UNIVERSITAT DE BARCELONA

Structure and biophysical studies of mitochondrial Transcription Factor A in complex with DNA

Anna Cuppari

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

Anna Cuppari

Structure and biophysical studies of mitochondrial Transcription Factor A in complex with DNA

Anna Cuppari, 2016

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

PROGRAMA DE DOCTORAT
Biotecnología

Structure and biophysical studies of mitochondrial Transcription Factor A in complex with DNA

Memòria presentada per Anna Cuppari per optar al títol de doctor per la universitat de Barcelona

Directora de la tesis Maria Solà i Vilarrubias

Doctoranda Anna Cuppari

Tutora de la tesis Josefa Badia Palacín

Anna Cuppari, 2016

"One never notices what has been done; one can only see what remains to be done."

Marie Curie, physicist and chemist (1867-1934)

INDEX

1. ABSTRACT

The mitochondrial transcription factor A, TFAM, has a dual function in the organelle: it activates mitochondrial DNA transcription by binding to the HSP and LSP promoters, while in higher concentrations compacts the mtDNA. In this thesis the mechanism of complex formation between the mitochondrial transcription factor A (TFAM) and its cognate DNA binding sequences is analysed. TFAM is a DNA binding protein that belongs to the HMG-box family. Previous crystallographic works have shown that, by its two HMG-boxes and the intervening linker, TFAM binds to the DNA minor groove of mitochondrial DNA promoters imposing severe DNA distortions. These include two sharp 90-degree kinks that bend the cognate DNA into a U-turn. We here present the crystallographic structure of TFAM in complex with site Y, which together with site X are protein binding sites alternative to the promoter binding regions at the control region of mitochondrial DNA (mtDNA). The structure of the TFAM/site Y complex shows the two HMG-box domains (HMG-box1 and 2) organized in an "L"-shape fold that bends the contacted DNA by 90 degrees. Each HMG-box domain inserts a leucine, Leu 58 from HMG-box1 and Leu182 from HMG-box2, into a base-pair step from respective DNA contacted regions. The two DNA steps are separated by a DNA helix turn. Each insertion disrupts the DNA stacking and, together with additional interactions, facilitates the 90º DNA bending, the two bends resulting in the U-turn conformation. A structural comparison between available TFAM/DNA complexes shows that the linker between HMG-domains is instrumental for the protein to adapt to a conformation variability induced by the different DNA sequences. In addition, while all other crystal structures are unambiguous in the assigned DNA sequence, TFAM/site Y electron density maps indicated a surprising DNA disorder that suggested to trace the DNA in an alternative, not predicted, orientation.

Thus in order to better characterize the binding mechanism of TFAM to the DNA and the role of the DNA properties in this process, we further studied the TFAM/site Y, TFAM/site X and TFAM/LSP complexes by molecular dynamics (MD) simulations,

isothermal titration calorimetry (ITC) and electrophoretic mobility shift assays (EMSA). All these techniques showed a recurrent result, which is that TFAM has a clear preference in binding and bending site Y over site X and LSP. The three DNAs present intrinsic distortions that facilitate binding, which occurs by a mechanism in all cases endothermic and spontaneous and TFAM presents similar affinities to all of them. However, site Y is intrinsically more rigid but easier to distort into the shape found in the crystal, it competes better for TFAM binding, and the enthalpy and entropy of binding are much higher than for the other two sequences. These results suggest a specific binding and bending mechanism significantly dependent on the DNA sequence.

Finally, by multi-angle laser light scattering (MALLS) and analytical ultracentrifugation the multimerization ability of TFAM detected by EMSA and size exclusion chromatography was analysed. The results indicate multimerization of the protein either alone or on the DNA in a cooperative manner at increased complex concentrations, which is consistent with the alternative function of TFAM as an mtDNA packaging protein. Altogether, our results suggest that the DNA sequence properties mediate TFAM binding, involving either specific interactions at the mtDNA control region, or non-specific contacts during mtDNA compaction. For this latter, the regulation of TFAM binding exerted by the DNA sequence might be combined with regulation of protein multimerization processes, all together determining mtDNA compaction, which is essential for cell life.

RESUMEN

Este trabajo de tesis doctoral está centrado en el análisis del mecanismo de unión del factor A de transcripción mitocondrial (TFAM) con sus secuencias de reconocimiento en la región control del ADN mitocondrial (mtADN). En la mitocondria TFAM está implicado en dos procesos fundamentales: la regulación de la trascripción del mtADN, cuando está unido a las secuencias promotoras del filamento ligero y pesado (HSP y LSP), y la compactación del mismo ADN cuando está presente en alta concentración. TFAM pertenece a la familia de los HMG-box y está constituida por dos dominios HMG conectados por un "linker" de 20 residuos. En este trabajo se presenta la estructura cristalográfica de TFAM en complejo con su sitio de reconocimiento alternativo a los promotores, site Y. Desde el análisis de la estructura se ha evidenciado que TFAM presenta el mismo plegamiento observado también cuando está en complejo con LSP, HSP, ADN no específico (nsADN) y su otro sitio de unión site X. Además en todos estos complejos el ADN resulta doblado 180° por medio de dos inserciones mediadas por LEU58 y 182, cada una responsable de un "kink" de 90°. La diferencia principal entre todas las estructuras se observa a nivel del linker que presenta una desviación en respuesta a las diferentes propiedades de los ADNs que contacta.

Para caracterizar mejor el mecanismo de unión de TFAM con sus secuencias de reconocimiento en la región control del mtADN (LSP, site Y and site X), se realizaron diferentes análisis de tipo biofísico y bioquímico. La flexibilidad de estas secuencias se estudió primero por dinámica molecular. Estudios de "isothermal titration calorimetry" y "electrophoresis mobility shift assays" permitieron evidenciar que también si TFAM tiene el mismo mecanismo de unión y la misma afinidad por las tres secuencias, la cinética de formación de los complejos parece ser diferente. Para el análisis de la estequiometria de la unión de TFAM a los diferentes ADN fueron empleadas las técnicas de "multi angle laser light scattering" y "analytical ultracentrifugation". Estos estudios evidenciaron la tendencia de TFAM de multimerizar, en presencia y ausencia de ADN, en respuesta a aumento de su concentración.

2. OBJECTIVES:

The global aim of this work was to understand the binding mechanism of TFAM to site Y and site X and compare it with other TFAM binding sequences. To achieve this goal the following objectives were established:

- Crystallization of TFAM in complex with site Y and site X separately and in a continuous DNA.
- Crystal structure determination of the TFAM/site Y complex.
- Analysis of TFAM binding mechanism to DNA and of the role of DNA sequences in complex formation and differential binding by molecular dynamics (MD) simulations, electrophoretic mobility shift assays (EMSA) and isothermal titration calorimetry (ITC).
- Analysis of the oligomeric state of TFAM alone or in complex with DNA by multi-angle laser light scattering (MALLS) and analytical ultracentrifugation.

3. INTRODUCTION

3. INTRODUCTION

3.1 The mitochondrion.

The name "Mitochondrion" comes from the Greek "μίτος" (mitos), thread, and "χονδρίον" (chondrion), granule. This name was coined by Benda in 1898 referring to granulose structures observed in cells undergoing spermatogenesis. Some years before, Altmann in 1890 was the pioneer in discovering and recognizing these structures as organelles , which he called "bioblasts"(Ernster and Schatz 1981). The mitochondrion is a double membrane organelle with its own DNA, found in eukaryotic organisms, characterized by a high degree of variability in function and features depending on cell type and cell metabolism. In general terms, mitochondria have a size between 0.75 and 3 μm in diameter but can vary considerable in both size and structure (Figure_1). This organelle is the primary source of chemical energy inside the cell, which supplies in the form of adenosine triphosphate (ATP) through mitochondrial respiration (Friedman and Nunnari 2014). There is no direct evidence on how this organelle originated inside the cell, and different hypotheses about its origin and evolution have been formulated. Some of these are presented below.

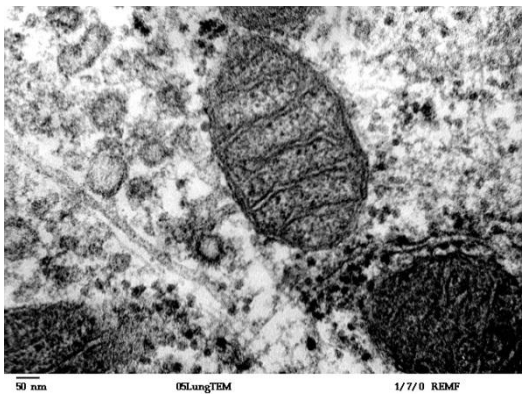


Figure_1. Micrography of a mitochondria obtained by transmission electron microscopy. Image from a thin section of mammalian lung tissue. This image is taken from: http://www.broadinstitute.org/education/glossary/mitochondria.

3.1.1 Mitochondrion origin and evolution: phylogenetic hypotheses.

Phylogenetic analysis of the mitochondrial proteome composition is a starting point to explore the mitochondrion origin and its implication in life evolution. Biochemical, molecular and bioinformatic sudies strongly suggest that a core of conserved genes that encode for mitochondrial proteins are related to genes from Alpha-proteobacteria, pointing to this class of bacteria as precursors of the organelle. This supports the hypothesis that the original mitochondrion was an Alpha-proteobacteria engulfed by another cell by endosymbiosis. However, from the total mitochondrial proteome, only 10%-20% of it can be confidently traced to Alpha-proteobacteria. The remaining 80%-90% non-Alpha-proteobacterial component seems to arise from eukaryotes or from other eubacteria (Gray, Burger et al. 2001; Gray 2014; Gray 2015). Therefore, so far various endosymbiotic models have been hypotesized that can only partially explain the high level of divergence of the mitochondrial proteome. These models can be divided into two groups that describe two different endosymbiotic scenarios: the "archezoan" and the "symbiogenesis" scenarios (Koonin 2010). The archezoan hypothesis assumes the existence of an ancestral amitochondriate eukaryotic host that engulfed an Alpha-proteobacterium, then converted into the mitochondrion. In contrast, in the symbiogenesis hypothesis, a cellular fusion occurred between an Alpha-proteobacterium and an Archaea bacterium followed by the evolution of the nucleus and compartmentalization of the eukaryotic cell. The Alpha-proteobacterium became then the mitochondrion (Koonin 2010). A third hypothesis was proposed by Grey in 2012: "the pre-endosymbiont hypothesis". Grey postulated the existence of a "pre-mitochondrion" inside the ancestral eukaryotic cell that had its own proteome but was not able to generate energy. Such ancestral eukaryotic cell phagocyted an aerobic Alpha-proteobacterium able to produce energy. The symbiont, once inside the host cell, incorporated the pre-mitochondrion proteome to generate the mitochondrion (Gray 2014). Apart from the multiple theories on the mitochondrion origin, it is well established that the core of the mitochondrial proteome was the same in both the mitochondrial common ancestor and in the contemporary mitochondria (Gray 2012). This suggests that the diversity of both the mitochondrial proteome and genome probably derives from function specialization of eukaryotic cells, involving both loss

and gain of protein components (Gray 2012; Gray 2014). Based on the theory of endosymbiosis, Lane and Martin in 2010 postulated a linear correlation between energy production by the mitochondrion and eukaryote complexity. In eukaryotic cells, the number of genes and the complexity of the genome correlate with higher availability of energy in the cell (Lane and Martin 2010).

3.2 Mitochondrion structure and function.

3.2.1 Fundamental components of the mitochondrion.

A mitochondrion is composed by two phospholipidic bilayer membranes, named "outer membrane" and "inner membrane", with distinct function and protein/lipid ratio composition. The existence of these double membranes results in the compartmentalization of the mitochondrion into the "intermembrane space", between the two membranes, and the "matrix space" surrounded by the inner membrane. Infolding of the inner membrane towards the matrix forms the "cristae" structures, which are present in variable amount in response to the energetic cell needs (Figure_2). A detailed analysis of the mitochondrial compartments shows a tight correlation between properties and functions.

The outer membrane, and this is maybe related to its endosymbiotic origin, has the same protein/phospholipid ratio as found at the cell membrane. One of its most abundant protein type is the integral membrane protein channel "porine" (Figure_2).

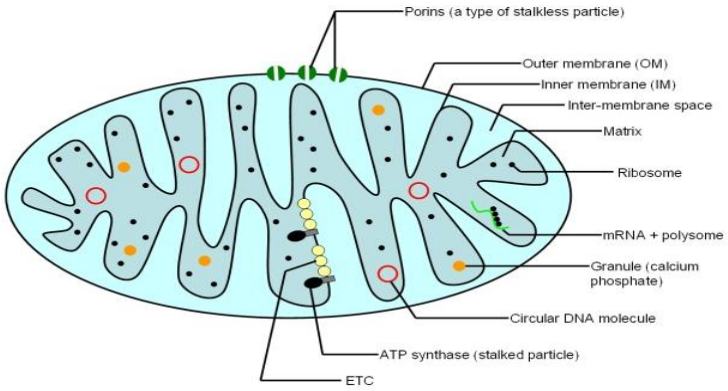


Figure_2. Principal features of the mitochondrion.
The fundamental mitochondrion components are schematically represented.
Modified from: http://cronodon.com/BioTech/Respiration.html.

Apart from these, there are protein transporters, the most important one being the "translocase of the outer membrane" (TOM). The TOM transporter binds specific mitochondrial leading sequences allowing the entrance of exogenous big proteins (>5 kDa) into the intermembrane space. Channels and transporters determine the chemical composition of the intermembrane space. In this regard, this organelle compartment presents the same ion concentration but a different set of proteins with respect to the cytosol. The outer membrane also presents enzymes involved in diverse activities such as elongation of fatty acids, oxidation of epinephrine and the degradation of the tryptophan. Association of the outer membrane with the endoplasmic reticulum (ER) forms the mitochondria-associated ER-membrane (MAM). MAM is of fundamental importance in the regulation of calcium homeostasis and in lipid transfer from the endoplasmatic reticulum to the mitochondrion.

In contrast to the outer membrane, the inner membrane is highly impermeable to ions and small molecules, so it has a specific set of transporters. No porins are present to allow ion diffusion. the ions pass through the "sodium-calcium exchanger" channel (NCX). With this, the mitochondrion maintains a membrane potential across the inner membrane that is necessary for ATP synthesis and mitochondrion and cell homeostasis (see below). Apart from the protein transporter other enzymes are also located in this internal membrane: enzymes of the elecron transport chain and the ATP synthase complex (all togeher shaping the oxidative phosphorylation pathway), specific metabolite transporters and mitochondrial fusion and fission proteins. Therefore, the inner membrane has a protein/phospholipid ratio higher than the outer membrane.
Finlly, most of mitochondrial proteins (around 1000) are located inside the matrix space. In this space important biochemical pathways are found, such as pyruvate decarboxylation, the essential fatty acids cycle and the citric acid cycle (Krebs cycle). Apart from the components of these pathways, other enzymes, but also ribosomes, rRNAs, tRNAs and several copies of mitochondrial DNA (mtDNA) together with mtDNA-processing proteins imported from the nucleus are also present (Figure_2) (Alberts B., Johnson A. et al. 2002).

3.2.2. Functions of mitochondria

The mitochondrion has a crucial role in the last steps of cell catabolism, which results in energy production. These include the glycolysis pathway, which takes place in the cytosol and renders pyruvate. Pyruvate is subsequently imported to mitochondria and converted, by pyruvate decarboxylation, to acetyl-CoA. In the mitochondrial matrix Acetyl-CoA enters the citric acid cycle, which generates the substrates for the electron transport system (ETS) and oxidative phosporylation (OXPHOS, or cellular respiration) (Figure_3).

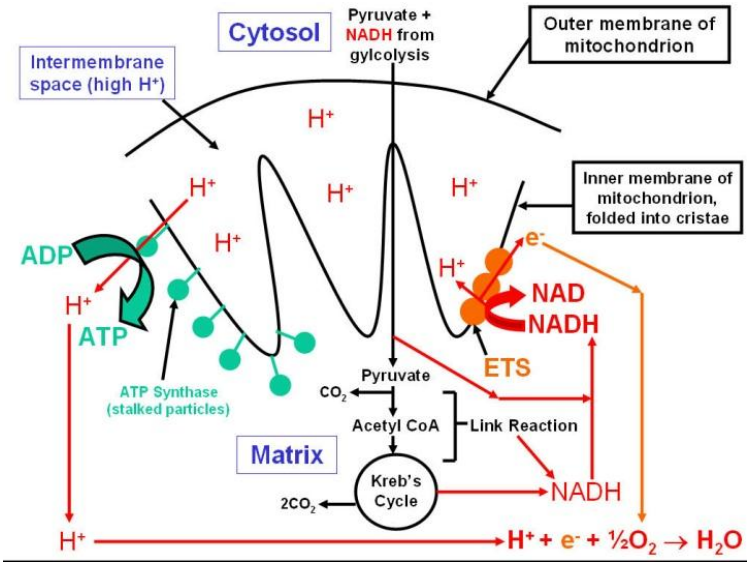


Figure_3. Schematic representation of energy production inside the mitochondrion.
This image was taken from: http://thegreatestgarden.com/do-plant-cells-have-mitochondria.

Thus, during glycolisis, pyruvate processing and Krebs cycle the cofactors nicotinamide adenine dinucleotide (NAD^+) and flavine adenine dinucleotide (FAD) are reduced to NADH and $FADH_2$, respectively. NADH and $FADH_2$ are substrates for the ETS complexes, which are located at the inner membrane. These consist of Complex I (NADH coenzyme Q reductase;) that accepts electrons from the electron carrier NADH. Complex I transfers the electrons to coenzyme Q (ubiquinone). At this step NADH converts to NAD^+ and one H^+ is released to the matrix. In addition, coenzyme Q also receives electrons from complex II (succinate dehydrogenase), which obtains H

atoms directly from the Krebs cycle using $FADH_2$ as a cofactor. At this step FADH2 transfers the electrons and converts to FAD releasing $2H^+$ to the matrix. Coenzyme Q passes the electrons to complex III (cytochrome bc1 complex), which in turn transfers them to cytochrome c (Cytc). Cytc passes electrons to Complex IV (cytochrome c oxidase), which uses the electrons and H^+ from the matrix to reduce molecular oxygen to water. In addition, Complex I, III and IV are proton pumps that couple electron transfer with proton pumping through the inner membrane towards the intermembrane space, generating a H^+ gradient or differential potential across the internal membrane. This promotes an protonmotive force by which the protons return back to the matrix through the ATP synthase (F_1F_0-ATPase or Complex V which together with the ETC forms the OXPHOS pathway). The ATP synthase produces ATP from ADP and Pi (Figure_3). The total amount of ATP obtained from complete glucose degradation throughout the glycolysis, Krebs cycle and oxidative phosphorylation pathways is of 36 molecules (Alberts B., Johnson A. et al. 2002).

Apart from these functions the mitochondrion is also implicated in the storage of calcium ions, programmed cell death (Green, John et al. 1998), the synthesis of certain heme groups (Oh-hama 1997), the production of steroids (Rossier 2006) and in hormonal signaling (Klinge 2008).

3.3 Mitochondria related diseases.

As explained above, the mitochondrion is the central player in energy production inside the cell. It is able to produce about 90% of cellular energy, regulates cellular redox status by producing ROS species, is involved in the Ca^{2+} homeostasis, synthesizes and degrades high energy biochemical intermediates, and it is responsible of apoptosis activation. All the known mitochondrial related diseases are associated with an alteration of mitochondrial bioenergetics regulation (Wallace 2013). This alteration is generally correlated with three factors: mutation of the mtDNA sequence, mutation of nuclear genes that codify for mitochondrial proteins, or variation in calories amount and in the caloric demands of the organism. Inside the organism different tissues have different energy needs, thus energy deficiency can result in

tissue-specific symptoms. For example the brain is the organ most sensitive to energy diminution. Other high-energy demand tissues are: heart, muscle (Wang, Xu et al. 2016), kidney and the endocrine system. Therefore, impairment of the mitochondrial activity can affect different parts of the body. In addition, it can alter with different intensity the different tissues and organs, and can appear at different ages. Such a high variability complicates extraordinarily the diagnosis of the associated health problem, which is often in the form of complex syndromes. In Figure_4 there is a schematic representation of the source of most common mitochondrial diseases correlated with alterations at the OXPHOS level, many of which are related with mtDNA impairing mutations (see MITOMAP) (Brandon, Lott et al. 2005).

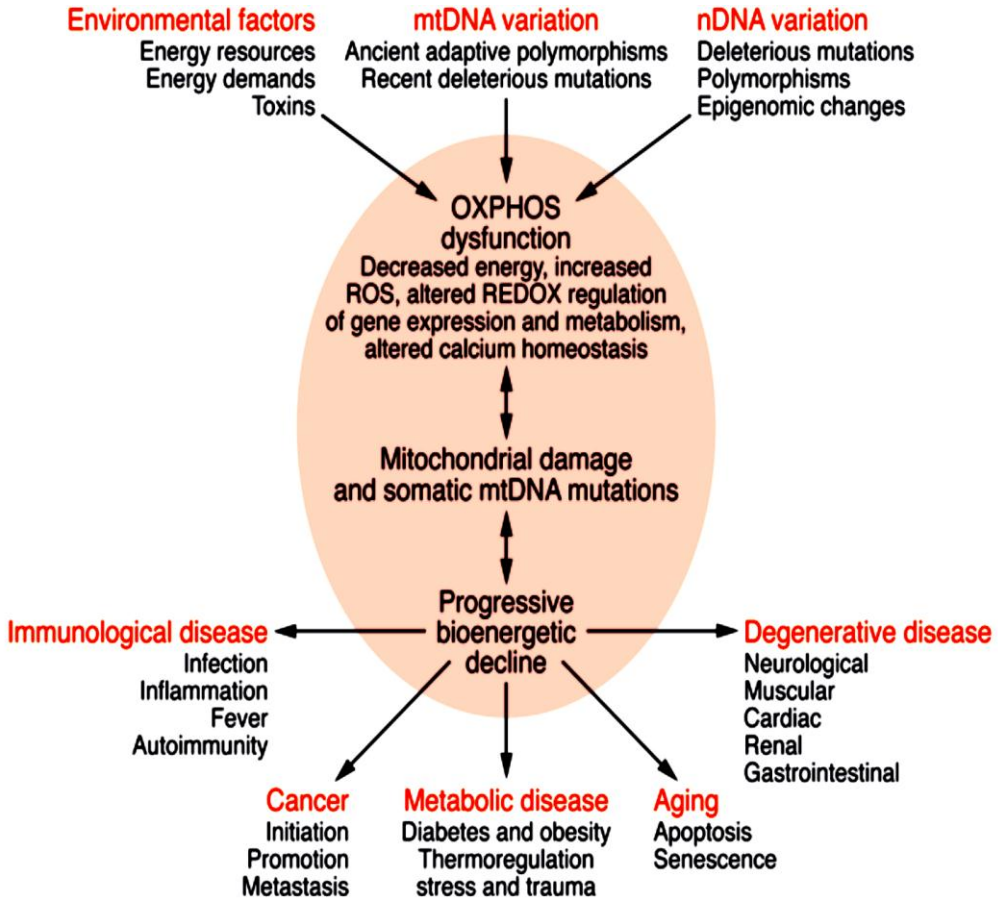


Figure_4. Most common mitochondrial diseases.
Alterations at the OXPHOS level determine progressive accumulation of somatic mtDNA mutations which account for immunological diseases, cancer, metabolic diseases, aging, degenerative diseases. Image taken from Wallace, 2013(Wallace 2013).

3.4 Human mitochondrial DNA metabolism and organization

3.4.1 Features of human mitochondrial DNA

During evolution the mitochondrial genome was subjected to extensive loss of most of its genes, which were partly relocated to the nucleus of the host cell. Such endosymbiosis gene transfer (EGT) to the nucleus could involve an entire mitochondrial gene or a portion of it. These events of EGT were the main source of mitochondrial gene variability among eukaryotes (Burger, Gray et al. 2003).

Regarding the mammalian mitochondrial genome, it appears to be highly conserved inside the mammalian lineage (Wolstenholme 1992). The human mitochondrial DNA (h-mtDNA) was the first genome completely sequenced (Anderson 1981). H-mtDNA is a circular double stranded DNA of 16,569 bp. The two strands present a different G/C content and show different density in denaturing caesium-chloride gradients. Due to this, they are named "heavy" (H) (guanine rich) and "light" (L) (cytosine rich) strand, respectively (Kasamatsu, Grossman et al. 1974). Both strands encode for proteins involved in the ETS and the rRNAs and tRNAs required for their synthesis. In particular, the heavy strand contains most of the coding sequences comprising 2 rRNAs, 14 tRNAs and 12 subunits of the ETS. In contrast, the light strand codifies for 8 tRNAs and for 1 protein also belonging to the ETS (Anderson 1981; Macreadie, Novitski et al. 1983; Chomyn, Cleeter et al. 1986) (Figure_4). One important feature of h-mtDNA is that it is a compact genome without any intron. MtDNA contains few non-coding regions. The major non-coding sequence corresponds to the mitochondrial control region (nt 576-16024), which clusters most of the mtDNA regulatory elements. This region is the most polymorphic fragment of the human mtDNA genome, the polymorphisms being clustered at two hypervariable regions (HVR1 and 2). In the revised Cambridge Reference Sequence of the human mitogenome, the most variable sites of HVR1 are at 16024-16383 positions (this subsequence is called HVR-I), while the most variable sites of HVR2 are at 57-372 (HVR-II) and 438-574 (HVR-III) positions. Inside HVR2, three Conserved Sequence Blocks are found, CSBI (nt 213-236), CSBII (nt 300-315) and CSBIII (nt 363-346) (Walberg and Clayton 1981) (Figure_4). These blocks are different from each other but each one is conserved throughout species (Sbisa, Tanzariello et al. 1997). The mtDNA control region

contains also the origin of replication of the heavy strand (O_H), and the promoters for transcription activation for both strands (HSP and LSP), which show opposite directions (Figure_5). In addition, a Termination-Associated Sequence (TAS) is found at the 3' end of the control region (Figure_5). Furthermore, a 7S R-loop superstructure is founded downstream LSP promoter derived from transcription premature termination event where the nascent RNA remains stably associated with template DNA (Gustafsson, Falkenberg et al. 2016).

Additional mitochondrial non-coding regions include the origin of replication of the light strand, at 2/3 genome distance from the control region and much shorter. Another difference with the nuclear DNA is that intergenic tracts are not present and some of the protein genes are overlapped. Genome sequencing showed that the mitochondrial genetic code is different from the standard genetic code (for example: TGA codon in h-mtDNA codifies for a Trp whereas is a stop codon in the nucleus). In addition, a simplified decoding system allows reading all codons using less than 32 tRNA (Anderson 1981). In this way 22 tRNAs are sufficient to decode all the mtDNA codons(Bonitz, Berlani et al. 1980).

In 1971 Kasamatsu et al. discovered, by electron microscopy, the existence of mtDNA that was constituted by three strands (Kasamatsu, Robberson et al. 1971). In 1978 Brown et al. found that these three strands result from a premature replication stop of the heavy strand that includes the sequence from O_H to TAS, generating a 7S strand stably bound to the template (Brown, Shine et al. 1978). The presence of this new strand creates a permanent displacement of the parental heavy ones. This three-strand region was called Displacement loop (D-loop) (Figure_5).

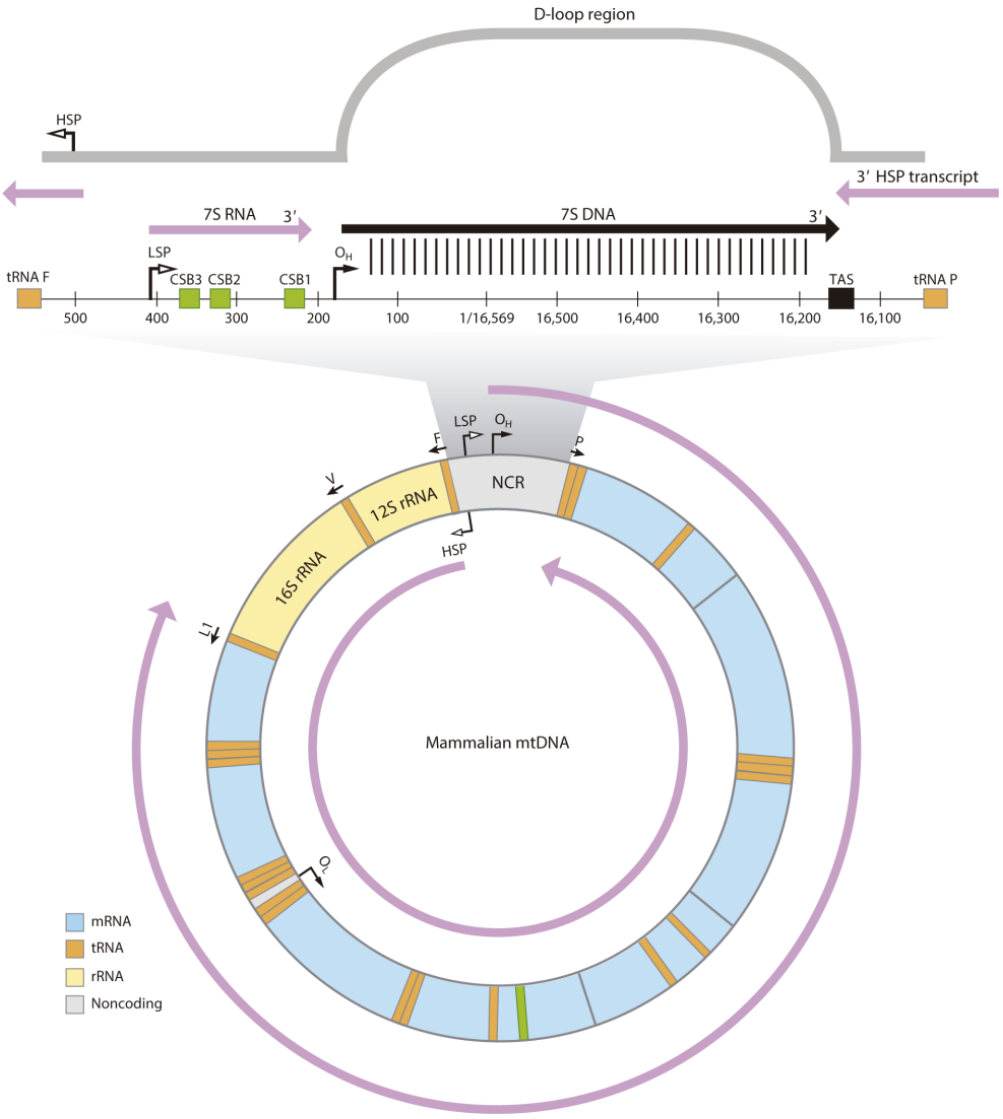


Figure_5. Map of human mitochondrial genome.
An enlarged view of the mitochondrial non coding region (NCR) together with a schematic representation of the gene disposition on heavy and light strands is presented. The 7S DNA within the control region and the 7S RNA from premature LSP transcription termination are also illustrated. The Termination associated sequence (TAS) is indicated at the 3'end of the D-loop. Full-length transcript from both HSP and LSP are indicated by violet arrows. O_H and O_L are the replication origins of the heavy and light strand, respectively. Note that O_L is at 2/3 genome distance of the NCR. CSBs are the conserved sequence blocks. Image taken from Gustafsson et al, 2016(Gustafsson, Falkenberg et al. 2016).

The triplex D-loop length is species-specific and is characterized by a huge variability at the 5' and 3' ends among the different vertebrates. However all vertebrates show that the 3' D-loop ends always map downstream TAS, suggesting that the DNA template sequence regulates the frequency of premature termination of the heavy strand replication (Shadel and Clayton 1997). The fact that the D-loop structure is

conserved in vertebrates suggested crucial roles in mtDNA metabolism. Extensive research demonstrated that the D-loop is involved in mtDNA maintenance, transcriptional regulation and correct segregation during mitochondrial division (King and Low 1987; Annex and Williams 1990; He, Mao et al. 2007).

3.4.2 Transcription promoter regions of human mitochondrial DNA.

The transcription initiation sites of both the light and heavy strands are found at the h-mtDNA control region. These are the regulatory sequences best characterized. Different techniques such as 5'mRNA labelling and nuclease S1 protection experiments, site directed mutagenesis and inter promoter sequences substitution of cloned fragments in run off transcription assays, footprinting analysis, are some of the techniques used to date to analyse these regions. The results obtained from these experiments agree with the conclusion that within the control region there are two major transcription initiation (IT) sites for the heavy and light strands, IT_{H1} and IT_{L} respectively, separated by 150bp (Montoya, Christianson et al. 1982; Yoza and Bogenhagen 1984; Hixson and Clayton 1985; Topper and Clayton 1989). These two starting sites correspond to nucleotides 407+/-1 at the light strand promoter (LSP) and 561+/-1 at the heavy strand promoter 1 (HSP1) (Figure_5). Close to these sites a consensus motif of 15bp was also detected: 5'- CANACC (G) CC (A) AAAGAYA -3' (N stands for any nucleotide, Y for pyrimidine) (Chang and Clayton 1984; Hixson and Clayton 1985). In addition to these transcription initiation sites, a second site for H-strand transcription initiation, IT_{H2}, was detected upstream to the 12S rRNA gene, around nucleotide 638 (Figure_6). IT_{H2} region doesn't show high similarity with the 15bp consensus sequence observed for the other two promoter regions, and it seems to be used less frequently than IT_{H1} (Montoya, Christianson et al. 1982; Chang and Clayton 1984; Yoza and Bogenhagen 1984; Martin, Cho et al. 2005). In addition, *in vitro* mapping of IT_{H2} differs from the *in vivo* one (Montoya, Christianson et al. 1982; Lodeiro, Uchida et al. 2012). IT_{H2} codifies for a long polycistronic mRNA that included 2 rRNA, 12 respiratory chain polypeptides and 13 tRNAs (Figure_6), with a transcription termination located inside an A/T reach region near the $tRNA^{Pro}$ gene

(Figure_5), named D-term region (Camasamudram, Fang et al. 2003). It was also observed that the transcription from IT_{H1} (Figure_6) generates a premature terminated mRNA that includes the $tRNA^{Phe}$ and $tRNA^{Val}$ and the 12S and 16S rRNAs (Figure_5). This premature termination was mediated by the mitochondrial transcription termination factor MTERF1 (Kruse, Narasimhan et al. 1989; Falkenberg, Larsson et al. 2007). On the other hand, transcripts initiated from the LSP are frequently terminated at CSB1, creating the 7S RNA. Premature transcription termination also occurs at CSB2 sequence, generating primers for heavy strand replication. Full-length LSP transcript termination occurs just downstream of the 16S rRNA (Gustafsson, Falkenberg et al. 2016) (Figure_5 and Figure_6).

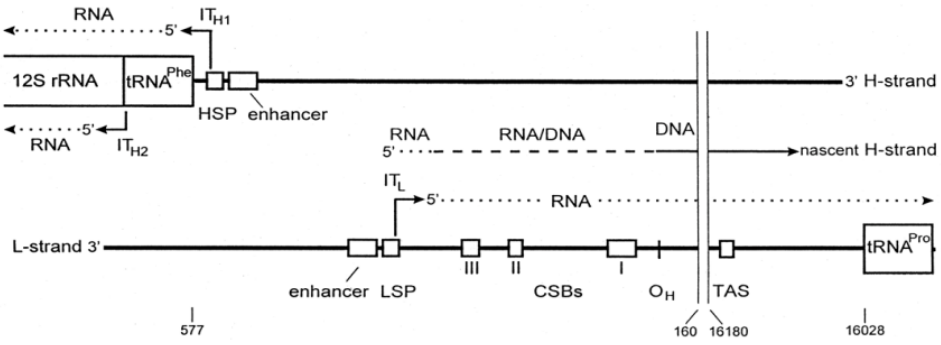


Figure_6. Schematic representation of D-loop regulatory region.
The two heavy strand transcription initiation sites, IT_{H1} and IT_{H2} are shown, together with the transcription initiation site at the light strand (IT_L). Both IT_H and IT_L mapped inside the respective heavy and light strand promoters (H/LSP). The new synthetized RNA is indicated by dots. Dashed lines indicate the switching point between the transcription and replication. Enhancer elements, conserved sequence blocks (CSBs), termination associated sequences (TAS) and heavy strand replication origin (O_H) are also shown. Image taken from Taanman, 1999 (Taanman 1999).

Interestingly, the analysis of the LSP and HSP1 regions showed two enhancer elements required for optimal transcription (Fisher and Clayton 1988). They were found between -10/-40 nt far from respective IT sites (Figure_6). These elements include the binding sites of the transcription factor A (mtTFA/TFAM)(Chang and Clayton 1984; Hixson and Clayton 1985; Fisher, Topper et al. 1987). The binding of TFAM to the transcription start site ensures transcription to proceed. In vitro transcription studies demonstrated that HSP1 and LSP promoters are functionally

independent (Walberg and Clayton 1983; Chang and Clayton 1984; Hixson and Clayton 1985; Topper and Clayton 1989). TFAM induced conformational changes to the IT sites that would allow the RNA polymerase to access to the template to initiate the transcription.

3.4.2.1 Transcription initiation in human mitochondria and RNA post-transcriptional modifications.

In addition to TFAM, other three proteins are essential for the transcription initiation: the mitochondrial RNA polymerase (mtRNAP/POLRMT), TFB1M and TFB2M.
Mammalian POLRMT in its mature form (without the mitochondrial targeting sequence) is a protein of 1,189 amino acids that binds to promoter elements in complex with other proteins. This protein is homologous to the RNA polymerase of T7 bacteriophage (T7 RNAP). The analysis by X-ray crystallography showed that human POLRMT contains the catalytic domain at the C-terminus of the protein (amino acids 647–1,230). In addition, it has an N-terminal domain (amino acids 368–647) with high similarity to the promoter-binding, AT-rich recognition loop of T7 RNAP (Ringel, Sologub et al. 2011). A unique feature is its N-terminal extension (NTE: amino acid 42-368) constituted by a pentatricopeptide repeat (PPR) domain located near the point from which the newly synthetized RNA leaves the catalytic POLRMT domain. The PPR domain hypothetically should prevent the reannealing of the mRNA to the template DNA and thus allow transcription (Schwinghammer, Cheung et al. 2013).
TFB2M and TFB1M proteins were identified due to their primary sequence similarity to the yeast transcription factor mtTFB. The high similarity between TFB1M and TFB2M proteins suggests a gene duplication event. TFB1M is a transcription factor and overall a methyltransferase. This protein less participate in transcription initiation to respect TFB2M (Falkenberg, Gaspari et al. 2002). Regarding TFB2M, it has a significantly weaker methyltransferase activity but stimulates transcription 10- to 100-fold more efficiently than its paralogue TFB1M (Falkenberg, Gaspari et al. 2002). It forms a transient heterocomplex with POLRMT at the mtDNA promoters. In addition, it interacts with the priming substrate while stabilizing the opening of the two parental

strands (Sologub, Litonin et al. 2009). Once the POLRMT leaves the promoter and elongation starts, TFB2M leaves the complex similarly to as its homologous mtTFB does in yeast (Mangus, Jang et al. 1994).

So far different transcription initiation models have been proposed depending on the formation and the composition of the transcription machinery. In 2003 McCulloch proposed a transcription model of three components: TFAM, TFBM and POLRMT (McCulloch and Shadel 2003). In this model TFAM binds first and wraps the mtDNA at its specific sequences at HSP or LSP. This causes a distortion that allows POLRMT binding. TBFM bridges TFAM and mtRNAP by interacting simultaneously with the C-terminal tail of TFAM and the N-terminal region of POLRMT. With these contacts TFB stabilizes mtRNAP on the promoter region and ensures a specific initiation of transcription (Figure_8).

