is a circular double stranded molecule, 16,569 bp in length (Figure 1.3) (23) (28). Each strand of the genome is referred to as the heavy strand (HS) or light strand (LS) as determined by their buoyant densities in a CsCl gradient (29,30). The genome not only encodes for the 13 proteins necessary for OXPHOS, but also encodes for 22 tRNAs and 2 rRNAs (Figure 1.3) that are distributed asymmetrically on both strands (23,28). The distribution of the genes is compact with little or no space in between (Figure 1.4). Interestingly, the genes lack introns and there are few noncoding regions (31). The

largest noncoding sequence corresponds to the control region known as the regulatory D-Loop (32,33). Here, transcription promoters for both strands as well as an origin for HS replication are present (34,35). Upon replication initiation of the HS, a short nascent strand is synthesized (complementary to the LS) that is terminated in a template directed manner (36). This nascent strand forms a stable hybrid with the LS. As a result of the hybrid, a DNA triplex structure is formed between the LS, the newly synthesized nascent DNA strand and the HS (Figure 1.5). The triplex structure causes the HS to bubble out forming the characteristic displacement loop from which the region gets its name (30). Interestingly, the site of LS replication is located far from the D-loop region in humans and may have alternate sites in other vertebrates.

Expression of the human mitochondrial genome.

Transcription initiation

The genes on the LS and HS are transcribed as long, continuous polycistronic transcripts (Figure 1.4) (34,37-39). The transcripts are organized such that the mRNA sequences for each of the 13 proteins and 2 rRNAs are flanked by tRNAs (Figure 1.4) (40). Thus, an extra step is required to excise the tRNAs, mRNAs and rRNAs from the continuous polycistronic transcripts. This excision process involves precise endonucleolytic cleavage on either side of the tRNA gene, which results in the formation of mature transcripts. Thus the large polycistronic mRNAs are subjected to an important RNA processing event associated with tRNA maturation and this is referred to as the "tRNA punctuation model" (41).

The formation of polycistronic mRNAs arises from several promoter sites (42-45). Transcription of the LS initiates at a promoter sequence within the D-loop region known as the light strand promoter (LSP) (Figure 1.4) (46). The long polycistronic mRNA produced from LSP encompasses almost the entire mitochondrial genome and includes eight tRNAs and one mRNA (47). Transcription of the HS is slightly more complicated and ultimately produces the other 14 tRNAs, 12 proteins and 2 rRNAs. Interestingly, the genes of the HS are not transcribed at the same level. The 2 rRNAs genes are found to be transcribed at a rate ~50-100 times higher than genes located downstream of the rRNAs on the HS (48). The mechanism by which the mitochondria are able to maintain differing ratios of HS transcripts *in vivo* is not entirely clear. However there are two general viewpoints. The first suggests that HS transcripts are all produced from one single promoter located within the regulatory D-Loop region (currently referred to as HSP). From this point of view, the different ratios of HS transcripts are maintained by a termination event at the $tRNA^{Leu}$ gene, which is located immediately downstream of the rRNA genes (Figure 1.4). Premature termination of transcripts at the $tRNA^{Leu}$ gene would prevent further transcription of HS genes downstream of this site. Therefore, a constant rate of transcription initiation from HSP would yield differing levels of HS transcripts because some will prematurely terminate immediately downstream of the rRNA genes. Thus, the termination event could serve as a mechanism for regulating HS transcription levels.

An alternative explanation to account for the differing amounts of HS transcripts is the existence of two promoters, referred to as HSP1 and HSP2. HSP1 is located within the regulatory D-loop and is the same promoter as mentioned above.

However HSP1, in this case, is thought to only transcribe the rRNA genes and terminate at the $tRNA^{Leu}$ gene. The transcripts for the rest of the 12 tRNAs and 12 mRNAs on the HS are thought to initiate from HSP2 (42,49,50). The site of HSP2 was identified in the early 1980s and mapped to a region outside the regulatory D-loop. Interestingly, it is found within the $tRNA^{Phe}$ gene located ~100bp downstream of the HSP1 site. Evidence suggests that transcription initiation from HSP2 is much weaker than from HSP1 *in vitro* (49). In this case, the weaker initiation event from HSP2 would account for the decreased amount of HS transcripts past the rRNA genes.

Transcription Initiation Machinery

It is widely accepted that transcription initiation in mitochondria is dependent on a basal level of machinery (51-53). All the transcription machinery is nuclear coded and imported into the mitochondria (54). In humans, transcription requires a specific mitochondrial RNA polymerase. This is known in humans as POLRMT and is distantly related to the T7 bacteriophage RNA polymerase (RNAP). Interestingly, comparisons between crystal structures of POLRMT and T7 RNAP reveal structural similarities in the C-terminal domain. However, unlike the T7 bacteriophage that is able to initiate transcription on its own, POLRMT requires two transcription factors (55). These two factors, TFAM and TFB2M are important for binding to and melting DNA at promoter sites (56). Furthermore, electron microscopy and molecular docking experiments have been able to associate all three factors at the LSP providing structural evidence supporting their role in the transcription initiation event (51).

TFAM is a highly conserved HMG box protein found in yeast, frogs and vertebrates (57). The crystal structure of TFAM bound to the LSP has been solved and demonstrated that it can bend DNA. The presence of two high mobility group boxes within the TFAM protein sequence mediates this bend, forming a U-turn in the DNA at the LSP. In addition to its role as a transcription factor, TFAM has implications in the packaging and maintenance of the mitochondrial genome. Biochemical studies have revealed that TFAM is found associated with nucleoid like protein complexes. These complexes are important for mtDNA packaging, maintenance and also associate with mitochondrial chaperones that assist in protein folding events (57). However, the mechanism by which TFAM maintains mtDNA levels is not entirely understood. Recent work has suggested that post-translational modifications of TFAM may regulate its levels in mitochondria (58). The association of TFAM with mitochondrial nucleoids indicates that TFAM is in close proximity with mtDNA (59-61). In fact, biochemical and crystallographic data have shown that TFAM is able to bind nonspecifically to mtDNA. In addition, the ability of TFAM to bend DNA suggests a possible mechanism by which TFAM may assemble mtDNA for packaging.

TFB2M, the second transcription factor of the basal machinery, is homologous to another transcription factor found in human mitochondria, named TFB1M. Both have been observed to initiate transcriptional activity *in vitro,* however, the activity of TFB2 is 1-2 orders of magnitude greater than TFB1M (62-64). In fact, biochemical and structural work has shown that TFB1M functions largely as a mitochondrial methyltransferase and can methylate the small mitochondrial ribosomal subunit, playing a critical role in ribosome assembly (65) (66).

Until recently, it has been widely accepted that mammalian mitochondrial transcription initiation requires POLRMT, TFB2 and TFAM as discussed above. However, recent work has demonstrated that transcription initiation from HSP and LSP can occur in the absence of TFAM (49,50,67,68). Furthermore, upon addition of a TFAM dosage, transcription levels are increased. This information suggests that the initiation event does not absolutely require TFAM but rather, acts as an enhancer for transcription. Curiously, the reproducibility of this observation varies throughout the mitochondrial transcription field. Thus, whether or not TFAM is absolutely essential for transcription initiation is still debated (49,68-71).

Transcription Termination at tRNA^{Leu} gene

RNA transcripts produced from HSP1 and LSP promoters have been shown to terminate at a region within the tRNA^{Leu} gene. Experiments in the early 1980s have demonstrated that the 3' ends of transcripts originating from HSP1 were heterogeneous (72). This finding suggested that these transcripts were not a result of precise excision events, but rather due to premature termination of the transcripts. Further experiments performed in the mid-1980s identified a factor from mitochondrial lysates promoting transcription termination at this site (73). In addition, DNA foot printing and S1 nuclease digestion assays established that the termination factor binds to a 28bp binding sequence within the tRNA^{Leu} gene (74,75). This factor is referred to as MTERF1, which stands for mitochondrial transcription termination factor. Strikingly, DNA methylation protection experiments have revealed that this site corresponds to one of the most protected sites of the entire genome and stresses the role of a binding event at this location (76).

Roles of MTERF1

The location of the MTERF1 binding site is immediately downstream of the 2 rRNA genes. This location has implications for the role of MTERF1 in regulating the quantity of rRNA transcripts produced from HSP1. An early model suggested that the DNA between HSP1 and the $tRNA^{Leu}$ looped out to form a structure known as the rDNA loop (Figure 1.6) (77). The rDNA loop would put the initiation (HSP1) and termination sites in close proximity, thus allowing the initiation machinery to be quickly recycled. Therefore, the authors proposed that this loop is important for maintaining higher levels of rRNA transcripts. Furthermore, it was suggested that MTERF1 would mediate the formation of this loop by having an additional binding site at HSP1. Thus the interaction of one MTERF1 molecule, one end at the canonical $tRNA^{Leu}$ site and the opposite end at the HSP1 site, would stabilize this rDNA loop (Figure 1.6).

Interestingly, MTERF1 exhibits termination polarity at the $tRNA^{Leu}$ gene (78). Reconstituted *in vitro* termination assays demonstrate that when the MTERF1 DNA binding sequence is oriented in reverse relative to the HSP promoter, termination activity is 2-3 fold higher than when the sequence is oriented in the forward direction (79-84). This would suggest that MTERF1 is able to terminate transcripts initiating from the LSP direction more strongly than from HSP. A reason for this observation might be to prevent the buildup of missense transcripts, as the LS does not encode for any genes beyond the termination site (Figure 1.4). More recently, an MTERF1 mouse knockout model has been constructed and demonstrates that in the absence of MTERF1 there is an increase in missense LS transcripts beyond the termination site

(82). In addition, the stronger termination of LSP products may also be important to prevent transcriptional interference at the initiation sites. Thus, MTERF1 would act as a roadblock to the transcription machinery, in order to prevent the machinery from entering into the D-loop region.

The role of the polar road blocking mechanism mediated by MTERF1 is not well understood. To recall, the regulatory D-Loop does not only contain promoter sites for transcription initiation, it also contains the origin of replication for the HS. It has been suggested that mtDNA replication and transcription initiation at the D-loop region are tightly coordinated as there is a high probability of potential collisions between the replication machinery and transcription machinery (85-87). These collisions are detrimental to the processivity of the transcription machinery (88-90). Interestingly, MTERF1 has been shown to bind multiple sites, including the regulatory D-Loop region. In addition, MTERF1 has been shown to regulate mitochondrial DNA replication pausing at the canonical $tRNA^{Leu}$ site and at alternative sites (91). It is suggested that the role of MTERF1 in this context is necessary in order to prevent collisions between the replication and transcription machinery. Although this role for MTERF1 in humans is not clearly understood, other systems, such as *E.coli* and sea urchin, have analogous approaches to prevent these collisions. For the case of *E.coli*, a DNA replication terminator named Tus is responsible for preventing these collisions by attenuating the DNA helicase DnaB (92). Tus is indirectly able to pause growing replication forks by attenuating the helicase. As a result, the transcription machinery is able to continue beyond the attenuation site. The ability of Tus to attenuate DnaB from one direction and allow passage of the transcription machinery from the other is

analogous to the MTERF1 polar termination mechanism. Similarly, mtDBP, a protein found in the mitochondria of sea urchin, is homologous to MTERF1 and also exhibits termination polarity much like MTERF1(93). In addition, mtDBP has been shown to have multiple binding sites in the sea urchin mitochondrial genome and has implications in regulating replication. Thus, the observation of multiple MTERF1 binding sites in the mitochondrial genome and its ability to pause replication, further suggests that MTERF1 is able to take on multiple roles in the maintenance of the mitochondrial genome, however the mechanisms by which MTERF1 can carry out these roles is not well understood.

The MTERF family of Proteins

Analysis of the MTERF1 protein sequence reveals significant sequence homology to a family of MTERFs that are shared amongst metazoans and plants (94). Interestingly, there can be a variable number of MTERFs found in different species (95). In some plants, there are upwards of 30 different types of MTERFs that are found in both the plastids and mitochondria (96). In all vertebrates, there are four MTERF proteins appropriately named MTERF1-4 (94,97,98). From an evolutionary standpoint, MTERF1 and MTERF2 are unique to vertebrates while MTERFs 3&4 are found in all metazoans, suggesting that MTERFs 3&4 are the ancestral genes in metazoans (96). In humans, MTERF1 was the first of the four MTERF proteins to be characterized (74). Currently, the functional roles of all four human MTERFs have been characterized through studies in mice. A summary of the putative roles for each human MTERF based on mouse knockouts have been assembled in Table 1.1.

Functional Roles of MTERFs 2, 3,4

The roles of MTERFs 2&3 are implicated in transcription regulation events and have been found to interact with DNA in the D-loop region. Individual mouse knockout studies of each MTERF reveal that they have opposite effects in the regulation of transcription initiation (94,99,100). For the case of MTERF2, the overall level of transcripts is reduced suggesting that it is a positive regulator of transcription initiation. Interestingly, a complete knockout of MTERF3 is embryonic lethal (Table 1.1), however tissue specific knockouts of MTERF3 in mouse heart are viable and reveal that mitochondrial RNA transcripts are actually increased in the absence of MTERF3. This suggests that MTERF3 is a repressor of transcription initiation (86,87,101). Furthermore, the absence of MTERF3 reveals decreased levels of mitochondrial translation. The phenotype of both the MTERF2 knockout (KO) and the MTERF3 conditional KOs reveal significant defects in the OXPHOS system. In the case of MTERF2, the observed defects in OXPHOS were only noticed in skeletal muscle (100). However, laboratory mice lead very sedentary lifestyles and thus do not require high ATP output from OXPHOS (100). To test for the effect of an MTERF2 knockout on a stressed OXPHOS system, mice were fed a ketogenic (high fat, low carb) diet (102). Interestingly, this experiment revealed more significant myopathies in tissues other than skeletal muscle. This suggests that the function of MTERF2, which did not have an initially severe phenotypic effect in the knockout, can present with severe deficiencies under conditions that mimic heavy use of the mitochondria, such as during exercise (100).

The mouse knockout of MTERF3 provides evidence for the significance of MTERFs for the viability of mouse. Interestingly another member of the MTERF family, MTERF4, the most divergent of the MTERFs, is the only other embryonic lethal KO out of the 4 human family members (103). Conditional knockouts of MTERF4 have revealed decreased levels of mitochondrial translation. This has been attributed to a reduction in ribosome assembly. Furthermore, studies show that MTERF4 is able to bind to the 12S rRNA, 16S rRNA and associates with a methyltransferase, NSUN4, found in mitochondria, that can methylate the 16S rRNA, which is a critical step for the assembly of the complete mitochondrial ribosome (103,104). Therefore MTERF4 is largely responsible for associating with NSUN4 and recruiting it to the 16S rRNA in preparation for assembly. These studies suggest a new role for the MTERFs that does not involve transcription, but rather in ribosome assembly. Curiously, MTERF4 was found to associate with a 7S RNA transcript (short transcript produced in D-loop region complimentary to LS) and thus puts MTERF4 in close proximity with the LSP. Therefore it was suggested that MTERF4 could be involved in crosstalk between transcription events at the D-loop and translation, although the molecular mechanism is not clear (103). Strikingly, the MTERF3 KO and knockdown studies in flies, has recently shown that it is involved in the biogenesis of the large ribosomal subunit. This may have implications for MTERF3 as not only a repressor of transcription initiation, as previously described, but also may mediate crosstalk between transcription and translation. Furthermore, investigation of the mouse Mterf3 KO also reveals alterations to the ribosome assembly process. This

provides a common function between MTERFs 3&4. The dual role in both transcription and translation may have implications for the lethality observed (87).

Structure of MTERF family of Proteins

As discussed above, the four human MTERF proteins are important for proper expression of the mitochondrial genome. However, the molecular mechanisms by which the MTERFs carry out their function can be characterized through X-ray crystallography. The structures of MTERFs 1, 3 and 4 have been solved and provide a wealth of structural and mechanistic information about the interactions that MTERF proteins mediate in humans (Figure 1.7).

The first members of the MTERF family to be structurally characterized were MTERF1 and MTERF3 (78,87,105). Interestingly, both exhibit highly homologous structures and are suited for their roles as DNA binding proteins. The crystal structure of MTERF1 has been solved by our lab and others, and reveals a modular protein made up of 8 mterf motifs that are all α helical. The mterf motif consists of 2 α helices followed by a 3_{10} helix and form a conserved triangular shape (Figure 1.7). The triangular shape is maintained by hydrophobic interactions between the helices. The overall MTERF fold is highly conserved amongst the family of proteins as MTERF3 and MTERF4 have a modular architecture and contain all α helical mterf motifs (104,106). The largest structural differences between MTERF1 and MTERFs 3 and 4 are in the N and C-terminal tails, respectively. For the case of MTERF4, the C-terminal tail is important for mediating interactions with NSUN4, the methyltransferase that is critical for mitochondrial ribosome assembly. Interestingly, the all α helical

folds of the MTERF proteins share structural homology to PUF domains (107). PUF domains are found in proteins that bind RNA. The domains are all α helical and form a crescent shape that is suited for binding RNA. The structure of PUF domains are very similar to the MTERF fold as they too are all α helical and exhibit a similar crescent shape. This suggests that MTERF proteins (other than MTERF 3 & 4 that have been biochemically shown to interact with RNA) that adopt this fold may be able to interact with an RNA substrate. However, the current structures of MTERF3 and MTERF4 do not contain nucleic acid substrates and are thus limited in providing the mechanical details regarding how this interaction may take place.

Implications for MTERF in pathogenesis of mitochondrial disease

There are a large number of mutations in the mitochondrial genome that are associated with pathogenesis of mitochondrial disease. Interestingly, many pathogenic mutations are localized within the tRNA genes (108). The $tRNA^{Leu}$ gene contains a high concentration of pathogenic mutations and interestingly is one of the most protected sites of the entire genome (76,109-113). As previously mentioned, the $tRNA^{Leu}$ region plays an important role in the regulation of transcription termination and includes the binding site for MTERF1. Strikingly, the binding sequence for MTERF1 alone contains nine pathogenic mutations. Among these include the most common mitochondrial mutation, A3243G that has been associated with multiple mitochondrial myopathies (Mitoweb). In addition, other mutations, such as G3249A result in a variant of Kearns Sayre Disease (Other mutations found in Table 4.3) (114). Interestingly, work done in our lab has shown that several of the pathogenic mutations found in this region alter MTERF1 mediated termination activity (78).

As mentioned above, our lab has solved the crystal structure of MTERF1 bound to the canonical $tRNA^{Leu}$ termination sequence. From this structure we discovered a unique DNA binding mode that involved helix unwinding and a rare three-nucleotide base-flipping event. However, we lack a clear understanding of how such a rare binding mode works and why it is necessary for MTERF1 termination activity. Furthermore, the presence of pathogenic mutations within the $tRNA^{Leu}$ gene has been shown to alter MTERF1 activity (78,111,115). This information may provide a link between mitochondrial transcription termination and the pathogenesis of mitochondrial disease. However, the molecular mechanisms that alter MTERF1 activity are not known. In order to learn about the unique binding mode adopted by MTERF1 and its link to mitochondrial pathogenesis, we approached this work in three different ways. First, we analyzed the mechanism by which MTERF1 is able to coordinate the flipping of three residues and assessed how alterations to this mechanism affect termination activity. Second, we structurally analyzed the role of pathogenic mutations found in the $tRNA^{Leu}$ gene and correlate the structural perturbations to functional defects in MTERF1 termination activity. Lastly, we studied the functional and structural alterations to the MTERF1 binding mode as a result of allelic variations found in MTERF1.

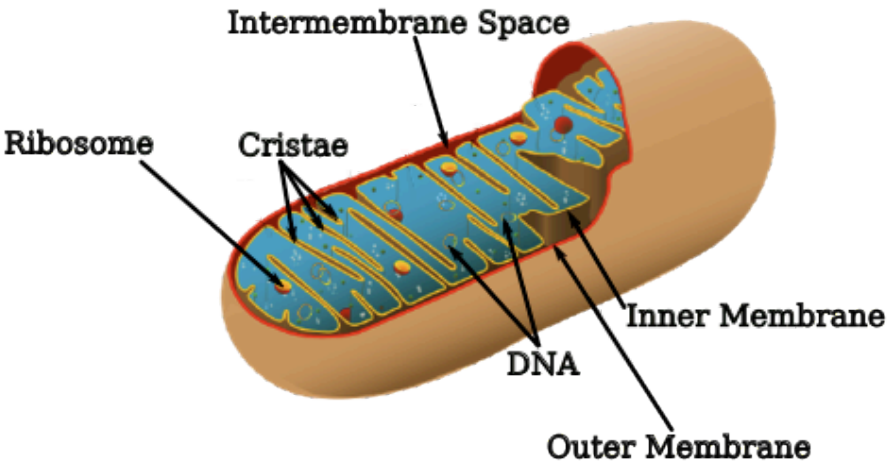

Figure 1.1 Structure of the mitochondria. *A single mitochondrion is a double membrane structure that consists of an inner and outer membrane. The space between the two membranes is referred to as the intermembrane space. The inner membrane is highly impermeable and consists of multiple invaginations known as cristae. The folds of the cristae increase the surface area of the membrane to the intermembrane space and matrix (matrix colored blue). The mitochondrial matrix contains the mitochondrial genome and ribosomes. In addition, many processes that occur in the mitochondria such as fatty acid oxidation and the TCA cycle take place in the matrix.*

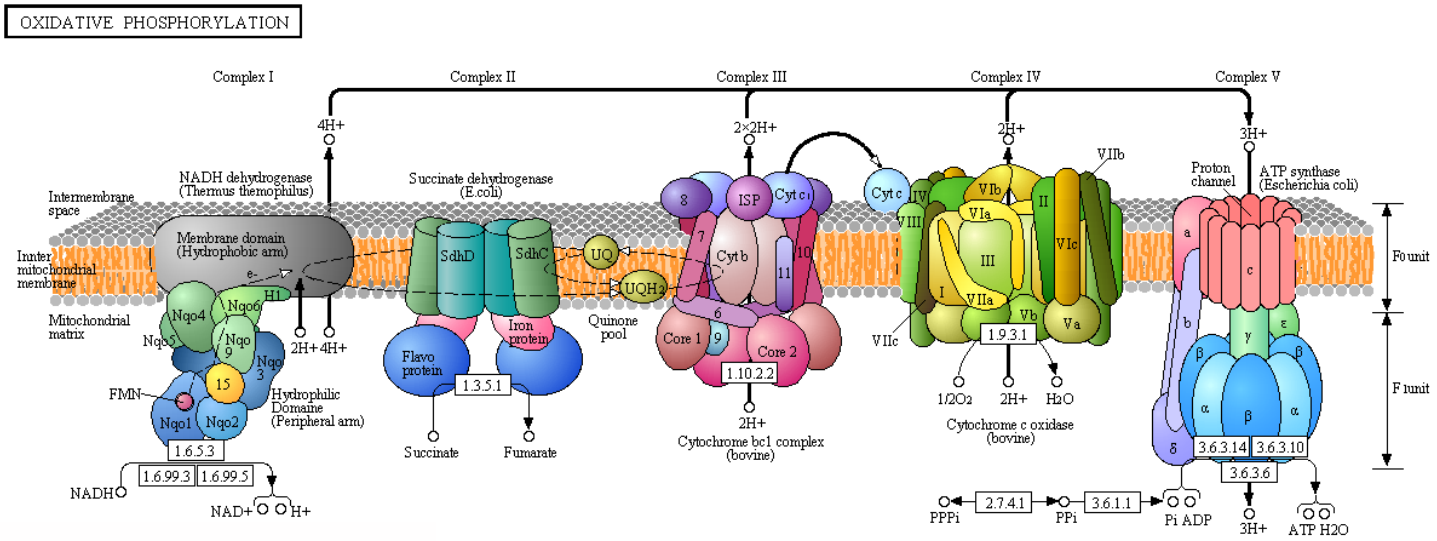

Figure 1.2 The Oxidative Phosphorylation System. *The OXPHOS system is made up of five protein complexes that are imbedded within the inner membrane of the mitochondria. Complexes I-IV catalyze a series of oxidation-reduction reactions that ultimately reduces molecular oxygen (O_2) to water. As a result of the oxidative reduction reactions, H^+ ions are pumped from the mitochondrial matrix to the intermembrane space creating an electrochemical gradient. Due to the impermeability of the inner membrane the H^+ ions move down their concentration gradient through Complex V. This drives a molecular motor in Complex V to generate energy in the form of ATP. The protein that make up the OXPHOS system are largely nuclear encoded, however core proteins for each complex, except Complex II, contain proteins encoded in the mitochondrial genome. * Image adapted from the Kyoto Encyclopedia of Genes and Genomes (KEGG), by Kanehisa Laboratories.*

.

