

# The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA

Huu B Ngo<sup>1</sup>, Jens T Kaiser<sup>2</sup> & David C Chan<sup>1,3</sup>

**Tfam (transcription factor A, mitochondrial), a DNA-binding protein with tandem high-mobility group (HMG)-box domains, has a central role in the expression, maintenance and organization of the mitochondrial genome. It activates transcription from mitochondrial promoters and organizes the mitochondrial genome into nucleoids. Using X-ray crystallography, we show that human Tfam forces promoter DNA to undergo a U-turn, reversing the direction of the DNA helix. Each HMG-box domain wedges into the DNA minor groove to generate two kinks on one face of the DNA. On the opposite face, a positively charged  $\alpha$ -helix serves as a platform to facilitate DNA bending. The structural principles underlying DNA bending converge with those of the unrelated HU family proteins, which have analogous architectural roles in organizing bacterial nucleoids. The functional importance of this extreme DNA bending is promoter specific and seems to be related to the orientation of Tfam on the promoters.**

Among eukaryotic protein complexes, the mitochondrial oxidative phosphorylation (OXPHOS) machinery is unique in having a bigenomic origin. Most of the OXPHOS machinery is encoded by the nuclear genome, but 13 essential subunits of respiratory chain complexes I, III, IV and V are encoded by the 16-kb mitochondrial genome. Tfam (also known as mtTFA), a DNA-binding protein in mitochondria, is a central player in expression and maintenance of mitochondrial DNA (mtDNA), and therefore is essential for ATP production via OXPHOS<sup>1,2</sup>. The mammalian mitochondrial genome contains three promoters—the light strand promoter (LSP), the heavy strand promoter 1 (HSP1) and the heavy strand promoter 2 (HSP2)—that drive expression of mtDNA transcripts. Transcription from LSP and HSP1 has been reconstituted *in vitro*, and normal levels of transcription require Tfam<sup>1–4</sup>. Moreover, because truncated RNA transcripts from LSP are used to prime DNA replication, Tfam is secondarily essential for mtDNA replication. Mice lacking *Tfam* therefore show impaired mtDNA transcription and loss of mtDNA, leading to bioenergetic insufficiency and embryonic lethality<sup>5</sup>.

Upstream of both the LSP and HSP1 transcriptional start sites, Tfam recognizes a binding site that has been defined by DNase I footprinting experiments<sup>3,4</sup>. Tfam contains two HMG-box domains followed by a short C-terminal tail<sup>6</sup>. HMG-box domains are DNA-binding motifs that bind to the minor groove of DNA and, in some cases, result in DNA bending<sup>7</sup>. Tfam belongs to the HMG-box subgroup that contains tandem HMG-box domains<sup>7</sup>. Several proteins in this subgroup, such as Tfam, have important structural roles in DNA organization, but there is currently no information about how two HMG-box domains can be spatially coordinated to affect DNA structure. The C-terminal tail of Tfam is essential for transcriptional activation<sup>8</sup> and also for its physical association with Tfb2m<sup>9</sup>, another

transcription factor required for mtDNA transcription. As a result, it has been proposed that Tfam binding allows recruitment of Tfb2m by the C-terminal tail.

In addition to its transcriptional function, Tfam is thought to have an important role in mtDNA packaging<sup>10,11</sup>. Although Tfam functions as a sequence-specific transcription factor, it also has high affinity for nonspecific DNA. Unlike nuclear DNA, mtDNA is not associated with histones. mtDNA genomes within the mitochondrial matrix are organized into compact DNA–protein complexes called nucleoids<sup>12</sup>. Tfam is one of the most abundant proteins associated with mtDNA nucleoids<sup>13</sup>, and its levels have been estimated to be sufficient to coat the entire mitochondrial genome<sup>14</sup>. The levels of Tfam correlate with the levels of mtDNA<sup>15</sup>. The yeast ortholog of Tfam, ARS-binding factor 2, mitochondrial (Abf2), has no role in transcription, and its major function is thought to be in the organization of the mitochondrial genome<sup>16</sup>.

To understand how Tfam mediates these multiple functions on mtDNA, we have solved the structure of human Tfam in complex with the LSP binding site. The structure shows how Tfam coordinates its two HMG-box domains to impose a dramatic U-turn on the DNA. To bend DNA, Tfam uses structural principles analogous to those used by the HU family of prokaryotic nucleoid proteins, which, like Tfam, have architectural roles in genome organization. Moreover, we find this DNA bending is more important for transcriptional activation at LSP than HSP1.

## RESULTS

### Structure determination

We solved the 2.5-Å crystal structure of human Tfam bound to a 28-bp DNA fragment derived from LSP (Table 1 and Fig. 1a–d).

<sup>1</sup>Division of Biology, California Institute of Technology, Pasadena, California, USA. <sup>2</sup>Division of Chemistry, California Institute of Technology, Pasadena, California, USA. <sup>3</sup>Howard Hughes Medical Institute, California Institute of Technology, Pasadena, California, USA. Correspondence should be addressed to D.C.C. (dchan@caltech.edu).

Received 20 June; accepted 13 September; published online 30 October 2011; doi:10.1038/nsmb.2159



**Table 1** Data collection, phasing and refinement statistics

	Crystal 1 <sup>a</sup>			Crystal 2 <sup>a</sup>
<b>Data collection</b>				
Space group	C222 <sub>1</sub>			C222 <sub>1</sub>
Cell dimensions	68.36, 81.35, 160.63			68.44, 81.91, 161.25
<i>a</i> , <i>b</i> , <i>c</i> (Å)	90, 90, 90			90, 90, 90
$\alpha$ , $\beta$ , $\gamma$ (°)				
	Peak	Inflection	Remote	
Resolution (Å)	28.8–3.0 (3.17–3.0) <sup>b</sup>	28.8–3.0 (3.17–3.0) <sup>b</sup>	28.8–3.0 (3.17–3.0) <sup>b</sup>	24.2–2.5 (2.64–2.50) <sup>b</sup>
<i>R</i> <sub>merge</sub>	0.048 (0.120)	0.046 (0.114)	0.049 (0.215)	0.060 (0.454)
<i>I</i> / $\sigma$ <i>I</i>	21.9 (9.7)	23.3 (10.0)	21.3 (5.8)	12.2 (2.8)
Completeness	97.7 (91.4)	97.4 (88.8)	98.5 (97.4)	98.3 (99.5)
Redundancy	5.8 (4.2)	5.8 (4.0)	5.8 (4.6)	3.5 (3.6)
<b>Refinement</b>				
Resolution (Å)				24.2–2.5
No. reflections				15,795
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>				19.8 / 24.7
No. atoms				
Total				2,888
Protein				1,641
DNA				1,148
Water				80
<i>B</i> -factors (Å <sup>2</sup> )				
Protein				59.1
DNA				68.7
Water				55.3
R.m.s. deviations				
Bond lengths (Å)				0.007
Bond angles (°)				1.03

<sup>a</sup>Two crystals were used for the structure. <sup>b</sup>Values in parentheses are for highest-resolution shell.

