SUPPLEMENTARY MATERIAL

DNA specificities modulate human transcription factor A binding at the mitochondrial DNA control region

Anna Cuppari[1,a], Pablo Fernández-Millán[1,a], Federica Battistini[2,a], Aleix Tarrés-Solé[1], Sébastien Lyonnais[1], Guillermo Iruela[3], Elena Ruiz-López[1], Anna Rubio-Cosials[1], Rafel Prohens[4], Miquel Pons[3], Carlos Alfonso[5], Katalin Tóth[6], Germán Rivas[5], Modesto Orozco[2,7,8], Maria Solà*[1]

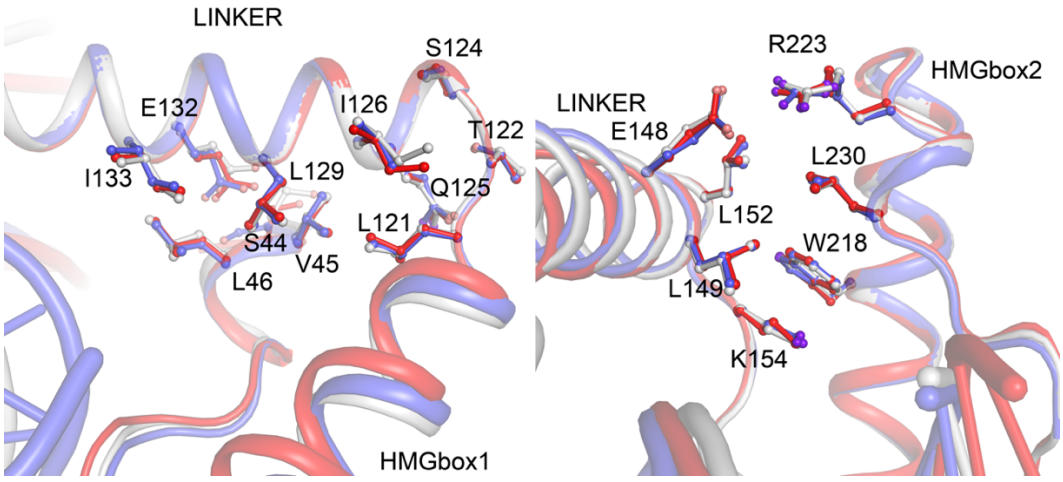
	TFAM/Site-X	TFAM/Site-Y
PDB code	6HC3	6HB4
Data Collection		
Wavelength	0.87260	0.87260
Resolution range (last shell)	42.21 - 3.10 (3.27 - 3.10)	86.25 – 3.00 (3.18 – 3.00)
Total reflections	132671 (19425)	229166 (36730)
Unique reflections	34552 (5000)	36317 (5776)
Space group	C2	C2
Cell dimensions (a, b, c, α, β, γ)	164.8, 145.1, 108.1, 90.0, 130.8, 90.0	158.8, 141.3, 109.0 90.0, 130.6, 90.0
Rmeas* (%)	7.6 (30.1)	12.6 (101.8)
I/σI	15.9 (4.7)	16.05 (2.4)
Completeness (%)	99.0 (99.1)	98.6 (98.3)
Redundancy	3.8 (3.9)	6.3 (6.4)
Refinement		
Resolution	42.2 - 3.10	86.3 – 3.05
Nr of reflections	34552	34312
Rwork / Rfree (%)	0.193 / 0.232	0.190 / 0.236
Nº of atoms:		
- protein	6510	6551
- DNA	3584	3793
water molecules	13	0
Bfactors:	65.5	92.5
Residues in favoured and allowed Ramachandran regions	100 %	100 %
R.m.s. deviations		
- Bond lengths (Å)	0.010	0.012
- Bond angles (°)	1.36	1.52

Supplementary Table 1. Data collection and refinement statistics.