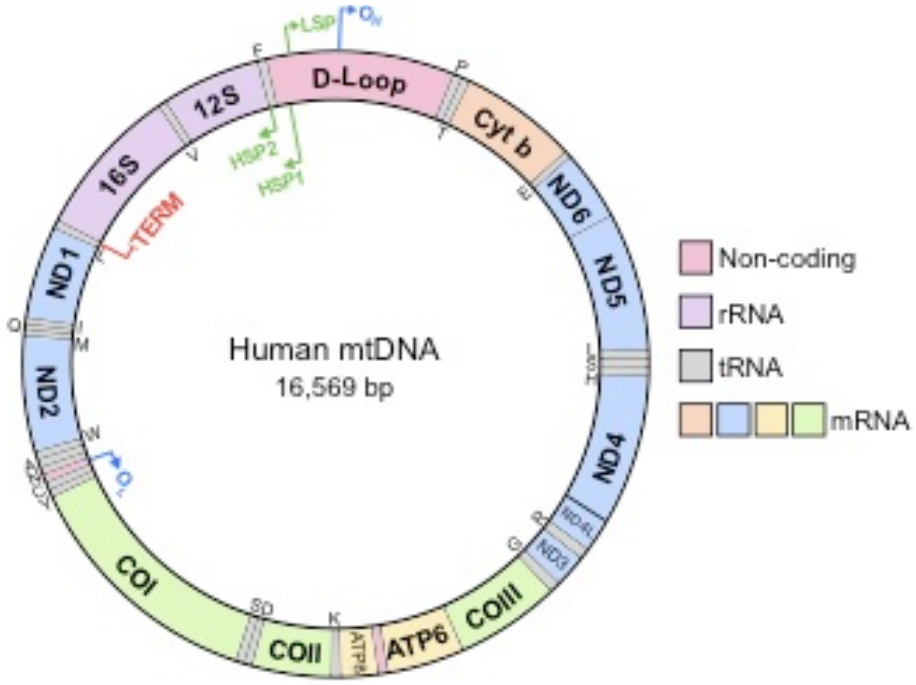

Figure 1.3 Structure and organization of the human mitochondria genome. *The genome encodes for 22 tRNAs 2rRNAs and 13 proteins that make up the mitochondrial contribution to OXPHOS. The molecule is made up of two strands referred to as the heavy and light strands. The genes are tightly clustered with little or no space between them. There is one major noncoding region referred to as the regulatory D-Loop, which contains promoters for each strand and an origin of heavy strand replication (O_H). There are two promoters for HS transcription. HSP1 is located within the regulatory D-Loop and HSP2 is located ~100bp downstream of HSP1 within the $tRNA^{Phe}$ gene. A transcription termination site is located immediately downstream of the 2 rRNA genes in the $tRNA^{Leu}$ gene (TERM).*

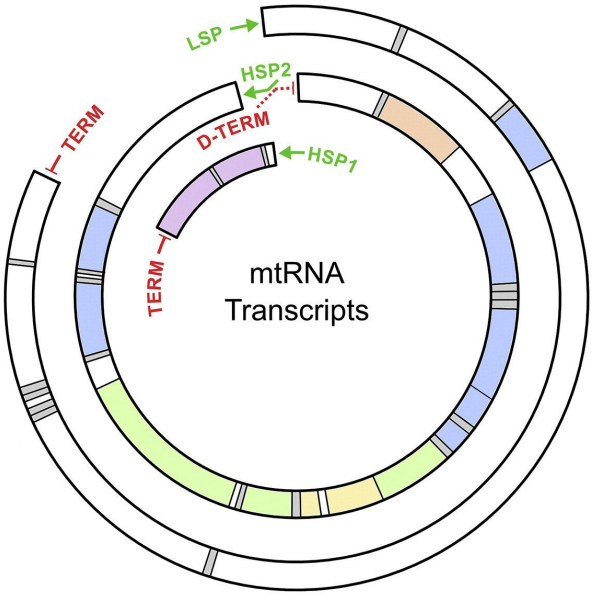

Figure 1.4 Mitochondrial Transcription products of the HS and LS. *Transcription of the mitochondrial genome produces polycistronic mRNAs that are then processed to produce mature mRNAs, tRNAs and rRNAs. Transcription from LSP produces a long polycistronic mRNA that terminates within the $tRNA^{Leu}$ gene. Transcription of the HS produces two transcripts. Transcription from HSP1 produces a short transcript that contains the 2rRNAs and terminates within the $tRNA^{Leu}$ gene. Initiation from HSP2 produces a long polycistronic mRNA that encompasses the entire mitochondrial genome.*

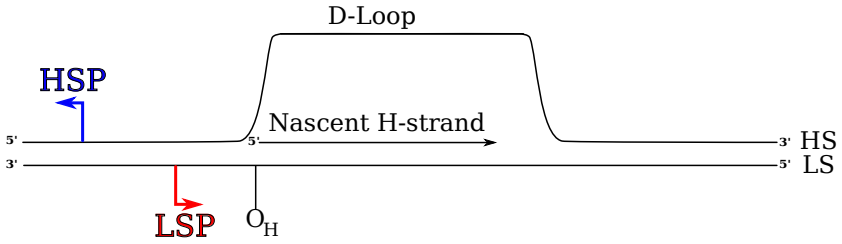

Figure 1.5 The D-loop structure. *The formation of the D-loop is a result of a triple helix that forms upon replication initiation at the origin of heavy strand replication O_H. The nascent H-strand forms a stable hybrid to the LS and results in the bubbling out of the HS. This bubble is referred to as the D-loop.*

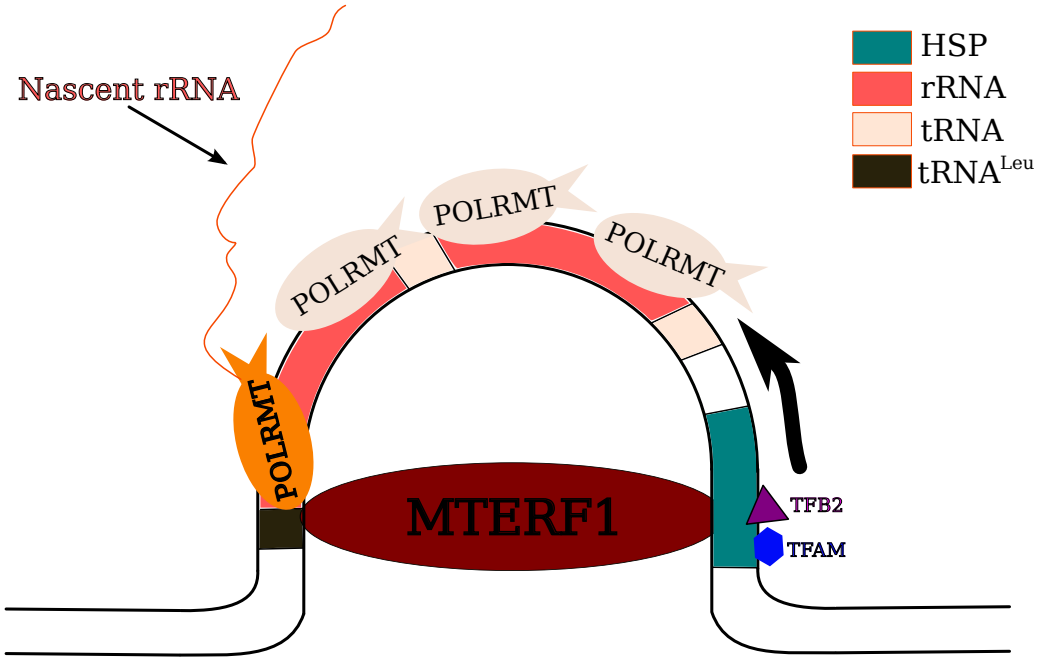

Figure 1.6 Structure of the rDNA loop. *Transcription from the HS has been shown to maintain higher levels of rRNA transcripts compared to genes downstream of the $tRNA^{Leu}$ gene. Electron microscopy data has provided evidence that a looping out of the DNA between HSP and the $tRNA^{Leu}$ gene forms the rDNA loop pictured above* (116). *This loop puts the HSP (aquamarine box) in close proximity to the $tRNA^{Leu}$ gene (dark brown box). Thus, POLRMT can be easily recruited back to the HSP site and initiate another round of transcription. MTERF1 is thought to stabilize the rDNA loop by forming simultaneous interactions with the $tRNA^{Leu}$ gene and HSP.*

	Mouse KO embryonic lethal?	X-ray Crystal Structure?	Function
MTERF1	No	Yes	Transcription terminator at $tRNA^{Leu}$
MTERF2	No	No	Positive Regulator of transcription initiation at D-Loop
MTERF3	Yes	Yes	Repressor of transcription initiation at D-Loop. Involved in ribosome biogenesis.
MTERF4	Yes	Yes	Associates with the methyltransferase NSUN4. Ribosome Biogenesis.

Table 1.1 Summary of the four human MTERF proteins

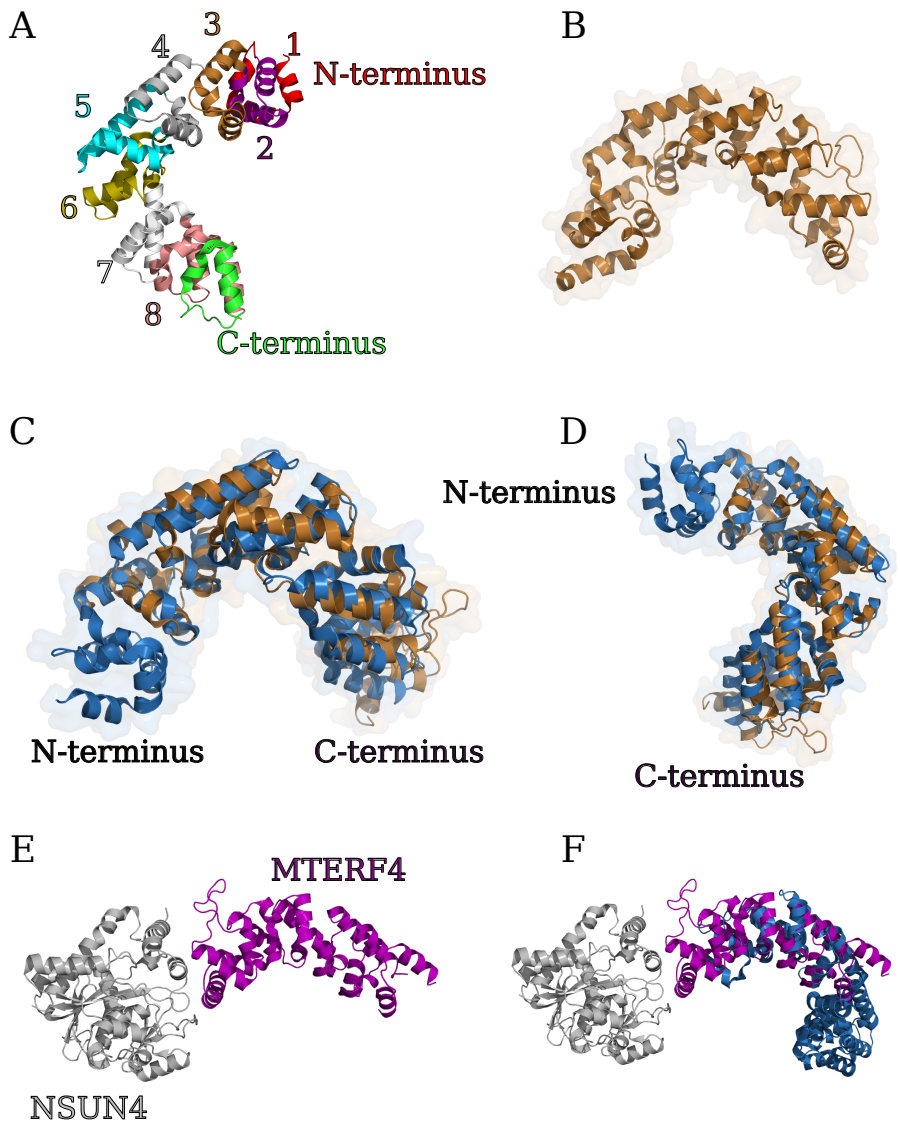

Figure 1.7 X-ray crystal structures of human MTERFs 1, 3 & 4. *(A) The crystal structure of MTERF1 (DNA not shown) reveals an all α-helical protein consisting of 8 mterf motifs (colored and numbered). The mterf motifs are conserved among MTERF1, MTERF3 and 4. (B) Crystal structure of MTERF3. (C) Overlay of MTERF1 (blue) with MTERF3 (copper) reveals a highly similar structure. The largest difference is at the N-terminus. (D) Same as (C) however this view shows the major difference in the N-terminus. (E) Crystal structure of the MTERF4:NSUN4 complex. MTERF4, highlighted in purple shares structural homology to MTERF1 as shown in (F). (F) MTERF1 colored blue in an overlay with MTERF4 (purple).*

Chapter 2 – Materials & Methods

The MD methods and simulations were performed in collaboration with Kevin Hauser and the Carlos Simmerling laboratory.

Mutagenesis

Wild type human MTERF1 (residues 57-399) was expressed as a fusion with maltose binding protein (MBP) (to improve solubility during expression) and a 6X His tag. The HMBP3 plasmid was used and the gene expression is controlled by binding of T7 polymerase to the T7 promoter upstream of the insert. MTERF1 substitution constructs were created using the QuikChange Site-Directed Mutagenesis protocol (Stratagene). PCR products were cleaned using the MinElute PCR purification cleanup kit (Qiagen) and successful mutagenesis was confirmed by DNA sequencing.

Protein Expression

MTERF1 constructs were electroporated into Arctic Express DE3 RIL cells using a standard protocol. Cells were grown overnight in LB with Kanamycin (50μg/μl Final concentration) at 37°C with shaking. The next day, the culture was scaled up 1:100 and grown at 37°C with shaking. During log growth phase, 500 μL of cells were mixed 1:1 with 100% glycerol and frozen at -80°C for glycerol stocks. Once an OD_{600}=0.6 was reached 0.2 mM IPTG (Final concentration) was added to induce expression of the T7 polymerase under the control of the lac promoter. The temperature was then reduced to 15°C incubated overnight. A sample was taken before and after induction with IPTG and ran on an SDS-PAGE gel to confirm expression of the 70 kD MTERF1/MBP/6XHIS fusion construct. Cells were harvested in 1L bottles at 4500 rpm using a Sorvall Evolution RC centrifuge for 7 minutes. After

the cells were collected, the pellet was transferred to a 50 mL tube, gently resuspended in a 0.9% NaCl solution and then spun down at 7000 rpm for 20 minutes. The supernatant was discarded and the pellet was frozen at -80°C.

Protein Purification

All steps for the purification of MTERF1 were done on ice or in the cold room as apo MTERF1 is highly unstable at room temperature. Pellets were thawed on ice and resuspended in Lysis Buffer containing the following: 1M KCl, 20mM Hepes pH8.0, 20mM Imidazole, pH8.0 and 5% glycerol and 1 protease inhibitor tablet (Roche cOmplete Mini, EDTA-free). After resuspension, cells were lysed by sonication using a microtip set at a power level of 5. The slurry was pulsed 4X for 30 seconds with a one-minute rest in between for a total process time of 2 minutes. After sonication, the cells were spun down at 18000 X g for 30 minutes.

The lysate was then incubated with Ni-NTA beads (Invitrogen) for one hour with rotation. After the incubation, the beads were washed with lysis buffer and then the protein was eluted in Elution Buffer consisting of: 0.5 M KCl, 0.5 M Imidazole, pH 8.0, 20 mM Hepes pH 8.0 and 5% glycerol.

The elution from the Ni-NTA column was subjected to an overnight TEV (tobacco etch virus) digestion at 4°C to remove the 6X His and MBP tags (1 mg of TEV/25 mL of elution buffer).

After digestion, the sample was carefully diluted with Buffer A (20 mM Hepes pH 8.0, 5% glycerol, 1 mM EDTA, 1 mM DTT) to a final salt concentration of 0.2 M KCl. The diluted MTERF1 solution was filter though a 0.45 μm filter. Further purification was carried out using an Akta FPLC. The protein was loaded on a pre-

packed 5 mL heparin column (HiTrap Heparin HP, GE healthcare), equilibrated in 20% Buffer B (1 M KCl, 20 mM Hepes pH 8.0, 5% glycerol, 1 mM EDTA, 1 mM DTT) washed with 5 column volumes (CV) of 20% Buffer B and then eluted using a linear gradient (20%-80% Buffer B) over 20 CV. Two peaks were observed, one at 30mS/cm corresponding to TEV and one at 50 mS/cm corresponding to MTERF1. After elution, the sample was filtered through a 0.2 μm filter and subjected to size exclusion chromatography using a HiLoad 16/600 Superdex 200 PG (Prep Grade) column (GE healthcare) equilibrated in Buffer C (500 mM KCl, 5% glycerol, 1 mM DTT, 1 mM EDTA, 20 mM Hepes pH 8.0). Fractions corresponding to monomeric MTERF1 were then concentrated using a 10 K MWCO spin concentrator (Corning) and spun down at 4000 rpm using a Sorvall Evolution RC centrifuge in 15 minute intervals with mixing in between. The samples were concentrated in a range from 15 mg/ml to 35 mg/ml then flash frozen and stored at -80°C.

Transcription termination

Transcription assays were adapted from Asin-Cayuela (83) and described in Yakubovskaya, et al (78). Briefly, 290 bp of the mitochondrial genome was cloned into a circular template DNA plasmid (pet 22b, Novagen) containing the heavy strand promoter (HSP1) and the MTERF1 binding sequence 100bp downstream in reverse orientation with respect to HSP1. The end of the 290 bp insert was flanked with a HindIII restriction site and the circular plasmid was linearized using HindIII. After digestion, the template DNA was then incubated with MTERF1 (8.0 pmole, 4.0 pmoles

or 2.0 pmoles), rNTPs (0.4 mM ATP, 0.15 mM CTP and GTP, 0.01 mM UTP), DNA template (30ng, 8.0 fmoles), 0.5ul (5 μCi) of α-^{32}P labeled UTP and the initiation machinery TFAM (7.5 pmoles), TFB2 (3.0 pmoles) and POLRMT (3.0 pmoles) and transcription buffer (150 mM KCl, 20 mM Hepes pH 8.0, 5 mM DTT, 1 mM EDTA and 10 mM $MgCl_2$) in a total reaction volume of 20 μl. The reaction was carried out for 30 minutes at 32°C and stopped using 100 μl of 1% SDS, 20 mM EDTA, 300 mM Sodium Acetate and 20 μg Calf thymus DNA. The samples were then ethanol precipitated and run on a 5% UREA-PAGE gel. The gel was dried and exposed to a phosphorimager (Amersham Biosciences) and scanned using the Typhoon FLA 9000 scanner and software package (V1.2, GE). Each experiment was repeated at least three times. Densitometry analysis was performed using the ImageQuant TL (GE) software. For each gel a control experiment was performed that did not contain MTERF1 but included the DNA template, initiation machinery, rNTPs and labeled UTP. The result of this control experiment reveal a large band corresponding to the runoff transcript, and a weak band that is similar in size to what would be the termination band and is proportional to the runoff band. However, since the control experiment does not contain any MTERF1, this weak band is not a result of MTERF1 mediated termination. Presumably, this band would contribute in part to the termination band observed when MTERF1 is present and falsely increase the signal from the actual termination band. In order to eliminate the contribution of this band from MTERF1 termination, using densitometry, the ratio of the weaker band to runoff transcription was obtained. This value was then multiplied by the runoff counts in each lane to obtain a background value. The background value was subtracted from the termination band in each lane.

In addition a correction factor was applied to account for the increased UTP incorporation observed (factor of 1.5) in the runoff band and the percentage termination was calculated. Graphs were made using Prism 5 for Mac OS X version 5.0c.

X-ray Crystallography

Crystallization

The wild type 22mer DNA oligo was added to each of the MTERF1 substitution proteins for a final ratio of 2.5:1 and diluted with buffer containing 200 mM KCl, 20 mM Hepes pH 8.0, 1 mM DTT and 1 mM EDTA. Crystals were grown using the hanging drop vapor diffusion method at room temperature in 2 µl drops containing a 1:1 ratio of reservoir solution (0.2 M Sodium acetate, 0.1 M Tris HCl pH 8.0, 15.5% Peg 4000 for R162A, Y288A, F243A and F322A, 0.1 M Bis-tris pH 5.0, 0.2 M NaK tartrate, 18% PEG3350 for G3244A, 0.03 M Citric Acid 0.07 M Bis-Tris pH 4.7, 17% PEG 3350 for G3249A, 0.05 M KPO_4, 16% PEG 3350 for A3243G and A3243T) to the protein:DNA mixture. Crystals formed overnight. Crystals were cryoprotected using 25-30% ethylene glycol and flash frozen in liquid nitrogen.

Data collection, processing and refinement

Diffraction data was collected using the x25 and x29 beamlines at the National Synchrotron Light Source (NSLS, Upton, NY). Data were then processed using XDS (117) and Scala, carried out within the autoPROC toolbox (118). Molecular replacement (MR) and refinement was then performed using MOLREP (119) and

REFMAC5 (120,121) in the ccp4 software suite (122) and manual model building was done using COOT (123). Figures were made using pymol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

MD simulations

System Preparation

We prepared chemical topologies and coordinates for the WT MTERF1-DNA complex. Coordinates of the MTERF1-DNA complex used those deposited in the PDB, 3MVA (78). N-terminal Glu and Asp residue sidechains were absent in the model. MOLPROBITY (124) was used to analyze all rotamers. His98 and His227 were assigned ε-H, neutral state, informed by preliminary simulations. Selecting the A conformation of Ser292, the 3499 solute and waters were built using standard procedures in AMBER (125).We used a truncated octahedron solvent container with a minimum 10Å buffer from any solute atom to the surface of container. This fully solvated system was 61341 atoms and its dimensions were used to calculate how many ions to add to mimic the in vitro 0.2 M concentration; neutralizing K^+ ions were also added. Ion placements were randomized.

Equilibration and production MD

We equilibrated the system: (1) Minimize the model with complete force evaluations, periodic boundary conditions, and non-bonded interaction cutoff of 8.0 Å, for 10000 0.01 kcal/mol Å steepest descent steps; atomic coordinates from the crystallographic model, excluding crystallographic waters, were restrained positionally with a force of 100.0 kcal/mol $Å^2$. (2) Ramp the temperature from 100.0 K to 300.0 K smoothly over 200 ps, with SHAKE (126), Berendsen thermostat, (127) positional restraint force of

100.0 kcal/mol Å^2 to the minimized structure, constant volume with bath coupling constants of 0.1 ps^{-1}, and 1 fs timestep while initializing velocities. (3) Resume the second stage velocities, at 300.0 K, constant pressure conditions with 0.5 ps^{-1} bath coupling constants, and a 100.0 kcal/mol Å^2 positional restraints to the final structure from (2), for 100 ps. (4) Resume conditions from (3) but with weaker positional restraint force of 10.0 kcal/mol Å^2 to final structure from (3), for 250 ps. (5) Change to restraining only MTERF1 backbone (CA, N, C) or DNA backbone (O5', C5', C4', C3', O3', O2P, O1P, P, O4', C2', C1') atoms to (4), with a force of 10 kcal/mol Å^2, otherwise identical to the initial minimization. (6) Use identical conditions as (4), with 10 kcal/mol Å^2 restraints (5), for 100 ps. (7) Resumed velocities and conditions, with 1 kcal/mol Å^2 restraints (6), for 100 ps. (8) Resume from (7) restraining to it with 0.1 kcal/mol Å^2, for 100 ps. (9) Resume for 250 ps free of restraints. Production dynamics resume (9), except the integration timestep was 2 fs, for 1 ns. Equilibration and production calculations utilized the single precision-double precision PMEMD implementation of SANDER, release 14 by Le Grand et al. (128) except for the initial minimization, which utilized pmemd-CPU. Four independent trajectories were generated by utilizing different random seeds for velocity initialization in stage (2) during equilibration.

Energy decomposition

For each of the four independent 1 ns production trajectories only the 6718 solute atomic coordinates were used. We used our in-house energy decomposition program (equivalent to AMBER idecomp=4) (129) to calculate the pairwise per-residue non-bonded interaction energy van der Waals plus electrostatics. For each of

the four residues R162, F243, Y288, and F322, we calculated the interaction energy of all the atoms within the residue to all the atoms in each of the nucleotides, leading to 44 residue-nucleotide energies for each of the 1 ns simulations. Each residue-nucleotide interaction was averaged over all four 500 MD frame trajectories. If the magnitude of these average energies was less than 1 kcal/mol, the value was automatically considered below the chemical limit and set to zero. Standard deviations were calculated to estimate data precision.