The recombinant Tfam used (residues 43–246) represents the full-length, mature Tfam after cleavage of the N-terminal mitochondrial leader sequence<sup>6</sup>. The DNA fragment includes a ~22-bp sequence that was identified as a high-affinity Tfam-binding site by DNase I footprinting and has two half-sites that interact with the HMG-box domains<sup>3</sup>. A selenomethionine-substituted Tfam–mtDNA complex was used for structure determination at 2.5 Å by multiwavelength anomalous diffraction (MAD) analysis. The crystallographic statistics of data collection and refinement are presented in **Table 1**. The electron density map was of sufficient quality to build almost all of the protein (residues 43–237) and all 28 base pairs of mtDNA. Model building and refinement produced a final structure with excellent stereochemistry, with an *R*<sub>free</sub> of 24.7% and an *R*<sub>work</sub> of 19.8%.

This crystal structure is the first one of a native tandem HMG-box protein in complex with DNA. In a previous study<sup>17</sup>, the NMR structure was solved of a chimeric molecule consisting of the HMG-box domain of Sex-determining region Y protein (SRY, a single HMG-box protein) fused to one of the two HMG-box domains of HMGB1 (a tandem HMG-box protein). This artificial molecule is nonphysiological, and its structure in complex with DNA does not resemble the structure described here.

### Tfam imposes a severe bend on LSP mtDNA

The most striking feature of the structure is that binding of a Tfam monomer dramatically distorts the DNA into a U-shape, causing a reversal in the direction of the DNA helical axis (**Fig. 1c,d**). Each

HMG-box folds into a three-helix motif with a concave surface that intercalates between the bases in the minor groove of an LSP half-site (**Fig. 1c**). These two intercalations result in two sharp kinks on one face of the DNA helix. The buried contact area of the first HMG-box domain (box A) with DNA is 1,566 Å<sup>2</sup>, and the corresponding surface area of the second HMG-box (box B) is nearly as extensive at 1,404 Å<sup>2</sup> (**Fig. 1e,f**). The linker connecting the two HMG-box domains forms an  $\alpha$ -helix around which the DNA wraps (contact area 864 Å<sup>2</sup>) (**Fig. 1c,d,f**). As described in detail later, basic side chains in the linker interact with the negatively charged phosphates in the bent DNA backbone. The C-terminal tail also contacts DNA (580 Å<sup>2</sup>), and the first part of this region extends the third helix of the second HMG-box domain. Therefore, all four regions (**Fig. 1a**) of Tfam—the two HMG-box domains, the linker and the C-terminal tail—make extensive contact with the DNA.

The structure agrees well with previous DNase I footprinting and methylation interference experiments probing the binding of Tfam to LSP DNA<sup>3,18</sup>. The Tfam monomer accounts for the large recognition site identified by a combination of DNase I footprinting and sequence analysis<sup>3,4,8</sup>. Each HMG-box domain binds to one of the two half-sites identified by sequence analysis<sup>8</sup>.

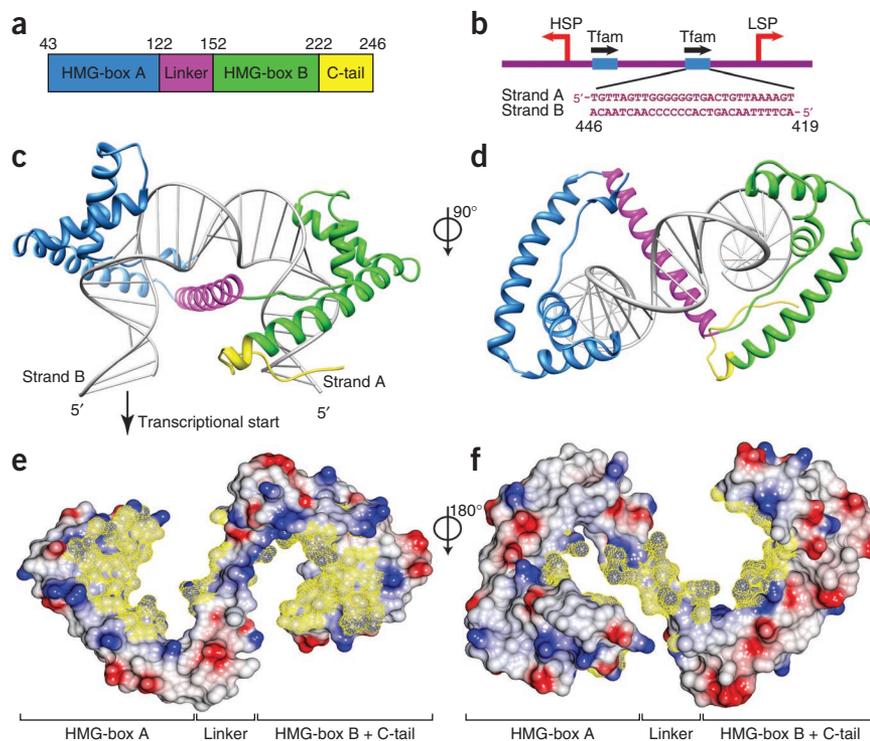
In previous methylation interference experiments<sup>18</sup>, the methylation by dimethylsulfate (DMS) of selected adenines was associated with reduced binding of Tfam. DMS methylates adenine at the N3 atom, which is located within the minor groove and would sterically block subsequent Tfam binding. In our crystal structure, all of the adenines identified by Clayton and colleagues<sup>18</sup> reside in a position where Tfam contacts the DNA minor groove and causes widening (**Supplementary Fig. 1**). By contrast, methylation of adenines located outside the contact area did not affect Tfam binding.

Our crystal structure indicates that Tfam binds mtDNA as a monomer. Human Tfam without DNA is monomeric, but it has been suggested that Tfam assembles into dimers on DNA binding<sup>19</sup>. The latter conclusion is tenuous, because it was based on a gel mobility assay that used extremely high concentrations of Tfam and DNA and does not give a definitive assessment of stoichiometry. To independently test the 1:1 stoichiometry found in our crystal structure, we analyzed Tfam and the Tfam–mtDNA complex in solution by size exclusion chromatography with in-line multi-angle light scattering analysis (SEC-MALS). The measured molar masses indicated that Tfam in isolation is monomeric and, when complexed with DNA, forms a 1:1 complex (**Fig. 2**).

### Protein-DNA interactions

As in other HMG-box structures, each Tfam HMG-box domain folds into an L shape composed of three  $\alpha$ -helices, with the third helix forming the long axis (**Fig. 1c**). A hydrophobic core composed of Tyr57, Phe60, Trp88 and Tyr99 stabilizes the L-shaped configuration

**Figure 1** Overview of the Tfam–mtDNA complex. (a) The domain structure of mature Tfam. Residues 1–42 constitute the mitochondrial targeting sequence that is cleaved upon import of Tfam into the mitochondrial matrix. (b) Organization of the LSP and HSP1 promoters. Comparative sequence analysis showed that the two Tfam binding sites are oriented in opposite directions relative to the direction of transcription<sup>4,8</sup>. The sequence of the LSP DNA fragment used for crystallization is indicated. (c) Side view of the Tfam–mtDNA complex. The Tfam domains are color coded as in a, and DNA is colored in gray. The LSP transcriptional start site would be located away from the DNA end on the left, as indicated by the arrow. Note that HMG-box B binds to the half-site further away from the transcriptional start site. (d) A view of the Tfam–mtDNA complex from the top. The protein and DNA are color coded as in c. (e) Electrostatic surface potential plot of Tfam. Surface areas of Tfam that are buried on DNA binding are highlighted in yellow mesh. The HMG-box A, linker, HMG-box B and C-terminal tail (C-tail) regions are labeled. Regions of negative electrostatic potential are indicated in red and regions of positive electrostatic potential in blue. (f) Electrostatic surface potential plot of Tfam, viewed in the same orientation as in d and flipped 180° from e. This view emphasizes that the surface of the linker contacts the DNA.



of the first HMG-box (Fig. 3a; Supplementary Fig. 2). Similarly, buried residues Tyr162, Tyr165, Trp189 and Tyr200 stabilize the second HMG-box domain (Fig. 3b; Supplementary Fig. 2). The overall folds of both HMG-box domains superimpose well with other HMG boxes whose structures in complex with DNA have been solved (Fig. 3c).