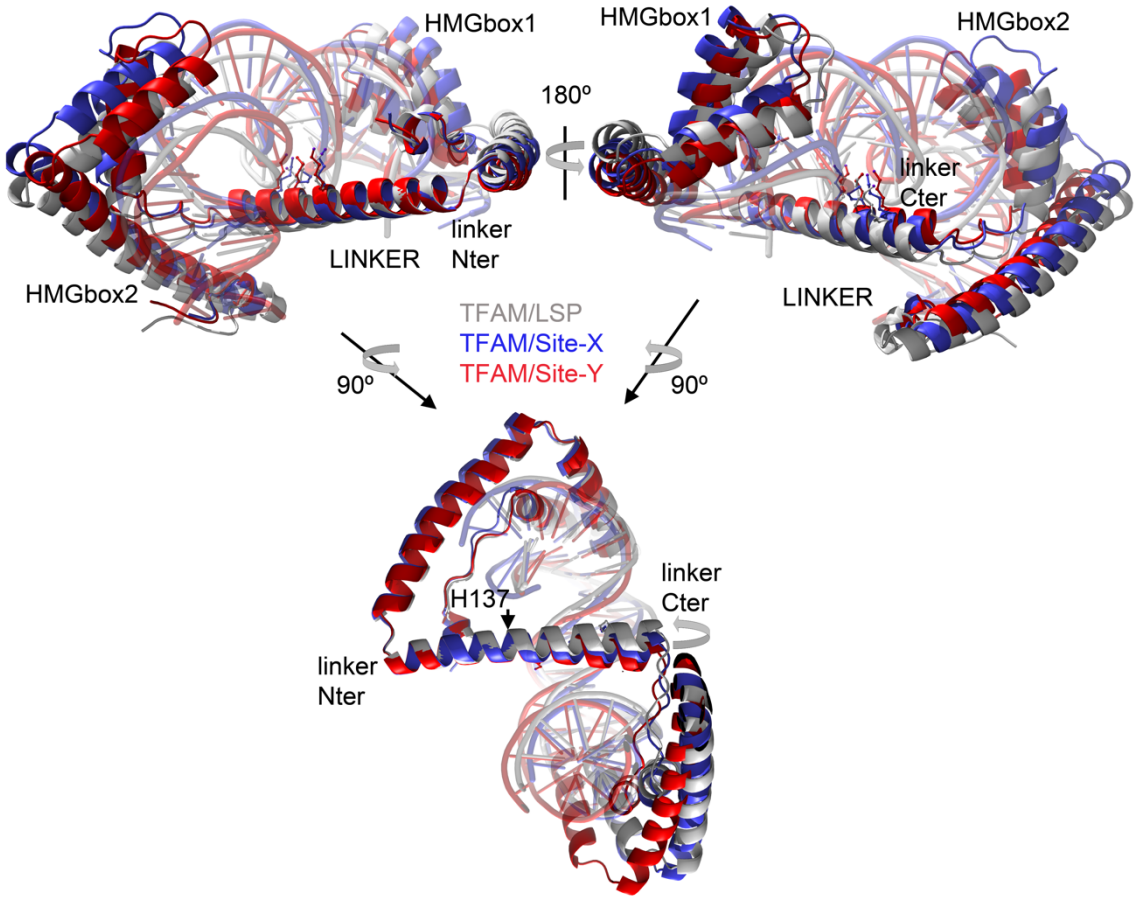
*Rmeas, redundancy independent R-factor

Moving chain to fixed chain (superposed aa)	RMSD (Å)	Max deviation between HMG-box1's (at aa), (Å)	Max deviation between HMG-box2's (at aa), (Å)
TFAM/Site-X HMG-box1			
Chain D to A (50-119)	0.21	0.80 (Lys 76)	9.53 (Asn 191)
Chain G to A (50-119)	0.21	-	9. 08 (Asn191)
Chain J on A (50-119)	0.04	-	-
TFAM/Site-Y HMG-box1			
Chain D to A (50-119)	0.37	0.84 (Ala 70)	3.84 (Ser 195)
Chain G to A (50-119)	0.87	2.53 (Asp 74)	9.69 (Glu 196)
Chain J on A (50-119)	0.63	1.05 (Thr 78)	6.75 (Lys 174)
TFAM/LSP HMG-box1			
Chain B to A (50-119)	0.28	0.61 (Lys 76)	4.18 (Ser 193)
TFAM/Site-X to LSP HMG-box1			
Chain A to A (50-119)	0.51	1.26 (Pro 50)	10.25 (Glu 196)
TFAM/Site-Y to LSP HMG-box1			
Chain A to A (50-119)	0.39	0.92 (Lys 76)	12.82 (Asp 184)
TFAM/Site-X to Site-Y HMG-box1			
Chain A to A (50-119)	0.57	1.23 (Val 54)	9.27 (Glu 172)
TFAM/nsDNA HMG-box1			
Chain A to B (50-119)	0.19	0.38 (Ala 75)	3.04 (Gly 175)
TFAM/nsDNA to LSP HMG-box1			
Chain A to A (50-119)	0.44	1.08 (Pro73)	15.54 (Ser 195)
TFAM/nsDNA to Site-X HMG-box1			
Chain A to A (50-119)	0.66	1.50 (Ser 55)	8.29 (Ser 193)
TFAM/nsDNA to Site-Y HMG-box1			
Chain A to A (50-119)	0.59	1.33 (Pro 73)	3.58 (Asp 176)

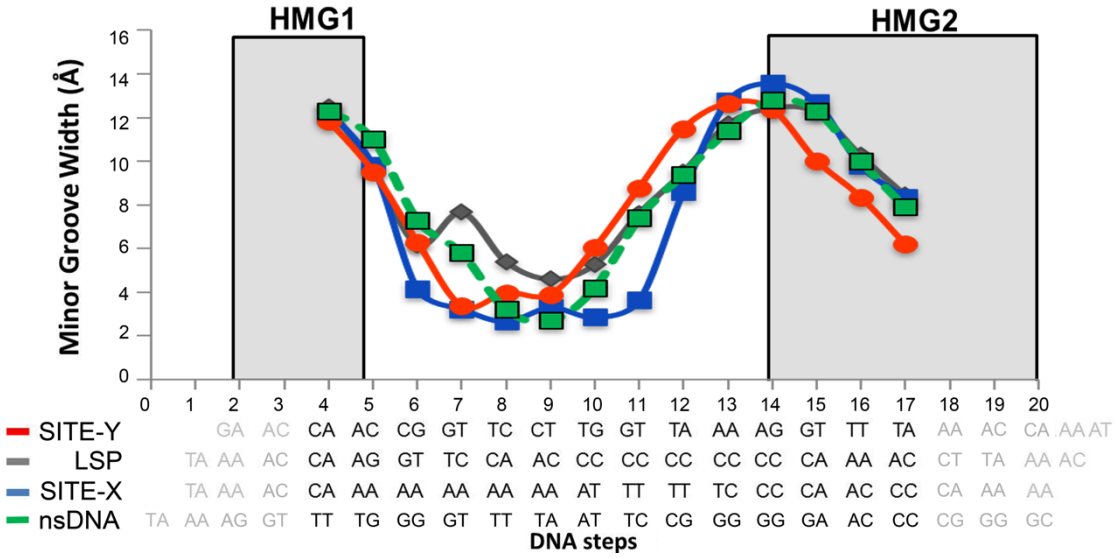
Supplementary Table 2. Structural variability of TFAM complexes. Root mean square deviation (RMSD) of superposed C^{α} of indicated aminoacids of Site-X, Site-Y, LSP (1) and nsDNA (2), and maximum deviations of the second HMG-box domain, calculated with Coot (3).