Identification of MTERF1 Allelic Variants

A search of the dbSNP (short genetic variations) database through NCBI (National Center for bioinformatics information) for human MTERF1 reveals ~121 SNPs at the time of this writing. We made mutant constructs for 6 missense mutations listed in Table 5.1. The reference SNP cluster report for each mutation reveals the frequency of the variant. All the variants, except R202H, have been validated in larger population cohorts such as the 1000 Genomes or Hap Map projects. The cohort population size is 4532 chromosomes.

Chapter 3 - The unique Binding Mode of MTERF1

The original wtMTERF1 crystal structure, ITC experiments and termination assays for the WT, RFY and Arginine substitution proteins were performed in collaboration with other members of the Garcia-Diaz lab and previously published by Yakubovskaya E, Mejia E, Byrnes J, Hambardjieva E, Garcia-Diaz M, published in *Cell 141,* 982-993, June 11, 2010. The MD simulations were conducted in collaboration with Kevin Hauser and the Carlos Simmerling laboratory.

As discussed in Chapter 1, the MTERF1 crystal structure bound to its recognition sequence was solved in our lab to 2.2Å. The structure reveals a modular architecture that is composed of 8 all α-helical mterf motifs. The MTERF1 fold is consistent with that observed for other MTERF proteins (see Figure 1.7) The superhelical MTERF1 fold is suited for binding dsDNA as the structure reveals positively charged grooves that interact nonspecifically with the negatively charged DNA backbone. However, the crystal structure demonstrates that MTERF1 can also form sequence specific contacts. As a result of specific binding, MTERF1 is able to elicit unwinding of the DNA duplex and the eversion of three nucleotides. The everted bases are stabilized by MTERF1 through π stacking interactions with three residues. This base-flipping event is important for stabilizing MTERF1 at its canonical binding site within the $tRNA^{Leu}$ gene. Furthermore, base flipping has been demonstrated to be important for the ability of MTERF1 to terminate at this site. However, the mechanism by which the flipping (or eversion) of three nucleotides takes place is not well

understood. Given the rarity and importance of base flipping for MTERF1 function, we decided to further investigate this mechanism. In this chapter, I will demonstrate, through structural and biochemical experiments that the flipping of three nucleotides is a complex multi-step process, yet occurs in an ordered stepwise fashion. First, we will begin by analyzing the details of the MTERF1 crystal structure.

DNA bending, helix unwinding and base flipping

The crystal structure reveals that MTERF1 is able to bind as a monomer to the major groove of double stranded DNA and can induce a slight bend (25°) in the duplex (Figure 3.1). Interestingly, the protein-DNA complex reveals that MTERF1 maintains a large footprint on the binding sequence as it covers twenty of the twenty-two bases present in the crystallization substrate. Importantly, the ends of the DNA duplex are found to be in a B-DNA conformation. However, at the center of the duplex a non B-DNA conformation is observed along with partial duplex melting (Figure 3.2).

At the site of duplex melting, a rare base flipping mechanism is observed that involves the eversion of three nucleotides: T3243, C3242 and A3243 (Figure 3.3). The flipped bases form π-stacking interactions with Arg162, Phe243 and Tyr288 of the protein (Figure 3.3). To assess the importance of the stacking interactions for base flipping, we have solved the crystal structure of a triple RFY to alanine (R162A, F243A, Y288A) substitution that eliminates the π-stacking interactions. This structure has been solved to 2.8Å and reveals that the overall structure of MTERF1 remains the same. However the base flipped nucleotides are in altered conformations (Figure 3.4). Interestingly, the conformations of A3243 and T3243 are intrahelical and no longer stabilized outside the DNA duplex. In addition, C3242 is still flipped out of the helix,

but in a completely different conformation and occupies the space where the Arg162 side chain was located. The RFY substitution structure stresses the importance of stacking interactions for stabilizing each nucleotide in the flipped conformation and that the absence of proper base flipping does not affect the overall MTERF1 fold. However, the mechanism through which the bases arrive to their final wild type conformation cannot be well understood based on the RFY substitution structure alone.

MTERF1 maintains non-specific contacts with DNA

MTERF1 forms a large number of interactions with the DNA backbone. The majority of interactions are electrostatic and nonspecific between the positively charged grooves in MTERF1 and the negatively charged phosphate backbone of the DNA (Figure 3.5). Interestingly, MTERF1 preferentially forms more electrostatic contacts with the LS than with the HS (78,130) . Since the majority of interactions with the DNA are nonspecific and electrostatic, it is possible that MTERF1 could form interactions with any DNA sequence. To test this, we performed isothermal calorimetry studies (ITC) to measure the affinity of MTERF1 for the wild type sequence and for an arbitrary sequence. The ITC experiments demonstrated that MTERF1 binds to wtDNA with 1:1 (DNA: protein) stoichiometry and a binding constant (K_d) of 1 μM. MTERF1 has a reduced affinity and lower DNA to protein ratio of binding in the presence of the arbitrary sequence (Table 3.1). This information suggests that in the absence of sequence specificity, as demonstrated by using the arbitrary sequence, binding of MTERF1 can still occur. However, the lower stoichiometry demonstrates that multiple MTERF1 molecules can associate with a single unspecific DNA sequence. In addition, the reduced stoichiometry also indicates

that MTERF1 does not associate with the DNA in any preferential conformation, further suggesting that MTERF1 can bind to multiple regions of the nonspecific DNA.

Base flipping is important for MTERF1 stability.

In the presence of the wild type binding sequence, MTERF1 is able to elicit DNA bending, helix unwinding and base flipping stabilized by the aforementioned stacking interactions. However, since we have demonstrated that MTERF1 has an alternative binding mode in the presence of a nonspecific DNA sequence. Thus we wanted to investigate if the base-flipping event is adopted upon stable specific binding. To further investigate this, we performed ITC experiments with the RFY triple substitution in the presence of wild type DNA and the arbitrary sequence previously mentioned (Table 3.1). Not surprisingly, the affinity of the RFY substitution for the WT DNA sequence is reduced, indicating that base flipping is indeed critical for specific stable binding. Moreover, the affinity loss is even greater in the presence of the arbitrary sequence. Interestingly, the RFY substitution maintains very similar stoichiometric ratios for both the WT and arbitrary sequences as wtMTERF1, indicating that the binding modes are similar. Again, this suggests that MTERF1 and the RFY substitution are able to adopt alternate conformations in the presence of an unspecific sequence. In addition, the crystal structure of the RFY substitution confirms that the binding mode is identical to wtMTERF1 except that the flipped bases are in altered conformations. This strongly suggests that base flipping is a consequence of sequence specific recognition and increases affinity for the binding site.

The binding mechanism of MTERF1 involves 5 specific contacts

Although the RFY substitution reveals a decreased affinity for the WT sequence, the stoichiometry is very similar to that of wtMTERF1. In addition, the crystal structure of the RFY triple substitution reveals that the binding mode is largely the same as wild type, suggesting that the role of base flipping has little to do with sequence recognition. Moreover, most interactions are electrostatic, established with the major groove of the DNA backbone and therefore cannot distinguish between different DNA bases. However, among the many protein-DNA contacts made, there are five conserved arginine residues that make specific contacts with bases in the WT DNA (Figure 3.6). This suggests a mechanism by which MTERF1 is able to exhibit sequence specificity. The sequence specific contacts are mediated by hydrogen bonds established between arginine side chains and guanine bases. As shown in Figure 3.6 three of the 5-arginine residues (Arg169, Arg251 and Arg387) make exclusive contacts with bases of the LS and one (Arg350) makes an exclusive contact with the HS. Interestingly, Arg202 mediates contacts between guanine bases on both strands while Arg387 and Arg251, in addition to their interactions with guanine, make contacts with the adjacent thymine and adenine bases respectively.

In order to investigate the importance of each arginine residue in the mechanism of MTERF1 sequence recognition, we constructed five, individual, arginine to alanine substitutions and using ITC determined the effect on MTERF1 stoichiometry and binding affinity for the WT DNA sequence. The results, shown in Table 3.1 show that with the exception of R350A, binding affinity is largely lost. Interestingly, with the exception of R387A, the substitutions exhibit a similar stoichiometry to that of wtMTERF1. This suggests that Arg387 is more critical for

establishing the proper bind mode and that the other arginine residues play a role in maintaining affinity for the sequence. Thus the contributions of each arginine to binding and sequence recognition are not equal. However, at this point, we lack a clear understating of how these residues affect MTERF1 termination activity.

Sequence recognition and base flipping are important for MTERF1 termination activity.

To assess termination activity mediated by MTERF1, a reconstituted *in vitro* transcription termination assay was performed (83). Briefly, the assays (described in detail in chapter 2) utilize a linear template containing the HSP and the 22 bp MTERF1 binding sequence inserted 100bp downstream. The DNA template is incubated with MTERF1 and then the transcription reaction is initiated upon addition of the transcription machinery (POLRMT, TFAM and TFB2). After the reaction is stopped, the transcripts are run on a gel, resulting in two bands corresponding to runoff transcripts that are full length (FL) and a distinct termination band (T) as a result of MTERF1 activity (Figure 3.7). Interestingly, others have shown that MTERF1 terminates more strongly from LSP than HSP. To simulate this, we inserted the termination sequence in reverse orientation with respect to HSP and observe much stronger termination, confirming previous results that suggest MTERF1 exhibits termination polarity. As an interesting aside, although the reason for polar termination is not entirely clear, the mechanism for this observation may involve the greater amount of contacts established between MTERF1 and the LS compared to the HS.

After establishing that purified MTERF1 is able to mediate termination, we then tested the roles of base flipping and the five-arginine residues for MTERF1

function. The results demonstrated that the RFY triple substitution is unable to mediate any termination activity. Similarly, all the arginine to alanine substitutions were unable to terminate transcription, with only residual activity found at high concentrations of the R350A substitution (Figure 3.8). These data confirm the importance of sequence recognition, the binding mode and base flipping for the ability of MTERF1 to mediate termination.

Model of MTERF1 termination

The structural, thermodynamic and biochemical assays performed with MTERF1 and the various substitution constructs have revealed mechanistic information about the protein-DNA binding mode. From these observations we have proposed a model by which MTERF1 is able to bind and stabilize itself to DNA. Since we have shown that MTERF1 is capable of interacting with DNA in an unspecific fashion, we suggest that MTERF1 initially samples arbitrary DNA sequences and adopts an alternative binding mode at these sites. Secondly, we have demonstrated that the arginine-guanine interactions are able to identify the sequence specific site and are critical for termination activity. Furthermore, the contribution of each arginine residue varies, whereas some contribute to establishing the correct binding mode and others increase affinity for the binding site. Lastly, base flipping is suggested to exclusively occur upon recognition of the specific binding sequence and serves to increase the affinity for the specific DNA sequence. Thus, we suggest that the final base-flipping step is important to anchor MTERF1 to the DNA, which serves as a roadblock to prevent transcription from continuing past the site.

The eversion of three nucleotides is a rare and complex event

As demonstrated above, the eversion of three bases from the DNA duplex is important for increasing the affinity of MTERF1 to the specific DNA sequence. However, the mechanism through which this unique mode of base flipping occurs is not well understood (80). Since there are three nucleotides involved, that would suggest that the flipping mechanism is a complex and highly ordered process. Therefore, we sought to understand the mechanical steps involved in nucleotide eversion. In order to investigate this, we decided to break down the process of base flipping and created individual substitutions that disrupt the stacking interactions one at a time. We have evaluated the structural and functional consequences of these substitutions and show that individual stacking interactions do not contribute equally to function. We have identified additional residues important for base flipping and propose a model for the base-flipping process. Our data also demonstrate the existence of alternate base flipped conformations, some of which are still compatible with termination activity. This could have implications for the ability of MTERF1 to perform specific functional roles at other sites in the mitochondrial genome.

Functional differences among Y288A, R162A and F243A substitutions.

As previously described, eliminating the three stacking interactions by an RFY to AAA triple substitution results in a severe functional loss (Figure 3.7) (78). However this does not provide information on the specific role of base flipping or the mechanism of eversion. To further investigate base-flipping in MTERF1 we created, expressed and purified R162A, Y288A and F243A individual substitutions, and performed termination assays to assess their function on a model substrate as previously reported (78). The resulting UREA –PAGE gels for each substitution is

shown in Figure 3.9. Surprisingly, the termination defect for each mutant is not equal. Whereas both R162A and F243A have a severe functional defect, the Y288A substitution results in a protein that is deficient at low concentrations, but closer to wild type at higher concentrations. Interestingly, R162A displays the strongest effect, with only residual termination activity. Thus, unexpectedly, the termination assays demonstrate that the functional contribution of each stacking residue is clearly not equal with respect to function.

Energy decomposition explains the importance of R162

Analysis of the MTERF1 crystal structure does not immediately explain the reason for the inequality in the functional differences amongst Arg162, Phe243 and Tyr288 substitutions. We reasoned that these differences could be explained by the presence of additional interactions between Arg162, Phe243, Tyr288 and the DNA backbone. To test this hypothesis and evaluate the strength of these interactions, we conducted MD simulations with wild type MTERF1 bound to wild type DNA.

Consistent with our expectations based on the crystal structure, the analysis of our MD simulations identified interactions expected to be favorable between Tyr288, Phe243, and Arg162, with the everted nucleotides (Figure 3.10). Thus, we analyzed the remaining interactions to better understand how Tyr288, Phe243, and Arg162 support the everted bases and stabilize the altered backbone conformation observed in the crystal structure.

Our analysis identifies a weaker interaction not obvious in the crystal structure between Phe243 and T3243 (Figure 3.10 B). Thus, together with the Arg162-T3243 interaction, this interaction might compensate partially for duplex destacking.

Interestingly, this analysis also reveals the importance of the nonspecific ionic interactions established between Arg162 and the DNA backbone (Figure 3.10 D). In addition to cation π- stacking with T3243, Arg162 forms an ionic interaction between its guanidinium group and the phosphate group of T3241 (2.6Å; Figure 3.11). Thus, in addition to stacking with one of the everted residues, like Phe243 or Tyr288, Arg162 also interacts with the DNA through electrostatic interactions, providing a potential explanation for the severity of the phenotype in the R162A substitution. Because they are nonspecific, these interactions might be important to define the conformation of the DNA backbone during the initial stages of DNA binding. Furthermore, the Arg162-T3241 interaction might be key to stabilizing the severe kink in the HS backbone that takes place in the base-flipping region (arrow in Figure 3.11). This kink is essential to facilitate the conformation observed in the crystal structure.

X-ray crystal structures of R162A, Y288A and F243A substitutions.

While explaining the importance of the Arg162 side chain, the MD analysis does very little to explain the drastic functional differences observed between Phe243 and Tyr288. In fact, the energetic contribution of the interactions established by both residues in the final state of base flipping appears to be very similar. This suggests that the nature of the difference might result from differential contributions of these residues to the base-flipping process. Indeed, base-flipping of three residues, combined with duplex unwinding, is likely to be a complex, multi-step process, and the importance of specific residues during the base-flipping mechanism is yet unknown.

We decided to crystallize each individual substitution in order to assess any conformational changes that would provide insight into the base flipping mechanism.

Each structure was solved using the wild type (PDB: 3MVA) structure as a search model for molecular replacement. The structures were solved to a resolution of 2.6Å, 2.6Å and 2.5Å for R162A, Y288A and F243A, respectively. Crystallographic and refinement statistics are shown in Table 3.2. The overlay of R162A (yellow) on wild type (gray) shows an altered conformation of the HS in the R162A substitution. As expected (since Arg162 stacks with T3243 in the wild-type crystal structure) the conformation of T3243 is altered and this residue is now located within the double-helix (Figure 3.12 A). This residue now interacts with Arg195 through a hydrogen bond and Phe243 through overlapping pi orbitals in a t-stacking conformation (Figure 3.13). Surprisingly, C3242 also swings out of the helix and occupies the space where the arginine side chain would normally be located (gradient arrows, Figure 3.12 A). Interestingly, the conformation of this residue is identical to that seen in the triple RFY mutant that is devoid of termination activity. Furthermore, we observe that A3243 in the light strand (LS) is stacked with Tyr288 in the same fashion as observed in the wild type structure (Figure 3.12 A).

By contrast, the loss of the stacking interaction with A3243 in the Y288A structure results in an altered conformation of the A3243 on the LS (Figure 3.12 B). The overlay of Y288A (cyan) and wild type (gray) shows that the A3243 is only partially flipped out from the helix. Moreover, the conformation of the HS in the Y288A structure is also altered, but not as drastically compared with what was observed in R162A. Whereas the conformations of Arg162 and C3242 are similar to those observed in the wild type structure, T3243 is located within the helix as in the R162A substitution (Figure 3.12 B, gradient arrow). The conformation of T3243 is

very similar to that observed in the R162A substitution, forming a t-stacking interaction with Phe243 (Figure 3.13 A). However, the hydrogen bond with Arg195 is not observed, but instead T3243 interacts with its neighboring nucleotide through a hydrogen bond between O2 of T3243 and N4 of C3244 (3.4Å; Figure 3.13 B). Interestingly, the perturbation of the HS conformation resulting from a substitution that only directly affects the LS implies a coupling between the conformations of the LS and the HS.

Not unexpectedly given the severe termination defect, the structure of F243A reveals multiple alterations of both the LS and HS (Figure 3.12 C). Phe243 forms a stacking interaction with C3242 in the wild-type structure and importantly, this residue is the only one of the three everted residues that is not fully extrahelical. As could be expected, in the F243A structure, the conformation of C3242 is severely altered. It is now in a position where it is still forming Watson-Crick hydrogen bonds with its G pair (Figure 3.12 D). Interestingly, the conformation of T3243 is very close to that in the wild type structure, stabilized by stacking with Arg162. However, further stressing the coupling between HS and LS conformations, A3243 is located in a helical position.

Overall, these structures reveal that the conformation of C3242 closely correlates with termination activity, suggesting that the conformation of this base is critical to stabilize the transcriptionally active conformation of the protein. The structures also suggest that some of the aberrant conformations adopted by MTERF1 can be stabilized by compensatory interactions, which presumably stabilize the final state sufficiently to enable close to wild-type termination activity in the case of Y288A. One of the most striking observations that result from the structures corresponds to the

conformation of C3242 in the F243A structure. As stated above, C3242 in this structure is still base-paired with its partner G (although the base pair is clearly perturbed; Figure 3.12 D). This structure therefore reveals a conformation that represents an intermediate in the base-flipping mechanism, and suggests a pathway for the movement of C3242 from its G-C base pair to the everted conformation observed in the wild type structure. An overlay between the F243A and wild-type structures suggests that this transition is mediated by a conformational change of the side chain of Asp283 (Figure 3.12 E). This residue interacts with the N4 of C3242 and appears to facilitate breaking the G/C pair and follow the movement of C3242 from its original position to its final position where it stacks with Phe243 and hydrogen-bonds to Glu280 (Figure 3.12 E).

Phe322 and Phe243 stabilize base-flipping intermediates

As mentioned above, the F243A structure represents an intermediate in the base-flipping mechanism where A3243 is still located in a helical conformation. Importantly, A3243 in this conformation is stabilized by a t-stacking interaction with a phenylalanine residue, Phe322 (Figure 3.12 F). This residue is not involved in any interactions with the DNA of the final state captured in the wild-type crystal structure and this is confirmed through MD simulations (Figure 3.14). This suggests that Phe322 might be important to stabilize intermediate steps during base flipping.

Interestingly, in the R162A and Y288A structures, T3243 is located within the helix, adopting conformations that are likely intermediate in the base-flipping process. Observing the environment of T3243 in those structures reveals a striking parallelism with the Phe322-A3243 interaction seen in the F322A structure: Phe243 forms an

analogous t-stacking interaction with T3243 (Figure 3.13 A). Similar to what is observed with Phe322, no interactions are observed between Phe243 and T3243 in the wild-type crystal structure.

This suggests that both Phe243 and Phe322 play an important role in the base-flipping process, helping stabilize base-flipping intermediates before the bases are flipped-out of the helix. Interestingly, an overlay between the Y288A and F243A structures reveals a hypothetical intermediate during base flipping where both T3243 and A3243 are still within the helix (Figure 3.15 A). This suggests a model where Phe243 and Phe322 intercalate between adjacent bases in the HS and LS, respectively, and stabilize their intrahelical conformations. Importantly, given the position of both phenylalanines, Phe322 and Phe243 appear to be important to help break the A3243/T3243 base pair (dashed arrow, Figure 3.15 A).

To test this hypothesis, we constructed an F322A substitution and tested its biochemical effect on MTERF1 function using a termination assay as above. As can be seen (Figure 3.15 B), F322A results in a large functional defect, almost of a similar magnitude to that observed with an F243A substitution. Given the analogous role that appears to be played by both residues, we then constructed a double F243A/F322A substitution. The resulting gels show that both substitutions have additive effects, as the double mutant results in a protein severely impaired in its ability to mediate termination activity (Figure 3.15 B).

Phe322 and Phe243 play an important role in coordinating the base-flipping mechanism.

To further investigate the importance of Phe322 for the base-flipping process, we decided to determine the structure of the F322A substitution. The structure was solved to 2.6Å resolution (Table 3.2). Interestingly, despite the fact that Phe322 stabilizes LS intermediate conformations, the F322A structure (Figure 3.15 C) reveals that the substitution results in a perturbed final state in which the LS is in a wild-type conformation, but where C3242 is located in an aberrant conformation outside the helix. Since Phe322 does not play any role stabilizing the final conformation, this result further supports that the function of this residue is important during the base-flipping process. Moreover, the fact that the F322A substitution results in alterations in the HS conformation implies that intermediate LS conformations affect the base-flipping process in the HS, further supporting that the mechanisms of base-flipping in the HS and LS are coupled.

Base flipping is an ordered stepwise process.

In this study we have biochemically and structurally characterized the importance of several residues in the mechanism of base flipping by MTERF1. Consistent with the fact that sequence recognition is mostly mediated by major groove interactions, base flipping is not strictly required for DNA binding by MTERF1. However, it is important for its transcriptional termination activity, presumably because it allows MTERF1 to anchor itself to the termination sequence, preventing the transcription machinery from moving past the termination site. The fact that this mechanism involves the eversion of three nucleotides implies that the base-flipping

mechanism is likely complex. We have discovered that residues Tyr288, Arg162 and Phe243 that interact with each of the everted nucleotides are all important to stabilize the final state of base flipping. Eliminating these side chains results in alterations to the final conformation not observed in the wild-type crystal structure, highlighting the importance of pi-stacking interactions to stabilize the flipped-out bases. However, our results also show that the three side chains are not equally critical for function. Importantly, in the absence of some of the key interactions observed in the wild-type conformation, MTERF1 is able to adopt alternative binding modes that while not to the same extent as the wild-type protein, still maintain an appreciable amount of activity. This suggests that the MTERF1 binding mode is sufficiently flexible to support binding and at least partial termination activity in different sequence contexts. This might have implications for the alternative MTERF1 binding sites that have been characterized in the mitochondrial genome (91), where the full consensus sequence required for the binding mode observed in the wild-type structure is not present.

In addition, our crystal structures reveal altered conformations that are representative of intermediate base flipping states, and thus provide information on the MTERF1 base-flipping mechanism. The structures reveal interactions not present in the final state of base flipping, and have allowed us to identify additional residues important for base flipping. Furthermore, the structures provide evidence that a high degree of coordination exists between base flipping events in the heavy and light strands.

A key mechanism that appears to be at the core of this coordination is related to the role of two phenylalanine side chains, Phe243 and Phe322. In addition to the role of

Phe243 stabilizing C3243 in the final state, these residues help stabilize intermediate conformations during base flipping. As we have shown, alterations to this mechanism lead to perturbations in the conformation of the final state, and reveal alterations in the coupling between HS and LS conformations. Furthermore, the structures provide details on the nature of this mechanism. For instance, Phe243 appears to be able to stabilize the intrahelical conformation of T3243. Once T3243 is flipped out of the helix, Phe243 interacts exclusively with C3242. However, Phe243 is able to simultaneously interact with both residues, as shown in the Y288A structure (Figure 3.13 B), implying that base flipping in the HS follows a stepwise mechanism. Thus, C3242 is first delivered to its final position through a network of interactions involving Asp283 (Figure 3.12 D), while F243 stabilizes the intrahelical conformation of T3243. Similarly, Phe322 stabilizes the intrahelical conformation of A3243, preceding base flipping of this residue. Given the importance of Phe322 for events taking place in the HS, this suggests that this conformation must be transiently maintained in order to properly coordinate the steps in HS base flipping. Although the exact mechanisms by which this coordination happens is still unclear, our data stress the importance of a network of hydrogen bonds and sequential t-stacking and π-stacking interactions between Phe243, Phe322, T3243 and C3242.

