

Unexpected sequences and structures of mtDNA required for efficient transcription from the first heavy-strand promoter

Akira Uchida^{1,4}, Divakaran Murugesapillai^{2,4}, Yao Wang¹, Maria F. Lodeiro¹, Shaan Prabhakar¹, Jamie J. Arnold¹, L. James Maher III³, Mark C. Williams², and Craig E. Cameron¹

¹The Pennsylvania State University, Department of Biochemistry and Molecular Biology, University Park, PA 16802

²Northeastern University, Department of Physics, Boston, MA 02115

³Mayo Clinic College of Medicine, Department of Biochemistry and Molecular Biology, Rochester, MN 55905

⁴These authors contributed equally to this study.

Address correspondence to: Craig E. Cameron, cec9@psu.edu; Mark C. Williams, ma.williams@northeastern.edu

Human mtDNA contains three promoters, suggesting a need for differential expression of the mitochondrial genome. Studies of mitochondrial transcription have used a reductionist approach, perhaps masking differential regulation. Here we evaluate transcription from light-strand (LSP) and heavy-strand (HSP1) promoters using templates that mimic their natural context. These studies reveal sequences upstream, hypervariable in the human population (HVR3), and downstream of the HSP1 transcription start site required for maximal yield. The carboxy-terminal tail of TFAM is essential for activation of HSP1 but not LSP. Images of the template obtained by atomic force microscopy show that TFAM creates loops in a discrete region, the formation of which correlates with activation of HSP1; looping is lost in tail-deleted TFAM. Identification of HVR3 as a transcriptional regulatory element may contribute to between-individual variability in mitochondrial gene expression. The unique requirement of HSP1 for the TFAM tail may enable its regulation by post-translational modifications.

INTRODUCTION

In spite of the absolute requirement of mitochondrial function for life, more is known about gene-regulatory mechanisms in prokaryotes than is known about corresponding mechanisms in the mitochondria of humans or other mammalian species. One reason for this knowledge gap is the inability to perform reverse-genetic analysis of mitochondrial DNA (mtDNA). Studies performed in cells and cell-free extracts have revealed the existence of three mitochondrial promoters: light-strand promoter (LSP), heavy-strand promoter 1 (HSP1) and HSP2 (Bogenhagen et al., 1984; Cantatore and Attardi, 1980; Chang and Clayton, 1984; Montoya et al., 1982; Montoya et al., 1983). Transcripts from each promoter are polygenic and all contain tRNAs (Ojala et al., 1980; Ojala et al., 1981). mtDNA encodes only 13 proteins, all components of the electron transport chain (ETC) or ATP synthase. The mRNA for ND6 is transcribed from LSP; the other 12 mRNAs are transcribed from HSP2 (Clayton, 1984). The primary role of HSP1 is transcription of rRNA genes (Clayton, 1984). We recently suggested that such a division of transcription would allow assembly of the core components of the ETC by activating HSP2. In this model, activation of complex I would not occur until LSP was activated, as ND6 mRNA is the sole mRNA transcribed from this promoter. Such an order could diminish reactive oxygen production that might be produced if ND6 subcomplexes formed (Lodeiro et al., 2012). Having rRNA transcription controlled distinctly by HSP1 would permit mitochondrial biogenesis to be regulated distinctly from the myriad other homeostatic functions of mitochondria (Gaines and Attardi, 1984; Gaines et al., 1987; Lodeiro et al., 2012).

For a regulated program of gene expression to exist, each promoter should exhibit some unique attribute relative to the others with respect to transcription (re)initiation and/or elongation, steps most frequently used to control transcription. The first promoter characterized biochemically was LSP. This initial characterization showed that the region 50 bp upstream of the transcription start site (TSS) was sufficient for transcription (Bogenhagen et al., 1984; Fisher and Clayton, 1985; Fisher et al., 1987; Gaines and Attardi, 1984). In going from cell-free extracts to highly purified systems, the genetic determinants for LSP did not change (Falkenberg et al., 2002; Gaspari et al., 2004; Lodeiro et al., 2010; Sologub et al., 2009). When the genetic architecture of HSP1 was deduced based on LSP, factor-dependent transcription could be observed (Bogenhagen et al., 1984). The factors required for initiation are mitochondrial transcription factor A (TFAM), mitochondrial transcription factor B2 (TFB2M), and the mitochondrial RNA polymerase (POLRMT) (Falkenberg et al., 2002; Fisher and Clayton, 1985; Fisher et al., 1987). The model that has emerged for transcription initiation from LSP posits binding of one molecule of TFAM at a specific binding site upstream of the TSS (Gaspari et al., 2004; Lodeiro et al., 2010; Morozov et al., 2015; Sologub et al., 2009). This binding event leads to a large bend, "U-turn," of the DNA (Ngo et al., 2011; Rubio-Cosials

et al., 2011). POLRMT is then recruited to the promoter by means of an interaction of its first 150 amino acids with TFAM (Morozov et al., 2014). TFB2M likely joins with POLRMT as these proteins form a stable complex (Gaspari et al., 2004). Worth noting, binding of TFB2M to POLRMT is not essential for its recruitment by TFAM (Morozov et al., 2014).

There is a consensus for this mechanism of transcription initiation from LSP (Gaspari et al., 2004; Lodeiro et al., 2010; Morozov et al., 2015; Sologub et al., 2009). However, how initiation from HSP1 and HSP2 occurs is actively debated (Litonin et al., 2010; Lodeiro et al., 2012; Morozov and Temiakov, 2016; Shi et al., 2012; Shutt et al., 2010; Zollo et al., 2012). Some investigators observe transcription only from HSP1 in the presence of all three factors, and observe no transcription from HSP2 (Litonin et al., 2010; Morozov and Temiakov, 2016; Shi et al., 2012). Other investigators observe transcription from HSP1 and HSP2 in the presence of only POLRMT and TFB2M (Lodeiro et al., 2012; Shutt et al., 2010; Zollo et al., 2012). Interestingly, these same investigators show that TFAM stimulates transcription from HSP1 but inhibits transcription from HSP2 (Lodeiro et al., 2012; Zollo et al., 2012). Therefore, some conclude that there is no regulation of transcription mediated by the core components of transcription (Litonin et al., 2010; Morozov and Temiakov, 2016; Shi et al., 2012). Others conclude that POLRMT and TFB2M represent the core transcription machinery that is regulated by TFAM (Lodeiro et al., 2012; Shutt et al., 2010; Zollo et al., 2012). In this latter scenario, TFAM is essential for any transcription from LSP, an activator of transcription from HSP1 and an inhibitor of transcription from HSP2 (Lodeiro et al., 2012).

Essentially all of studies of mitochondrial transcription using purified proteins rather than cell-free extracts rely on minimal promoters in isolation. The objective of this study was to evaluate the mechanism of transcription of LSP and HSP1 simultaneously using a dual-promoter template (DPT) in the same context in which these promoters appear in mtDNA. Data obtained with this DPT show unambiguously that LSP and HSP1 exhibit unique properties that permit each to be regulated independent of the other. Even though TFAM activates both promoters in this context, the concentration dependence and TFAM domain-dependence for this activation differ between the two promoters. In the absence of TFAM, only a single round of transcription occurs from HSP1, suggesting that TFAM activates HSP1 by promoting reinitiation. Unexpectedly, we identified sequences of mtDNA upstream and downstream of the TSS for HSP1 that contribute substantially to both basal and activated transcription from this promoter. The sequences upstream of the TSS are a part of a region of mtDNA that is hypervariable in the human population (HVR3), the variability of which may predispose to disease (<http://www.mitomap.org/MITOMAP>). Using atomic force microscopy, we show that TFAM bound to these elements of HSP1 multimerize in a manner dependent on the carboxy-terminal tail of TFAM.

Formation of the resulting DNA loops correlated directly to maximal activation of transcription from HSP1. Similarly, template compaction observed at higher TFAM concentrations correlates with transcription inhibition. Our data point to the potential for post-translational modifications of the TFAM carboxy-terminal tail to regulate mitochondrial gene expression. We further suggest that HVR3 polymorphisms may contribute to disease by interfering with regulated expression of the mitochondrial genome.

RESULTS

The light-strand promoter (LSP) is the most extensively studied human mtDNA promoter. All of the cis-acting elements required for transcription factor-dependent initiation are located within 50 bp upstream of the transcription start site (TSS). The primary role of this upstream sequence is to bind mitochondrial transcription factor A (TFAM), a DNA binding, bending and wrapping protein related to mammalian high-mobility-group proteins (Malarkey and Churchill, 2012; Murugesapillai et al., 2016). It has been assumed by some that the organization and TFAM-dependence of the HSP1 and HSP2 promoters are identical. Our previous studies, however, suggested that TFAM regulation of HSP1 and HSP2 differs (Lodeiro et al., 2012).

Sequences more than 50 bp upstream of the TSSs of LSP and HSP1 have largely been ignored in studies of mitochondrial transcription, at least over the past decade. This inter-promoter region (IPR) is the third of three hypervariable regions (HVR3) of mtDNA (**Fig. 1a**). We reasoned that direct comparison of the two promoters would be facilitated by studying a relevant mtDNA fragment. We chose a sequence that began 35 bp downstream of LSP, producing a 35-nt transcription product, and ended 45 bp downstream of HSP1, producing a 45-nt transcription product (**Fig. 1b**). We refer to this DNA fragment as the dual-promoter template (DPT). Transcription using the DPT employed a ³²P-labeled adenylate homotrimer as primer in order to quantify transcription products after separation by denaturing polyacrylamide gel electrophoresis (**Fig. S1**). Reaction conditions used were stringent and consistent with those used to study a variety of nucleic acid polymerases (**Fig. S1**). Under these conditions, transcription products formed linearly for at least 30 minutes (**Fig. S1**). Most single time-point assays were performed between 5 and 15 minutes.

The first experiment used the DPT to evaluate transcription from LSP and HSP1 as a function of TFAM concentration. Transcription from HSP1 was readily detectable in the absence of TFAM (**Fig. 1c**); the yield ranged from 70-100 nM, which is essentially stoichiometric with template (**Fig. 1d**). The TFAM-independent yield from HSP1 did not change even after a 90 min incubation (data not shown). Titration of TFAM into the reaction resulted in an increase in transcription from LSP that reached a maximum value of ~500 nM transcript at a TFAM concentration of 300 nM (**Figs. 1c and 1d**). Over this same range, output

from HSP1 changed minimally. Transcription from HSP1 reached a maximum value of ~300 nM transcript at a TFAM concentration of 1000 nM (**Figs. 1c and 1d**). For TFAM concentrations at which transcription from LSP was inhibited almost entirely, transcription from HSP1 proceeded to levels no less than 50% of maximum levels (**Figs. 1c and 1d**). Together, these observations demonstrate different requirements for maximal transcription from LSP and HSP1. Some of these differences are due to differential requirements for and response to TFAM.

In order to determine the extent to which the IPR contributed to LSP and HSP1 function, we randomized the IPR sequence. Interestingly, randomization exhibited a much more pronounced effect on transcription from HSP1 than from LSP (**Fig. S1e**). These data suggest that the IPR is a functional component of the HSP1 core promoter. In order to further delimit the sequences of the IPR that contribute to HSP1 and/or LSP function, we constructed single-promoter templates that included the IPR through the distal TFAM-binding site (**Fig. 2a**). For HSP1, 138 bp upstream of the TSS were present; for LSP, 146 bp upstream of the TSS were present (**Fig. 2a**). The LSP TFAM-binding site (region 1 in **Fig. 2a**) could be deleted without any effect on transcription from HSP1, maintaining the 4-fold activation by TFAM observed for the DPT (-114 in **Fig. 2b**). However, additional deletions interfered with TFAM-activated transcription from HSP1 (-60 and -50 in **Fig. 2b**). In contrast, extending LSP by adding sequences beyond -50, exhibited marginal increases in TFAM activation, at best. We conclude that the sequence extending to 114-bp upstream of the HSP1 TSS should be considered an integral component of the TFAM-responsive element for this promoter. For LSP, the long-known TFAM-binding site appears sufficient.

As discussed above, activation of LSP by TFAM occurs by binding to the TFAM-binding site (region 1 in **Fig. 2a**), bending the DNA, and recruiting POLRMT and TFB2M (Gaspari et al., 2004; Lodeiro et al., 2010; Morozov et al., 2015; Rubio-Cosials et al., 2011). Transcription from LSP is strictly dependent on TFAM. The longstanding paradigm for TFAM-activated transcription from HSP1 suggested a comparable mechanism, with the TFAM-binding site being located between positions 532 and 553 of mtDNA (region 5 in **Fig. 2a**). However, here we show that transcription from HSP1 can occur independent of TFAM, and that the TFAM-responsive element includes more than the proposed HSP1 TFAM-binding site (region 5 in **Fig. 2a**). We therefore addressed the role of regions 1 and 5 using the DPT. Deletion of region 1 had no impact on transcription from HSP1 or its activation by TFAM, but essentially silenced transcription from LSP (**Fig. 3a**). Similarly, deletion of region 5 had no impact on transcription from LSP (**Fig. 3b**). Interestingly, region 5 was not required for TFAM-independent transcription from HSP1 (**Fig. 3b**). Surprisingly, deletion of region 5 converted TFAM from an activator of HSP transcription to a repressor (**Fig. 3b**). One possible explanation is that the deletion created a negative element by sequence

juxtaposition. To test this possibility, we inverted region 5 or randomized the sequence of region 5. Regardless of the change, TFAM continued to repress transcription from HSP1 instead of activate transcription (**Figs. 3c** and **3d**). Worth noting, inverting region 5 caused a significant increase in the yield from LSP (**Fig. 3c**); the basis for this activation requires further study. Our general conclusion from this line of investigation is that region 5 may be a transcription "insulator," a sequence that prevents TFAM bound upstream of region 5 from influencing transcription downstream where POLRMT-TFB2M assembles.

Historically, region 5 was first hypothesized to be a TFAM binding site based on its location at a position equivalent to that of the LSP TFAM-binding site, similarity to the sequence of the LSP TFAM binding site, and the finding that deletion of the sequence interfered with transcription from HSP1 in cell-free extracts (Bogenhagen et al., 1984). Later, more direct evidence for TFAM binding was obtained by DNase I footprinting (Fisher et al., 1987). In order to determine when and where TFAM binds to the DPT under transcription conditions, and as a function of TFAM concentration, we performed a DNase I footprinting experiment (**Fig. 4**). When the HSP1 template strand was labeled, protection was observed in region 1 as expected. A DNase I hypersensitive site appears to be due to TFAM bending of DNA at this site (**Figs. 4a** and **S2a**). Another hypersensitive site was observed at the edge of region 2 (**Fig. 4a** and **S2a**). There was clear nuclease protection between regions 2-4 (**Fig. 4a** and **S2a**). Protection in regions 2-4 was observed at lower TFAM concentrations than for region 5 (**Fig. 4a** and **S2a**). Indeed, even at the highest concentration of TFAM employed, complete protection of region 5 was not observed (**Fig. 4a** and **S2a**). When the LSP template strand was labeled, similar conclusions were reached (**Fig. 4b** and **S2b**). In addition to protection in regions 2-4, many sites of hypersensitivity were noted (**Fig. 4b** and **S2b**). Changes in region 5 were observed at the highest TFAM concentrations (**Fig. 4b** and **S2b**). These changes included weak protection and the presence of a hypersensitive site (**Fig. 4b** and **S2b**). These data are consistent with binding of TFAM to regions 2-4 correlating to activation of transcription from HSP1 and binding of TFAM to region 5 correlating to transcription inhibition.

Biophysical and structural studies of TFAM have shown that its carboxy-terminal tail (CTT), defined here as the final 26 amino acid residues (221-246, **Fig. 5a**), is not only able to interact with DNA (**Fig. 5b**) but also able to contribute to interactions with a second molecule of TFAM to form a dimer (**Fig. 5c**) (Wong et al., 2009). Such intermolecular dimerization of separate DNA-bound TFAM molecules could create DNA loops. Looping could impose DNA strain within the loop and perhaps even induce the DNase I hypersensitivity observed in regions 2-4 (**Fig. 4b** and **S2b**). TFAM deleted for its CTT remains competent to bind DNA (Wong et al., 2009). We evaluated the ability of tail-deleted TFAM (CT Δ 26) to stimulate transcription using the DPT. This derivative activated transcription from LSP with the same concentration

dependence as observed for the wild-type (WT) protein (compare LSP in **Figs. 5d** and **5e** to **Figs. 1c** and **1d**). Activation of this promoter was, however, reduced by approximately three-fold (compare LSP in **Fig. 5e** to **Fig. 1d**). In contrast, TFAM-CTΔ26 was completely unable to activate transcription from HSP1 (compare HSP1 in **Figs. 5d** and **5e** to **Figs. 1c** and **1d**). The transcription-repression activity of TFAM observed at higher ratios of TFAM:DPT was unchanged for both promoters by deleting the TFAM CTT (compare **Figs. 5d** and **5e** to **Figs. 1c** and **1d**). We also evaluated the interaction of TFAM-CTΔ26 with the DPT by using DNase I footprinting (**Fig. S4**). TFAM CTT truncation eliminated all of the DNase I-protected and hypersensitive sites in regions 2-4 (**Fig. S4**). Protection of the LSP TFAM-binding site in region 1 was not eliminated but was reduced several fold (**Fig. S4**). Together, these data are consistent with a model in which the TFAM CTT enables a specific interaction of TFAM with regions 2-4 that induces a specific conformation of regions 2-4, perhaps a DNA loop, that is absolutely essential for TFAM activation of HSP1.