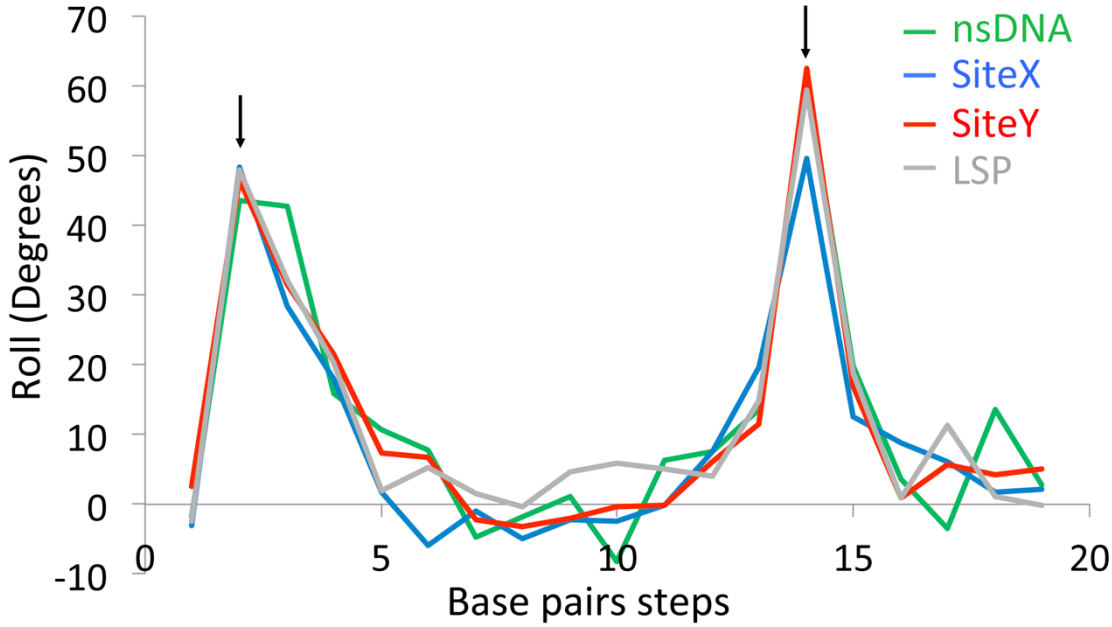
Supplementary Figure S1.
Representation of the side-chains of amino acids of TFAM/LSP (in grey, (1)), TFAM/Site-X (in blue), and TFAM/Site-Y (in red), located at the loops connecting HMG-box1 helix 3 with the linker (left panel) and the linker with helix3 from HMG-box2. Note the highly similar orientation of side chains in both regions, indicating high structural regularity. Same colour code as in Supplementary Figure S2.



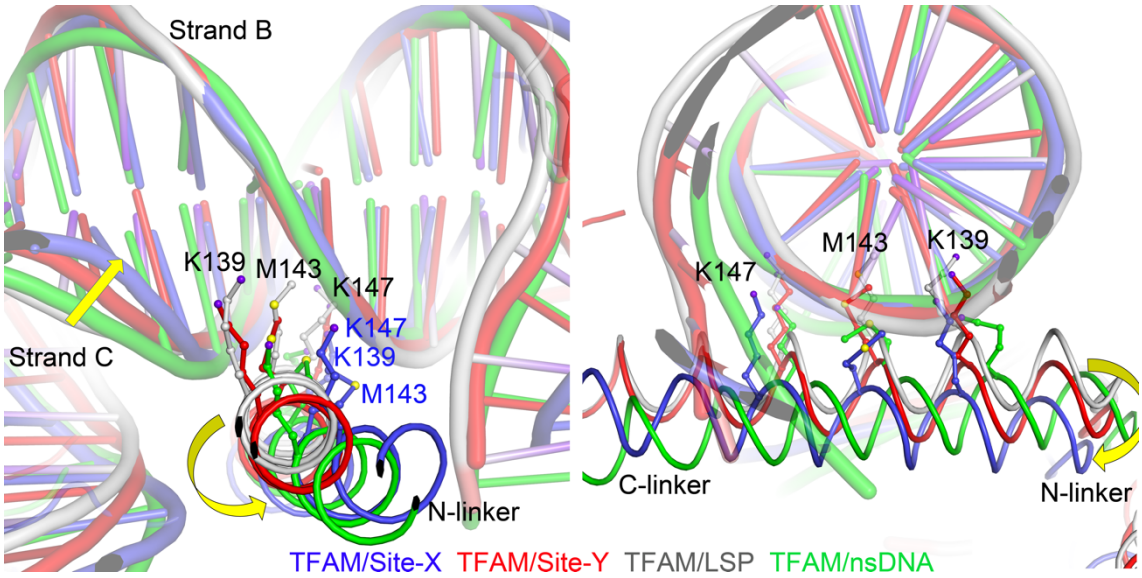
Supplementary Figure S2. The structural differences arise from the linker region. Top: two orientations of the superposition of the three complexes TFAM/LSP (in grey), TFAM/Site-X (in blue), and TFAM/Site-Y (in red) by the N-terminal region of the linker (linker N-ter). Bottom, an additional orientation that shows the displacement of the linker C-terminal end (linker C-ter), which starts with a distortion at His 137, as indicated.



