Supplementary Materials

Appendix I. DNA reference frames and conventional coordinates

Analysis of stress fields and molecular deformation was carried out with our in-home NA-BAKE code, which allows to extract the principal components of the stress, compute invariants and other derived quantities, compare stress fields from different simulations, and write the outputs in the portable Gaussian-cube format for visualization. The code also includes the set of subroutines CURVES+, [40] for a detailed analysis of the DNA helical parameters and coordinates. These latter are defined according to the so-called Tsukuba conventional reference frame (Suppl.Fig.S1a), with the three orthonormal basis vectors $\{\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}\}$ such that $\mathbf{e_3}$ is the tangent unit vector along the center line defining the DNA helical symmetry axis, and $\mathbf{e_1}$ is the unit vector pointing at the major groove.

It is however worth noting that the base-pair reference frame defined in Ref. [39] and used by CURVES+ can become extremely distorted, given the large deformations between bases observed in the present work. For this reason, NA-BAKE also uses a second, "smoother" Cartesian reference frame, centered at the midpoint of each pair of P atoms in a base-pair (Suppl.Fig.S1b), the phosphate backbone being more rigid than the stacked base-pairs. At each midpoint, the three basis vectors are defined as follows: the tangent vector τ lies on the direction joining two midpoints in adjacent base-pairs; the normal vector $\mathbf{n}$ points at the center of the major groove; and the binormal vector is $\mathbf{b} = \mathbf{n} \times \tau$. (This can be seen as the approximation of a Frenet-Serret continuous frame for the broken polyline connecting the midpoints of the base-pairs.) The origin of this reference frame does not coincide with the $\{\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}\}$, which depends on the type of base-pair, whereas in the P-centered frame this is uniquely defined as the mid-point along the P-P ideal segment; the origin is also shifted along $\tau || \mathbf{e_3}$ by about half a repeat distance ("rise") with respect to the Tsukuba base-pair plane.

The ensemble of midpoints ordered according to the bp numbering, defines a pathlength $0 \leq s \leq S_0$ along the DNA contour length. For a strand of N bp, with midpoints defined by the set of position vectors $\{\mathbf{r}_1, ..\mathbf{r}_N\}$, s is a sum of discrete segments (a polyline):

$$S_0 = \sum_{i=1}^{N-1} \mathbf{s}_i = \sum_{i=1}^{N-1} \mathbf{r}_{i+1} - \mathbf{r}_i \tag{1}$$

The pathlength vectors $\mathbf{s}_i$ are parallel to the local tangent τ. The variation of the three vectors $\mathbf{n}(s)$, $\mathbf{b}(s)$, $\tau(s)$ along the DNA pathlength can be used to identify the local bending and torsion of the line. From the discrete variation we compute three Euler angles, defining the 3D rotation of the local reference frame between two bp i and $i+1$; in NA-BAKE this is done by using quaternion algebra.

Because of the double-helix geometry there are two bending modes with different elastic constants A and B (stiffness, usually given in units of nm^{-1}), i.e. the DNA polyline has an effective thickness, and is anisotropic. By looking at the definitions of the helical parameters, it turns out that the variation of $\mathbf{n}(s)$ corresponds to the "tilt" θ, that of $\mathbf{b}(s)$ corresponds to the "roll" ρ, and that of $\tau(s)$ corresponds to the "twist". The first two parameters also give the local curvature between consecutive bp along the line, as

$$\kappa = \frac{1}{s}\sqrt{\theta^2 + \rho^2} \tag{2}$$

However, such a definition of "strictly local" curvature may miss the ample bending deformations that extend over lengths longer than just 1-2 bp (see Suppl.Fig.S1c). For this reason, we introduced a more geometric-minded, "global" notion of curvature, by calculating the radius R of the best fitting circle to a

series of midpoints along s, projected in the best-fitting common plane. Note that the direction vector of R does not necessarily coincide with $\mathbf{b}$, which instead varies from point to point: this is the very reason to introduce a "global", rather than "local" notion of curvature. The numerical procedure is described in the following Appendix II. In practice, the geometric curvature $\kappa = 1/R$ is estimated by computing R for a series of equal lengths $S_n = \sum_{i=k}^{k+n} \mathbf{s}_i$, the starting point k spanning the whole DNA length S_0; each set overlaps with the next one by ± 1 midpoint, therefore the values of $\kappa(s)$ behave quite smoothly for a sufficiently large n. In the calculations shown in Suppl.Fig.S1 we used n=5-8, such that each set spans a length approximately equal to one half-turn of the double helix. However, by using larger values for averaging, large bending movements can be captured. Suppl.Fig.S1c shows an example of curvature calculated with n=25 points.