We demonstrate that the absence of Phe322 leads to perturbed base-flipping suggesting that this residue is important for coordinating the base flipping steps and in conjunction with Phe243 seem to be important in separating the A3243/T3243 base pair.

Based on our observations we are able to propose an ordered step-wise model for the base flipping mechanism. Helped by the helix unwinding that occurs upon MTERF1 binding to its target sequence, the first crucial step is the intercalation of Phe322 and Phe243, which stabilize the intrahelical conformations of A3243 and T3243 and presumably facilitate breaking this base pair. The formation of a hydrogen bond network between Asp283 and C3242 might help break the 3242 base pair and is important for delivering the C to its final conformation. Since the final conformation of C3242 is incompatible with the intrahelical conformation of A3242, it is likely that the motion of C3242 towards Phe243, is coupled with A3243 flipping. Lastly, the interaction between Phe243 and C3242 is likely coupled to the movement of T3243 outside the helix (Figure 3.16).

Our structural observations recapitulate our step-wise mode for the mechanism of MTERF1 base flipping. For example, the coupling of the C3243 motion to A3243 flipping is consistent with the conformations observed in the F243A structure, where in the absence of the motion of C3242 towards Phe243, A3243 is located within the helix and stabilized by Phe322. Furthermore, any alteration to this mechanism results in an altered final conformation. For instance, decoupling C3242 from flipping of A3243, like in the F322A, or premature base flipping of C3242 because of the absence of the Arg162 side chain results in an aberrant conformation of C3242 that prevents T3243 from flipping out of the helix. Moreover, incomplete base flipping of A3243 due to the absence of the Tyr288 side chain is coupled to T3243 base flipping through a yet unclear mechanism.

In summary, we have shown that the eversion of three nucleotides by MTERF1 is a complex multi-step process that is important for transcriptional termination. However, we have demonstrated that this process is relatively flexible, and that some conformations are consistent with a certain degree of termination activity. This might explain the residual termination activity observed in pathogenic mtDNA mutations like A3243T (78,131),expected to affect base-flipping. Importantly, base flipping involves not only residues that stabilize the final base flipped state but also residues responsible for establishing intermediate interactions that are important in coordinating this tightly coupled process but not evident from an MTERF1 crystal structure. Our work begins to elucidate the complex mechanism underlying the process of base flipping that takes place in MTERF1 and is essential for its transcriptional termination activity. Since the base flipping process might be more tolerant of sequence alterations than previously thought, it also has implications for the ability of MTERF1 to perform putative additional roles at alternative binding sites as has been previously suggested.

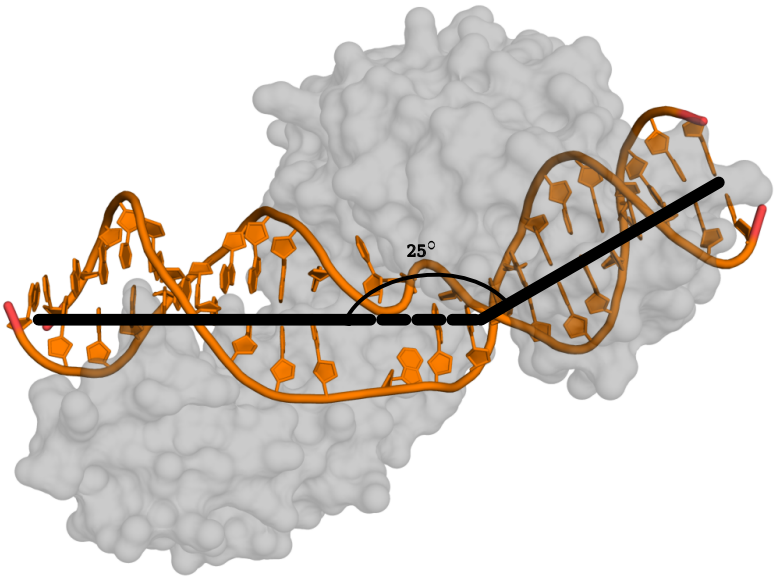

Figure 3.1 MTERF1 binding to its recognition sequence induces a 25° bend in the DNA duplex.

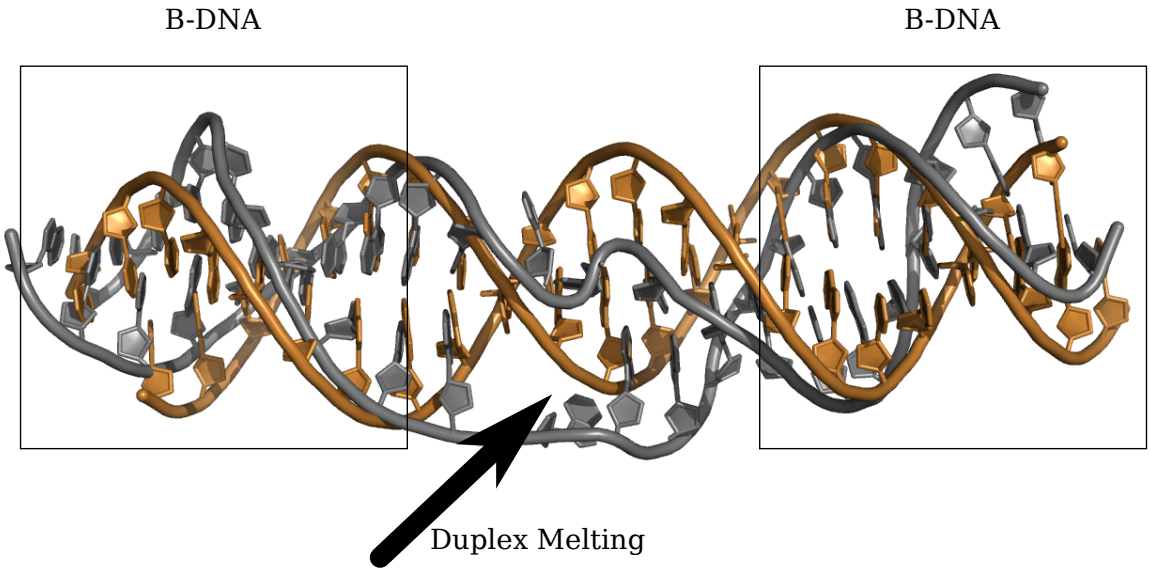

Figure 3.2 MTERF1 induces duplex melting at the center of the binding sequence. *The wild DNA binding sequence from the crystal structure (colored gray) shows duplex melting in the center and a B-DNA conformation at each end outlined with boxes. The copper colored DNA sequence is the same as the gray, however is entirely B-form DNA. Notice how the gray colored DNA is distorted at the center (black arrow).*

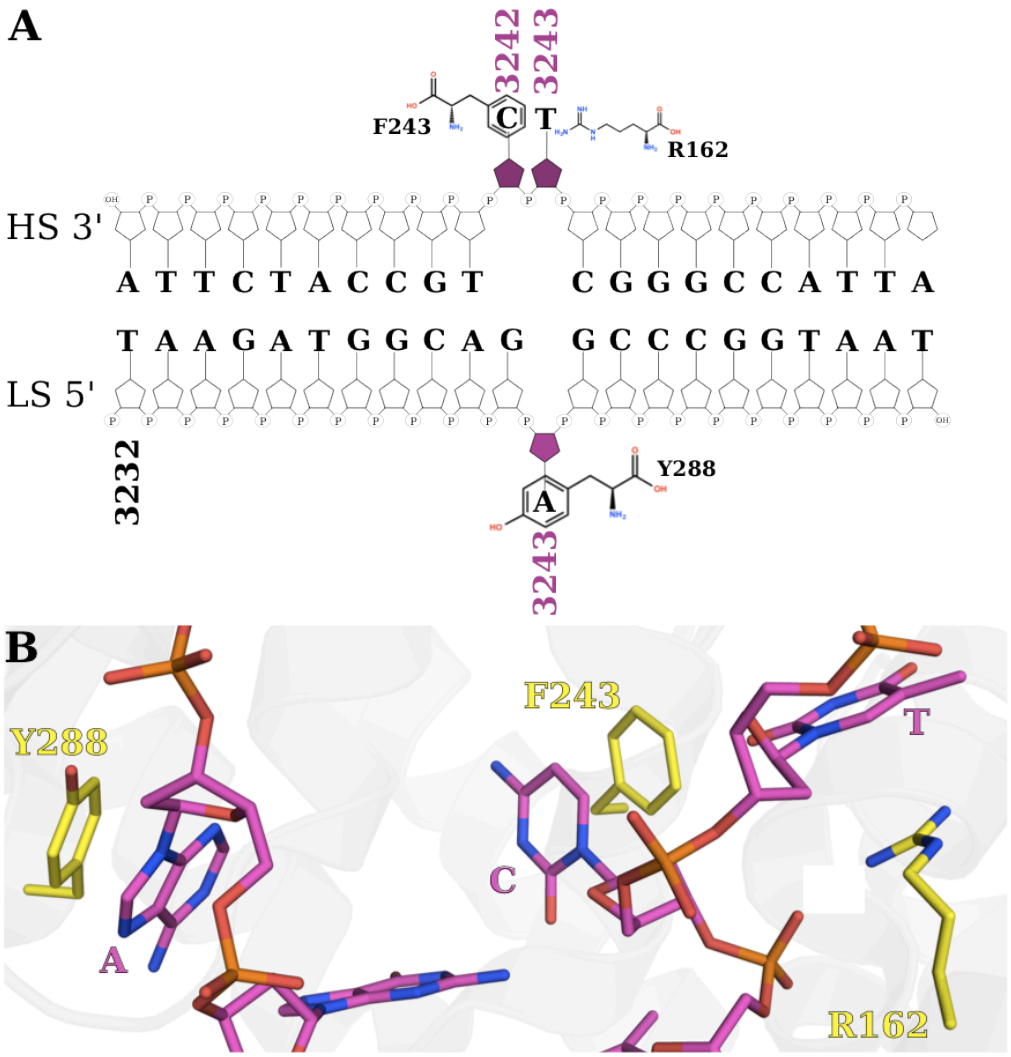

Figure 3.3 MTERF1 base flipping is stabilized by π stacking interactions. *(A) Schematic representation of the MTERF1 binding sequence. The three everted nucleotide A3243 on the LS, T3243 and C3242 on the HS are stabilized outside the DNA helix through stacking interactions with Y288, R162 and F243. (B) Crystal structure of wtMTERF1 (PDB:3MVA) shows the stacking interactions that stabilize the three flipped bases.*

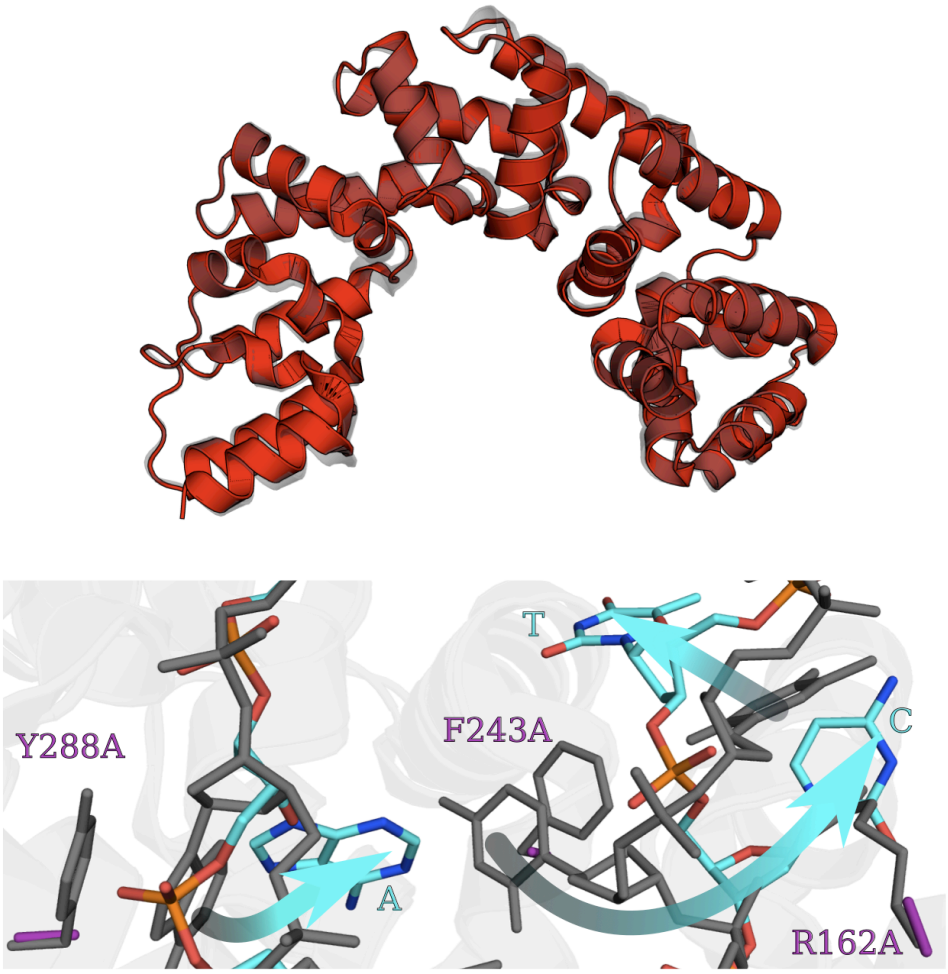

Figure 3.4 Overlay of wtMTERF1 and the RFY triple substitution. *The top panel shows an overlay of the RFY triple substitution solved to 2.8Å (colored red) with wtMTERF1 (transparent, colored grey). The overall fold of the RFY substitution is the same as wtMTERF1. The bottom panel is an overlay of wtMTERF1 (colored gray) with the RFY substitution (protein colored purple, DNA colored cyan). Gradient arrows represent the altered conformations of the flipped bases from their wild type positions to that observed in the RFY structure.*

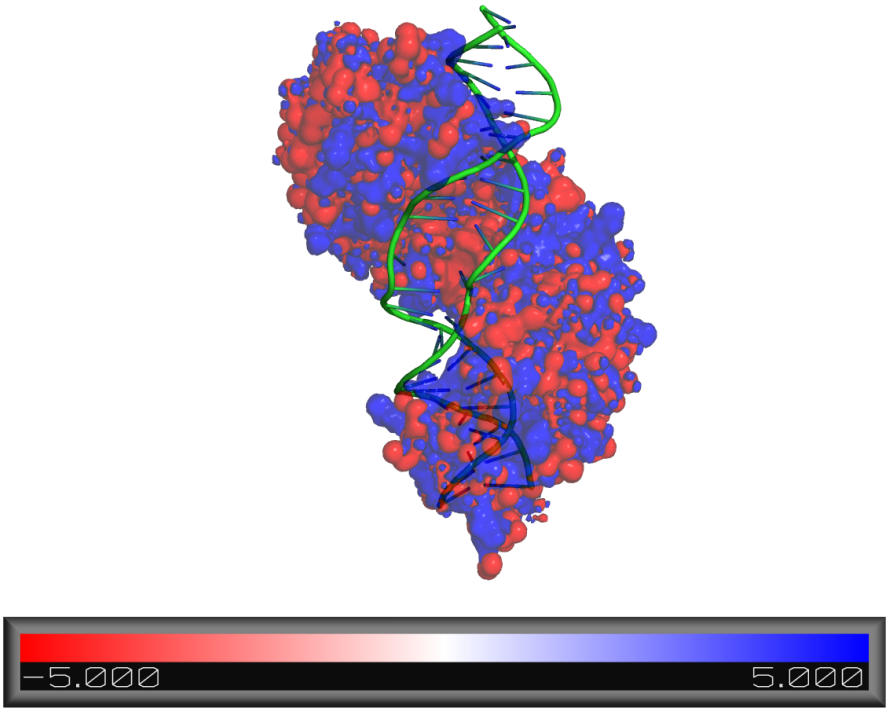

Figure 3.5 Electrostatic Surface potential of MTERF1. *The electrostatic potential surface map was calculated using the APBS Tools plugin with pymol. The protein surface is colored from -5kT/e (red, negatively charged) to +5kT/e (blue, positively charged). The negatively charged DNA backbone interacts nonspecifically with positively charged grooves located within the MTERF1 fold.*

WT 5' TAAGATGGCAGAGCCCGGTAAT 3'
ARBITRARY 5' TAAGATAATCACGCTTCATAAT 3'

Protein	Stoichiometry (N)	K_d (μM)
WT (WT sequence)	1.06 ± 0.08	1.05 ± 0.18
RFY (WT sequence)	0.80 ± 0.09	13.04 ± 1.31
WT (arbitrary sequence)	0.47 ± 0.20	12.03 ± 2.02
RFY (arbitrary sequence)	0.67 ± 0.04	21.18 ± 2.43
R169A	1.05 ± 0.15	24.75 ± 2.83
R202A	0.95 ± 0.14	11.52 ± 1.65
R251A	1.03 ± 0.01	3.90 ± 0.32
R350A	1.09 ± 0.01	2.23 ± 0.15
R387A	0.26 ± 0.12	25.64 ± 1.11

Table 3.1 Isothermal Calorimetry (ITC) experiments for WT, RFY and arginine substitution proteins. *The double stranded WT and arbitrary sequences (LS shown only) used in the ITC experiments are displayed in the top panel. The differences between the sequences are highlighted in red. The arbitrary sequence alters the bases involved in sequence recognition and base-flipping region. The results of the ITC experiments are shown in the table. ITC experiments with the arginine substitutions were tested with WT DNA sequence. The stoichiometry represents the ratio of DNA:protein.*

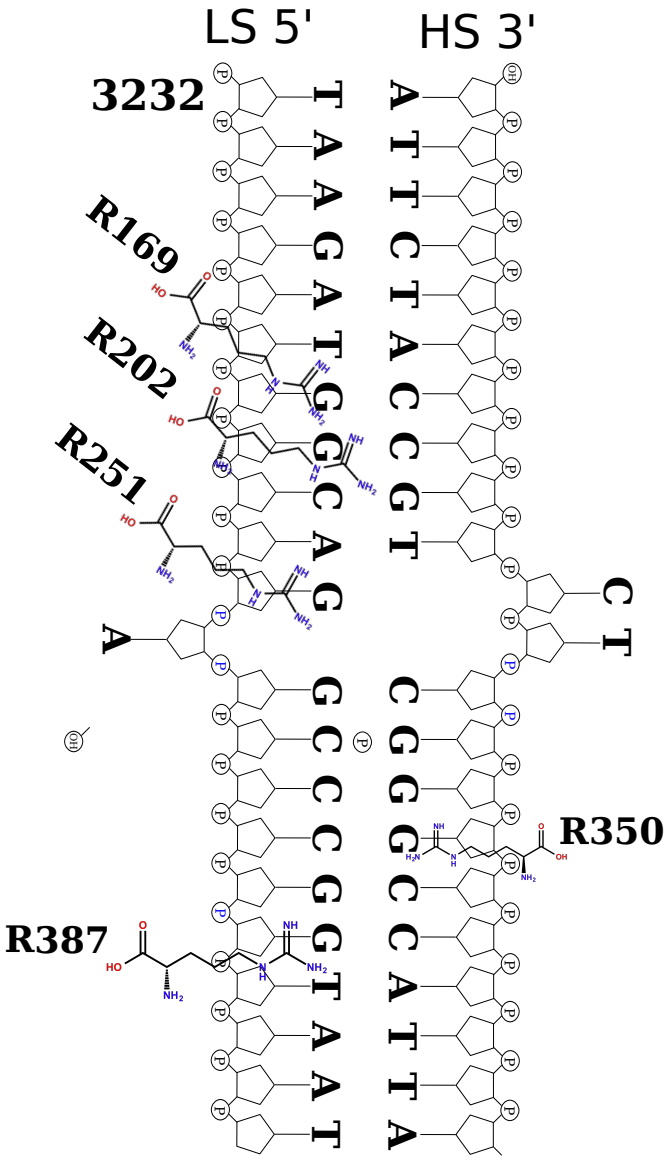

Figure 3.6 Location of the specific arginine-guanine contacts within the MTERF1 binding site. *The five arginine-guanine contacts are shown in the figure above. R202 hydrogen bonds with guanines on both strands, while R387 and R251 in addition to their interactions with guanines, also form hydrogen bonds with the adjacent nucleotides.*

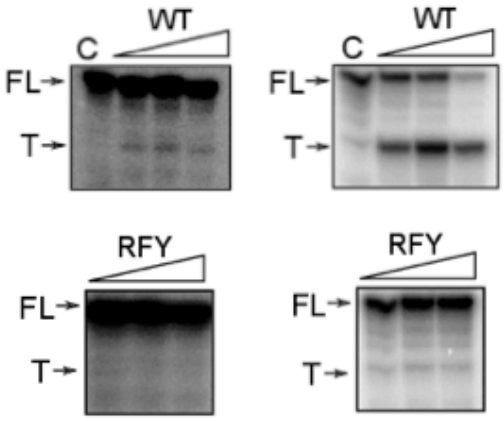

Figure 3.7 In vitro termination activity mediated by MTERF1. *Runoff transcription termination assays were conducted and result in two distinct bands of a 5% UREA-PAGE gel. The top band corresponds to full-length runoff transcripts (FL) and the lower band corresponds to the termination (T) transcript. The control lane C, does not contain MTERF1. Transcription termination was assessed with the termination sequence oriented in a forward direction with respect to HSP (2 gels in left panel) and with the sequence oriented in the reverse direction (2 gels in right panel). Stronger termination activity is observed with the termination sequence oriented in the reverse orientation as seen in the WT gel (top right panel). Thus, MTERF1 exhibits termination polarity. The ability of MTERF1 to terminate is lost in a triple RFY substitution that eliminates stacking interactions important for stabilizing the three everted nucleotides.*

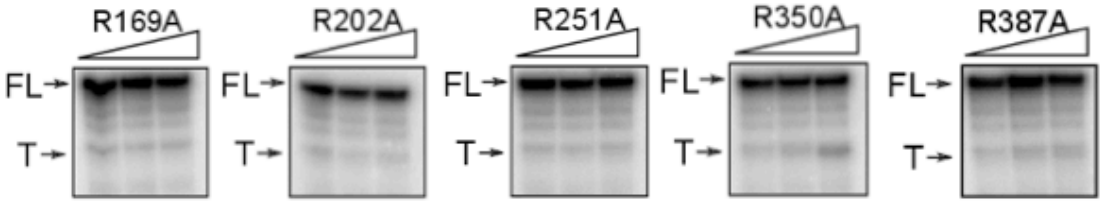

Figure 3.8 Termination activity for five arginine to alanine substitution constructs. *All assays were performed with the termination sequence oriented in reverse with respect to the HSP promoter. All constructs demonstrate severely impaired MTERF1 termination activity.*

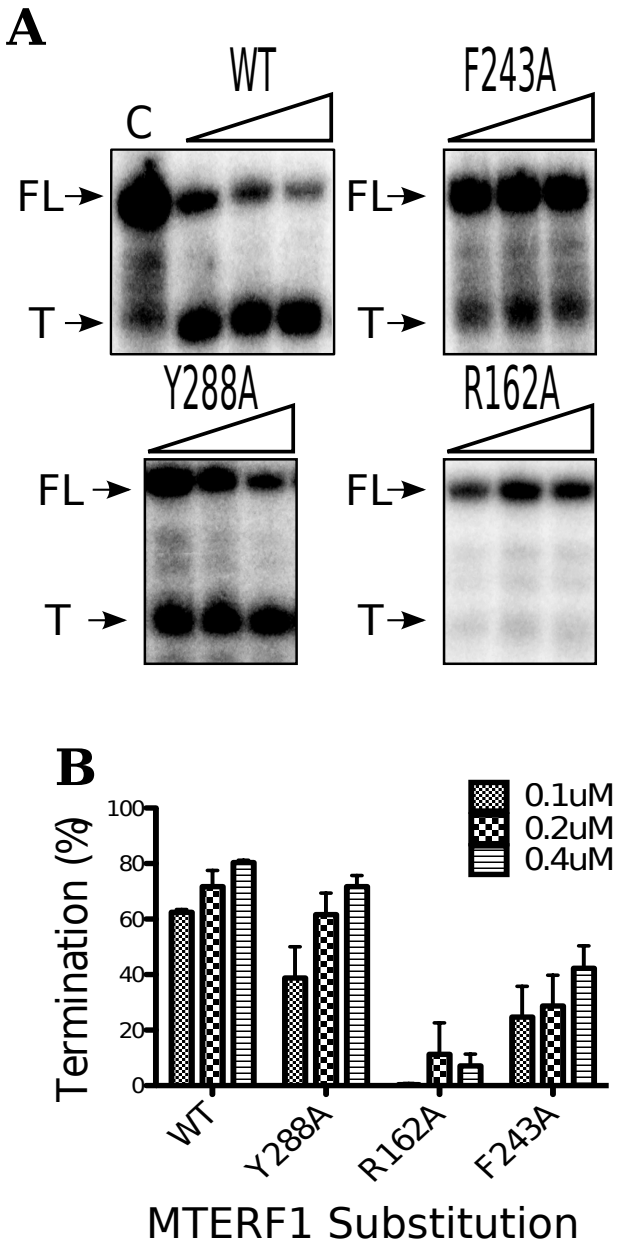

Figure 3.9 Termination assays for WT, F243A, Y288A and R162A substitutions. *Panel (A) shows the termination activity for each substitution at 3 concentrations of MTERF1 (0.1uM, 0.2uM and 0.4uM) as exemplified by the gradient triangle. Panel (B) shows the percentage termination for each substitution as determined by densitometry. Error bars represent the standard error of the mean (SEM) for at least 3 independent experiments.*