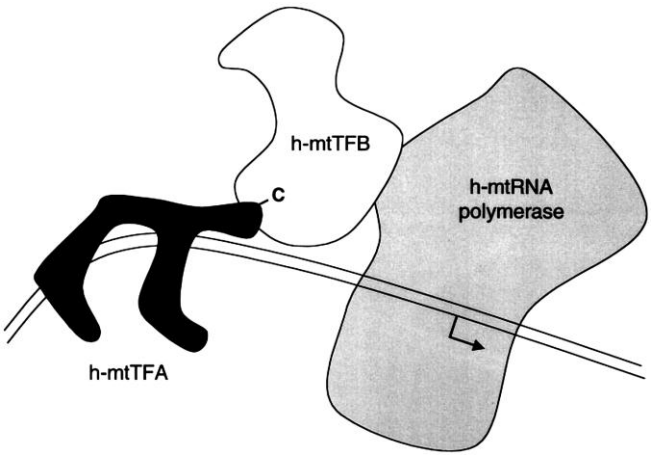


Figure_8. Model of interaction of the essential transcription components.
Human TFAM (coloured in black) binds and bends the promoter at a specific sequence upstream the transcription initiation site, indicated by an arrow. The C-terminal TFAM tail is showed contacting h-mtTFB. H-mtTFB serves as an adapter that bridges TFAM and h-mtRNA polymerase. Image taken from McCulloch and Shadel, 2003 (McCulloch and Shadel 2003).

In 2009 Sologub and collaborators discovered that a bubble created by non-complementary nucleotides covering the region from -3 to +4 (being +1 the transcription initiation site), induced POLRMT to efficiently start transcription (Sologub, Litonin et al. 2009). By crosslinking assays they observed that during transcription initiation TFB2M was cross-linked to both the priming NTP and the

DNA template base +1. They also observed that TFBM2 stabilizes POLRMT at the exact position for transcription initiation. Thus Sologub et al. hypothesized that TFB2M stabilizes the open promoter and, together with TFAM, "cements" the transcription machinery onto it and prevents competing re-annealing of the non-transcribed strand during transcription initiation. During elongation, when the 3' end of the nascent RNA occupies the binding site for the priming substrate (and thus is stabilized by complementary interactions with the template DNA), TFB2M is no longer required and likely dissociates from the complex (Figure_9).

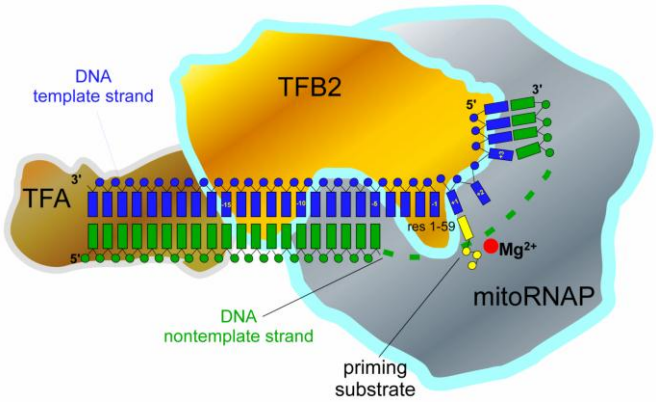


Figure_9. Model of the transcription initiation complex.
TFB2M binds to the promoter region and contacts TFAM, POLRMT and the priming nucleotide. The N-terminus of TFB2M is shown in the vicinity of the POLRMT active site. Image taken from Sologub et al, 2009 (Sologub, Litonin et al. 2009).

However, in 2011 Shutt et al. proposed a transcription initiation model of two components: POLRMT and TFB2M (Shutt, Bestwick et al. 2011). This model proposes that the transcription initiation depends on the amount of TFB2M and that TFAM acts just as a transcription activator instead of being part of the transcription initiation complex. Shutt et al. showed that HSP and LSP in vitro had the same transcription efficiency in presence of TFB2M and POLRMT even if these promoters possess a different sensitivity for TFAM, which in minimal amounts activate LSP but not HSP. Shutt et al. also postulated that TFAM could act as transcription regulator: At

low amounts TFAM activates transcription of LSP, and by increasing the TFAM concentration HSP is also activated.

Another alternative role of TFAM as transcription activator was proposed by Lodeiro et al. in 2012. They showed that TFAM binding to HSP2 at position -10, +1 and /or +5 inhibits transcription (Lodeiro, Uchida et al. 2012). This inhibition was due to the incapability of the transcription machinery to assemble when TFAM was binding the HSP2 promoter (Lodeiro, Uchida et al. 2012). Competition experiments demonstrated that when the amounts of POLRMT and TFB2 are higher than the amount of TFAM the transcription inhibition is released.

Finally, in 2015 Morozov et al., proposed a pre-initiation complex that includes mtRNAP, TFAM and the promoter DNA (Morozov, Parshin et al. 2015). By protein-protein crosslinking they showed that TFAM contacts the N-terminal domain of mtRNAP in presence of DNA, thus the complex is formed on the nucleic acid. TFAM induces a sharp bend in the promoter DNA, which helps to juxtapose POLRMT at the −50 to −60 region. In the next step, TFB2M joints the mRNAP/TFAM/DNA ternary complex and melts the promoter. This results in the open initiation complex (Figure_10).

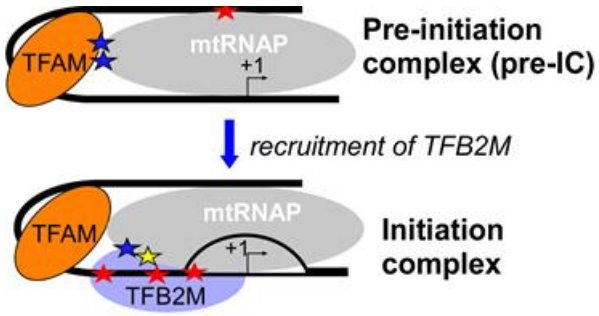


Figure_10. Schematic model of transcription pre-initiation and initiation complex.
TFAM recruits mtRNAP to the promoter forming the pre-initiation complex (pre-IC). The binding of TFB2 to the pre-initiation complex results in promoter melting and thus formation of an open initiation complex. Blue stars indicate protein-protein crosslinking by the pBpa cross-linker. Yellow stars indicate TFB2/mtRNAP crosslinking by DSG cross-linker. Red stars indicate TFB2/DNA crosslinking by 4-thioUMP and 6-thioGMP cross-linkers. Image taken from Morozov et al, 2015(Morozov, Parshin et al. 2015).

Once the transcription starts, a mitochondrial transcription elongation factor, TEFM, interacts with the catalytic C-terminal part of POLRMT and helps the polymerase to bypass regions of highly structured RNA, such as tRNA clusters and the G-quadruplex at CSB2. TEFM also stimulates bypass of oxidative lesions, such an 8-Oxo-2-deoxyguanosine, which otherwise may terminate transcription (Agaronyan, Morozov et al. 2015; Posse, Shahzad et al. 2015).

Transcription termination initiated at HSP or LSP occurs downstream the 16S rRNA, at the $tRNA^{LEU}$ gene, and is induced by transcription termination factor mTERF1, which bends the DNA (Jimenez-Menendez, Fernandez-Millan et al. 2010) and causes base pair flipping after binding at specific sequences (Yakubovskaya, Mejia et al. 2010). The action of mTERF1 is bidirectional. Therefore, at $tRNA^{LEU}$, MTERF1 attenuates the H-strand transcription, whilst it halts the L-strand transcription precisely at this point where no L-strand encoded genes are found downstream (Asin-Cayuela, Schwend et al. 2005; Terzioglu, Ruzzenente et al. 2013) (see Figure_5).

The transcription products of HS or LS are polycistronic mRNAs. These are processed, thus single and mature forms of tRNA, rRNAs and protein mRNAs are generated. Interestingly, the tRNA genes are found between protein-encoding genes, which suggested the "tRNA punctuation model" for mtDNA (Ojala, Montoya et al. 1981). Following this model, the tRNA secondary structures are specifically recognized by dedicated endonuclease proteins that cleave at their 5' and 3' ends. By this, the tRNAs and the flanking mRNAs or rRNAs are released. In mammalian mitochondria the 5' end of tRNAs is processed by the ribonuclease P (RNaseP), constituted by three subunits (MRPP1-3), while the 3' end is processed by RNaseZ (Hallberg and Larsson 2014). Once tRNAs, rRNAs and mRNAs are released, they are post-transcriptionally modified. Maturation of tRNAs involves addition of the CCA sequence to their 3' end by the ATP(CTP):tRNA nucleotidyltransferase (Rossmanith, Tullo et al. 1995). The mature form of rRNAs presents a short adenyl stretch at their 3' end (Dubin, Montoya et al. 1982). Finally, for mRNAs the post transcriptional modifications consist of a 3' end polyadenylation, during or immediately after cleavage, mediated by poly (A) polymerase (Amalric, Merkel et al. 1978). Upon mRNA 3' polyadenylation, translation takes place at the mitoribosomes. Mitoribosomes have a sedimentation coefficient of ~55S and are constituted by a big

(~39S) and a small (~28S) subunit that contain, respectively, the 16S and 12S rRNA species encoded by the mtDNA (Attardi and Ojala 1971; Brega and Baglioni 1971). The ~28S subunit recognizes and tightly binds the mRNA at its 5' end (Denslow, Michaels et al. 1989). Mitochondrial initiation factor-2 (mtIF-2) crucially promotes binding of fMet-tRNA to the small ribosomal subunit in presence of GTP. Hydrolysis of GTP facilitates the release of mtIF-2 and the concomitant association of the large (39S) ribosomal subunit to form the 55S initiation complex (Liao and Spremulli 1991; Ma and Spremulli 1996). Once the translation starts, the elongation step is under control of the mitochondrial elongation factors mtEF-Tu, mtEF-Ts, and mtEF-G (Schwartzbach and Spremulli 1989; Chung and Spremulli 1990).

3.4.3 Human mitochondrial DNA replication.

The mitochondrial DNA polymerase POLγ is the main component of the human mtDNA replisome. POLγ is a heterotrimer of 140 kDa that belongs to the family-A DNA polymerases. It is constituted by one catalytic subunit, POLγA, and two POLγB accessory subunits that enhance the activity of POLγA (Gray and Wong 1992; Yakubovskaya, Chen et al. 2006). POLγ present a 3' to 5' proofreading activity and a double strand binding ability (Carrodeguas, Pinz et al. 2002). During replication POLγ is functionally associated with the mitochondrial helicase TWINKLE (Korhonen, Pham et al. 2004). TWINKLE catalyses ATP-dependent unwinding of the mtDNA duplex in the 5' to 3' direction at the replication fork (Korhonen, Pham et al. 2004). The 16 kDa mitochondrial ssDNA binding protein (mtSSB) is another important component of the replisome. It is active as a tetramer and increases TWINKLE's helicase activity and stabilizes the single H-strand displaced at the replication fork (Korhonen, Gaspari et al. 2003; Korhonen, Pham et al. 2004).

So far different mtDNA replication models have been proposed. Among these, two are the most accredited: the "strand-displacement model" (SDM) and the "ribonucleotide incorporation throughout the lagging strand" (RITOLS) model (Clayton 2003; Holt and Jacobs 2003) (Figure_11). Both models propose asynchrony during replication of the heavy and light strand. The replication of the heavy strand is primed by the 3' end

of RNAs that are synthesized from LSP and prematurely terminated by a G-quadruplex structure at CSB2 (Wanrooij, Uhler et al. 2010; Wanrooij, Uhler et al. 2012). G-quadruplex structures are assemblies of guanine quartets. In this case, the newly synthesised RNA and the displaced parental DNA strand provide the guanine quartets that form the G-quadruplex at CSB2. Regarding SDM, during the H-strand replication, the parental H strand becomes exposed and is stabilized by mtSSB. Once the replication machinery reaches the origin of replication of the light strand, O_L, this latter becomes single-stranded due to the displacement of the parental H-strand. At this moment the O_L sequence folds into a stem-loop structure that cannot be bound by mtSSB. This allows POLRMT to initiates primer synthesis from a poly-T stretch present at O_L, located close to the top of the stem (Fuste, Wanrooij et al. 2010). After 25nt POLRMT is substituted by POLγ and the L-strand synthesis progresses by displacing mtSSB from the H strand, which is now the template (Figure_11).

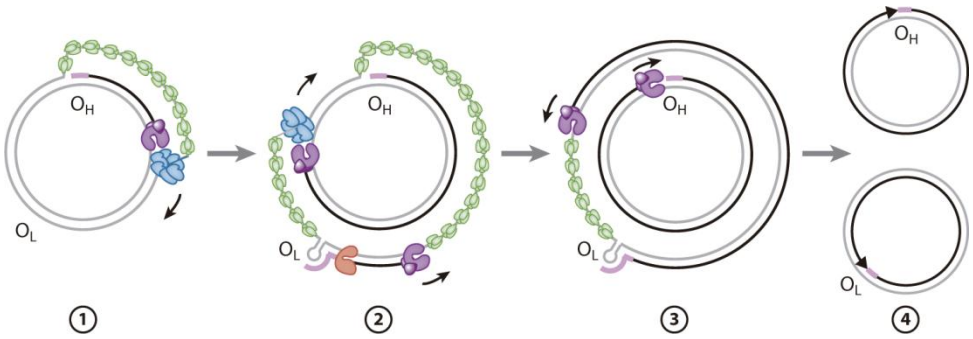


Figure_11. Mitochondrial DNA replication: the strand displacement model.
The replication starts with the assembly of the replisome at the O_H site and proceeds unidirectionally, producing a new H-strand that displaces the old one. In order to stabilize the parental H-strand, mtSSB proteins (in green) bind it (step1). When the H-strand replication machinery passes by the light-strand origin (O_L), a stem-loop structure is formed in the displaced strand (Step2). POLRMT (in orange) starts to synthesize short primers at the stem-loop, next used by Polγ (violet) for L-strand synthesis. For L-strand synthesis Twinkle (in blue) is not necessary anymore (step3). Once the two strands are completely replicated two DNA molecules are formed with a nick at the H or in the L strand in OH or OL proximity, respectively. Image taken from Gustafsson, et al 2016 (Gustafsson, Falkenberg et al. 2016).

The RITOLS model proposes a mechanism of replication similar to SDM. The difference is that during H-strand synthesis the parental H-strand is not bound by mtSSB but by the processed mRNA that remains hybridized until is displaced or degraded during lagging strand DNA synthesis (Reyes, Kazak et al. 2013).

In both models, once either strand has been replicated, two DNA molecules are obtained with a nick near O_H or O_L, respectively (Figure_11). The required ligation is catalysed by DNA ligase III (Lakshmipathy and Campbell 1999). The degradation of the RNA primers is performed by ribonuclease H1 (RNaseH1) while mitochondrial genome maintenance exonuclease-1 (MGME1) is involved in DNA primer degradation starting 100 nt upstream to the O_H site (Holmes, Akman et al. 2015; Uhler and Falkenberg 2015).

3.4.4 Human mitochondrial DNA forms nucleoids.

Mammalian mtDNA has a contour length of approximately 5 μm (Nass 1966). In order to fit inside the mitochondria, which typically have a width of 0.5 μm, mtDNA is compacted into nucleoprotein structures named "nucleoids" due to their similarity to the bacterial nucleoid. The organization and maintenance of these structures is fundamental for the correct mtDNA expression and transmission. TFAM is the major contributor to nucleoid architecture. Different microscopic techniques have been employed to describe the mammalian mitochondrial nucleoids. Satoh and Kuriowa in 1991 provided one of the earliest microscopy characterizations of nucleoids in mammalian cells using DAPI as a fluorescent DNA-binding compound. They observed an average of 3.2 nucleoids per organelle and estimated an average of 4.6 mtDNA molecules per mitochondrion (Satoh and Kuroiwa 1991). In 1996, Bereiter-Hahn and Voth (Bereiter-Hahn and Voth 1996) used Pico-Green dsDNA staining to visualize the mtDNA and estimated 1 to 6 mtDNA molecules per nucleoid. Other studies employed antibodies against mtDNA or nucleoid components (Alam, Kanki et al. 2003; Garrido, Griparic et al. 2003; Iborra, Kimura et al. 2004; Legros, Malka et al. 2004), or used fluorescent proteins fused to TFAM and Twinkle (Spelbrink, Li et al. 2001; Garrido, Griparic et al. 2003; Goffart, Martinsson et al. 2007). In these works conventional light microscopy techniques appeared to be a good strategy for localization of proteins and nucleoids *in vivo.* However, these techniques were not able to resolve a structure smaller than 150 nm (Figure_12). In 2011 Kukat and collaborators used stimulated emission depletion (STED) microscopy, a super

resolution technique with a resolution limit much higher (~20 nm in the lateral dimensions and ~50 nm in the axial dimension) respect to the canonical light microscopy (lateral resolution limit: ~300 nm) (Figure_12). Fibroblast cell cultures grown in presence of BrdU and exposed to anti-TFAM and anti-mtDNA antibodies were employed. In addition, different mammalian cell lines such as human primary culture fibroblasts (Fibro), human cervix adenocarcinoma cells (HeLa), human osteosarcoma cells (U2OS), human glioblastoma cells (U373), culture mouse embryonic fibroblasts (MEF), African green monkey (*Cercopithecus aethiops*) kidney epithelial cells (Vero), and potoroo (*Potorous tridactylus*) kidney cells (PtK2) were also used. All these cell lines were incubated with antibody anti-TFAM and anti-mtDNA, and confocal microscopy images were compared to the ones of STED. While confocal microscopy showed a nucleoid size of 271 ± 23 nm, STED microscopy revealed a very uniform nucleoid size of 85 to 111 nm as the largest dimension. Furthermore, based on these STED images the number of mtDNA molecules per nucleoid and the number of TFAM molecules per mtDNA was recalculated, yielding 1.4 mtDNAs/nucleoid and 1,400 TFAM/nucleoid on average. They also reported that TFAM is the major nucleoid component, with one molecule every 10-20bp (Kukat, Wurm et al. 2011).

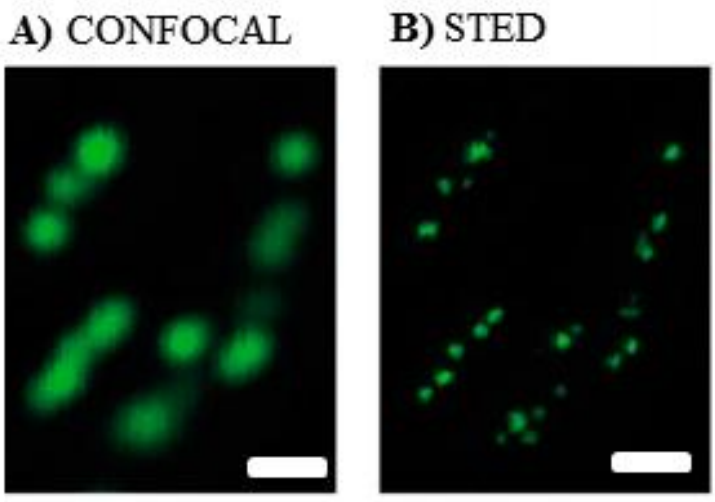


Figure_12. Confocal and STED resolution comparison.
A) and **B**) show the same section of human fibroblast cells. Nucleoids were labelled with DNA antibodies. In A) nucleoids are detected employing a confocal microscope, and they appear like solid structures with a diameter of ~300 nm. By applying super-resolution STED microscopy (B), the nucleoids observed in A) appear as agglomerates of several smaller structures with a diameter of ~99 nm. Image taken from Kukat et al, 2011(Kukat, Wurm et al. 2011).

Studies from another group that employed alternative super resolution techniques such as photoactivatable localization microscopy (PALM), interferometric PALM (iPALM), and direct stochastic optical reconstruction microscopy (dSTORM) confirmed that the mean diameter of nucleoids is ~110 nm in human cells (Brown, Tkachuk et al. 2011). Brown and collaborators, by using the PALM technique, observed that nucleoids present different size and shape. A three-dimensional volumetric analysis indicated that, on average, the mtDNA is organized in ellipsoidal nucleoids extraordinarily compacted from where matrix proteins are largely excluded. Furthermore, they found that nucleoids are closely associated with the inner membrane and often appear to be wrapped around cristae or crista-like inner membrane invaginations (Brown, Tkachuk et al. 2011). In 2015, Kukat and collaborators used STED to analyse the nucleoids shape and dynamics in mouse embryonic fibroblast (MEF) cells. They compared "wild type" MEFs with MEFs that over-expressed TFAM. Cells with TFAM over-expression showed an increase of mtDNA copy number, which correlated with an increase in the number of nucleoids but not in the number of mtDNA in the same nucleoid (which was estimated to be 1.4 mtDNA/nucleoid). Furthermore, the nucleoids in both lines present essentially the same irregular ellipsoidal shape (Brown, Tkachuk et al. 2011; Kukat, K.M. et al. 2015).

3.5 Human mitochondrial TFAM: structure and function.

3.5.1 TFAM: from the gene to the 3D structure.

TFAM belongs to the superfamily of high-mobility group (HMG) proteins, described for the first time in 1973 by Goodwin (Goodwin, Sanders et al. 1973). The term 'high mobility group' is because these proteins, discovered by acidic extraction of mammalian cellular chromatin, had a high electrophoretic mobility(Goodwin, Sanders et al. 1973). HMG superfamily includes three structurally and functionally distinct classes of proteins, defined as: HMG–nucleosome-binding family (HMGN), HMG-AT-hook family (HMGA) and HMG-box family (HMGB). TFAM belongs to the HMGB family, so we will focus on this group. The HMGB proteins have different

roles: in the nucleus and mitochondria are found as architectural DNA binding proteins, whereas in the cytoplasm function as signalling regulators, and in the extracellular environment as inflammatory cytokines (Park, Gamboni-Robertson et al. 2006; Malarkey and Churchill 2012). In the early 90's it was shown that HMGBs are characterised by an archetypical DNA binding domain, the HMG-box, with an "L" shape constituted by 3 α-helices and an N-terminal extended segment (Weir, Kraulis et al. 1993). Helix 1 and the N-terminal segment form the long arm of the L, while the two helices 2 and 3 form the short L-arm. Mammalian HMGB proteins can further be subdivided into two major groups: HMGB sequence-specific and non-sequence-specific DNA binding proteins, on the basis of their capability to produce DNaseI footprints on specific DNA sequences (Landsman and Bustin 1993). HMGB sequence-specific proteins usually have a single HMGB domain and no acidic C-terminal tail (e.g. the Sex-determining Region Y (SRY), or the Sry-related HMG box (SOX)). In contrast, the HMGB non sequence-specific type usually presents two HMGB domains and an acidic C-terminal tail (e. g. high-mobility group box 1 (HMGB1)). HMGB proteins are characterised by their interaction with the minor groove of the DNA double helix and provoke a strong DNA distortion induced by the protein L-shape and by electrostatic and hydrophobic interactions (Grosschedl, Giese et al. 1994) (Figure_13). They bind and widen the minor groove, inducing a strong bend towards the major groove. Both types of HMGBs present at least one hydrophobic residue at the HMG-box domain that makes an insertion between DNA base pairs, further stabilizing the interaction and contributing to DNA bending (Churchill, Klass et al. 2010)(Figure_13). Additional residues may also perform partial insertions, but these second insertions are not as deep as the major one.

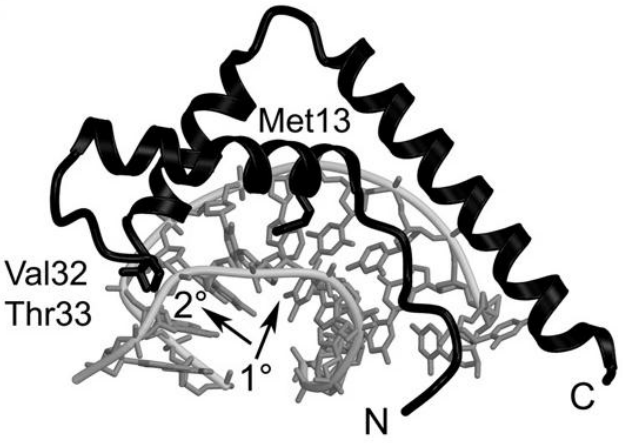


Figure_13. Ribbon diagram of a HMGB (HMGD) protein in complex with DNA.
When HMGD is bound to the DNA induces a severe kink stabilized by the insertion of three residues: Met13, Val32 and Thr33. The insertion sites are indicated by arrows, while the inserting residues are labelled. Image taken from Churchill et al, 2010 (Churchill, Klass et al. 2010).

TFAM, in its mature form (without the mitochondrial leading sequence), is a protein of 203 aa (24.4 kDa). It is unique in that it preferentially binds to specific sequences (Fisher and Clayton 1988; Alam, Kanki et al. 2003). However, it contains two HMG box domains (HMG1 and HMG2), which is typical of non-specific DNA binders (see above). Therefore, it combines features of specific and non-specific DNA recognition. The gene codifying for TFAM is located inside the nucleus and spans about 10 kb. It consists of seven exons and six introns (Reyes, Mezzina et al. 2002). Exon 5 can splice alternatively, resulting in a second TFAM isoform (Tominaga, Hayashi et al. 1993). The TFAM-Δexon5 isoform binds DNA non-specifically and compact it. However, this isoform only slightly binds the LSP region and cannot activate transcription, (Dairaghi, Shadel et al. 1995). This isoform lacks the first helix of HMG2.

In 2011, the crystallographic structure of full-length human TFAM in complex with the cognate binding site at LSP (TFAM/LSP complex) was described for the first time (Ngo, Kaiser et al. 2011; Rubio-Cosials, Sidow et al. 2011). The two reports showed highly similar complexes in which the protein-DNA contacts are almost identical despite the crystals contained DNAs of different length (22 bp in Rubio-Cosials 2011, PDB (3TQ6); and 28bp in Ngo et al 2011, PDB (3TMM)). TFAM is an all-alpha modular protein constituted by two HMG domains, HMG-box1 and 2, each one of approximately 75 aminoacids (positions 44-120 and 153-225, respectively). The two domains have the same fold, which consists of three helices, helix1, helix2 and the

longest helix3, all together organized in an "L" shape (Figure_14). An extended N-terminal region of 13 residues packs antiparallel to helix 3, both regions forming the long L-arm of the L-shape. In their turn, helices 1 and 2 form the short L-arm. Both HMGbox1 and 2 domains show similar dimensions except for helix1, which in HMGbox2 is one turn shorter. The two HMG domains are connected by a linker of helical conformation (aminoacids 124-152), and are followed by a C-terminal, positively charged tail that shows an almost extended conformation (aminoacids 226-246) (Figure_14). TFAM contacts the DNA minor groove by the inner surface of the L-shape of the HMG boxes, by polar and non-polar interactions. In addition, two specific residues, Leu58 from HMG-box1 and Leu182 from HMG-box2, insert in DNA steps of the respective contacted regions (Figure_13). In particular, Leu58 inserts at the contacted region $T_1A_2A_3||C_4A_5G_6T_7$ (vertical bars correspond to Leu58 insertion between $A_3||C_4$ base pairs) whereas Leu182 inserts at $C_{14}C_{15}||A_{16}A_{17}C_{18}T_{19}A_{20}A_{21}$. These insertions stabilize respective ~90º DNA kinks, resulting in an overall 180º (U-turn) total bent. With respect to an ideal DNA, the two inserted steps show both an important increase in roll and a decrease in twist. Alignment of the two DNA regions contacted by HMG-box1 and HMG-box2 show an inverted consensus of the inserted DNA binding motif, which is separated by 10 bp: $A_3|C_4$ –10 bp– $C_{15}|A_{16}$. This follows the symmetry head to head of the HMG boxes, as shown by Rubio-Cosials et al in 2011 (Rubio-Cosials, Sidow et al. 2011). In the Ngo structure the DNA includes 5 more nucleotides at the 5' of the LSP recognition region and 1 more at the 3' respect to the sequence used by Rubio-Cosials. Despite this difference in sequence length, the Leu insertions happen at the same DNA steps (Ngo, Kaiser et al. 2011; Rubio-Cosials and Sola 2013). In both structures the positive charged helix of the linker between HMG boxes plays a fundamental role in stabilizing the bending of the DNA in a U turn conformation.

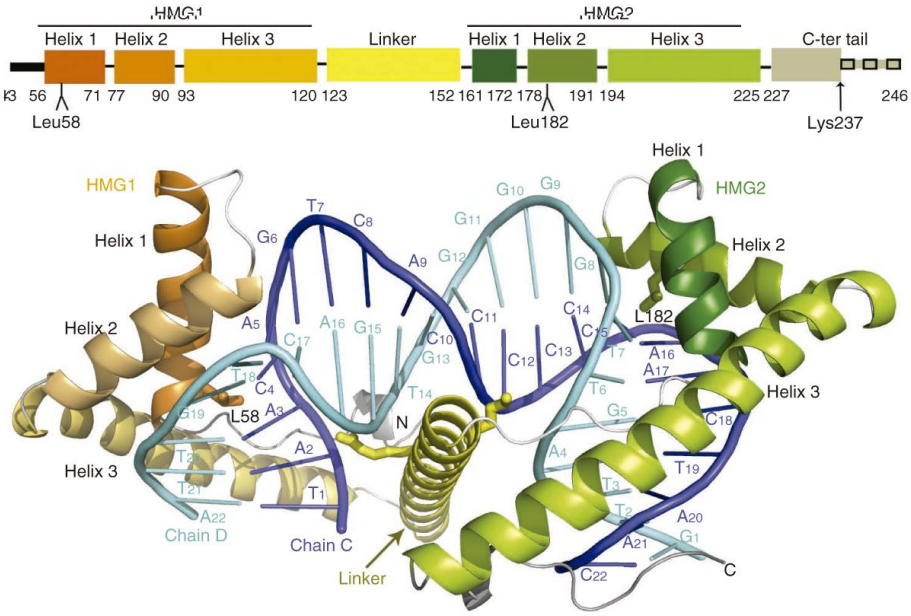


Figure_14. Crystal structure of the TFAM/LSP complex.
At the top is presented a schematic representation of TFAM domains, with both the intercalating residues and the N- and C-terminal traced ends indicated. At the bottom a ribbon diagram of the TFAM/LSP complex crystal structure (pdb code: 3TQ6) is shown. HMG1 is colored in orange while HMG2 is colored in green. The linker is colored in yellow and the N and C-terminal ends are also indicated. The intercalating residues are shown as sticks. Image taken from Rubio-Cosials 2011 (Rubio-Cosials, Sidow et al. 2011).

Analysis of unbound TFAM in solution was performed by small angle X-ray diffraction (SAXS). Unexpectedly, these studies showed that the linker between the two HMG boxes is highly unstructured in solution, is not folded into a α-helix as in the crystal structure but extended with random conformations. Due to this, the two HMG-box domains show a high degree of freedom with respect to each other (Figure_15).

Figure_15. SAXS TFAM model.
Ribbon representation of a 50 models subensemble where all TFAM molecules are superimposed by their HMG2 domain (in green) (left and right panels are two 90 degrees rotations of the ensemble). Note that the HMG1 domain (in orange) shows a great degree of freedom. Image taken from Rubio-Cosials 2011 (Rubio-Cosials, Sidow et al. 2011).

This suggests a mutual rearrangement and fitting of both TFAM and LSP during binding, which involves bending of the DNA into a U-turn and folding of the linker into a α-helix (Rubio-Cosials, Sidow et al. 2011).

TFAM bends DNA by 180º independently from the sequence, as shown by the structure in complex with a non-specific 22bp DNA (Ngo, Lovely et al. 2014). This complex also formed a U-turn similar to the one in the TFAM/LSP crystal. In addition, another structure of TFAM in complex with part of the TFAM binding site at HSP also shows structural features similar to the previous ones (Ngo, Lovely et al. 2014). These results substantiated formation of U-turns during TFAM recognition of non-specific DNA sequences, for example during packaging of the DNA.

3.5.2 TFAM binding sites at the mtDNA control region.

In 1984 Chang and Clayton identified the minimal mtDNA region necessary for transcription initiation of both the H and L strands, which correspond to the heavy and light strand promoters (HSP and LSP, respectively). The promoters span nucleotides -16 to +7 for HSP and -28 to +10 for LSP. In both regions a consensus sequence of 15 bp 5'-CANACC(G)CC(A)AAAGAYA-3' was found (Chang and Clayton 1984). The transcription initiation site is an integral part of each promoter and each promoter can function in the absence of the other. In 1985 Fisher and collaborators found that transcription was regulated by a mitochondrial transcription factor, which recognizes the promoter and confers specificity to POLRMT. This factor was named mtTF and corresponds to current mtTFA/TFAM (Fisher and Clayton 1985). By using partially purified TFAM, the same group mapped the specific recognition sites in the D-loop regions by run off transcription assays with truncated promoters and DNaseI protection experiments (Fisher, Topper et al. 1987). The corresponding TFAM binding sites were found inside the two promoter regions: one at nt530-550 in HSP, and the other one at nt440-420 in LSP. Notably, these two sequences are in the same orientation in the mtDNA sequence and can be aligned. However, the promoters initiate transcription in opposite sense. Therefore, it was deduced that TFAM is active in both orientations (Fisher, Topper et al. 1987). The actual orientation of TFAM on

the promoters is still under discussion, since biochemical studies suggest not tandem but opposite binding orientations on the DNA, following the sense of respective promoters (Morozov and Temiakov 2016).

Apart from the binding sites determined at LSP and HSP, two additional protected regions were found downstream of LSP, between CSBII and CSBI, called site X (nt 276-303) and site Y (nt 233-260) (Fisher, Topper et al. 1987) (Figure_16).

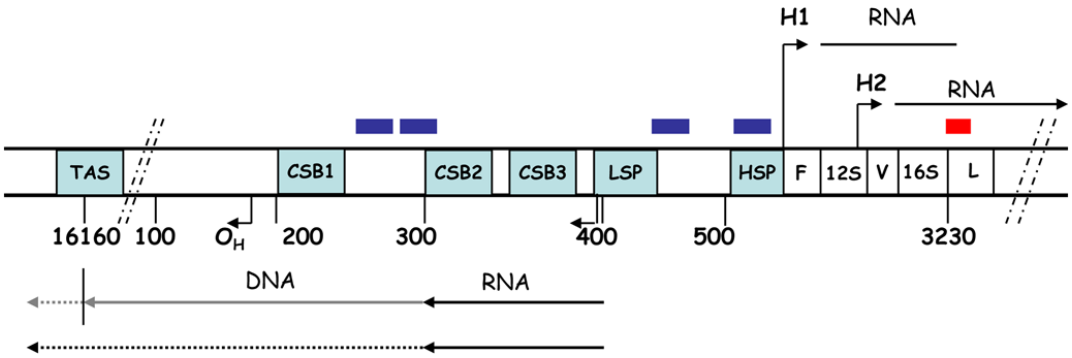


Figure_16. Schematic representation of TFAM binding sites inside the control region.
The blue rectangles represent TFAM binding sites, while the red one is the MTERF1 binding site. The two promoters, LSP and HSP, are indicated, together with the conserved sequence blocks (CSB1-3) and the termination associated sequences (TAS). The arrows pointing left and right show the transcription orientation of the heavy (H1 and H2) and L strands, respectively. The heavy strand origin of replication (O_H) is represented together with genes encoding tRNAs for phenylalanine, valine and leucine (F, V, and L, respectively). Image taken from Suissa et al, 2009 (Suissa, Wang et al. 2009).

Alignment of these four sequences showed a conserved 5' region (TAAC) and a limited conservation at the 3' region (Fisher, Topper et al. 1987). (Figure_17).

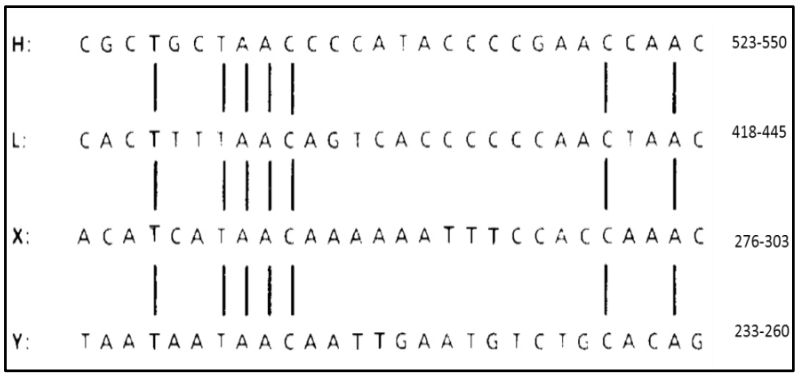


Figure_17. Alignment of the four TFAM D-loop binding sites.
HSP (H), LSP (L), site X and site Y sequences are compared with their map position indicated at right. Vertical lines connect the conserved nucleotides. Image taken from Fisher et al, 1987 (Fisher, Topper et al. 1987).

In 1995, Ghivizzani and collaborators analysed the control region by *in organello* footprinting, performed on isolated mitochondria (Ghivizzani, Madsen et al. 1994). They discovered the same TFAM binding sites at HSP, LSP and X and Y described previously. In 2009 Suissa and collaborators showed in a screen of more than 2500 human mitochondrial genomes that represented all major populations worldwide, the natural variation in the identified TFAM binding sites inside the control region (Suissa, Wang et al. 2009). They showed that the Caucasian haplogroup J presents a C295T mutation that enhances both the binding of TFAM and the levels of *in vitro* L-strand transcription. This mutation maps inside site X and seems to be also correlated with an increase of mtDNA copy number *in vivo*. Since site X is downstream to LSP, Suissa et al., proposed that this mutation could evidence a role of site X in transcription. They also suggested that TFAM binding downstream to LSP could cause important structural changes to the promoter that enhanced the transcription (Suissa, Wang et al. 2009).

3.5.3 TFAM multimerization and mitochondrial DNA compaction.

In order to better understand the molecular mechanism of TFAM mtDNA maintenance, the knowledge of its functional stoichiometry is fundamental. In 2007 Kaufmann et al. showed by surface plasmon resonance that TFAM in complex with DNA is a homodimer (Kaufman, Durisic et al. 2007). Gangelhoff et al. in 2009 showed, by analytical ultracentrifugation, that TFAM in solution and in the absence of dsDNA is a monomer, even if by size exclusion chromatography it elutes as a dimer (Gangelhoff, Mungalachetty et al. 2009). In the same work they showed that TFAM in complex with LSP dimerizes (Gangelhoff, Mungalachetty et al. 2009). In contrast Wong et al in 2009 showed also by analytical ultracentrifugation that TFAM in the absence of DNA is in equilibrium between monomeric and dimeric states (Wong, Rajagopalan et al. 2009). Additionally, by fluorescence anisotropy they observed that TFAM binds cooperatively to specific and no specific DNA as a dimer (Wong, Rajagopalan et al. 2009). Both Gangelhoff and Wong concluded that the DNA-binding ability of TFAM is mainly driven by its first HMG-box because, in both cases, the

second HMG-box showed a drastically lower DNA-binding capability than HMG-box1. Between 2011 and 2014 crystallographic structures of TFAM in complex with LSP, a truncated HSP form and non-specific sequences were published (Ngo, Kaiser et al. 2011; Rubio-Cosials, Sidow et al. 2011; Ngo, Lovely et al. 2014). In all these structures the ratio TFAM:DNA was 1:1. Moreover, Rubio-Cosials et al in 2011 demonstrated by small angle X-ray scattering that TFAM in solution is a monomer even at high concentrations (Rubio-Cosials, Sidow et al. 2011). In 2014 Ngo et al showed by FRET that TFAM dimerization occurs in presence of DNA in solution. These authors analysed the crystal packing throughout different crystals and observed that Lys95, Tyr99, Glu106, Glu112 and Arg116 are systematically involved in a recurrent protein interface. These aminoacids are located within helix3 of HMGbox1, which in the crystal performs antiparallel contacts with itself (Ngo, Lovely et al. 2014). In 2015 Kasashima et al demonstrated the existence of TFAM dimers and multimers *in vivo*, by using in organello crosslinking techniques and fusion of TFAM with fluorescent dyes in HeLa cells. In their experiments they tested the dimer mutant reported by Ngo et al., and confirmed that dimerization occurs in TFAM helix3 of HMGbox1 (Kasashima and Endo 2015).