Given the requirement of the TFAM CTT for activation of HSP1, we were curious to know if the TFAM CTT contributed to the repression of HSP1 observed when region 5 of the DPT was altered (**Fig. 3**). We performed a transcription reaction using the DPT containing an inversion of the region-5 sequence (**Figs. 5f** and **5g**). Interestingly, by using TFAM-CTΔ26 instead of TFAM, the inhibitory nature of the region-5 inversion was eliminated (**Figs. 5f** and **5g**). Unfortunately, DNase I footprinting failed to contribute additional insight to our understanding of the mechanism of repression or anti-repression (**Fig. S3**).

The preceding studies suggest that sequences upstream of HSP1 contribute more to the regulation of this promoter than previously recognized. For completeness, we asked if sequences downstream of the HSP1 TSS contributed to promoter regulation. The template used for this experiment contained only regions 2-5 upstream of the TSS. The downstream sequence varied from 45-bp, the length used in the DPT, to 145-bp (**Fig. 6a**). Unlike most prior experiments of this type (Litonin et al., 2010; Morozov and Temiakov, 2016; Shi et al., 2012), a ³²P-labeled primer was used instead of an α-³²P-labeled ribonucleotide. Therefore, in our experiments, an increase in labeled product signifies an increased transcript yield rather than more incorporation of labeled ribonucleotide substrate. A clear increase in yield was visible in templates as long as 95 bp; thereafter, the change was more subtle (**Fig. 6b**). Yield was dependent on the TFAM concentration (**Fig. 6b**). A quantitative perspective is provided in **Fig. 6c**. The optimal ratio of TFAM:template for maximal activation was 0.020-0.030 for all templates employed. By increasing the template to include the entirety of the tRNA^F gene, the TFAM-dependent yield from HSP1 was more than 8-fold higher than the TFAM-independent yield (**Fig. 6c**). We conclude that sequences downstream of the HSP1 TSS contribute to the output of this promoter, and perhaps also its regulation.

Because the TFAM-independent yield remained stoichiometric with template for all constructs tested, downstream sequence may increase RNA yield by somehow promoting template reutilization.

Our data are consistent with a model in which mtDNA sequences from 447 to 656 (see **Fig. 1a**) contribute to HSP1 transcriptional output in the presence of TFAM. This stretch of mtDNA begins with HVR3, includes the tRNA^F gene and ends near the 5' terminus of the 12S rRNA gene. To our knowledge, HVR3 has never been implicated in any aspect of mitochondrial gene expression. Because the same carboxy-terminal 26 amino acid residues of TFAM required for HVR3-dependent transactivation of HSP1 (**Fig. 5**) are required for dimerization and DNA looping (Ngo et al., 2011; Rubio-Cosials et al., 2011), it was possible that TFAM-mediated looping in HVR3 contributed to HSP1 transactivation. We used atomic force microscopy to investigate the complex formed upon addition of TFAM or TFAM-CTΔ26 to a 1653-bp DNA fragment, comprising the DPT (153 bp) extended upstream of LSP by 500 bp and downstream of HSP1 by 1000 bp. We will refer to this fragment here as mtDNA. An irrelevant sequence of similar length (1650-bp) was derived from pUC18 and used as a control for specificity. This DNA is termed pUC. We imaged samples of mtDNA in the absence (**Fig. S5**) and presence (**Fig. 7**) of TFAM. In the absence of TFAM, we measured an average contour length of 1684 ± 21 bp, consistent with the sequence used. We titrated mtDNA with TFAM, from 0.05 to 2.7 molecules of TFAM per bp. The lowest ratio is near optimal for transcription from HSP1; the highest ratio is inhibitory (**Fig. 6c**). At 0.05 molecules of TFAM per bp, loops were clearly visible (**Fig. 7a**). Further addition of TFAM converted loops into more compact structures (**Fig. 7a**). We used contour length to monitor this process quantitatively. A 4-fold reduction in contour length was observed in the presence of the highest concentration of TFAM (**Fig. 7b**).

The observation of loops at TFAM concentrations that promote maximal transcription is provocative. We then sought to determine the location(s) of the loops. We repeated the experiment at the low ratio of TFAM: mtDNA to obtain sufficient looped molecules for detailed characterization and statistical analysis (**Fig. S6a**). Formation of a DNA loop leads to a conformation that is elevated relative to the mica surface used for AFM (**Fig. S6b**). For each mtDNA observed, we measured the distance from the short end to the crossing of the strands (**Fig. S6c**), the size of the loops (**Fig. S6d**), and the sum of these two measurements, referred to as the total length (**Fig. S6e**). This analysis showed that the average distance from the short end of the molecule to the start of the loop was 245 ± 22 bp, and the average loop size was 409 ± 47 bp (**Fig. S6f**). A representative image illustrating a loop is shown in **Fig. 7c**. Interestingly, the position of the loop and its overall size are consistent with the entire transcription control region being located in the loop (**Fig. 7c**). The sequence of the average loop is highlighted in yellow in **Fig. 7d**.

We were then in a position to determine the importance of the TFAM CTT for DNA looping. Under conditions in which looped mtDNA molecules were prevalent in the presence of TFAM, looped mtDNA molecules were rare in the presence of TFAM-CT Δ 26 (**Fig. 8a**). In fact, only 24% (N=29) of molecules showed any loop-like conformations, similar to mtDNA alone (**Fig. 8b**). A trivial explanation for this observation is that TFAM-CT Δ 26 does not bind to DNA in the range from 10 to 100 nM. To test this, we used AFM imaging in liquid to determine the extent of TFAM-CT Δ 26 binding under the conditions of this study. Because the persistence length of mtDNA decreases upon TFAM binding, persistence length can be used to determine if and when TFAM-CT Δ 26 is bound to mtDNA (Murugesapillai et al., 2014; Rivetti et al., 1996; Zhang et al., 2009). As expected, a value of 50 ± 2 nm was measured in the absence of TFAM (**Fig. 8d**). The persistence length decreased to 30 ± 2 nm in the presence of 10 nM TFAM (**Fig. 8d**), consistent with previous measurements (Farge et al., 2012). The persistence length decreased only to 46 ± 1 nm in the presence of 10 nM TFAM-CT Δ 26, consistent with a somewhat lower binding affinity to mtDNA than TFAM (**Fig. 8d**). In contrast, the persistence length of mtDNA decreased to 33 ± 1 nm in the presence of 100 nM TFAM-CT Δ 26, similar to that observed for 10 nM TFAM (**Fig. 8d**), suggesting that TFAM-CT Δ 26 has approximately ten-fold weaker binding to mtDNA relative to TFAM. To determine if this decrease in binding affinity could be responsible for the observed inability of TFAM-CT Δ 26 to induce mtDNA looping, we also measured the looping fraction at 100 nM TFAM-CT Δ 26, and the looping probably was still similar to that observed for mtDNA only (**Fig. 8c**). We therefore conclude that the lack of looping enhancement by TFAM-CT Δ 26 is due to the properties of the CTT and not due to reduced DNA binding affinity upon removal of the CTT. Worth noting, the efficiency with which TFAM-CT Δ 26 condenses mtDNA into nucleoids is reduced comparably with the reduced efficiency of DNA binding (**Fig. S7c**). Interestingly, compaction occurs in the absence of loops for TFAM-CT Δ 26 (**Fig. S7c**), thus ruling out loops as obligatory intermediates in the mechanism of DNA compaction by TFAM.

Next, we determined the extent to which the mtDNA sequence influences the size and location of the loops observed in the presence of TFAM. To do so, we performed an analogous experiment using pUC. The average contour length was 1624 ± 96 (N=47), consistent with the known length of the construct. In the presence of 10 nM TFAM, looping was observed. The average of the shortest distance to the loop was 265 ± 43 bp, and the average loop size was 235 ± 43 bp. These values are similar to those measured for mtDNA; however, the region that corresponds to the IPR location on mtDNA remains outside the average loop.

Finally, we determined the extent to which the mtDNA sequence influences the number of loops per molecule observed by AFM in the presence of TFAM. This experiment required interpretation

of each observed crossover. Three types of crossovers (or bridges) can be identified: protein-bound DNA (purple arrow in **Fig. S8a**); TFAM-mediated DNA bridges or crossovers (blue arrow in **Fig S8a**); and naked DNA crossovers (green arrow in **Fig. S8a**). Each of these conformations exhibits a diagnostic height in the AFM image (**Fig. S8b**). For the mtDNA molecules evaluated, 77% (N=15) of the loops were mediated by TFAM (**Fig. S8c**). However, only 47% (N=19) of pUC molecules contained TFAM-mediated loops (**Fig. S8c**). These data are consistent with loop formation resulting from sequence-specific recruitment of TFAM to mtDNA.

To confirm that TFAM-induced DNA looping occurs in the absence of a surface, we performed optical-tweezers experiments (Chaurasiya et al., 2010; Murugesapillai et al., 2014). In such experiments, a 48.5 kbp DNA is tethered at each of its termini to beads; one bead is held in an optical trap while the other is held on a glass micropipette. At low force and extension [less than one piconewton (pN)], the DNA is free to sample multiple conformations. A force can be exerted to pull one bead away from the other and extend the tethered DNA. In the absence of TFAM, the force-extension curve was smooth (**Fig. 9a**). However, in the presence of TFAM, jumps were present in the force-extension curve, consistent with the force breaking loops created by TFAM dimers (**Fig. 9b**). Loops were readily detected when the DNA was extended at a rate of 970 nm/s but not when the DNA was extended at a rate of 100 nm/s, suggesting that TFAM binds, dimerizes and dissociates during the experiment with the slower pulling rate. The DNA was relaxed and immediately extended, yielding a force-extension curve resembling naked DNA (**Fig. 9c**). However, when the DNA was relaxed, held for several minutes, and then extended, a jump in the force-extension curve was apparent (**Fig. 9d**). These data suggest that the DNA looping process is an equilibrium process, with TFAM dimerization likely representing the rate-limiting step. Finally, we determined the strength of the TFAM dimer by measuring the average force required to break a loop (**Fig. 9e**). This value was 17 ± 2 pN (N=15). This is similar to the force required to break loops formed by the yeast HMG-box protein, HMO1 (Murugesapillai et al., 2014).

DISCUSSION

The pioneering work of the Attardi and Clayton laboratories identified three transcription start sites on mitochondrial DNA, one on the light strand and two on the heavy strand (Chang and Clayton, 1984; Montoya et al., 1982; Montoya et al., 1983). These studies also revealed that transcription from the first heavy-strand promoter (HSP1) produced far more RNA than transcription from the second heavy-strand promoter (HSP2) (Montoya et al., 1983). Collectively, this early work, much of which was performed using cell-based and/or cell-free experiments, made a good case for differential regulation of mitochondrial

transcription. Fast forward 30 years to the current era of studying mitochondrial transcription using purified components and a case has now been made for only two promoters: LSP and HSP1, and the absence of differential regulation (Litonin et al., 2010; Morozov and Temiakov, 2016; Shi et al., 2012). The goal of this study was to resolve this paradox.

Studies that were unable to observe differential regulation of LSP and HSP1 in reconstituted systems shared several features of concern (Litonin et al., 2010; Shi et al., 2012). First, "minimal" promoters were used. Second, reaction products were labeled with a radioactive nucleotide, a condition requiring distorted nucleotide concentrations and associated caveats. Third, absolute transcript quantification was not performed, so it was unclear how many initiation events were being monitored or if all templates were (re)used. In this study, we address all of these concerns. We show that transcription from HSP1 can occur in the absence of TFAM (**Fig. 1**). Quantitative analysis shows that in the absence of TFAM only one transcript is produced from HSP1 per template (**Fig. 1**). Thus, the inability of others to detect a single round of transcription would explain their inability to observe the TFAM-independent transcription from HSP1.

TFAM functions differently at LSP and HSP1 (**Fig. 10**). TFAM is essential for transcription initiation at LSP (**Fig. 1**). Here TFAM functions stoichiometrically with the template and supports multiple rounds of reinitiation from LSP as 5-10 LSP transcripts are produced from each template (**Figs. 1 and 2**). In contrast, TFAM appears to be essential only for reinitiation from HSP1 (**Fig. 1**). We suggest that the nascent transcript may not be efficiently displaced in the absence of TFAM. In the absence of TFAM, regardless of the amount of DNA upstream or downstream of the HSP1 transcription start site (TSS) and/or the duration of the incubation, transcript yield was always stoichiometric with template (**Figs. 1, 2 and 6**). The first 20 nt of the HSP1 transcript is GC rich (see 561-581 in **Fig. 7d**), in contrast to the AT-rich LSP transcript. Perhaps the failure of nascent RNA to be displaced leads to formation of an R-loop (Wanrooij et al., 2012; Wanrooij et al., 2010). It is well known that POLRMT elongation can be impeded by stretches of guanylate residues, for example conserved sequence box II (CSB II, 299-315 in **Fig. 7d**), perhaps because of their propensity to form quadruplexes (Wanrooij et al., 2012; Wanrooij et al., 2010). Arrest at CSB II can be prevented by the mitochondrial transcription elongation factor, TEFM (Agaronyan et al., 2015). TFAM may prevent arrest at guanylate stretches during initiation. Deep-sequencing approaches that permit identification of sites of transcription pausing and/or arrest identified the region between positions 600 and 700 of the human mtDNA as sites of POLRMT pausing or arrest (Blumberg et al., 2017).

The carboxyl terminus of TFAM is essential for its transcription-activation function at HSP1 but not at LSP (**Fig. 5**). We and others have previously made similar observations (Lodeiro et al., 2012; Ngo et al.,

2014); however, a debate remains (Morozov and Temiakov, 2016). We now show that production of more than one transcript from HSP1 is impossible when the TFAM is truncated by removal of its carboxy-terminal 26 amino acid residues, though there is no impact on the first round of transcription from HSP1 (**Fig. 5**). In contrast, transcription from LSP is only modestly reduced (**Fig. 5**). Transcription repression activity of TFAM is unaltered by loss of the carboxyl terminus (**Fig. 5**). The carboxy-terminal tail has been shown to contribute to intermolecular interactions that could give rise to DNA looping (**Fig. 5c**) (Ngo et al., 2011; Rubio-Cosials et al., 2011). DNA looping is a well-established paradigm for transcription activation in prokaryotes (Cournac and Plumbridge, 2013). Looping could contribute to displacement of nascent RNA by inducing sufficient strain in the template to prevent stable hybridization of the transcript, thus facilitating multiple rounds of initiation at HSP1. If the role of looping is to increase yield from HSP1, then post-translational modifications in the carboxy-terminal tail could regulate loop formation and therefore transcriptional output from HSP1. While numerous post-translational modifications of TFAM have been observed (Grimsrud et al., 2012), none of these map to the carboxy-terminal tail. It should be noted that most of the proteomic studies have been performed using rapidly-dividing cancer cell lines; mitochondrial transcription and biogenesis are quite active under these conditions.

TFAM-mediated looping of pUC19 plasmid DNA has been observed using atomic force microscopy (AFM) imaging in air (Kaufman et al., 2007), although this result was later suggested to be due to random crossovers rather than TFAM-mediated looping (Farge et al., 2012). Here we show using both AFM imaging in liquid and optical tweezers that TFAM actively mediates DNA looping, and our AFM studies show that it creates loops on mtDNA sequences as well (**Fig. 7c**). The average size of the loops formed on mtDNA sequences is larger than observed on plasmid DNA (**Fig. S6**), and more loops are found per molecule with mtDNA sequences than with plasmid DNA (**Fig. S8**). TFAM-mediated loops appear to be localized to the control region, and most loops include HVR3 (**Fig. S6**). Looping is strictly dependent on the carboxy-terminal tail of TFAM (**Fig. 8**). Interestingly, the inability to form loops has little impact on the ability of TFAM to compact DNA (**Fig. S7**), suggesting that loops are not obligatory intermediates on the path for compaction as suggested previously (Kaufman et al., 2007). We are intrigued by the possibility that TFAM-mediated looping represents another layer of transcriptional regulation (**Fig. 10**). TFAM binding induces a U-turn into the DNA at LSP (Ngo et al., 2011; Rubio-Cosials et al., 2011). At HSP, in addition to a U-turn, looping is required.