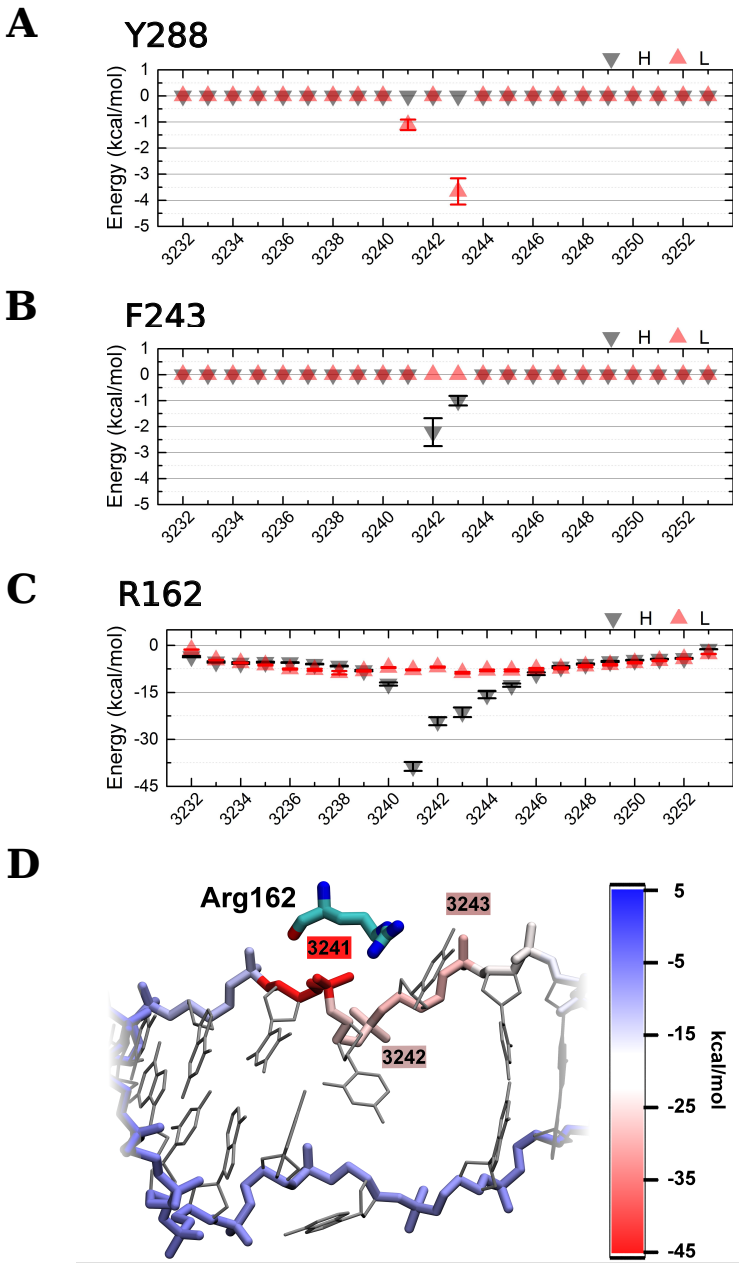

Figure 3.10 *Energy Decomposition diagrams showing the interaction energies between Tyr288 (A), Phe243 (B) or Arg162 (C) and the DNA in the wild type structure. Red triangles represent an interaction between the residue and the light strand (L) DNA or with heavy strand (H) DNA, represented by gray triangles. (D) Color gradient showing the strongest electrostatic interactions (red) of Arg162 with the DNA backbone.*

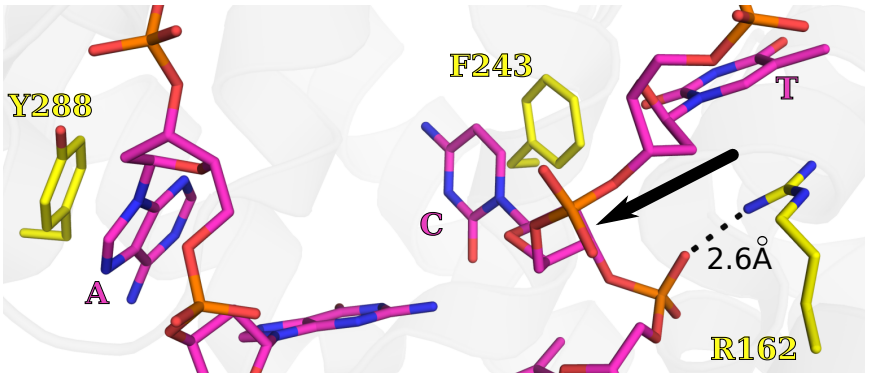

Figure 3.11 Arg162 stabilizes a kink in the DNA backbone at the site of base flipping. *At the site of base flipping in wild type MTERF1, Arg162 forms a hydrogen bond with the backbone phosphate of T3242. This interaction may be key to stabilizing the severe kink in the DNA backbone as a result of duplex melting and base flipping (black arrow). DNA backbone carbon atoms are colored magenta and the protein backbone carbon atoms are colored yellow.*

	R162A	Y288A	F243A	F322A
Data Collection				
PDB Code				
R_{merge}	0.066 (0.807)	0.054 (0.786)	0.037 (0.708)	0.042 (0.825)
R_{meas}	0.073 (0.873)	0.060 (0.840)	0.041 (0.773)	0.047 (0.89)
R_{pim}	0.037 (0.444)	0.031 (0.437)	0.021 (0.388)	0.024 (0.452)
No. of total observations	143906 (1384)	137707 (1338)	166530 (1592)	149721 (1562)
I/σI	25.4 (2.8)	27.3 (2.6)	34.1 (2.9)	29.1 (2.5)
$CC_{1/2}$	0.999 (0.823)	1.0 (0.793)	0.999 (0.851)	1.0 (0.787)
Completeness %	99.9 (100)	99.8 (100)	99.3 (100)	100 (100)
Multiplicity	7.3 (7.4)	7.2 (7.2)	7.3 (7.5)	7.2 (7.4)
Cell Dimensions (a,b,c)	88.22, 91.6, 159.10	87.53, 90.43, 169.50	89.25, 90.14, 161.43	89.15, 90.19, 160.85
α,β,γ	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Space Group	C2221	C2221	C2221	C2221
Resolution Å	2.62	2.59	2.48	2.59
Refinement				
R_{work}/R_{free}	0.21/0.26	0.21/0.27	0.24/0.27	0.21/0.28
R.m.s.d Bond Angles Å	1.698	1.588	1.692	1.580
R.m.s.d. Bond Lengths Å	0.012	0.012	0.013	0.011
Average B-factor	53.18	60.866	69.086	72.032

Table 3.2 X-ray crystallography Data collection and refinement statistics. *Data in parentheses is for highest resolution shell.*

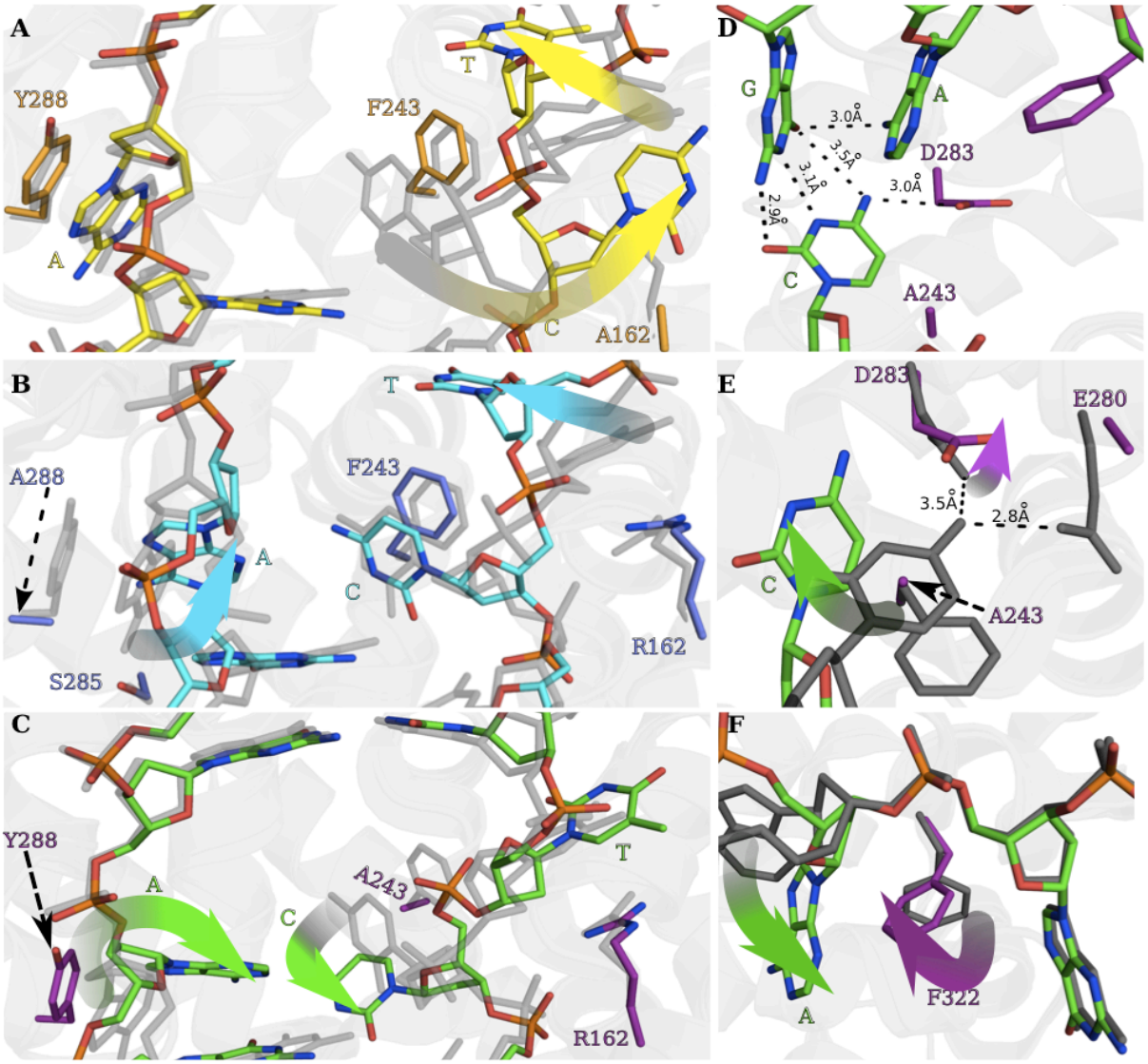

Figure 3.12 X ray crystal structures of the R162A, Y288A and F243A substitutions*. An overlay of wild type (gray) with (A) R162A (yellow and orange), (B) Y288A (cyan and blue) and (C) F243A (green and magenta). Alterations of the base-flipped nucleotides from their wild type conformations are highlighted with gradient arrows. (D) The F243A structure demonstrates the inability of C3243 to form a stacking interaction with Phe243 and reveals C3242 in a perturbed conformation yet, still associating with its G base pair as denoted by the hydrogen bond distances. In addition, a new contact with Asp283 and C3242 forms as the G/C base pair is being broken. (E) An overlay of wild type (gray) with the F243A structure (green and magenta) demonstrates a pathway for the movement of C3242 from its G/C base pair to its wild type conformation stacked with Phe243 and hydrogen bonded to Glu280. Note the alteration of the Asp283 side chain in the F243A structure (magenta, gradient arrow) that follows the movement of C3242, indicating its role to help break its G/C base pair.*

(F) An overlay of wild type (gray) with the F243A structure that demonstrates the intrahelical conformation of A3243 (green gradient arrow) that is stabilized by an altered, intrahelical conformation of Phe322 (magenta gradient arrow).

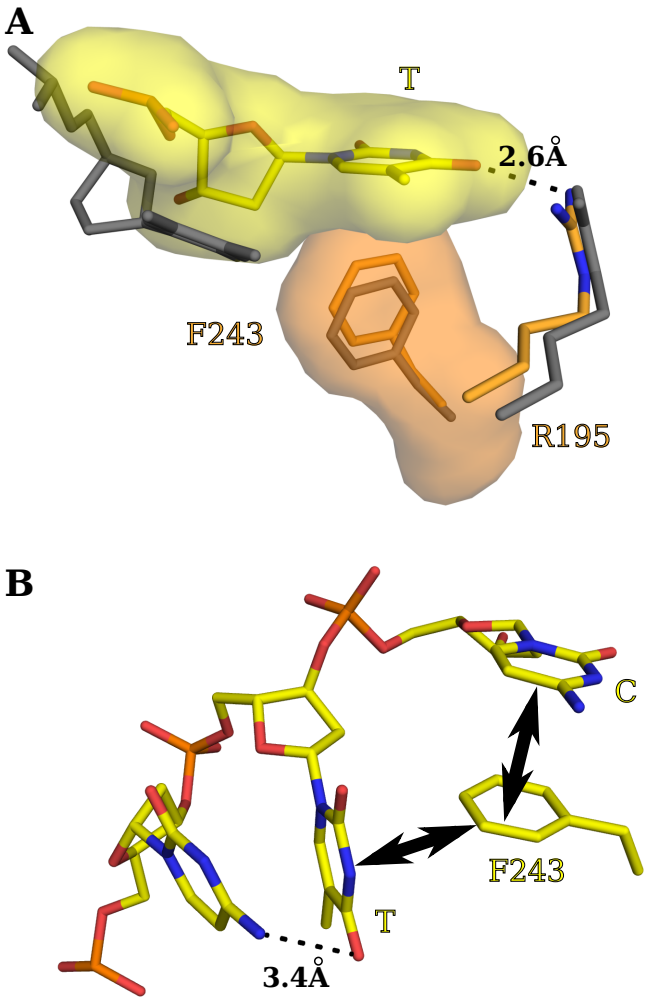

Figure 3.13 *(A) An overlay of wild type (gray) and the R162A substitution (yellow and orange), demonstrating a t-stacking interaction between Phe243 and T3243 that is absent in the wild type structure and the formation of a hydrogen bond between O2 of T3243 and N4 of Arg195. (B) The Y288A substitution structure show that Phe243 stacks with both C3242 and T3243 simultaneously in the heavy strand.*

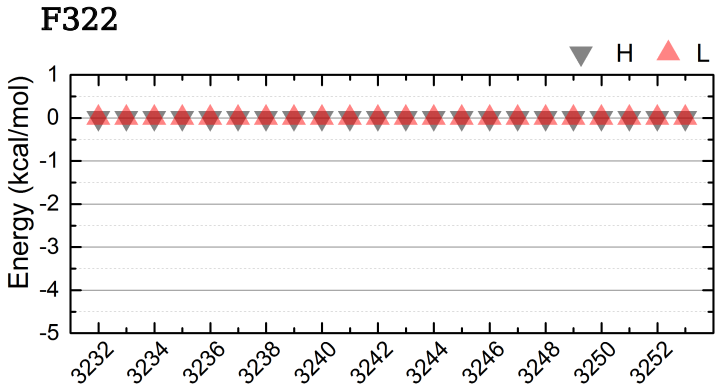

Figure 3.14 *Energy Decomposition Diagram showing that Phe322 has no interactions with the light strand (red triangles) or heavy strand (gray triangles) DNA in the wild type MTERF1 structure.*

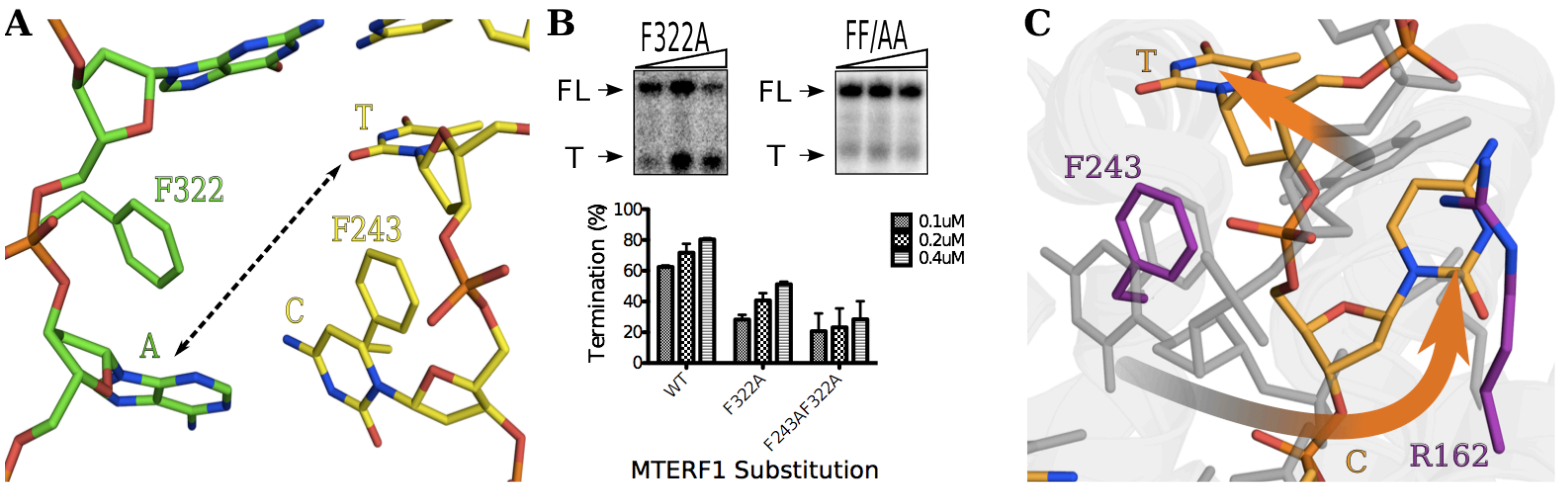

Figure 3.15 *(A) A composite structure of the F243A light strand DNA (green) with the Y288A heavy strand DNA (yellow) demonstrates that both Phe322 and Phe243 play a role in A/T base pair separation. (B) Termination assays for the F322A substitution and the double F243A/F322A substitutions. The bar graph values correspond the mean +/- SEM percentage termination from three independent experiments. (C) X-ray crystal structure of the F322A substitution (orange and magenta) with wild type (gray). The perturbed base flipping observed in this structure (gradient arrows) suggests Phe322 plays a role in coordinating proper flipping of the heavy strand nucleotides.*

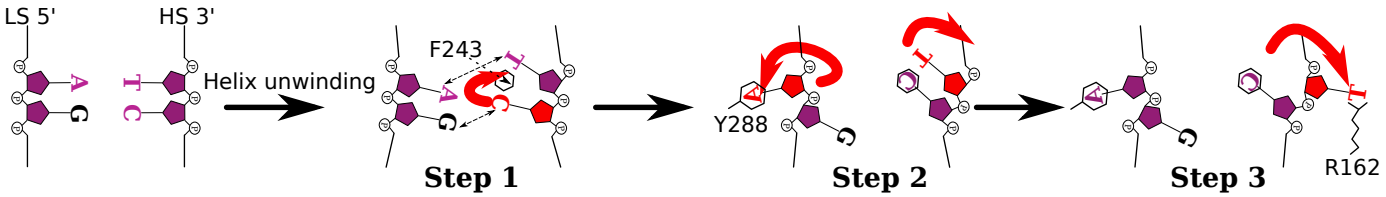

Figure 3.16 Proposed model for the stepwise order of base flipping by MTERF1. *Assisted by MTERF1 mediated binding and helix unwinding, the first step involves the breaking of the A/T base pair mediated by Phe322 and Phe243. This is a result of the structural conformation observed in both the F243A and Y288A structures (see Figure 3.15A). In addition, the first step also involves the movement of C3242 away from its G3242 partner on the LS. This is observed in the F243A structure (see Figure 3.12 D&E), which is an intermediate conformation to base flipping. In Step 2, we propose that A3243 moves out as C3242 moves to its wild type position stacked with Phe243. The movement of A3243 is coupled to the movement of C3242 as shown by the intrahelical conformation of both A3243 and C3242 in the F243A structure (see Figure 3.12 C&D). Finally, Step 3 is the movement of T3243 to its position stacked with Arg162. This final step is a result of the observations from the Y288A and R162A crystal structures that show an intrahelical conformation of T3243, suggesting that the orientation of C3242 and A3243 must occur before T3243 is able to flip out of the helix.*

Chapter 4 – Pathogenic Mutations alter the MTERF1 binding mode

In this chapter, the termination assays and ITC experiments for each of the pathogenic mutations were done in collaboration with Yakubovskaya E, Mejia E, Byrnes J, Hambardjieva E, Garcia-Diaz M, published in *Cell 141,* 982-993, June 11, 2010.

The crystal structures, thermodynamic and biochemical assays discussed so far have demonstrated that MTERF1 function is dependent on recognition of the specific $tRNA^{Leu}$ sequence and base flipping. Recall from Chapter 1, the mitochondrial $tRNA^{Leu}$ gene contains a very high concentration of pathogenic mutations. Interestingly, several of the pathogenic mutations are found within the 22mer MTERF1 binding site (Figure 4.2A). Based on our knowledge of the wtMTERF1 crystal structure, several of these mutations occur at locations involved in sequence recognition or base flipping. In fact, one of the most common mitochondrial mutations, A3243G present in ~0.02% of the adult population (16.3/100000 people) (110), corresponds to one of the bases that is flipped out of the helix. Although evidence suggests that the pathogenic effects of the A3243G mutation are likely a result of perturbations to the $tRNA^{Leu}$ structure, it has also been demonstrated that MTERF1 has a reduced affinity for the A3243G binding site (115,132). Thus, to investigate the roles by which these mutations may alter the thermodynamic and biochemical properties of MTERF1, we conducted ITC experiments and termination

assays with several of the mutations predicted to alter either sequence recognition or base flipping.

Pathogenic mutations within the MTERF1 binding site alter binding affinity and termination activity.

The results of the ITC experiments demonstrate that wtMTERF1 has only a small reduction in affinity for all of the mutations except G3249A (Table 4.1). For the case of A3243G, our data confirms other work demonstrating that MTERF1 only has a small decrease in affinity (115) for the sequence. Interestingly, MTERF1 shows the greatest decrease in affinity for G3249A (14uM) compared to WT (1uM). The stoichiometry for the G3249A mutation is much lower than the 1:1 (DNA:protein) ratio of binding observed for the WT sequence. These data suggests that in the presence of the G3249A mutation, more than one MTERF1 molecule can associate with the single DNA substrate. Thus, MTERF1 can adopt a binding mode that is different from WT, which only has one MTERF1 molecule bound to the single DNA substrate. Moreover, these data suggest that nucleotide 3249 is important for the ability of MTERF1 to adopt the wild type binding mode.

To investigate the functional consequences of the DNA mutations on MTERF1 activity, we conducted termination assays. Overall, the MTERF1 termination activity is reduced in the presence of each mutation. Not surprisingly, the largest defect was present for G3249A suggesting that the reduction in affinity and altered binding mode is important for function. In addition, MTERF1 was able to mediate some termination activity in the presence of the A3243G and G3244A mutations, while termination

defects were only apparent at low concentrations of MTERF1 with A3243T (Figure 4.1).