TFAM wraps and bends circular DNA, causing negative supercoiling (Fisher, Lisowsky et al. 1992). As shown above, it is widely reported that TFAM binds non-specifically the mitochondrial DNA. In contrast, its role as specific-sequence regulatory factor seems to be confined inside the D-loop. Ghivizzani et al., in 1994 proposed that inside the D-loop TFAM performs a phased binding that starts at the LSP region, for which has the highest affinity. From there, multimerization on the DNA continues, reaching sites Y, X and CSBI. They did not detect binding to CSBII and CSBIII, thus they interpreted that these sequences exclude the interaction with TFAM. This suggests that the local structural organization limits the areas in which TFAM can bind, wrap and bend the DNA (Ghivizzani, Madsen et al. 1994).

In 2011 Shutt et al., proposed a model in which TFAM is a transcription modulator more than a component of the transcription complex. In their model TFAM activity is regulated by its own amounts in the mitochondrial matrix: at low concentrations TFAM activates transcription from LSP. HSP is activated only at higher TFAM concentrations. Finally, saturation by TFAM blocks transcription and replication due

to high DNA compaction (Shutt, Bestwick et al. 2011). This model was corroborated by Farge et al., in 2014 by using transcription and replication assays *in vitro*. Farge showed that TFAM is able to bind to the DNA cooperatively (Farge, Laurens et al. 2012) and that at a ratio of 1 TFAM per 8 bp the replication is inhibited (Farge, Mehmedovic et al. 2014). Indeed, at a ratio of 1 TFAM per 12bp, the DNA was already fully compacted, as shown by EMSA and atomic force microscopy (AFM) (Farge, Mehmedovic et al. 2014). Regarding transcription, increasing amounts of TFAM also abolished it. In order to discriminate if the transcription was inhibited at the elongation step or it never started, *in vitro* transcription assays were performed at ratio 1TFAM:15-18bp using different template lengths. Since the shorter templates were less affected the transcription blockage were related with the elongation step and so to DNA compaction (Farge, Mehmedovic et al. 2014). In addition, Ngo and collaborators showed that dimerization was important for DNA compaction (Ngo, Lovely et al. 2014).

Based on the results of Farge et al., Ngo and collaborators proposed that mtDNA compaction mediated by TFAM could control the number of mtDNA molecules metabolically active (Ngo, Lovely et al. 2014).

The TFAM mechanism of DNA compaction is based on cooperative binding. Electron microscopy and AFM studies showed that, at low TFAM concentrations, clusters of protein (rather that single molecules) bind to the DNA and induce DNA bending and looping. Thus increasing TFAM amounts results in full DNA compaction (Figure_18) (Kaufman, Durisic et al. 2007; Farge, Mehmedovic et al. 2014; Kukat, K.M. et al. 2015). It has additionally been shown that TFAM performs cross-strand binding by bridging DNA duplex regions (Figure_18) (Kaufman, Durisic et al. 2007; Kukat, K.M. et al. 2015).

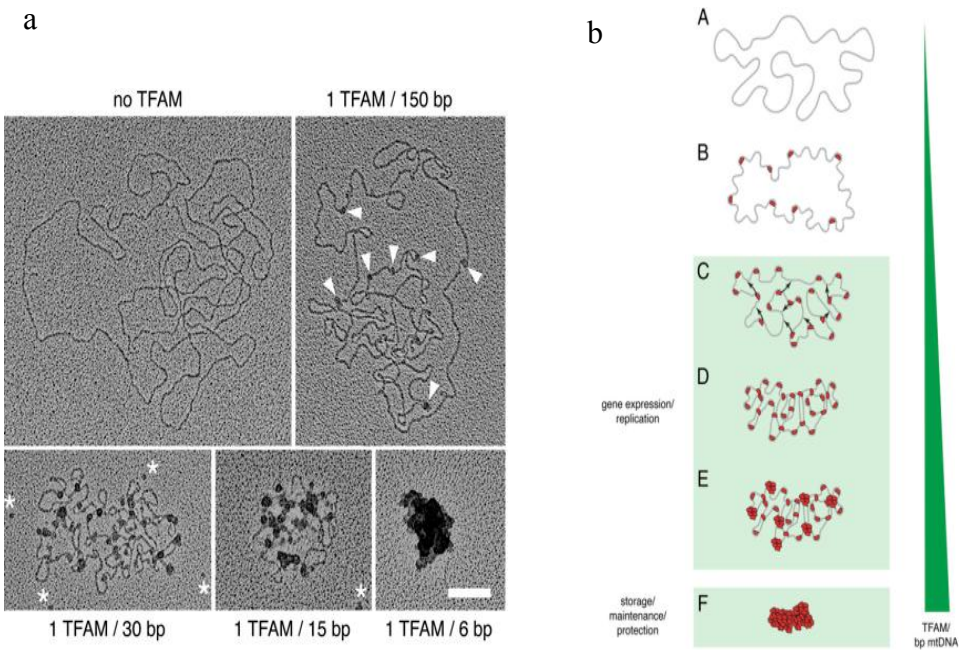


Figure_18. EM micrografies show that TFAM induces compaction of mtDNA.
In "a" electron micrographs of increasing TFAM concentration in presence of mtDNA are showed. White arrowheads indicate TFAM molecules bound to the DNA. White asterisks mark the unbound TFAM. Scale bar: 100 nm. In "b" a model for mtDNA packaging is presented. MtDNA duplex is indicated in grey, TFAM in red. The arrows represent cross-strand TFAM binding. From step "A" to step "F" a sequential increasing TFAM amount (green triangle) is postulated until a fully compacted mtDNA molecule is obtained. Images taken from Kukat et al, 2015 (Kukat, K.M. et al. 2015).

Kasashima in 2015 observed that TFAM dimerization is essential for nucleoid compaction *in vivo*, and that expression of dimer-defective mutant in HeLa cells induced enlarged mtDNA nucleoids (Kasashima and Endo 2015).

3.6 Structural and thermodynamical features of DNA binding

Protein DNA recognition and binding is fundamental for genome maintenance and gene expression. Protein/DNA interactions can be divided in two main categories. In one case the protein recognizes specific DNA bases in a process named direct readout or base-readout. In the second case, the protein recognizes the geometry of the DNA rather than performing direct contacts with the bases; this second type is called indirect

readout or shape readout (Rohs, Jin et al. 2010). The base readout includes hydrogen bonds or hydrophobic interactions between the protein and the bases at the major or the minor groove. The shape readout includes electrostatic interactions with DNA phosphate backbone of the minor or the major groove and usually a high level of flexibility of the DNA. The DNA shape types recognized by the protein include DNA kinks, narrowing of the minor groove (e.g. A-tract sequence), or a global shape modification such as DNA conformations alternative to the ideal DNA B-form. Shape (or indirect) readout of A-tract sequences often shows an ordered layer of water molecules that can mediate the interactions between the protein and the DNA. Most of the time, shape readout processes include a further shape modification after protein binding that consists in the intercalation of side chains between base pairs. Such a contact can be crucial to stabilize protein binding. (Rohs, Jin et al. 2010).

Fuxreiter et al in 2011 postulated the importance of unstructured protein regions, far from the binding motifs, for the protein selectivity and the DNA binding affinity (Fuxreiter, Simon et al. 2011). They named these region "intrinsically disordered" (ID) segments. In general proteins can use more than one mechanism to recognize and bind to the DNA. Often nonspecific contacts can be converted in specific binding after protein/DNA mutual rearrangement. Thus in general a high level of flexibility and adaptability of a protein/DNA interface is necessary for a stable interaction. In this context the ID domains may regulate DNA binding at different levels: can facilitate protein diffusion along the DNA; can be implicated in the transition from a nonspecific to a specific interaction; can modulate protein selectivity; can fold upon contact with DNA; can contribute to the affinity; and, ultimately, the presence of ID segments reduce the loss of conformational entropy upon binding (Fuxreiter, Simon et al. 2011).

The energetics of the binding is also useful to understand the molecular mechanism that governs the target DNA recognition and binding. Privalov et al., in 2007 analysed the Gibbs energy associated with the binding to the minor and major groove and discovered that, even if the Gibbs free energy (ΔG) was comparable, there was a huge difference in the enthalpic and entropic contribution in the two different binding that correlate with their dynamics (Privalov, Dragan et al. 2007). Gibbs free energy (ΔG), enthalpy (ΔH) and entropy (ΔS) are related by the equation: $\Delta G=\Delta H-T\Delta S$. For

spontaneous processes, ΔG must be <0. Changes in the enthalpy are due to changes in van de Waals, hydrogen bonding and charge interactions. Changes in entropy involve changes in the arrangement of the solvent or counterions, increase in the degree of freedom and due to rotational and translational changes. In spontaneous processes the entropy always increases. In general, in the case of proteins that interact with the DNA major groove there is a heat release during the contact reaction, the process is exothermic and enthalpy-driven ($\Delta H<0$). Instead, in the case of proteins binding to the minor groove the process is endothermic with an unfavorable enthalpy ($\Delta H>0$), which is compensated by a favourable entropy (that contributes to $-T\Delta S$). This increase in entropy has been related to the presence of an ordered spine of water molecules along the DNA minor groove in the case of A-T rich sequences. When the protein substitutes the water molecules there is an increase in enthalpy (bonds are formed and broken) but also in entropy (the water molecules dissolve, contributing to disorder), this latter drives the process (Privalov, Dragan et al. 2007). In addition, the removal of water molecules requires minimal work and results in significant loss of rigidity in the DNA, which can then be specifically modelled by the protein binding interface. Thus the bend induced by the proteins is not energetically expensive (Privalov, Dragan et al. 2009). In 2011 Privalov et al., proposed the counter ion-condensation model to better explain the protein/DNA interaction at the energetic level. Protein/DNA binding energy can be split into two qualitative components: the salt-dependent, electrostatic, and the salt-independent, non-electrostatic components. The salt-dependent component can be experimentally measured. If it is considered that the enthalpy is independent from salt concentration, the salt effect reflects just the entropy of the system. Thus, the electrostatic component of the binding is fully entropic. So this component participates and may drive the binding, and influence the affinity. Instead, the non-electrostatic component is related to the selectivity (Privalov, Dragan et al. 2011).

4. MATERIALS AND METHODS

4. MATERIALS AND METHODS

4.1 TFAM protein production.

4.1.1 TFAM cDNA cloning and protein expression protocol.

TFAM coding sequence without the mitochondrial targeting sequence (residues from 43 to 246) was previously cloned in the lab in vector pET28 (pET28-TFAM) by using NcoI and XhoI restriction sites. This vector is Kanamycin resistant.

TFAM expression protocol starts with transformation of BL21-DE3 pLys-S *E. coli* expression strain with pET28-TFAM. 2 uL (usually at 100 ng/uL) of pET28-TFAM are deposited at the bottom of the microtube, onto which 50 ul BL21-DE3 pLys-S strain are added and incubated 30 min on ice. It follows a heat-shock step (1 min at 42º C) and 5 min of incubation on ice. Next, 900 ul of LB medium are added to the transformation reaction and the cells are shacked at 200 rpm for 1h at 37º C. After this step 100 ul of reaction are plated on an agar plate with kanamycin and incubated overnight (O/N) at 37º C. One colony of the transformation plate is inoculated into a pre-culture of 4 ml of LB and shaked at 250 rpm O/N at 37º C. The pre-culture is used to inoculate a 500 ml of LB large scale culture, which is shacked at 200 rpm at 37º C, until it reaches an OD of 0.6-0.7. At this point the protein expression is induced with 1mM IPTG. After induction the culture are let grow for 4-5h at 37º and subsequently are harvested by centrifugation at 4000 rpm at 4º C for 25 min. The bacterial pellet can be stored at -80º C.

4.1.2 Chromatography techniques and TFAM purification protocol**.**

Chromatography is an analytical technique commonly used for separating a mixture of chemical substances into its individual components, so that they can be thoroughly analysed. There are different types of chromatography such as: liquid chromatography, gas chromatography, ion-exchange chromatography, size exclusion chromatography and affinity chromatography, among other types. In this work we employed affinity and size exclusion chromatography.

Affinity chromatography is based on a specific interaction between the protein in the mobile phase and a specific component of the matrix of the stationary phase. In order to release the protein from the matrix a competitor that has a higher affinity for the stationary phase than the protein has to be loaded into the column. Frequently the protein of interest is fused to tags that have high specificity for a particular matrix. Such tags fused to the target protein include a tail of six histidines (his-tag), and/or the maltose binding protein (MBP), or glutathione s-transferase (GST), among others. For each of these tags there is a specific resin, such as the nickel chelating, maltose or GSH resins, respectively for the cases mentioned. During purification of his-tagged proteins usually a gradient (linear or in steps) of imidazole is applied to avoid that the protein target co-elutes with contaminants. For GST and MBP tags, the affinity and specificity of the tag for the resin is such that a washing step followed by a single step of elution is sufficient to obtain a highly pure sample.

When the purity level of the target protein is not high enough, a polishing size exclusion chromatography (SEC) is employed after the affinity purification step. SEC can be also useful for analytical purposes. This chromatographic method separates the molecules in solution by their hydrodynamic radius (or volume), which in some cases directly correlates with the molecular weight. One subtype of SEC is gel filtration which has a stationary phase constituted by a resin made of polymers of different chemical composition. These polymers are crosslinked, creating pores whose dimensions depend on the resin. The particles at the mobile phase that fit into the pores are trapped in them, delaying their elution and thus are fractionated from the rest of the components to be purified. The dimension of the resin pore determines the resolution range of the column. Therefore, its size is fundamental for appropriate purification of the target protein.

TFAM purification consists in resuspension of the harvested cells from 500 ml culture in 50 ml of lysis buffer (750 mM NaCl, 20 mM Imidazole, 50 mM Hepes pH7.5, 1 mM DTT, DNAse, one protease inhibitor pill from Roche) and sonication (0.3 seconds on, 0.3 seconds off, 10 minutes at an amplitude of 28%. This is followed by a 13000 rpm centrifugation and subsequent filtration of the supernatant. The filtrated sample is then loaded on a His-tag 5ml affinity column (GE Healthcare), mounted on a FPLC

machine. Next a washing step of 5 column volumes in buffer A (750 mM NaCl, 50 mM Hepes pH 7.5, 20 mM imidazole, 1 mM DTT) is performed, which is followed by a linear gradient of imidazole from 20 to 500 mM, using 750 mM NaCl, 50 mM Hepes pH 7.5, 500 mM imidazole and 1 mM DTT as buffer B. The eluted fractions are analysed by SDS-PAGE using 15% polyacrylamide gels. The purest fractions are selected and pooled and their concentration measured by nanodrop (nanodrop 2000 by Thermo Fisher Scientific). The absence of nucleic acids is assessed by the O.D. relation at 280 and 260 nm (OD^{260}/OD^{280}) of the peak fractions, which must be below 0.7. It follows a concentration step using microfiltrators (Vivaspin). A second purification step by gel filtration is performed to polish the sample, in which the size of the column depends on the amount of protein and performance in separating the purification peaks. Taking into account the size of TFAM, Superdex 75 10-300 (GE Healthcare Life Sciences) or Superdex 75 26-60 (GE Healthcare Life Sciences) could be used. The size exclusion purification buffer is 750 mM NaCl, 50 mM Hepes pH 7.5 and 5 mM DTT.

4.2 TFAM-DNA complex formation and crystallization.

4.2.1 TFAM-DNA complex formation protocol.

The DNA molecules used were purchased from Sigma Aldrich (for EMSA and crystallization) and "Biomers" (for crystallization). Complementary strands are dissolved in milliQ water to a final concentration of 400 uM; the concentration is measured by a nanodrop device (nanodrop 2000 by Thermo Fisher Scientific). The annealing of complementary strands is done in a thermo-block by heating at 90º C for 10 minutes the sample and let it cooling down O/N.

In order to make the protein-DNA complex, at least 1mg/ml of purified TFAM (2ml or more total volume) is mixed with the duplex DNA by adding the appropriate volume from a DNA stock (at 400 uM) to obtain the desired protein:DNA molar ratio (for example 2:1 or 4:1). Once the protein and the DNA are mixed they are transferred to dialysis tubing with pore dimensions of 3500 Da and dialyzed step-wise against the gel filtration buffer solution containing decreasing NaCl concentration. The mild decrease

in salt favours complex formation. The first dialysis buffer consists of 500 mM NaCl, 50 mM Hepes pH 7.5 and 5 mM DTT for 2h. Afterwards, the dialysis bag containing the complex is plunged for another 2h to a second buffer that contains 250 mM NaCl together with the rest of components. The third and last dialysis step is O/N in 20 mM NaCl. All these steps are done at 4º C. Based on the amounts of TFAM needed to completely shift all the DNA molecules in electrophoresis mobility shift assay (see section 5.1.2.1 and 5.1.2.2), the TFAM/site Y or /site X complexes used for crystallization were formed at a ratio of 2 protein:1 DNA. For TFAM in complex with the large X-Y sequence of 71bp and its variants, ratios of 4:1 and 6:1 were tested.

4.2.2 Crystallization techniques.

Crystallization is the process that brings to the formation of a solid in which all the atoms or molecules are highly organized in an ordered structure known as crystal. Crystallization should start when the concentration of a compound in a solvent is higher than the solubility product of this compound. However, in general crystallization is kinetically hindered and macromolecules require a supersaturated solution to start to precipitate and form crystals (Figure_19).

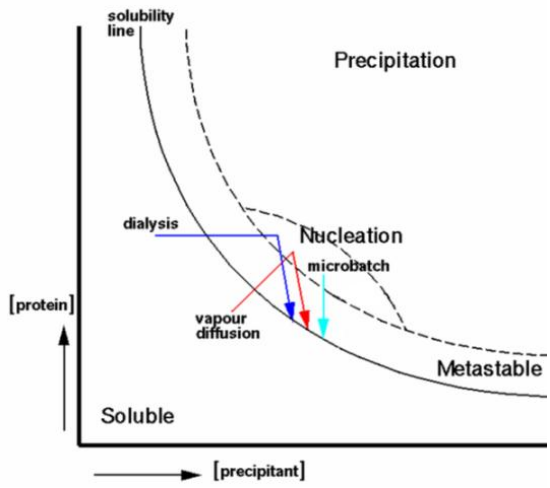


Figure _19. Protein solubility phase diagram.
Four zones representing different degree of supersaturation are shown. A zone of high supersaturation where the protein will precipitate (precipitation zone); a zone of moderate supersaturation where spontaneous nuclei take place (nucleation zone); the metastable zone where crystals can grow but cannot form; and the zone of undersaturation (soluble zone) where the protein is fully dissolved and will never crystallize. The principal crystallization methods are also indicated (dialysis, vapour diffusion and microbatch).

Nucleation precedes crystallization. Considering the phase diagram shown in Figure_19, nucleation happens just at the border between the metastable zone and the precipitation zones. The formation of nuclei and the start of crystallization reduce the concentration of soluble protein. Crystals can only grow in the supersaturated metastable zone (Figure _19). In order to promote a supersaturating state, different crystallization techniques can be employed such as dialysis, microbatch and vapour diffusion in its two different set up, hanging or sitting drop. In the vapour diffusion technique, which is the one used in this thesis, a drop containing usually 1 μL (or 100 nL for the automated set up) of protein (in its purification buffer) + 1 μl (or 100 nL) of crystallization solution is equilibrated against the crystallization solution in the reservoir. The excess of water present in the protein buffer is spontaneously transferred to the reservoir solution by vapour diffusion, so that the protein drop is dehydrated and the supersaturating state is reached, promoting protein precipitation. Successful protein crystallization often needs a considerable number of trials in which a variety of chemical and biophysical parameters are explored such as: type of precipitant (PEG, salts), concentration of precipitant, concentration of protein (in general from 5mg/ml to 25mg/ml, but it could be less o more depending on the solubility of the sample), type of buffer (Tris-HCl, Hepes, citrate, etc), pH, volume of the crystallization drop and of the reservoir, temperature (e.g., 4º, 17º, 20º), small molecule additives (salts, PEGs, glycerol, acetonitrile, DTT), biological additives, among others. The type of crystallization set up (dyalisis, microbath or vapor diffusion) can also be instrumental for crystallization. Nowdays it is usual to start with crystallization screenings using automated systems to perform the highest number of trials in the shortest time and by using very small volumes (at the nano level) of protein. Once the crystallization condition is found, it is often necessary to optimize it to obtain bigger and more ordered crystals. At this step, frequently larger volumes (microliter level) of both protein and crystallization conditions are used.

4.2.3 Crystallization protocol of TFAM/DNA complexes.

In this work crystals were obtained using the vapour diffusion technique by the sitting drop set up. Crystallization conditions were found by extensive screening of chemical conditions employing automated dispensing methods. For this, 96 wells plates (Hampton Research) were dispensed by a nanodrop dispensing robot (Cartesian model by "Genomic solutions" or Phoenix by "Rigaku"). Nanodrops containing 100 nl of protein + 100 nl of crystallization buffer from the reservoir were prepared and monitored during time. The successful or promising crystallization conditions were then scaled up to µl volumes using 24 well plates (Hampton Research) also using the sitting drop set up. For these, 500 µl of crystallization solution was dispensed in the reservoir and let equilibrated against a drop of 1 µl of protein + 1 µl of crystallization solution. Both the screening and optimization plates were incubated at 20º C. Different TFAM concentrations were tested. In the case of TFAM/site Y and TFAM /site X complexes the crystals preferentially grew at a TFAM concentration of 12-13mg/ml. Regarding TFAM in complex with the large X-Y sequence, 15mg/ml of protein was dispensed. The concentration of the sample is sometimes a crucial parameter for crystallization. Therefore, it needs to be determined with precision. However, because DNA and proteins absorb light at 280 nm the standard method of concentration determination by spectrophotometric methods using this wavelength is not possible for protein/DNA complexes. As an alternative, the protein-specific Bradford colorimetric method, which modifies proteins chemically, was used. TFAM/site X complex crystals grew in 18-24% PEG 3350, 0.1-0.25 M sodium potassium tartrate, 0.1 M Hepes pH 7.5. TFAM/site Y native and derivative crystals grew in 23-28% PEG 3350, 0.08-0.2 M ammonium acetate, and 0.1 M Bis-Tris pH 6.5 or Hepes pH 7.5. For TFAM in complex with the large X-Y sequence the most promising conditions were: PEG 3350 24-26%, 0.1M NaCl, 0.1M Hepes pH7.5; PEG 400 33-36% $MgCl_2$ 0.25M, imidazole pH 8.0; isopropanol 25%, $MgCl_2$ 0.25 M, Hepes pH 7.5. The protein:DNA ratios used were 2:1 for TFAM/site X and TFAM/site Y. For TFAM in complex with the large X-Y sequence the ratios 4:1 and 6:1 were tested. TFAM/site Y and TFAM/site X crystals were cryoprotected by adding 10-20% of PEG 400 to the crystallization condition.

4.3 Crystal structure determination of the TFAM/Site Y complex.

4.3.1 X-ray diffraction.

X-ray crystallography is used to determine the atomic and molecular structure of a crystal and its components. When a beam of X-rays hits a crystal the electrons within the crystal diffract the incident X-rays into specific directions. By knowing the phase angle and intensity of the diffracted X-rays, it is possible to generate a three-dimensional representation of the electron density within the crystal. Therefore, when the X-rays hit the crystal, an elastic scattering is produced by the ordered atoms or scatterers, disposed in planes. A regular array of scatterers produces a regular array of spherical waves. Only waves that add constructively will contribute to the diffraction pattern and are those described by the Bragg's law (equation 1):

Equation 1: the Bragg's law

$$2d*\sin\theta = n\lambda.$$

In this equation, d is the spacing between diffracting planes, θ is the incident angle, n is any integer, and λ is the wavelength of the beam.

Therefore, this equation predicts the position in space of any diffracted X-ray. During data collection, the diffracted X-rays (or reflected beam) hit a detector, which collects the X-ray associated intensity (or reflection). The diffraction is in three dimensions; however X-ray diffraction intensities are collected using a bi-dimensional, flat detector. Therefore, a highly precise rotation of the crystal is required to collect all diffracted X-rays. This results in consecutive images that contain the reflection intensities (a spot in an image) at consecutive angles of oscillation. Therefore, from a two-dimensional data collection, a three-dimensional space, called the reciprocal space, that contains all intensities, is reconstituted. In other terms, the reciprocal space lattice is obtained. The coordinates of the reflections in the lattice are defined by the Miller indexes (hkl). These indexes define each point that constitutes an hkl reflection with a specific intensity I_{hkl}. The mathematic Ewald's sphere model shows the correlation between the crystal diffraction and its reciprocal space lattice construction. This sphere is a representation in reciprocal space of all the possible points where planes (reflections) could satisfy the Bragg's equation. After data collection, in order to define the Miller index for each reflection, the reflections have to be firstly indexed.

The indexing permits to identify the unit cell dimension and the symmetry of the crystal (its space group). The unit cell is the smallest unit of volume that contains all the structural and symmetry informations and that by translation produces a 3D pattern generating the crystal. After indexing, the spot area of each reflection is integrated giving the intensity value to each single reflection (I_{hkl}). It follows the scaling of the intensity by using a scale factor. The scale factor is the attempt to put all of the observations onto a common scale, by accounting for errors and inconsistencies caused by the instrument or the crystal. Once the scaled intensities (I) are obtained, they are truncated to amplitudes (F) by applying the premise of $I=F^2$. The amplitude together with the phase constitutes the structure factor. The structure factors are waves, expressed as complex numbers and, as waves, can be described by their module |F(h,k,l)| and their phase Φ(hkl). The Fourier transform (equation 2) uses the structure factors to calculate the three-dimensional distribution of electron density in the real space.

Equation 2: electron density function:

$$\boldsymbol{\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|F(h,k,l)|exp[-2\pi i * ((hx + ky + lz) - i\Phi(hkl))]}$$

Here ρ is the electron density; *x*, *y*, and *z* are the cartesian coordinates in the real space, *V* is s the volume of the unit cell, F_{hkl} the structure factor for a given set of Miller indices, $\boldsymbol{i}$ is an imaginary number and Φ the phase**.** Thus structure factors are complex numbers whose real part is related to the measured intensity, available from the experiment, while the imaginary part is related to the phase angle that cannot be directly measured. Although the previous Fourier transform requires both the intensities and the phases for each structure factor**,** from the experimental data it is possible to know only the module (|F(h,k,l)|) while phases (Φ(hkl)) are not recorded during data collection. This introduces the "phase problem", which needs to be solved. In order to overcome this problem different crystallographic techniques can be used, such as: multiple isomorphous replacement (MIR), single or multi-wavelengths anomalous dispersion (SAD or MAD) and molecular replacement (MR) methodologies.

4.3.2 Multiple Isomorphous Replacement (MIR).

This method consists in soaking the crystal with a heavy atom solution or co-crystallization with heavy atoms. The inclusion of heavy atoms should not modify the crystal unit cell composition and dimension, the derivate crystal has to be isomorphic to the native one. With this technique datasets of both the native and heavy atom derivative crystals are collected. By this it is possible to calculate a Patterson difference map to reveal the location of the heavy atoms in the unit cell (heavy atom sub-lattice). A Patterson map is a map of peaks that represent the interatomic distance vectors weighted by the product of the number of electrons from the concerned atoms. The Patterson difference map allows determining both the amplitudes and the phases of the reflections generated by the heavy atoms. Since the structure factor of the heavy atom derivative (Fph) of the crystal is the vector sum of the heavy atom alone (Fh) and the native crystal (Fp), then the phase of the native Fp and Fph vectors can be solved geometrically taking into account that:

$$\boldsymbol{Fph = Fp + Fh}$$

At least two isomorphous derivatives must be evaluated, since using only one will give two possible phases that will need to be discerned.

4.3.3 Single or Multi-wavelengths Anomalous Dispersion (SAD or MAD).

MAD and SAD techniques require the incorporation inside the macromolecule of "labelling" atoms (Se-Met or heavy atoms) that produce anomalous scattering when irradiated at specific X-ray wavelengths (Hendrickson 2014). Anomalous scattering happens when the X-ray energy is just below or equal to the energy of the electrons of the irradiated atom (absorbing edge). In this case the atom absorbs part of the X-ray energy changing the phase of the scattered wave. Such an anomalous scattering modifies the atomic scattering factor for the heavy atom present in the crystal. The atomic scattering factor, f(S), is a real quantity that falls off with resolution and, since it depends on the heavy atom absorption energy, it is usually wavelength-independent. Due to the effect of the anomalous scattering, the scattering factor is changed to:

$$\mathbf{f(S) = f_0(S) + f'(\lambda) + if''(\lambda)}$$

Here $f_0(S)$ is the usual energy-independent value (i.e. away from the absorption edge). The "f '(λ) + if "(λ)" term is the anomalous scattering correction after absorption and is wavelength (λ) dependent. f 'is the real, or dispersive, component of the correction (inflection point) and f " is the imaginary, or absorptive, component (peak). The symbol "S" refers to the diffraction vector whose length is characteristic of resolution (actually 1/d). The values of f ' and f " can be looked up in tables, or, better, can be extracted from empirical fluorescence scan before data collection. Due to the anomalous atom scattering, the diffraction is not anymore centrosymmetric since the Friedel's law (equivalent intensity for symmetrically equivalent pairs) is broken.

In the case of multi-wavelength anomalous diffraction, MAD, the same crystal is diffracted three times at three different X-ray wavelengths. These correspond to the energy of the absorption edge inflection point, the energy of the absorption edge peak (maximal absorption) and to an energy value away from absorption edge (remote energy) all referred to a specific anomalous scatterer. The three datasets obtained are processed, scaled and merged. It follows the anomalous scatterer position identification using Patterson function and phase determination. An advantage to use MAD instead of MIR is that in the former the isomorphism problem between crystals is eliminated.

It is also possible to obtain limited phase information from a single wavelength set at the maximal anomalous scatterer absorption. This technique is called single wavelength anomalous diffraction, SAD. In this case once determined the anomalous peak positions, density modification programs can be used to improve the phase and solve the structure.

4.3.4 Molecular replacement (MR).

Molecular replacement (MR) is another method of solving the phase problem in X-ray crystallography. MR is a technique that can be employed when exist a previously solved protein structure that is expected to be highly similar to the target protein from

which the diffraction data is obtained (Rossmann 1990). The similarity between the homologous and the target protein should be at least of the 35%, not below, in order to decide to use MR, otherwise the search becomes highly complicated with low success rate. The fundamental principle of MR is to *borrow* the phase from the homologous protein structure (model) and attribute it to the reflections of the target protein. In order to do this the Patterson function is calculated for the model protein and for the target.

By the MR approach the position of the homologous model in the crystal unit cell of the unknown structure is found. To define the correct orientation and position of the model inside the crystal asymmetric unit, three rotational angles (α, β, γ) and three translations axis (t_x, t_y, t_z) have to be found (6 dimensional search). If the asymmetric unit is constituted by N molecules then a total of 6N parameters are needed to define the solution. In order to reduce the calculations most programs split the search in two parts: a rotational search is performed firstly and then from the best solutions of the rotation search a translation search is tested. Rotation and translational search can be performed by following two different approaches: the traditional Patterson method and the probability approach (Evans and McCoy 2008). In the traditional Patterson method the rotation searches are based on the Patterson function, scoring the overlap between target and model Patterson maps in the region close to the origin, where the function is dominated by intramolecular self-vectors that are independent from translation. After rotation, in the translation search, the now correctly oriented known model can be correctly positioned by translating it to the correct co-ordinates within the asymmetric unit. This is accomplished by moving the model, calculating a new Patterson map, and comparing it to the target Patterson map. The translation search is not required for space group P1 where there is no translational symmetry.

The modern probability approach uses the 'Maximum-likelihood' statistical methods. This method take into account the similarity grade between the model and the target protein in order to be able to estimate the error probability that is due to a low similarity of the model and not to a wrong solution. In a maximum-likelihood rotation search, the model is rotated sequentially on an angular grid through the unique angular space and the orientation that predicts the data with highest probability is selected. As for the rotation search, the translation search for any given orientation consists in

placing the model sequentially in grid points throughout the translationally unique volume of unit cell. A formula based on the Rice function (Read 1990) gives the probability for each putative translation, from which the most likely is selected as the solution to the translation problem.

If the MR process gives one solution that clearly stands out in scores from the next best result, it is likely to be correct. The principal test for a correct and useful solution is that the maps calculated from the solution model should show new and plausible information that was not present in the model. This might be side chains or loops that were different in the model and are present in the target protein.

4.3.5 Phase optimization and structure validation.

Once the model is correctly placed it follows a cyclic process of model optimization at the real space and phase improvement at the reciprocal space. This is assessed by a good fitting between the experimental data (F_{obs}) and the calculated model (F_{calc}). This is achieved by iterative cycles of manual model building combined with automatic refinement of the phases. An indicator of good fitting between the model and the experimental data is the R-factor. The R-factor (also called residual factor, reliability factor, or R-value) is a measure of the agreement between the crystallographic model and the experimental X-ray diffraction data. It is defined by the following equation:

$$R = \frac{\sum_{hkl} |F_{hkl}^{obs} - F_{hkl}^{calc}|}{\sum_{hkl} F_{hkl}^{obs}}$$

Where F is the structure factor and the sum extends over all the reflections measured and their calculated counterparts. The R_{Free} parameter is also used, to assess possible over fitting of the data. R_{Free} is computed according to the same formula given above, but on a small, random sample of data that are set aside for the purpose and never included in the refinement. R_{Free} will always be greater than R-factor because the model is not fitted to the reflections that contribute to it but the two statistics should be similar because a correct model should predict all the data with uniform accuracy. If

the two statistics differ significantly then that indicates the model has been over-parameterized. The quantities R_{sym} and R_{merge} are similarly used to describe the internal agreement of measurements in a crystallographic data set (Kleywegt and Brunger 1996).

The last step of macromolecule structure refinement is the stereochemical validation by checking interatomic distances, main chain torsion angles, *Ramachandran* outliers, correct rotamers conformation and water molecules position.

4.3.6 TFAM/Y22 data processing and structure determination.

The datasets of TFAM/site X and site Y complexes were collected at beamline ID23-2 of the European Synchrotron Radiation Facility (ESRF), Grenoble (France). In the case of TFAM in complex with Site Y, brominated derivative crystals diffraction was collected with Friedel's pairs in the same image. Native and derivative TFAM/Y22 diffraction data were processed with XDS program (Kabsch 2010) . Intensities were scaled with SCALA (Evans 2006) and truncated to structure factors F with C-truncate, both from CCP4 suite (Winn, Ballard et al. 2011). The phase problem was solved by MR technique employing TFAM/LSP (PDB: 3TQ6) as a searching model. Phaser as implemented in the Phenix suite was used as MR program (McCoy, Grosse-Kunstleve et al. 2007; Adams, Afonine et al. 2011). The model was built manually with COOT (Emsley 2004) and automatically refined with Phenix.refine (Afonine, Grosse-Kunstleve et al. 2010) and BUSTER (Branden and Jones 1990). TLS and non-crystallographic symmetry (NCS) refinement were applied together with Watson and Crick double strand DNA restrains. For the derivative data, the anomalous signal was estimated by using XPREP (Sheldrick 2008) while for the anomalous atoms positions determination ANODE was employed (Sheldrick 2008).

4.4 DNA physical properties and molecular dynamics computational analysis.

The early representations of DNA presented the molecule as a static straight rod. In reality the DNA double helix is a highly dynamic structure. It has three significant degrees of freedom; bending, twisting, and compression. DNA molecules often have a

preferred direction to bend (anisotropic bending). Protein/DNA interactions and recognition depend crucially on the local structural variations of DNA, in particular in bending and torsion (twisting) degrees of freedom. In general DNA flexibility is correlated to its sequence-dependent properties. The geometry of a dinucleotide inter-base pair (base pair step) can be described by 6 parameters that define six movements: shift, slide, rise, tilt, roll, and twist. Translational (Shift, Slide, and Rise) and rotational (Tilt, Roll, Twist) parameters are defined below (Figure_20):

Shift: Translation around the X-axis.

Slide: Translation around the Y-axis.

Rise: Translation around the Z-axis.

Tilt: Rotation around the X-axis.

Roll: Rotation around the Y-axis.

Twist: Rotation around the Z-axis.

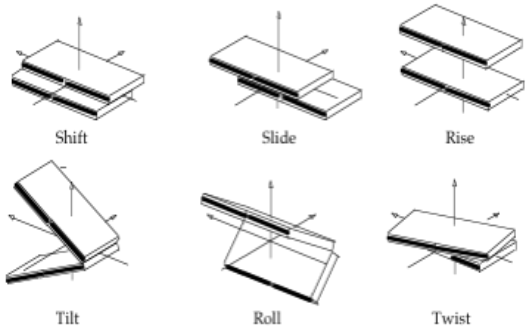


Figure_20: Base-pair parameters.

These parameters determine the global DNA fragment geometry.

In this thesis molecular dynamics (MD) simulations were employed to calculate DNA equilibrium and stiffness values of each parameters for each base pairs in the three ligand molecules (site Y, site X, LSP) (Drsata, Perez et al. 2013). MD calculations were performed in collaboration with Dr. Federica Battistini, from Prof. Modesto Orozco laboratory, Institut de Recerca Biomèdica (Barcelona). MD simulations allow studying the physical movements at atomic level in the protein-DNA complex in time, giving a view of the dynamical evolution of the system. The trajectories are determined by numerically solving Newton's equations of motion for a system of interacting particles where forces between the particles and their potential energies are calculated using interatomic potentials or molecular mechanics force fields.

For each ligand, it was run a 500ns MD simulation using AMBER 12 package and parmBSC1 force field (Ivani, Dans et al. 2016). The naked sequences was placed in a box of water, using the TIP3P (Jorgensen, Chandrasekhar et al. 1983) water model with a minimum 10-Å buffer solvation layer beyond the solute, and the negatively

charged DNA was neutralized with Na^+ ions (Smith and Dang 1994). 0.5 M concentration was reached by adding salt (NaCl). All structures were first optimized, thermalized and pre-equilibrated for 1 ns using a standard protocol (Lavery, Zakrzewska et al. 2010) and subsequently equilibrated for an additional 10-ns period. DNA deformability along these six directions was calculated and described by the associated stiffness constant matrix (Drsata, Perez et al. 2013). From the ensemble of MD simulations, the covariance matrix describing the deformability of the B-DNA parameters for a given DNA fragment is computed and is inverted to generate the 6x6 stiffness matrix for each fragment. Pure stiffness constants corresponding to the six parameters mentioned above (kshift, kslide, krise, ktilt, kroll and ktwist) are extracted from the diagonal of the matrix. The total stiffness (K_{tot}) is obtained as a product of these six constants and provides a rough estimate of the flexibility of each base pair step. The deformation energy was calculated using a mesoscopic energy model (Drsata, Perez et al. 2013; Portella, Battistini et al. 2013), which is based on a harmonic approximation to describe deformability along DNA helical parameters. The equilibrium geometry and stiffness force constants were extracted from a dataset built from long all-atoms MD simulations of short oligonucleotides in water using the parmbsc1 force field. The equilibrium geometry and stiffness force constants were extracted from a dataset built from long all-atoms MD simulations of short oligonucleotides in water using the parmbsc1 force field. For each ligand, the reference structure of the protein-bound DNA conformation was taken directly from the X-ray crystal structure.