Appendix II. Numerical approximation of global curvature

We describe the method to fit a circle to a cluster of points in 3D space, used to estimate the global curvature along the DNA backbone. Consider a set of n points $\{P_1, .., P_n\}$, where $P_i = (x_i, y_i, z_i)^T \in R^3$, obtained by extracting the centers of the pairs of P atoms of each bp along the DNA backbone. For a DNA segment containing N base-pairs, we consider subsets of $n < N$ consecutive points, typically $n \simeq 6-12$, covering the whole DNA length, possibly with some overlap between adjacent subsets. We want to find a circle that fits as close as possible to each subset of points. The circle fitting method can be split into three steps:

1. use the least-squares method to find the best fitting plane to the set of points;
2. project all the points perpendicularly onto the fitting plane in 2D;
3. use again the least-square method, to fit a circle in the 2D plane to the set of projected points, and obtain the circle center and radius.

Eventually, the circle center can be transformed back to 3D coordinates, if we want to collect the ensemble of curvatures in the common reference frame other than just the scalar value of κ. We will use two different implementations of the least-squares method to obtain the radius of the best fitting circle.

Step1. Given the subset of n points in 3D, the best plane can be found by a least-squares method that minimizes the distance of each point to the plane. The equation for a generic plane is: $ax + by + c = z$. Hence, we build an eigenvalue problem like $\mathbf{Ax} = \mathbf{B}$, or:

$$\begin{pmatrix} x_0 & y_0 & 1 \\ x_1 & y_1 & 1 \\ .. & .. & .. \\ x_n & y_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} z_0 \\ z_1 \\ .. \\ z_n \end{pmatrix} \tag{3}$$

Then, solve for the $\mathbf{x}$ vector of coefficients. However, since each subset is made up by more than just 3 points, the system is over-determined. Therefore, we use the left pseudo-inverse matrix: $A^+ = (A^T A)^{-1} A^T$. Finally, the coefficients of the plane are found as:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = (A^T A)^{-1} A^T B \tag{4}$$

Note that such a method is not entirely general. A more robust method should be to subtract out the centroid of the subset of points, to form a $n \times 3$ matrix with the resulting coordinates, and calculate its singular value decomposition. As a result, the normal vector of the best-fitting plane should be found as the left singular vector corresponding to the least singular value. However, we preferred the method of least squares because of its simpler implementation in this case.

Step2. Once the best fitting plane is obtained, the 3D coordinates of the subset of points are projected onto this plane according to their respective perpendicular distances, thus obtaining a 2D representation of the same subset in the new coordinates.

Step3. The implicit equation in 2D for a circle with radius R and center $(x_c, y_c)^T$ can be arranged as:

$$\begin{aligned} (x-x_c)^2 + (y-y_c)^2 &= R^2, \\ 2x_c x + 2y_c y + (R^2 - x_c^2 - y_c^2) &= x^2 + y^2, \\ c_1 x + c_2 y + c_0 &= x^2 + y^2 \end{aligned} \tag{5}$$

with $c = (c_1, c_2, c_0)^T$ the vector of unknown parameters. Then, by applying this definition to all the input points P_i, it yields to a system of linear equations:

$$\mathbf{Ac} = \mathbf{b} \tag{6}$$

with:

$$\mathbf{A} = \begin{pmatrix} x_1 & y_1 & 1 \\ .. & .. & .. \\ x_n & y_n & 1 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} x_1^2 + y_1^2 \\ .. \\ x_n^2 + y_n^2 \end{pmatrix}, \tag{7}$$