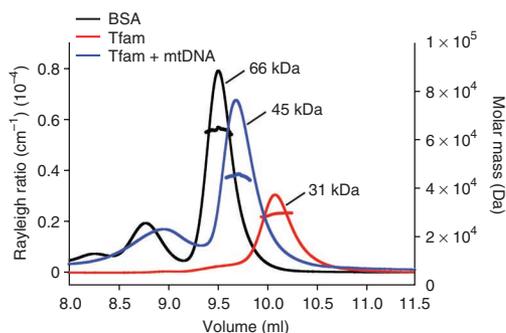
In the Tfam–mtDNA complex, most of the side-chain–DNA contacts are not sequence specific and occur on the sugar–phosphate backbone of the DNA. However, a small number of contacts to bases within the minor groove can be seen. HMG-box domains generally contain one or two hydrophobic residues that intercalate into the minor groove (highlighted in Supplementary Fig. 3). Consistent with this generalization, the HMG-box A of Tfam contains the first of these intercalating residues at position 58 (Leu58), which interacts with A8 (strand B) (Fig. 3d, red residue). A previous crystal

structure of the isolated, HMG-box B of Tfam raised the issue of whether it was a noncanonical HMG-box domain with unusual binding properties<sup>19</sup>, because it seemed to lack both intercalating hydrophobic residues. Our Tfam–mtDNA structure clarifies this issue by showing that HMG-box B does contain DNA-binding residues at these same positions, even though the residues are not nonpolar. In the first position, HMG-box B contains Asn163, which reaches into the minor groove and contacts the underlying thymine (T7, strand A). In the second position, Pro178 similarly inserts into the minor groove and contacts a guanine (G9, strand A) (Fig. 3e, red residues). In comparison to the previous structure of HMG-box B without DNA<sup>19</sup>, Pro178 has shifted >2 Å to make this contact with the DNA base.

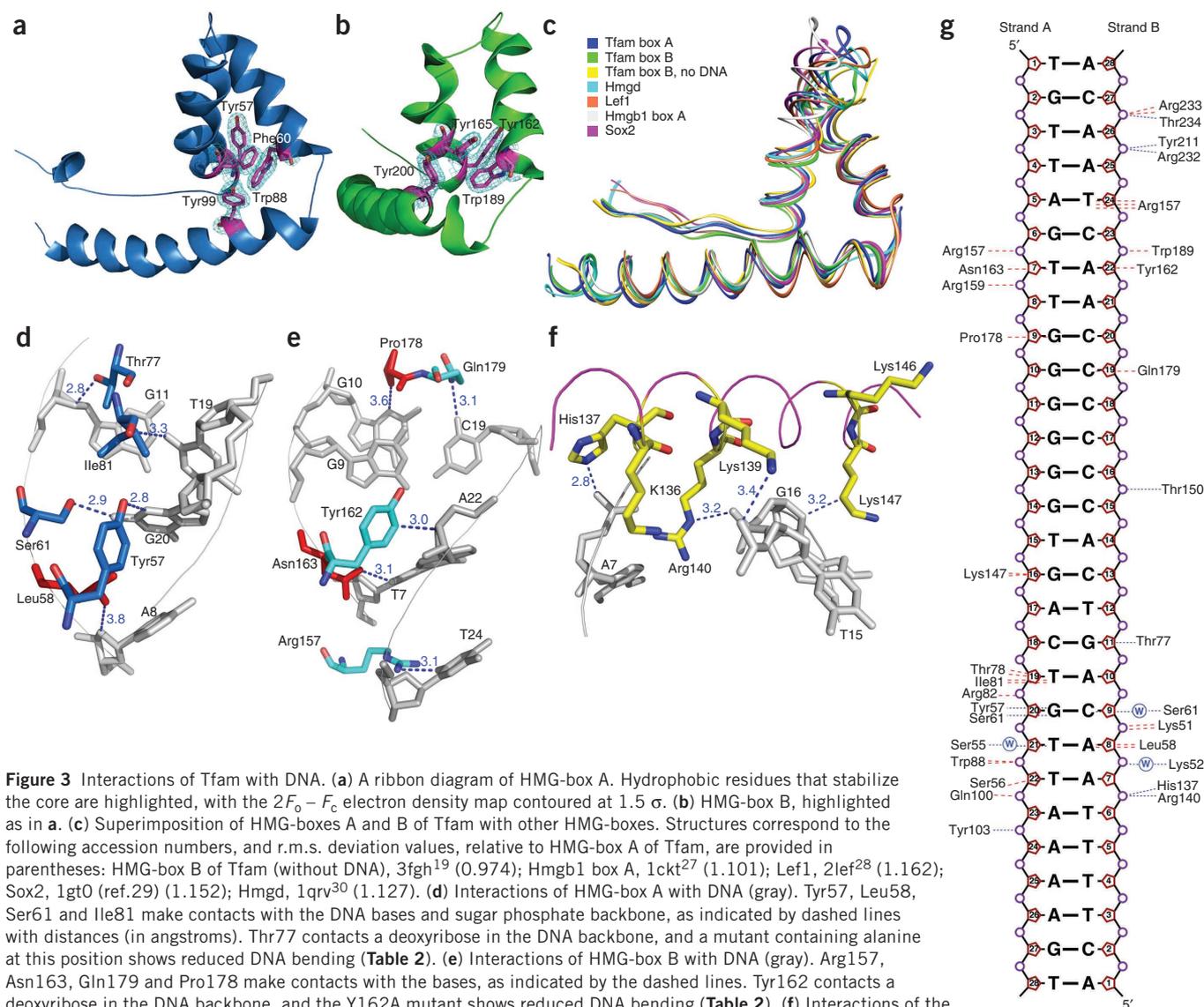
Besides the interactions indicated above, several other contacts to DNA bases are apparent. In HMG-box A, contacts are observed between Ile81 and T19 (strand A), Tyr57 and G20 (strand A), and Ser61 and G20 (strand A) (Fig. 3d; Supplementary Fig. 2c). In addition, Ser61 and Ser55 indirectly interact with C9 (strand B) and T21 (strand A), respectively, through water molecules. In the HMG-box B (Fig. 3e; Supplementary Fig. 2d), contacts are observed between Arg157 and T24 (strand B), and Gln179 and C19 (strand B). The linker does not directly interact with DNA bases. However, it makes substantial contacts with DNA via charged or polar interactions (Fig. 3f,g). Lys147 contacts G16 (strand A). His137 and Arg140 both make contacts to the phosphate backbone. Other lysine residues in the linker region (Lys136, Lys139 and Lys146) make longer-range contacts (>3.35 Å) with the sugar–phosphate backbone.