4.5 Gel Electrophoresis techniques and protocols.

Gel electrophoresis is a method for separation and analysis of macromolecules (DNA, RNA and proteins), based on their size, shape and charge. In this method both an electric field applied along a gel and the dimensions of the gel pores determine and influence the mobility of a polar sample. Two different kind of gel material are mostly used for macromolecules separation: polyacrylamide (polyacrylamide gel electrophoresis, PAGE) and agarose. In order to increase the resolution power of the technique and to be able to separate a mixture of small macromolecules, higher

percentage of polyacrylamide or agarose have to be used. Another important element that affects the macromolecules migration is the electrophoresis buffer, since the charge of the sample can vary in function of the pH and ionic strength of the electrophoretic buffer employed. The gel preparation can include denaturing agent or not (native gel) depending on the goal of the experiment. In this thesis work, several separation techniques were used:

- SDS-PAGE in Tris-Tricine electrophoresis buffer was routinely used to run protein samples after their expression and to assess the protein purity level after each purification step. The gels were stained with Coomassie blue.

- Non-denaturing PAGE was used to analyse TFAM/DNA complexes formation by electrophoresis mobility shift assay (EMSA). In this technique, a constant amount of DNA is generally titrated with an increasing amount of protein at higher concentrations. During electrophoresis the bound protein is separated from free DNA, which indicates the proper formation of the complexes.

For EMSA of TFAM/site Y and /site X complexes, the protein and the DNA were mixed in the TFAM-stabilising buffer used during purification of the protein (50 mM HEPES pH7.5, 750 mM NaCl, 5mM DTT). Since high concentration of NaCl impairs complex formation, the mixture was dialyzed stepwise at 4° C with buffers containing less salt amounts. To this end, the sample was dispensed in 20 ul dialysis buttons (Hampton Research) that were subsequently plunged in buffer A (50 mM Hepes pH7.5, 500 mM NaCl, 5 mM DTT), buffer B (50mM Hepes pH7.5, 250 mM NaCl, 5mM DTT) and buffer C (50mM Hepes pH7.5, 100 mM NaCl, 5mM DTT). The buttons contained a constant concentration of DNA (5 µM) and increasing TFAM amounts. 10ul of the dialyzed complexes were supplemented with a loading dye and the reactions mixtures were loaded onto 11% polyacrylamide native gel buffered with 0.5x TBE (0.5x). 8x10 cm gels were run on a Miniprotean II (Biorad) at 10 v/cm. The gels were next stained with SYBR safe (Molecular Probes) and scanned for fluorescence with a Typhoon 8600 imager (Marshall Scientific)

By labelling the DNA with a radio-isotope (^{32}P), the DNA band intensity can be quantified to provide quantitative information on protein-DNA binding. K_d

measurements using radio-isotope (^{32}P) labelled DNA were realized in triplicate with 1nM DNA and increasing concentrations of TFAM with the help of the radio-competent person in the laboratory. After migration, the gels were dried, exposed to a phosphoimager screen and quantified with a Typhoon 8600.

The titration of radio-isotope (^{32}P) labelled DNA bands was described by a modified form of the Hill equation, that can compensate for deviations from ideal conditions, including incomplete binding caused by loss of protein sample at low concentrations, cooperative binding, oligomerization, or other more complex mechanisms of binding:

$$\bar{Y} = b + \left[\frac{m - b}{1 + (K_d / [P_t])^n} \right]$$

where $\bar{Y}$ is the fractional saturation of the binding site, P_t is the total protein concentration used, K_d the equilibrium dissociation constant, m and b are normalization factors that represent the fraction of bound DNA at the upper and lower asymptotes of the titration and n is the Hill coefficient. The Hill coefficient measures the cooperativity of binding. Deviations from unity may indicate cooperative binding of multiple proteins or binding reactions that have not reached equilibrium. Small deviations from integer values are commonly caused by the adherence of protein or DNA to the equilibration vessel. This phenomenon occurred with TFAM, which, in addition, aggregated around the 500nM-1µM range, both effects being suppressed to some extent by the addition of 0.01% v/v of Tween 20 to the reaction buffer.

Competitions reactions using 5' fluorescein-labelled and unlabelled DNAs were also performed to compare the differential affinities between TFAM ligands. The gels containing the fluorescein-DNA were scanned directly after electrophoresis with Typhoon 8600 imager.

4.6 Isothermal Titration Calorimetry (ITC) technique and protocol.

Isothermal Titration Calorimetry (ITC) is a biophysics technique used in quantitative studies of a wide variety of biomolecular interactions. It works by directly measuring at constant temperature and pressure the heat that is either released or absorbed during a biomolecular binding event. Measuring heat transfer during binding enables accurate determination of binding constants (K_d), reaction stoichiometry (n), enthalpy (ΔH) and entropy (ΔS). This provides a complete thermodynamic profile of the molecular interaction.

An ITC apparatus is constituted by two cells thermally insulated, one of which contains water and acts as a reference cell, and the other contains the sample. The reference cell and the sample cell are set to the desired experimental temperature. The microcalorimeter needs to keep these two cells at exactly the same temperature. A heat sensing device detects temperature difference between the cells and activates the heaters that provide to restore the same temperature in both of them.

An ITC experiment consists in serial injection of small ligand volume (2-50μl) into the sample cell containing the protein of interest. If there is a binding of the ligand to the protein, heat changes (*qi*) are detected and measured. This heat change is proportional to the enthalpy of the reaction:

$$\boldsymbol{qi = v \times \Delta H \times \Delta Li}$$

Where *qi* is the heat change; *v* is the cell volume and ΔLi is the increasing ligand concentration in each injection *i*. For each injection a peak of heat change is registered in a graph. In the case of exothermic reactions downward peaks are collected (Figure_21); while in the case of endothermic reaction upward peaks are plotted. For each ligand injection, the protein gets more and more saturated, less binding of the ligand occurs and the heat change starts to decrease until ultimately the sample cell contains an excess of ligand versus protein, bringing the reaction towards saturation. The area of each peak is then integrated and plotted versus the molar ratio of ligand to protein. The resulting isotherm can be fitted to a binding model from which the affinity (K_d) is derived. The molar ratio at the center of the binding isotherm gives the reaction stoichiometry (Figure_21).

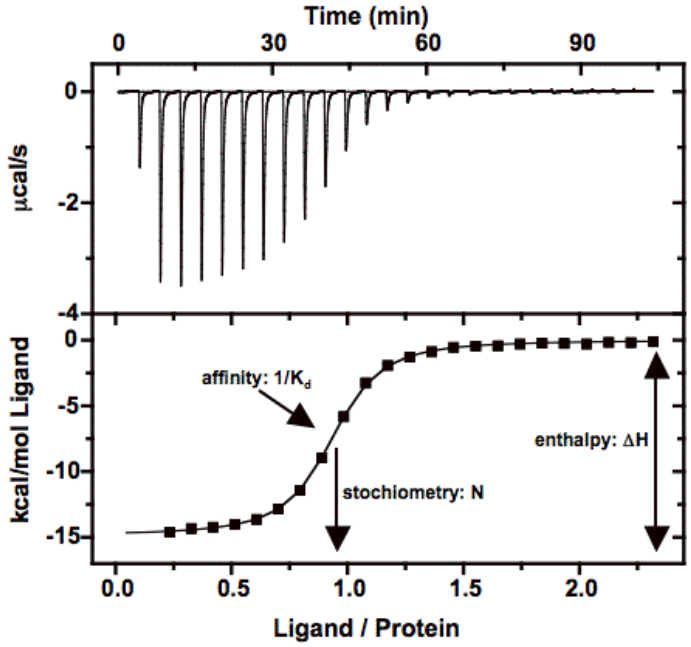


Figure_21. ITC measurement of an exothermic reaction.
In the upper graph, exothermic reaction heat change peaks versus time are shown. The graph in the bottom represents the integration of the area of each peak. The resulting isotherm is then fitted to a binding model that permits to extrapolate the stoichiometry of the reaction (n) the enthalpy variation (ΔH) and the affinity constant ($1/k_d$). This picture was taken from http://www2.mrc-lmb.cam.ac.uk/groups/hmm/techniqs/ITC.html.

ITC experiments of this thesis work were performed at the "Centre Científics i Tecnològics" of Barcelona University, in collaboration with Dr. Rafel Prohens. *VP-ITC microcalorimeter system* was employed for samples measurements, while VPViewer2000 was the program used for data collection. The experiments were performed at 20ºC. TFAM and the three DNAs alone (site Y, site X and LSP) where dialyzed with a dialysis tubing pore of 3500 Da in the same buffer (150mM NaCl, 50mM Hepes pH7.5) changed four times in 36 h at 4ºC. Before starting the experiment, TFAM (6-9uM) was loaded in the sample cell at 20ºC. Once the system reached the equilibrium, the DNA injection started. In the case of site Y and LSP 30 injections of 10ul, 20s/injection were performed; while in the case of site X the protein was titrated with 15 injections of 20ul, 30s/injection. The three maximal DNAs concentration were around 40-50uM. The data were analyzed with the program "MicroCal Origin". Each experiment was repeated three times.

4.7 Size Exclusion-Multi Angle Laser Light Scattering (SEC-MALLS) technique and protocol.

Multi angle laser light scattering (MALLS) is a technique to determine the absolute molar mass of particles in solution. In case that several macromolecular assemblies are present in the sample, flow-mode MALLS (SEC-MALLS), in which the MALLS instrument is coupled to a size exclusion column, can be used to separate the different species. A SEC-MALLS measurement requires a set of ancillary elements. The most important among them is a laser source producing a collimated beam of monochromatic light that illuminates a region of the sample. Another required element is an optical cell to hold the sample. Cells incorporate tools to permit measurement of flowing samples. Furthermore, a size exclusion column with a correct resolution range is needed together with the following flow detector: UV detector, light scattering (LS) detectors located at different angle with respect to the sample, and a refracting index (RI) detector. The amount of light scattered is directly proportional to the product of the weight-average molar mass and the solute concentration: LS ~ M *x* c. This relationship is based on Zimm's formalism of the Rayleigh-Debye-Gans light scattering model for dilute polymer solution. The relation between the excess scattered light and MW is given by equation 3:

Equation 3: Debye-Zimm formalism for R(θ):

$$\frac{K^* c}{R(\theta)} = \frac{1}{MwP(\theta)} + 2A_2 c$$

Where $R(\theta)$ is the excess intensity of scattered light at angle θ; c is the sample concentration (g/ml); Mw is the weight-average molecular weight (molar mass); A_2 is the second virial coefficient (ml-mol/ g^2); $P(\theta)$ is the form factor (describes the angular dependence of scattered light); K^* is an optical parameter equal to $4\pi^2 n^2 (dn/dc)^2 / (\lambda_0^4 N_A)$. In this last equation "n" is the solvent refractive index and "*dn/dc*" is the refractive index increment of the sample respect to the baseline (buffer only). "N_A" is the Avogadro's number while "λ_0" is the wavelength of the scattered light in vacuum (cm).

In this thesis the SEC-MALLS experiments were performed by using a DAWN EOS Multi-angle light scattering instrument coupled to an Optilab rEX refracting index detector (Wyatt Technology). The molecular weight of the different species separated by the column was calculated employing ASTRA software (Wyatt Technology). ASTRA uses scattering data acquired across the entire peak of the sample and the Mw is determined with Debye fitting method. A *dn/dc* value of 0.178 mL/g and of 0.185 mL/g was used for all TFAM-DNA complexes and protein alone respectively; while a dn/dc value of 0.166 mL/g was used for the DNA alone. The apparatus calibration was performed using Ovalbumin (Calibration constant obtained: 9.2). The experiments were performed at two different fluxes: 0.5 ml/min and 0.3 ml/min, both at RT. The TFAM-DNA complexes were measured by loading a ratio of 2 proteins: 1 DNA into the gel filtration column. 100 μL of samples was loaded into the column. The sample buffer was the same for the three analysed complexes, the protein and the DNA: 100 mM NaCl, 50 mM Hepes pH7.5, 5 mM DTT. For the three complexes and the protein alone the protein concentration was measured by Bradford method. In the case of TFAM/site Y complex, TFAM concentration was 9 mg/ml (351.5 uM), while for TFAM/site X complex TFAM concentration was 8 mg/ml (321.5 uM). In the case of TFAM in complex with LSP two different TFAM concentrations were tested: 5.6 mg/ml (218.7 uM) and 10 mg/ml (390.6 uM). The protein alone concentration was 2.5mg/ml (97.6uM) and the DNA alone had a concentration of 200 uM.

4.8 Analytical Ultra Centrifugation technique and protocol.

Analytical ultracentrifugation (AUC) is a versatile and powerful method for the quantitative analysis of macromolecules in solution. AUC can be used to study macromolecules in a wide range of solvents and over a wide range of solute concentrations. It provides useful information on the size and shape of molecules in solution with very few restrictions on the sample or the nature of the solvent. The fundamental requirements for the sample are: it should have an optical property that distinguishes it from other solution components; it should sediment or float at a reasonable rate at an experimentally achievable gravitational field; and it should be chemically compatible with the sample cell. The solvent must be chemically

compatible with both the sample cell and the optical systems. The range of molecular weights suitable for AUC exceeds that of any other solution technique, from a few hundred Daltons (e.g. peptides, dyes, oligosaccharides) to several hundred-million Daltons. There are three optical systems available for the analytical ultracentrifuge: absorbance, interference and fluorescence that permit precise and selective observation of sedimentation in real time. The analytical rotor holds sample containers commonly called cells. Each cell contains a centerpiece with chambers (called channels) that contain the liquid samples. The centerpiece, in turn, is sealed between windows, so that the light passes through the channels allowing the detection of the cell content. Essentially there are two AUC complementary approaches: sedimentation velocity (SV) and sedimentation equilibrium (SE). In sedimentation velocity the macromolecules are subjected to a high centrifugal force and fractionated according to their differences in buoyant mass and shape. In sedimentation equilibrium a lower centrifugal field is applied. By this technique is possible to study the interaction of the macromolecules: mass, assembly stoichiometry and association constant when the equilibrium is reached.

In this work the sedimentation velocity method was employed. In sedimentation velocity experiments act three essential forces:

1) the Gravity force $F_{sed} = M_p\omega^2 r$, where M_p is the protein mass, $\omega^2$is the rotor speed in radians and r is the distance from the center of the rotor;

2) the buoyancy force $F_b = Mp(1 - \bar{v}\rho$, where M_p is the protein mass, $\bar{v}$ is the protein partial specific volume and ρ is the solvent density;

3) the frictional force developed by the motion of the particle through the solvent is given by fv, where f is the frictional coefficient and v is the velocity.

Balancing these three forces is obtained the sedimental coefficient s, defined as the ratio of the velocity to the centrifugal field:

$$s \equiv \frac{v}{\omega^2 r} = \frac{M_b}{f}$$

Diffusion causes the sedimenting boundary to spread with time. The translational diffusion coefficient, D, is described by the equation:

$$D = RT/N_a f$$

Where R is the gas constant, T is the absolute temperature, and N_a is the Avogadro's number. s and D values can be extracted from AUC data. By taking the ratio s/D, the frictional contribution to these parameters is removed and the result is proportional to the buoyant molar mass, Mb, through the Svedberg equation:

$$s/D=M_b/RT$$

Svedberg equation constitutes the basis of AUC technique. In this thesis work TFAM/site Y, site X and LSP complexes were analyzed at two different protein:DNA ratios, 4:1 and 2:1. In both cases the concentration of TFAM for each complex was calculated using the Bradford method. For the 2:1 ratio, the TFAM concentration was 13 mg/ml (508 uM); while for the 4:1 ratio, the concentration of TFAM was 11 mg/ml (429 uM). The concentration of TFAM alone was of 2 mg/ml (78 uM). The site Y, site X and LSP dsDNAs the concentration was 200 uM. All samples were equilibrated in 100 mM NaCl, 50 mM Hepes pH 7.5, 2 mM TCEP. The experiments were performed at 48 krpm and 20ºC in a Beckman Optima XL-I analytical ultracentrifuge (Beckman-Coulter) equipped with absorbance and interference optics, using an An50Ti rotor. Absorbance scans (0.003 cm step size) were taken at 295 nm. Differential sedimentation coefficient distributions, c(s), were calculated by least squares boundary modeling of sedimentation velocity data using the program SEDFIT (Schuck 2000; Schuck, Perugini et al. 2002). The only requirement for this analysis to be accurate is that all association reactions are at equilibrium prior to the start of sedimentation. The experimental values obtained from this analysis was corrected for solvent composition and temperature to obtain the corresponding standard S values ($S_{20,w}$) using the public domain software SEDNTERP, retrieved from the RASMB server (Laue, Shah et al. 1992).

5. RESULTS

5. RESULTS

5.1 X-ray Crystallography of TFAM in complex with long site Y-X DNA and site X and site Y short fragments.

5.1.1 TFAM protein production**.**

The cDNA that included the TFAM genetic sequence corresponding to the mature protein (residues 43 to 246, without the mitochondrial targeting sequence) was cloned into plasmid pET28. This plasmid adds a tag of 6 histidines at the C-terminus of the protein to facilitate purification by affinity chromatography. The protein was expressed in BL21 (DE3) pLys-S (see section 4.1.1 for a detailed protocol). This strain is lysogenic for λ-DE3, which contains the T7 bacteriophage gene I, encoding T7 RNA polymerase. It also contains a plasmid, pLys-S, which carries the gene encoding T7 lysozyme. T7 lysozyme lowers the background expression level of target genes under the control of the T7 promoter but does not interfere with the level of expression achieved following induction by IPTG. Previous work in the lab established optimal conditions for protein solubilisation during cell lysis, which consists in sonication of the bacterial cells in a buffer (lysis buffer) that contains 750 mM NaCl, 20 mM imidazole, 50 mM Hepes pH 7.5, 1 mM DTT, DNAse, and protease inhibitors (see section 4.1.2). After supernatant clearance by centrifugation, a two-step purification protocol follows. First the filtrated supernatant is passed through a nickel-chelating column that traps histidine-tagged proteins, which are eluted by a gradient of imidazole. The purest eluting fractions are pooled, concentrated and injected to a gel filtration column (see section 4.1.2 for details). The protein is obtained with a level of purity higher than 95% as assessed by SDS-PAGE (Figure_22). The final yield after protein purification is 10 mg of TFAM per 1L of culture.

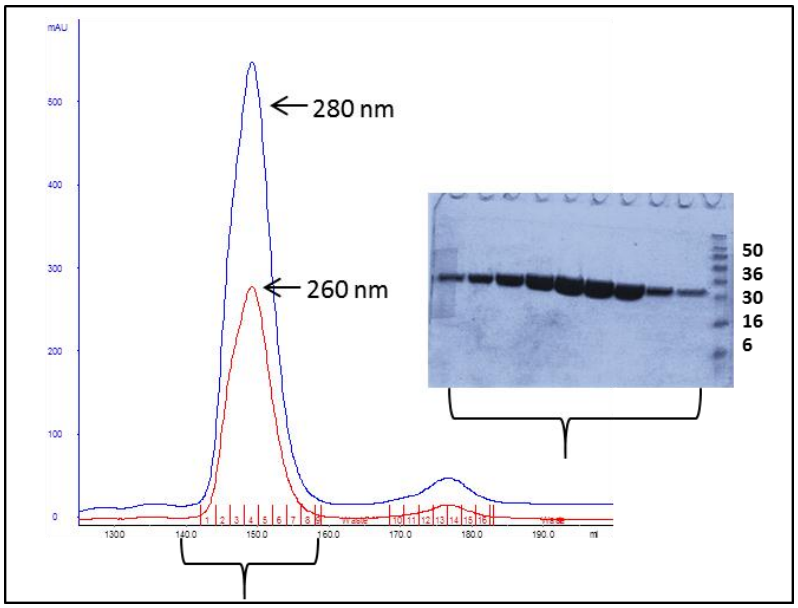


Figure_22. TFAM purification chromatogram and SDS-gel.
TFAM is regularly purified with superdex 75 26-60 column (GE Healthcare). The absorbance at 280 and 260 nm are shown in blue and red lines respectively. The eluted fractions from the first peak are loaded into a poly-acrylamide denaturing gel (on the right). As it is possible to see just one pure band of TFAM is obtained that migrate as a molecule of 30 kDa. The second peak in the elution graph corresponds to a degradation product of TFAM. It is noteworthy the slight shoulder at the beginning of the main peak, indicating heterogeneity of the sample.

5.1.2 Crystallization of TFAM/site Y-X, TFAM/site Y and TFAM/site X complexes.

The first crystallographic structure of TFAM was described in 2011 by two different groups (Ngo, Kaiser et al. 2011; Rubio-Cosials, Sidow et al. 2011). These reports showed TFAM crystallized in complex with its cognate binding sequence at the light strand promoter (LSP).

In order to know if the binding mechanism of TFAM at binding sites Y (nt 233-260) and X (nt 276-303) from the control region (Figure_23) was the same as in LSP, the corresponding protein/DNA complexes were crystallized. At the mtDNA control region these two sites map between CSBI and CSBII (see Introduction). Thus two different experiments were planned: i) crystallization of TFAM in complex with a 71 bp DNA sequence ("Y-X long fragment") that includes both binding sites: Y and X and the 15 bp intermediate region; and ii) crystallization of TFAM in complex with site Y or site X, separately ("site Y " and "site X"). This kind of study should permit

not only unveil a possible alternative mechanism of TFAM binding to the individual recognition sites, but also identify the type of binding in a longer fragment were both sites may be bound by TFAM simultaneously.

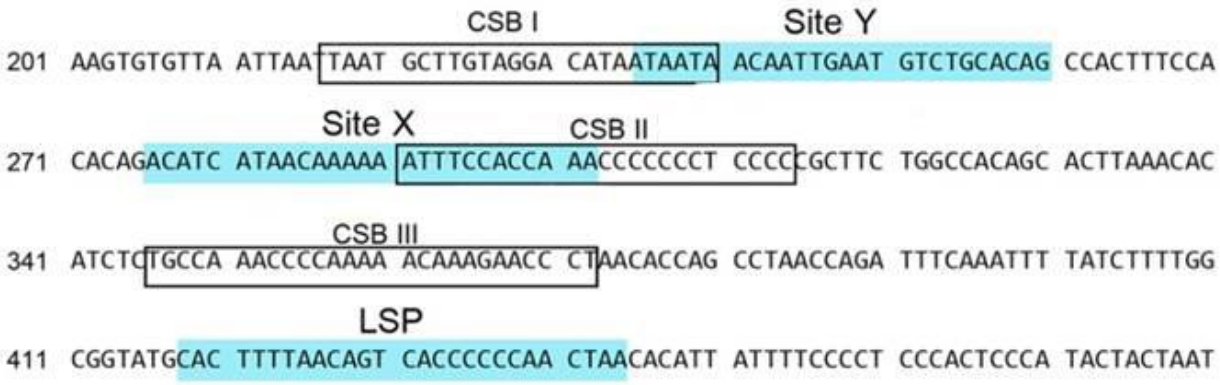


Figure_23. Fragment of the mtDNA control region that includes site X and site Y.
MtDNA TFAM-protected regions, found by Fisher et al.in 1987 by footprinting analysis, are shown in light blue (Fisher, Topper et al. 1987). The name of each TFAM binding site is indicated together with the location of the conserved sequences blocks I to III (CSBI-III).

Taking into account i) the footprinting analysis which showed the DNase I protection of HSP, LSP, X and Y sequences by TFAM (Figures_23); ii) the alignment of LSP, HSP, X and Y proposed by Fisher et al in 1987 (Fisher, Topper et al. 1987) (Figures_17) and iii) the sequence of LSP (5'TAACAGTCACCCCCCAACTAAC3') that was previously crystallized in complex with TFAM (Ngo, Kaiser et al. 2011; Rubio-Cosials, Sidow et al. 2011), a list of DNAs was designed to perform crystallization trials in complex with TFAM. This list included sequences that contained the Y-X long fragment, and site Y and site X separately, with blunt or cohesive ends, and of different lengths. These DNAs include the complete TFAM DNase I protected region, or part of it. In Table_1 the Y-X long fragment sequences are reported, whereas Table_2 reports the short fragments site Y and site X.

Table_1. Y-X long fragment sequences tested during crystallization trials in complex with TFAM. Site Y and site X regions are indicated in light blue.

SiteY-X71bp	5′ TAATAATAACAATTGAATGTCTGCACAGCCACTTTCCACACAGACATCATAACAAAAAATTTCCACCAAAC 3′ 3′ ATTATTATTGTTAACTTACAGACGTGTCGGTGAAAGGTGTGTCTGTAGTATTGTTTTTTAAAGGTGGTTTG 5′
SiteY-X65bp	5′ TAACAATTGAATGTCTGCACAGCCACTTTCCACACAGACATCATAACAAAAAATTTCCACCAAAC 3′ 3′ ATTGTTAACTTACAGACGTGTCGGTGAAAGGTGTGTCTGTAGTATTGTTTTTTAAAGGTGGTTTG 5′
SiteY-X65cbp	5′ TAACAATTGAATGTCTGCACAGCCACTTTCCACACAGACATCATAACAAAAAATTTCCACCAAAC 3′ 3′ GATTGTTAACTTACAGACGTGTCGGTGAAAGGTGTGTCTGTAGTATTGTTTTTTAAAGGTGGTTT 5′
SiteY-X65c2bp	5′ GTAACAATTGAATGTCTGCACAGCCACTTTCCACACAGACATCATAACAAAAAATTTCCACCAAAC 3′ 3′ ATTGTTAACTTACAGACGTGTCGGTGAAAGGTGTGTCTGTAGTATTGTTTTTTAAAGGTGGTTTGC 5
SiteY-X58$_A$bp	5′ CAATTGAATGTCTGCACAGCCACTTTCCACACAGACATCATAACAAAAAATTTCCACC 3′ 3′ GTTAACTTACAGACGTGTCGGTGAAAGGTGTGTCTGTAGTATTGTTTTTTAAAGGTGG 5
SiteY-X50bp	5′ TAATAACAATTGAATGTCTGCACAGCCACTTTCCACACAGACATCATAAC 3′ 3′ ATTATTGTTAACTTACAGACGTGTCGGTGAAAGGTGTGTCTGTAGTATTG 5′

Table_2. Site Y and site X short fragments sequences tested for crystallization in complex with TFAM. *Sequence highlighted in red gave the best diffracting crystals (highest resolution), from which the TFAM/site Y and TFAM/site X structures were determined.

SITE X SEQUENCES	
X28-A	5′ ACATCATAACAAAAAATTTCCACCAAAC 3′ 3′ TGTAGTATTGTTTTTTAAAGGTGGTTTG 5′
X24-A	5′ ATAACAAAAAATTTCCACCAAACC 3′ 3′ TATTGTTTTTTAAAGGTGGTTTGG 5′
X22-A	5′ TAACAAAAAATTTCCACCAAAC 3′ 3′ ATTGTTTTTTAAAGGTGGTTTG 5′
X22-A2	5′ ATAACAAAAAATTTCCACCAAC 3′ 3′ TATTGTTTTTTAAAGGTGGTTG 5′
X22c*	**5′ TAACAAAAAATTTCCACCAAAC 3′** **3′GATTGTTTTTTAAAGGTGGTTT 5′**
X22-Ac2	5′ TAACAAAAAATTTCCACCAAAC 3′ 3′ TTGTTTTTTAAAGGTGGTTTGA 5′
SITE Y SEQUENCES	
Y22Ac	5′ TAACAATTGAATGTCTGCACAG 3′ 3′ TTGTTAACTTACAGACGTGTCA 5′
Y22*	**5′ TAACAATTGAATGTCTGCACAG 3′** **3′ ATTGTTAACTTACAGACGTGTC 5′**
Y22B	5′ AACAATTGAATGTCTGCACAGC 3′ 3′ TTGTTAACTTACAGACGTGTCG 5′

5.1.2.1 Crystallization of TFAM in complex with site Y-X long sequence.

TFAM site Y-X long sequence complex (TFAM/Y-X complex) formation was analysed by electrophoretic mobility shift assay (EMSA), before starting crystallization trials. For EMSA experiments TFAM protein stock was diluted in 500 mM NaCl, 50 mM Hepes pH 7.5, 1 mM DTT buffer and added, in appropriate amounts to DNA diluted in water. Dilution buffer (50 mM Hepes pH 7.5 and 1mM DTT) was then added up to the reaction to a final volume of 10 ul. By this, 0.2 uM DNA was titrated with increasing TFAM concentration, from 0.2 uM to 6.4 uM. In Figure_24 a representative EMSA performed with the long Y-X sequence ("SiteY-X65bp", Table 1) in complex with TFAM is shown. A total DNA shift is observed at a ratio 4TFAM per1 DNA molecule, suggesting that TFAM may bind to site X and site Y as a dimer. Multiple shifted DNA bands appear at increasing TFAM concentrations that correlate with its tendency to oligomerize when present in huge (uM range) amount in a gel grid confined space.

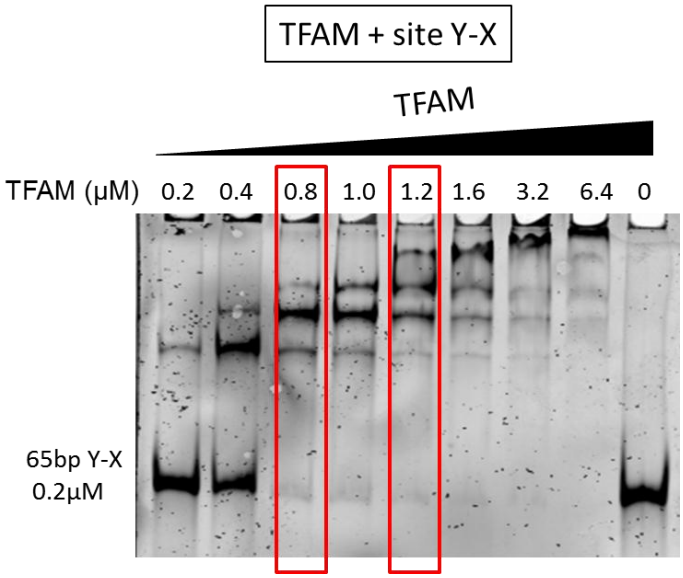


Figure_24. Gel shift assay of TFAM in complex with site Y-X longer fragment.
In the first lane on the right, indicated with a "0", a 65bp long Y-X DNA fragment was loaded (SiteY-X65bp in Table 1). All lanes contain 0.2 uM DNA that was titrated, from the left to the right, with increasing amounts of TFAM. The final concentration of TFAM in each lane is indicated above the gel. The red rectangles indicate the protein:DNA ratios of 4:1 (TFAM 0.8 uM) and 6:1 (TFAM 1.2 uM).

Since the gel showed full binding of the free DNA at a protein:DNA ratio of 4:1 (left vertical rectangle in Figure_24) , this ratio was used for crystallization. In addition, to make sure the protein bound all DNA, an excess of protein was also tried by using the 6:1 ratio (right vertical rectangle in the same figure). TFAM/X-Y complexes were prepared by mixing protein and DNA at the appropriate amounts as explain in Materials and Methods section 4.2.1. In general terms the protein/DNA complexes are stabilized by electrostatic interactions that can be competed by high salt concentrations disrupting the complex. TFAM was purified at 750 mM NaCl. Therefore, to promote complex formation with pure, purchased DNA, the protein/DNA mixture was stepwise dialyzed against buffers with lower salt concentration to a final concentration of 20 mM NaCl. TFAM/X-Y complex in a final amount of 8 until 15 mg/ml was then automatically dispensed in 96 well nanodrop plates containing crystallization conditions from Hampton by using the robots available at the Automated Crystallography Platform (PAC) at the Barcelona Science Park. Initial crystallization screenings included PAC screens 1, 2, 3, 8, 9, 10 and 16. After inspection of the crystallization drops, additional 96-well plates with customized optimization screens were tested to find more favourable crystallization conditions. After these first trials, the most promising conditions that presented change of phase, nucleation or microcrystals were manually scaled up to drops of microliter volumes by using hanging and sitting drop 24 well plates. The SiteY-X 65 and 50 bp sequences (Table 1) were the only ones that yielded crystals. In order to improve the crystals obtained from the 24 well plates, different reservoir and drop volumes were additionally tried (reservoir volume: 500 ul, 700 ul, 900 ul; drop volume: 2ul and 3ul). In order to reduce nucleation and promote crystals growing, different ratios (1:1, 1.5:1 and 2:1) of paraffin oils and silicon oils were added to 500 ul of reservoir until a volume of 800 ul. The microseeding technique was also tried. Unfortunately TFAM in complex with the site Y-X long fragment, even after multiple attempts, yielded only microcrystals (Figure_25) that diffracted like DNA fibres.

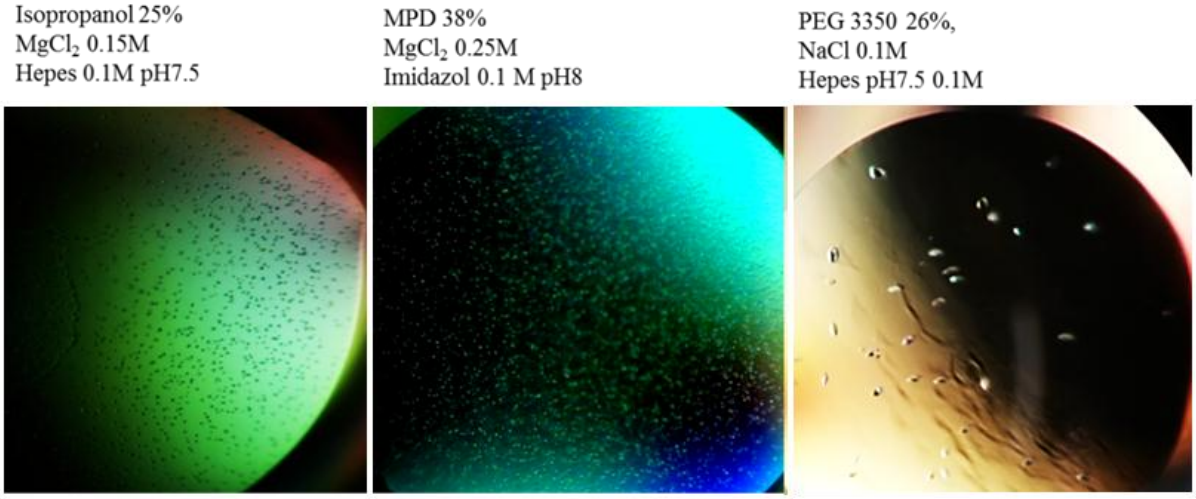


Figure_25. Drops containing micro or little crystals are shown.
The first two drops on the right were set up using "SiteY-X65bp" sequence (see table_1) in complex with TFAM at 10 mg/ml. The last drop was set up using "SiteY-X50bp" sequence (see table_1) in complex with TFAM at 15 mg/ml. On the top of each picture the crystallization conditions are indicated.

5.1.2.2 Crystallization of TFAM in complex with sequences Y and X.

TFAM in complex with site Y and site X yielded much successful results. The sequences tested in complex with TFAM are shown in red in Table_2. Also in this case the complex formation was firstly monitored by gel shift analysis. TFAM and the DNA were mixed in a buffer at high salt concentration and dialyzed against buffers containing lower salt concentration by using, in this case, dialysis buttons as described in Materials and Methods section 4.5. A constant concentration (5 uM) of DNA titrated with increasing TFAM concentration (from 2.5 uM to 30 uM) was used (Figure_26). The ratio at which the DNA was first fully shifted by TFAM was the one employed for crystallization. In both cases (site Y and site X) this ratio corresponded to 2 proteins per 1 DNA molecule, as indicated with a red rectangle in Figure_26 (Figure_26).

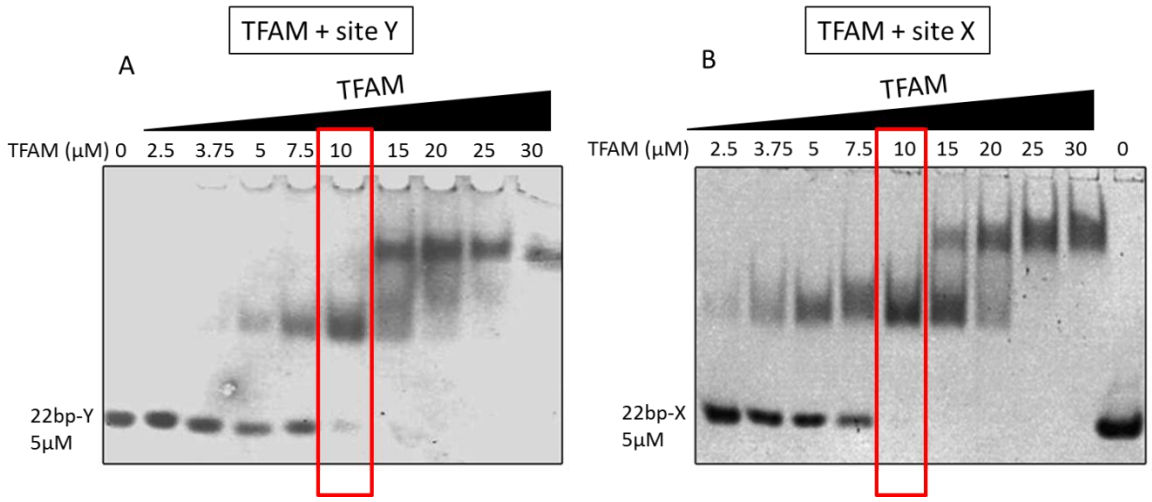


Figure_26. Gel shift assay of TFAM in complex with Site Y and Site X.
A, an EMSA of TFAM in complex with site Y is shown. **B**, EMSA of TFAM in complex with site X. In both cases the free DNA is indicated with "0". In both gels a serial increasing amount of TFAM concentration (indicated above the gel) was added to a constant concentration of 5uM 22 bp DNA. In both gels the red rectangle indicate the ratio 2 protein:1 DNA.

As for the Y-X long sequence, crystallization trials of TFAM in complex with site Y or site X started with extensive screenings employing the automated systems implemented in the PAC that dispense nanodrops into 96 well plates (ratio 1:1 sample and crystallization solution in the crystallizing drop). With this method nuclei or small crystals were observed, thus the most promising crystallization trials were scaled up in 24 well plates employing micro-scale volumes by using either a hanging or sitting drop set up. Crystals were obtained for both TFAM/site X and TFAM/site Y complexes, by employing 22 bp sequences with cohesive ends in the case of site X (X22c) or blunt ends in the case of site Y (Y22) (both duplexes are indicated in red in Table_2). Considering the crystallization condition, for the TFAM/X22c complex the best X-ray diffraction with highest resolution were obtained from crystals grown in 18-24% PEG 3350, 0.1-0.25M of sodium potassium tartrate and 0.1 M of Hepes, pH 7.5. For TFAM/Y22 complex the best diffraction pattern was obtained from crystals grown in 23-28% PEG 3350, 0.08-0.2M of ammonium acetate, and 0.1 M of bis-Tris pH 6.5 or Hepes pH 7.5 (Figure_27).