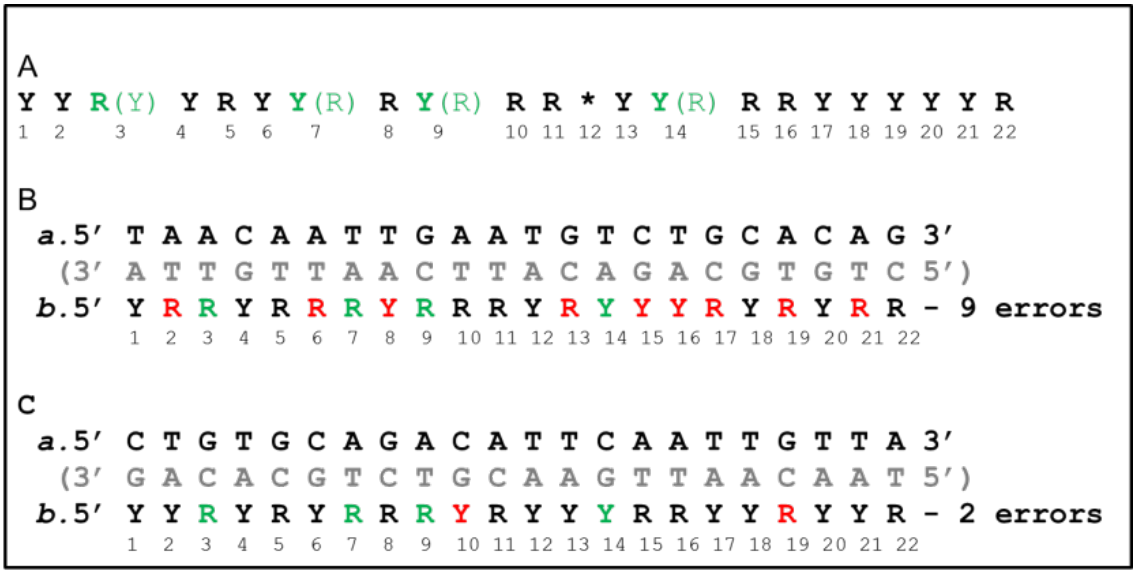
Supplementary Figure S3. **A)** Minor groove width values of nsDNA (2) (green line) compared to TFAM/Site-X (blue), TFAM/Site-Y (red) and TFAM/LSP (grey, (1)) structures.



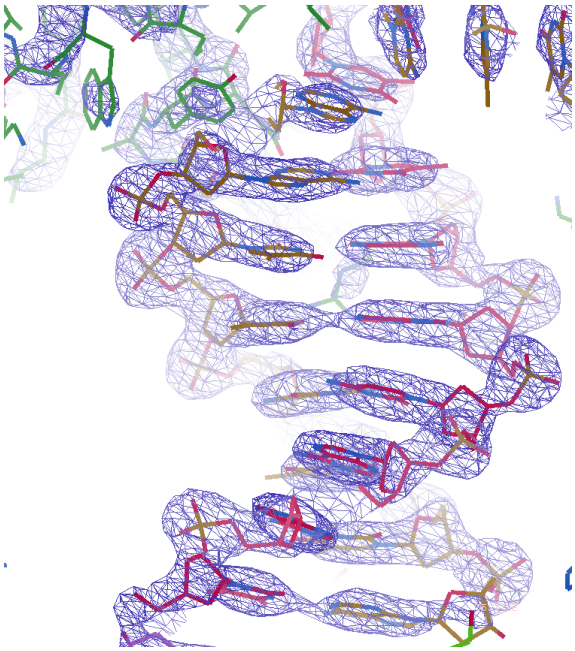
Supplementary Figure S4. Comparison of roll values of TFAM/nsDNA structure with the different structures analysed in this article. Vertical arrows indicate the site of insertion.