### Similarity to HU and IHF nucleoid proteins

The conformations of the two half-sites bound by Tfam deviate substantially from canonical B-DNA (Fig. 4a–d). At each location, intercalation by the HMG box results in substantial widening of the minor groove (Fig. 4a). There is local DNA unwinding, as indicated by sharp



**Figure 2** Molecular mass of the Tfam–mtDNA complex determined by SEC-MALS. Elution profiles of Tfam, the Tfam–mtDNA complex and BSA (control) examined by SEC-MALS. The horizontal black, red and blue lines correspond to SEC-MALS calculated masses for BSA, Tfam and Tfam–mtDNA, respectively. The corresponding theoretical masses are 28,075 Da (Tfam), 45,410 Da (Tfam–mtDNA; 1:1 complex) and 66,776 Da (BSA).



**Figure 3** Interactions of Tfam with DNA. **(a)** A ribbon diagram of HMG-box A. Hydrophobic residues that stabilize the core are highlighted, with the  $2F_o - F_c$  electron density map contoured at  $1.5\sigma$ . **(b)** HMG-box B, highlighted as in **a**. **(c)** Superimposition of HMG-boxes A and B of Tfam with other HMG-boxes. Structures correspond to the following accession numbers, and r.m.s. deviation values, relative to HMG-box A of Tfam, are provided in parentheses: HMG-box B of Tfam (without DNA), 3fgh<sup>19</sup> (0.974); Hmgb1 box A, 1ckt<sup>27</sup> (1.101); Lef1, 2lef<sup>28</sup> (1.162); Sox2, 1gt0 (ref.29) (1.152); Hmgd, 1qrv<sup>30</sup> (1.127). **(d)** Interactions of HMG-box A with DNA (gray). Tyr57, Leu58, Ser61 and Ile81 make contacts with the DNA bases and sugar phosphate backbone, as indicated by dashed lines with distances (in angstroms). Thr77 contacts a deoxyribose in the DNA backbone, and a mutant containing alanine at this position shows reduced DNA bending (**Table 2**). **(e)** Interactions of HMG-box B with DNA (gray). Arg157, Asn163, Gln179 and Pro178 make contacts with the bases, as indicated by the dashed lines. Tyr162 contacts a deoxyribose in the DNA backbone, and the Y162A mutant shows reduced DNA bending (**Table 2**). **(f)** Interactions of the  $\alpha$ -helical linker with DNA (gray). The backbone of the linker helix is traced in magenta. **(g)** Interactions between Tfam and DNA, analyzed by NUCPLOT<sup>31</sup>. Blue (dotted) and red (dashed) lines represent hydrogen-bonded and unbonded contacts ( $<3.35\text{ \AA}$ ) to DNA, respectively. Circles labeled W indicate water-mediated interaction with DNA. Stereo views of **a**, **b**, **d** and **e** are provided in **Supplementary Figure 2**.

minima in the base twist (base step parameter) at the sites of intercalation (**Fig. 4c**). Globally, however, the DNA is not underwound, with an average helical twist of  $\sim 36^\circ$ . The roll angle profile (**Fig. 4b**) shows two sharp peaks, which reflect distortions of base stacking owing to acute DNA bending.

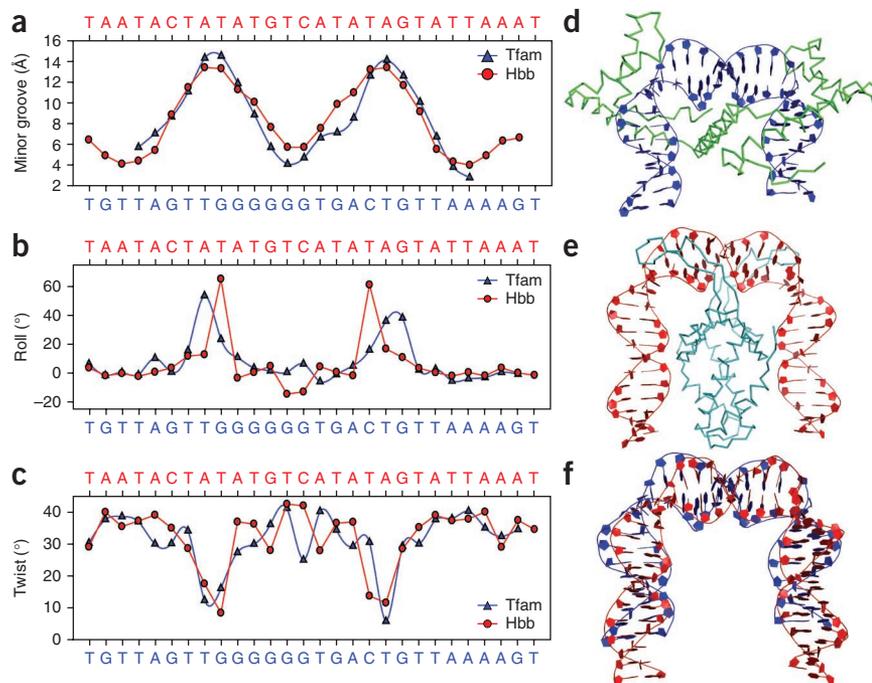
The mode of DNA bending in the Tfam–mtDNA structure shows remarkable parallels with the HU protein family, which consists of DNA minor groove-binding proteins that have architectural roles in prokaryotic DNA nucleoids<sup>20,21</sup>. Integration host factor (IHF), HU and Hbb are HU-family proteins that contort their bound DNA into a U-shape<sup>20,21</sup>. These proteins form dimers in which each subunit uses a ‘ $\beta$ -ribbon arm’ to intercalate into the DNA minor groove (**Fig. 4a,e**). The dimerization interface between the two subunits is rich in positive residues and serves to neutralize the negative charges of the bent DNA backbone. The DNA fragments in the Tfam and Hbb complexes show similar profiles in the minor groove width, with two broad peaks corresponding to minor groove intercalations (**Fig. 4a**). The roll angles also show two peaks that signify the sharp

bending of DNA. The peaks are slightly closer together in the Hbb ( $\sim 9$  bp apart) versus the Tfam structure (**Fig. 4b**). Superimposition of the DNA fragments reveals the similarity in overall geometry (**Fig. 4f**).

### Both HMG boxes and the linker are crucial for DNA bending

To monitor DNA bending by Tfam, we developed a fluorescence resonance energy transfer (FRET)-based assay. The crystal structure shows that after binding of Tfam, the ends of the 28 bp LSP DNA fragment are brought to within  $55\text{ \AA}$  of each other (measuring from the 5'-phosphate of one strand to the 5'-phosphate of the other strand), whereas there is a  $95\text{-\AA}$  separation in a rod-like DNA fragment of identical length. To construct the FRET sensor, Cy3 (donor) and Cy5 (acceptor) fluorophores were covalently attached to opposite ends of the LSP fragment. Addition of Tfam to the labeled, double-stranded DNA resulted in a dose-dependent increase in acceptor emission and a decrease in donor emission (**Table 2**; **Supplementary Fig. 4a,b**). Control experiments confirmed that the acceptor

**Figure 4** Comparison of Tfam and Hbb structures. (a) Profiles of minor groove width in the Tfam–mtDNA (blue) and Hbb–DNA (red) structures. (b) Roll angle profiles in the Tfam–mtDNA (blue) and Hbb–DNA (red) structures. (c) Twist angle profiles in the Tfam–mtDNA (blue) and Hbb–DNA (red) structures. (d) Side view of the Tfam–mtDNA complex. The protein is shown in green, and DNA is shown in blue. (e) Side view of the Hbb–DNA complex. Hbb is shown in light blue, and DNA is shown in red. (f) Manual overlay of DNA in the Tfam–mtDNA and Hbb–DNA structures. DNAs from the two structures are color coded as in d and e. Analyses of the helical parameters of the DNA molecules were carried out using 3DNA<sup>32</sup>.



emission depended on the presence of both the donor fluorophore and Tfam. The maximum FRET efficiency measured with wild-type Tfam corresponds to a calculated fluorophore separation of 59 Å, in good agreement with the crystal structure. The ability of Tfam to bend DNA is not restricted to the LSP template. Tfam was able to bend a DNA template lacking promoter sequences, and also one corresponding to HSP1 (Supplementary Fig. 4c,d).