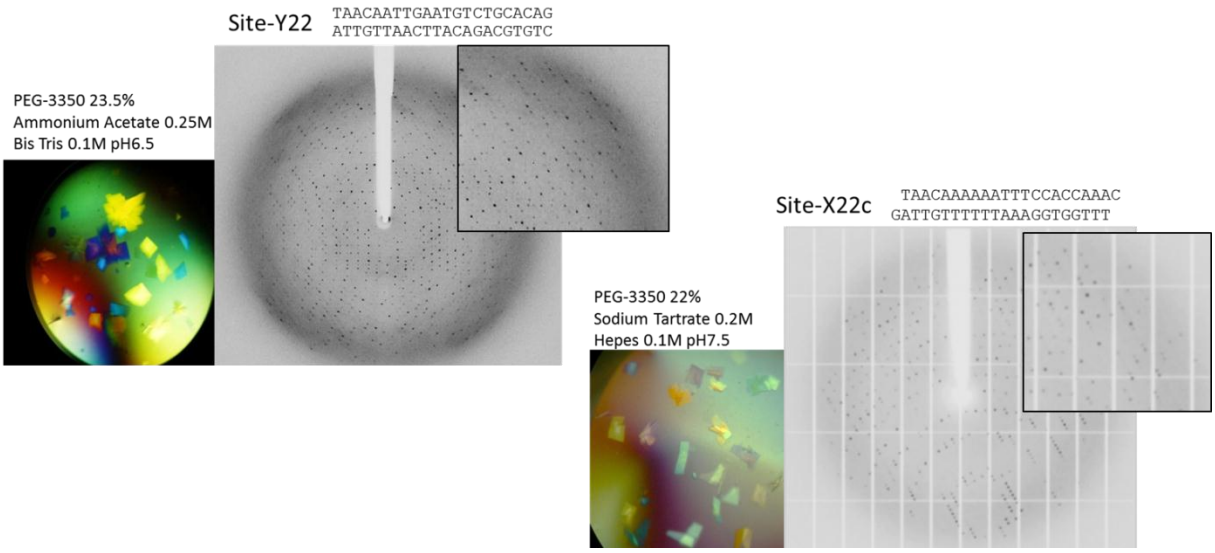


Figure_27. Crystals of TFAM/X22$_c$ and TFAM/Y22 and X-ray diffraction patterns.
Images of the crystallization drops and the diffraction pattern of the best crystals of the two complexes TFAM/site Y22 and /site X22c are shown. The two dsDNA sequences and the crystallization conditions are also reported.

TFAM/X22c and TFAM/Y22 complexes diffracted to a nominal resolution of 3Å. Diffraction data were collected at microfocus beamline ID23-2 of the European Synchrotron Radiation Facility (ESRF), Grenoble. The structure of TFAM in complex with site X22c was solved and described by Pablo Fernández Millán in his PhD thesis: "Estudio estructural de proteínas implicadas en el metabolismo del genoma mitocondrial: Helicasa y Factor de Transcripcíon A mitocondrial, TFAM", Universitat de Barcelona). Regarding TFAM in complex with site Y22, data processing and three-dimensional structure are described below.

5.1.3 TFAM/site Y22 data processing, crystal structure determination and refinement.

Diffraction data obtained from TFAM/Y22 complex was indexed, integrated, scaled and reduced to unique reflections with XDS (Kabsch 2010). The crystal belongs to C2 space group with unit cell dimensions a=157.67(Å), b=140.62(Å), c=108.92 (Å) and α=90.0º, β=130.73 and γ= 90.0º, at a resolution of 3.05 Å. The calculated Matthew's coefficient was of 2.59 Å^3/Da, which is consistent with four complexes in the asymmetric unit (a.u.). The phase was solved by molecular replacement (MR) method

using Phaser from the CCP4 suite (McCoy, Grosse-Kunstleve et al. 2007) and employing TFAM-LSP structure (PDB: 3TQ6) (Rubio-Cosials, Sidow et al. 2011) as a searching model. All values of the Phaser "Z-score" associated with rotation (RFZ) and translation peaks (TFZ) relative to each molecule in the asymmetric unit are reported in Table_3. The solution obtained presented a "TF Z-score" of 11.6 that, according to the criteria of this program, indicate that it is the correct ones. The number of clashes (PAK) between molecules of the a.u. and the ones related by symmetry is reported. In addition, a visual inspection of the packaging confirmed the correct distribution of the molecules within the crystal.

Table_3. "z-score" values associated to Phaser solutions

	RFZ	TFZ	PAK
Solution 1	5.7	4.5	4
Solution 2	4.5	9.8	4
Solution 3	4.8	12.3	1
Solution 4	4.3	9.2	17

The a.u. contains four TFAM molecules A, D, G and J, each in complex with one dsDNA molecule, strands B-C, E-F, H-I, K-L respectively. Automatic refinement with Phenix.Refine (Adams, Afonine et al. 2011) and Buster (Branden and Jones 1990) followed, which was alternated with visual inspection of the refined structure and manual model building with COOT (Emsley 2004), until the model was completed (Figure_28). At the beginning of the refinement non-crystallographic restrictions were applied, together with TLS and Watson-and-Crick restraints. In all four molecules, TFAM structure was traced from residue 44 to residue 236, since beyond residue 236 to 246 the density was too poor to trace further.

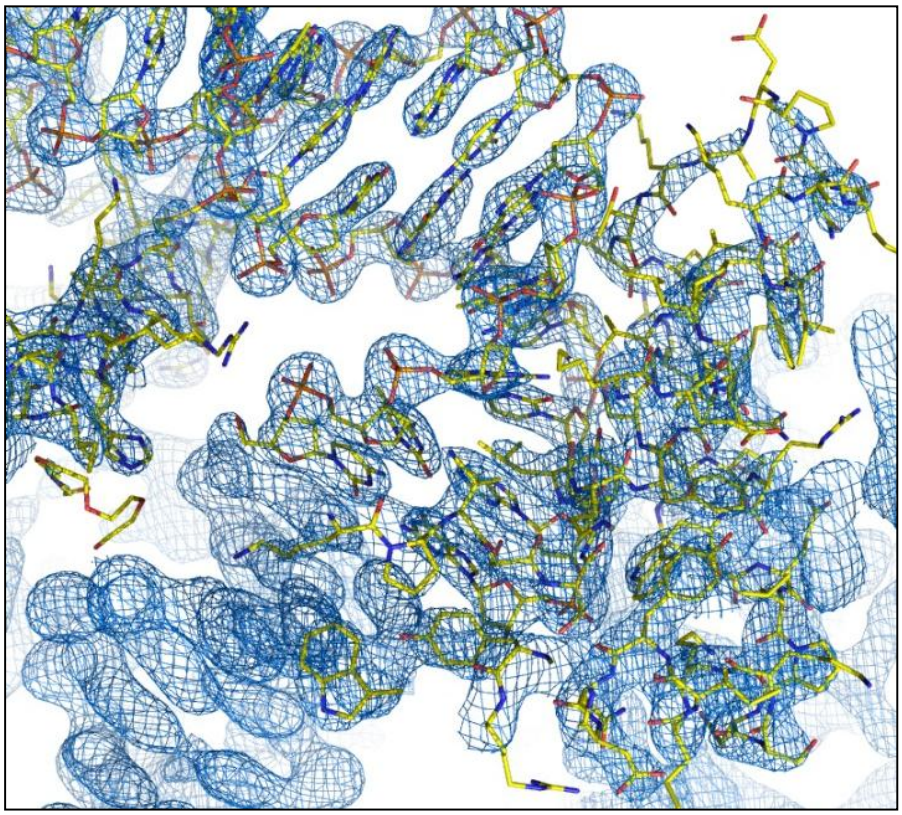


Figure_28. TFAM/Y22 crystal structure and electron density map.
The protein and DNA bonds are shown as sticks. The electron density shows a clear DNA stacking pattern. At the top of helices 1 and 2 of HMG-box1 the map is less well defined, suggesting disorder at this region.

Regarding the tracing of the DNA, the original LSP sequence was substituted by the Y22 sequence according to the alignment provided by Fisher and collaborators (Fisher, Topper et al. 1987). This orientation implied similar positioning of the DNA ends in both TFAM/LSP and TFAM/Y22 structures. However, after refinement the 2Fo-Fc and Fo-Fc electron density maps around the DNA bases were very poor or not consistent with the base type (purine, pyrimidine). This suggested an incorrect assignment of the sequence. In addition the oligos were purchased without 5' phosphate. However positive density map (Fo-Fc map), suggesting the presence of a phosphate, appeared at the 5' end of either strands (Figure_29).

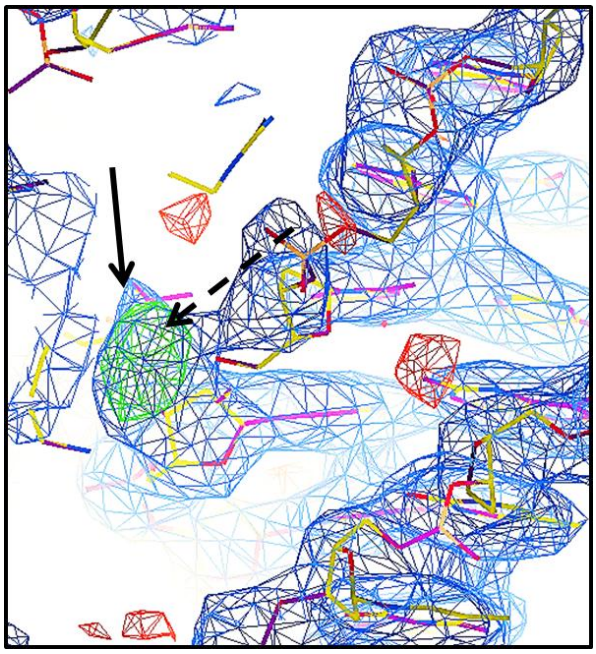


Figure_29. Electron density maps of the Y22 DNA model at 5' end of molecule B.
Fo-Fc map clearly indicates that the initial model was not correct in describing the experimental data. Both maps suggested a shift of the DNA molecule (black arrow) to fill the positive density with a phosphate. The discontinuous arrow indicates the shift.

As a consequence the DNA was shifted. In addition, to minimize model bias with the original DNA sequence, all the bases of the four a.u. structures, were mutated to cytosine and the resulting model refined. The 2Fo-Fc and Fo-Fc maps corresponding to this structure suggested a purine (R) and pyrimidine (Y) sequence that was compared to the sequences of the purchased oligonucleotides. Figure_30 A shows the Y-R 5'-3'predicted sequence of one DNA strand, obtained after map interpretation (Figure_30A). By the alignment of the predicted Y-R sequence and the 5'-3' Y22 strands (Figure_30B and C) one duplex orientation showed significant more mismatches (9 errors) than the other one (2 errors) (Figure_30Bb and Cb).

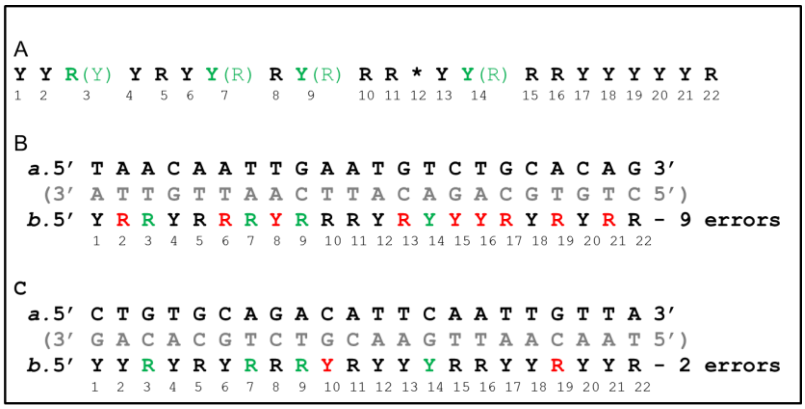


Figure_30. Sequence assignment of the bases based on the electron density map.
A) The purine (R) and pyrimidine (Y) sequence suggested by the electron density map is shown. Black characters correspond to bases assigned with more confidence, green characters are preferred guesses (could also be the alternative in brackets). The asterisk * refers to a non-assignable position. **B)** In ***a.*** the crystallized 5' to 3' direct sequence (below in grey the complementary in the 3' to 5' sense). ***b.*** Purine (R)-pyrimidine (Y) version of the sequence in *a*. From comparison with sequence in A, bases depicted in green are at 'guess' position (see panel A), while in red are the non-coincidences (9 errors in total). **C)** Same as in B but for the 5' to 3' complementary sequence (in grey in B).

Therefore site Y22 was traced in the opposite orientation with respect to LSP and one step shifted. However, still after subsequent steps of refinement and model building an extra density at the ends of all the four Y22 DNAs is showed, as if a continuous 44bp ring shape DNA molecule was present in the a.u. of the crystal (Figure_31).

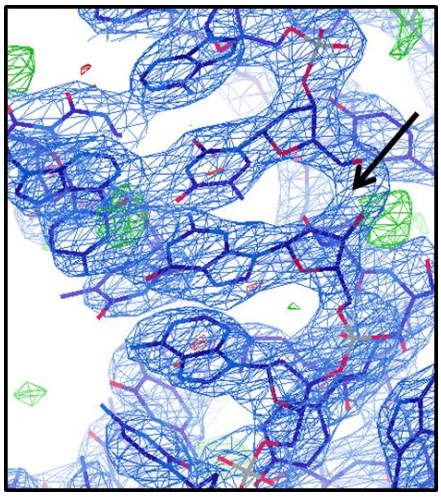


Figure_31. **Final DNA electron density maps calculated using Y22 reoriented model.**
Continues electron density map is indicated by a black arrow.

All this suggests that the DNA probably has a preferred orientation but also a second, alternative orientation might be possible. Nevertheless DNA stacking was clear from the maps and the sequence was consistent with the 2Fo-Fc and Fo-Fc maps at the purine and pyrimidine level.

The X-ray diffraction data and the final structure validation parameters of the TFAM/Y22 complex structure are summarized in Table_4. The values of both R-free (0.2381) and R-work (0.1964) are reasonable for a 3.05 Å resolution data set (Kleywegt and Brunger 1996). All the protein residues are within allowed Ramachandran regions (97.8% of the residues are in favourable regions, 2.2% in allowed regions and "0" outliers) (Figure_32). The geometry values are appropriate (bond lengths=0.012Å; bond angles=1.47°) and the not common rotamers are consistent with the electron density.

Table_4. crystallographic parameters of the TFAM/Site Y complex crystal.
Values within parentheses correspond to the last resolution shell.

	TFAM-site Y22 complex
Data collection	
Space group	C2
Cell dimensions	
A, b, c (Å)	157.67-140.62-108.92
a, β, γ (°)	90.0 - 130.73 - 90.0
Resolution (Å)	43.70 - 3.05 (3.21-3.05)
R_{sym} (%)	10.3 (53.6)
I/σI	19.4 (2.8)
Completeness (%)	99.96 (100)
Redundancy	13.6 (5.1)
Refinement	
Resolution (Å)	43.05 - 3.05
No. reflections	34312
R_{work} / R_{free}	0.1964 / 0.2381
No. atoms	
Protein	6522
DNA	3584
Water	0
B-factors	
Protein	87.5
Ligand/ion	85.5
Ramachandran outlier	0
R.m.s. deviations	
Bond lengths (Å)	0.012
Bond angles (°)	1.47

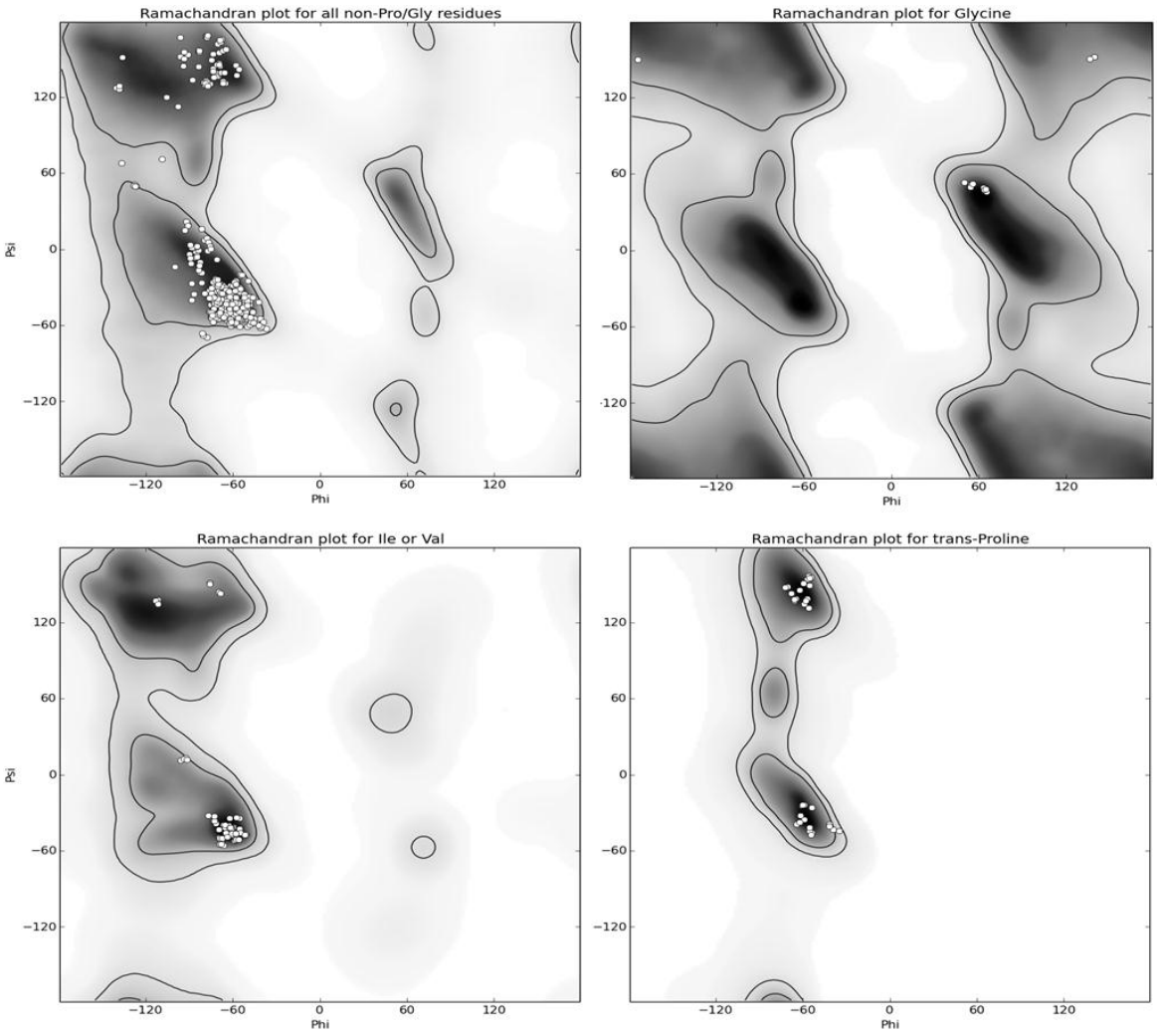


Figure_32. Ramachandran plots.
Four Ramachandran plots are showed. On the top left a general Ramachandran plot for all TFAM residues is presented. On the top right the Gly in the bottom the Ile orVal and Pro Ramachandran plots are shown. These plots were calculated by Phenix suite (Adams, Afonine et al. 2011).

5.1.4 TFAM/site Y22Br crystallization, X-ray diffraction, data processing and anomalous signal analysis.

In order to assess the orientation of the DNA sequence in the TFAM/Y22 complex, crystals with Bromo-Uracyl (BrU) substituting one thymine within the site Y sequence was generated. In total, four DNAs with a thymine substitution at a different position (Y22Br1 to 4) were tested (Table_5).

Table_5. Y22 brominated DNAs

Y22Br1	5′**BrU**AACAATTGAATGTCTGCACAG 3′ 3′A TTGTTAACTTACAGACGTGTC 5′
Y22Br2	5′TAACAATTGAATGTCTGCACA G 3′ 3′ATTGTTAACTTACAGACGTG**BrU**C5′
Y22Br3	5′TAACA ATTGAATGTCTGCACAG 3′ 3′ATTG**BrU**TAACTTACAGACGTGTC 5′
Y22Br4	5′TA ACAATTGAATGTCTGCACAG 3′ 3′A**BrU**TGTTAACTTACAGACGTGTC 5′

TFAM in complex with brominated Y22 crystallized in the same condition as TFAM/Y22 native complex. The crystals presented the same shape as the native ones. Best diffraction patterns were obtained with Y22Br1 and Y22Br2. Y22Br3 did not crystallize, while Y22Br4 crystals diffracted at lower resolution. Data of derivative TFAM/Y22Br2 with maximal nominal resolution of 3 Å were collected at ALBA synchrotron, (Cerdanyola del Vallès, Spain). Prior to data collection a X-ray absorption near-edge scan (XANES, commonly called energy scan) was performed that showed a peak of fluorescence close to the Br edge and thus confirmed the presence of BrU inside the crystals. Accordingly, data were collected at the peak (λ=0.919183).

Diffraction data was processed with XDS (Kabsch 2010), with no merging of Friedel's pairs. The crystal belongs to the orthorhombic primitive space group. The analysis of systematic absences suggested two helical axes, thus a $P2_12_12$ space group with unit cell dimensions of: a=112.3 Å, b=121.6 Å, c= 55.4 Å and α=90.2º, β=90.0°, γ= 90.0º. The analysis of anomalous differences between Bijvoet pairs suggested a weak anomalous signal (which in XPREP was less than 1 or equal to 1 in all datasets collected). The structure determination and refinement of the TFAM/Y22Br2 data set followed. The Matthew's coefficient was of 3.21 $Å^3$/Da, indicating two complexes per a.u. The phase was solved by the molecular replacement (MR) method using Phaser from the CCP4 suite (McCoy, Grosse-Kunstleve et al. 2007) and employing TFAM/Y22 structure as a searching model. The solutions obtained from Phaser presented a "TFZ" value of 13.4 indicating the solutions were correct. All the values of the "z-score" associated with rotation and translation peaks relative to each molecule

of the asymmetric unit are reported in table_6 together with the number of clashes between the molecules of the a.u. and the ones related by symmetry. As for the previous crystal, visual inspection of the MR solution indicated a minimal number of clashes between complexes within the crystal.

Table_6. "z-score" associated to Phaser MR solutions

	RFZ	TFZ	PAK
Solution 1	6.4	10.2	0
Solution 2	3.2	9.5	4

The a.u. contains two TFAM molecules labelled A and D, each in complex with one dsDNA, B-C and E-F strands, respectively. Automatic refinement with Phenix.Refine (Adams, Afonine et al. 2011) and Buster (Branden and Jones 1990) followed, alternated with manual model building with COOT (Emsley 2004) until the model was completed. During refinement it was apparent an important disorder at helices 1 and 2 of HMG-box2 of molecule B, which show a slight reorientation. Upon model building and refinement a Fourier synthesis using the anomalous differences as coefficients did not show any significant peak, indicating absence of any anomalous signal. Consequently, the DNA sequence could not be verified and the analysis of the DNA arrangement was again inconclusive, since the position of the 5' and 3' ends showed again a continuous density. Therefore, the disorder of the DNA most probably precluded the anomalous signal. The results from processed data and the Rwork/Rfree validation parameters of the TFAM/Y22Br2 complex are summarized in Table_7.

Table_7. TFAM/Y22Br2 complex crystallographic table.

Data collection	
Space group	$P2_12_12$
Cell dimensions	
A, b, c (Å)	a=112.3 Å, b=121.6 Å, c= 55.4 Å
a, β, γ (°)	α=90.2º, β=90.0, γ= 90.0º
Resolution (Å)	43.70 – 2.95
R_{sym} (%)	10.3 (53.6)
I/σI	10.6 (1.2)
Completeness (%)	99.3%
Redundancy	3.3
Refinement	
Resolution (Å)	43.70 – 2.95
No. reflections	
R_{work} / R_{free}	0.3001/0.3407

5.1.5 TFAM/Y22 crystal structure.

As mentioned above (section 5.1.3) TFAM/Y22 a.u. is constituted by four TFAM molecules A, D, G and J each one in complex with a dsDNA: B-C, E-F, H-I and K-L strands respectively. Since the four complexes show the same TFAM folding and DNA shape, the description will refer to molecule A represented in Figure_33 in a ribbon plot (Figure_33). The three-dimensional structure of TFAM in complex with Y22 shows two HMG-box domains, HMG-box1 and HMG-box2, contacting the narrow groove at one side of the DNA and inducing a strong 180º bend to the double chain (Figure_33). Between the HMG-boxes, the DNA is straight, in a B-DNA conformation. A linker of 30 aa joints the two HMG-boxes. This linker connects the C-terminal end of HMG-box1, located at one side of the DNA, with the N-terminal end of HMG-box2, at the other side, by passing close to the narrow groove at the DNA face opposite to the HMG-boxes (Figure_33). In Figure_33 the DNA orientation is the one assigned considering the best fitting of purine and pyrimidine of Y22 to 2Fo-Fc and Fo-Fc maps.

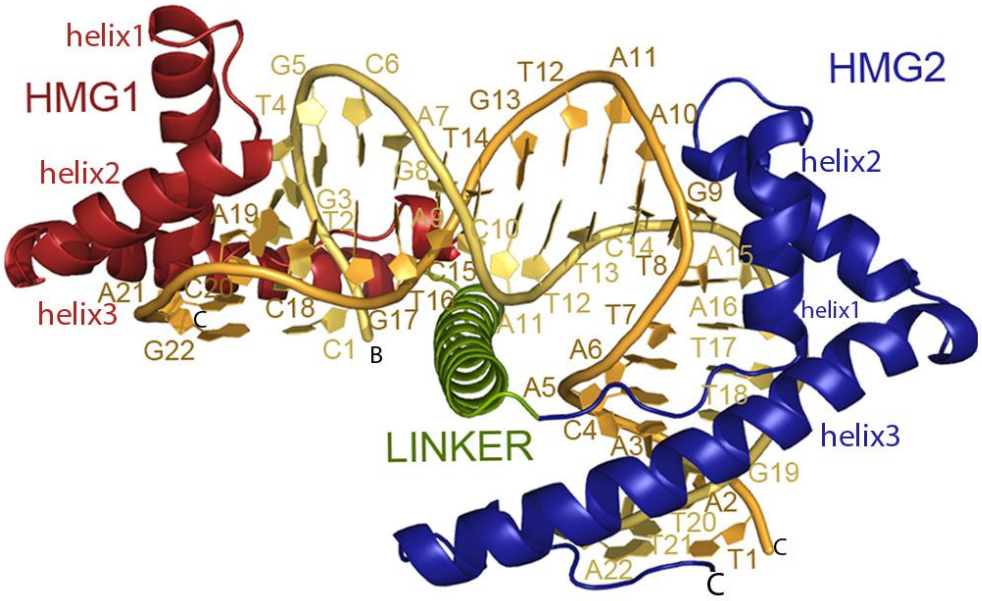


Figure_33. Three-dimensional structure of TFAM/Y22.
TFAM/Y22 complex crystal structure is presented. In blue and in red are represented the two HMG-box domains 1 and 2, respectively. The DNA is represented in orange, (strand C) and yellow (stand B). The linker is in green. The three helices of the two HMG-box domains are indicated (helix1, 2, 3). The DNA orientation is the one assigned considering the highest fitting of purine and pyrimidine to 2Fo-Fc and Fo-Fc maps.

Both HMG-boxes have the same fold, which consists of three α-helices organized in the archetypical "L" shape found in HMG domains. HMG-box1 starts with an N-terminal elongated segment (aminoacids, aa, 44 to 56) followed by three helices: helix1 (aa 56-71) and helix2 (aa 77-90) that are antiparallel to each other (forming the short L-arm), and the longest helix 3 (aa 93-120) that is oriented antiparallel to the extended N-terminal region (forming the long L-arm). A linker (aa 123 to 152) with α-helix conformation connects HMG-box1 with HMG-box2. HMG-box2 is also constituted by the three α-helices, helix1 (aa 161-172), helix2 (aa 178-191) and helix3 (aa 194-225) that are in an "L" shape conformation too. The C-terminal region of TFAM is constituted by a positively charged tail, traced from residue 227 to residue 236 (residues from 237 to 246 were not traced due to very weak electron density maps). The three helices of both HMGB domains present the same length except for helix1, which in HMG-box2 is one turn shorter, a feature unique in TFAM. Furthermore, in both domains the loops connecting helix1 to helix2 (loop 1), and helix 2 to helix 3 (loop 2) are same times tentatively traced in weak density, which suggests

flexibility. Within the crystal, the four HMG-box1 (or HMG-box2) from A, D, G and J protein structures are highly similar, with an overall root mean square deviation (r.m.s.d.) of 0.33Å (the same between the four HMG-box2 domains).

In all the previous crystallographic structures TFAM impose a strong bend to the DNA (Ngo, Kaiser et al. 2011; Rubio-Cosials, Sidow et al. 2011; Ngo, Lovely et al. 2014). This also occurs in TFAM/Y22 complex structure. Each HMG-box domain induce a 90º kink on Y22, resulting in a total DNA bending of 180º or U-turn conformation. The analysis of the structure reveals that the majority of interactions between TFAM and site Y22 are not specific since it involves aa side chains that contact sugars and phosphates of the DNA backbone. Only a small number of residues contact the DNA bases, whose sequence assignment is based on the orientation suggested by the electron density maps (in brackets in Figure_34).

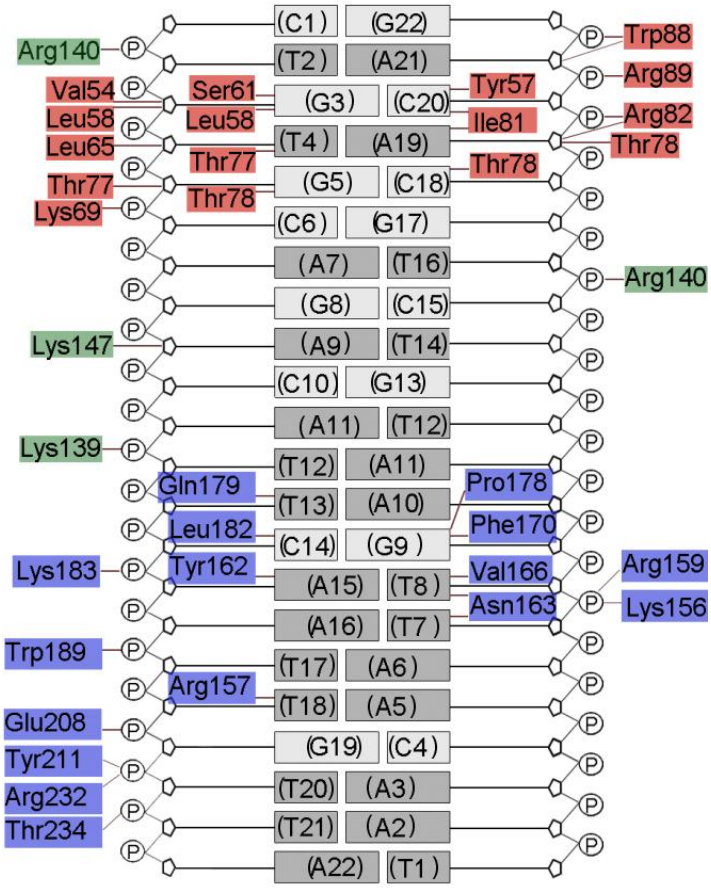


Figure_34. Scheme of TFAM/Y22 contacts considering the putative site Y22 orientation.
In this scheme the nature of the bases is tentative. In red are indicated the residues from HMG-box1, in green the residues from the linker and in blue the residues from the HMG-box2 and C-terminal tail that contact the site Y22 phosphate backbone or bases (in brackets due to its ambiguous orientation).

The Y22 sequence is contacted by all protein domains: HMG-box1 and 2, the linker and the C-terminal tail. HMG-box1 most important interactions are mediated by helix1 and helix2, which together with helix3 create a concave surface that, by hydrophobic interactions with the bases and polar contacts with the phosphate backbone separate the two DNA strands at the narrow groove (Figure_35). Through these interactions HMG-box1 induces a 90º bend to the DNA double helix towards the major groove. This bending involves high roll values (see section 5.2.1) between base pairs together with a decrease of their twist value. The residues involved in interactions include: Tyr57, Leu58, Ser61, and Leu65 from helix1; and Thr77 and 78, Ile81, Trp88 and Arg89 from helix2. Importantly, Leu58 side chain inserts between two consecutive DNA base pairs, at step $T_2G_3/C_{20}A_{21}$ (chain B and C respectively) of the assigned sequence, disrupting the stacking interactions between bases and causing an, important increase of the roll value (Figure_43, section 5.2.1). This insertion is stabilized by hydrophilic contacts of the G_3/C_{20} pair from this step with Tyr57 and Ser61 (C_{20} O2···OH Tyr57; G_3 N2···OH Ser61, chains B and C respectively) (Figure_35). At the neighbouring $G_3T_4/A_{19}C_{20}$ step, the Ile81 side chain (from helix2) causes a 'half insertion' that further contributes to the roll increase and global bend (*half insertion* is a term widely used in the HMG-box structural field and indicates that the DNA stacking is disturbed by a side chain that does not fully penetrate between the two base pairs). At the N-terminus of helix 2 (at the tip of the helix1/helix2 triangle), the Thr78 OH contacts the O2 of assigned C_{18} (chain C). Finally, helix1 Leu65, and Lys69 and helix 2 Thr77, Arg82, Trp88 and Arg89 are close or directly involved in sugar and phosphate backbone contacts, further stabilizing the kink.

Between HMG-boxes, the linker contacts the minor groove at the side of the DNA opposite to the HMG-boxes (Figure_33). In particular, the residues from the C-terminal region of the linker Lys139, Arg140 and Lys147, as well as Met143, point their side chains towards the minor groove contacting the phosphate backbone but not the base atoms (Figure_35). After the linker, the protein continues intertwining the DNA eventually positioning HMG-box2 on it. HMG-box2 imposes the second 90º kink (Figure_35). In there, Lys156 and Arg157 from the HMG-box2 N-terminal elongated fragment stabilize DNA strands separation. It is noteworthy that Arg157

gets deep into the narrow groove, where it contacts O2 from T_{18} (chain B). This contributes to the opening of the two strands and stabilizes the insertion of Leu182 (at helix3) into step $C_{14}A_{15}/T_8G_9$ (chains B/C) of the assigned sequence. In addition, Tyr162 and Asn163 from helix1 stabilize the orientation of the T_8/A_{15} pair (Tyr162 OH···N3 A_{15}; Asn163 ND2···O2 T_8, chains B and C, respectively) (Figure_35). Another polar contact involves the N-terminus of helix2 (at the tip of the short L arm), where Gln179 NE2 contacts O2 from T_{13} (chain B). Arg159, Val166, Ala167 and Phe170 from helix1 together with Lys183, Lys186 and Trp189 from helix2 stabilize strand separation by contacting the phosphate backbone. Finally, residues from helix3 and the C-terminal tail (Glu208, Try211, Arg232, Arg233, and Thr234) contact the downstream DNA, from nucleotide 19 to 22.

Notably, the insertions of Leu58 and Leu182 at respective steps $T_2G_3/C_{20}A_{21}$ $C_{14}A_{15}/T_8G_9$ (chains B/C) are separated by 10 bp, thus one DNA helical turn.

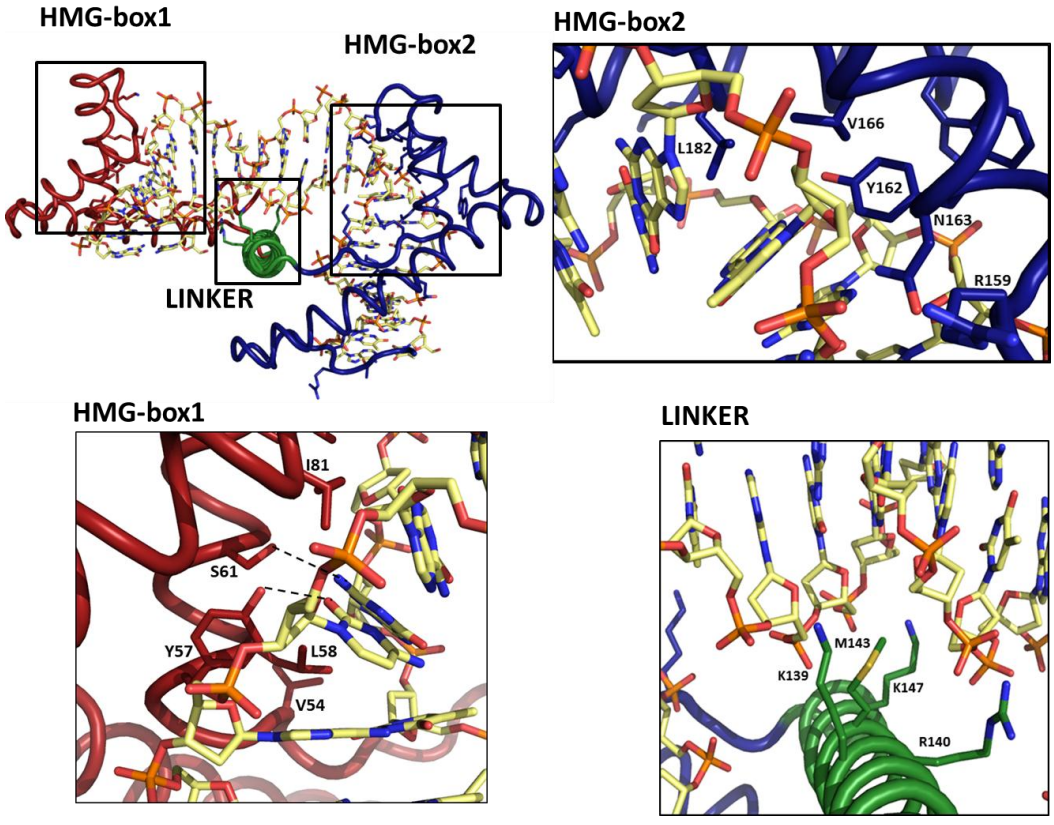


Figure_35. Iteractions between TFAM and the Y22 assigned sequence.
The three principal TFAM domains involved in DNA contacts are indicated with black frames. In blue is depicted HMG-box2, in red HMG-box1 and in green the linker. The residues that interact with the DNA are presented using one letter code.

5.1.6 Analysis of the TFAM/ Y22 asymmetric unit.

The four complexes (protein/DNA strands A/BC, D/EF, G/HI, and J/KL) in the a.u. of the TFAM/Y22 crystal interact with each other in two couples: A/BC+D/EF and G/HI+J/KL (Figure_36A and B). The DNA molecules from a couple constitute a pseudo-continuous DNA ring with the DNA ends performing a perfect stacking. Within a ring the two proteins interact with each other through the N-terminal part of the linker and the loop between HMG-box2 helix3 and the C-terminal tail. The α-helices of the two linkers in one ring cross over each other and are much closer at one side of the ring and much distant at the other side (Figure_36). Furthermore the two couples are oriented with respect to each other head to back, with the N-terminal part of the linker of molecule G in close proximity to the C-terminal tail of molecule D (distance between C^{α} Pro123/G and Glu224/D is 3.31Å). Importantly, all HMG-box1 domains from both rings perform antiparallel contacts with HMG-box1 domains from symmetry-related partner. This interaction involves residues: Lys95, Tyr99, Glu106, Glu112, Arg116 (Figure_36B and C), it has been found in all crystals of TFAM bound to DNA and it was described as essential for TFAM dimer formation in solution (Ngo, Lovely et al. 2014).