A particularly unexpected outcome of this study is the finding that sequences as far as 114-bp upstream (**Fig. 2**) and 95-bp downstream (**Fig. 6**) of the HSP1 TSS contribute to the transcriptional output of HSP1. Collectively, our data suggest that TFAM binds within HVR3 and the tRNA^F gene to create loops

that are used to facilitate transcription. TFAM protection of HVR3 is readily observed by DNase I footprinting (**Fig. 4**, **Fig. S2**); however, the footprint is lost completely when the TFAM carboxy-terminal tail is deleted (**Fig. S4**).

Dogma in the field has been that transcription initiation from LSP and HSP1 is regulated by similar or identical mechanisms (Litonin et al., 2010). Because of this perspective, there has been an effort to force HSP1 to fit the paradigm established for LSP. One example is the widely-held belief that the TFAM-binding site of HSP1 (referred to here as region 5 and shown in red at 532-553 in **Fig. 7d**) resides approximately 50-bp upstream of the start site, as does the TFAM-binding site of LSP (referred to here as region 1 and shown in red at 425-446 in **Fig. 7d**). Here we show that region 5 functions differently than region 1. Binding of region 5 by TFAM occurs at concentrations of TFAM in which transcription from HSP1 is inhibited (**Fig. 4** and **Fig. S2**). Alterations of region 5 converts TFAM into a potent repressor of transcription from HSP1 (**Fig. 3**). Repression by TFAM requires its carboxy-terminal tail (**Fig. 5**). It is possible that region 5 has evolved to prevent TFAM binding, bending and looping from occluding the site used by POLRMT and TFB2M to bind to the promoter. If this is the case, then it might also be possible to use DNA modifications, for example cytosine methylation or even guanine oxidation, of region 5 to control transcription from HSP1.

Our studies provide the first suggestion for the role of HVR3, the LSP-HSP1 inter-promoter region (IPR), in mitochondrial transcription. We suggest that HVR3 sequences from 450 to 560 are an integral component of HSP1. The transcriptional function of HVR3 is mediated by TFAM binding. Footprinting demonstrates that binding of TFAM to HVR3 is not random (**Fig. 4**). After binding, TFAM-TFAM interactions lead to the formation of loops that permit multiple rounds of transcription to occur from HSP1. Somatic mutations in HVR3 have been linked to diseases, such as cancer (**Fig. S9**). The mechanism is completely unknown. Our studies suggest that these mutations could possibly alter mitochondrial gene expression. Also worth noting is the fact that many of the somatic mutations linked to disease actually are present as polymorphisms in "normal" individuals (**Fig. S9**). It is possible that disease-associated polymorphisms predispose to disease. We conclude that natural variation in HVR3 may lead to between-individual differences in mitochondrial transcription or, in the worst cases, misregulated mitochondrial transcription that contributes to disease.

In conclusion, the experiments described here demonstrate differential regulation of transcription (re)initiation from LSP and HSP1 mediated by TFAM (**Fig. 10**). Initiation at LSP likely begins with TFAM binding and bending promoter DNA followed by recruitment of POLRMT and TFB2M, likely already in complex. In the presence of nucleoside triphosphates, di- and/or tri-nucleotide products are produced

and further extended into full-length RNA. LSP templates are efficiently reutilized and factors recycled. Initiation at HSP can occur with POLRMT and TFB2M only. However, after production of transcript, the template fails to be reutilized. As stated above, this could be due to formation of R-loops, triplexes or some other heteroduplex. Additional studies will be required to understand this phenomenon. In the presence of TFAM, HSP1 template reutilization becomes feasible. We attribute template reutilization to the ability of TFAM to form loops in a manner dependent on its carboxy-terminal tail. The molecular basis for this property of TFAM-mediated loops remains unclear. Collectively, these studies reveal clear differences in the cis-acting elements and trans-acting factors required for transcription initiation from LSP and HSP1. These differences are presumably exploited for differential promoter regulation, and may play roles in disease.

MATERIALS AND METHODS

Materials. The proteins used to perform this study can be obtained from Indigo Biosciences (State College, PA). DNA oligonucleotides were from Integrated DNA Technologies. Oligonucleotide primers used in this study are listed in Table 1. RNA oligonucleotides were obtained from Dharmacon. Restriction endonucleases were from New England Biolabs. T4 polynucleotide kinase was from USB. RQ1 DNase was from Promega. Ultrapure NTP solutions were from GE Healthcare. [γ - 32 P]ATP (7000 Ci/mmol) was from Perkin Elmer. HPLC purified [γ - 32 P] ATP (3000 Ci/mmol) was from MP Biomedical. 10-bp DNA ladder was from Invitrogen. All other reagents were of the highest grade available from Sigma, Fisher or VWR.

Preparation of dual promoter template. Dual promoter DNA templates were prepared by PCR. The D-loop region of mtDNA (CRS, NC_012920) was cloned into pUC18 and used as template for PCR. Seven 100 μ L reactions contained final concentration of 0.5 ng/ μ L plasmid template, 1 μ M LSP+35-For primer, 1 μ M HSP1+45-Rev primer, 0.3 mM dNTP, 1 \times Thermo Pol Buffer (NEB), and DeepVent DNA polymerase (NEB). Dual promoter deletion, inverted, randomized and AFM constructs were also prepared by PCR but using the appropriate forward and reverse primers listed in Table 1. PCR cycling conditions are as follows: 1 cycle at 95°C for 4 min; 40 cycles of denaturing at 95°C for 30 sec, annealing at 57°C for 30 sec, and extension at 72°C for 20 sec; 1 cycle at 72°C for 10 min. PCR products were purified with Wizard SV gel and PCR Clean-up System (Promega), DNA was eluted in 80 μ L TE (10 mM Tris-HCl, pH 8.0 buffer and 0.1 mM EDTA) and finally diluted to 1 μ M in TE buffer. Extinction coefficients for each DNA construct were calculated with IDT DNA technologies tool (<http://biophysics.idtdna.com/UVSpectrum.html>).

In vitro transcription assays. Reactions were performed in 1× reaction buffer (10 mM HEPES pH 7.5, 100 mM NaCl, 10 mM MgCl₂, 1 mM TCEP and 0.1 µg/µL BSA) with 10 µM ³²P-end-labeled RNA primer (pAAA), 500 µM NTP and 100 nM DNA template. Reactions were performed by incubating template DNA in reaction buffer at 32°C for 5 min and then adding in the following order: TFAM (varying concentrations), TFB2M and POLRMT (1 µM each). Between each addition of protein to the reaction there was an incubation time of 1 minute. After addition of POLRMT, the reaction was allowed to incubate at 32°C for 5 to 60 min. At each time point 4 µL of the reaction mix were quenched into 8 µL of stop buffer (79.2 % formamide, 0.025% bromophenol blue, 0.025% xylene cyanol and 50 mM EDTA final). Products were resolved by denaturing 20% (37:3, acrylamide:bis-acrylamide ratio) PAGE. Proteins were diluted immediately prior to use in 10 mM HEPES, pH 7.5, 1 mM TCEP, and 20% glycerol. The volume of protein added to any reaction was always less than or equal to one-tenth the total volume. Any deviations from the above are indicated in the appropriate figure legend.

DNase I footprinting. Footprinting reaction probes of dual promoter template were prepared as described above except that the PCR was performed with either LSP+35-For or HSP1+45-Rev primer which were ³²P-labeled prior to PCR. DNase I footprinting was performed in the presence of TFAM alone or in the presence of TFAM, TFB2M and POLRMT. Reaction buffer consisting of 10 mM HEPES, pH 7.5, 10 mM MgCl₂, 10 mM CaCl₂, 100 mM NaCl, 1 mM TCEP, and 0.1 mg/mL BSA at final concentration was prepared and incubated for 1 min at 32°C prior to addition of proteins in the following order: TFAM (varying concentrations), TFB2M (1 µM) and POLRMT (1 µM). Between each addition of protein to the reaction there was an incubation time of 1 min. Upon addition of all proteins, the reaction was incubated for 5 min followed by adding mixture of RQ1 DNase (0.002 units/µL, final) and CaCl₂ (1 mM, final). The DNA digestion for 2 min at 32°C was quenched by addition of EDTA and urea at final 40 mM and 840 mM, respectively, in 5 µL. Proteinase K was mixed with each reaction at final 0.04 µg/µL, and incubated at 50°C for 15 min to digest TFAM and other proteins. Finally, 20 µL of loading buffer (79.2% formamide, 50 mM EDTA, 0.025% xylene cyanol and 0.025% bromophenol blue, and 5 µM trap oligonucleotide DNA, a 90-bp oligo of either the LSP or HSP strand, whichever was ³²P-labeled) were added to each reaction. Samples were heated at 95°C for 5 min prior to loading PAGE. Products were resolved by denaturing PAGE on 7% gels (37:3, acrylamide:bis-acrylamide ratio).

Characterization of TFAM-DNA interactions using atomic force microscopy in liquid. The sample was prepared using a freshly cleaved mica surface coated with 3-aminopropyl-triethoxy silatrane (APTES)

(Shlyakhtenko et al., 2013). The mtDNA construct used for AFM was prepared by PCR as described above but with extension times of 2 min and was a fragment of mtDNA that began 510 bp downstream of LSP and 1000 bp downstream of HSP1. The concentration of mtDNA and pUC18 (pUC) used for the experiment is 0.11 nM. The concentration of protein ranges from 10 nM to 500 nM, corresponding to a ratio of protein to DNA varying from 0.05 to 2.7 [TFAM]/[bp]. The experiment was carried out using 25 mM NaCl, 25 mM Hepes at pH 7.5. We used a Bruker Nanoscope V MultiMode 8 with PeakForce Tapping™ mode atomic force microscope (AFM) to image in liquid TFAM proteins bound to DNA. We use peak force as an imaging signal, in which force information is collected at every pixel of the image. To image in a liquid environment, a silicon cantilever was used (resonance frequency=70 kHz, spring constant =0.4N/m and tip radius = 2 nm). Image processing was done using Nanoscope Analysis software. The scan range was varied from 1 μm X 1μm to 2μm X 2μm at 512 × 512 pixels and at 1024 × 1024 pixels, respectively. The AFM images were quantified using NCTracer, a software program developed by the Neurogeometry Lab at Northeastern University (Chothani et al., 2011; Gala et al., 2014). To determine the persistence length p the orientation differences θ along the DNA as a function of contour length L were fit to the 3D wormlike chain (WLC) model

$$\langle \cos(\theta) \rangle = e^{-L/p} \quad (1)$$

Characterization of TFAM-DNA interactions using optical tweezers. Optical tweezers were used to further investigate TFAM-DNA interactions. In this experiment, two high power laser beams are focused into a small spot of approximately 1 μm size, which acts as a trap for high refractive index beads compared to the surrounding medium. A single phage-λ DNA molecule (48,500 base pairs) labeled on the termini of its opposite strands with biotin can be attached by its termini to streptavidin-coated polystyrene beads (Bangs Labs). This allows TFAM-DNA interactions to be studied under tension (McCauley et al., 2013; Murugesapillai et al., 2014). The experiments were carried out in a buffer containing 10 mM Hepes, pH 7.5, and 100 mM Na⁺. The forces are measured in picoNewtons and extension in nm/bp, the total DNA extension divided by the total number of DNA base pairs. To observe and study loop formation in the presence of TFAM, DNA was brought to a very low extension at a force less than one picoNewton (Murugesapillai et al., 2014).

ACKNOWLEDGMENTS

We appreciate the thoughtful, constructive comments on various drafts of the manuscript by our colleagues. This study was funded, in part, by the endowment from the Eberly Family Chair in

Biochemistry and Molecular Biology to C.E.C. L.J.M. is supported by a grant from NIH, GM075965. M.W. and his laboratory are supported by a grant from NSF, MCB-1243883.

AUTHOR CONTRIBUTIONS

A.U., M.F.L, J.J.A and C.E.C designed the biochemical aspects of the study. D.M, L.J.M. and M.W. designed the biophysical aspects of the study. A.U., Y.W., M.F.L., S.P. and J.J.A conducted experiments. All authors analyzed the data and wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

- Agaronyan, K., Morozov, Y.I., Anikin, M., and Temiakov, D. (2015). Mitochondrial biology. Replication-transcription switch in human mitochondria. *Science (New York, NY)* 347, 548-551.
- Blumberg, A., Rice, E.J., Kundaje, A., Danko, C.G., and Mishmar, D. (2017). Initiation of mtDNA transcription is followed by pausing, and diverges across human cell types and during evolution. *Genome research* 27, 362-373.
- Bogenhagen, D.F., Applegate, E.F., and Yoza, B.K. (1984). Identification of a promoter for transcription of the heavy strand of human mtDNA: in vitro transcription and deletion mutagenesis. *Cell* 36, 1105-1113.
- Cantatore, P., and Attardi, G. (1980). Mapping of nascent light and heavy strand transcripts on the physical map of HeLa cell mitochondrial DNA. *Nucleic acids research* 8, 2605-2625.
- Chang, D.D., and Clayton, D.A. (1984). Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. *Cell* 36, 635-643.
- Chaurasiya, K.R., Paramanathan, T., McCauley, M.J., and Williams, M.C. (2010). Biophysical characterization of DNA binding from single molecule force measurements. *Physics of life reviews* 7, 299-341.
- Chothani, P., Mehta, V., and Stepanyants, A. (2011). Automated tracing of neurites from light microscopy stacks of images. *Neuroinformatics* 9, 263-278.

Clayton, D.A. (1984). Transcription of the mammalian mitochondrial genome. *Annual review of biochemistry* 53, 573-594.

Cournac, A., and Plumbridge, J. (2013). DNA looping in prokaryotes: experimental and theoretical approaches. *Journal of bacteriology* 195, 1109-1119.

Falkenberg, M., Gaspari, M., Rantanen, A., Trifunovic, A., Larsson, N.G., and Gustafsson, C.M. (2002). Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. *Nature genetics* 31, 289-294.

Farge, G., Laurens, N., Broekmans, O.D., van den Wildenberg, S.M., Dekker, L.C., Gaspari, M., Gustafsson, C.M., Peterman, E.J., Falkenberg, M., and Wuite, G.J. (2012). Protein sliding and DNA denaturation are essential for DNA organization by human mitochondrial transcription factor A. *Nature communications* 3, 1013.

Fisher, R.P., and Clayton, D.A. (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. *The Journal of biological chemistry* 260, 11330-11338.

Fisher, R.P., Topper, J.N., and Clayton, D.A. (1987). Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements. *Cell* 50, 247-258.

Gaines, G., and Attardi, G. (1984). Highly efficient RNA-synthesizing system that uses isolated human mitochondria: new initiation events and in vivo-like processing patterns. *Molecular and cellular biology* 4, 1605-1617.

Gaines, G., Rossi, C., and Attardi, G. (1987). Markedly different ATP requirements for rRNA synthesis and mtDNA light strand transcription versus mRNA synthesis in isolated human mitochondria. *The Journal of biological chemistry* 262, 1907-1915.

Gala, R., Chapeton, J., Jitesh, J., Bhavsar, C., and Stepanyants, A. (2014). Active learning of neuron morphology for accurate automated tracing of neurites. *Frontiers in neuroanatomy* 8, 37.

Gaspari, M., Falkenberg, M., Larsson, N.G., and Gustafsson, C.M. (2004). The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells. *The EMBO journal* 23, 4606-4614.

Grimsrud, P.A., Carson, J.J., Hebert, A.S., Hubler, S.L., Niemi, N.M., Bailey, D.J., Jochem, A., Stapleton, D.S., Keller, M.P., Westphall, M.S., *et al.* (2012). A quantitative map of the liver mitochondrial phosphoproteome reveals posttranslational control of ketogenesis. *Cell metabolism* 16, 672-683.

Kaufman, B.A., Durisic, N., Mativetsky, J.M., Costantino, S., Hancock, M.A., Grutter, P., and Shoubridge, E.A. (2007). The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures. *Molecular biology of the cell* 18, 3225-3236.

Litonin, D., Sologub, M., Shi, Y., Savkina, M., Anikin, M., Falkenberg, M., Gustafsson, C.M., and Temiakov, D. (2010). Human mitochondrial transcription revisited: only TFAM and TFB2M are required for transcription of the mitochondrial genes in vitro. *The Journal of biological chemistry* 285, 18129-18133.

Lodeiro, M.F., Uchida, A., Bestwick, M., Moustafa, I.M., Arnold, J.J., Shadel, G.S., and Cameron, C.E. (2012). Transcription from the second heavy-strand promoter of human mtDNA is repressed by transcription factor A in vitro. *Proceedings of the National Academy of Sciences of the United States of America* 109, 6513-6518.