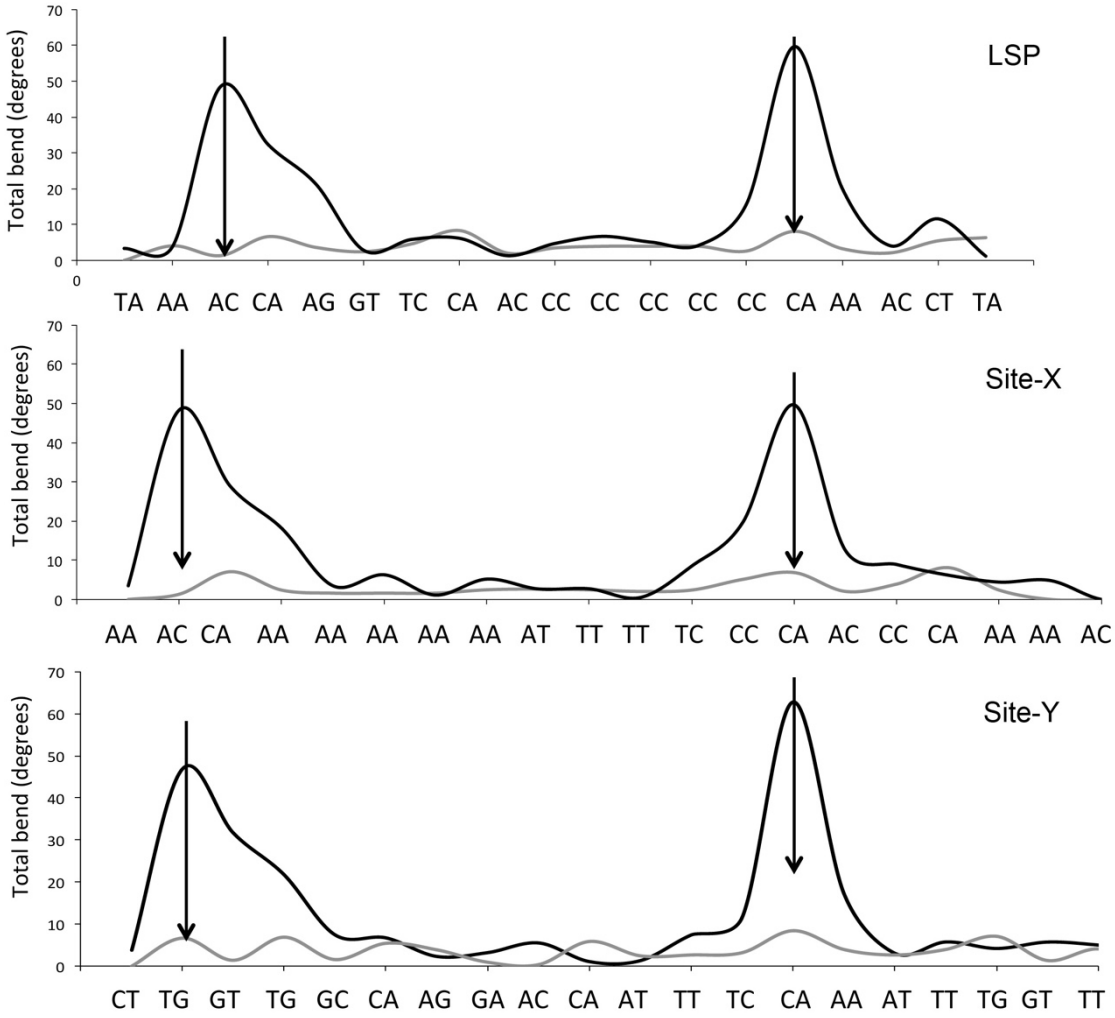
Supplementary Figure S5. **Displacement of the linker due to narrowing of the minor groove**. Left panel: frontal view of complexes TFAM/Site-X (in blue), TFAM/Site-Y (in red), TFAM/LSP (in grey, (1)) and TFAM/nsDNA (in green, (2)) superposed by DNA strand B. The straight yellow arrow shows the displacement of strand C in the different structures. The curved yellow arrow shows the displacement of the linker. Note that the displacement of Site-X strand C and, more locally, the same strand in the TFAM/nsDNA complex, cause the largest displacements of the linker due to narrowing of the minor groove. The three aminoacids from the linker that enter the most into the minor groove in the TFAM/LSP structure Lys 139, Met 143 and Lys 147 (represented as grey balls and sticks) are shown for all complexes, but labelled only in TFAM/LSP and TFAM/Site-X complexes (with black and blue labels respectively), for comparison. Right panel: lateral view of the same representation, showing the displacement of the linker and the residues; C and N-linker indicates the N-and C-terminal ends of the linker.