Analysis of a panel of Tfam mutants indicated that coordinated binding of both HMG-box domains is important for effective bending of DNA (Table 2). HMG-box A alone bound to LSP DNA with the same affinity as full-length Tfam, consistent with previous studies<sup>19,22</sup>. However, it showed a large reduction in DNA bending. HMG-box B alone showed much weaker affinity for LSP DNA ( $K_d \sim 400$  nM) and also showed a large reduction in DNA bending. In addition, we tested Tfam mutants with single point mutations in HMG-box residues that contact DNA (Table 2; Supplementary Fig. 4b). T77A, which contains a mutation in HMG-box A, and Y162A, which contains a mutation in HMG-box B, showed moderate reductions in DNA bending. Each of these residues makes contacts with the DNA backbone (Fig. 3d,e). Finally, we tested the effect of mutations in the positively charged residues in the linker helix. Single point mutations had little effect (data not shown). We therefore made a mutant, L6, in which six positively charged residues in the linker region were replaced by alanine. The L6 mutant showed a >30% reduction in FRET, indicating that the linker region is important for DNA bending. All of these mutants were well folded, as established through secondary structure analysis by circular dichroism (Supplementary Fig. 4e). In addition, the T77A, Y162A and L6 mutants retained high affinity for DNA, as indicated by an assay monitoring the quenching of intrinsic tryptophan fluorescence upon DNA binding (Table 2; Supplementary Fig. 5).

### Tfam bending mutants show promoter-specific defects

With mutants that reduce DNA bending by Tfam, we used *in vitro* transcription reactions containing mitochondrial RNA polymerase (Polrmt) and Tfb2m to test whether DNA bending is important for its transcriptional activation function (Fig. 5). Neither HMG-box A nor HMG-box B alone was able to activate transcription from LSP or HSP templates. In addition, Tfam containing both HMG boxes but lacking the C-terminal tail was unable to activate transcription (Fig. 5a,b). These results are expected, because previous studies indicated that the C-terminal tail of Tfam is essential for transcriptional activation<sup>8</sup>.

Notably, we found that both the T77A and Y162A mutants were less efficient in promoting transcription from the LSP template. Y162A, which has a more severe bending defect, was more affected. Finally, the mutant L6, which has the strongest bending defect, showed a severe defect in transcriptional activation. The transcriptional defects were similar whether full-length or truncated LSP transcripts were quantified (Fig. 5c,d).

To test whether these Tfam mutants were generally defective in transcriptional activation, we examined their activity with an HSP1 template (Fig. 5b,e). In DNA-bending measurements, these mutants showed defects in bending the HSP1 DNA template (Supplementary Fig. 4d), as was found previously with the LSP template. In transcriptional activation assays, however, the Y77A, Y162A and L6 mutants were all efficient at stimulating transcripts from HSP1 (Fig. 5e). Quantification showed that all three mutants showed a similar transcriptional activation profile compared to wild-type Tfam.

**Table 2** DNA bending and binding of Tfam variants

Tfam mutant	FRET efficiency <sup>a</sup> (%)	Distance <sup>b</sup> (Å)	$K_d^c$ (nM)
No protein	6.5 ± 0.7	84	-
Wild type	36.5 ± 0.8	59	6.0 ± 0.9
T77A	31.4 ± 0.2	61	7.6 ± 1.1
Y162A	29.4 ± 0.6	63	12.3 ± 1.8
L6	26.1 ± 0.6	64	10.1 ± 1.2
HMG-box A	12.2 ± 0.4	75	6.5 ± 1.2
HMG-box B	6.2 ± 0.4	84	411.3 ± 46

The DNA bending and binding properties of Tfam and the indicated mutants were measured. DNA bending was measured with a FRET assay using Cy3–Cy5-labeled LSP DNA. The measured FRET efficiency was used to calculate the distance between the DNA ends. The affinity of Tfam mutants to DNA was monitored through the change in intrinsic tryptophan fluorescence on DNA binding. s.d. from three independent experiments are indicated. L6: K136A, H137A, K139A, R140A, K146A and K147A; HMG-box A: residues 43–122; HMG-box B: residues 153–222.

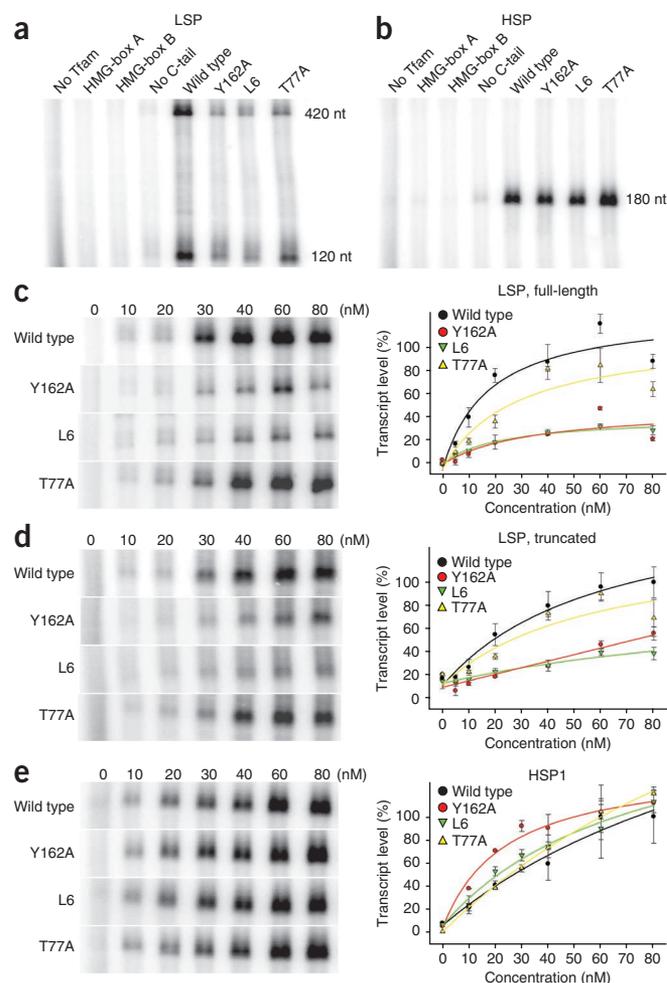
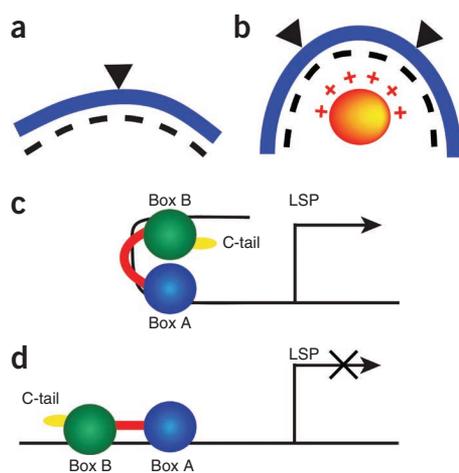
<sup>a</sup>The FRET efficiency ( $E$ ) was calculated with the following equation:  $E = (F_{corr}) / (F_{corr} + D_{corr})$ , where  $F_{corr}$  and  $D_{corr}$  are the corrected FRET and donor signals at 662 and 562 nm, respectively. <sup>b</sup>The distance was calculated from the FRET efficiency using the following equation:  $E = R_0^6 / (R_0^6 + R^6)$  with  $R_0 = 54$  Å. <sup>c</sup> $K_d$  was calculated from an LSP DNA-binding assay, as detailed in the Supplementary Methods.