Lodeiro, M.F., Uchida, A.U., Arnold, J.J., Reynolds, S.L., Moustafa, I.M., and Cameron, C.E. (2010). Identification of multiple rate-limiting steps during the human mitochondrial transcription cycle in vitro. *The Journal of biological chemistry* 285, 16387-16402.

Malarkey, C.S., and Churchill, M.E. (2012). The high mobility group box: the ultimate utility player of a cell. *Trends in biochemical sciences* 37, 553-562.

McCauley, M.J., Rueter, E.M., Rouzina, I., Maher, L.J., 3rd, and Williams, M.C. (2013). Single-molecule kinetics reveal microscopic mechanism by which High-Mobility Group B proteins alter DNA flexibility. *Nucleic acids research* 41, 167-181.

Montoya, J., Christianson, T., Levens, D., Rabinowitz, M., and Attardi, G. (1982). Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* 79, 7195-7199.

Montoya, J., Gaines, G.L., and Attardi, G. (1983). The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. *Cell* **34**, 151-159.

Morozov, Y.I., Agaronyan, K., Cheung, A.C., Anikin, M., Cramer, P., and Temiakov, D. (2014). A novel intermediate in transcription initiation by human mitochondrial RNA polymerase. *Nucleic acids research* **42**, 3884-3893.

Morozov, Y.I., Parshin, A.V., Agaronyan, K., Cheung, A.C., Anikin, M., Cramer, P., and Temiakov, D. (2015). A model for transcription initiation in human mitochondria. *Nucleic acids research* **43**, 3726-3735.

Morozov, Y.I., and Temiakov, D. (2016). Human Mitochondrial Transcription Initiation Complexes Have Similar Topology on the Light and Heavy Strand Promoters. *The Journal of biological chemistry* **291**, 13432-13435.

Murugesapillai, D., McCauley, M.J., Huo, R., Nelson Holte, M.H., Stepanyants, A., Maher, L.J., 3rd, Israeloff, N.E., and Williams, M.C. (2014). DNA bridging and looping by HMO1 provides a mechanism for stabilizing nucleosome-free chromatin. *Nucleic acids research* **42**, 8996-9004.

Murugesapillai, D., Micah, J., McCauley, L., Maher III, J., and Williams, M.C. (2016). Single-molecule studies of high mobility group B architectural DNA bending proteins. *Biophysical Reviews*.

Ngo, H.B., Kaiser, J.T., and Chan, D.C. (2011). The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA. *Nature structural & molecular biology* **18**, 1290-1296.

Ngo, H.B., Lovely, G.A., Phillips, R., and Chan, D.C. (2014). Distinct structural features of TFAM drive mitochondrial DNA packaging versus transcriptional activation. *Nature communications* **5**, 3077.

Ojala, D., Merkel, C., Gelfand, R., and Attardi, G. (1980). The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. *Cell* **22**, 393-403.

Ojala, D., Montoya, J., and Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**, 470-474.

Rivetti, C., Guthold, M., and Bustamante, C. (1996). Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. *Journal of molecular biology* **264**, 919-932.

Rubio-Cosials, A., Sidow, J.F., Jimenez-Menendez, N., Fernandez-Millan, P., Montoya, J., Jacobs, H.T., Coll, M., Bernado, P., and Sola, M. (2011). Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter. *Nature structural & molecular biology* *18*, 1281-1289.

Shi, Y., Dierckx, A., Wanrooij, P.H., Wanrooij, S., Larsson, N.G., Wilhelmsson, L.M., Falkenberg, M., and Gustafsson, C.M. (2012). Mammalian transcription factor A is a core component of the mitochondrial transcription machinery. *Proceedings of the National Academy of Sciences of the United States of America* *109*, 16510-16515.

Shlyakhtenko, L.S., Gall, A.A., and Lyubchenko, Y.L. (2013). Mica functionalization for imaging of DNA and protein-DNA complexes with atomic force microscopy. *Methods in molecular biology* *931*, 295-312.

Shutt, T.E., Lodeiro, M.F., Cotney, J., Cameron, C.E., and Shadel, G.S. (2010). Core human mitochondrial transcription apparatus is a regulated two-component system in vitro. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 12133-12138.

Sologub, M., Litonin, D., Anikin, M., Mustaev, A., and Temiakov, D. (2009). TFB2 is a transient component of the catalytic site of the human mitochondrial RNA polymerase. *Cell* *139*, 934-944.

Wanrooij, P.H., Uhler, J.P., Shi, Y., Westerlund, F., Falkenberg, M., and Gustafsson, C.M. (2012). A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. *Nucleic acids research* *40*, 10334-10344.

Wanrooij, P.H., Uhler, J.P., Simonsson, T., Falkenberg, M., and Gustafsson, C.M. (2010). G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 16072-16077.

Wong, T.S., Rajagopalan, S., Freund, S.M., Rutherford, T.J., Andreeva, A., Townsley, F.M., Petrovich, M., and Fersht, A.R. (2009). Biophysical characterizations of human mitochondrial transcription factor A and its binding to tumor suppressor p53. *Nucleic acids research* *37*, 6765-6783.

Zhang, J., McCauley, M.J., Maher, L.J., 3rd, Williams, M.C., and Israeloff, N.E. (2009). Mechanism of DNA flexibility enhancement by HMGB proteins. *Nucleic acids research* *37*, 1107-1114.

Zollo, O., Tiranti, V., and Sondheimer, N. (2012). Transcriptional requirements of the distal heavy-strand promoter of mtDNA. *Proceedings of the National Academy of Sciences of the United States of America* *109*, 6508-6512.

Table 1. DNA Oligonucleotides used in this study.

DNA Oligo	Sequence
LSP+35-For	5 'AACACCAGCCTAACCAGATTTC3 '
HSP1+45-Rev	5 'ATTGCTTTGAGGAGGTAAGC3 '
HSP1-138-For	5 'TTAACAGTCACCCCCCAACTAAC3 '
HSP1-114-For	5 'CATTATTTTCCCCTCCCCTC3 '
HSP1-60-For	5 'CCATCCTACCCAGCACACAC3 '
HSP1-50-For	5 'CAGCACACACACACCGCTGC3 '
LSP-146-Rev	5 'GGTTGGTTCGGGGTATGGG3 '
LSP-115-Rev	5 'GTGTGTGTGCTGGGTAGGATG3 '
LSP-86-Rev	5 'TTGTATTGATGAGATTAGTAG3 '
LSP-69-Rev	5 'GTAGTATGGGAGTGGGAGG3 '
LSP-40-Rev	5 'GTGTTAGTTGGGGGGTGAC3 '
IPR-Reg2-For	5 'ATTATTTTCCCCTCCCCTCCC3 '
HSP1+70-Rev	5 'GAGCCCGTCTAAACATTTTC3 '
HSP1+95-Rev	5 'AACCTATTTGTTTATGGGGTG3 '
HSP1+120-Rev	5 'AGAGCTAATAGAAAGGCT3 '
HSP1+145-Rev	5 'GATGCTTGCATGTGTAATCTTACTAAGAGCTAAT3 '
□-Reg1-For	5 'GCACTTTTATTATTTTCCCCTCCCAC3 '
□-Reg1-Rev	5 'GGAAAATAATAAAAGTGCATACCGCC3 '
□-Reg5-For	5 'CAGCACACACACACCAAACCCCAAAGACAC3 '
□-Reg5-Rev	5 'GGTTTGGTGTGTGTGTGCTGGGTAG3 '
Inv-Reg5-For	5 'TGGTTCGGGGTATGGGGTTAGCAGCGGTACCAAACCCCAAAGACACCC3 '
Inv-Reg5-Rev	5 'ACCGCTGCTAACCCCATACCCGAACCAGTGTGTGTGCTGGGTAGGATGG3 '
Rdm-Reg5-For	5 'GACCCGACCTCCCTACCACCACTCCGAACGACCACCCACCAAACCCCAAAGACACCC3 '
Rdm-Reg5-Rev	5 'TGGTGGTAGGGAGGTCGGGTCGGGGTGTATGAGGTGGGTCTTTGTATTGATGAGATTAGTAGTATGG3 '
LSP+510-For	5 'CTTCCGGCTCGTATGTTGTGTGGAATTG3 '
HSP1+1000-Rev	5 'ACTTTAAAAGTGCTCATCATTGG3 '

Figures and Figure Legends

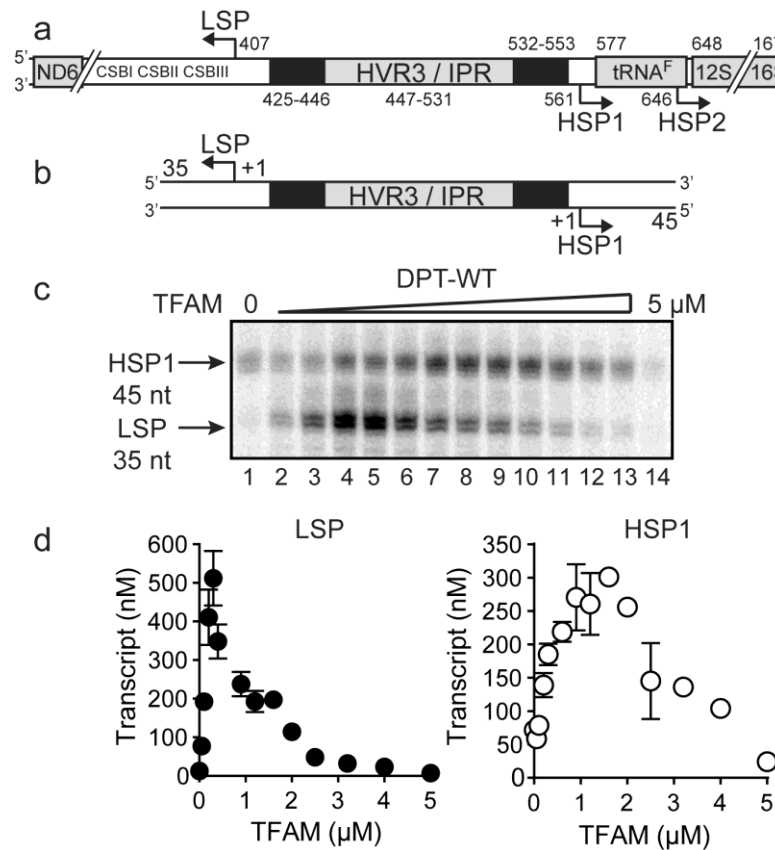


Figure 1. Dual promoter construct. (a) Schematic of the mtDNA transcriptional control region including the three promoters: LSP, HSP1 and HSP2; and interpromoter region (IPR), also known as hypervariable region 3 (HVR3). The relative positions of conserved sequence boxes (CSB I, II and III), ND6, tRNA^F, 12S and 16S rRNA genes are shown. Numbering according to the standard Cambridge mtDNA sequence. Black boxes indicate the putative TFAM binding sites. (b) Dual promoter DNA oligonucleotide (234 bp) template containing LSP, HSP1 and HVR3/IPR used for in vitro transcription reactions. This oligo gives rise to LSP and HSP1 derived RNA transcripts 35 and 45 nts long, respectively. (c) Run-off transcription products using the dual promoter DNA oligo template (DPT-WT) and increasing concentrations of TFAM (0-5 μ M) resolved by denaturing PAGE. (d) Amount of LSP (35 nt) and HSP1 (45 nt) RNA transcripts produced using the dual promoter construct plotted as a function of TFAM concentration. Data are means from three independent experiments. Error bars represent \pm S.E.M.

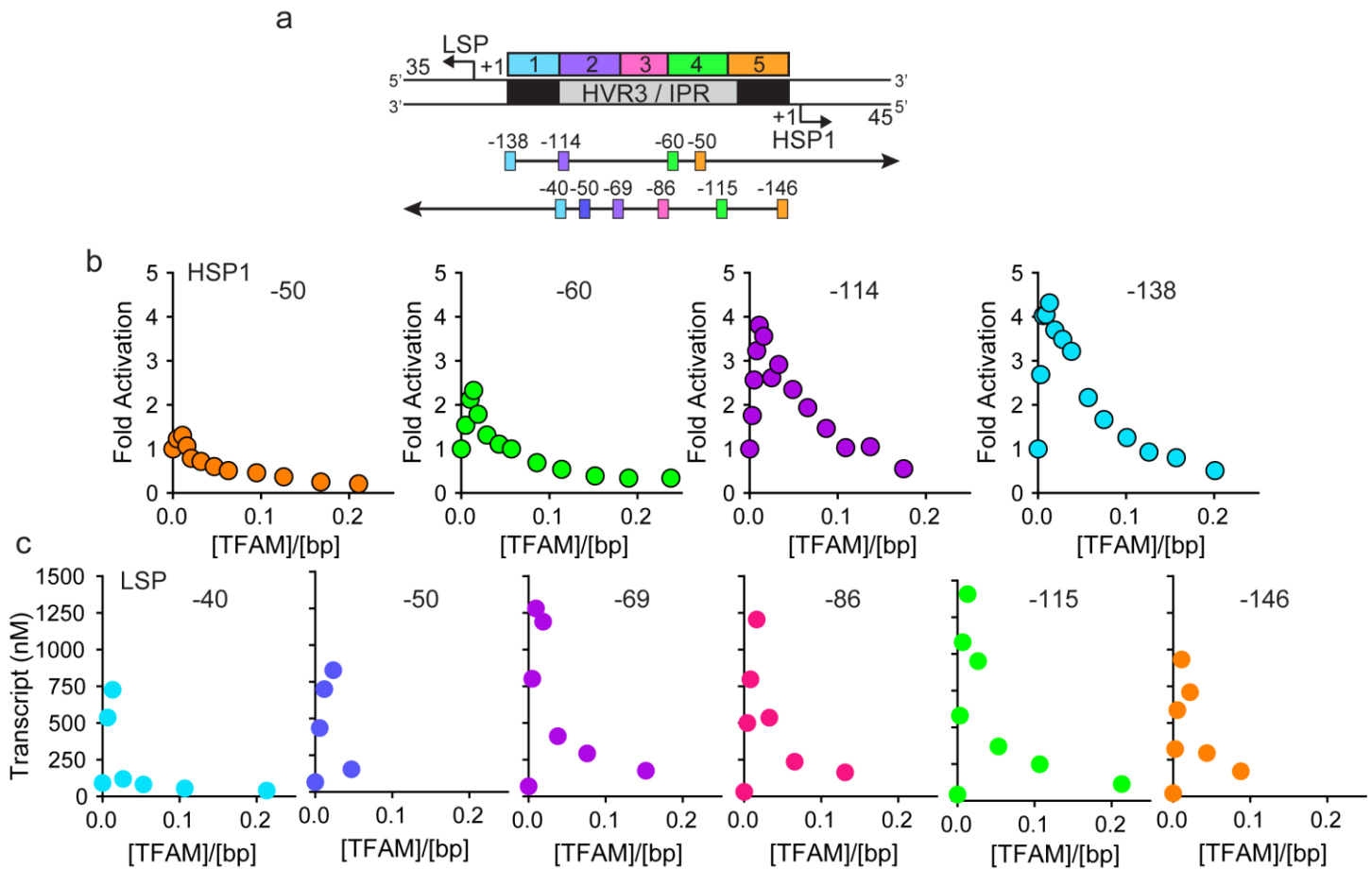


Figure 2. HVR3/IPR contributes to transcription from HSP1 but not from LSP. (a) Deletion constructs used to assess the impact of HVR3/IPR on LSP and HSP1 transcription. The section between and including the putative TFAM binding sites was divided into 5 distinct regions (1-5; 1:425-447; 2:448-476; 3:447-493; 4:494-522; 5:523-550) and used to guide the deletion design strategy. Four (-50, -60, -114 and -138) and six (-40, -50, -69, -86, -115 and -146) different deletion constructs were used to assess HSP1 and LSP transcription, respectively. (b) Fold activation on HSP1 transcription plotted as a function of TFAM concentration per bp for the different HSP1 deletion constructs. (c) Amount of LSP transcript product plotted as a function of TFAM concentration per bp for the different LSP deletion constructs.