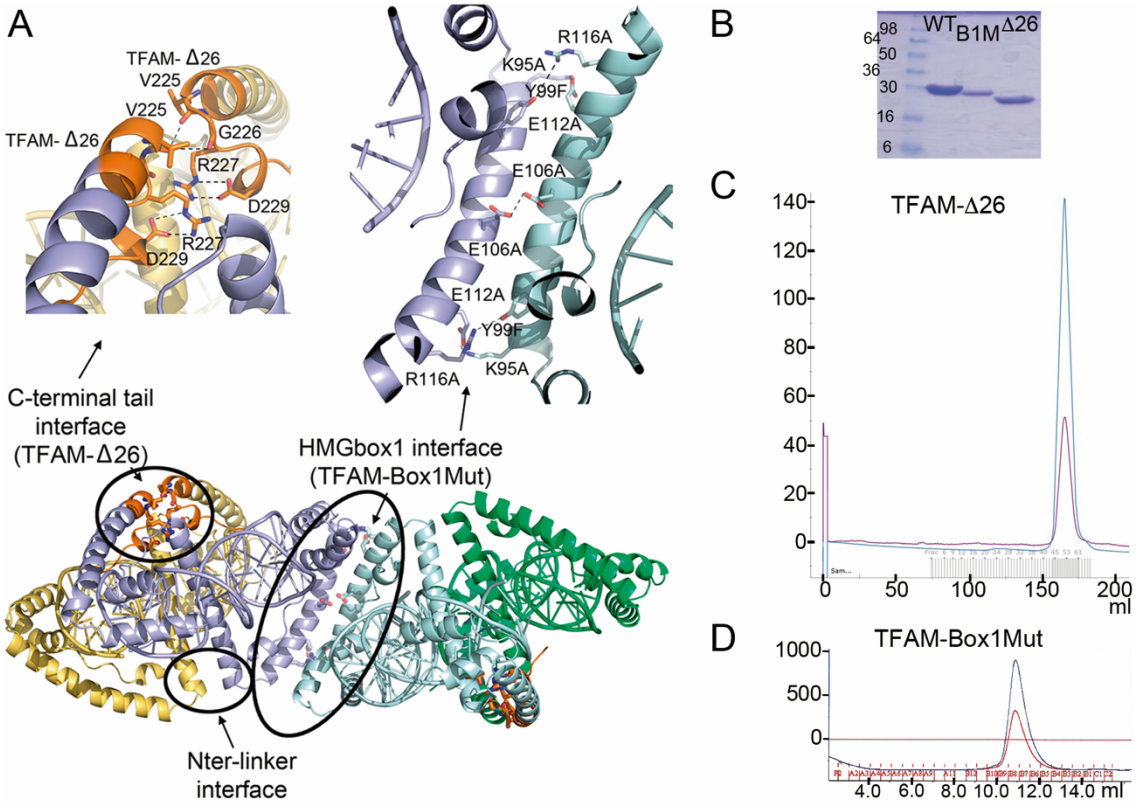
Supplementary Figure S6. Sequence assignment criteria based on the electron density map. **A)** The purine (R) and pyrimidine (Y) sequence suggested by the electron density map after refinement with an all-cytosines sequence is shown. Black characters correspond to R or Y bases assigned with more confidence, green characters are preferred guesses (which could also be the alternative in brackets). The asterisk * refers to a non-assignable position. **B)** In ***a,*** the 5'-3' Site-Y sequence (in grey, the complementary) assigned as aligned to LSP by Fisher *et al.* (4). ***b.*** R-Y translation of the sequence in *a*. From comparison with sequence in A, bases depicted in green are at "guess" (dubious) positions (see panel A), while the non-coincidences are in red (9 errors in total). **C)** Same as in B but for the new assignment based on the electron density map.



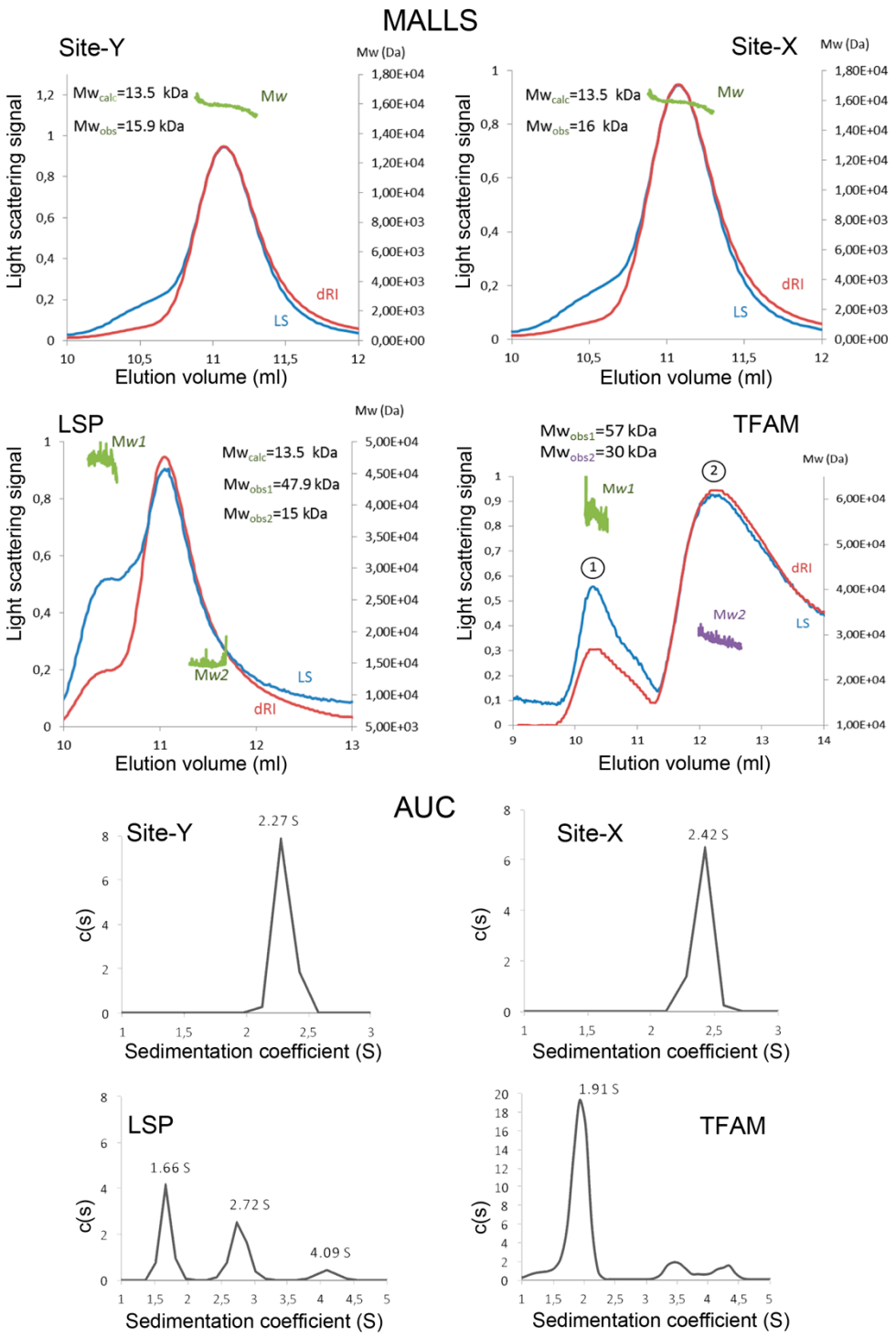
Supplementary Figure S7. Fragment of the final electron density map showing the quality in the assignment of the DNA sequence.