Curiously, the reasons for the termination defects are not obvious based on the biochemical and structural data thus far. For instance, the affinity and binding modes of MTERF1 for A3243T and A3243G is similar, however the ability of MTERF1 to terminate in the presence of these mutations is quite different. The wtMTERF1 crystal structure provides clues that we can use to predict what part of the MTERF1 binding mode is affected as a consequence of these mutations (Figure 4.2A). In the case of the G3249A mutant, the functional defect observed may be a result of a perturbation to a sequence specific contact between Arg387 and G3249. Similarly, for A3243G and A3243T, we can surmise that somehow, base flipping is perturbed as these residues are directly involved in the flipping event. However, the G3244A mutation has been shown to decrease termination activity but the location of the mutation does not directly involve specific contacts with MTERF1 or base flipping. Therefore, in order to understand the structural perturbations that result in the functional defects, we decided to crystalize wtMTERF1 bound to four pathogenic mutations: A3243G, A3243T, G3249A and G3244A.

A3243G and A3243T alter the base-flipped state.

In addition to A3243G being one of the most frequent mitochondrial mutations (109), its location within the MTERF1 binding sequence corresponds to two of the nucleotides that are everted from the DNA double helix. Due to its high frequency in

the population and key location at the site of base flipping, we decided to crystallize wild type MTERF1 bound to its binding sequence containing the A3243G mutation.

The crystal structure was solved to 2.9Å (Table 4.2) and reveals altered conformations of the flipped bases (Figure 4.3A). Interestingly, the most significant conformational change is the intrahelical conformation of C3243 on the HS. In, wild type position 3243 on the HS is a thymine and forms a cation π- stacking interaction with Arg162, however, the mutation results in the loss of this interaction (gradient arrow, Figure 4.3A). The intrahelical conformation of C3243 does not affect the conformation of the adjacent C3242 nucleotide on the HS. C3242 remains stacked with Phe243 in a conformation similar to that found in wild type. The conformation of G3243 on the LS is very similar to wild type and maintains the stacking interaction with Tyr288 (Figure 4.3A).

In the wild type structure, the flipped bases form hydrogen bonds with MTERF1 residues that act to stabilize the base in a binding pocket. This suggests selectivity of the binding pocket for specific types of nucleotide bases. Although the A3243G mutation alters base flipping, the mutation from an adenine to guanine is a transition mutation, which maintains the purine ring of the base. Oppositely, a transversion mutation would alter the ring structure from a purine to a pyrimidine. Thus, a transversion mutation such as A3243T drastically alters the size of the ring structure. To assess the structural changes as a result of the pathogenic A3243T transversion mutation we crystallized the mutated DNA sequence with wtMTERF1 and solved it to 2.3Å (Table 4.2). The structure demonstrates a severe alteration to the conformations of all three everted nucleotides (gradient arrows, Figure 4.3B). On the

HS, T3243A shows an intrahelical conformation that no longer interacts with Arg162. The intrahelical conformation of T3243A on the HS is stabilized by a t-stacking interaction with Phe243 that is not observed in the wild type structure. Furthermore, Phe243 simultaneously forms another t-stacking interaction with C3242 of the HS. In wild type, C3242 forms a π-stacking interaction with Phe243 (Figure 4.2D). However, even though the identity of this base has not been altered, the conformation of C3242 has changed. In this structure, C3242 forms simultaneous interactions with its G/C base pair and with Phe243. In the LS, we observe that T3243 (gradient arrows, Figure 4.3B) is located intrahelically (or within the helix) and no longer forms a stacking interaction with Tyr288. In addition, T3243 is stabilized by an altered conformation of Phe322. In the wild type structure, Phe322 is buried within the protein. In the A3243T structure, it now intercalates the LS between nucleotides T3243 and G3244 (gradient arrow, Figure 4.3C).

The structure of the A3243T mutation has revealed that all three nucleotides that are normally flipped out of the double helix are located intrahelically. This observation appears to represent an intermediate conformation before the bases are flipped out of the helix. Furthermore, the intrahelical conformations of the bases form simultaneous interactions with Phe322 and Phe243 (Figure 4.3D). This suggests that the phenylalanines, working in concert, may act to stabilize the intermediate conformation observed in this structure and may play a role in coordinating the flipping event. Additionally, we observe that the A3243T transversion mutation perturbs the base-flipping region more severely when compared to the transitional A3243G mutation mentioned above. This suggests that altering the bases perturbs the

interactions that would allow for normal base flipping. Moreover, altering the ring structure changes the size of the base and as a result, can no longer be accommodated in the binding pocket.

G3249A disrupts a key arginine-guanine interaction.

We have demonstrated that both the A3243G and A3243T mutations perturb the conformation of the base-flipping region. However, our ITC work shown in Chapter 3 has demonstrated that base flipping is most likely dependent upon recognition of the specific $tRNA^{Leu}$ sequence. Thus, sequence recognition is a critical first step before base flipping. Five arginine residues that make specific contacts with five guanine nucleotides within the MTERF1 binding site mediate the process of sequence recognition. (Figure 4.2A) Importantly, one such interaction is established between Arg387 and G3249. Previous work has shown that an R387A substitution severely alters function and changes the affinity and binding mode of MTERF1 for the WT DNA sequence (Table 3.1) Furthermore, the pathogenic G3249A is present at this location and has been shown to alter the affinity and binding mode of MTERF1 for the mutated DNA sequence (Table 4.1 and Figure 4.1) To establish the structural consequences that would result from this mutation, we crystallized the G3249A mutant sequence with wild type MTERF1 and solved it to 2.6Å (see Chapter 2 and Table 4.2). The structure reveals that the interaction between Arg387 and G3249A is lost (Figure 4.4A). In fact, the interaction of the Arg387 side chain with the base is so unfavorable, presumably due to the clashing of hydrogen atoms, that the side chain is bent away from all contact with the DNA (cyan sticks, Figure 4.4A). Interestingly, MTERF1 is

able to elicit normal base flipping in the presence of the G3249A mutation (Figure 4.4B). Therefore, the only alteration observed in this structure is the absence of the arginine-guanine interaction. Furthermore, the loss of this interaction does not preclude MTERF1 from adopting a presumably functional conformation of the DNA. This suggests that the altered binding mode observed (Table 4.1) and the functional defect is strictly due to the decrease in affinity of MTERF1 for the G3249A sequence.

G3244A does not affect base flipping or sequence specificity.

Thus far the pathogenic mutations we have examined played a role in altering either the base flipping mechanism or specific interactions formed between MTERF1 and the DNA. Interestingly, the G3244A mutation does not seem to be involved in either base flipping or sequence recognition based on the wtMTERF1 crystal structure. Yet, MTERF1 termination activity is reduced in the presence of this mutation. In addition, ITC experiments have shown that the stoichiometry and affinity are relatively similar to that of wild type (Table 4.1). Since the wild type crystal structure does not directly explain the possible role of G3244, we crystallized and solved the G3244A structure to 2.6Å (Table 4.2). The structure confirms that there are no alterations to the base flipping event (overlay, Figure 4.5A) or any sequence specific contacts between the DNA and MTERF1. At the site of the transversion mutation, however, an altered conformation of the 3244 base pair can be observed (gradient arrow, Figure 4.5A). The LS A3244 is in a normal helical conformation, however, T3244 on the HS has two conformations. One conformation forms a partial Watson-Crick base pair with A3244 on the LS and the other is located extrahelically (Figure 4.5B). However, the melting

of the DNA duplex as a result of the G3244A mutation does not change any of the expected base flipped conformations. Thus, this structure reveals that DNA distortion due to the unwinding imposed by MTERF1 extends beyond the two base pairs that are affected in the wild type structure. Hence, the presence of a G:C base pair at this position is presumably crucial to constrain the distortion to the preceding base pairs.

In summary, the presence of pathogenic mutations within the MTERF1 binding site results in functional and structural perturbations that alter sequence specificity or base flipping. We have identified both Phe322 and Phe243 as important intermediates in the steps that lead to base flipping and we have demonstrated that the G/C base pairs around the base-flipping region play an important role in preventing further DNA melting of the binding sequence. This information suggests the high dependence of MTERF1 on the specific $tRNA^{Leu}$ sequence environment for function.

Pathogenic mtDNA mutations affect MTERF1 binding and function.

The underlying molecular mechanisms of mitochondrial gene expression are poorly understood. However, different alterations to this process are linked to the pathogenesis of mitochondrial disorders. Although transcription is a critical event that is absolutely necessary for the production of the 13 proteins necessary for OXPHOS, no direct association have been found between alterations to the transcription process and mitochondrial disease. Nevertheless, as others and we have demonstrated, numerous pathogenic mtDNA mutations affect the capacity of MTERF1 to bind to its target site and mediate transcriptional termination (132). This suggests that defects in the ability of MTERF1 to terminate might play a role in pathogenesis.

We have thoroughly characterized the structural and functional consequence of pathogenic mutations found within the $tRNA^{Leu}$ gene. Binding of MTERF1 to its target sequence relies on sequence recognition, which relies on sequence specific contacts formed between MTERF1 and the DNA. Binding then results in the eversion of 3 nucleotides outside the helix. This event relies on sequence specific contacts formed between MTERF1 and the DNA sequence. In this work, we observed that different pathogenic mutations affect different aspects of this mechanism. For instance, A3243G and A3243T alter the base-flipping region, but do not seem to affect the initial sequence recognition step. G3244A, on the other hand, seems to affect the thermodynamic characteristics of the binding sequence, altering the bound conformation of MTERF1. Finally, G3249A directly affects one of the key sequence-specific interactions established by MTERF1 for initial sequence recognition. Interestingly, this mutation does not prevent MTERF1 from adopting a functional conformation on the target sequence, but reduces the affinity of MTERF1 for the binding site (and therefore its binding specificity). Furthermore, our work has also demonstrated that these mutations severely alter the termination ability of MTERF1 *in vitro*. These severe termination defects suggest that MTERF1 function is affected in carriers of these mutations and argue that this defect might play a role in pathogenesis of mitochondrial disease.

Relevance of alterations of MTERF1 function for pathogenesis.

Since the MTERF1 binding site contains the most common A3243G mutation, the association between MTERF1 activity and the clinical presentations of the mutation

have been thoroughly investigated. *In vivo* studies have suggested that the clinical presentations of the A3243G mutation may be a result of alteration to the $tRNA^{Leu}$ structure (115). However, previous work suggests that the binding affinity of MTERF1 for the A3243G mutated sequence is reduced (132). Our ITC work in this study has confirmed this result. One of the observations that supported the hypothesis that MTERF1 effects played no pathogenic role in A3243G mutation carriers was the fact that the level of upstream and downstream transcripts produced from HSP does not change in the presence of A3243G. While this indeed suggests that the mutation has no effect on MTERF1 termination ability on HSP transcription, our work, and others, has shown that termination activity mediated by MTERF1 occurs in a polar fashion and acts preferentially on LSP transcription. In support of this, a mouse knockout model of MTERF1 shows that transcription termination events from LSP are altered and that HSP termination is not affected. Furthermore, in the presence of the A3243G mutation, the ability of MTERF1 to terminate when the sequence is oriented in reverse to HSP is severely decreased. Thus it is possible that the main consequence of the A3243G mutation is the alteration of LSP termination events. To the extent that these alterations might have deleterious effects, our biochemical work suggests that the alterations to MTERF1 function caused by several pathogenic mtDNA mutations may play a role in the pathogenesis of mitochondrial disease.

DNA	Stoichiometry (N)	K_d (μM)
WT	1.06 ± 0.08	1.05 ± 0.18
A3243G	1.04 ± 0.08	3.02 ± 0.39
A3243T	0.98 ± 0.06	3.12 ± 0.33
G3244A	1.06 ± 0.04	1.91 ± 0.13
G3249A	0.20 ± 0.12	14.04 ± 2.32

Table 4.1 *ITC data reveals the stoichiometry and binding constants (K_d) of wtMTERF1 for the WT DNA sequence and four pathogenic mutations found within the MTERF1 binding sequence. *Table adapted from figure 6 of Yakubovskaya E, Mejia E, Byrnes J, Hambardjieva E, Garcia-Diaz M published in Cell 141, 982-993, June 11, 2010*

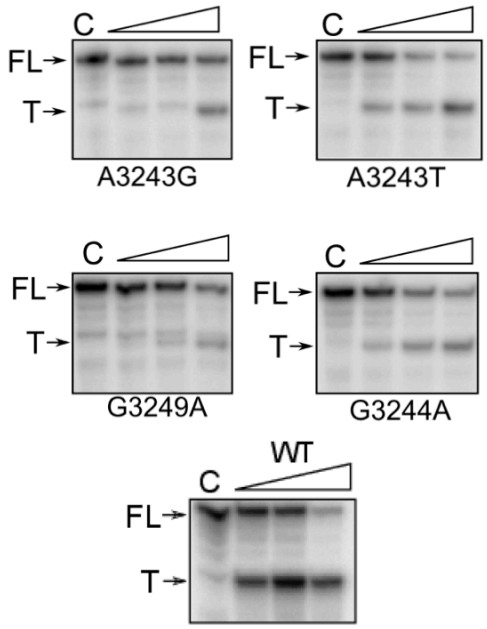

Figure 4.1 Termination ability of wtMTERF1 in the presence of 4 pathogenic mutations found within the binding site. *All assays were carried out with 0.1uM, 0.2uM and 0.4uM of wtMTERF1 with the termination sequence oriented in reverse with respect to the HSP. The control lane, C, contains the DNA template, POLRMT, TFAM and TFB2 but does not contain MTERF1. *Figure adapted from figure 6 of Yakubovskaya E, Mejia E, Byrnes J, Hambardjieva E, Garcia-Diaz M published in Cell 141, 982-993, June 11, 2010*

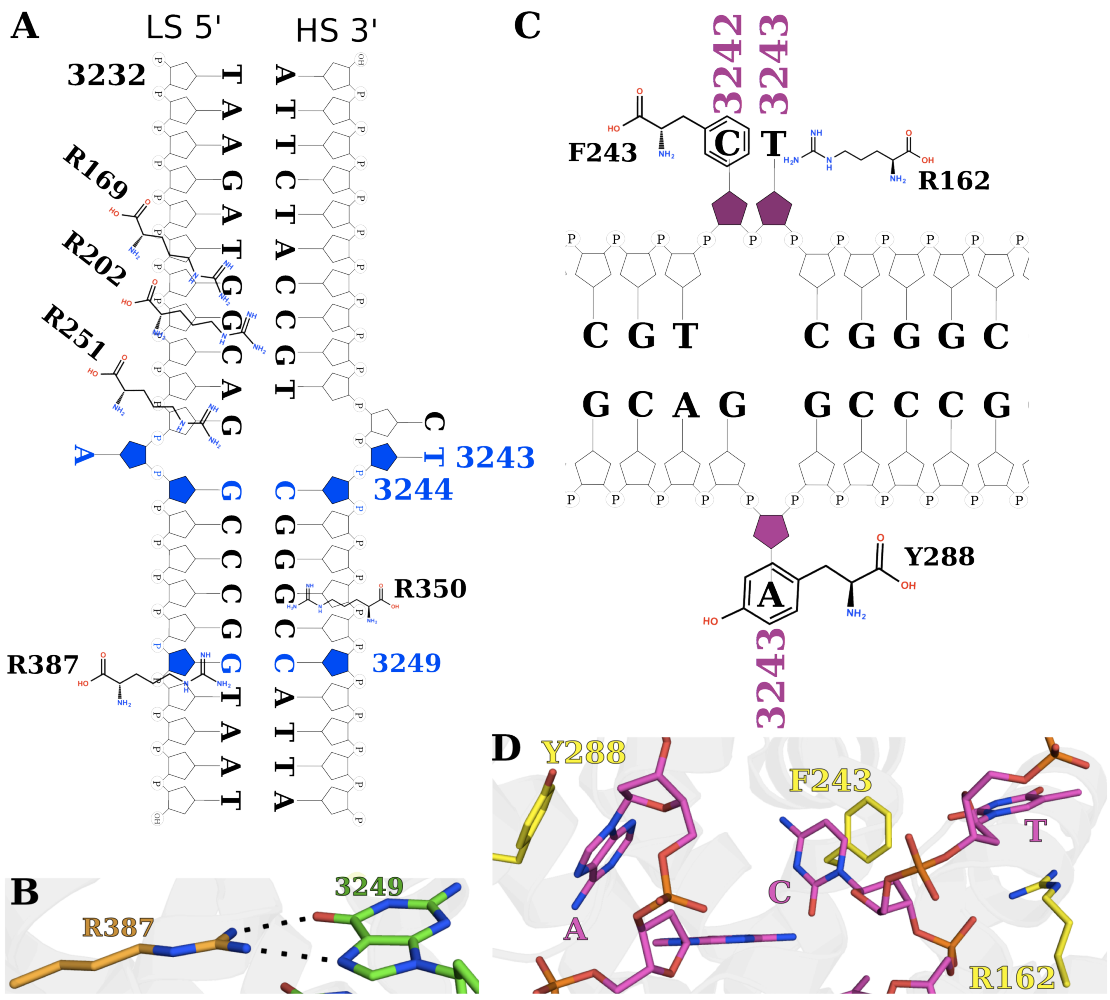

Figure 4.2 Pathogenic mutations located within the MTERF1 DNA binding site. *(A) The 22mer DNA binding site for MTERF1. Highlighted in blue is the location of four pathogenic mutations. Furthermore, the five arginine-guanine interactions are labeled, which are critical for sequence recognition. (B) Zoomed in view of the Arg387/G3249 hydrogen bond interaction. (C) Schematic showing the 3 flipped bases stabilized by stacking interactions. (D) View of the base-flipping region in the wtMTERF1 crystal structure.*

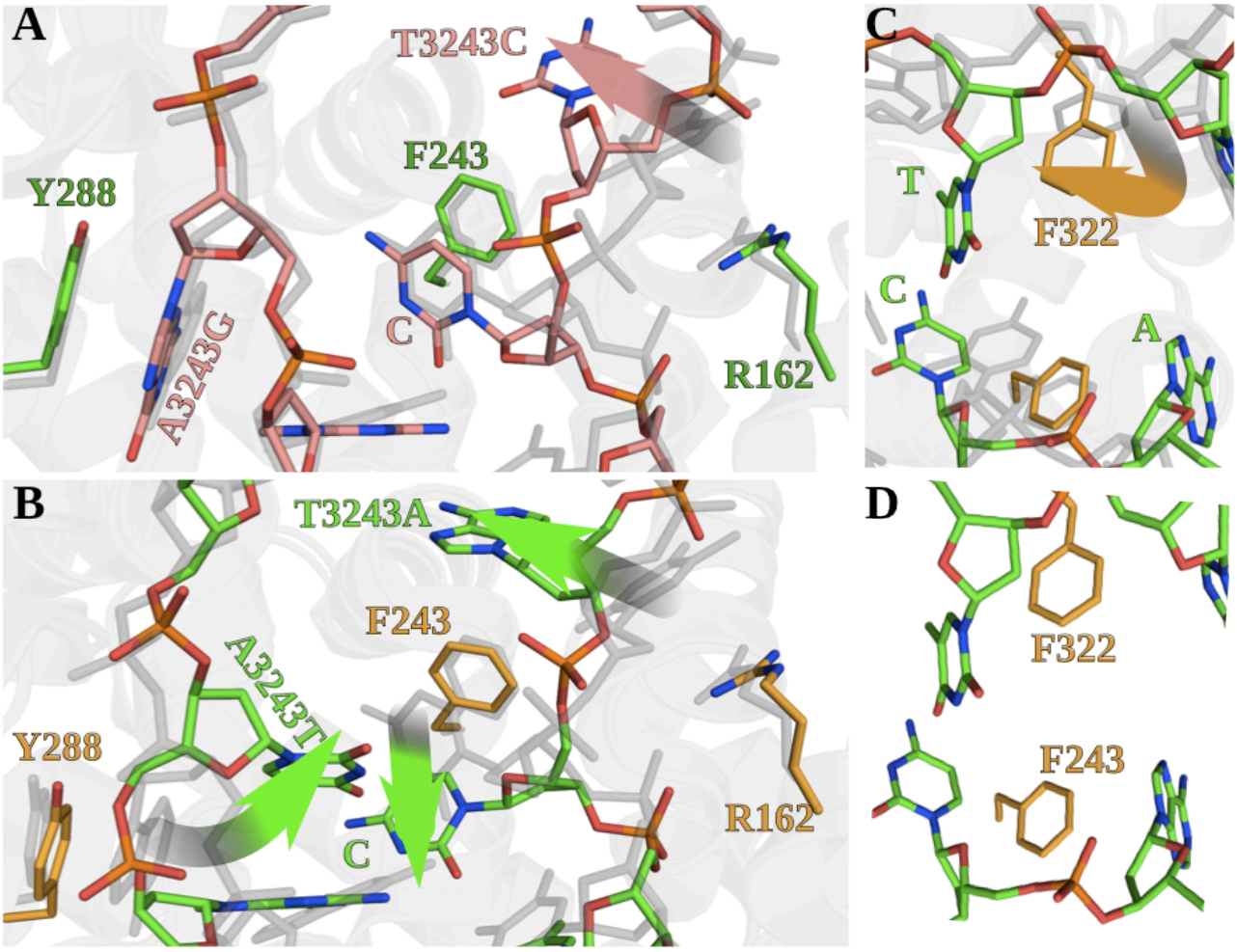

Figure 4.3 The crystal structures of wtMTERF1 bound to the A3243G and A3243T mutations. *In the overlays (A-C) wtMTERF1 is colored gray and transparent. (A) Structure of the A3243G mutation (protein colored green and DNA colored salmon). The base-flipping region reveals that T3243C is now located intrahelical (gradient arrow) as a result of the transition mutation. (B) Crystal structure of the A3243T mutation (DNA colored green, protein colored orange). As a result of the transition mutation, A3243T and T3242A are now located intrahelical (gradient arrows). The C no longer forms a stacking interaction with F243 and is now in a conformation more closely associated with its Watson-Crick base pair. (C) The A3243T structure reveals an altered conformation of F322, a residue that is not involved in stabilizing the final base flipped state (Chapter 3). In this structure F322 intercalates the LS and stabilizes the intrahelical T3243. (D) The base-flipping region of the A3243T structure shows the simultaneous intercalation of F243 and F322 into the HS and LS respectively. The role of these two phenylalanines is critical for coordinating the flipping mechanism (Chapter3).*

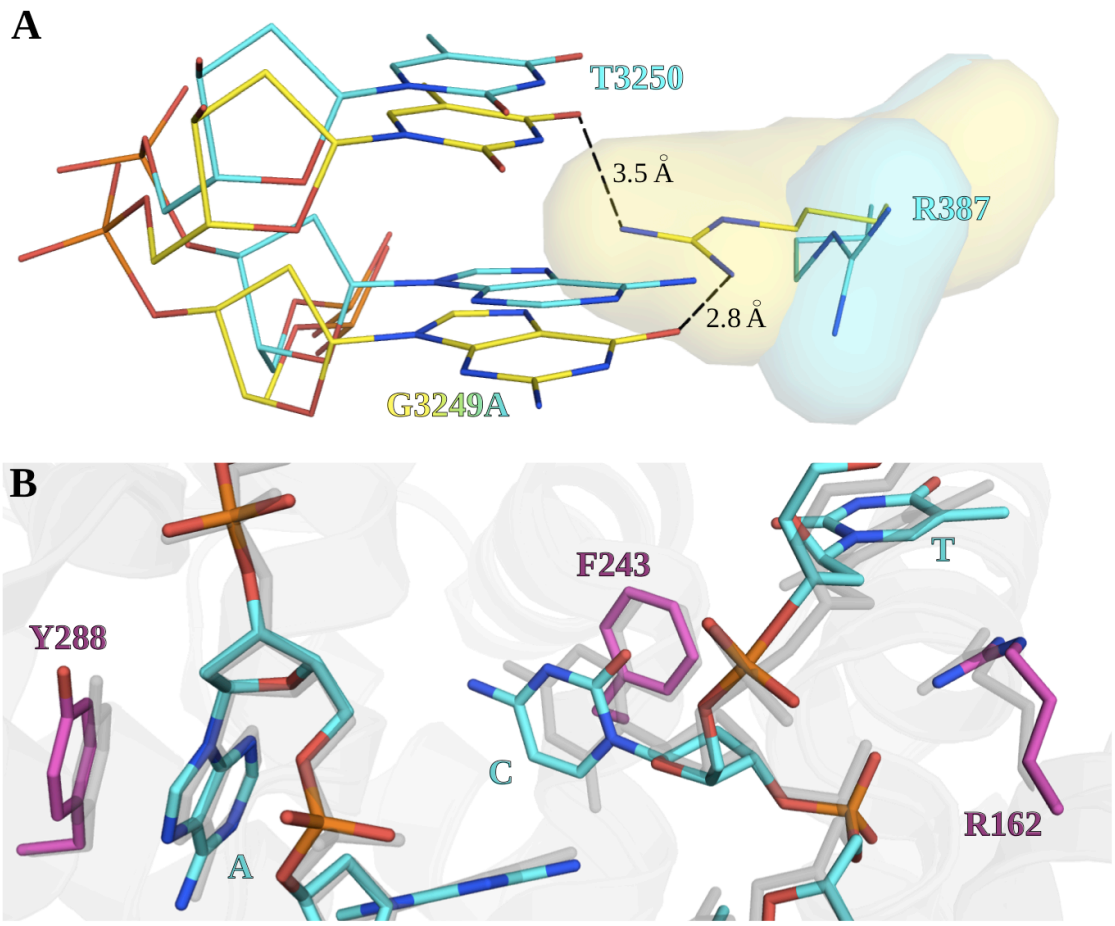

Figure 4.4 The G3249A structure perturbs MTERF1 sequence recognition. *(A) An overlay of the wild type DNA and wild type MTERF1 (backbone carbon atoms colored yellow) with the G3249A mutation and wild type MTERF1 (backbone carbon atoms colored cyan). Arg387 normally makes contacts with G3249A and T3250 in wild type. However, in the G3249A structure, the Arg387 side chain no longer forms contacts with the DNA. (B) Interestingly, the G3249A structure is able to elicit wild type base flipping. This structural information along with the ITC data (Table 4.1) suggest that the sequence recognition mechanism is perturbed as a result of the G3249A mutation.*

	G3244A	G3249A	A3243G	A3243T
Data Collection				
PDB Code				
R_{merge}	0.054 (0.843)	0.044 (0.828)	0.072 (0.972)	0.033 (0.513)
R_{meas}	0.062 (0.910)	0.049 (0.870)	0.079 (1.087)	0.038 (0.586)
R_{pim}	0.033 (0.509)	0.026 (0.454)	0.043 (0.589)	0.026 (0.392)
No. of total observations	128009 (1159)	143694 (1474)	183048 (1940)	67339 (654)
I/σI	20.2 (2.2)	22.1 (2.3)	19.7 (2.1)	20.6 (2.3)
$CC_{1/2}$	0.999 (0.798)	1 (0.744)	0.999 (0.741)	0.999 (0.758)
Completeness %	100 (100)	99.8 (100)	99.8 (100)	92.6 (91.9)
Multiplicity	6.3 (5.8)	6.5 (6.9)	6.4 (6.3)	2.8 (2.7)
Cell Dimensions (a, b, c)	89.1, 89.6, 164.8	88.4, 90.8, 163.3	55.0, 83.4, 133.6	88.3, 89.3, 161.9
α, β, γ	90, 90, 90	90, 90, 90	90, 99.7, 90	90, 90, 90
Space Group	C2221	C2221	P21	C2221
Resolution Å	2.6	2.5	2.8	2.3
Refinement				
R_{work}/R_{free}	0.20/0.27	0.23/0.27	0.21/0.26	0.22/0.27
R.m.s.d. Bond Length Å	0.013	0.012	0.01	0.014
R.m.s.d. bond Angle	1.6	1.6	1.5	1.7
Average B-factor	76.3	50.3	77.0	63.3

Table 4.2 X-ray crystallography data collection and statistics.