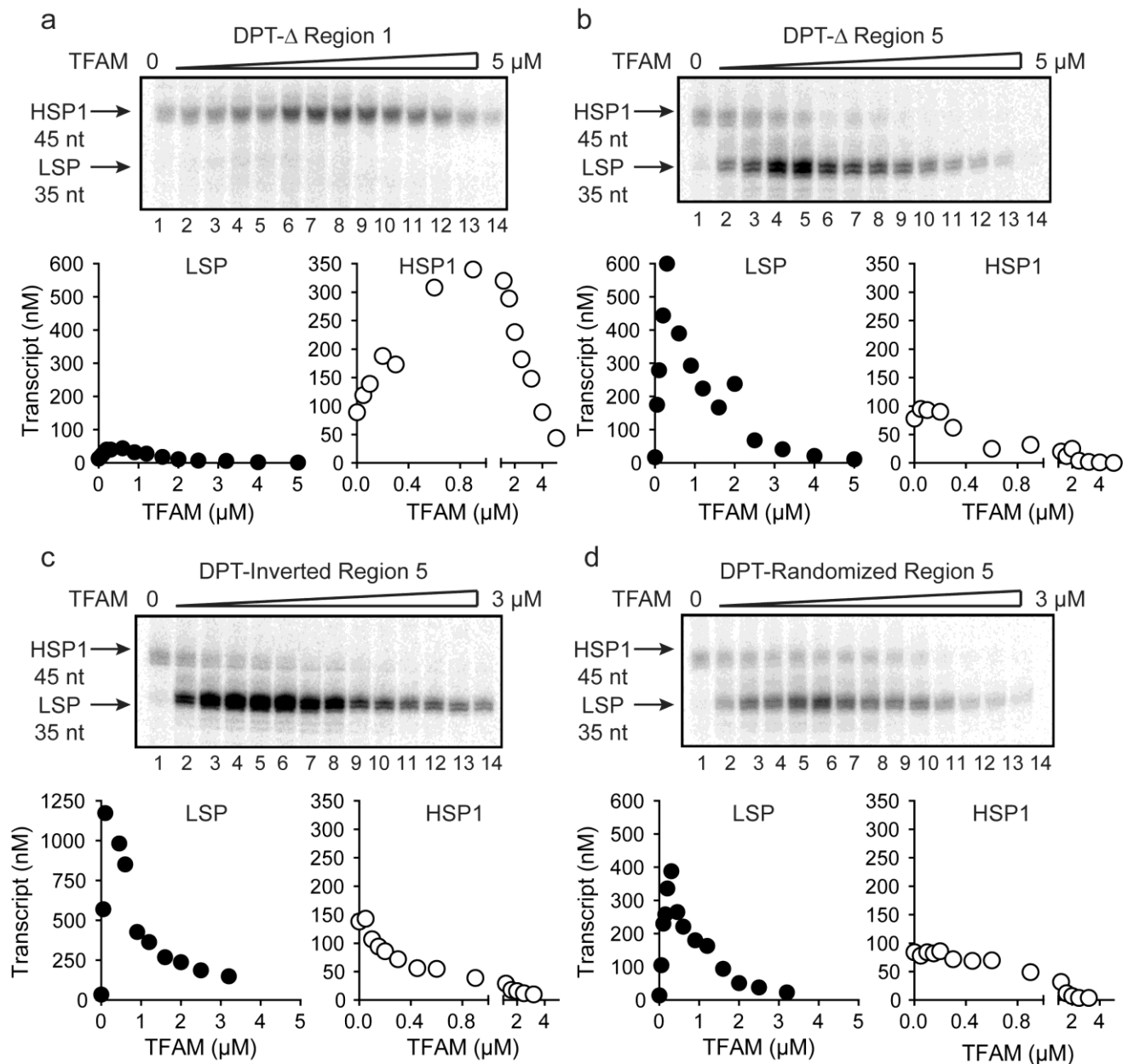


Figure 3. TFAM-binding site of record for HSP1 may not actually bind TFAM to contribute to HSP1 transcription. (a) Deletion of LSP TFAM-binding site, Δ Region 1, from dual promoter template precludes transcription from LSP. (b) Deletion of HSP1 TFAM-binding site, Δ Region 5, from dual promoter template does not interfere with TFAM-independent activity but converts TFAM from an activator to an inhibitor. (c,d) Inversion and randomization of region 5 also converts TFAM into an inhibitor. Shown are the run-off transcription products of LSP and HSP1 promoter-dependent transcription resolved by denaturing PAGE and the amount of transcription product plotted as a function of TFAM for the Δ Region 1, Δ Region 5, Inverted Region 5 and Randomized Region 5 constructs.

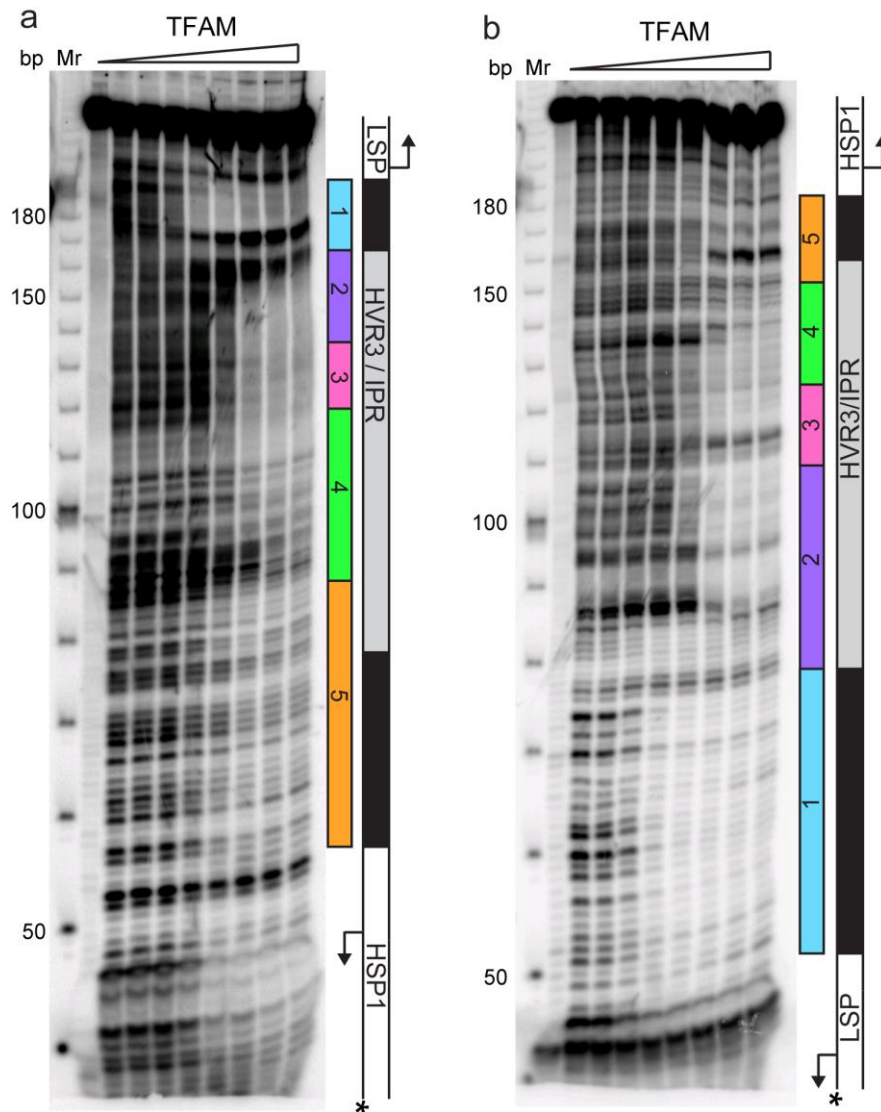


Figure 4. Footprinting of the dual promoter template in the presence of TFAM confirms protection or sensitivity in regions 2 and 4 before region 5. DNase I footprinting of the dual promoter template with increasing concentrations of TFAM. (a) HSP1 template strand ^{32}P -labeled. (b) LSP template strand ^{32}P -labeled. A schematic of the transcriptional control region is shown to the right of the denaturing PAGE gels to indicate regions of protection.

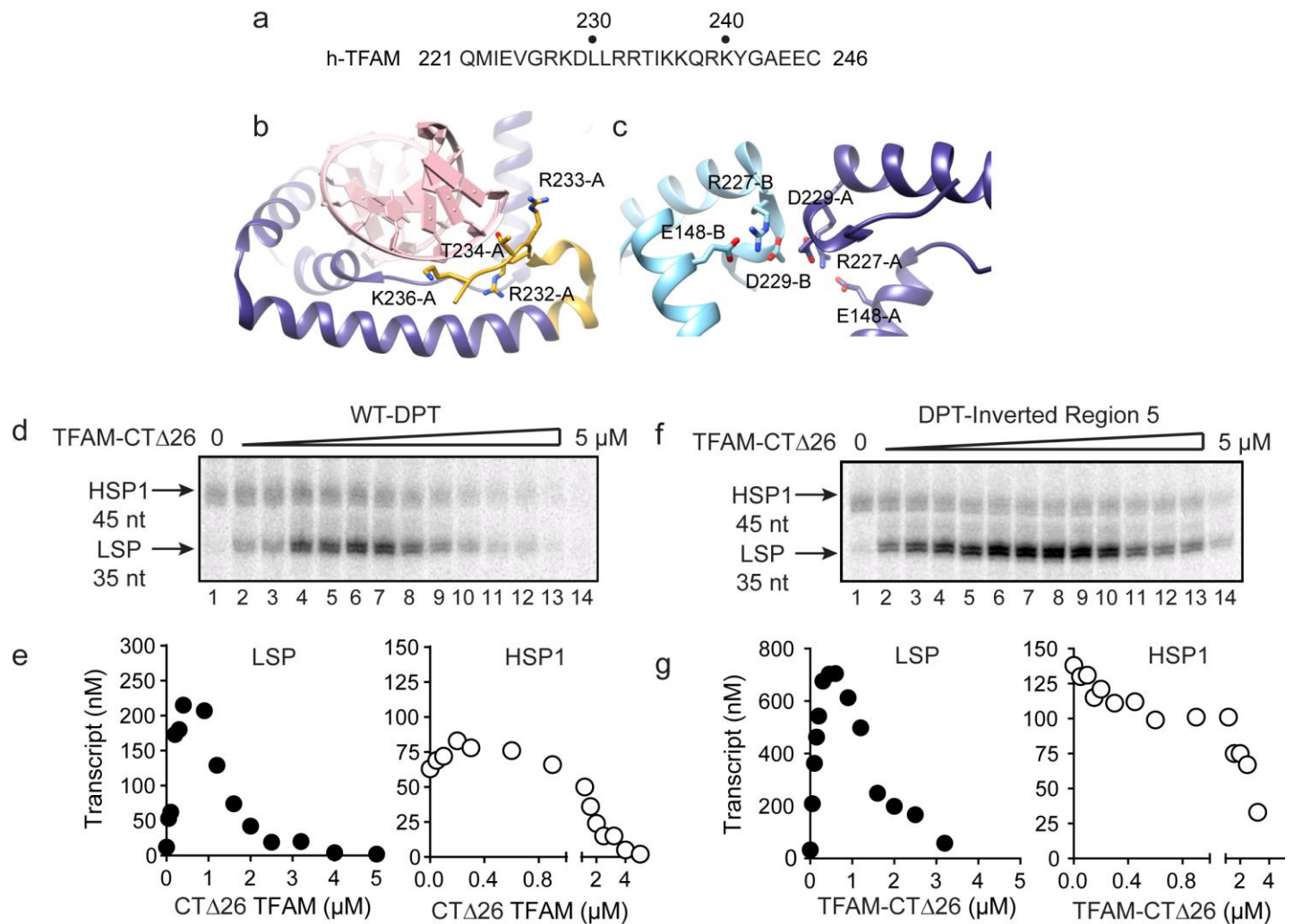


Figure 5. Carboxy-terminal tail of TFAM is essential for transcription from HSP1. (a) Carboxy-terminal tail (CTT) primary sequence of human TFAM. The last 26 amino acid residues are shown. (b,c) Interactions of TFAM CTT. Structural models were produced using PDB 3TQ6 [ref]. In panel b, TFAM residues 221-236 are colored yellow. Residues 237 to 246 are disordered and absent in the structure. Residues 232-236 interact with the phosphodiester backbone of bound DNA (red). In panel c, two TFAM-DNA complexes are present in the asymmetric unit and designated here as chain A (dark blue) and chain B (light blue). Structural integrity of the CTT of each monomer benefits from interaction of Arg-227 in each monomer with both Asp-229 and Glu148 of the same monomer. The CTT of one monomer packs against that of a second, perhaps creating a mechanism for association between TFAM-DNA complexes. (d) Run-off transcription products using the dual promoter DNA oligo template and increasing concentrations of TFAM-CT Δ 26 (0-5 μ M) resolved by denaturing PAGE. (e) Amount of LSP (35 nt) and HSP1 (45 nt) RNA transcripts produced using the dual promoter construct plotted as a function of TFAM-CT Δ 26 concentration. (f) Run-off transcription products using the dual promoter DNA oligo template with region 5 inverted and increasing concentrations of TFAM-CT Δ 26 (0-5 μ M) resolved by denaturing PAGE. (g) Amount of LSP (35 nt) and HSP1 (45 nt) RNA transcripts produced using the dual promoter construct with region 5 inverted plotted as a function of TFAM-CT Δ 26 concentration.

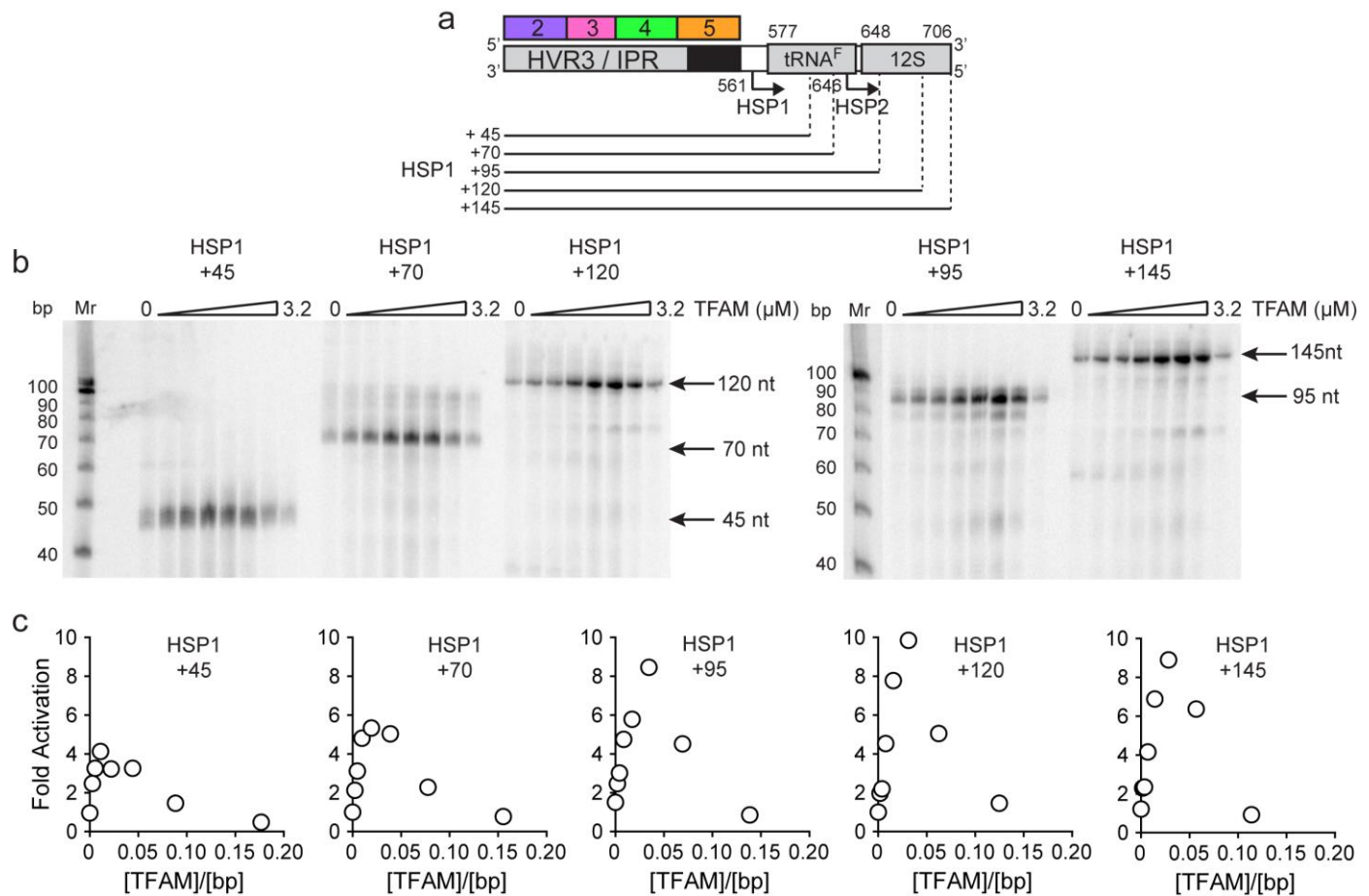
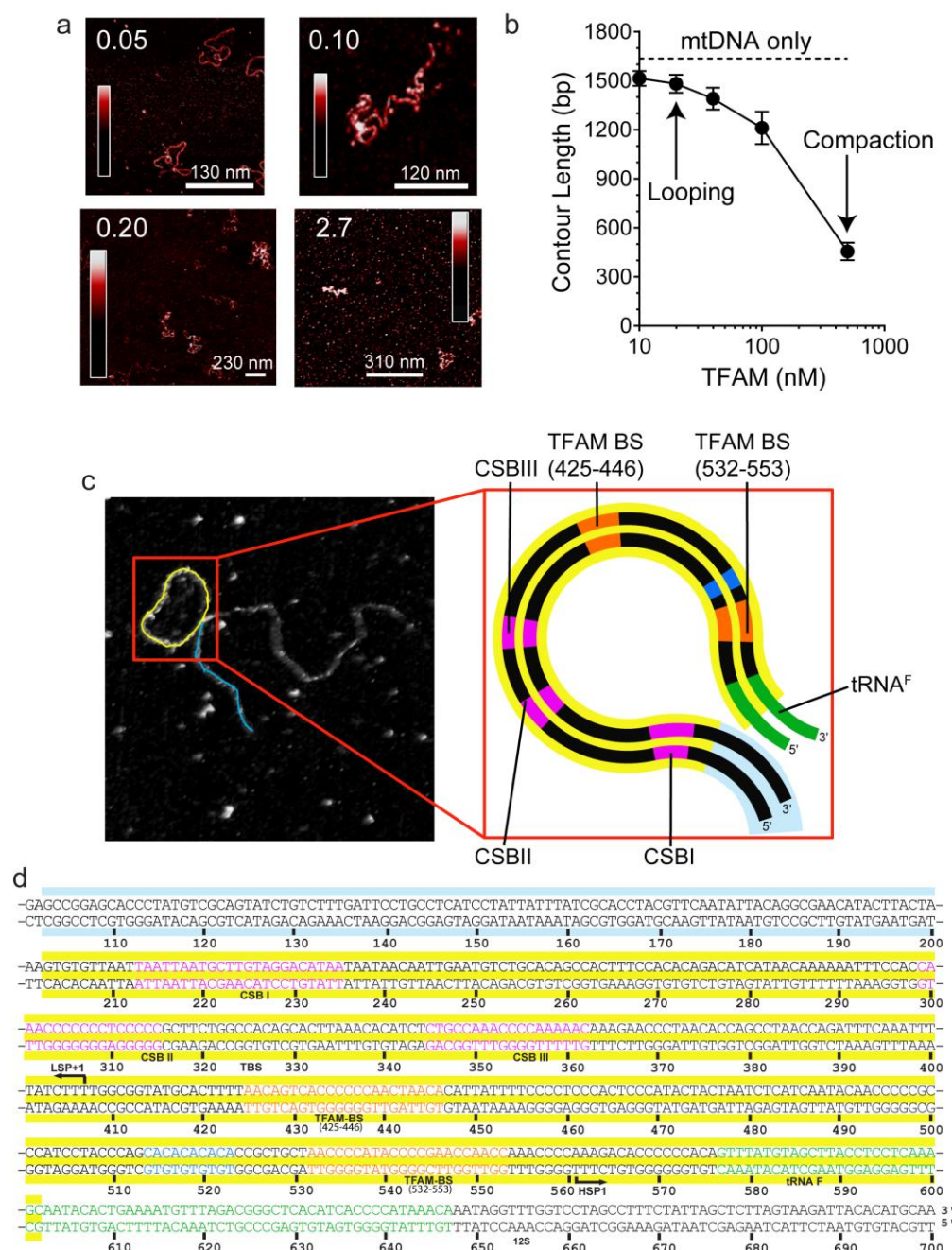