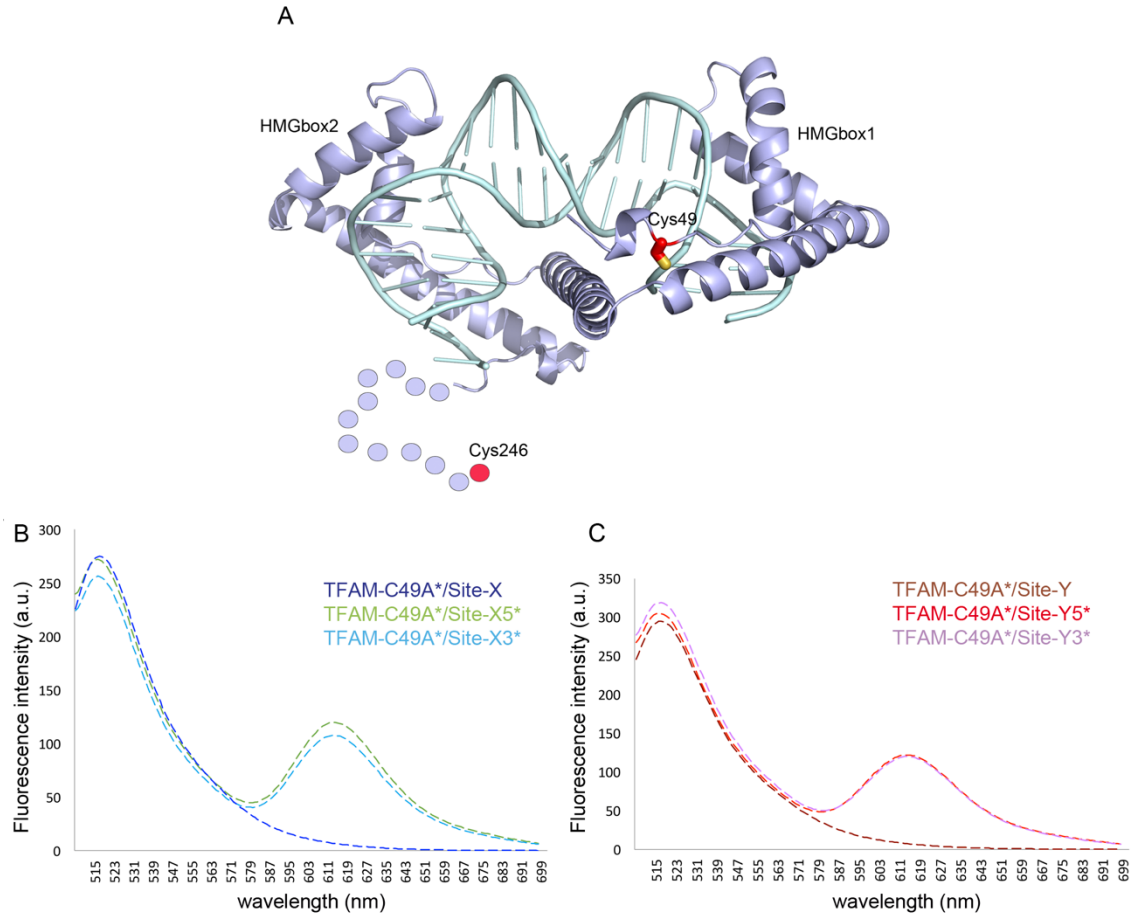
Supplementary Figure S8. Total bend (from crystals) or naked (MD-derived) LSP (top), Site-X (middle) and Site-Y (bottom). Total bend values, calculated from the roll and tilt contribution at each base pair step (in degrees) for protein-bound (black) and naked DNA (grey) are shown. The high bend peaks in the crystal structures correspond to the insertion sites, indicated by the arrows (left arrow, Leu58 insertion site; right arrow Leu182 site). For Site-Y, the tentative orientation assigned in the crystal is shown.