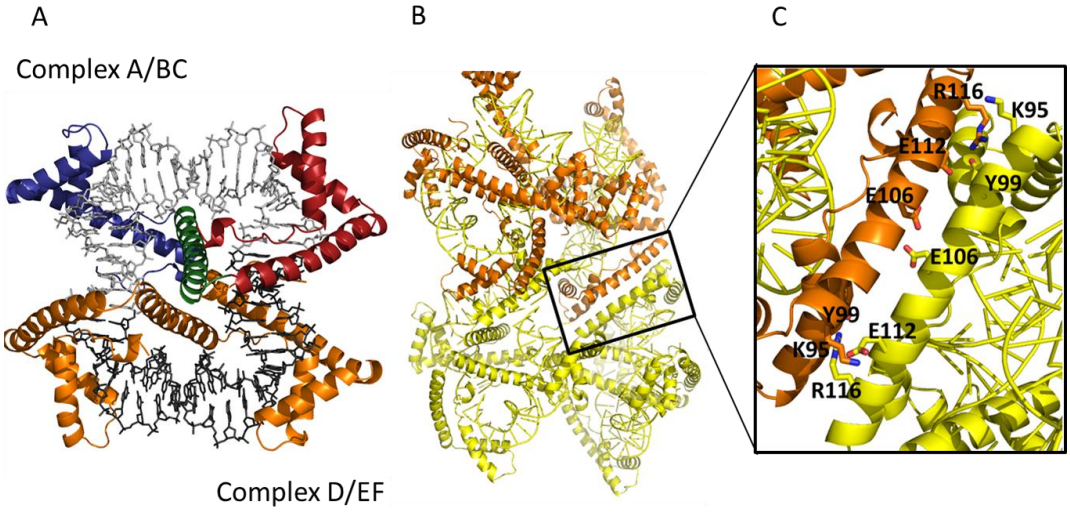


Figure_36. Components of the asymmetric unit and their contacts with symmetry-related molecules. **A**, Complexes A/BC and D/EF are presented. The domain colors of molecule A are the same as in previous figures. In **B**, the symmetry-related antiparallel interaction between HMG-box1 domains is framed. C) Zoom in of the antiparallel HMG-box1 interaction with involved residues highlighted.

The structural comparison between complexes in the a.u. was performed using complex A/BC as a reference. The root mean square deviation (r.m.s.d.) values of A/BC with the rest of complexes superposed by C^{α} atoms from residues 44 to 236 are reported in Table_8. The r.m.s.d. values indicate that complexes A/BC and D/EF, from the same ring, are the most similar ones (r.m.s.d.= 0.5Å), while the most different one is J/KL (r.m.s.d.= 2.36Å). However, such an overall superimposition did not show differences at a single site (and thus easily interpretable) but differences all along the structure. Therefore, a superposition by each different TFAM domain was performed.

Table_8. R.m.s.d. values calculated from superposition of aminoacids 44 to 236 C^{α} complexes D/EF, G/HI and J/KL onto A/BC, from.

	Overall r.m.s.d. against complex A/BC (Å)
Complex D/EF	0.48
Complex G/HI	1.42
Complex J/KL	2.36

In Figure_37, complexes D/EF, G/HI and J/KL are superposed onto A/BC by HMG-box1 (residues 44 to 120, Figure_37 panels A and B), or by HMG-box2, (residues 161 to 236, Figure_37 panels C and D). Superposition of the complexes by HMG-box1 results in different orientations of HMG-box2 (the highest distance of 12.2Å between HMG-box2 was found between A/BC and J/KL complexes, at Asp194) (Figure_37A). If the superposition is done by HMG-box2, the largest distance (10.52 Å) is found at Leu91 between HMG-box1 domains of A/BC and J/KL complexes (Figure_37C). This variability between HMG-box orientations among the four complexes in the asymmetric unit originates midway in the linker, at His137 (Figure_37).

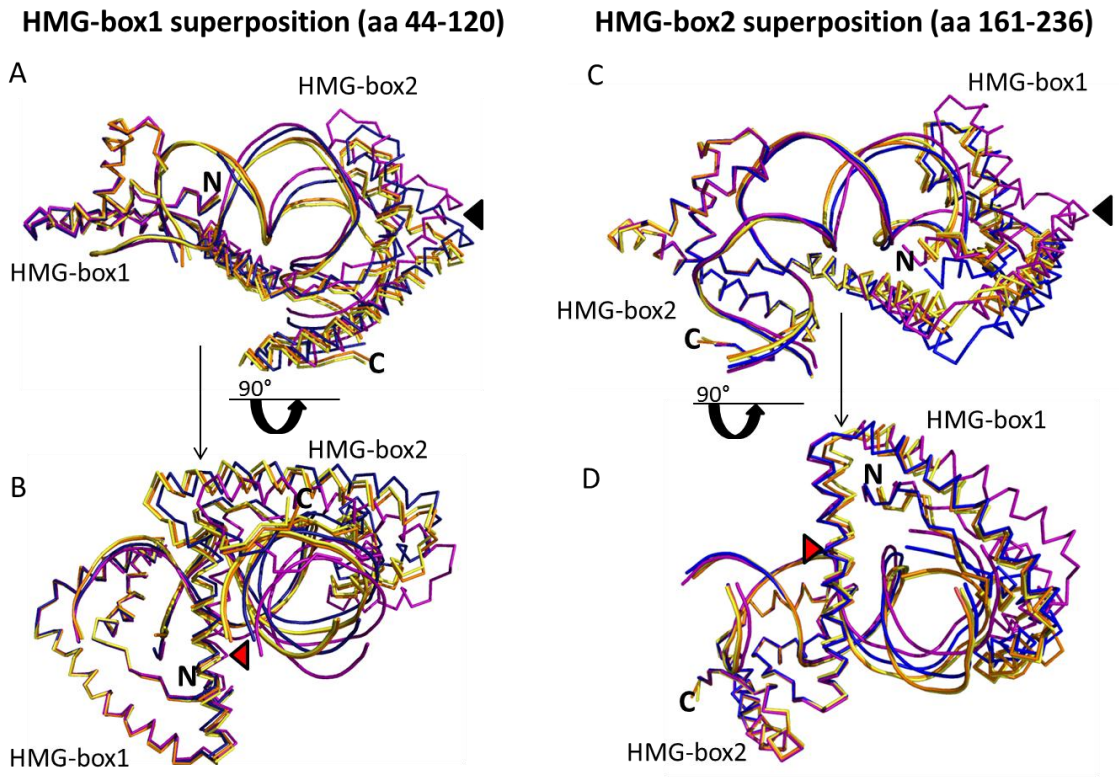


Figure_37. Superposition by the HMG-boxes of the four molecules of the a.u.
In **A** and **B** the same superposition of all a. u. structures by HMG-box1 is shown in two, 90º related orientations. In **C** and **D** superposition of the same molecules by HMGbox2 is shown, also in two 90º related orientations. The axis of rotation is indicated aside the vertical arrow. With a red arrowhead is indicated the position of His 137. With a black arrowhead is indicated the point of maximal distance between the complexes, at Asp194 and Leu91. In both superpositions the same color code was used: complex A/BC is represented in yellow, complex D/EF in orange, complex G/HI in blue and complex J/KL in purple.

From the N-terminus of the linker, His137 is the first residues closest to the DNA and, from there on, contacts occur with the minor groove backbone, mediated by Lys139, Arg140, Met 143 and Lys147 (see 5.1.5 section Figure_35). Analysis of the crystal packing suggests that different symmetry environments cause the variation in the position of HMG-box2, which involves the DNA too. In accordance to this, the linker would accommodate the DNA movements by means of a flexible hinge at position 137. Despite the torsion of the linker structure, the relative position between the linker N-terminus and HMG-box1, or between the linker C-terminus and HMG-box2 is highly similar in the four complexes (Figure_38). By linker N-terminus (aa 123-136) and C-terminus (aa 146-153) superposition it was seen that HMG-box1 and HMG-box2 are *glued* to the linker by respective hydrophobic cores with almost identical side

chain orientations in all complexes. These residues include Val45, Leu46 and Leu121 from HMG-box1 that interact with Ile126, leu129 and Ile133 of the linker. On the other hand HMG-box2/linker contacts involve: Leu149 and Leu152 from the linker and Met222 and Try218 from helix3 (Figure_38). Apart from these, additional polar interactions contribute to stabilize the interactions between the linker and the HMG boxes (Figure_38).

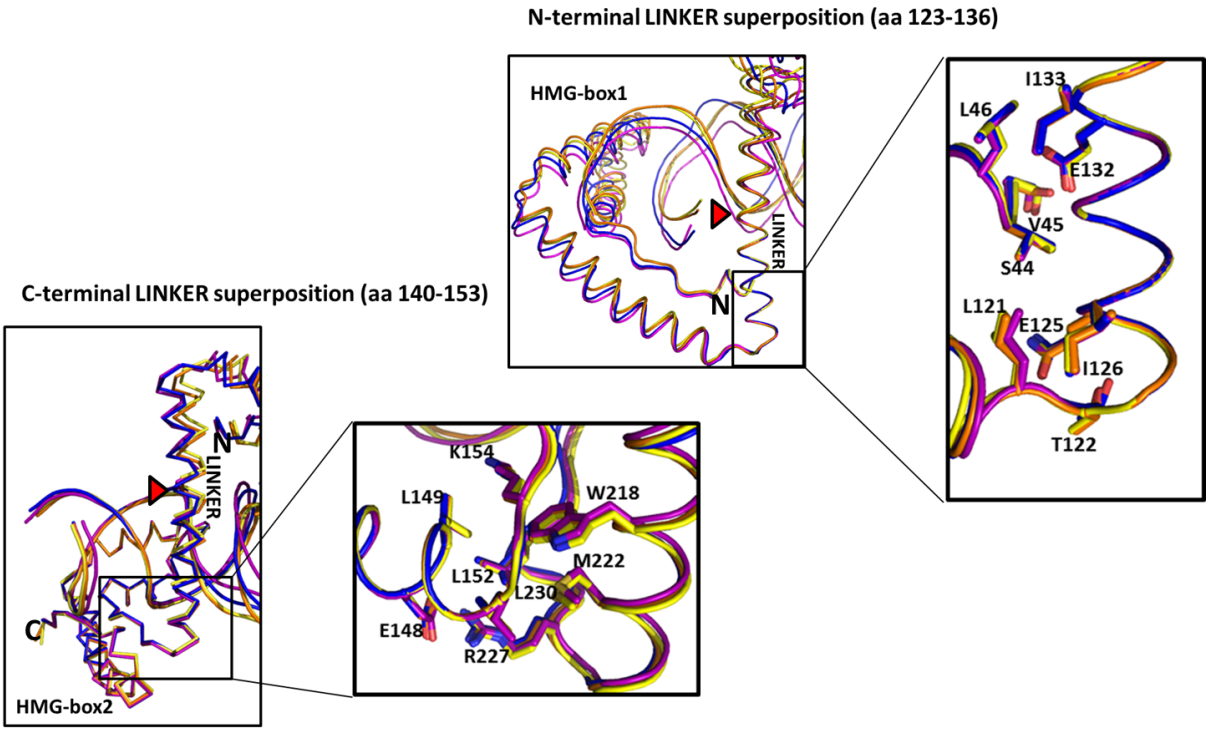


Figure_38. N-terminal and C-terminal linker/HMG-box 1 and 2 interactions.
N-terminal (aa123-136) and C-terminal (aa140-153) linker superposition of the four molecules in the a.u. is shown. A red arrowhead indicates the His137 position. The residues implicated in interactions are indicated using one letter code at the magnification panels. The color code of the four complexes is the same as for the previous pictures: complex A/BC is represented in yellow, complex D/EF in orange, complex G/HI in blue and complex J/KL in purple.

5.1.7 Comparison between the TFAM/Y22 structure and previous crystallized TFAM-DNA complexes.

The first crystal structure of a TFAM/DNA complex was published in 2011 by Ngo and collaborators, and Rubio and collaborators (Ngo, Kaiser et al. 2011; Rubio-Cosials, Sidow et al. 2011). In these works TFAM crystallized in complex with DNA

fragments harboring the cognate binding sequence, of 22 bp (LSP22; PDB code 3TQ6) (Rubio-Cosials, Sidow et al. 2011) or 28 bp (LSP28; PDB code 3TMM) (Ngo, Kaiser et al. 2011)). In 2014 Ngo et al., published TFAM in complex with a fragment of the HSP transcription regulatory region (TFAM/HSP, PDB code 4NOD), and in complex with a non-specific random DNA (TFAM/nsDNA, PDB code 4NNU) (Ngo, Lovely *et al.* 2014). Furthermore, TFAM in complex with the alternative binding site X (TFAM/X22c) was solved in our laboratory (data not published). Despite the different space groups and DNA sequences, all these structures present similar protein folding, the DNAs are always bent in a U-turn and the steps of Leu58 and Leu182 insertion are separated by 10 bp in all cases. In addition, all complexes show the HMG-box1 antiparallel interaction between symmetry-related partners.

It follows a comparison of the best refined complexes of these five TFAM/DNAs structures (represented by molecule A in all crystals) with complex A/BC from TFAM/Y22 structure. Superposition (from aa 44 to 236) of complex A/BC of TFAM/Y22 to the other structures yielded variable r.m.s.d. values reported in Table_9, with no major differences between complexes except for molecule A from TFAM/X22c, with highest r.m.s.d. value of 2 Å. In this case, the whole polypeptide chain presents slight positional differences with most deviated regions located at the loop between HMG-box1 and the linker, and at the tip of HMG-box2 short L-arm (loop between helix 1 and helix2). In all cases no big changes in protein fold or protein/DNA major contacts were apparent. However, at the level of the DNA, the position of the DNA ends varies from one sequence to another.

Table_9. R.m.s.d. values obtained from superposition of all known TFAM/DNA structures, (Molecule A) against complex A/BC of TFAM/site Y22 (aminoacids 44 to 236).

	Overall r.m.s.d. with complex A/BC (Å)
TFAM/LSP28	1.18
TFAM/LSP22	1.6
TFAM/HSP	1.12
TFAM/nsDNA	1.2
TFAM/site X22c	2.0

In order to better analyze structural changes between the different complexes, the HMG-box1 (residues 44-120) or HMG-box2 (residues 161-236) of the five aforementioned TFAM/DNA structures was superposed onto the equivalent domain of TFAM/Y22 A/BC complex. The r.m.s.d. values are reported in Table_10. Overall, these superpositions showed a good fitting of HMGbox1 from A/BC with HMGbox1 from either TFAM/LSP22 or LSP28. Slightly higher r.m.s.d. appeared for TFAM/HSP, /nsDNA and /X22c complexes, which are mostly due to crystal contacts (e.g. at TFAM/HSP Pro73 C^{α}, that is 1.51 Å far from TFAM/Y22 Pro73 C^{α}), a main chain displacement correlated to a DNA strand movement (which e.g. affects Val54 C^{α} in TFAM/Site X, 1.50 Å) or differences at the N-terminus (e. g. TFAM/nsDNA, 3.28 Å Val45 C^{α}). Similarly few and not significant differences were found by superposition by HMG-box2.

Table_10. R.m.s.d. values obtained by superposition of HMGbox1 and HMGbox2 of all TFAM/DNA structures.

	HMG-box1 A/BC (Å)	HMG-box2 A/BC (Å)
TFAM/LSP28	0.43	0.46
TFAM/LSP22	0.45	0.58
TFAM/HSP	0.78	0.77
TFAM/nsDNA	0.72	0.49
TFAM/site X22c	0.73	0.39

Furthermore superposition by one HMG-box shows important changes in the relative positions between both boxes (Figure_39). By superposing HMG-box 1, the most prominent ‘displacement’ of HMG-box 2 occurs between the TFAM/Y22 and TFAM/X22c structures, at Asp176 C^{α} (7.8 Å) located at the loop between helix 1 and helix 2 (the tip of the L-arm) (Figure_39). This movement correlates with a displacement of the DNA, which at the Asp176 region is kinked. The structures TFAM/HSP, TFAM/nsDNA, TFAM/LSP22 and TFAM/LSP28 are similar to TFAM/Y22. The difference with TFAM/X22c is probably the long, poly-A tract in site X ($A_5A_6A_7A_8A_9A_{10}T_{11}T_{12}T_{13}$), which confers rigidity to this sequence that alters the geometry of the binding (described by Pablo Fernández Millán in his PhD thesis:

"Estudio estructural de proteínas implicadas en el metabolismo del genoma mitocondrial: Helicasa y Factor de Transcripcíon A mitocondrial, TFAM", Universitat de Barcelona). On the other hand by superposing HMG-box 2, the most prominent 'displacement' of HMG-box 1 occurs between TFAM/LSP28 and TFAM/Y22c complexes at residues Asp93 (13.5Å) (Figure_39).

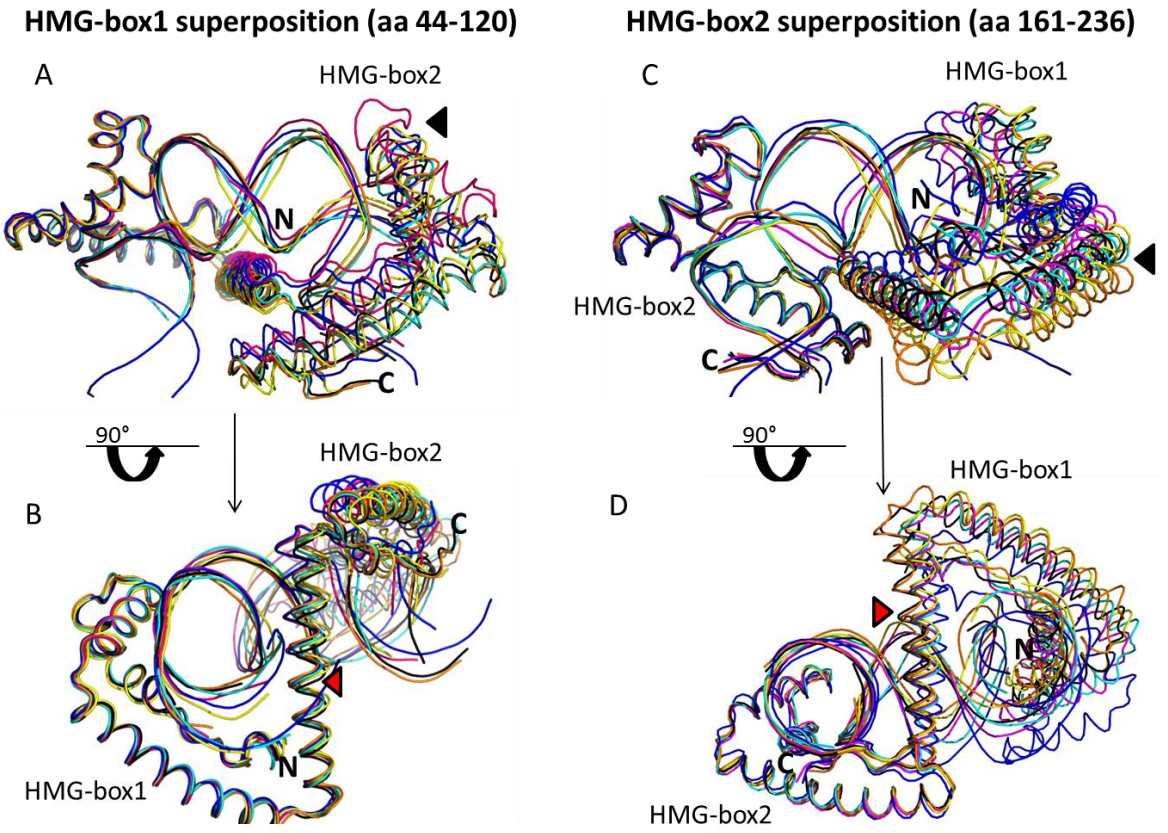


Figure_39. Superposition for HMG-box1 and 2 of the six TFAM/DNA structures.
In **A** and **B** the superposition by HMG-box1 is shown in two different, 90º-related orientations. In **C** and **D** superposition by HMG-box2 is also shown in two 90º-related orientations. The axis of rotation is indicated. The point of maximal distance between the complexes is indicated with a black arrowhead. With a red arrowhead is indicated the position of His137. In both superpositions it was used the same color code: site Y is represented in black, site X in yellow, LSP22 (3TQ6) in purple, HSP (4NOD) in light blue, LSP28 (3TMM) in dark blue and nsDNA (4NNU) in orange.

One possible reason for the variability of the relative positions between HMG-boxes among crystals could be differences at the connection between each HMG-box and the linker, where hinge points could exist. However, the analysis of the interactions

between these regions suggested a surprising stability of this contact. The N- and C-terminal ends of the linker form a small hydrophobic core with HMG-box1 and HMG-box2, respectively. The hydrophobic core between HMGbox1 and linker involve Val45, Leu46 from the N-terminus of the protein, Phe117 and Leu121 from helix3, and Ile126, leu129 and Ile133 from the linker. In addition, the carboxylate group of Glu132 interacts with the main chain amides (aa 44 and 45) of the protein N-terminus. The same occurs between the linker and HMG-box2: Leu149 and Leu152 from the linker, Met222 and Try218 from helix3, and Leu230 from the C-terminal tail form a second hydrophobic core, which is stabilized by the side chain methylene groups from Glu148 (linker) and Arg227 (helix3) which, in addition, form a salt bridge. Taking into account the important movements between HMG-boxes, the orientation of the side chains in these hydrophobic cores is unexpectedly very similar among structures (TFAM/site Y, TFAM/site X, TFAM/HSP, TFAM/nsDNA, TFAM/LSP22 and TFAM/LSP28 Figure_40). Therefore, the connections between the linker and the HMG-boxes are not structural hinges. Superposition of the linker N-terminal region (aa 123 to 136) onto TFAM/site Y showed again (as for the four molecules in the site Y crystal) that His 137 is indeed a hinge point where structural divergences start and progress towards the linker C-terminus (5.6 Å at Leu152). As pointed out above, at His137 start the contacts with the DNA narrow groove, which in all structures are essentially hydrophilic and non-specific since they occur only with the DNA backbone. Therefore, the different positions of HMG-box2, which grabs the contacted DNA, correlate with a structural rearrangement at the linker.

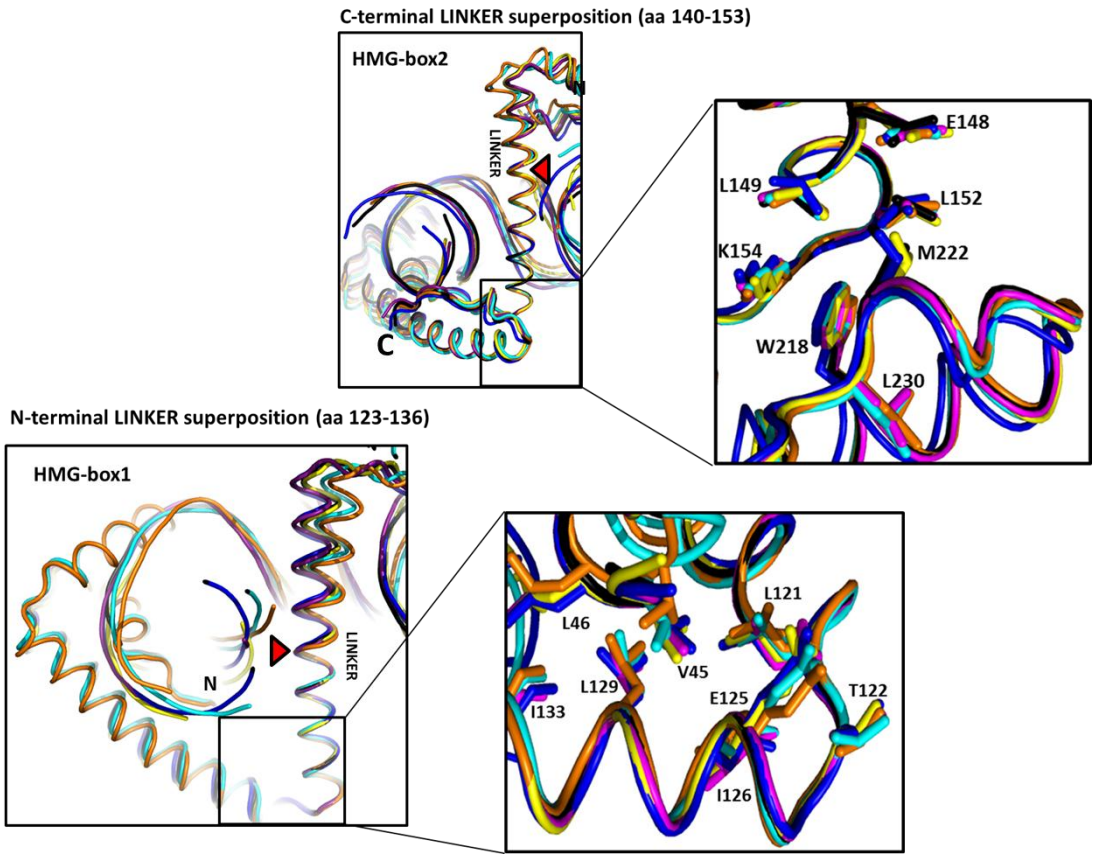


Figure_40. N-terminal and C-terminal linker/HMG-box 1 and 2 interactions.
The N-terminal (aa123-136) and C-terminal (aa140-153) linker superposition of the six TFAM/DNA complex structures is shown. The residues that are implicated in interactions are indicated using one letter code. With a red arrowhead is indicated the position of His137. The color code of the six complexes is black for TFAM/Y22, yellow for TFAM/X22c, purple for LSP22 (PDB ID 3TQ6), light blue for HSP (4NOD), dark blue for LSP28 (3TMM) and orange dor nsDNA (4NNU).

In addition to the linker torsion at the His137, superposition of all the six TFAM/DNAs structures by the linker N-terminal region (aa Pro123 to Leu152), shows that there is also a reorientation of HMG-box1 with respect to the linker. This reorientation is more severe, with a maximal distance at residue Glu90 for both TFAM/HSP and TFAM/nsDNA (12.6Å, 6.14Å respectively with respect to TFAM/Y22). Such displacement of HMG-box1 involves residues from Cys48 to Lys118 and is responsible of its pendulum-like behaviour to respect the linker (Figure_41). HMG-box1 mobility may be related to its binding and bending activity.

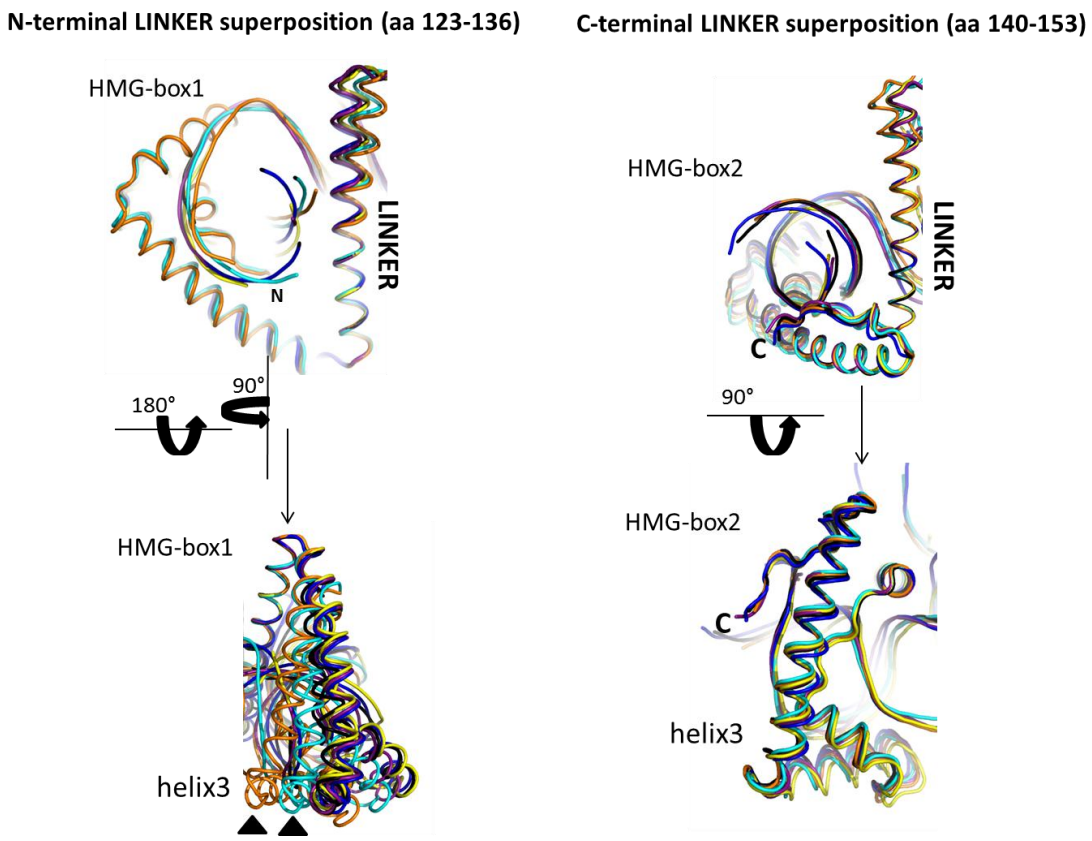


Figure_41. HMG-box1 and HMG-box2 orientation with respect to the linker.
N-terminal (aa 123-136) and C-terminal (aa 140-153) linker superposition of the six TFAM/DNA complex structures are shown, evincing the HMG-box1 reorientation. By superposition of the C-terminal part of the linker no reorientation of HMG-box2 is observed. With black arrowheads the highest displaced regions of the HMG-box1 domains are indicated. The color code of the six complexes is black for TFAM/Y22, yellow for TFAM/X22c, purple for LSP22 (PDB ID 3TQ6), light blue for HSP (4NOD), dark blue for LSP28 (3TMM) and orange dor nsDNA (4NNU).

As observed in the superposition of HMG-boxes domain1 and 2 in Figure_39, it appears that the DNA region that superimposes better is the one contacting HMG-box1. This may agree with what reported in literature that the binding is driven by HMG-box1 and that HMG-box2 has a secondary role in binding and an important function in DNA bending and complex stabilization (Gangelhoff, Mungalachetty et al. 2009; Wong, Rajagopalan et al. 2009; Ngo, Kaiser et al. 2011; Rubio-Cosials, Sidow et al. 2011).

5.2 Computational, biochemical, and biophysical studies of TFAM in complex with site Y, site X and LSP DNAs.

Crystallographic studies of TFAM reveal that this protein systematically induces a 180° DNA bending when bound to short DNA fragments (from 22 to 28 bp) spanning the cognate binding sites at the mtDNA promoters (HSP and LSP), or when is bound to additionally identified binding sites (such as sites X and Y) or unspecific sequences (Ngo, Kaiser et al. 2011; Rubio-Cosials, Sidow et al. 2011; Ngo, Lovely et al. 2014). However, the unexpected disorder of sequence Y22 in the crystal prompted us to better characterize the molecular mechanism and the stoichiometry of TFAM binding by computational, biochemical and biophysical studies. In this chapter, the DNA properties and bending energy are analyzed by molecular dynamics (MD) studies. Next, a differential recognition of LSP22, X22c and Y22 sequences by TFAM was identified by EMSA. Additional characterization of formation of TFAM/LSP22, TFAM/Y22 and TFAM/X22c complexes by isothermal titration calorimetry (ITC) is further shown. This is followed by the analysis of the oligomerization of TFAM either alone or in complex with Y22, X22c and LSP22 by multi-angle laser light scattering (MALLS) and analytical ultra-centrifugation (AUC). The DNA sequences used in all these studies are reported in Table_11 and are the ones crystallized in complex with TFAM described in the above TFAM/Y22, TFAM/X22c and TFAM/LSP22 structures.

Table_11. TFAM cognate sequences.

Y22	5′TAACAATTGAATGTCTGCACAG 3′ 3′ATTGTTAACTTACAGACGTGTC 5′
X22c	5′TAACAAAAAATTTCCACCAAAC 3′ 3′GATTGTTTTTTAAAGGTGGTTT 5′
LSP22	5′TAACAGTCACCCCCCAACTAAC 3′ 3′ATTGTCAGTGGGGGGTTGATTG 5′

5.2.1 Flexibility of site Y, site X and LSP sequences and TFAM bending energies determined by molecular dynamics.

Molecular dynamic studies of DNA Y22, X22c and LSP22 were performed in collaboration with Federica Battistini from Modesto Orozco laboratory, at the Institute of Research in Biomedicine (IRB), Barcelona. Physical properties of these three

sequences at the equilibrium were firstly calculated. Six parameters describe the base pair step translational and rotational movements: shift, slide, rise, tilt, roll and twist (see section 4.4, Figure_20). DNA flexibility depends on DNA base pair step behaviour along those translational and rotational coordinates in time (Travers 2004). Thus an average value of each base pair parameter along the trajectory for each base pair step was obtained. For each DNA, all these 6 parameters were calculated. In order to provide a rough estimate of flexibility at base pair step level, stiffness constants corresponding to the six parameters mentioned above (kshift, kslide, krise, ktilt, kroll and ktwist) were extracted from the diagonal of the stiffness matrix and the total stiffness (K_{tot}) was obtained as a product of these six constants (Figure_42).

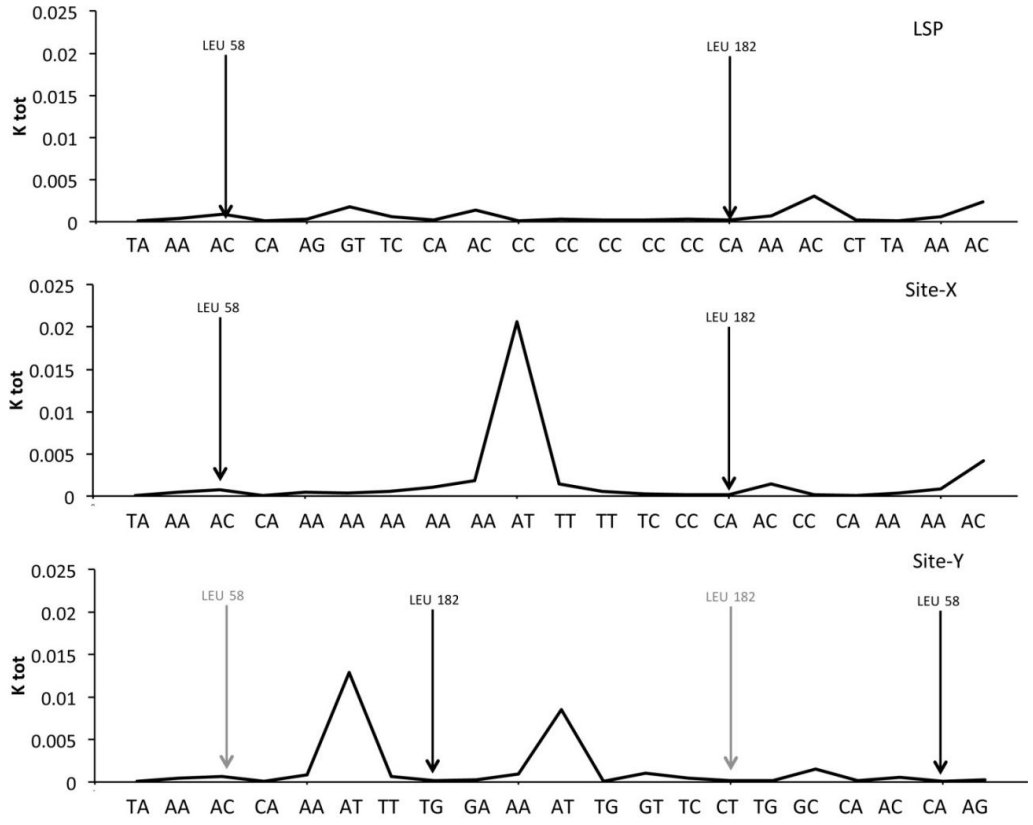


Figure_42. Stifness (K_{tot}) of LSP, Site X and Site Y sequences.
K_{tot} values of site Y (represented by the Y22 sequence in this work), site X (X22c) and LSP (LSP22) are shown. The peaks correspond to the stiffer base pair steps. For the three sequences the insertion sites found in the X-ray structures (by Leu58 and Leu182) are indicated by black arrows. For the sequence Y22 the crystallographic insertion sites and the predicted (Fisher, Topper et al. 1987) sites are indicated by black and grey arrows, respectively.

The K_{tot} of the three DNAs was compared (Figure_42). The higher is the value of the K_{tot}, the stiffer is the base pair step. As shown in Figure_42, the LSP22 sequence has rather flexible steps alternated by mildly stiff steps. In the case of site X22c, it is mainly flexible except at steps that correspond to the asymmetric A-tract $A_5A_6A_7A_8A_9A_{10}T_{11}T_{12}T_{13}$, with highest stiffness at the AT junction. This stiff region is located between the HMG-boxes, which only contact the A-tract ends. In addition, both LSP22 and site X22c insert Leu58 in an AC/GT step, which is slightly stiffer than the CA/TG step inserted by Leu182. Regarding Y22, the steps inserted by TFAM are both flexible. On the other hand, Site Y presents two rigid points at steps A_6T_7 and $A_{11}T_{12}$. In the TFAM/Y22 structure, precisely at the rigid region $A_5A_6T_7$ the narrow groove widens and allocates Arg157. This residue indeed goes deep into the groove, where its methylene groups "push" T7 phosphate, whereas its NH2 contacts both O2 from T_{18} (chain B and complementary to A_5) and O4 from the G_{19} ribose. Whilst this rigid region is deformed, the inserted step is highly flexible. The second rigid region $A_{10}A_{11}T_{12}$ is found between the two HMG-boxes.

The average base pair parameters (shift, slide, rise, tilt, roll and twist) along the MD trajectories were calculated for the three naked DNA sequences LSP22, Y22 and X22c, and the values were compared to the ones of the DNA in complex with the protein in the corresponding crystal structures. Since the analysis of all X-ray crystal structures showed that the parameter most affected by the binding is the roll (Figure_43), it was decided to analyse its variation. In this case, higher roll angles reflect more bent steps. This analysis identified the steps that are either more bent in the complex or are intrinsically bent in the naked sequence. Thus, the naked sequences showed that they do not follow a structure of an ideal B-DNA conformation since some steps have different degrees of bending, and thus the three sequences are intrinsically deformed (Figure_43). All naked DNAs show that the step that HMG-box2 Leu182 inserts has a high roll (is opened). Instead, whereas in LSP22 and X22c Leu58 inserts at closed steps with lower roll, in Y22 the roll is high (is opened). Therefore, only in Y22 TFAM inserts the two leucines in opened steps.

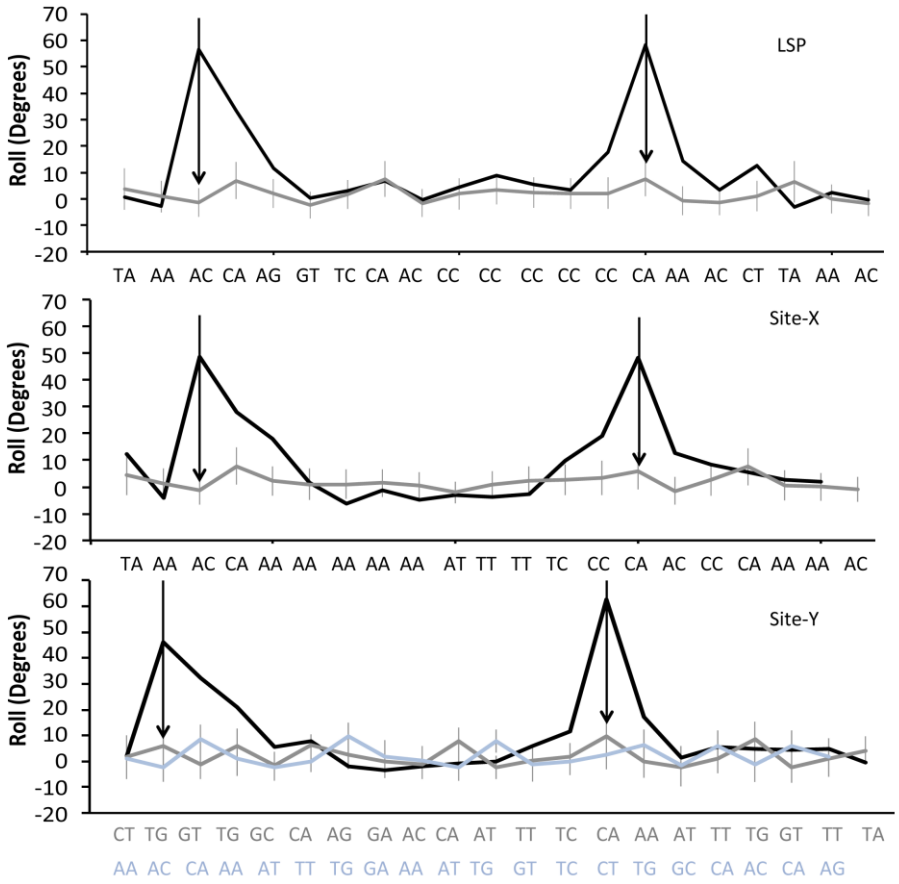


Figure_43. Roll values for naked and protein-bound DNA.
Roll values (in degrees) of Y22 (site Y), X22c (site X) and LSP22 (LSP), for the naked (grey) and protein-bound DNA (black) is shown. The high roll peaks in the crystal structures correspond to the insertion sites, indicated by the arrows (left arrow, Leu58 insertion site; right arrow Leu182 site). For Y22 both the orientation of the sequence as assigned in the crystal (grey) or as previously predicted (light blue) (Fisher, Topper et al. 1987) are indicated.