Figure 6. mtDNA sequences downstream of HSP1 increase the fold activation on HSP1 transcription. (a) Schematic of HSP1 dsDNA oligonucleotide templates used for in vitro transcription reactions which contained additional mtDNA sequences, either 45, 79, 95, 120 or 145 bp, downstream of the HSP1 promoter start site. Each template contained the HVR3/IPR, starting from Region 2, and ending at the indicated bp from the HSP1 start site. The dotted lines show the approximate position each DNA oligo template end relative to the tRNA^F and 12S RNA genes. (b) Run-off transcription products using HSP1 DNA oligo templates and increasing concentrations of TFAM (0-5 μ M) resolved by denaturing PAGE. Ten-bp markers are indicated on each gel. (c) Fold activation on HSP1 transcription plotted as a function of TFAM concentration using the different HSP1 DNA oligo templates.



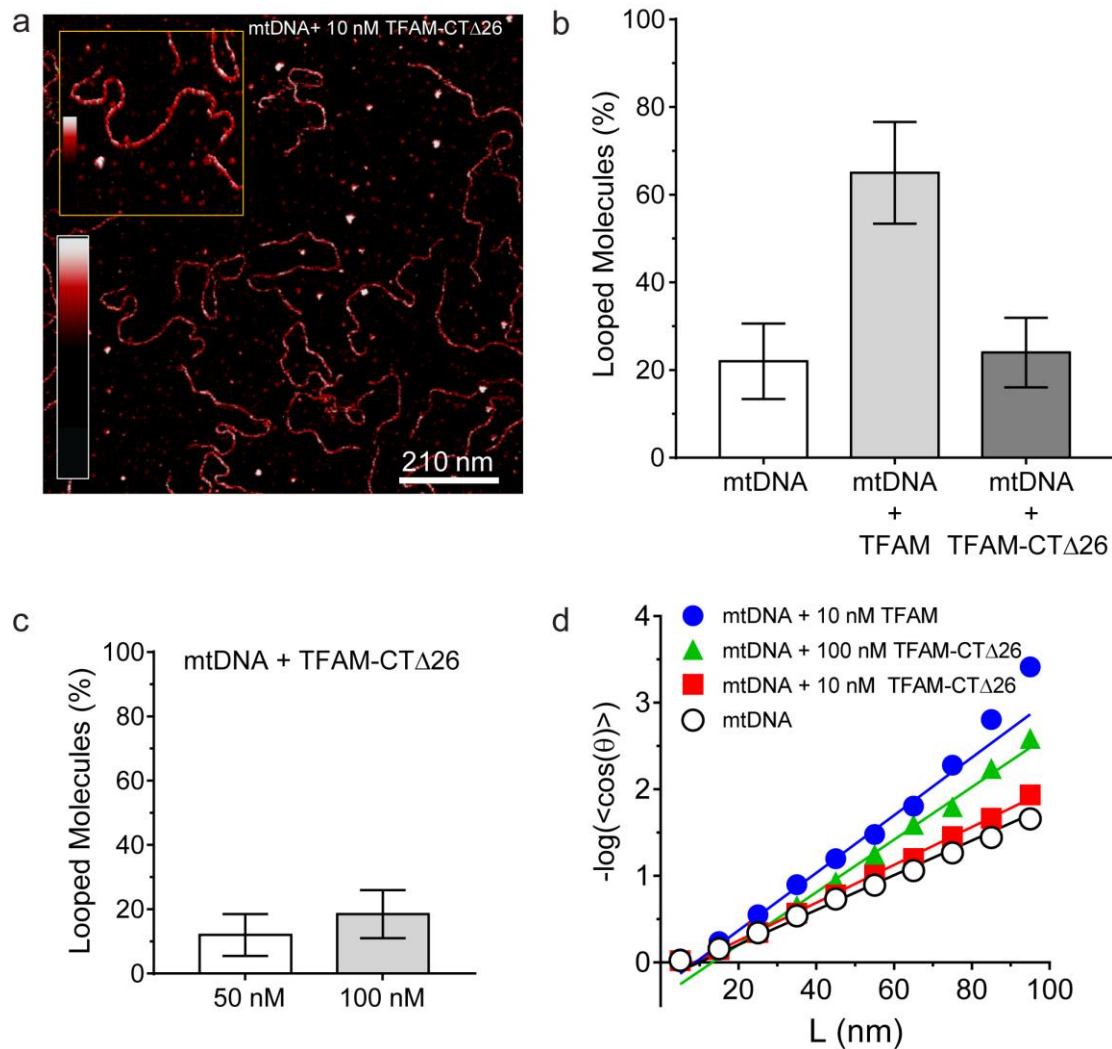


Figure 8. DNA looping is enhanced in the presence of TFAM, but not in the presence of TFAM-CT Δ 26, even when strongly bound. (a) Two-dimensional representation of mtDNA bound to TFAM-CT Δ 26. The inset shows a three-dimensional representation of a selected region from the same image. (b) The bar graph shows the percentage of looped molecules that were observed for mtDNA constructs in the absence and presence of 10 nM TFAM or TFAM-CT Δ 26. (c) Percentage of looped mtDNA molecules in the presence of 50 nM and 100 nM TFAM-CT Δ 26, showing that even when strongly bound, TFAM-CT Δ 26 does not increase DNA looping. (d) Fits to the 3D wormlike chain model for mtDNA construct in the absence and presence of TFAM or TFAM-CT Δ 26, showing 100 nM TFAM-CT Δ 26 is equivalent in DNA binding to 10 nM TFAM.

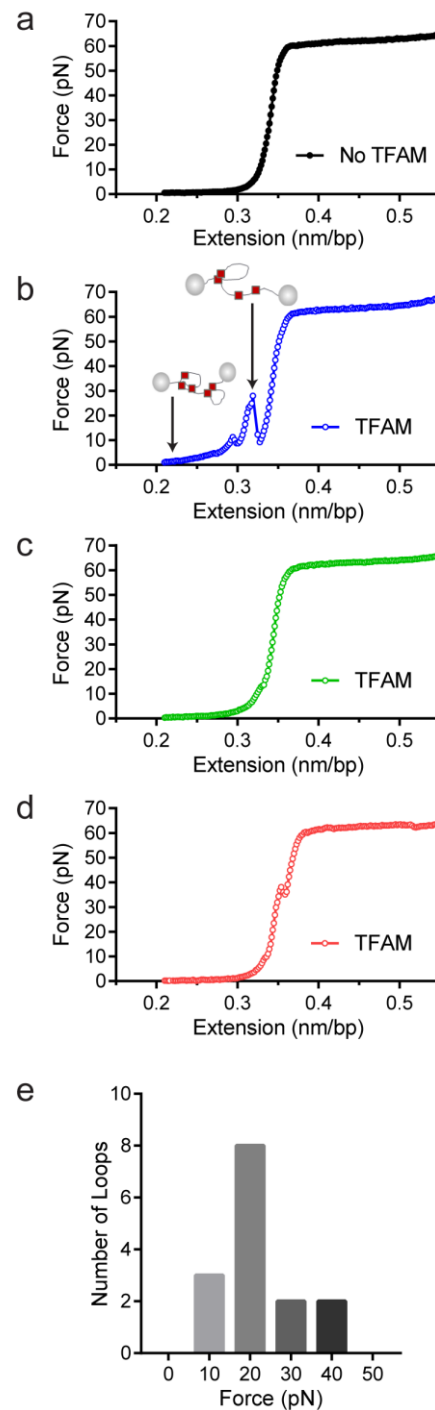


Figure 9. Optical tweezers data confirm DNA looping by TFAM. (a) Force-extension curve of bacteriophage- λ molecule of 48,500 base pairs (bp) in the absence of TFAM. (b) Initial force-extension curve of bacteriophage- λ molecule in the presence of 50 nM TFAM (blue open circles). When held at low extension loops are mediated in the presence of TFAM and as the molecule of DNA is extended we observe jumps revealing the breaking of a loop previously formed. The cartoon inset illustrates the formation of loops mediated by TFAM and the breaking of loops as the DNA is extended. (c,d) Consecutive force-extension curves in the presence of TFAM are shown in green (panel c, extended immediately after the initial extension shown in panel b) and red (panel d, extended after waiting 7 minutes). (e) The histogram illustrates the forces involved in breaking the loops mediated by TFAM. The most probable loop breaking force is 20 pN.

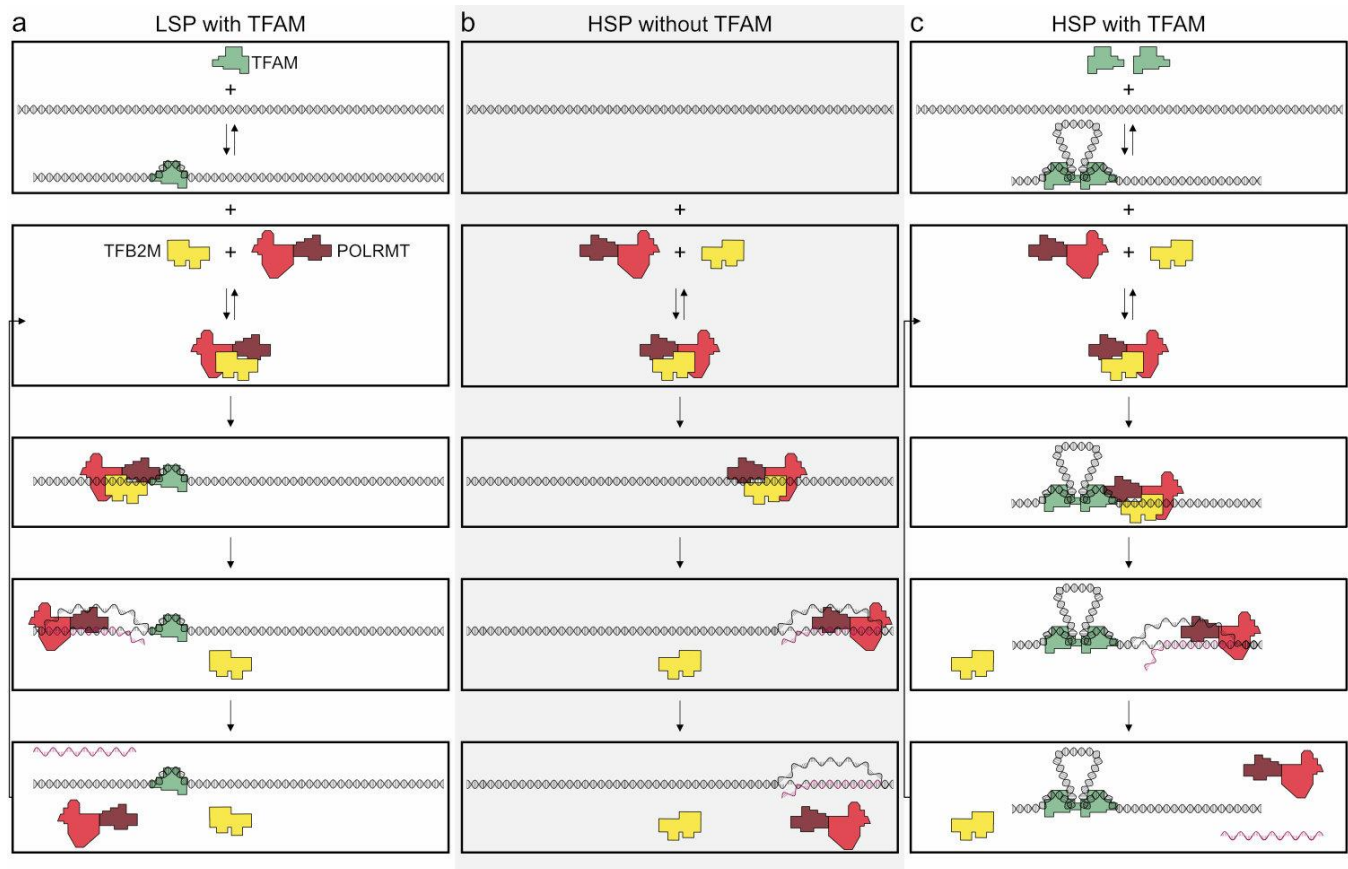


Figure 10. Regulation of mitochondrial transcription by TFAM. (a) Initiation of transcription at light-strand promoter (LSP) requires TFAM. A TFAM monomer binds to a specific site upstream of the LSP transcription start site (TSS) and bends the DNA. POLRMT and TFB2M associate and add to LSP, perhaps directed by TFAM. Initiation requires all three components at this promoter; elongation only requires POLRMT. Elongation to the end of template leads to dissociation of POLRMT and RNA product from template, thus enabling another round of transcription. (b) Initiation of transcription at heavy-strand promoter 1 (HSP) does not require TFAM. POLRMT and TFB2M associate and are sufficient to recognize and bind to HSP. Initiation requires only these two components, and elongation requires only POLRMT. Elongation to the end of template leads to dissociation of POLRMT, but additional rounds of transcription are not supported, perhaps because RNA product remains hybridized to template. (c) Reutilization of HSP requires TFAM. Binding of a TFAM dimer to the inter-promoter region creates loops of the DNA. Formation of the loops requires the carboxy-terminal tail of TFAM. POLRMT and TFB2M associate and add to TFAM-bound HSP. Initiation and elongation occur as described above; except TFAM facilitates multiple turnovers from HSP. Looping is required for this function as deletion of the carboxy-terminal tail precludes TFAM activation at HSP.