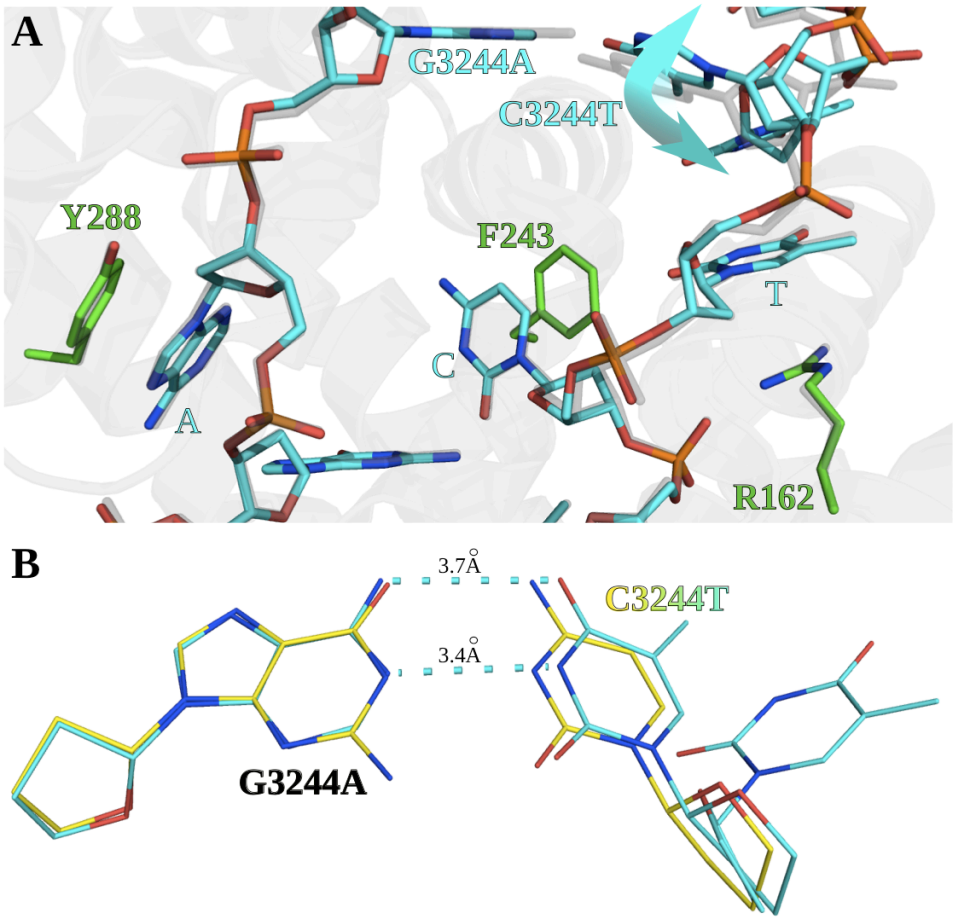

Figure 4.5 G3244 is important to prevent further duplex melting. *G3244 is not directly involved in base flipping or the formation of specific contacts with MTERF1. The crystal structure of G3244A overlaid with wild type (A) shows that the flipped bases are in a wild type conformation however, C3244T has two occupancies (cyan double arrow). (B) An overlay of the 3244 base pair with wild type (backbone carbon atoms colored yellow) with the G3244A structure (backbone carbon atoms colored cyan). The mutation eliminates the stronger G/C base pair to the weaker A/T base pair, which results in duplex melting beyond the base flipping site.*

Chapter 5 – MTERF1 allelic variation has deleterious effects on transcription termination

Identification of MTERF1 allelic variants within sample populations.

The presence of known pathogenic mutations within the DNA binding sequence has been shown to alter several mechanical and functional aspects mediated by wild type MTERF1 as discussed in Chapter 4. We have also shown that synthetically engineered substitution constructs of the MTERF1 protein, such as the R162A substitution that destabilizes the base-flipping region, has functional consequences in the presence of wild type DNA (Chapter 3). However, unlike the pathogenic mutations located within the DNA binding sequence, the MTERF1 substitution constructs do not represent mutations that may exist in the population. Therefore, we decided to investigate point mutations present in MTERF1 by searching for single nucleotide polymorphisms (SNPs) within the human MTERF1 gene that have been reported in a sample human population.

A search of the dbSNP database (NCBI) for MTERF1 resulted in ~138 SNPs. To focus our search, we decided to look for non-synonymous missense mutations that alter the amino acid sequence and had the potential to interfere with DNA binding, base-flipping, or protein folding, or presented with an unusually high allelic frequency. The missense mutations that met these criteria are shown in Table 5.1 along with their associated allelic frequencies from the sample populations (see Table 5.1 legend). Most interestingly, the highest frequency (~40%) observed within the sample populations correspond to A294T, a residue that does not immediately appear to be

involved in the sequence recognition or base-flipping processes mediated by MTERF1. In addition, one rare (~0.1%) non-sense mutation was found that results in a premature stop codon at Leu79. Furthermore, the locations of R202H, R169G/Q, R251Q and P242A, indicate that they may interfere with the base-flipping or binding event. The strategic locations of each variant and the fact that they are found in the population, motivated us to investigate and characterize the structural and functional consequences of each mutation. Termination assays were performed to gauge the functional ability of the mutation in the presence of the wild type DNA binding sequence. In addition, we predicted the structural consequences for each variant, by modeling the amino acid mutation in the wild type crystal structure

Functional consequences of Allelic variants within MTERF1 are dependent on their location.

A294T

The high frequency of the A294T (~40%, Table 5.1) mutation motivated us to investigate the consequences of this variant on the overall structure of MTERF1 and its termination activity. The position of A294 is found within an α - helix on the outer surface of the protein that is solvent exposed (Figure 5.1). To determine the possible environmental changes on the local and global structure, we modeled A294T in the wild type structure and the predicted conformation suggests that the threonine would not perturb the structure of the α helix. In addition, since it is located far away from any protein-DNA contacts or the base-flipping region, the mutation should not alter

any of these critical mechanisms (overlay, threonine colored white, Figure 5.1A). Thus, we hypothesized that A294T should terminate in a similar fashion as wtMTERF1. To verify this, we performed a termination assay and confirmed that indeed the functional ability of A294T to terminate is the same as wild type (Figure 5.2 A&B) and indicates that the variant has no functional consequences for MTERF1.

P242A

The location of Pro242 is in close proximity to the base-flipping region and mediates a sharp turn between two α helices (gradient arrow, Figure 5.1B). The turn is important for orienting Phe243 in order to form a key stacking interaction with C3242 (Figure 5.1B), which stabilizes base flipping. The P242A variant would most likely disrupt this turn and locally alter the base-flipping region. However the disruption of this turn would compromise the triangular structure of the mterf motif. In fact, all the proline residues in MTERF1 mediate turns that maintain the triangular shape for each of the eight MTERF1 motifs (Chapter 1). Not surprisingly, purification of the P242A variant has failed, as it co-purifies with the chaperone used in the expression system. Since chaperones are important in protein folding, co-purification of the P242A mutant with the chaperone suggests that the mutant is unable to fold properly.

R251Q, R169Q, R169G and R202H

The next four variants are found at locations that involve specific protein-DNA contacts that are important for MTERF1 binding and function (Chapter 3) and present at much lower frequencies than A294T. In the case of R251, a hydrogen bond is

formed between the amine group of R251 and O6 of G3242. The predicted orientation of R251Q reveals an increased distance of ~1.5Å between the two residues due to the shorter glutamine side chain, that eliminates the hydrogen bond interaction (white overlay, Figure 5.1C). Similarly, the model for R169Q suggests that the shorter glutamine side chain may interfere with the ability of MTERF1 in making a sequence specific contact with G3238 (Figure 5.1D). The functional consequences verify the importance of these arginine-guanine interactions, as both R251Q and R169Q are severely impaired of their ability to terminate (Figure 5.2 A&B). Although the R169G and R202H variants have not been tested *in vitro* yet, our current data strongly suggests that these variants will exhibit functional defects. Hence, a mutation to a glycine or histidine might prevent MTERF1 from being able to form specific interactions with the DNA at these locations.

Functional Implications for allelic frequency

It is interesting to note that the allelic variant frequencies for mutations that affect key MTERF1 mechanisms (e.g., R169Q) are very rare (Table 5.1). As we have demonstrated, alterations in key processes as a result of these variants, suggest that their low frequency in the sample populations reflect the importance of selecting against them throughout the evolution of MTERF1. Oppositely, the frequency of A294T is very high (40%) and there is no functional difference from wild type, indicating that the high frequency of the A294T allele is most likely tolerable and thus is not selected against in evolutionary terms. Although this is highly speculative, more work is needed to understand the *in vivo* role and clinical presentations of the variants

in order to associate an evolutionary selective advantage for or against a particular MTERF1 variant. Moreover, since the MTERF1 DNA binding site is a frequent site of mutations associated with disease, it would be interesting to look for compounding effects between individuals that might express both an allelic variant and a pathogenic mitochondrial DNA mutation in the MTERF1 binding site.

Missense Mutation (MTERF coding strand)	Amino Acid Change	Allelic Frequency (%)*
C/G	P242A	0.20
G/A	R202H	0.00**
G/A	R169Q	NA
C/G	R169G	0.02
C/T	A294T	40.2
G/A	R251Q	2.50
A/C	L79Ter (nonsense)	0.1

Table 5.1 List of Allelic Variants within MTERF1 that alter key mechanisms. *Allelic Frequencies reported have been validated by 1000 Genomes or Hap Map (See Chapter2).*
**Reported from NCBI dbSNP database*
***Reported in dbSNP but not validated in larger population cohorts such as 1000 Genomes or Hap Map.*

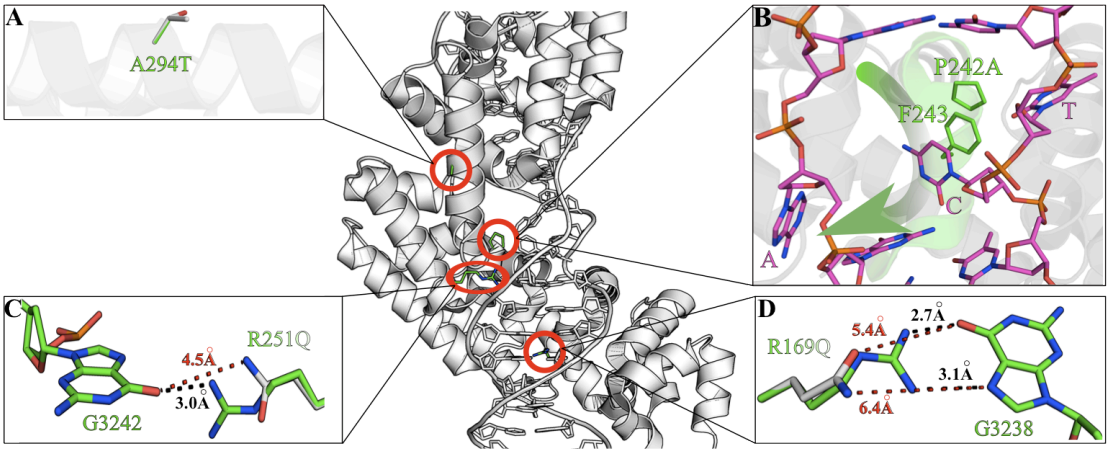

Figure 5.1 Location of four allelic variants within the MTERF1 structure. *(A) An overlay of Ala294, colored green, with the most probable side chain rotamer of Thr294 (backbone atoms colored white). (B) The location of P242 mediates a turn between two helices (gradient arrow) and orients Phe243 to stack with C, which stabilizes base flipping. (C) Interaction between R251 and G3242 (backbone atoms colored green) in an overlay with R251Q (white). The mutation to Gln increases the distance between the side chain and G3242 and is thought to no longer be able to mediate a critical sequence specific contact. (D) Same as (C) except for the R169Q variant.*

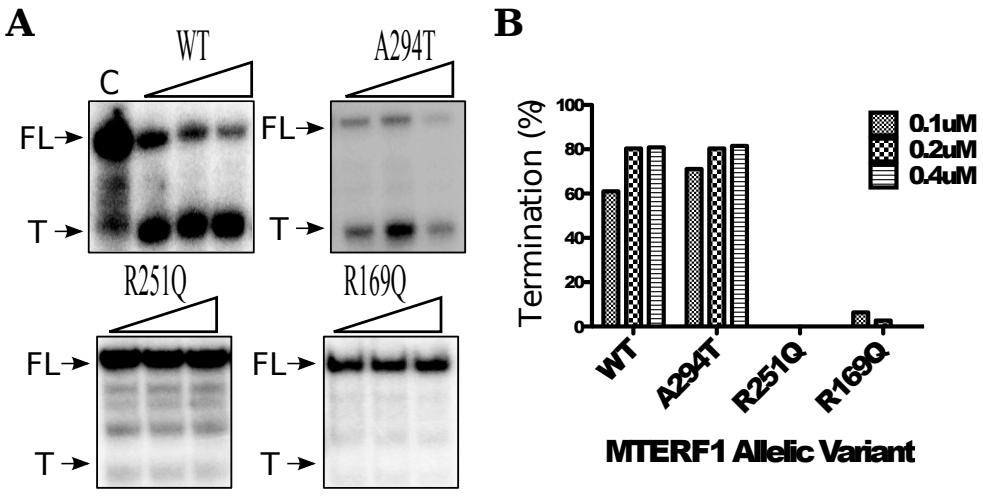

Figure 5.2 Termination ability of MTERF1 allelic variants. *(A) Termination assays for WT MTERF1 and three variants reveal a band corresponding to full-length transcripts (FL) or a termination transcript (T). The control lane, C, contains no MTERF1. (B) Percentage of termination activity determined from densitometry analysis of the gels in (A).*

Chapter 6 – Conclusions & Future Directions

The work contained herein describes the mechanical underpinnings of MTERF1, which are ultimately responsible for proper termination activity at the $tRNA^{Leu}$ site. The information obtained has broadened our knowledge regarding the rare 3-nucleotide base flipping mechanism and the role of MTERF1 in pathogenesis of mitochondrial diseases.

Base-Flipping Mechanism

We have determined that MTERF1 utilizes a base-flipping mechanism for stable specific binding to its location within the $tRNA^{Leu}$ gene. In this work, we demonstrate that the eversion of 3 nucleotides from the DNA duplex to pockets outside the DNA duplex is a complex mechanism that involves the coordination of several steps. We have demonstrated that the coordination of these steps involves two specific phenylalanine residues. Moreover, perturbations to the stepwise mechanism result in aberrant conformations of the everted nucleotides. Not surprisingly, aberrant base flipping has consequences for MTERF1 termination activity. Our biochemical work has assessed the functional contribution of each stacking interaction to the ability of MTERF1 to terminate. Strikingly, we have identified that each interaction does not contribute equally to function. Hence, in addition to the fact that alterations to the stacking interactions result in aberrant base flipping, MTERF1 is able to accommodate an altered base flipped state and maintain residual activity. This suggests that MTERF1 might be able to bind to alternate sequences that vary from the consensus. Although termination activity is not as strong as wild type, this has implications for the role of alternate MTERF1 activities at other sites in the mitochondrial genome, such as

replication pausing or regulating events at the HSP. However, structural and biochemical studies are needed to determine the ability of MTERF1 to bind alternate sites such as HSP.

Implications for the MTERF Family of Proteins

As previously discussed, the fold of MTERF1 is conserved amongst other members of the MTERF family. The crystal structures of MTERF1 and MTERF3 are highly similar and maintain the same mterf motif and overall shape indicative of DNA binding. Importantly, biochemical evidence suggests that MTERF3 interacts with DNA within the regulatory D-Loop (86). Interestingly, Phe243, the residue that is important for stabilizing one of the flipped bases in MTERF1, is conserved in MTERF3 (and MTERF2). This, in addition to the fact that MTERF3 interacts with DNA and shares structural homology with MTERF1 suggests that other MTERF family members may be able to adopt a similar binding mode as MTERF1. Moreover, we have established the importance of arginine-guanine contacts for the ability of MTERF1 to recognize its specific DNA binding sequence and elicit base flipping. It is plausible that different locations of the arginine residues in MTERF3 (or 2) could result in alternative binding sites that may elicit a base flipping mechanism.

Interestingly, the MTERF1 fold shares structural homology to PUF domains that are found in proteins that bind RNA. Much like the mterf motifs found in MTERFs 1, 3 and 4, PUF domains are all α-helical and triangular in shape. In addition, MTERF 3 and 4 have been shown biochemically to interact with mitochondrial rRNAs. Hence it is plausible that MTERF1 and other MTERF proteins

could interact with an RNA substrate and may not be strictly limited to binding dsDNA duplexes. Future biochemical and structural studies would be necessary to establish the possibility of alternative MTERF1 substrates.

MTERF1 has implications for Pathogenesis of Mitochondrial Disease

The molecular mechanisms of mitochondrial gene expression are poorly understood. Perturbations of mitochondrial gene expression have been shown to be associated with certain mitochondrial diseases. However, transcription of the mitochondrial genome is absolutely critical for OXPHOS. Yet, currently, deficiencies in the transcription process have not been associated with mitochondrial disease. We have shown that the presence of pathogenic mutations within the $tRNA^{Leu}$ gene have altered the ability of MTERF1 to bind, elicit normal base-flipping and terminate transcription at this site. Thus, our observations suggest an association between transcription defects at the $tRNA^{Leu}$ site and the pathogenesis of mitochondrial disease.

Interestingly, we and others have shown that MTERF1 terminates in a polar fashion and acts preferentially upon LSP transcripts. However, there is a lack of *in vivo* evidence regarding how pathogenic mutations present in the $tRNA^{Leu}$ gene may alter MTERF1 termination activity on LSP transcripts.

Finally, we have observed that allelic variants present within the MTERF1 protein can have deleterious effects on the ability of MTERF1 to mediate transcription termination. We have also shown that some of these variants occur at locations that

prevent the ability of MTERF1 to form sequence specific contacts with the DNA. However, more structural work is needed in order to correlate the structural alterations with the termination defects observed. The severe termination deficiencies observed for some of the MTERF1 allelic variants suggests a possible role in the pathogenesis of mitochondrial disease.

References

1. McBride HM, Neuspiel M, Wasiak S. Mitochondria: More than just a powerhouse. Curr Biol. 2006;16(14):R551–60.

2. Hatefi Y. The Mitochondrial Electron Transport and Oxidative Phosphorylation System. 1985. 1 p.

3. Munnich A, Rustin P. Clinical spectrum and diagnosis of mitochondrial disorders. Am J Med Genet. 2001;106(1):4–17.

4. Wallace DC. Mitochondrial DNA mutations in disease and aging. Environ Mol Mutagen. 2010 Jun;51(5):440–50.

5. Shadel GS. Coupling the mitochondrial transcription machinery to human disease. Trends in genetics. 2004.

6. Adzic M, Djordjevic A, Demonacos C, Krstic-Demonacos M, Radojcic MB. The role of phosphorylated glucocorticoid receptor in mitochondrial functions and apoptotic signalling in brain tissue of stressed Wistar rats. Int J Biochem Cell Biol. 2009 Nov 1;41(11):2181–8.

7. Zamora M, Meroño C, Viñas O, Mampel T. Recruitment of NF-kappaB into mitochondria is involved in adenine nucleotide translocase 1 (ANT1)-induced apoptosis. J Biol Chem. 2004 Sep 10;279(37):38415–23.

8. Talabér G, Boldizsár F, Bartis D, Pálinkás L, Szabó M, Berta G, et al. Mitochondrial translocation of the glucocorticoid receptor in double-positive thymocytes correlates with their sensitivity to glucocorticoid-induced apoptosis. Int Immunol. 2009 Nov 1;21(11):1269–76.

9. Wang X, Peralta S, Moraes CT. Mitochondrial alterations during carcinogenesis: a review of metabolic transformation and targets for anticancer treatments. Adv Cancer Res. 2013;119:127–60.

10. Mathupala SP, Ko YH, Pedersen PL. The pivotal roles of mitochondria in cancer: Warburg and beyond and encouraging prospects for effective therapies. ACTA-BIOENERG. 2010;1797(6-7):1225–30.

11. Lemasters JJ, Holmuhamedov E. Voltage-dependent anion channel (VDAC) as mitochondrial governator—Thinking outside the box. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2006 Jan 1;1762(2):181–90.

12. Majewski N, Nogueira V, Robey RB, Hay N. Akt inhibits apoptosis downstream of BID cleavage via a glucose-dependent mechanism involving mitochondrial hexokinases. Mol Cell Biol. 2004 Jan;24(2):730–40.

13. Otto Warburg FWEN. The metabolism of tumors in the body. The Journal of General Physiology. The Rockefeller University Press; 1927 Mar 7;8(6):519.

14. Schapira AHV, Cooper JM, Dexter D, Clark JB, Jenner P, Marsden CD. Mitochondrial Complex I Deficiency in Parkinson's Disease. J Neurochem. 1990 Mar;54(3):823–7.

15. Pinto M, Moraes CT. Mitochondrial genome changes and neurodegenerative diseases. Biochim Biophys Acta. 2014 Aug;1842(8):1198–207.

16. Trifunovic A, Wredenberg A, Falkenberg M, Spelbrink JN, Rovio AT, Bruder CE, et al. Premature ageing in mice expressing defective mitochondrial DNA polymerase. Nature. 2004 May 27;429(6990):417–23.