Supplementary Figure S9. Protein-protein interfaces identified in the crystal structures. **(A) Bottom**, crystal structure of TFAM/Site-X, which includes four protein/DNA complexes in the asymmetric unit. Each complex is represented in a different colour and performs three types of interactions, which is exemplified by the violet molecule. A first interaction occurs between the C-terminal tails (in orange) of the violet and yellow complexes and gives rise to the C-terminal tail interface (inset above left shows the involved side chains, represented as sticks). This interface is cancelled by deletion of the last 26 aa in mutant TFAM-Δ26 (5,6). A second interface involves the N-terminal region of the linker, indicated with a circle. A third interaction occurs between the violet and blue complexes and involves the respective HMG-box1 domains. Mutated residues in TFAM-Box1Mut (2), which cancel this interaction, are indicated in the above right inset. **(B)** Both TFAM-Δ26 (Δ26) and TFAM-Box1Mut (B1M) were obtained in high purity (15% SDS-PAGE). **(C)** TFAM-Δ26 eluted at the usual TFAM WT volume in size exclusion chromatography (in Superdex 75, 26/60, GE Healthcare). The absorption at 260 nm (red curve) was less than 70% of that at 280 nm (blue curve), which indicates absence of nucleic acids. **(D)** Same as in (C), TFAM-Box1Mut eluted at the expected volume (Superdex 75, 10/300, GE Healthcare), and no aggregates were present in the sample.



Supplementary Figure S10. Control experiments of the protein and DNA alone. Top four panels: the SEC-MALLS control experiments are shown for Site-Y (top left), Site-X (top right), LSP (bottom left) and TFAM (bottom right) alone. In each graph the light scattering curve (LS-curve) is shown in red, the differential refracting index curve (dRI-curve) in blue and the molecular weight peaks in green. On the left Y axis is indicated the light scattering while on the right Y axis the molecular weight of the sample is shown. The experimental molecular weight (*Mwobs*) was estimated to be 16 kDa, which is higher than the one calculated *Mw* (*Mwcalc*) for each of them (13.5 kDa). The Mw error estimation could be due to co-elution of some small amount of DNA dimers or higher order oligomers. These contaminants were the origin of the peaks at 10.5 ml elution volume. The extra shoulder in LSP corresponds to the G-quadruplex form of LSP (involving four strands (7)). For TFAM, the two molecular weight peaks correspond to the protein dimer (in green) and monomer (in purple). Bottom: four analytical ultracentrifugation assays corresponding to Site-Y, Site-X, LSP and TFAM (same order as for SEC-MALLS) alone are shown.



Supplementary Figure S11. Cysteine labelling and FRET experiment. (A) The positions of the cysteines used for labelling TFAM-Box1Mut (Cys49) or full-length TFAM-C49A (Cys246) are indicated. **(B)** Fluorescence emission spectra of single-labelled TFAM-C49A*/Site-X (dark blue curve) and double-labelled TFAM-C49A*/Site-X5* (in green), and TFAM-C49A*/Site-X3* (in sky blue) complexes excited at 495 nm (a.u., arbitrary units). **(C)** Same as in (A) but for complexes TFAM-C49A*/Site-Y (dark red curve), TFAM-C49A*/Site-Y 5* (red curve) and TFAM-C49A*/Site-Y 3* (in violet).

Supplementary Materials and Methods

FRET calculations

The energy transfer (ET) was calculated with the formula ET= [(S2-S1·(D2/D1)-S4·(A3/A4))·b]/[S4·A3/A4)]·a (8). S2 is the emission of the acceptor during excitation of the donor (FRET effect, P*/D* sample excited at 495 nm, scan 604 to 700 nm). S1·(D2-D1) subtracts from S2 the donor emission tail that spontaneously adds to the acceptor emission peak, and this value is obtained by scaling the donor emission peak in the FRET experiment (S1) with a quotient of the donor-only (P*/D) emissions at both donor (D1) and acceptor (D2) peak regions (D1 values averaged at 511–525 nm, and D2 at 610–624 nm). An additional subtraction was performed, S4·(A3/A4), of the *spontaneous* (not FRET) emission of the acceptor via direct excitation of the acceptor with 495 nm photons. This was estimated by scaling the P*/D* emission peak at 590 nm excitation (S4) by the quotient of average emission spectrums of the P/D* sample at the 610–624 nm range when excited at 495 nm (A3) or 590 nm (A4). The corrected FRET value is normalized by S4·(A3/A4). In addition, normalization by the excitability of the two dyes at 495 nm is achieved with the absorbance values of the complexes, in which $a = OD^{495} - b$ and

b=OD^{590}·A3/A4. The error in the FRET calculation was estimated to be 0.01, the errors in the r distances were calculated according to the Förster relation (i.e. based on the minimum and maximum FRET values).

Supplementary References

1. 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.

2. 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. *Nat Commun*, **5**, 3077.

3. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography*, **60**, 2126-2132.

4. 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.

5. 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.

6. Uchida, A., Murugesapillai, D., Kastner, M., Wang, Y., Lodeiro, M.F., Prabhakar, S., Oliver, G.V., Arnold, J.J., Maher, L.J., Williams, M.C. *et al.* (2017) Unexpected sequences and structures of mtDNA required for efficient transcription from the first heavy-strand promoter. *Elife*, **6**.

7. Lyonnais, S., Tarres-Sole, A., Rubio-Cosials, A., Cuppari, A., Brito, R., Jaumot, J., Gargallo, R., Vilaseca, M., Silva, C., Granzhan, A. *et al.* (2017) The human mitochondrial transcription factor A is a versatile G-quadruplex binding protein. *Sci Rep*, **7**, 43992.

8. Toth, K., Brun, N. and Langowski, J. (2001) Trajectory of nucleosomal linker DNA studied by fluorescence resonance energy transfer. *Biochemistry*, **40**, 6921-6928.