SI Figure Legends

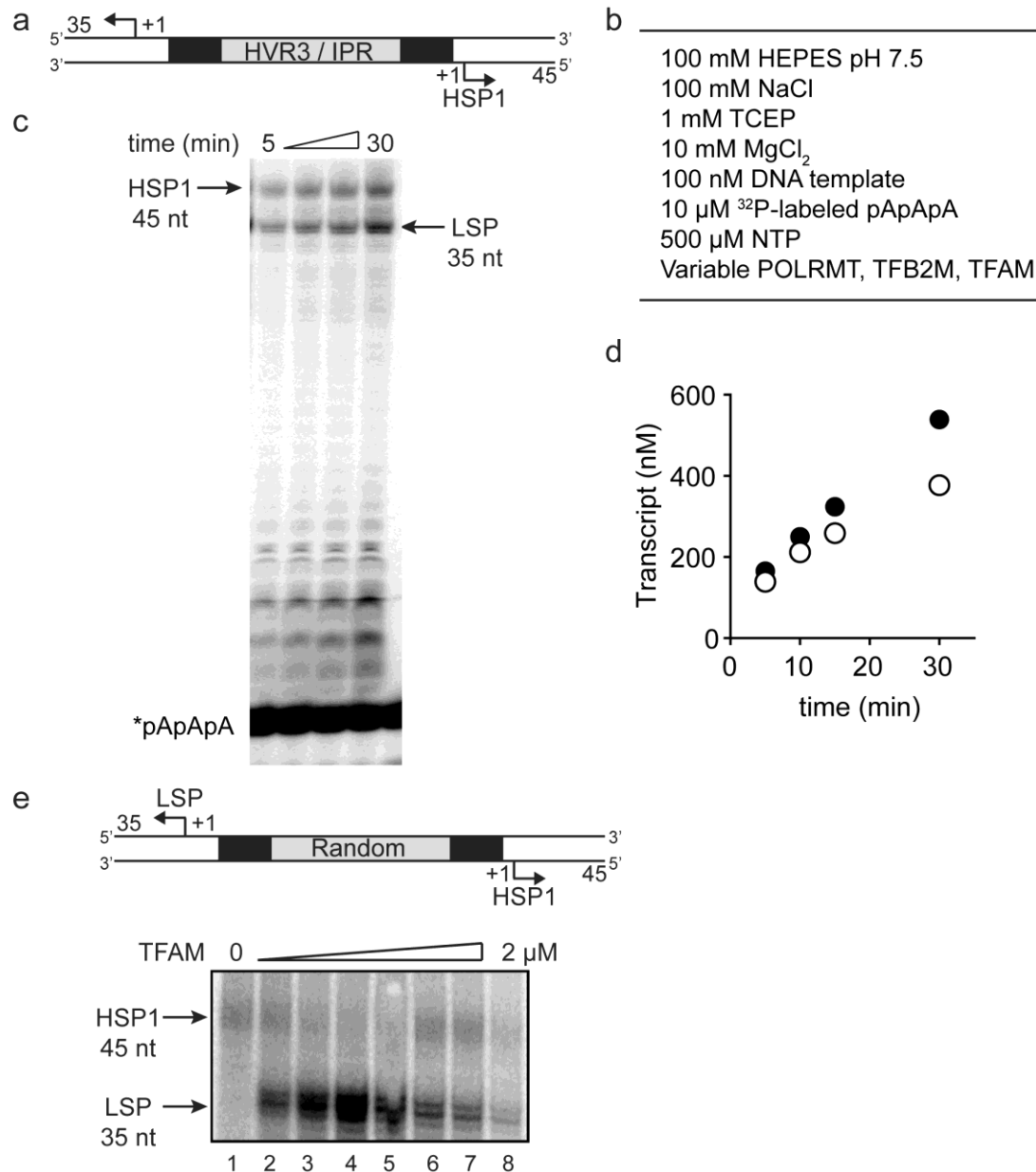


Figure S1. Dual promoter construct: Time course and reaction conditions. (a) Dual promoter DNA oligonucleotide (234 bp) template containing LSP, HSP1 and HVR3/IPR used for in vitro transcription reactions. This oligo gives rise to LSP- and HSP1-derived RNA transcripts 35 and 45 nt long, respectively. (b) In vitro reaction conditions employed for mitochondrial transcription assays. (c) Run-off transcription products using the dual promoter DNA oligo template as a function of time resolved by denaturing PAGE. Concentrations of mitochondrial transcription proteins were 600 nM TFAM, 1000 nM TFB2M and 1000 nM POLRMT. (d) Amount of LSP (35 nt) and HSP1 (45 nt) RNA transcripts produced using the dual promoter construct plotted as a function of time. Time courses were linear over 30 min. (e) Dual promoter DNA oligonucleotide template with the IPR sequence randomized. Run-off transcription products using the IPR randomized dual promoter DNA oligo template and increasing concentrations of TFAM (0-2 μM) resolved by denaturing PAGE.

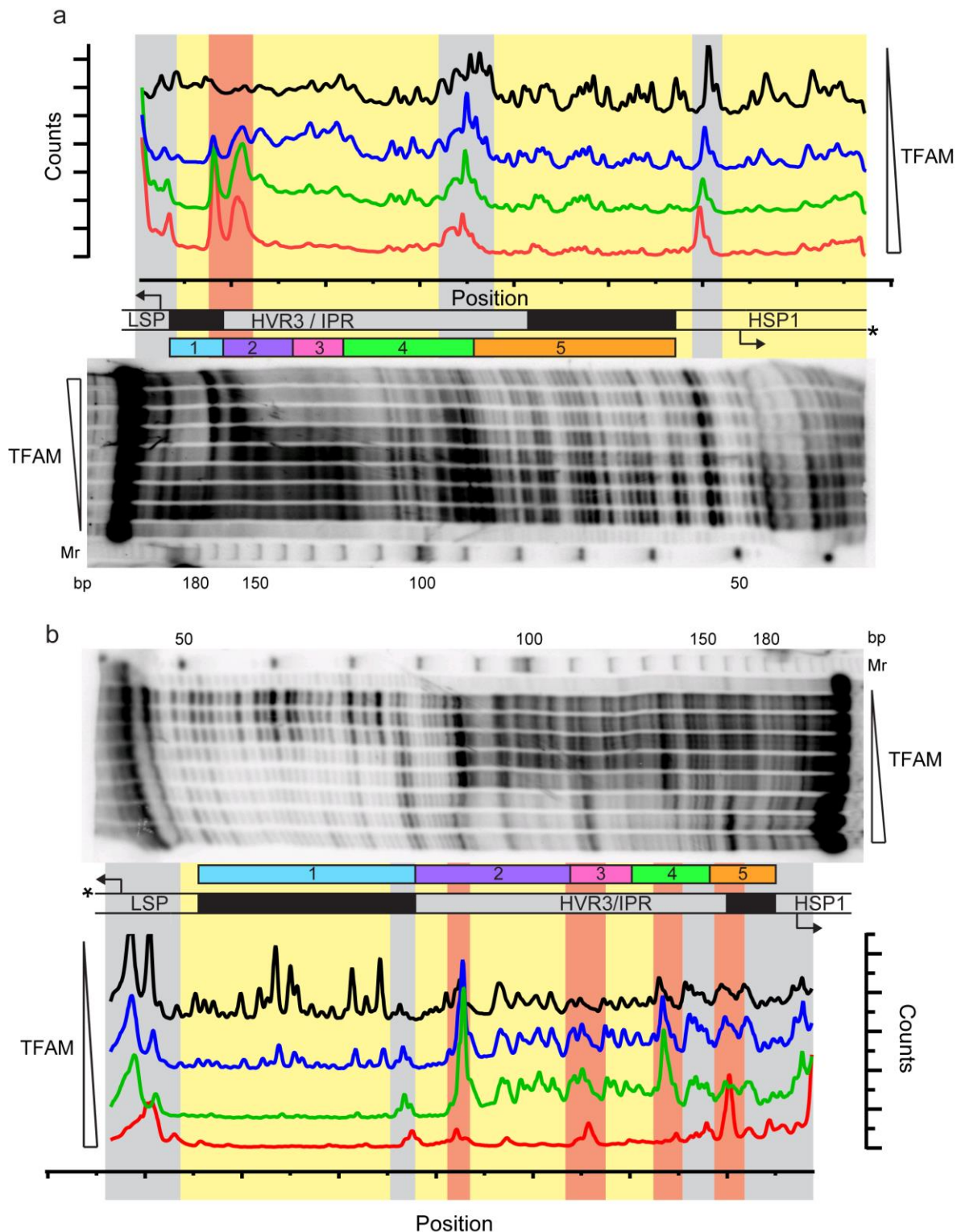


Figure S2. Footprinting of the dual promoter template in the presence of TFAM confirms protection or sensitivity in regions 2 and 4 before region 5. DNase I footprinting of the dual promoter template with increasing concentrations of TFAM. (a) HSP1 template strand ^{32}P -labeled. (b) LSP template strand ^{32}P -labeled. A schematic of the transcriptional control region is shown above or below the denaturing PAGE gels to indicate regions of protection. The graph indicates the cleavage and protection pattern across the dsDNA template as quantified using ImageQuant TL software. The yellow shaded areas indicate regions of protection, the red shaded areas indicate regions of enhanced cleavage and the grey shaded areas are regions essentially unchanged.

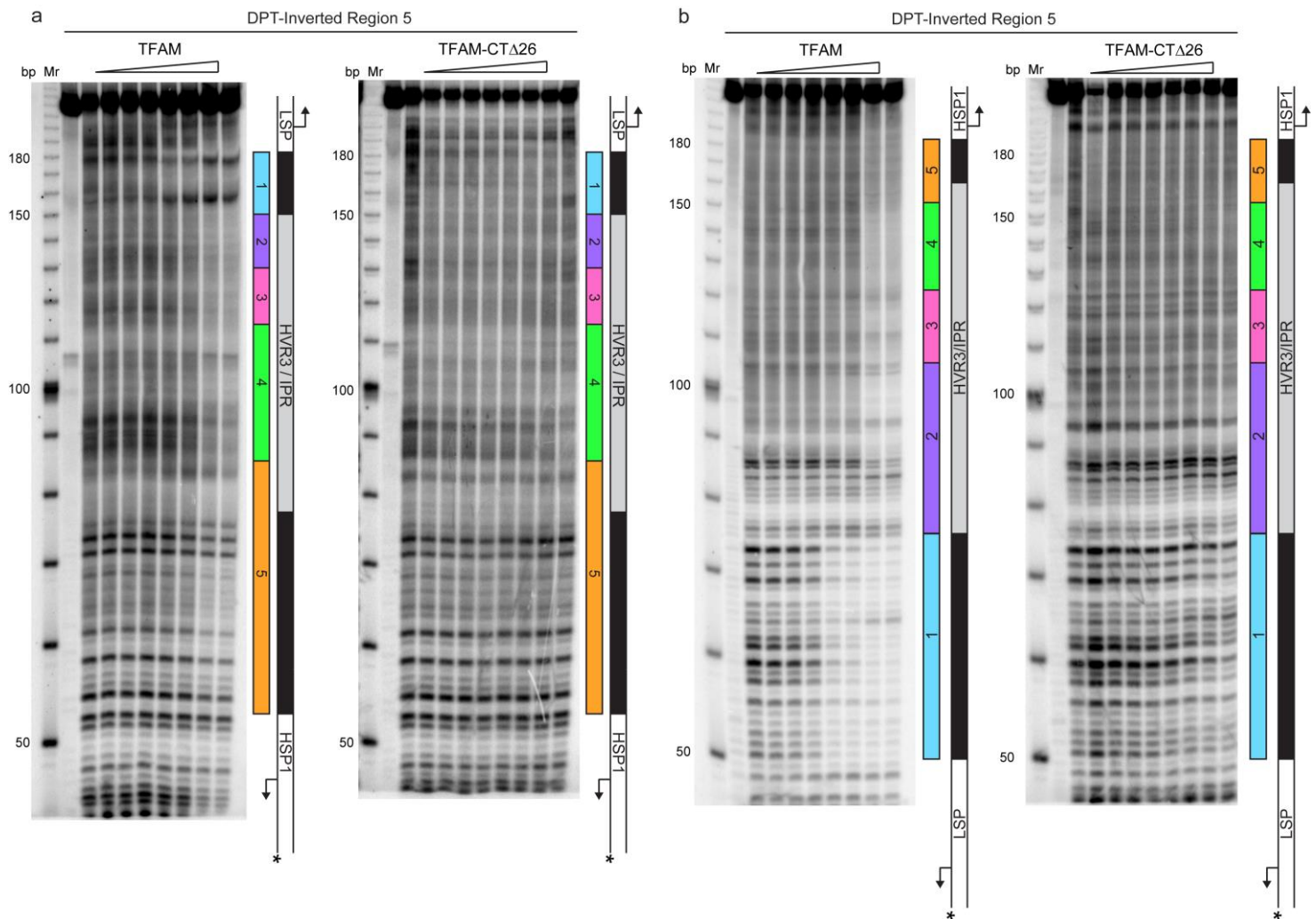


Figure S3. Footprinting of the dual promoter template with Inverted Region 5 in the presence of WT and TFAM-CTΔ26. DNase I footprinting of the dual promoter template with Inverted Region 5 with increasing concentrations of WT TFAM and TFAM-CTΔ26. (a) HSP1 template strand ³²P-labeled. (b) LSP1 template strand ³²P-labeled. A schematic of the transcriptional control region is shown to the right of each denaturing PAGE gel to indicate regions of protection.

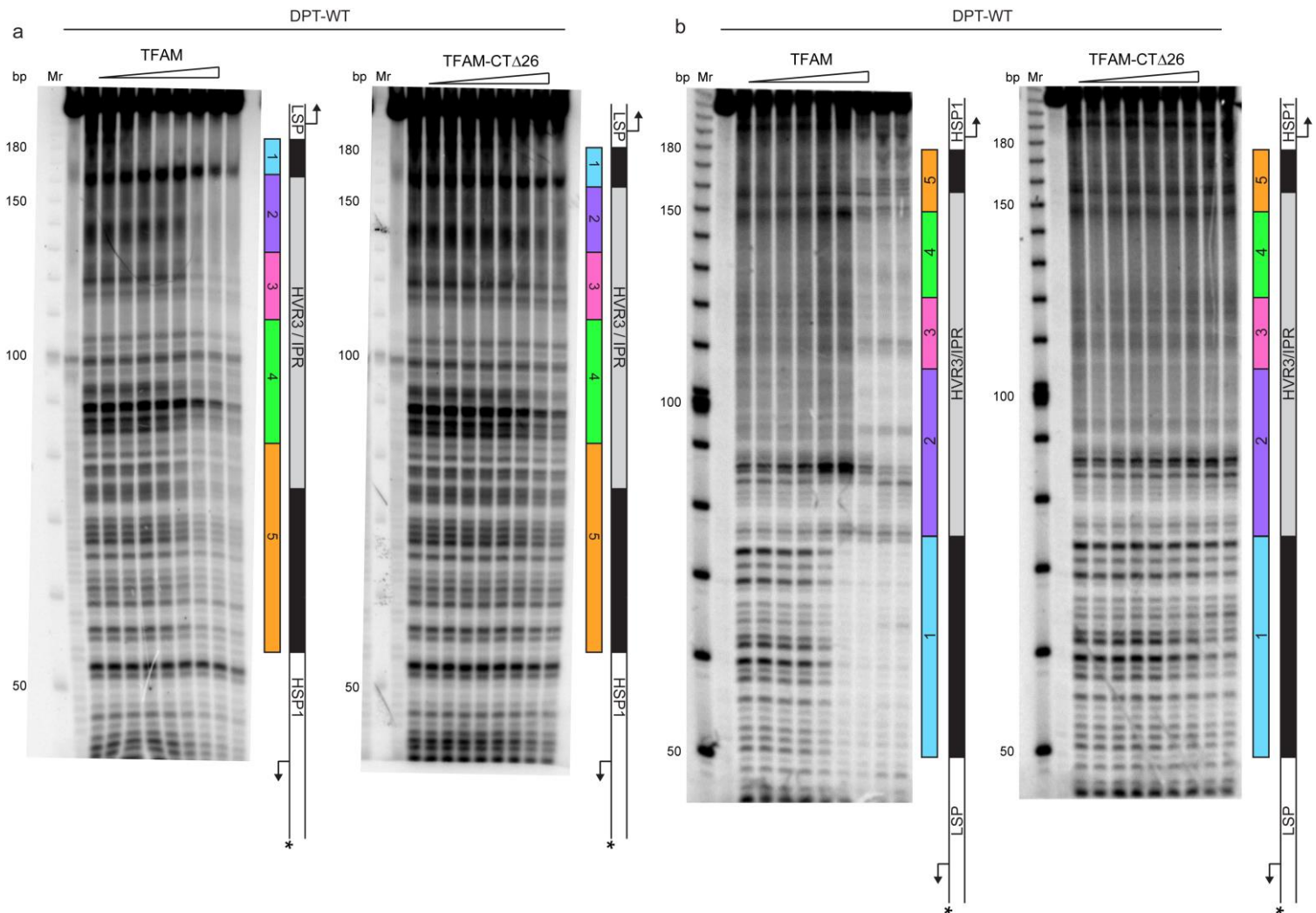


Figure S4. Footprinting of the dual promoter template in the presence of either WT and TFAM-CT Δ 26 with POLRMT and TFB2M. DNase I footprinting of the dual promoter template with increasing concentrations of WT and TFAM-CT Δ 26 with POLRMT and TFB2M. In both cases the HSP template strand is 32 P-labeled. **(a)** HSP1 template strand 32 P-labeled. **(b)** LSP1 template strand 32 P-labeled. A schematic of the transcriptional control region is shown to the right of each denaturing PAGE gel to indicate regions of protection.