17. Siekevitz P, Van R Potter. Biochemical structure of mitochondria. ii. Radioactive labeling of intramitochondrial nucleotides during oxidative phosphorylation. J Biol Chem. 1955 Jul 1;215(1):237–55.

18. Scarpulla RC. Transcriptional activators and coactivators in the nuclear control of mitochondrial function in mammalian cells. Gene. 2002;286(1):81–9.

19. Stroud DA, Ryan MT. Mitochondria: organization of respiratory chain complexes becomes cristae-lized. Curr Biol. 2013 Nov 4;23(21):R969–71.

20. Smeitink JAM. Oxidative Phosphorylation in Health and Disease. Springer; 2007. 1 p.

21. Janssen RJRJ, Nijtmans LG, van den Heuvel LP, Smeitink JAM. Mitochondrial complex I: Structure, function and pathology. J Inherit Metab Dis. 2006 Aug;29(4):499–515.

22. Kukat C, Larsson N-G. mtDNA makes a U-turn for the mitochondrial nucleoid. Trends in Cell Biology. 2013 Sep;23(9):457–63.

23. Anderson S, Bankier AT, Barrell BG, De Bruijn MHL, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature. 1981 Apr 9;290:457–65.

24. Salinas T, Larosa V, Cardol P, Maréchal-Drouard L, Remacle C. Respiratory-deficient mutants of the unicellular green alga Chlamydomonas: a review. Biochimie. 2014 May;100:207–18.

25. Chandrasekaran K, Hatanpää K, Rapoport SI, Brady DR. Decreased expression of nuclear and mitochondrial DNA-encoded genes of oxidative phosphorylation in association neocortex in Alzheimer disease. Brain Res Mol Brain Res. 1997 Jan 1;44(1):99–104.

26. DiMauro S, Tanji K, Bonilla E, Pallotti F, Schon EA. Mitochondrial abnormalities in muscle and other aging cells: Classification, causes, and effects. Muscle Nerve. 2002 Nov;26(5):597–607.

27. Attardi G, Schatz G. Biogenesis of mitochondria. Annu Rev Cell Biol. 1988;4:289–333.

28. Andrews RMR, Kubacka II, Chinnery PFP, Lightowlers RNR, Turnbull DMD, Howell NN. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 1999 Oct 1;23(2):147–7.

29. Cech TR, Reisler F, Hearst JE. Partial denaturation of mouse DNA in preparative CsCl density gradients at alkaline pH. Biochemistry. 1976 May 4;15(9):1865–73.

30. Shadel GS, Clayton DA. Mitochondrial dna maintenance in vertebrates. Annu Rev Biochem. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA; 1997 Jun;66(1):409–35.

31. Schneider A, Ebert D. Covariation of mitochondrial genome size with gene lengths: evidence for gene length reduction during mitochondrial evolution. J Mol Evol. 2004 Jul 1;59(1):90–6.

32. Kasamatsu H, Robberson DL, Vinograd J. A Novel Closed-Circular Mitochondrial DNA with Properties of a Replicating Intermediate. Proceedings of the …. 1971.

33. Donald L Robberson HKJV. Replication of Mitochondrial DNA. Circular Replicative Intermediates in Mouse L Cells. Proc Natl Acad Sci USA. National Academy of Sciences; 1972 Mar 1;69(3):737.

34. Bogenhagen DF, Applegate EF, Yoza BK. Identification of a promoter for transcription of the heavy strand of human mtDNA: In vitro transcription and deletion mutagenesis. Cell. 1984 Jan 1;36(4):1105–13.

35. Bogenhagen DF, Yoza BK, Cairns SS. Identification of initiation sites for transcription of Xenopus laevis mitochondrial DNA. J Biol Chem. 1986 Jun 25;261(18):8488–94.

36. MacKay SL, Olivo PD, Laipis PJ, Hauswirth WW. Template-directed arrest of mammalian mitochondrial DNA synthesis. Mol Cell Biol. 1986 Apr;6(4):1261–7.

37. Peralta S, Wang X, Moraes CT. Mitochondrial transcription: Lessons from mouse models. Biochimica et Biophysica Acta (BBA)-Gene …. 2012.

38. Aloni Y, ATTARDI G. Symmetrical In Vivo Transcription of Mitochondrial DNA in HeLa Cells. Proceedings of the National Academy of …. 1971.

39. Murphy WI, Attardi B, Tu C, Attardi G. Evidence for complete symmetrical transcription in vivo of mitochondrial DNA in HeLa cells. Journal of molecular biology. 1975 Dec;99(4):809–14.

40. Ojala D, Merkel C, Gelfand R, ATTARDI G. The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. Cell. 1980.

41. Ojala D, Montoya J, ATTARDI G. tRNA punctuation model of RNA processing in human mitochondria. Nature. 1981 Apr 9;290(5806):470–4.

42. Montoya J, Christianson T, Levens D, Rabinowitz M, Attardi G. Identification of

initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. Proc Natl Acad Sci USA. 1982 Dec;79(23):7195–9.

43. Chang DD, Clayton DA. Identification of primary transcriptional start sites of mouse mitochondrial DNA: accurate in vitro initiation of both heavy- and light-strand transcripts. Mol Cell Biol. 1986 May;6(5):1446–53.

44. Montoya J, Gaines GL, Attardi G. The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. Cell. 1983 Jan 1;34(1):151–9.

45. Chang DD, Clayton DA. Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. Cell. 1984 Jan 1;36(3):635–43.

46. Fernandez-Silva P, Enriquez JA, Montoya J. Replication and transcription of mammalian mitochondrial DNA. Exp Physiol. 2003 Jan;88(1):41–56.

47. Walberg MW, Clayton DA. In vitro transcription of human mitochondrial DNA. Identification of specific light strand transcripts from the displacement loop region. J Biol Chem. 1983.

48. Gelfand R, Attardi G. Synthesis and turnover of mitochondrial ribonucleic acid in HeLa cells: the mature ribosomal and messenger ribonucleic acid species are metabolically unstable. Mol Cell Biol. 1981 Jun;1(6):497–511.

49. Zollo O, Tiranti V, Sondheimer N. Transcriptional requirements of the distal heavy-strand promoter of mtDNA. Proc Natl Acad Sci USA. 2012 Apr 24;109(17):6508–12.

50. Lodeiro MF, Uchida A, Bestwick M, Moustafa IM, Arnold JJ, Shadel GS, et al. Transcription from the second heavy-strand promoter of human mtDNA is repressed by transcription factor A in vitro. Proc Natl Acad Sci USA. 2012 Apr 24;109(17):6513–8.

51. Yakubovskaya E, Guja KE, Eng ET, Choi WS, Mejia E, Beglov D, et al. Organization of the human mitochondrial transcription initiation complex. Nucleic Acids Res. 2014 Apr;42(6):4100–12.

52. Gaspari M, Larsson N-G, Gustafsson CM. The transcription machinery in mammalian mitochondria. ACTA-BIOENERG. 2004 Jan 1;1659(2):148–52.

53. Sondheimer N, Fang J-K, Polyak E, Falk MJ, Avadhani NG. Leucine-rich pentatricopeptide-repeat containing protein regulates mitochondrial transcription. Biochemistry. 2010 Sep 7;49(35):7467–73.

54. Scarpulla RC. Nuclear control of respiratory chain expression in mammalian cells. J Bioenerg Biomembr. 1997 Apr;29(2):109–19.

55. Asin-Cayuela J, Gustafsson CM. Mitochondrial transcription and its regulation in mammalian cells. Trends Biochem Sci. 2007 Jan 1;32(3):111–7.

56. Schwinghammer K, Cheung ACM, Morozov YI, Agaronyan K, Temiakov D, Cramer P. Structure of human mitochondrial RNA polymerase elongation complex. Nat Struct Mol Biol. 2013 Nov;20(11):1298–303.

57. Wang Y, Bogenhagen DF. Human mitochondrial DNA nucleoids are linked to protein folding machinery and metabolic enzymes at the mitochondrial inner membrane. J Biol Chem. 2006 Sep 1;281(35):25791–802.

58. Lu B, Lee J, Nie X, Li M, Morozov YI, Venkatesh S, et al. Phosphorylation of human TFAM in mitochondria impairs DNA binding and promotes degradation by the AAA+ Lon protease. Mol Cell. 2013 Jan 10;49(1):121–32.

59. Miyakawa I, Sando N, Kawano S, Nakamura S, Kuroiwa T. Isolation of morphologically intact mitochondrial nucleoids from the yeast, Saccharomyces cerevisiae. J Cell Sci. 1987 Nov 1;88 (Pt 4):431–9.

60. Bogenhagen DF, Rousseau D, Burke S. The layered structure of human mitochondrial DNA nucleoids. J Biol Chem. 2008 Feb 8;283(6):3665–75.

61. Kaufman BA, Durisic N, Mativetsky JM, Costantino S, Hancock MA, Grutter P, et al. The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures. Molecular Biology of the Cell. 2007 Sep;18(9):3225–36.

62. Sologub M, Litonin D, Anikin M, Mustaev A, Temiakov D. TFB2 Is a Transient Component of the Catalytic Site of the Human Mitochondrial RNA Polymerase. Cell. 2009 Jan 1;139(5):934–44.

63. Cotney J, Shadel GS. Evidence for an early gene duplication event in the evolution of the mitochondrial transcription factor B family and maintenance of rRNA methyltransferase activity in human mtTFB1 and mtTFB2. J Mol Evol. 2006 Nov;63(5):707–17.

64. Cotney J, Wang Z, Shadel GS. Relative abundance of the human mitochondrial transcription system and distinct roles for h-mtTFB1 and h-mtTFB2 in mitochondrial biogenesis and gene expression. Nucleic Acids Res. 2007 Jun;35(12):4042–54.

65. Guja KE, Venkataraman K, Yakubovskaya E, Shi H, Mejia E, Hambardjieva E, et al. Structural basis for S-adenosylmethionine binding and methyltransferase activity by mitochondrial transcription factor B1. Nucleic Acids Res. 2013 Sep;41(16):7947–59.

66. Seidel-Rogol BL, McCulloch V, Shadel GS. Human mitochondrial transcription factor B1 methylates ribosomal RNA at a conserved stem-loop. Nat Genet. 2003 Jan;33(1):23–4.

67. Litonin D, Sologub M, Shi Y, Savkina M, Anikin M, Falkenberg M, et al. Human mitochondrial transcription revisited: only TFAM and TFB2M are required for transcription of the mitochondrial genes in vitro. J Biol Chem. 2010 Jun 11;285(24):18129–33.

68. Shutt T, Bestwick M, Shadel G. The core human mitochondrial transcription initiation complex: It only takes two to tango. Transcription. 2011 Mar 1;2(2):55–9.

69. Kang D, Kim SH, Hamasaki N. Mitochondrial transcription factor A (TFAM): Roles in maintenance of mtDNA and cellular functions. Mitochondrion. 2007 Jan 1;7(1):39–44.

70. Shutt TE, Lodeiro MF, Cotney J, Cameron CE, Shadel GS. Core human mitochondrial transcription apparatus is a regulated two-component system in vitro. Proc Natl Acad Sci USA. 2010 Jul 6;107(27):12133–8.

71. Shi Y, Dierckx A, Wanrooij PH, Wanrooij S, Larsson N-G, Wilhelmsson LM, et al. Mammalian transcription factor A is a core component of the mitochondrial transcription machinery. Proc Natl Acad Sci USA. 2012;109(41):16510–5.

72. Dubin DT, Montoya J, Timko KD, Attardi G. Sequence analysis and precise mapping of the 3' ends of HeLa cell mitochondrial ribosomal RNAs. Journal of molecular biology. 1982 May;157(1):1–19.

73. Christianson TW, Clayton DA. A tridecamer DNA sequence supports human mitochondrial RNA 3'-end formation in vitro. Mol Cell Biol. 1988 Oct 1;8(10):4502–9.

74. Kruse B, Narasimhan N, Attardi G. Termination of Transcription in Human Mitochondria - Identification and Purification of a Dna-Binding Protein Factor That Promotes Termination. Cell. 1989;58(2):391–7.

75. Daga A, Micol V, Hess D, Aebersold R, Attardi G. Molecular Characterization of the Transcription Termination Factor From Human Mitochondria. J Biol Chem. 1993;268(11):8123–30.

76. Rebelo AP, Williams SL, Moraes CT. In vivo methylation of mtDNA reveals the dynamics of protein-mtDNA interactions. Nucleic Acids Res. 2009 Nov 1;37(20):6701–15.

77. Martínez-Azorín F. The Mitochondrial Ribomotor Hypothesis. IUBMB Life (International Union of Biochemistry and Molecular Biology: Life). 2005 Jan 1;57(1):27–30.

78. Yakubovskaya EE, Mejia EE, Byrnes JJ, Hambardjieva EE, Garcia-Diaz MM. Helix unwinding and base flipping enable human MTERF1 to terminate mitochondrial transcription. Cell. 2010 Jun 11;141(6):982–93.

79. Shang J, Clayton DA. Human mitochondrial transcription termination exhibits RNA polymerase independence and biased bipolarity in vitro. J Biol Chem. 1994 Nov 18;269(46):29112–20.

80. Byrnes J, Garcia-Diaz M. Mitochondrial transcription: How does it end? Transcription 2:1, 32-36; January/February 2011 2011.

81. Guja KE, Garcia-Diaz M. Hitting the brakes: termination of mitochondrial transcription. Biochim Biophys Acta. 2012 Sep 1;1819(9-10):939–47.

82. Terzioglu M, Ruzzenente B, Harmel J, Mourier A, Jemt E, López MD, et al. MTERF1 Binds mtDNA to Prevent Transcriptional Interference at the Light-Strand Promoter but Is Dispensable for rRNA Gene Transcription Regulation. Cell Metab. 2013 Apr 2;17(4):618–26.

83. Asin-Cayuela J, Schwend T, Farge G, Gustafsson CM. The human mitochondrial transcription termination factor (mTERF) is fully active in vitro in the non-phosphorylated form. J Biol Chem. 2005 Jul 8;280(27):25499–505.

84. FernandezSilva P, MartinezAzorin F, MICOL V, ATTARDI G. The human mitochondrial transcription termination factor (mTERF) is a multizipper protein but binds to DNA as a monomer, with evidence pointing to intramolecular leucine zipper interactions. EMBO J. 1997;16(5):1066–79.

85. Shutt TE, Shadel GS. A compendium of human mitochondrial gene expression machinery with links to disease. Environ Mol Mutagen. 2010 Jun 1;51(5):360–79.

86. Park CB, Asin-Cayuela J, Cámara Y, Shi Y, Pellegrini M, Gaspari M, et al. MTERF3 Is a Negative Regulator of Mammalian mtDNA Transcription. Cell. 2007 Jan 1;130(2):273–85.

87. Wredenberg A, Lagouge M, Bratic A, Metodiev MD, Spåhr H, Mourier A, et al. MTERF3 regulates mitochondrial ribosome biogenesis in invertebrates and mammals. PLoS Genet. 2013;9(1):e1003178.

88. Gerber JK, Gögel E, Berger C, Wallisch M, Müller F, Grummt I, et al. Termination of mammalian rDNA replication: polar arrest of replication fork movement by transcription termination factor TTF-I. Cell. 1997 Aug 8;90(3):559–67.

89. Helmrich A, Ballarino M, Tora L. Collisions between Replication and Transcription Complexes Cause Common Fragile Site Instability at the Longest Human Genes. Mol Cell. 2011;44(6):966–77.

90. Omont N, Kepes F. Transcription/replication collisions cause bacterial transcription units to be longer on the leading strand of replication. Bioinformatics. 2004;20(16):2719–25.

91. Hyvärinen AK, Pohjoismäki JLO, Reyes A, Wanrooij S, Yasukawa T, Karhunen PJ, et al. The mitochondrial transcription termination factor mTERF modulates replication pausing in human mitochondrial DNA. Nucleic Acids Res. 2007 Jan 1;35(19):6458–74.

92. Bastia D, Zzaman S, Krings G, Saxena M, Peng X, Greenberg MM. Replication termination mechanism as revealed by Tus-mediated polar arrest of a sliding helicase. Proc Natl Acad Sci USA. 2008 Sep 2;105(35):12831–6.

93. Fernandez-Silva P, Polosa PL, Roberti M, Di Ponzio B, Gadaleta MN, Montoya J, et al.

Sea urchin mtDBP is a two-faced transcription termination factor with a biased polarity depending on the RNA polymerase. Nucleic Acids Res. 2001 Nov 15;29(22):4736–43.

94. Linder T, Park CB, Asin-Cayuela J, Pellegrini M, Larsson N-G, Falkenberg M, et al. A family of putative transcription termination factors shared amongst metazoans and plants. Curr Genet. 2005 Oct;48(4):265–9.

95. Robles P, Micol JL, Quesada V. Unveiling plant mTERF functions. Mol Plant. 2012 Mar 1;5(2):294–6.

96. Babiychuk E, Vandepoele K, Wissing J, Garcia-Diaz M, De Rycke R, Akbari H, et al. Plastid gene expression and plant development require a plastidic protein of the mitochondrial transcription termination factor family. Proc Natl Acad Sci USA. 2011;108(16):6674–9.

97. Roberti M, Polosa PL, Bruni F, Deceglie S, Gadaleta MN, Cantatore P. MTERF factors: a multifunction protein family. BioMolecular Concepts. 2010;1(2).

98. Roberti M, Polosa PL, Bruni F, Manzari C, Deceglie S, Gadaleta MN, et al. The MTERF family proteins: mitochondrial transcription regulators and beyond. Biochim Biophys Acta. 2009 May;1787(5):303–11.

99. Pellegrini M, Asin-Cayuela J, Erdjument-Bromage H, Tempst P, Larsson N-G, Gustafsson CM. MTERF2 is a nucleoid component in mammalian mitochondria. ACTA-BIOENERG. 2009 Jan 1;1787(5):296–302.

100. Wenz T, Luca C, Torraco A, Moraes CT. mTERF2 Regulates Oxidative Phosphorylation by Modulating mtDNA Transcription. Cell Metab. 2009 Jan 1;9(6):499–511.

101. Hyvärinen AK, Pohjoismäki JLO, Holt IJ, Jacobs HT. Overexpression of MTERFD1 or MTERFD3 impairs the completion of mitochondrial DNA replication. Mol Biol Rep. 2011 Feb 1;38(2):1321–8.

102. Laffel L. Ketone bodies: a review of physiology, pathophysiology and application of monitoring to diabetes. Diabetes Metab Res Rev. 1999;15(6):412–26.

103. Cámara Y, Asin-Cayuela J, Park CB, Metodiev MD, Shi Y, Ruzzenente B, et al. MTERF4 Regulates Translation by Targeting the Methyltransferase NSUN4 to the Mammalian Mitochondrial Ribosome. Cell Metab. 2011 Jan 1;13(5):527–39.

104. Yakubovskaya E, Guja KE, Mejia E, Castano S, Hambardjieva E, Choi WS, et al. Structure of the Essential MTERF4:NSUN4 Protein Complex Reveals How an MTERF Protein Collaborates to Facilitate rRNA Modification. Structure. 2012 Nov 7;20(11):1940–7.

105. Jiménez-Menéndez N, Fernández-Millán P, Rubio-Cosials A, Arnan C, Montoya J, Jacobs HT, et al. Human mitochondrial mTERF wraps around DNA through a left-handed superhelical tandem repeat. Nat Struct Mol Biol. 2010 Jul;17(7):891–3.

106. Spåhr H, Habermann B, Gustafsson CM, Larsson N-G, Hallberg BM. Structure of the human MTERF4-NSUN4 protein complex that regulates mitochondrial ribosome biogenesis. Proc Natl Acad Sci USA. 2012 Sep 18;109(38):15253–8.

107. Lu G, Dolgner SJ, Hall TMT. Understanding and engineering RNA sequence specificity of PUF proteins. Curr Opin Struct Biol. 2009 Feb;19(1):110–5.

108. Florentz C, Sohm B, Tryoen-T th P, P tz J, Sissler M. Human mitochondrial tRNAs in health and disease. Cellular and Molecular Life Sciences (CMLS). Birkhäuser-Verlag; 2003 Jul 1;60(7):1356–75.

109. Lehtonen MS, Uimonen S, Hassinen IE, Majamaa K. Frequency of mitochondrial DNA point mutations among patients with familial sensorineural hearing impairment. Eur J Hum Genet. 2000 Apr 1;8(4):315–8.

110. Elliott HR, Samuels DC, Eden JA, Relton CL. Pathogenic Mitochondrial DNA Mutations Are Common in the General Population. The American Journal of Human Genetics *83,* 254-260, August 8, 2008.

111. Chomyn A, Martinuzzi A, Yoneda M, Daga A, Hurko O, Johns D, et al. Melas Mutation in Mtdna Binding-Site for Transcription Termination Factor Causes Defects in Protein-Synthesis and in Respiration but No Change in Levels of Upstream and Downstream Mature Transcripts. Proc Natl Acad Sci USA. 1992;89(10):4221–5.

112. Chomyn A, Enriquez JA, Micol V. The MELAS syndrome-associated human mitochondrial tRNALeu (UUR) mutation causes aminoacylation deficiency and concomitant reduced association of mRNA with ribosomes. Journal of Biological Chemistry, April 14,2000.

113. Majamaa K, Turkka J, Kärppä M, Winqvist S, Hassinen IE. The common MELAS mutation A3243G in mitochondrial DNA among young patients with an occipital brain infarct. Neurology. 1997 Nov;49(5):1331–4.

114. Zeviani M, Moraes CT, DiMauro S, Nakase H, Bonilla E, Schon EA, et al. Deletions of mitochondrial DNA in Kearns-Sayre syndrome. Neurology. 1998 Dec 1;51(6):1525–5.

115. Chomyn A. The Mitochondrial Myopathy, Encephalopathy, Lactic Acidosis, and Stroke-like Episode Syndrome-associated Human Mitochondrial tRNALeu(UUR) Mutation Causes Aminoacylation Deficiency and Concomitant Reduced Association of mRNA with Ribosomes. J Biol Chem. 2000 Apr 14;275(25):19198–209.

116. Martin M, Cho J, Cesare AJ, Griffith JD, Attardi G. Termination Factor-Mediated DNA Loop between Termination and Initiation Sites Drives Mitochondrial rRNA Synthesis. Cell. 2005 Jan 1;123(7):1227–40.

117. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010 Feb 1;66(Pt 2):125–32.

118. Vonrhein C, Flensburg C, Keller P, Sharff A, Smart O, Paciorek W, et al. Data processing and analysis with the autoPROC toolbox. Acta Crystallogr D Biol

Crystallogr. 2011 Apr;67(Pt 4):293–302.

119. Vagin A, Teplyakov A. Molecular replacement with MOLREP. Acta Crystallogr D Biol Crystallogr. 2010 Jan 1;66(Pt 1):22–5.

120. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011 Apr 1;67(Pt 4):355–67.

121. Nicholls RA, Long F, Murshudov GN. Low-resolution refinement tools in REFMAC5. Acta Crystallogr D Biol Crystallogr. 2012 Apr 1;68(Pt 4):404–17.

122. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr. 2011 Apr 1;67(Pt 4):235–42.

123. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004 Nov 26;60(12):2126–32.

124. Chen VB, Arendall WB3, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010 ed. 2010 Jan;66(Pt 1):12–21.

125. Darden TA, Cheatham III TE, Simmerling CL, Wang J, Duke RE, Luo R, et al. AMBER12. University of California, San Francisco. 2012.

126. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys. 1977;23(3):327–41.

127. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Molecular-Dynamics with Coupling to an External Bath. J Chem Phys. 1984;81(8):3684–90.

128. Le Grand S, Götz AW, Walker RC. SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. Computer Physics Communications. 2013.

129. Gohlke H, Kiel C, Case DA. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. Journal of molecular biology. 2003rd ed. 2003 Jul 18;330(4):891–913.

130. Nam S-C, Kang C. DNA light-strand preferential recognition of human mitochondria transcription termination factor mTERF. J Biochem Mol Biol. 2005 Nov 30;38(6):690–4.

131. Helm MM, Florentz CC, Chomyn AA, Attardi GG. Search for differences in post-transcriptional modification patterns of mitochondrial DNA-encoded wild-type and mutant human tRNALys and tRNALeu(UUR). Nucleic Acids Res. 1999 Feb 1;27(3):756–63.

132. Hess JF, Parisi MA, Bennett JL, Clayton DA. Impairment of mitochondrial transcription termination by a point mutation associated with the MELAS subgroup of mitochondrial encephalomyopathies. Nature. 1991 May 16;351(6323):236–9.