**Figure 5** Tfam mutants with a selective defect at LSP. (a) *In vitro* transcription reactions using an LSP template. Reactions contained 100 nM Tfam or the indicated mutant. HMG-box A, residues 43–122; HMG-box B, residues 153–222; no C-tail, residues 43–222; L6, K136A, H137A, K139A, R140A, K146A and K147A. The LSP template generates a 420-nucleotide (nt) full-length (run-off) transcript and a truncated 120 nt transcript. (b) Same as a, except using an HSP1 template. (c) Generation of full-length LSP transcripts by Tfam and mutants. The left panel shows representative reactions, using the indicated concentrations of protein. Quantification is presented in the right panel, with error bars representing s.d. from three independent experiments. (d) Same as in c, except that truncated LSP transcripts are shown and quantified. A fraction of LSP transcripts are known to terminate prematurely at the conserved sequence block II (CSBII) site located downstream of the start site<sup>33</sup>. (e) Same as in c, except that an HSP1 template was used.

## DISCUSSION

Previous structural studies have indicated that a single HMG-box domain can bind to the DNA minor groove and sometimes cause bending of the DNA double helix. For example, the prototypical HMG-box protein Sry, which contains a single HMG-box domain, bends DNA  $\sim 70\text{--}80^\circ$  on binding to the minor groove<sup>23</sup>. This mode of DNA bending (Fig. 6a) superficially resembles that of TATA-box-binding protein (Tbp), in which binding of a  $\beta$ -sheet to the DNA minor groove again induces moderate bending toward the opposite direction<sup>24,25</sup>.

In comparison to these structures, the Tfam–mtDNA complex illustrates how spatial coordination of tandem HMG-box domains can be harnessed to impose even more extreme distortion onto DNA (Fig. 6b). Tfam belongs to the subset of HMG-box proteins that contain tandem HMG-box domains. These HMG-box proteins generally show broad DNA binding and have important roles in regulating chromatin structure and function<sup>7</sup>. For example, Hmgb1 is an architectural protein on chromatin that has been implicated in transcription, chromatin organization and genome stability<sup>26</sup>. In Tfam, the  $\alpha$ -helical linker plays a key part by spatially coordinating the two HMG-box domains, so that they bind the DNA minor groove at sites located approximately one helical turn apart. Moreover, the linker further facilitates DNA bending by neutralizing the negative charges on the DNA backbone. Intriguingly, all of the other dual HMG-box proteins in the human genome contain a cluster of 5–8 positively charged residues in the short region between the HMG-box domains (Supplementary Fig. 6). It will be interesting to determine whether these residues have a role analogous to that of the linker region in Tfam.



Although Tfam and the HU family of nucleoid proteins do not share sequence or structural homology, our studies indicate that they use remarkably analogous strategies to impose extreme bending onto DNA (Fig. 6b). The similarities between the Tfam–DNA and HU–family–DNA structures are intriguing, given that both proteins are thought to control the architecture of DNA in nucleoids. The DNA in our structure is from LSP and therefore is more directly related to mitochondrial transcriptional activation. However, the structure is likely to also be relevant for the role of Tfam in nucleoid organization, given the ability of Tfam to bend generic DNA (Supplementary Fig. 4c).

Our results indicate that the relative importance of extreme DNA bending by Tfam depends on the mitochondrial promoter.

**Figure 6** Models for DNA bending and transcriptional activation.

(a) DNA bending by a single HMG box. The DNA (blue) is moderately bent by wedging of the HMG box (triangle) on one face of the DNA. Dashes indicate negative charges on the opposite face of the DNA backbone. (b) Extreme DNA bending by Tfam and HU family proteins. Two wedges (triangles) applied to one face of DNA result in two acute kinks. A positively charged platform (circle) on the opposite face helps to neutralize the negative charges of the DNA backbone. (c) Transcriptional activation at LSP. Based on our crystal structure, HMG-box B binds the half-site further away from the transcriptional start site. The C-terminal tail (C-tail) nevertheless faces the transcriptional start site because of the extreme DNA bend. (d) With Tfam mutants, we suggest that the defect in DNA bending prevents proper orientation of the C-terminal tail.

Previous studies indicated that the C-terminal tail of Tfam is essential for transcriptional activation<sup>8</sup> and physical interaction with Tfb2m<sup>9</sup>. In the crystal structure, when Tfam is bound to the LSP promoter, the HMG-box B domain binds at the half-site further upstream from the transcription start site (Figs. 1c and 6c). Without DNA bending, the C-terminal tail would face away from the transcriptional start site (Fig. 6d). However, the DNA U-turn redirects the C-terminal tail toward the transcriptional machinery (Fig. 6c). We speculate that one of the functions of DNA bending by Tfam is to enable the C-terminal tail to interact with the rest of transcriptional machinery. Based on previous results<sup>9</sup>, Tfb2m is a favored candidate for such an interaction. Remarkably, transcription from HSP1 is much less sensitive to DNA bending by Tfam. Based on sequence analysis, the Tfam binding sites in HSP1 versus LSP are in reverse orientations relative to the direction of transcription<sup>4,8</sup> (Fig. 1b). When Tfam is bound to the HSP1 promoter, HMG-box B would be expected to bind the half site adjacent to the transcriptional start. The C-terminal tail is therefore in proximity to the transcriptional machinery, regardless of whether the DNA is bent or not. In future studies, it will be important to test this proposal by determining the structure of Tfam in complex with HSP1 promoter DNA.

## METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/nsmb/>.

**Accession codes.** Protein Data Bank: atomic coordinates and structure factors for the Tfam–mtDNA complex have been deposited under the accession code 3TMM.

*Note: Supplementary information is available on the Nature Structural & Molecular Biology website.*

## ACKNOWLEDGMENTS

We thank N. Chan (California Institute of Technology) for making some mutant constructs, Y. Zhang and Z. Liu (California Institute of Technology) for suggestions on phase determination and structure refinement, T. Walton (California Institute of Technology) for advice on SEC-MALS, S. Shan (California Institute of Technology) for use of equipment and insightful discussions, the staff at the Stanford Synchrotron Radiation Lightsource (SSRL) for technical support with crystal screening and data collection, and members of the Chan laboratory for critical reading of the manuscript. We acknowledge the Gordon and Betty Moore Foundation for support of the Molecular Observatory at Caltech. SSRL is supported by the US Department of Energy and National Institutes of Health (NIH). This work was supported by NIH grants GM083121 (D.C.C.) and GM062967 (D.C.C.).