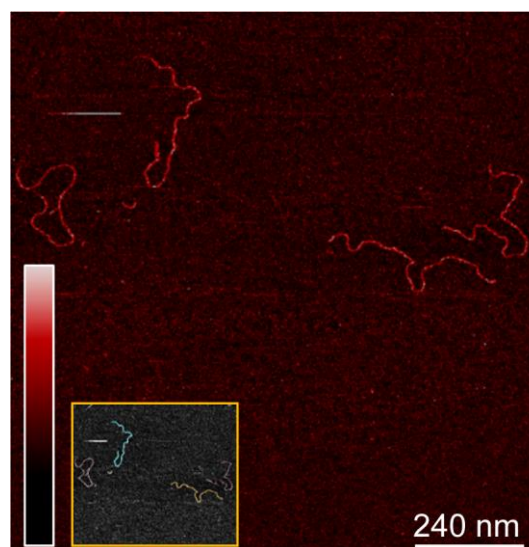


Figure S5. Imaging of mtDNA using AFM in liquid. Two-dimensional representation of mtDNA in the absence of protein. The inset shows traces along the contour of the DNA, which is used to calculate the DNA persistence length (Figure 8).

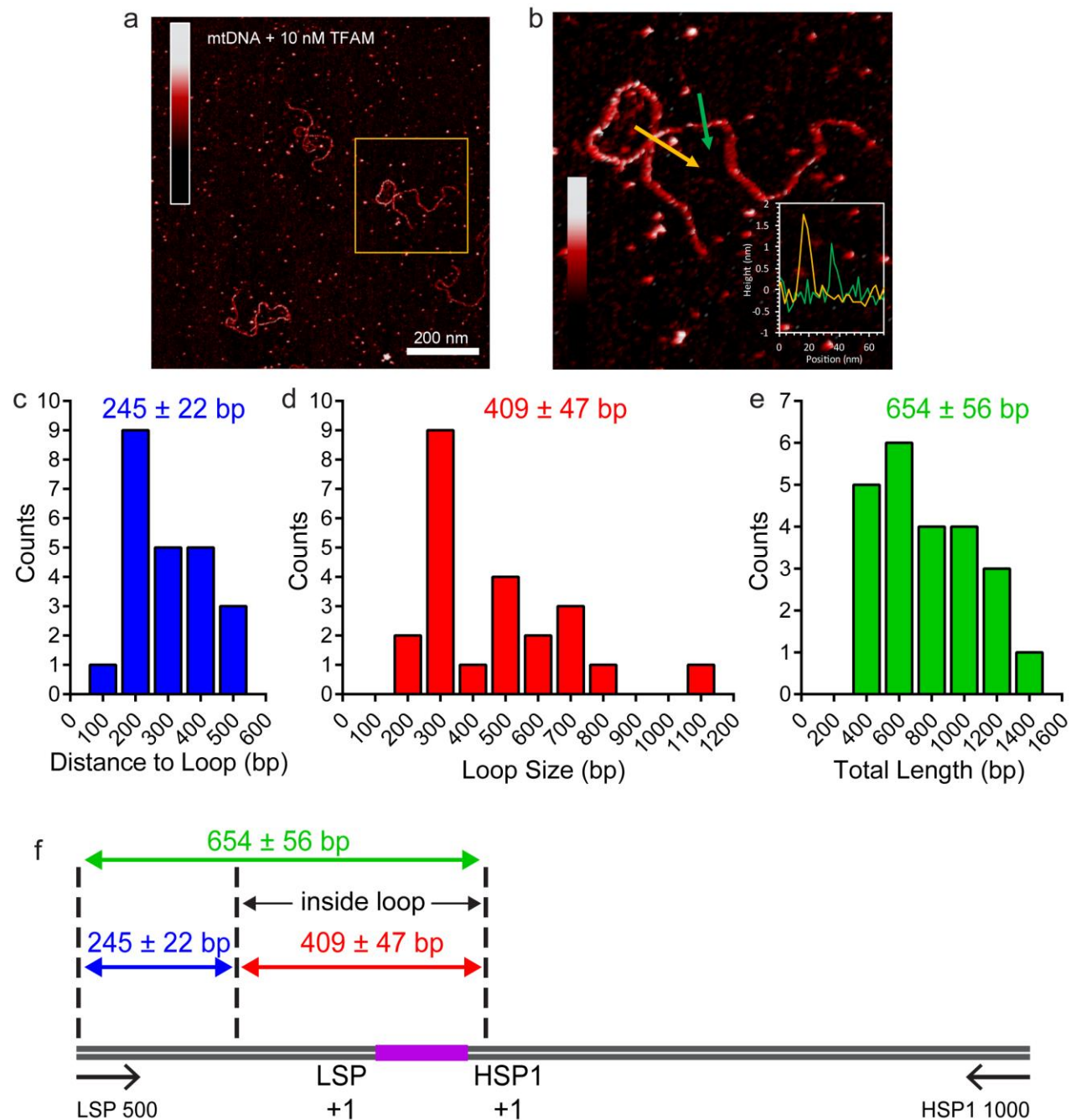


Figure S6. TFAM-induced looping characterized by AFM in liquid. (a) Two-dimensional representation of loops and bridges mediated by TFAM bound to mtDNA. The concentration used is 10 nM TFAM and 0.11 nM mtDNA. (b) Three-dimensional representation of looped mtDNA mediated by TFAM, an example from the yellow box in (a). (c) Histogram of the short distance to the loop, illustrated in blue in Fig. 7c. The average distance to the loop is 245 ± 22 bp. (d) Histogram of loop sizes in base pairs. The average loop size is 409 ± 47 bp. An example loop is illustrated in yellow in Fig. 7c. (e) Histogram showing the sum of the distance to loop and loop size. The average distance is 654 ± 56 bp. (f) The mtDNA construct, with a 153 bp IPR region (purple), shows that the average location of the IPR region is inside the loop.

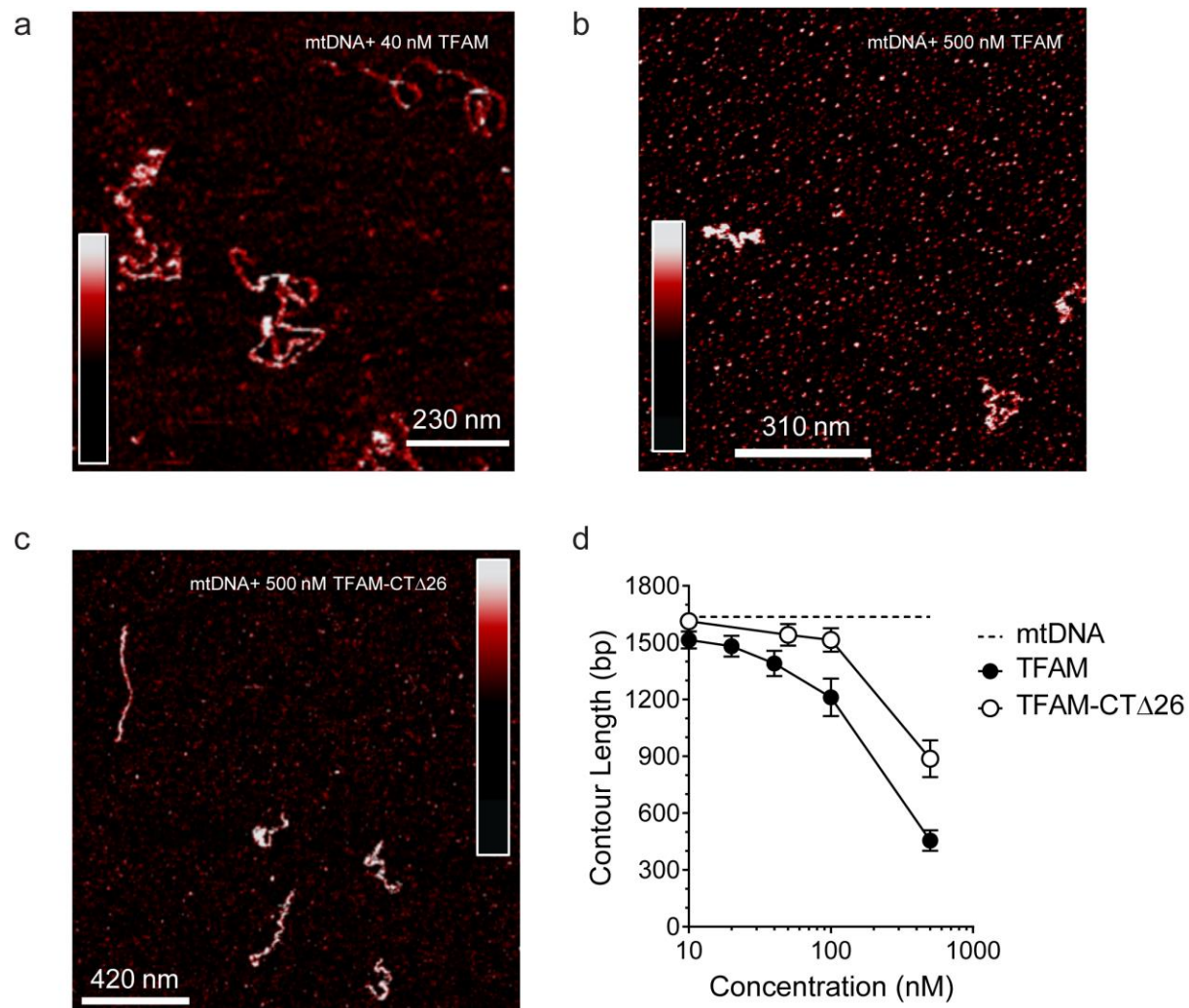


Figure S7. Both TFAM and TFAM-CT Δ 26 compact DNA at high concentrations. (a) Two-dimensional representation of 40 nM TFAM bound to 0.11 nM mtDNA. The color bar indicates the height of the sample ranging from -1.5 to 2.0 nm. (b) Two-dimensional representation of 500 nM TFAM bound to 0.11 nM mtDNA. The color bar indicates the height of the sample ranging from -2.1 to 3.0 nm. (c) Two-dimensional representation of 500 nM TFAM-CT Δ 26 bound to 0.11 nM mtDNA. The color bar indicates the height of the sample ranging from -1.5 to 2.0 nm. (d) Contour length of the mtDNA as a function of concentration of TFAM and TFAM-CT Δ 26. The black dotted line represents the contour length of the mtDNA only. The open circles (with connecting black line) represent the measured contour length of the mtDNA in the presence of TFAM-CT Δ 26. The filled black circles (with connecting black line) represent the measured contour length of the mtDNA in the presence of TFAM.

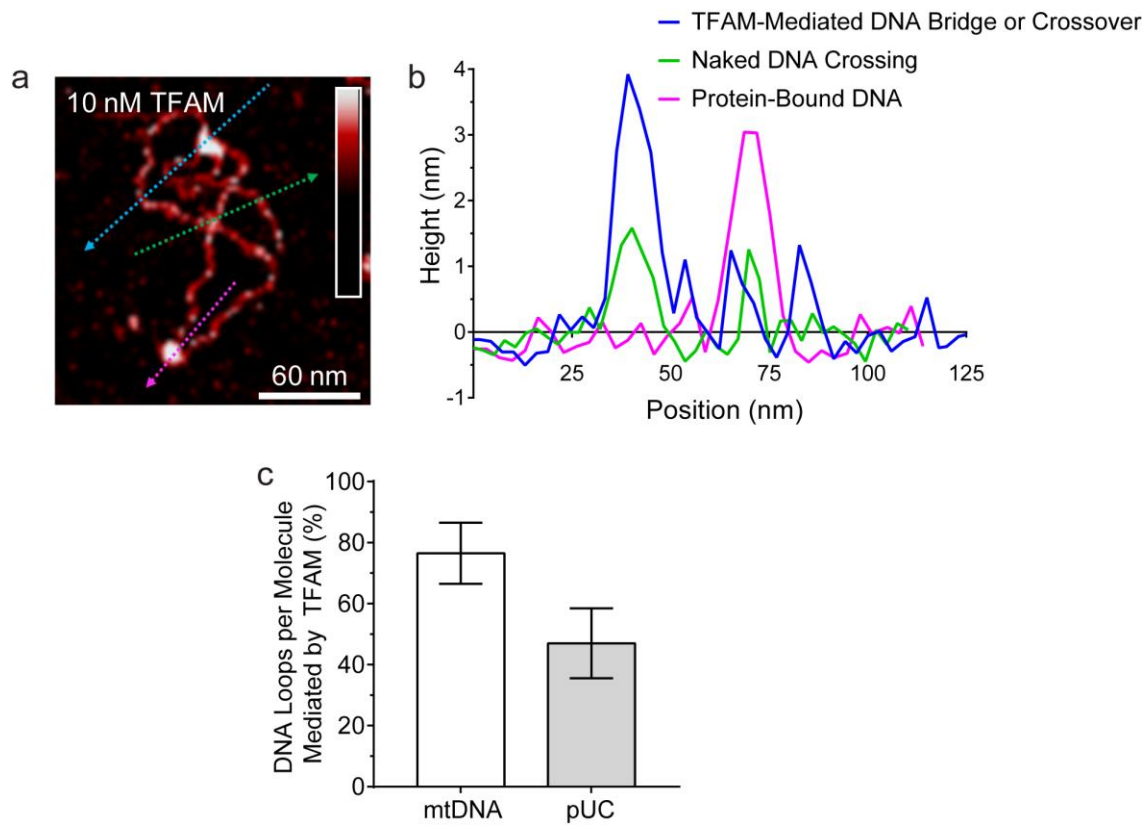


Figure S8. Protein-mediated DNA looping by TFAM is sequence-specific. (a) Two-dimensional representation of DNA bridges and loops mediated by TFAM. We imaged 16 molecules at 10 nM TFAM and counted the number of loops. The lines represent the locations of cross-section presented in panel (b). Three types of crossovers or bridges can be identified: protein-bound DNA (purple arrow); TFAM-mediated DNA bridge or crossover (blue arrow); and naked DNA crossover (green arrow). (b) Height of the cross-sections on the image (a) as color coded. (c) Percent of measured DNA loops that are mediated by TFAM as determined from the cross-sections, illustrated in (b). The mtDNA construct is much more likely to have protein bound at the crossover point.

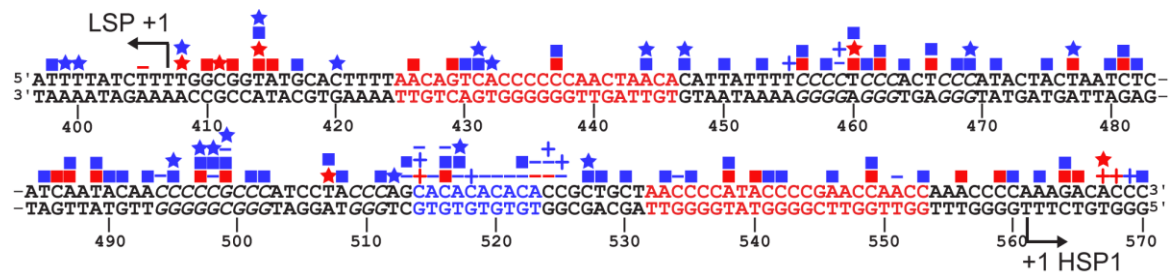


Figure S9. SNPs and somatic mutations associated with disease in LSP, HSP1 and HVR3. The mtDNA sequence of LSP, HSP1 and HVR3/IPR is shown. SNPs and mutations associated with disease are shown as blue and red symbols, respectively; deletions (-); insertions (+), transitions (■); transversions (★). Location of putative TFAM-binding sites are colored red. CA repeats implicated in disease are colored blue. Possible TFAM binding sites in HVR3/IPR are italicized.