Finally, all the equilibrium and stiffness values of each sequence were used to calculate the deformation energy (ΔE_{def}), which is the elastic energy necessary for the naked DNA to adopt the conformation in the corresponding crystal. Thus, four values of ΔE_{def} were determined:

ΔE_{def} (LSP) = 8.4 Kcal/(mol base pair)

ΔE_{def} (X) = 6.3 Kcal/(mol base pair);

ΔE_{def} (Y) = 5.5 Kcal/(mol base pair);

ΔE_{def} (Y_{Fisher}) = 8.7 Kcal/(mol base pair).

In the case of site Y two different ΔE_{def} were calculated: one considering the orientation of the sequence assigned based on the density (ΔE_{def}(Y)) and the other following the orientation predicted according to the alignment of Fisher in 1987 (ΔE_{def}

(Y_{Fisher}), shown in Figure_17). Interestingly, the assigned site Y orientation is more prone to be bent to the conformation observed in the crystal than the orientation initially proposed (Fisher, Topper et al. 1987).

5.2.2 Differential binding of TFAM to the 22bp DNA fragments Y22, X22c and LSP22 analysed by EMSA.

The apparent K_d of TFAM for fragments X22c, Y22 and LSP22 was estimated by EMSA using 1nM ^{32}P-labelled DNA probes titrated with TFAM at very low concentration (at the nM range). This was done since TFAM at higher concentrations changed its DNA binding properties, which is a phenomenon typically observed in proteins influenced by crowding conditions. Therefore The experiments at such low concentration avoided transient and unspecific self-assembly effects (quinary structures) due to crowding (Wirth and Gruebele 2013). In gerenal terms, the phenomenon of macromolecular crowding alters the properties of molecules in a solution when high concentrations of macromolecules such as proteins are present. Such conditions occur routinely in living cells. The high concentrations of macromolecules reduce the volume of solvent available for other molecules in the solution, which determine an increase of their effective concentrations.

EMSA gels using 1nM ^{32}P-labelled DNA probes encoding X22c, Y22 and LSP22 are reported in Figure_44A. By using a modified Hill equation (see section 4.5), the estimated apparent K_d for LSP22 was found to be ≈ 9.21 nM (calculated error: +/- 0.98), thus in aggrement with the 1-10 nM range measured by other teams (Gangelhoff, Mungalachetty et al. 2009; Malarkey and Churchill 2012; Brown, Tkachuk et al. 2015). For Y22 sequence the apparent K_d was ≈ 4.41 nM (calculated error: +/- 2.32) and for the X22c sequence the apparent K_d was ≈ 13.63 nM (calculated error: +/- 2.89), suggesting a slightly higher affinity for the Y22 sequence (Figure_44B). Each experiment was repeated three times. For the three ligands the binding of TFAM was not cooperative (The Hill coefficient was always around 1).

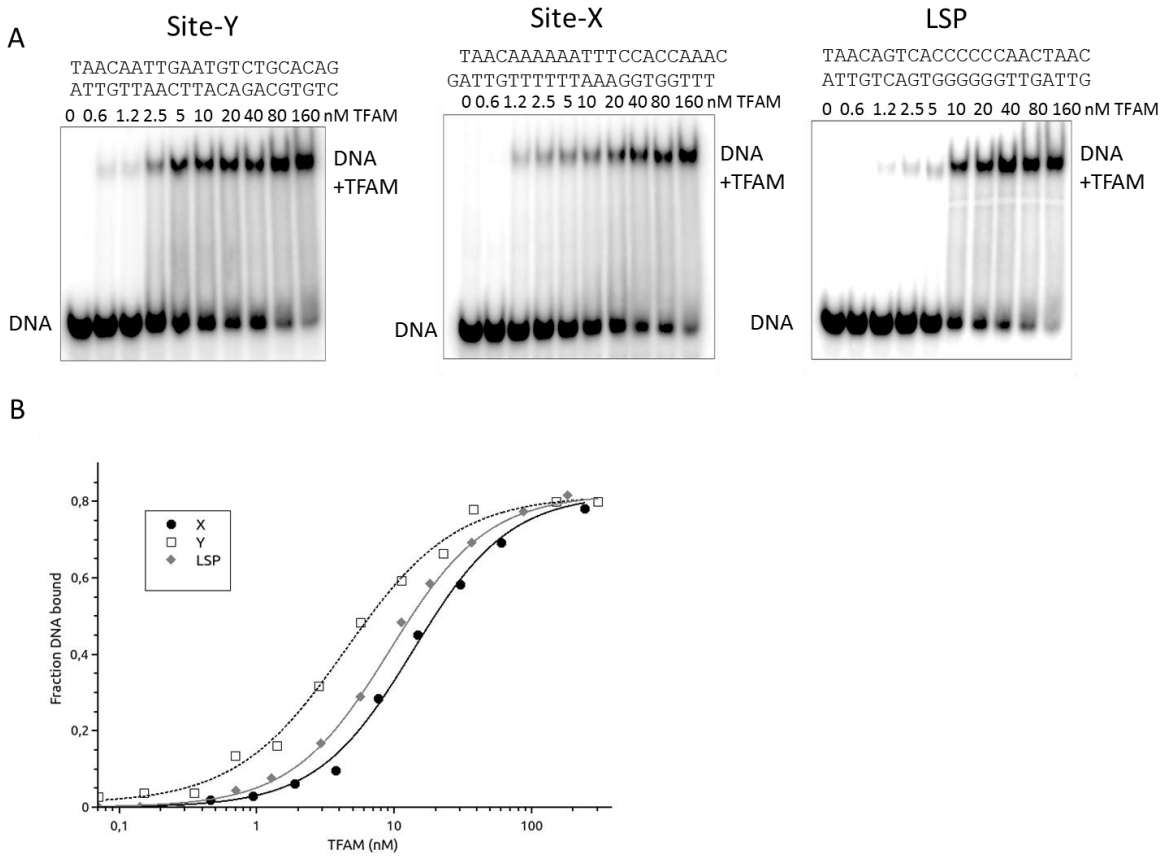


Figure_44.TFAM/site Y, TFAM/siteX and TFAM/LSP binding and affinity.
In **A**, gel-shift assays (EMSA) show TFAM inducing equivalent DNA migration to all sequences. Free DNA (1nM) was loaded in lane "0". In subsequent lanes from left to right, 1 nm of ^{32}P-labeled DNA was titrated with increasing amounts of TFAM indicated above the corresponding lanes. The DNA shift corresponding to the complexes is indicated as "DNA+TFAM". In **B,** the measurements of TFAM/DNA binding are reported in a graph. The data were analysed with the program "MicroCal Origin". As shown in the graph legend TFAM in complex with site X is indicated by black circles, TFAM in complex with site Y is indicated by squares, while TFAM in complex with LSP is indicated by rhombus.

To further compare the binding of TFAM to the different sequences and confirm that the protein bound to all of them equally, competitions experiments were realized using 5' fluorescently labelled X22c and Y22 DNA fragments, as shown in Figure_45. TFAM complexes with fluorescein-X (F-X) or fluorescein-Y (F-Y) sequences (0.4 μM) were assembled with an excess of TFAM, at a 2 TFAM:1 DNA ratio chosen based on the DNA full shift observed in Figure_26. These complexes were competed with increasing amounts of unlabelled LSP22 fragments from 0.2 μM to 3.2 μM (Figure_45, upper panels). This experiment resulted in displacement of TFAM from F-X to LSP22 at the second point of LSP22 titration. The displacement was not completed, since remnant TFAM/F-X complex was found in subsequent lanes

suggesting eventual equal amounts of TFAM/F-X and TFAM/LSP22 complexes. Thus, this confirmed similar affinities of TFAM for both fragments. Instead, and to our surprise, displacement of TFAM from F-Y to LSP22 was importantly delayed; indeed the complex TFAM/F-Y was still present at 4x-fold excess of competitor. This indicates that TFAM forms a more stable complex with fragment Y22 than with LSP22 fragment. Since LSP22 competed X22c and Y22 differently, despite all had similar apparent K_d's at the nanomolar range, we analysed the competition between Y22 and X22c for TFAM binding. The lower panel of Figure_45 shows that a non-labelled fragment Y22 strongly competed the TFAM/F-X complex, while a non-labelled X22c fragment poorly displaced F-Y from the TFAM/F-Y complex. Therefore, TFAM showed a clear preference for Y22 over both X22c and LSP22 despite all of them have similar apparent affinities.

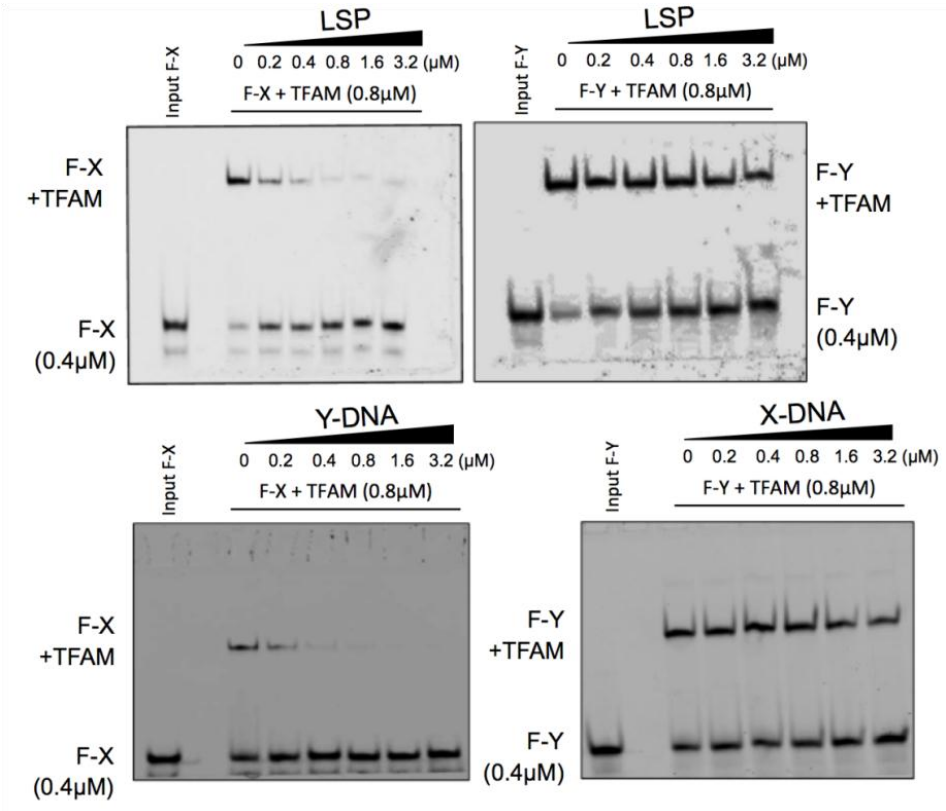


Figure_45. Differential binding of TFAM to site X, site Y and LSP.
The upper panel shows competition by unlabelled LSP22 of X22c (F-X, labelled with 5'fluorescein) or Y22 (F-Y, also labelled with 5'fluorescein) bound to TFAM (lane 0). Serial increasing concentrations of LSP22 (lanes 0.2 to 3.2μM) were added as indicated. The lower panel shows competition of F-X (left) or F-Y (right) by unlabelled Y22 or X22c fragments respectively. Also in these cases a serial increasing amount of unlabelled Y22 or X22c (lanes 0.2 to 3.2μM) were added to the labelled complexes. The labelled complexes were formed with 0.4μM labelled DNA and 0.8μM TFAM in 50mM Tris-HCl pH 7.5, 60mM NaCl, 5mM DTT. The input labelled DNA control is indicated as well as the concentrations of the unlabelled competitor.

5.2.3 Binding analysis of TFAM to site Y, site X and LSP by Isothermal Titration Calorimetry (ITC).

In order to analyse the thermodynamic properties during formation of the different TFAM/DNA complexes, ITC experiments were performed. ITC calorimetric measurements of TFAM/X22c and /LSP complexes were already described by Pablo Fernández Millán in his PhD thesis: "Estudio estructural de proteínas implicadas en el metabolismo del genoma mitocondrial: Helicasa y Factor de Transcripcíon A mitocondrial, TFAM", Universitat de Barcelona). Follow ITC analysis of TFAM/Y22 compared with the other two complexes. As shown in Figure_46, these assays showed that TFAM binds to LSP22, X22c and Y22 in an endothermic process driven by favourable entropy (ΔS) that compensates an unfavourable enthalpy (ΔH). In this process, the variation of free energy of the system is negative (ΔG<0) (Table_12), indicating a spontaneous process. Considering the enthalpy variation, the TFAM/Y22c complex is the one that presents the highest (unfavourable) value of ΔH, also in this case compensated by an increase in entropy. The data collected from the three complexes fit well with "one set of site" algorithm (Figure_46, bottom). This means that the two TFAM HMG-box domains can bind without affecting each other to one DNA molecule in just one single step and always in the same manner. From the intercept of the fitting curve it was possible to determine the stoichiometry of the binding, indicated with *n* in Table_12, that corresponds to the molar ratio of DNA per TFAM molecules and in the three cases was around 0.36-0.4 (Figure_46). This is consistent with at least two molecules of TFAM bound to one molecule of DNA. From the values of ΔG it was possible to calculate the affinity constant K_a reported in Table_12. The three constant values are very similar and do not indicate any preference of TFAM for one of these sequences. K_a is the inverse of K_d. The values of K_d obtained by ITC compared with the K_d values obtained by EMSA (see section 5.2.2) are 10 fold or more higher. This can be explained considering that ITC data were collected using macromolecules in high concentration (at the micromolar range, versus nanomolar range for EMSA, see material and method section 4.6) and thus subjected to crowding or exclusion volume force. This phenomenon affects the enthalpy and entropy variation related not only to the binding but also to the

environment changes, substantiating the discrepancy in the affinity measurements of TFAM binding by ITC and EMSA. Furthermore, since the value of enthalpy has been correlated with the release of water molecules from the DNA minor groove when this latter interacts with a protein, it can be speculated that the more rigid Y22 DNA may coordinate more water molecules, so is more hydrophilic. It is also possible that the higher value of enthalpy correlates with the kinetics of this complex formation. Thermodynamic analyses with previous HMGB proteins show that endothermic reactions are a specific feature of this protein family (Dragan, Read et al. 2004; Malarkey and Churchill 2012). In addition, endothermic reactions generally involve proteins that bind to the minor groove of the DNA (Jen-Jacobson, Engler et al. 2000) as TFAM does .

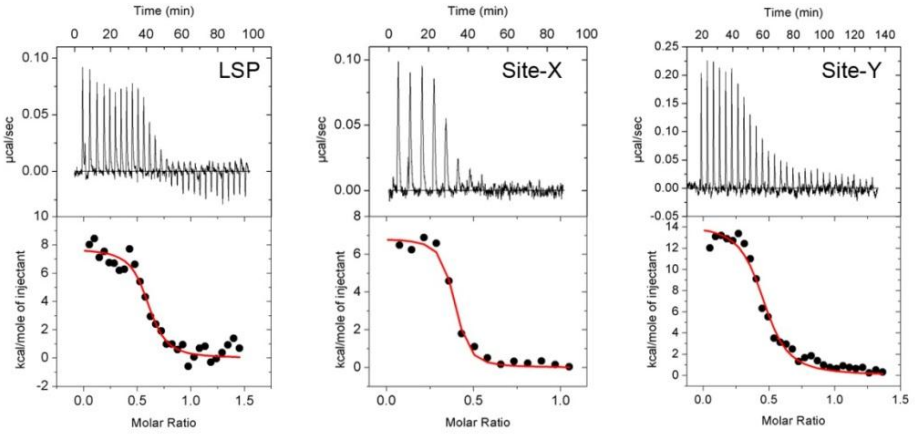


Figure _46. ITC results.
The graphs in the upper part of the figure show a representative result of the three measurements done with the three complexes. The graphs represented at the bottom of the figure derive from the integration of the three series of data obtained for the three complexes.

Table_12. ITC results

DNA Sequences	K_a (M^{-1})	K_d (nM)	n	ΔH (kcal.mol^{-1})	ΔS (kcal.mol^{-1}.K^{-1})	ΔG (kcal.mol^{-1})
LSP	1.30E7	76.92	0.59	7737	58.9	**-9521**
LSP	1.32E7	75.75	0.22	10430	68.2	**-9553**
LSP	1.09E7	91.74	0.39	6939	55.9	**-9440**
SiteX	4.25E6	235.2	0.32	7721	56.7	**-8892**
SiteX	3.10E7	32.25	0.36	6851	57.6	**-10026**
SiteX	3.36E7	29.76	0.41	8566	63.6	**-10069**
SiteY	1.11E6	900.9	0.42	19270	93.4	**-8096**
SiteY	6.77E6	147.7	0.46	14270	79.9	**-9141**
SiteY	9.89E6	10.11	0.22	12840	75.8	**-9369**

5.2.4 TFAM oligomerization: Size exclusion chromatography combined with multi angle laser light scattering (SEC-MALLS), and analytical ultra-centrifugation (AUC) analyses.

Our EMSAs with TFAM/Y22 and TFAM/X22c complexes suggested assembly of more than one protein with one binding site (Figure 24 and 26). Both the length of the DNA molecules (22 bp) and the dimensions of the U-turns found in the crystals suggested that it is difficult to envisage a mechanism by which more than one protein could contact one single DNA molecule. EMSAs of TFAM in complex with either the long linked X-Y form or the isolated short X22c and Y22 forms, and at a 4x-fold or higher excess of TFAM, showed a progressive conversion of the first single band shift into species of lower mobility, suggesting TFAM multimerization as observed in the past by other groups (Gangelhoff, Mungalachetty et al. 2009; Wong, Rajagopalan et al. 2009; Ngo, Lovely et al. 2014). This multimerization might be induced by crowding forces that promote transient and not specific protein-protein interactions also known as quinary structures (Wirth and Gruebele 2013). To clarify the content of protein and DNA in TFAM/Y22, X22c and LSP22 complexes, experiments with size exclusion chromatography coupled with multi angle laser light scattering (SEC-MALLS) and analytical ultracentrifugation (AUT) were performed.

5.2.4.1 SEC-MALLS analysis of TFAM in complex with site Y, site X and LSP DNAs.

In order to determine the molecular weight (*Mw*) of the different TFAM/LSP22, TFAM/Y22 and TFAM/X22c complexes, SEC-MALLS experiments were performed in collaboration with Drs. German Rivas Caballero and Carlos Alfonso Botello (Centro de Investigaciones Biológicas, CIB-CSIC, Madrid). Firstly each 22bp DNA fragment was injected into the size exclusion column and in all cases the fragments eluted at 11 ml elution volume (Figure_47). The experimental molecular weight (Mw_{obs}) was estimated to be 16 kDa, higher than the one calculated *Mw* (Mw_{calc}) for each of them, of 13.5 kDa. *Mw* error estimation is due to co-elution of contaminants

with the target DNA. These contaminants were the origin of the peaks at 10.5 ml elution volume (Figure_47). In particular in the case of the LSP22 fragment, it presented a DNA contamination due to the tendency of its G-rich single strand to self-assemble into G-Quadruplexes (Figure_47).

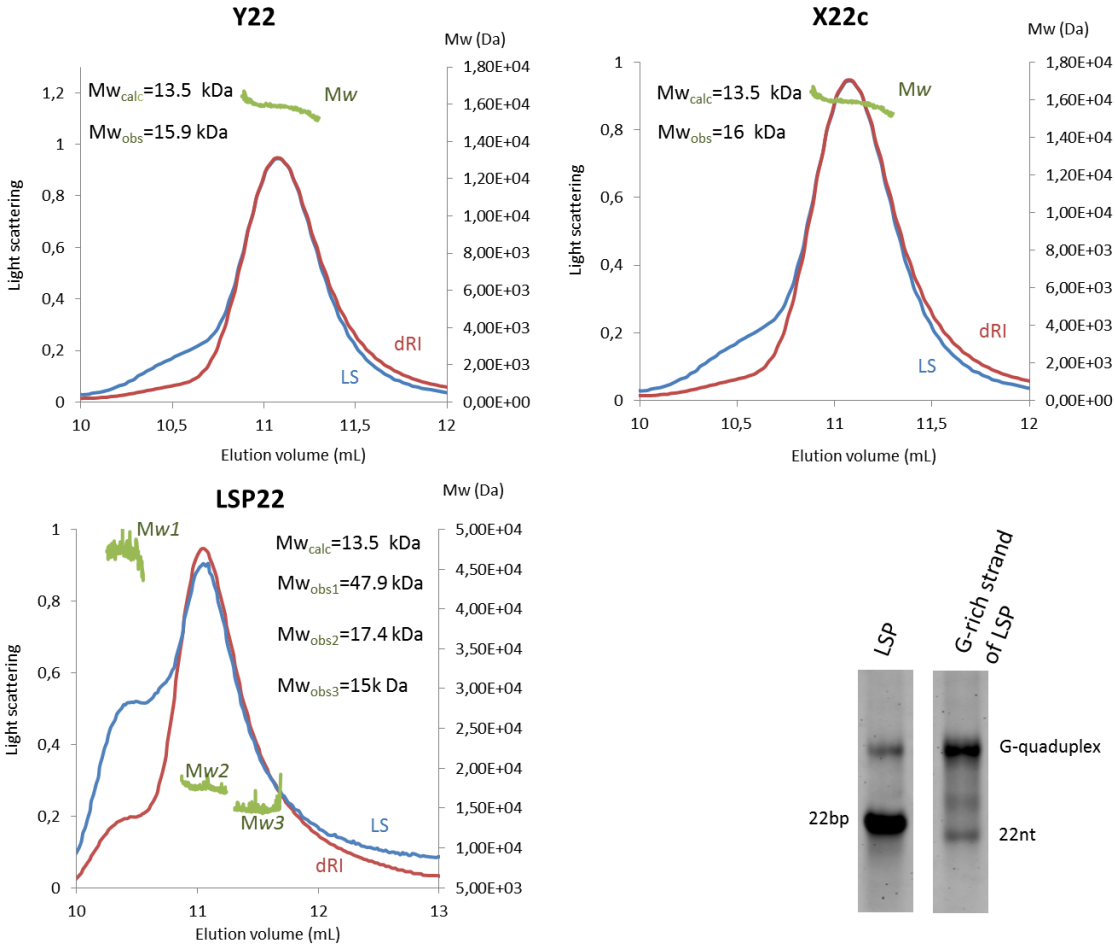


Figure_47. SEC-MALLS analysis of Site Y, Site X and LSP.
In each graph the light scattering curve (LS-curve) is shown in red, the differential refracting index curve (dRI-curve) in blue and the molecular weight peaks in green. On the left Y axis is indicated the light scattering while on the right Y axis the molecular weight of the sample is shown. A native gel that shows the migration of the LSP duplex demonstrates the band corresponding to the G-quadruplex.

TFAM alone was next injected, which eluted in two peaks at 10.2 ml and 12 ml with a Mw_{obs} of 57 kDa and 30 kDa respectively. The Mw_{cal} of TFAM monomer is of 25.6 kDa. Thus, the 56 kDa species corresponds to a stable dimer formed at the concentration of the assay ([TFAM] = 97.6 uM) (Figure_48) whilst the 30 kDa

corresponds to the monomer. Therefore, in the tested buffer conditions monomers and dimers of TFAM coexist.

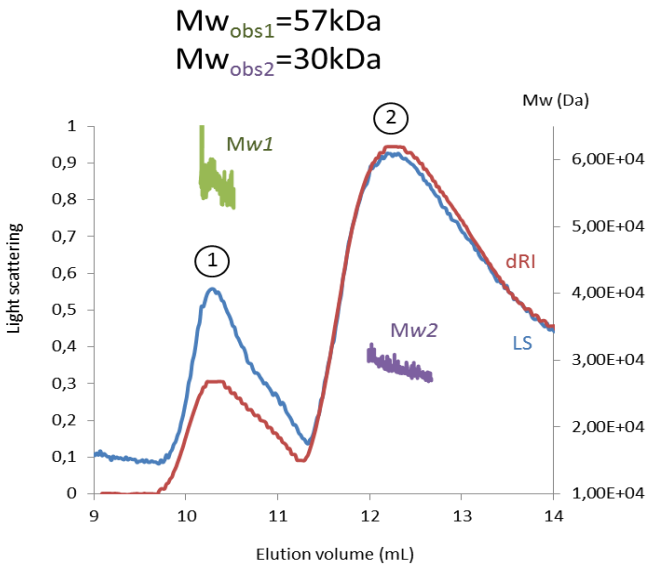


Figure_48. SEC_MALLS analysis of TFAM.
In this graph the molecular weight of TFAM in the absence of DNA is shown. On the left Y axis the light scattering is indicated, while on the right Y axis the molecular weight of the sample is shown. The light scattering curve appears in blue, while the differential refracting index curve is in red. In green are represented the molecular weight values of elution peaks 1 (Mw1) and 2 (Mw2). As indicated above the curve, the two molecular weight peaks correspond to the protein dimer (in green) and monomer (in purple).

Once the Mw of the DNAs and TFAM alone were experimentally determined, the analysis of complexes TFAM/Y22 (351.5 uM), TFAM/X22c (321.5 uM) and TFAM/LSP22 (218.7 and 390.6 uM) followed (Figure_49). The analysis of TFAM in complex with either Y22 or LSP22 showed the existence of 62 kDa species in solution with an elution volume of 9.8 ml. These observed molecular weights are compatible with two proteins bound to one DNA molecule (Mw_{calc}=64.7 kDa). In the case of TFAM/LSP22 complex, a 40 kDa species at an elution volume of 10.6 ml is also present and should correspond to one TFAM interacting with one DNA molecule (Mw_{calc}=39.1 kDa). In addition, a peak of 47 kDa appears in the elution profile of TFAM in complex with X22c and Y22 too, with an elution volume of 10.2 ml (Figure_49). This population with a Mw_{obs} of 47 kDa should include a mixture of the two species of 62 and 40 kDa or should constitute a third new species of one protein

bound to two DNAs (Mw_{calc}=52.6 kDa). In order to be sure about the integrity of the complexes, before and after the column injection, the input samples together with the peak fractions were loaded into native gels and the native gels were stained before with SYBR safe and then with Coomassie (Figure_49).

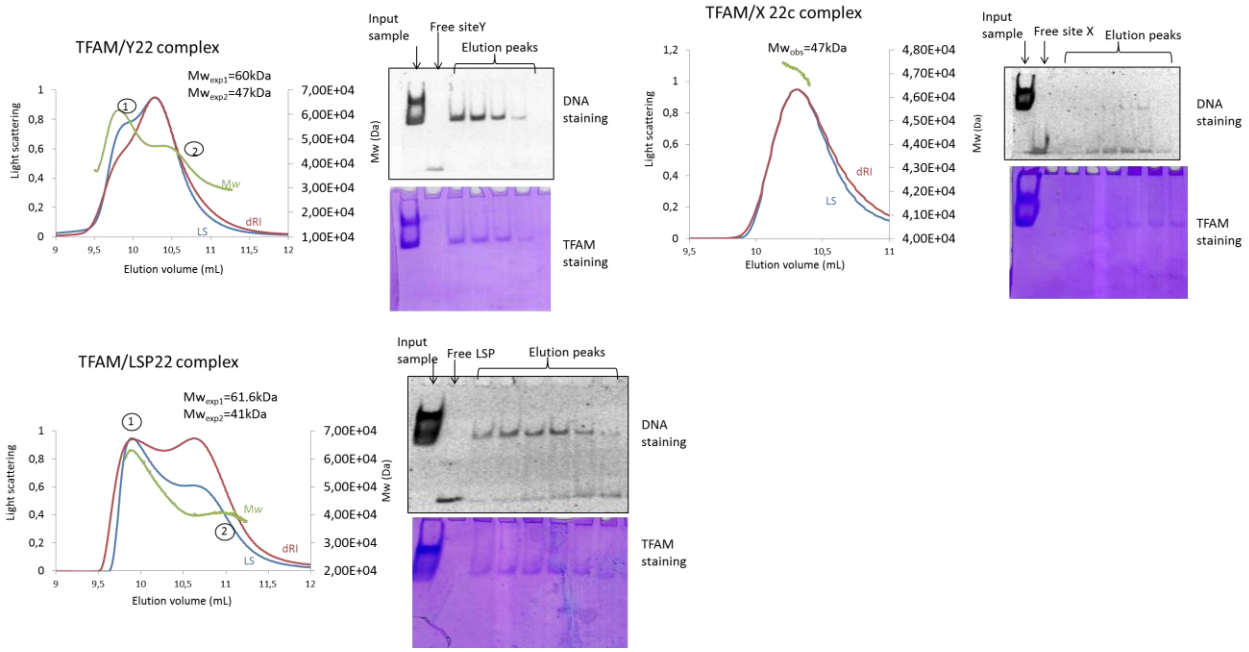


Figure_49. SEC-MALLS analysis of TFAM in complex with site X, site Y and LSP.
In each graph the light scattering curve (LS-curve) is shown in red, the differential refracting index curve (dRI-curve) in blue and the molecular weight peaks in green. The Y axis on the left reports the light scattering while on the right shows the molecular weight. For each complex is shown a native gel with the input sample in the first lane, the free DNA in the second lane and the samples from the elution peaks in the rest of the lanes. These gels was first stained with Sybr Safe and then with Coomassie.

Remarkably, these three different peaks of 62 kDa, 47 kDa and 40 kDa appeared when the experiment was conducted at a low flux of 0.3 ml/min. At a high flux of 0.5 ml/min the high molecular weight species disappeared and just the 40 kDa peaks eluted (Figure_50A). This suggests that higher pressure caused disassembly of the complex. Furthermore, a concentration-dependent oligomeric state was seen for the TFAM/LSP22 complex. Injection of TFAM/LSP22 at a concentration of 218.7 uM yielded a single peak of 41 kDa (one protein bound to one DNA) (Figure_50B). Increasing the concentration to 390.6 uM resulted in a second peak of 62 kDa,

corresponding to two proteins bound to one DNA molecule (Mw_{calc}=64.7 kDa) (Figure_49).

These results indicate there is an equilibrium between TFAM monomer and dimer on the DNA in the time-course of the experiment, which is displaced by the flow of the elution buffer. This monomer-dimer equilibrium appeared also to be concentration dependent as shown in Figure_50B. The only remaining question was to what correspond the species of 47 kDa. To clarify this point we embarked to analytical centrifugation experiments.

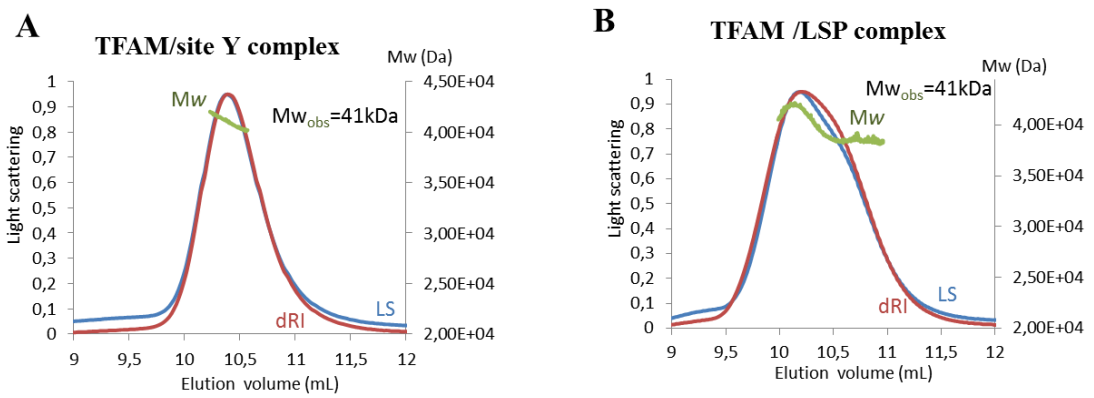
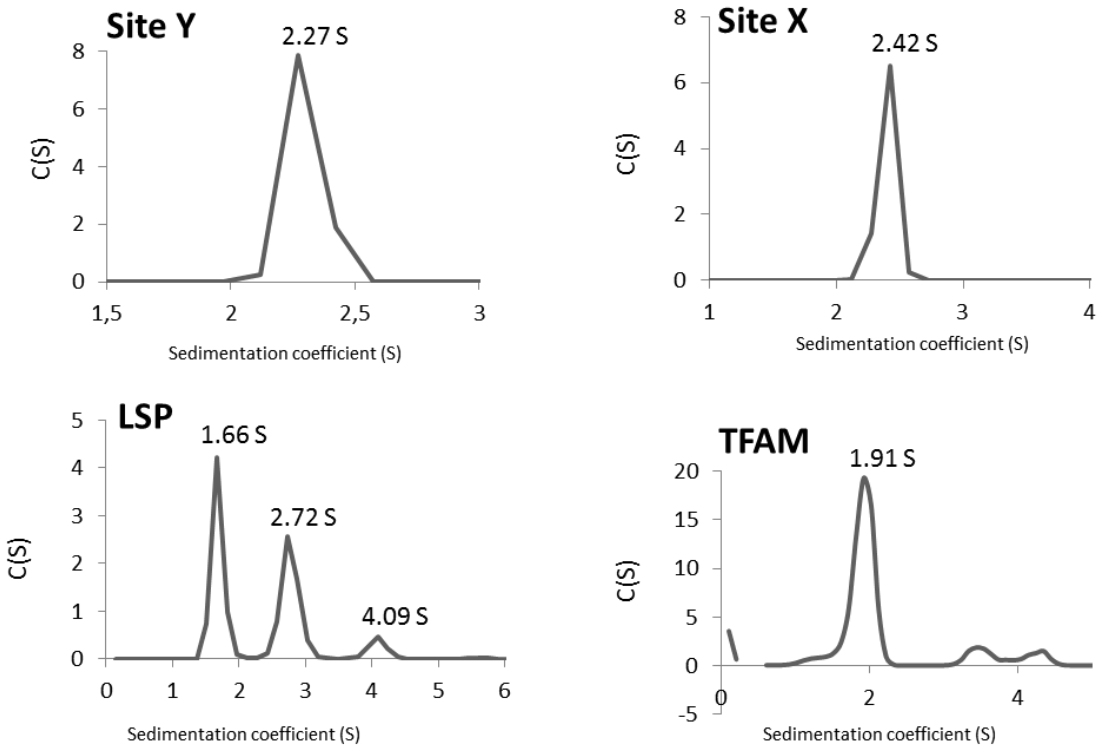


Figure_50. SEC-MALLS analysis of the effect on the TFAM/DNA oligomeric state due to (A) variation in the flux of the chromatography, or in (B) TFAM concentration.

5.2.4.2. Analytical ultracentrifugation analysis of TFAM in complex with site Y, site X and LSP DNA.

In order to better characterize the species obtained by MALLS, analytical ultra-centrifugation was performed. Sedimentation coefficients were obtained by the sedimentation velocity (SV) technique. The data were collected using interference signal and corrected to standard conditions (S_{20w}). Two different ratio of protein-DNA were used for each complex: 2TFAM:1DNA (508 uM) and 4TFAM:1DNA (429 uM). For the DNA fragments alone (200 uM), a peak between 2.3S and 2.7S was observed compatible with the *Mw* of the free DNAs (Mw_{calc}=13.5 kDa) (Figure_51). LSP22

presents minor additional peaks likely corresponding to the G-quadruplex discussed previously (Figure_47). The protein alone (78 uM) shows a sedimentation coefficient of 1.8S compatible with the Mw_{exp} of 25.6 kDa (Figure_51). Regarding the complexes analysis, the peaks with a sedimentation coefficient between 4.6S and 5.0S (S_{20w} correction: 4.5 +/-0.2 S) are compatible with the ratio 2 proteins to 1 DNA with a *Mw* expected (Mw_{exp}) of 64.8 kDa; while the peaks with a sedimentation coefficient bigger than 5.0S are compatible with 4 proteins bound to 1 DNA molecule with a Mw_{exp} of 116.1 kDa (Figure_51). In the analysis of the ratio 2 proteins: 1 DNA, also another peak around 3.2S and 3.4S was observed (S20w corrections: SiteY: 3.6 +/0.2S; Site X: 3.65 +/- 0.1S; LSP: 3.75 +/- 0.1S). This peak is compatible to the Mw_{exp} of 39.1 kDa, corresponding to one protein bound to one DNA (Figure_51). In the case of the ratio 4 protein: 1 DNA a peak compatible with free DNA is observed that is not present in the ratio 2 proteins:1 DNA (except in the case of TFAM/site Y 2:1 ratio complex where a really small fraction of free DNA was observed too).