## AUTHOR CONTRIBUTIONS

H.B.N. and D.C.C. designed the experiments, analyzed the data and wrote the paper. H.B.N. carried out the crystallography and performed the experimental work. J.T.K. helped with the crystallographic analysis.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/nsmb/>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Falkenberg, M., Larsson, N.G. & Gustafsson, C.M. DNA replication and transcription in mammalian mitochondria. *Annu. Rev. Biochem.* **76**, 679–699 (2007).
2. Bonawitz, N.D., Clayton, D.A. & Shadel, G.S. Initiation and beyond: multiple functions of the human mitochondrial transcription machinery. *Mol. Cell* **24**, 813–825 (2006).
3. Fisher, R.P. & Clayton, D.A. Purification and characterization of human mitochondrial transcription factor 1. *Mol. Cell. Biol.* **8**, 3496–3509 (1988).

4. Fisher, R.P., Topper, J.N. & Clayton, D.A. Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements. *Cell* **50**, 247–258 (1987).
5. Larsson, N.G. *et al.* Mitochondrial transcription factor A is necessary for mtDNA maintenance and embryogenesis in mice. *Nat. Genet.* **18**, 231–236 (1998).
6. Parisi, M.A. & Clayton, D.A. Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. *Science* **252**, 965–969 (1991).
7. Stros, M., Launholt, D. & Grasser, K.D. The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cell. Mol. Life Sci.* **64**, 2590–2606 (2007).
8. Dairaghi, D.J., Shadel, G.S. & Clayton, D.A. Addition of a 29 residue carboxyl-terminal tail converts a simple HMG box-containing protein into a transcriptional activator. *J. Mol. Biol.* **249**, 11–28 (1995).
9. McCulloch, V. & Shadel, G.S. Human mitochondrial transcription factor B1 interacts with the C-terminal activation region of h-mtTFA and stimulates transcription independently of its RNA methyltransferase activity. *Mol. Cell. Biol.* **23**, 5816–5824 (2003).
10. Kang, D., Kim, S.H. & Hamasaki, N. Mitochondrial transcription factor A (TFAM): roles in maintenance of mtDNA and cellular functions. *Mitochondrion* **7**, 39–44 (2007).
11. Kaufman, B.A. *et al.* The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures. *Mol. Biol. Cell* **18**, 3225–3236 (2007).
12. Spelbrink, J.N. Functional organization of mammalian mitochondrial DNA in nucleoids: history, recent developments, and future challenges. *IUBMB Life* **62**, 19–32 (2010).
13. Bogenhagen, D.F., Rousseau, D. & Burke, S. The layered structure of human mitochondrial DNA nucleoids. *J. Biol. Chem.* **283**, 3665–3675 (2008).
14. Alam, T.I. *et al.* Human mitochondrial DNA is packaged with TFAM. *Nucleic Acids Res.* **31**, 1640–1645 (2003).
15. Ekstrand, M.I. *et al.* Mitochondrial transcription factor A regulates mtDNA copy number in mammals. *Hum. Mol. Genet.* **13**, 935–944 (2004).
16. Kucej, M., Kucejova, B., Subramanian, R., Chen, X.J. & Butow, R.A. Mitochondrial nucleoids undergo remodeling in response to metabolic cues. *J. Cell Sci.* **121**, 1861–1868 (2008).
17. Stott, K., Tang, G.S., Lee, K.B. & Thomas, J.O. Structure of a complex of tandem HMG boxes and DNA. *J. Mol. Biol.* **360**, 90–104 (2006).
18. Fisher, R.P., Parisi, M.A. & Clayton, D.A. Flexible recognition of rapidly evolving promoter sequences by mitochondrial transcription factor 1. *Genes Dev.* **3**, 2202–2217 (1989).
19. Gangelhoff, T.A., Mungalachetty, P.S., Nix, J.C. & Churchill, M.E. Structural analysis and DNA binding of the HMG domains of the human mitochondrial transcription factor A. *Nucleic Acids Res.* **37**, 3153–3164 (2009).
20. Mouw, K.W. & Rice, P.A. Shaping the *Borrelia burgdorferi* genome: crystal structure and binding properties of the DNA-bending protein Hbb. *Mol. Microbiol.* **63**, 1319–1330 (2007).
21. Rice, P.A., Yang, S., Mizuuchi, K. & Nash, H.A. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* **87**, 1295–1306 (1996).
22. Wong, T.S. *et al.* Biophysical characterizations of human mitochondrial transcription factor A and its binding to tumor suppressor p53. *Nucleic Acids Res.* **37**, 6765–6783 (2009).
23. Werner, M.H., Huth, J.R., Gronenborn, A.M. & Clore, G.M. Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell* **81**, 705–714 (1995).
24. Kim, J.L., Nikolov, D.B. & Burley, S.K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527 (1993).
25. Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520 (1993).
26. Stros, M. HMG proteins: interactions with DNA and chromatin. *Biochim. Biophys. Acta* **1799**, 101–113 (2010).
27. Ohndorf, U.M., Rould, M.A., He, Q., Pabo, C.O. & Lippard, S.J. Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. *Nature* **399**, 708–712 (1999).
28. Love, J.J. *et al.* Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791–795 (1995).
29. Reményi, A. *et al.* Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev.* **17**, 2048–2059 (2003).
30. Murphy, F.V.IV., Sweet, R.M. & Churchill, M.E. The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition. *EMBO J.* **18**, 6610–6618 (1999).
31. Luscombe, N.M., Laskowski, R.A. & Thornton, J.M. NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. *Nucleic Acids Res.* **25**, 4940–4945 (1997).
32. Lu, X.J. & Olson, W.K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31**, 5108–5121 (2003).
33. Pham, X.H. *et al.* Conserved sequence box II directs transcription termination and primer formation in mitochondria. *J. Biol. Chem.* **281**, 24647–24652 (2006).

## ONLINE METHODS

**Tfam purification.** The human *TFAM* gene was cloned into the pET28a expression vector (Novagen) between the BamHI and XhoI sites. This construct encodes residues 43–246, corresponding to full-length Tfam after cleavage of the N-terminal mitochondrial leader sequence (residues 1–42). Tfam mutants were constructed using PCR with oligonucleotides encoding mutations. Plasmids were transformed into BL21 (DE3) *Escherichia coli* (Invitrogen). LB medium (20 ml) containing 50  $\mu\text{g ml}^{-1}$  kanamycin was inoculated with a single colony and grown overnight at 37 °C. The overnight culture was diluted to 4 l and grown until an  $\text{OD}_{600}$  it reached of 1.0. After induction with 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside, the culture was grown overnight at room temperature (24 °C). The cells were harvested and stored at –80 °C. Five grams of cells were resuspended in 50 ml lysis buffer (20 mM Tris-HCl, 500 mM NaCl, pH 7.5) and sonicated for 5 min (10 s on and 20 s off) on ice. After centrifugation at  $4.3 \times 10^4\text{g}$  for 30 min at 4 °C, His-tagged Tfam was purified from the supernatant with 3 ml of Talon Cobalt resin (Clontech). The protein was eluted (20 mM Tris-HCl, 500 mM NaCl, 300 mM imidazole, pH 7.5) and further purified by gel filtration chromatography using a Hi-Load Superdex 200 16/60 column (GE Healthcare) pre-equilibrated with running buffer (20 mM Tris-HCl, 300 mM NaCl, 1 mM dithiothreitol (DTT), pH 7.5) in an AKTA Purifier (Amersham). The peak fraction was collected and concentrated to 17–20  $\text{mg ml}^{-1}$  using Amicon Ultra-15 concentrators (Millipore) with a molecular weight cutoff of 10 kDa. The protein was flash-frozen in liquid nitrogen and stored at –80 °C. Selenomethionine-substituted Tfam was produced by the metabolic inhibition method<sup>34</sup>, and preparative buffers contained 5 mM  $\beta$ -mercaptoethanol instead of DTT. Proteins were analyzed by DNA binding and circular dichroism analysis, as detailed in the **Supplementary Methods**.