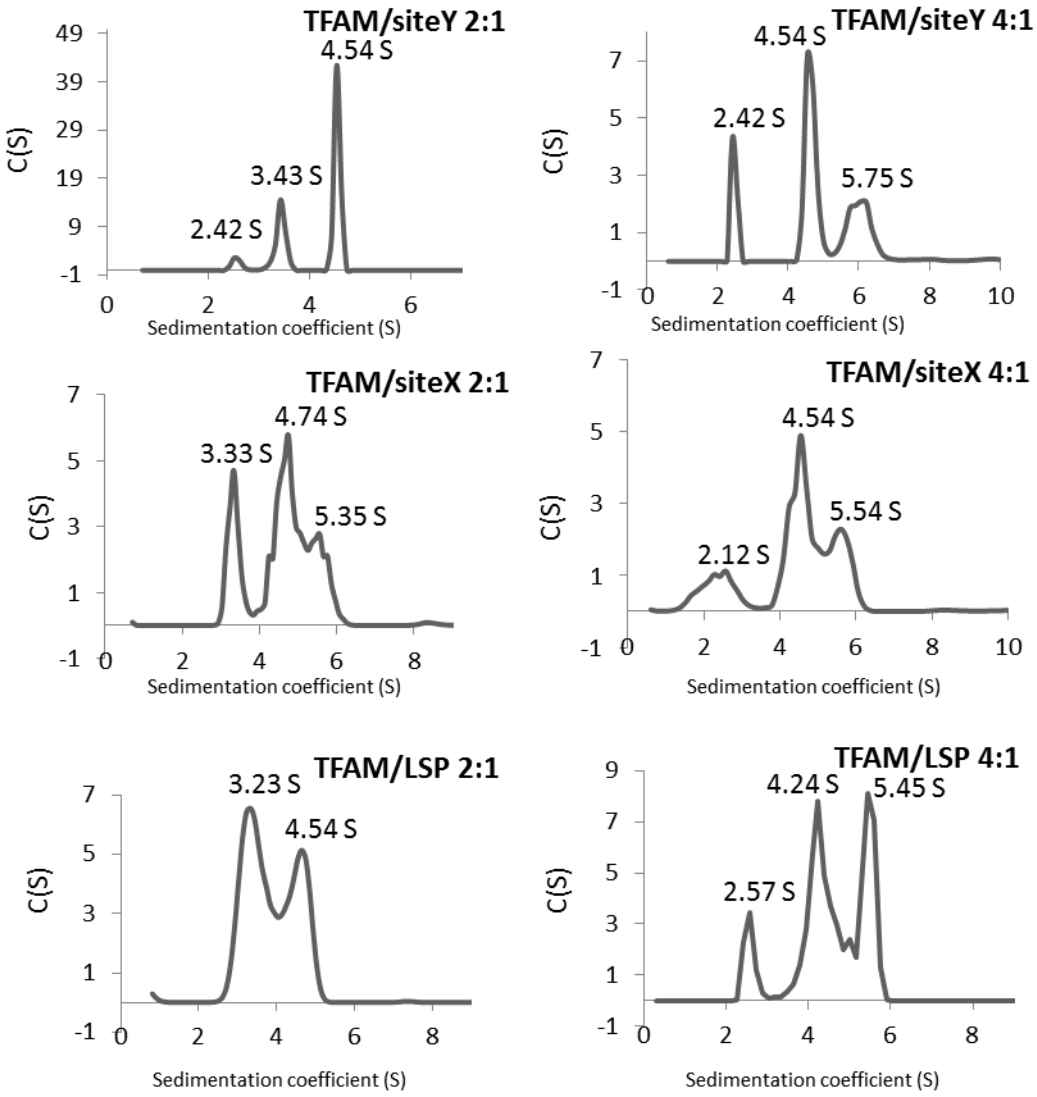


Figure_51. AUC analysis of TFAM, site X, Site Y and LSP alone and in complex
A series of graphs is reported in which is indicated the name and the ratio of the complex and the name of the DNA or the protein analysed separately. In each graph in the y axis is reported the continuous size distribution coefficient, while in the x axis is reported the sedimentation coefficient.

TFAM in complex with Y22 yielded the most sharped peaks for both 2:1 and 4:1 ratios, with respect to the other two complexes that resulted in diffuse peaks. These results may reflect the stability of the three complexes. The TFAM/Y22 complex seems the most stable in time.

6. DISCUSSION

6. DISCUSSION

The overall structure of TFAM in complex with site Y reveals a protein/DNA arrangement similar to the one observed in the previously solved crystal structures, which consists in a DNA U-turn induced by both TFAM HMG-boxes that create two 90-degree bends (Ngo, Kaiser et al. 2011; Rubio-Cosials, Sidow et al. 2011; Ngo, Lovely et al. 2014). In our crystals, two U-turns form a pseudocontinous DNA ring due to perfect stacking between DNA ends. Initially, the site Y sequence was traced as suggested by previous alignments (Fisher, Topper et al. 1987), thus the DNA ends were positioned as in TFAM/LSP crystal structure. However, during TFAM/site Y crystallographic refinement the calculated maps indicated important errors, and a thorough inspection of the electron density suggested reorientation and displacement of the sequence and, hence, of DNA ends. However, in the final Y22 structure the electron density is weaker but still continuous between DNA ends of independent DNA molecules in the a.u. Thus, the assigned Y sequence is not fully conclusive and suggests a certain disorder of Y22, i.e. additional orientations of the DNA might be present in the crystal. This idea is supported by the fact that crystals containing brominated DNA did not render any peak in the calculated anomalous differences maps, probably reflecting the DNA disorder. Thus, this is the first case where a DNA sequence seems not to be recognized in a unique manner by TFAM.

Computational studies performed with naked site X, site Y and LSP gave hints into the mechanism that underlies complex formation with TFAM. The analysis of the three naked DNAs showed a common pattern at the site inserted by Leu182: this base pair step is a CA in the three DNAs, and is characterised by high flexibility and higher roll angle, thus it is a *spontaneously* flexible open step. In the X-ray crystal structures, this insertion site has very high roll angles (between 40 and 50 degrees). This suggests that HMG-box2 introduces Leu182 into a step with an intrinsic distortion suitable for insertion, which simply increases its bending upon protein binding. In contrast, the HMG-box1 insertion site in both LSP and site X (but not in site Y) is a mildly flexible and closed AC step, which suggests that additional factors may facilitate its opening, such as the other interactions. Interestingly, in the crystal structure LSP shows the

most distorted DNA at these two inserted positions (by about 10-15 degrees higher roll at the insertions sites than Site X and Y). Site Y shows a very interesting behaviour. The positioning of the Y sequence predicted by Fisher in 1987 would locate two stiff steps 4-5 bps far from the insertion sites in the structure, and two low roll angle steps at the insertion sites. Instead, the Y sequence orientation suggested by the electron density positions two flexible steps with high roll angle (thus, opened) at both inserted sites. This suggests that TFAM inserts aminoacids into suitable open flexible steps at sequence Y, which simply increase their roll angles upon protein binding. These favourable conditions may explain the site Y orientation observed in the crystal.

The DNA sequence confers additional remarkable differences to LSP, site X and site Y. The six consecutive guanines at the B-DNA region between the two HMG boxes at LSP result in alternated flexible and mildly stiff steps that narrow the major groove and widen a shallow minor groove. Instead, site X has a asymmetric A_6T_3 A-tract not contacted but just flanked by the HGM-boxes, with highest rigidity at the central AT junction. A-tracts are rigid and induce a curvature to the DNA due to a narrowing of the major groove. Finally, site Y presents two separated stiff points. Interestingly, the higher number of stiff regions in the DNA (LSP none, site X presents one, site Y two) correlates with a marginal decrease of the deformation energy needed to bend each of the sequences into the corresponding U-turn: slightly less energy is required to bend site Y (in the suggested orientation) than site X, being LSP the most difficult to bend. The site Y orientation predicted by Fisher et al. 1987 is the most expensive to bend into a U-turn, further supporting the Y22 orientation hereby proposed.

In general terms, the curvature of the DNA, the roll angle and flexibility of the steps, which depend on the DNA sequence, shape the DNA structure for protein recognition (Rohs, Jin et al. 2010). Thus, it is possible that a mechanism of shape (indirect) readout drives TFAM binding, which is expected to vary depending on the sequence conformation and deformability.

The particularities of site Y were further backed up by the ITC studies, which showed that TFAM/site Y complex has thermodynamic properties different from TFAM/LSP and TFAM/site X, these two being similar to each other. As stated above, in all cases

the binding process is spontaneous, endothermic, and entropy driven as typically found for DNA minor-groove binding proteins such as HMG-box proteins. However, formation of TFAM/site Y complex results in a higher (unfavourable) enthalpy compensated by higher entropy than the other two complexes, indicating that the DNA sequence definitely influences the binding process. The special features of site Y was indeed evidenced by DNA competition assays. While the affinity of TFAM for site X, site Y and LSP is at the nano-molar range (slightly higher for site Y), the competition assays demonstrated a clear preference of TFAM for site Y over the other two sequences. This might be related to different kinetics during TFAM/site Y complex formation. All together, this data suggests that the DNA sequence modulate the binding mechanism of TFAM: similar affinities position TFAM on the DNA but the time of residence together with the ability in distorting the contacted sequence varies among binding sites. This draws a non-systematic, uneven scenario of TFAM binding at the mtDNA control region, which should have implications in DNA regulation.

According to what aforementioned, the DNA features influence the TFAM binding mode. From the protein side, the analysis of all available TFAM/DNA structures shows variability on the relative orientation between HMG-boxes, which is achieved by means of the linker plasticity and also by a reorientation of HMG-box1 to respct the linker. The linker torsion right before the region contacting the DNA toghether with HMG-box1 desplacement may reflect adaptation of the protein structure to the different DNA. Therefore, TFAM is a DNA packaging protein that imposes U-turns to the contacted DNA molecule and by its flexibility accommodates the specificities of the different mtDNA sequences.

Regarding multimerization of TFAM on the DNA, all available crystal structures of TFAM/DNA complexes present a protein:DNA ratio of 1:1, but ITC, gel shift and size-exclusion chromatography analyses suggest multimerization at high protein concentrations. Both MALLS and analytical ultra-centrifugation confirmed this tendency by using short DNAs which, indeed, showed a protein:DNA ratio up to 4:1. Since this increase in the protein:DNA ratio was dependent on protein concentration, the results suggest cooperative binding of additional TFAM proteins to the initial complex. On the other hand, we detected a monomer/dimer equilibrium of TFAM in

solution. Gangelhoff et al in 2009 reported analytic ultracentrifugation studies showing that TFAM in solution is a monomer (Gangelhoff, Mungalachetty et al. 2009). In contrast, Wong et al in 2009 demonstrated, by analytical ultracentrifugation, that unbound TFAM is in equilibrium between a monomeric and dimeric states and that binds to DNA as a dimer (Wong, Rajagopalan et al. 2009). Moreover, Rubio-Cosials et al in 2011 demonstrated by SAXS that TFAM in solution is a monomer at high protein concentrations whereas in 2014 Ngo et al showed by FRET that TFAM dimerization occurs in the presence of DNA (Rubio-Cosials, Sidow et al. 2011; Ngo, Lovely et al. 2014). We hereby show that at high concentrations TFAM forms multimers on the DNA. Our MALLS studies confirmed the multimeric state of TFAM alone and in complex with DNA up to a ratio 2protein:1DNA. Higher ratios were obtained by analytical ultracentrifugation, where the sample is not diluted but confined into a constant volume. Multimerization of TFAM is consistent with its functional role in mitochondria. This protein is involved in mtDNA biogenesis regulation but also in DNA packaging. One or another function seems to be activated by the protein levels in the organelle: at low concentration it binds only to mtDNA promoters thereby activating transcription and transcription-dependent replication. At high concentration it induces a strong compaction of the mitochondrial genome, inhibiting transcription and replication (Shutt, Bestwick et al. 2011; Farge, Laurens et al. 2012; Farge, Mehmedovic et al. 2014). Our multimerization results suggest that during mtDNA compaction protein-protein interactions occur and mediate the genome architecture. Thus, in such circumstances, not only the DNA sequence is predicted to modulate DNA binding and bending by TFAM but also the protein-protein flexible contacts are expected to intervene and might be regulated during the dynamics of this essential process.

7. **CONCLUSIONS**

7. CONCLUSIONS

- TFAM crystallizes in complex with a 22 bp DNA comprising site Y or site X sequences, but could not be crystallized in presence of a long DNA including both sequences together.
- Crystal structure of TFAM in complex with site Y was solved by molecular replacement, and shows an 180º DNA bending induced by contacts from the three helices of the TFAM HMG-boxes (which additionally insert two leucines, Leu58 and Leu182, at two 10 bp separated steps) and which is stabilized by the protein linker.
- In addition, the crystal structure of TFAM in complex with site Y suggests that the bound DNA is in a non-predicted, alternative orientation and presents disorder, which is a feature specific to this complex.
- Molecular dynamics (MD) simulations with naked site X and LSP sequences show that the insertion sites of Leu58 and Leu182 are semi-rigid or flexible, respectively. In contrast, the corresponding insertion sites in site Y (in the alternative orientation) are flexible.
- MD simulations show that naked site Y, site X and LSP present a positive roll value at the base pair steps involved in the Leu58 and Leu182 insertions, suggesting a shape readout mechanism for target DNA recognition.
- In addition, the MD simulations show that site Y presents a lower deformation energy than site X and LSP.
- Binding of TFAM to site Y, site X or LSP is a spontaneous process ($\Delta G<0$) in which the increase of enthalpy is compensated by the entropy.
- TFAM site Y complex formation presents different thermodynamic properties compared to TFAM/site X and TFAM/LSP complexes.
- Differential binding analysis shows that TFAM/site Y complex formation suggest different kinetics compared to TFAM/site X and TFAM/LSP complexes.
- TFAM is in concentration-dependent monomer/dimer equilibrium in solution and forms multimers on the DNA, possibly by potential protein-protein interactions involved in mtDNA packaging.

8. BIBLIOGRAPHY

8. BIBLIOGRAPHY

Adams, P. D., P. V. Afonine, et al. (2011). "The Phenix software for automated determination of macromolecular structures." Methods **55**(1): 94-106.

Afonine, P. V., R. W. Grosse-Kunstleve, et al. (2010). "phenix.model_vs_data: a high-level tool for the calculation of crystallographic model and data statistics." J Appl Crystallogr **43**(Pt 4): 669-676.

Agaronyan, K., Y. I. Morozov, et al. (2015). "Mitochondrial biology. Replication-transcription switch in human mitochondria." Science **347**(6221): 548-51.

Alam, T. I., T. Kanki, et al. (2003). "Human mitochondrial DNA is packaged with TFAM." Nucleic Acids Res **31**(6): 1640-5.

Alberts B., Johnson A., et al. (2002). Molecular Biology of the Cell. New York, Garland Science.

Amalric, F., C. Merkel, et al. (1978). "Fractionation of mitochondrial RNA from HeLa cells by high-resolution electrophoresis under strongly denaturing conditions." J Mol Biol **118**(1): 1-25.

Anderson, S. (1981). "Sequence and organization of the human mitochondrial genome." Nature **290**.

Annex, B. H. and R. S. Williams (1990). "Mitochondrial DNA structure and expression in specialized subtypes of mammalian striated muscle." Mol Cell Biol **10**(11): 5671-8.

Asin-Cayuela, J., T. Schwend, et al. (2005). "The human mitochondrial transcription termination factor (mTERF) is fully active in vitro in the non-phosphorylated form." J Biol Chem **280**(27): 25499-505.

Attardi, G. and D. Ojala (1971). "Mitochondrial ribssome in HeLa cells." Nat New Biol **229**(5): 133-6.

Bereiter-Hahn, J. and M. Voth (1996). "Distribution and dynamics of mitochondrial nucleoids in animal cells in culture." Experimental Biology Online **1**(4): 1-17.

Bonitz, S. G., R. Berlani, et al. (1980). "Codon recognition rules in yeast mitochondria." Proc Natl Acad Sci U S A **77**(6): 3167-70.

Branden, C. and A. Jones (1990). "BUSTER." Nature **343**: 687-689.

Brandon, M. C., M. T. Lott, et al. (2005). "MITOMAP: a human mitochondrial genome database--2004 update." Nucleic Acids Res **33**(Database issue): D611-3.

Brega, A. and C. Baglioni (1971). "A study of mitochondrial protein synthesis in intact HeLa cells." Eur J Biochem **22**(3): 415-22.

Brown, T. A., A. N. Tkachuk, et al. (2015). "Mitochondrial Transcription Factor A (TFAM) Binds to RNA Containing 4-Way Junctions and Mitochondrial tRNA." PLoS One **10**(11): e0142436.

Brown, T. A., A. N. Tkachuk, et al. (2011). "Superresolution fluorescence imaging of mitochondrial nucleoids reveals their spatial range, limits, and membrane interaction." Mol Cell Biol **31**(24): 4994-5010.

Brown, W. M., J. Shine, et al. (1978). "Human mitochondrial DNA: analysis of 7S DNA from the origin of replication." Proc Natl Acad Sci U S A **75**(2): 735-9.

Burger, G., M. W. Gray, et al. (2003). "Mitochondrial genomes: anything goes." Trends in Genetics **19**(12): 709-716.

Camasamudram, V., J. K. Fang, et al. (2003). "Transcription termination at the mouse mitochondrial H-strand promoter distal site requires an A/T rich sequence motif and sequence specific DNA binding proteins." Eur J Biochem **270**(6): 1128-40.

Carrodeguas, J. A., K. G. Pinz, et al. (2002). "DNA binding properties of human pol gammaB." J Biol Chem **277**(51): 50008-14.

Chang, D. D. and D. A. Clayton (1984). "Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA." Cell **36**(3): 635-43.

Chomyn, A., M. W. J. Cleeter, et al. (1986). "URF6, Last Unidentified Reading Frame of Human mtDNA, Codes for an NADH Dehydrogenase Subunit." Science **234**.

Chung, H. K. and L. L. Spremulli (1990). "Purification and characterization of elongation factor G from bovine liver mitochondria." J Biol Chem **265**(34): 21000-4.

Churchill, M. E., J. Klass, et al. (2010). "Structural analysis of HMGD-DNA complexes reveals influence of intercalation on sequence selectivity and DNA bending." J Mol Biol **403**(1): 88-102.

Clayton, D. A. (2003). "Mitochondrial DNA replication: what we know." IUBMB Life **55**(4-5): 213-7.
Dairaghi, D. J., G. S. Shadel, et al. (1995). "Addition of a 29 residue carboxyl-terminal tail converts a simple HMG box-containing protein into a transcriptional activator." J Mol Biol **249**(1): 11-28.
Denslow, N. D., G. S. Michaels, et al. (1989). "Mechanism of mRNA binding to bovine mitochondrial ribosomes." J Biol Chem **264**(14): 8328-38.
Dragan, A. I., C. M. Read, et al. (2004). "DNA binding and bending by HMG boxes: energetic determinants of specificity." J Mol Biol **343**(2): 371-93.
Drsata, T., A. Perez, et al. (2013). "Structure, Stiffness and Substates of the Dickerson-Drew Dodecamer." J Chem Theory Comput **9**(1): 707-721.
Dubin, D. T., J. Montoya, et al. (1982). "Sequence analysis and precise mapping of the 3' ends of HeLa cell mitochondrial ribosomal RNAs." J Mol Biol **157**(1): 1-19.
Emsley, P. (2004). "Coot: model-building tools for molecular graphics." **60**(Pt 12 Pt 1): 2126-32.
Ernster, L. and G. Schatz (1981). "Mitochondria: A Historical Review." Cell Biology.
Evans, P. (2006). "Scaling and assessment of data quality." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 72-82.
Evans, P. and A. McCoy (2008). "An introduction to molecular replacement." Acta Crystallogr D Biol Crystallogr **64**(Pt 1): 1-10.
Falkenberg, M., M. Gaspari, et al. (2002). "Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA." Nat Genet **31**(3): 289-94.
Falkenberg, M., N. G. Larsson, et al. (2007). "DNA replication and transcription in mammalian mitochondria." Annu Rev Biochem **76**: 679-99.
Farge, G., N. Laurens, et al. (2012). "Protein sliding and DNA denaturation are essential for DNA organization by human mitochondrial transcription factor A." Nat Commun **3**: 1013.
Farge, G., M. Mehmedovic, et al. (2014). "In vitro-reconstituted nucleoids can block mitochondrial DNA replication and transcription." Cell Rep **8**(1): 66-74.
Fisher, R. P. and D. A. Clayton (1985). "A transcription factor required for promoter recognition by human mitochondrial RNA polymerase. Accurate initiation at the heavy- and light-strand promoters dissected and reconstituted in vitro." J Biol Chem **260**(20): 11330-8.
Fisher, R. P. and D. A. Clayton (1988). "Purification and characterization of human mitochondrial transcription factor 1." Mol Cell Biol **8**(8): 3496-509.
Fisher, R. P., T. Lisowsky, et al. (1992). "DNA wrapping and bending by a mitochondrial high mobility group-like transcriptional activator protein." J Biol Chem **267**(5): 3358-67.
Fisher, R. P., J. N. Topper, et al. (1987). "Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements." Cell **50**(2): 247-58.
Friedman, J. R. and J. Nunnari (2014). "Mitochondrial form and function." Nature **505**(7483): 335-43.
Fuste, J. M., S. Wanrooij, et al. (2010). "Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication." Mol Cell **37**(1): 67-78.
Fuxreiter, M., I. Simon, et al. (2011). "Dynamic protein-DNA recognition: beyond what can be seen." Trends Biochem Sci **36**(8): 415-23.
Gangelhoff, T. A., P. S. Mungalachetty, et al. (2009). "Structural analysis and DNA binding of the HMG domains of the human mitochondrial transcription factor A." Nucleic Acids Res **37**(10): 3153-64.
Garrido, N., L. Griparic, et al. (2003). "Composition and dynamics of human mitochondrial nucleoids." Mol Biol Cell **14**(4): 1583-96.
Ghivizzani, S. C., C. S. Madsen, et al. (1994). "In organello footprint analysis of human mitochondrial DNA: human mitochondrial transcription factor A interactions at the origin of replication." Mol Cell Biol **14**(12): 7717-30.
Goffart, S., P. Martinsson, et al. (2007). "The mitochondria of cultured mammalian cells: II. Expression and visualization of exogenous proteins in fixed and live cells." Methods Mol Biol **372**: 17-32.
Goodwin, G. H., C. Sanders, et al. (1973). "A new group of chromatin-associated proteins with a high content of acidic and basic amino acids." Eur J Biochem **38**(1): 14-9.

Gray, H. and T. W. Wong (1992). "Purification and identification of subunit structure of the human mitochondrial DNA polymerase." J Biol Chem **267**(9): 5835-41.

Gray, M. W. (2012). "Mitochondrial evolution." Cold Spring Harb Perspect Biol **4**(9): a011403.

Gray, M. W. (2014). "The pre-endosymbiont hypothesis: a new perspective on the origin and evolution of mitochondria." Cold Spring Harb Perspect Biol **6**(3).

Gray, M. W. (2015). "Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria." Proc Natl Acad Sci U S A **112**(33): 10133-8.

Gray, M. W., G. Burger, et al. (2001). "The origin and early evolution of mitochondria." Genome Biology.

Green, C. John, et al. (1998). "Mitochondria and Apoptosis." Science **281**(5381): 1309-1312.

Grosschedl, R., K. Giese, et al. (1994). "HMG domain proteins: architectural elements in the assembly of nucleoprotein structures." Trends Genet **10**(3): 94-100.

Gustafsson, C. M., M. Falkenberg, et al. (2016). "Maintenance and Expression of Mammalian Mitochondrial DNA." Annu Rev Biochem **85**: 133-60.

Hallberg, B. M. and N. G. Larsson (2014). "Making proteins in the powerhouse." Cell Metab **20**(2): 226-40.

He, J., C. C. Mao, et al. (2007). "The AAA+ protein ATAD3 has displacement loop binding properties and is involved in mitochondrial nucleoid organization." J Cell Biol **176**(2): 141-6.

Hendrickson, W. A. (2014). Anomalous diffraction in crystallographic phase evaluation. Quarterly Reviews of Biophysics. New York, USA, Cambridge University Press. **47:** 49-93.

Hixson, J. E. and D. A. Clayton (1985). "Initiation of transcription from each of the two human mitochondrial promoters requires unique nucleotides at the transcriptional start sites." Proc Natl Acad Sci U S A **82**(9): 2660-4.

Holmes, J. B., G. Akman, et al. (2015). "Primer retention owing to the absence of RNase H1 is catastrophic for mitochondrial DNA replication." Proc Natl Acad Sci U S A **112**(30): 9334-9.

Holt, I. J. and H. T. Jacobs (2003). "Response: The mitochondrial DNA replication bubble has not burst." Trends Biochem Sci **28**(7): 355-6.

Iborra, F. J., H. Kimura, et al. (2004). "The functional organization of mitochondrial genomes in human cells." BMC Biol **2**: 9.

Ivani, I., P. D. Dans, et al. (2016). "Parmbsc1: a refined force field for DNA simulations." Nat Methods **13**(1): 55-8.

Jen-Jacobson, L., L. E. Engler, et al. (2000). "Structural and thermodynamic strategies for site-specific DNA binding proteins." Structure **8**(10): 1015-23.

Jimenez-Menendez, N., P. Fernandez-Millan, et al. (2010). "Human mitochondrial mTERF wraps around DNA through a left-handed superhelical tandem repeat." Nat Struct Mol Biol **17**(7): 891-3.

Jorgensen, W. L., J. Chandrasekhar, et al. (1983). "Comparison of simple potential functions for simulating liquid water " J Chem Phys **79**(2): 204-512.

Kabsch, W. (2010). "Xds." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 125-32.

Kasamatsu, H., L. I. Grossman, et al. (1974). "The replication and structure of mitochondrial DNA in animal cells." Cold Spring Harb Symp Quant Biol **38**: 281-8.

Kasamatsu, H., D. L. Robberson, et al. (1971). "A novel closed-circular mitochondrial DNA with properties of a replicating Intermediate." Proc. Nat. Acad. Sci. USA **68**: 2252-2257.

Kasashima, K. and H. Endo (2015). "Interaction of human mitochondrial transcription factor A in mitochondria: its involvement in the dynamics of mitochondrial DNA nucleoids." Genes Cells **20**(12): 1017-27.

Kaufman, B. A., N. Durisic, et al. (2007). "The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures." Mol Biol Cell **18**(9): 3225-36.

King, T. C. and R. L. Low (1987). "Mitochondrial DNA displacement loop structure depends on growth state in bovine cells." J Biol Chem. **262**(13): 6214-20.

Kleywegt, G. J. and A. T. Brunger (1996). "Checking your imagination: applications of the free R value." Structure **4**(8): 897-904.

Klinge, C. M. (2008). "Estrogenic control of mitochondrial function and biogenesis." J Cell Biochem **105**(6): 1342-51.

Koonin, E. V. (2010). "The origin and early evolution of eukaryotes in the light of phylogenomics." Genome Biology.

Korhonen, J. A., M. Gaspari, et al. (2003). "TWINKLE Has 5' -> 3' DNA helicase activity and is specifically stimulated by mitochondrial single-stranded DNA-binding protein." J Biol Chem **278**(49): 48627-32.

Korhonen, J. A., X. H. Pham, et al. (2004). "Reconstitution of a minimal mtDNA replisome in vitro." Embo J **23**(12): 2423-9.

Kruse, B., N. Narasimhan, et al. (1989). "Termination of transcription in human mitochondria: identification and purification of a DNA binding protein factor that promotes termination." Cell **58**(2): 391-7.

Kukat, C., D. K.M., et al. (2015). "Cross-strand binding of TFAM to a single mtDNA molecule forms the mitochondrial nucleoid." Proc Natl Acad Sci U S A **112**(36): 11288-93.

Kukat, C., C. A. Wurm, et al. (2011). "Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA." Proc Natl Acad Sci U S A **108**(33): 13534-9.

Lakshmipathy, U. and C. Campbell (1999). "The human DNA ligase III gene encodes nuclear and mitochondrial proteins." Mol Cell Biol **19**(5): 3869-76.

Landsman, D. and M. Bustin (1993). "A signature for the HMG-1 box DNA-binding proteins." Bioessays **15**(8): 539-46.

Lane, N. and W. Martin (2010). "The energetics of genome complexity." Nature **467**(7318): 929-34.

Laue, T. M., B. D. Shah, et al. (1992). "Computer aided interpretation of analytical sedimentation data for proteins. In Analytical ultracentrifugation in biochemistry and polymer science." The Royal Society of Chemistry: 90-125.

Lavery, R., K. Zakrzewska, et al. (2010). "A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA." Nucleic Acids Res **38**(1): 299-313.

Legros, F., F. Malka, et al. (2004). "Organization and dynamics of human mitochondrial DNA." J Cell Sci **117**(Pt 13): 2653-62.

Liao, H. X. and L. L. Spremulli (1991). "Initiation of protein synthesis in animal mitochondria. Purification and characterization of translational initiation factor 2." J Biol Chem **266**(31): 20714-9.

Lodeiro, M. F., A. Uchida, et al. (2012). "Transcription from the second heavy-strand promoter of human mtDNA is repressed by transcription factor A in vitro." Proc Natl Acad Sci U S A **109**(17): 6513-8.

Ma, J. and L. L. Spremulli (1996). "Expression, purification, and mechanistic studies of bovine mitochondrial translational initiation factor 2." J Biol Chem **271**(10): 5805-11.

Macreadie, I. G., C. E. Novitski, et al. (1983). "Biogenesis of mitochondria: the mitochondrial gene (aap1) coding for mitochondrial ATPase subunit 8 in Saccharomyces cerevisiae." Nucleic Acids Res **11**(13): 4435-51.

Malarkey, C. S. and M. E. Churchill (2012). "The high mobility group box: the ultimate utility player of a cell." Trends Biochem Sci **37**(12): 553-62.

Mangus, D. A., S. H. Jang, et al. (1994). "Release of the yeast mitochondrial RNA polymerase specificity factor from transcription complexes." J Biol Chem **269**(42): 26568-74.

Martin, M., J. Cho, et al. (2005). "Termination factor-mediated DNA loop between termination and initiation sites drives mitochondrial rRNA synthesis." Cell **123**(7): 1227-40.

McCoy, A. J., R. W. Grosse-Kunstleve, et al. (2007). "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.

McCulloch, V. and G. S. Shadel (2003). "Human Mitochondrial Transcription Factor B1 Interacts with the C-Terminal Activation Region of h-mtTFA and Stimulates Transcription Independently of Its RNA Methyltransferase Activity." Molecular and Cellular Biology **23**(16): 5816-5824.

Montoya, J., T. Christianson, et al. (1982). "Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA." Proc Natl Acad Sci U S A **79**(23): 7195-9.

Morozov, Y. I., A. V. Parshin, et al. (2015). "A model for transcription initiation in human mitochondria." Nucleic Acids Res **43**(7): 3726-35.

Morozov, Y. I. and D. Temiakov (2016). "Human Mitochondrial Transcription Initiation Complexes Have Similar Topology on the Light and Heavy Strand Promoters." J Biol Chem **291**(26): 13432-5.

Nass, M. M. (1966). "The circularity of mitochondrial DNA." Proc Natl Acad Sci U S A **56**(4): 1215-22.

Ngo, H. B., J. T. Kaiser, et al. (2011). "The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA." Nat Struct Mol Biol **18**(11): 1290-6.

Ngo, H. B., G. A. Lovely, et al. (2014). "Distinct structural features of TFAM drive mitochondrial DNA packaging versus transcriptional activation." Nat Commun **5**: 3077.

Oh-hama, T. (1997). "Evolutionary consideration on 5-Aminolevulinate synthase in nature." Origins of Life and Evolution of the Biosphere **27**: 405-412.

Ojala, D., J. Montoya, et al. (1981). "tRNA punctuation model of RNA processing in human mitochondria." Nature **290**(5806): 470-4.

Park, J. S., F. Gamboni-Robertson, et al. (2006). "High mobility group box 1 protein interacts with multiple Toll-like receptors." Am J Physiol Cell Physiol **290**(3): C917-24.

Portella, G., F. Battistini, et al. (2013). "Understanding the connection between epigenetic DNA methylation and nucleosome positioning from computer simulations." PLoS Comput Biol **9**(11): e1003354.

Posse, V., S. Shahzad, et al. (2015). "TEFM is a potent stimulator of mitochondrial transcription elongation in vitro." Nucleic Acids Res **43**(5): 2615-24.

Privalov, P. L., A. I. Dragan, et al. (2009). "The cost of DNA bending." Trends Biochem Sci **34**(9): 464-70.

Privalov, P. L., A. I. Dragan, et al. (2011). "Interpreting protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components." Nucleic Acids Res **39**(7): 2483-91.

Privalov, P. L., A. I. Dragan, et al. (2007). "What drives proteins into the major or minor grooves of DNA?" J Mol Biol **365**(1): 1-9.

Read, R. (1990). "Structure-factor probabilities for related structures." Acta Crystallographica Section A: Foundations of Crystallography **46**(11): 900-912.

Reyes, A., L. Kazak, et al. (2013). "Mitochondrial DNA replication proceeds via a 'bootlace' mechanism involving the incorporation of processed transcripts." Nucleic Acids Res **41**(11): 5837-50.

Reyes, A., M. Mezzina, et al. (2002). "Human mitochondrial transcription factor A (mtTFA): gene structure and characterization of related pseudogenes." Gene **291**(1-2): 223-32.

Ringel, R., M. Sologub, et al. (2011). "Structure of human mitochondrial RNA polymerase." Nature **478**(7368): 269-73.

Rohs, R., X. Jin, et al. (2010). "Origins of specificity in protein-DNA recognition." Annu Rev Biochem **79**: 233-69.

Rossier, M. F. (2006). "T channels and steroid biosynthesis: in search of a link with mitochondria." Cell Calcium **40**(2): 155-64.

Rossmanith, W., A. Tullo, et al. (1995). "Human mitochondrial tRNA processing." J Biol Chem **270**(21): 12885-91.

Rossmann, M. G. (1990). "The molecular replacement method." Acta Crystallographica Section A: Foundations of Crystallography **46**(2): 73-82.

Rubio-Cosials, A., J. F. Sidow, et al. (2011). "Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter." Nat Struct Mol Biol **18**(11): 1281-9.

Rubio-Cosials, A. and M. Sola (2013). "U-turn DNA bending by human mitochondrial transcription factor A." Curr Opin Struct Biol **23**(1): 116-24.

Satoh, M. and T. Kuroiwa (1991). "Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell." Exp Cell Res **196**(1): 137-40.

Sbisa, E., F. Tanzariello, et al. (1997). "Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications." Gene **205**(1-2): 125-40.

Schuck, P. (2000). "Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling." Biophys J **78**(3): 1606-19.

Schuck, P., M. A. Perugini, et al. (2002). "Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems." Biophys J **82**(2): 1096-111.

Schwartzbach, C. J. and L. L. Spremulli (1989). "Bovine mitochondrial protein synthesis elongation factors. Identification and initial characterization of an elongation factor Tu-elongation factor Ts complex." J Biol Chem **264**(32): 19125-31.

Schwinghammer, K., A. C. Cheung, et al. (2013). "Structure of human mitochondrial RNA polymerase elongation complex." Nat Struct Mol Biol **20**(11): 1298-303.

Shadel, G. S. and D. A. Clayton (1997). "Mitochondrial DNA maintenance in vertebrates." Annu Rev Biochem **66**: 409-35.

Sheldrick, G. M. (2008). "A short history of SHELX." Acta Crystallogr A **64**(Pt 1): 112-22.

Shutt, T. E., M. Bestwick, et al. (2011). "The core human mitochondrial transcription initiation complex: It only takes two to tango." Transcription **2**(2): 55-59.

Smith, D. and L. Dang (1994). "Computer simulations of NaCl association in polarizable water." J Chem Phys **100**(20): 3757-3766.

Sologub, M., D. Litonin, et al. (2009). "TFB2 is a transient component of the catalytic site of the human mitochondrial RNA polymerase." Cell **139**(5): 934-44.

Spelbrink, J. N., F. Y. Li, et al. (2001). "Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria." Nat Genet **28**(3): 223-31.

Suissa, S., Z. Wang, et al. (2009). "Ancient mtDNA genetic variants modulate mtDNA transcription and replication." PLoS Genet **5**(5): e1000474.

Taanman, J. W. (1999). "The mitochondrial genome: structure, transcription, translation and replication." Biochimica et Biophysica Acta: 103-123.

Terzioglu, M., B. Ruzzenente, et al. (2013). "MTERF1 binds mtDNA to prevent transcriptional interference at the light-strand promoter but is dispensable for rRNA gene transcription regulation." Cell Metab **17**(4): 618-26.

Tominaga, K., J. Hayashi, et al. (1993). "Smaller isoform of human mitochondrial transcription factor 1: its wide distribution and production by alternative splicing." Biochem Biophys Res Commun **194**(1): 544-51.

Topper, J. N. and D. A. Clayton (1989). "Identification of transcriptional regulatory elements in human mitochondrial DNA by linker substitution analysis." Mol Cell Biol. **9**(3): 1200-11.

Travers, A. A. (2004). "The structural basis of DNA flexibility." Philos Trans A Math Phys Eng Sci **362**(1820): 1423-38.

Uhler, J. P. and M. Falkenberg (2015). "Primer removal during mammalian mitochondrial DNA replication." DNA Repair (Amst) **34**: 28-38.

Walberg, M. W. and D. A. Clayton (1981). "Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA." Nucleic Acids Res **9**(20): 5411-21.

Walberg, M. W. and D. A. Clayton (1983). "In vitro transcription of human mitochondrial DNA. Identification of specific light strand transcripts from the displacement loop region." J Biol Chem **258**(2): 1268-75.

Wallace, D. C. (2013). "A mitochondrial bioenergetic etiology of disease." J Clin Invest **123**(4): 1405-12.

Wang, P., L. Xu, et al. (2016). "Energy Remodeling, Mitochondrial Disorder and Heart Failure." Curr Pharm Des **22**(31): 4823-4829.

Wanrooij, P. H., J. P. Uhler, et al. (2012). "A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop." Nucleic Acids Res **40**(20): 10334-44.

Wanrooij, P. H., J. P. Uhler, et al. (2010). "G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation." Proc Natl Acad Sci U S A **107**(37): 16072-7.

Weir, H. M., P. J. Kraulis, et al. (1993). "Structure of the HMG box motif in the B-domain of HMG1." Embo J **12**(4): 1311-9.

Winn, M. D., C. C. Ballard, et al. (2011). "Overview of the CCP4 suite and current developments." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 235-42.

Wirth, A. J. and M. Gruebele (2013). "Quinary protein structure and the consequences of crowding in living cells: leaving the test-tube behind." Bioessays **35**(11): 984-93.

Wolstenholme, D. R. (1992). "Animal mitochondrial DNA: structure and evolution." Int Rev Cytol **141**: 173-216.

Wong, T. S., S. Rajagopalan, et al. (2009). "Biophysical characterizations of human mitochondrial transcription factor A and its binding to tumor suppressor p53." Nucleic Acids Res **37**(20): 6765-83.

Yakubovskaya, E., Z. Chen, et al. (2006). "Functional human mitochondrial DNA polymerase gamma forms a heterotrimer." J Biol Chem **281**(1): 374-82.

Yakubovskaya, E., E. Mejia, et al. (2010). "Helix unwinding and base flipping enable human MTERF1 to terminate mitochondrial transcription." Cell **141**(6): 982-93.

Yoza, B. K. and D. F. Bogenhagen (1984). "Identification and in vitro capping of a primary transcript of human mitochondrial DNA." J Biol Chem **259**(6): 3909-15.

ACKNOWLEDGMENTS

I would like to acknowledge all the people that are involved in this thesis work. First of all, my thesis director Dr. Maria Solà for giving me the opportunity to work in her group and for supervising my projects during these years; Dr. Federica Battistini from Modesto Orozco laboratory, at the Institute of Research in Biomedicine (IRB), Barcelona, for the fundamental molecular dynamics studies provided; Dr. Rafel Prohens, from the "Centre Científics i Tecnològics" of Barcelona University, for his help in ITC data collection and interpretation; Drs. German Rivas Caballero and Carlos Alfonso Botello, from the "Centro de Investigaciones Biológicas", CIB-CSIC, Madrid, for receiving me in their laboratory and for their help in data collection and interpretation of MALLS and AUC analysis.

Special acknowledgments go to the people of my group: to Pablo and Anna, my first lab mates and my first mentors in science a part that good friends, I learned a lot from them; to Sebastian, his scientific support was essential for my thesis development; to Reicy, Arko , Claus, Aleix and Javi, for being precious friends and colleges.

I'd like also to acknowledge people from crystallization and purification platform: Sonia, Joan, Sandra, Isabel, Laia and Roman, for their efficiency and their helpfulness.

And for sure I want to acknowledge all the people from the structural laboratories: Cristina F. and Diego for having always answers to all my scientific doubts; Pablo Victor, Cristina S., Laura, for the reactives that they shared with me, now I will stop to ask!

I want also to thank Tibi and Rosa for teaching me all the secrets of crystal grow and fishing; Theo for continuing to follow my scientific growth after the end of my internship; Nereida, David, Luca, Barbara, Marieta, Sergio, for remembering me how many years passes from when I started my thesis!

A special acknowledgment goes also to all the people of cri4 lab: Giovanna, Massimo, Claudia, Alfonso, their support was fundamental and irreplaceable.

Finally but not because less important I want to thank all the friends not mentioned above and my family for being always a reference point in my life and for always encouraging and supporting me in the most difficult moments.

THE END