**Crystallization, data collection and structure determination.** The duplex LSP fragment was made by annealing complementary oligonucleotides (5'-TGTTA GTTGGGGGGTGACTGTTAAAAGT-3' and 5'-ACTTTTAACAGTCACCCC CCAACTAACA-3') in buffer (10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM EDTA) at a concentration of 0.9 mM. The mixture was incubated at 95 °C for 5 min, 75 °C for 5 min and room temperature for >5 h.

To form Tfam–mtDNA complexes, Tfam was mixed with duplex DNA in a 1.3:1 molar ratio. The mixture was incubated at room temperature for 30 min and then on ice for 2 h. Crystallization trials by hanging drop-vapor diffusion at room temperature identified a condition (29% (w/v) PEG 400, 0.15 M calcium acetate, 0.1 M sodium acetate (pH 4.2), 400 mM NDSB211 (dimethyl-2-(hydroxyethyl)-(3-sulfopropyl)-ammonium)) that yielded rod-shaped crystals. Diffraction data were collected on frozen crystals on beamline 12-2 at the Stanford Synchrotron Radiation Lightsource. All data were processed with IMOSFLM<sup>35</sup> or XDS<sup>36</sup>, and merged using SCALA<sup>37</sup> as implemented in CCP4 (ref. 38). A selenomethionine-substituted Tfam–mtDNA complex was used for phasing. Using intensity data at 3.0 Å from three wavelengths, all five selenium sites were located with PHENIX<sup>39</sup>. After solvent flattening and density modification in PHENIX, the map revealed clear density for the protein and DNA. Manual model building in COOT<sup>40</sup> using the 3.0-Å experimental map generated a starting model. Refinement of the best solutions was carried out using PHENIX, with an initial round of rigid body refinement followed by a round of simulated annealing. Refinement against a 2.5-Å data set produced an excellent map with density for most of the side chains. After a few rounds of model adjustment and refinement with TLS obtained from the TLSMD server<sup>41</sup>, the  $R_{\text{work}}$  converged to 19.8% and the  $R_{\text{free}}$  to 24.7%. The final model includes residues 43–237 of Tfam and all the nucleotides. The current model has excellent stereochemistry with no Ramachandran outliers, as assessed by MOLPROBITY<sup>42</sup>.

**FRET experiments.** To generate LSP, HSP and non-promoter templates, the following complementary oligonucleotides were annealed as described above:

LSP, 5'-Cy3-TGTTAGTTGGGGGGTGACTGTTAAAAGT-3' and 5'-Cy5-ACTTTTAACAGTCACCCCCCAACTAACA-3'; HSP1, 5'-Cy3-GGTTGGTTCGGGGTATGGGGTTAGCAGC-3' and 5'-Cy5-GCTGCTAACCCCATACCCCGA ACCAACC-3'; non-promoter DNA, 5'-Cy3-GACATTGGGAACACTATACCTA TTATTCG-3' and 5'-Cy5-cgaataataggtatgttccaatgctc-3'.

Additional details of the FRET measurements and analysis of the FRET data are described in the **Supplementary Methods**.

**SEC-MALS.** SEC-MALS experiments were performed at room temperature by loading samples on a Shodex KW 803 column with a Dawn Heleos MALS detector (Wyatt Technology). The column was eluted with buffer containing 20 mM Tris-HCl (pH 7.5), 300 mM NaCl and 1 mM DTT. A  $dn/dc$  (refractive index increment) value of 0.185  $\text{ml mg}^{-1}$  was used. Bovine serum albumin was used as an isotropic scatterer for detector normalization. The light scattered by a protein is directly proportional to its weight-average molecular mass and concentration.

**In vitro transcription reactions.** DNA fragments corresponding to LSP (positions 1–477) and HSP1 (positions 499–741) of human mtDNA were cloned into the pSP65 vector at the BamHI and SalI sites. After digestion with BamHI for LSP and SalI for HSP1, the linearized plasmids were used as templates in a transcriptional run-off assay. Transcription reactions were carried out as described<sup>43</sup> with modifications. Template DNA (5 nM) was added to the reaction mix (10 mM HEPES (pH 7.5), 10 mM  $\text{MgCl}_2$ , 1 mM DTT, 100  $\mu\text{g ml}^{-1}$  BSA and 40 units of RNaseOut (Invitrogen)) for 5 min, and then Tfam, Tfb2m (30 nM, Enzymax) and Polrmt (30 nM, Enzymax) were sequentially added, with a 1-min incubation between each addition. After addition of rNTPs (400  $\mu\text{M}$  rATP, 150  $\mu\text{M}$  rCTP, 150  $\mu\text{M}$  rGTP, 15  $\mu\text{M}$  rUTP (Promega), 0.2  $\mu\text{M}$  [ $\alpha$ -<sup>32</sup>P]rUTP (3,000 Ci  $\text{mmol}^{-1}$ , PerkinElmer)), the reaction was incubated for 3 h at 33 °C, and stopped by addition of 25  $\mu\text{l}$  of stop buffer (80% formamide (v/v), 10 mM EDTA, pH 8.0, 0.025% xylene cyanol (w/v), 0.025% bromophenol blue (w/v)). Samples were heated to 90 °C for 5 min and separated on 5% polyacrylamide gels (w/v) containing 8 M urea in 1 $\times$  TBE buffer. The gels were fixed in 7% (v/v) acetic acid, dried and exposed to a phosphorimager screen. The data were collected on a Storm 880 phosphorimager (Molecular Dynamics) and quantified using ImageQuant 5.2 Software.

34. Van Duyn, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. & Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124 (1993).
35. Leslie, A.G.W. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* (Warrington WA4 4AD, UK, 1992).
36. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
37. Evans, P.R. Data reduction. in *Proceedings of the CCP4 Study Weekend. Data Collection and Processing* (eds. Sawyer, L., Isaacs, N. & Bailey, S.) 114–122 (Daresbury Laboratory, Warrington, UK, 1993).
38. Collaborative Computational Project. N. The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
39. Adams, P.D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954 (2002).
40. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
41. Painter, J. & Merritt, E.A. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **39**, 109–111 (2006).
42. Davis, I.W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
43. Lodeiro, M.F. *et al.* Identification of multiple rate-limiting steps during the human mitochondrial transcription cycle in vitro. *J. Biol. Chem.* **285**, 16387–16402 (2010).