Since there are more equations than unknowns, an approximate solution is obtained by the method of least-squares, which minimizes the squared sum of the residuals $||\mathbf{b} - \mathbf{Ac}||^2$; we implemented for this purpose the family of subroutines QR_SOLVE from LINPACK. Therefore, the center of the fitting circle is $x_c = c_1/2$, $y_c = c_2/2$, and the radius is $R = (c_0 + x_c^2 + y_c^2)^{1/2}$. Then, the line curvature for the subset $\{P_1, .., P_n\}$ is just $\kappa = 1/R$.

Appendix III. Stress calculation for atomistic systems

In classical macroscopic continuum elasticity, stress is constructed as a continuous field at each point $\mathbf{r}$ in a homogeneous domain where a distribution of forces exists, namely a 3x3 tensor, $\sigma_{\alpha\beta}(\mathbf{r})$, with the dimensions of a force per unit surface. In practice, for a force vector $\mathbf{f}$ and a boundary S with normal unit vector $\mathbf{n}$, the Cartesian components of force and stress are linked as:

$$f_\alpha = \sum_\beta \int_S (\sigma_{\alpha\beta} \cdot n_\beta) dS, \qquad \alpha, \beta = x, y, z \tag{8}$$

In atomistic simulations, however, we deal with material points exerting forces across empty space. Hence, the definition of the analogous of stress at the atomic level is complicated by several issues, such as: discretization of a continuum field; what is the volume around each atom; how to precisely define a boundary surface; last but not least, the fact that atoms have a velocity [66, 67]. With a much simplified notation for the sake of clarity, Hardy's definition of a continuous stress that can be mapped onto atomic

positions, velocities and forces (that is, accelerations) of a system of N point particles ("atoms") can be written as [68]:

$$\sigma_{\alpha\beta}(\mathbf{r}) = -\sum_{i<j=2}^{N}(r_{ij}^{\alpha}\cdot f_{ij}^{\beta})b_{ij}(\mathbf{r}) + \frac{1}{2}\sum_{i=1}^{N} m_i(v_i^{\alpha}\cdot v_i^{\beta})g_{ij}(\mathbf{r}) \tag{9}$$

where $\mathbf{r}_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the scalar distance between atoms i, j, and $\mathbf{f}_{ij}$ is the force on atom i from any other atom j; the b_{ij} and g_{ij} are spatial weighting functions (usually derived by variational arguments). The first term in the equation represents the potential (or "virial") contribution, and the second term is the kinetic contribution to the molecular stress.

In this work we use the recently developed covariant central-force decomposition scheme (CCFD) for the intra- and intermolecular forces [65], which is based on thermodynamical arguments rather than on momentum conservation. We aim to extract such information from our simulations and couple it to the state of deformation, along the lines of our first application of this method to a single nucleosome [51].

In the condition of uniaxial loading used in this work, the component of the stress parallel to the loading condition, $\sigma_{zz}(\mathbf{r})$, may be a first, interesting quantity, at least at the beginning of the simulation; in the early stage of loading, the compressive force (directed along the line joining the centers of mass of the groups C and B, in Fig. 1a) is parallel to the z-direction. Therefore, the $\sigma_{zz}(\mathbf{r})$ component is practically coincident with the σ_3 principal component of the diagonal stress.

At later times, however, the configuration quickly becomes so much deformed that other components of the stress tensor describe the redistribution of the internal forces in response to the compression. Next to the purely compressive component, also mixed transverse and shearing components arise. A quantity that can be useful to characterise the complex deformations occurring, instead of looking at the different components one by one, is the distribution of deviatoric, or Von Mises, stress:

$$\varepsilon^{dev}(\mathbf{r}) = \frac{1+\nu}{3E}\left[\frac{1}{2}\sum_{i<j}(\sigma_{ii}-\sigma_{jj})^2 + 3\sum_{i<j}\sigma_{ij}^2\right] \tag{10}$$

where all the $\sigma_{ij} = \sigma_{ij}(\mathbf{r})$, and E=300 MPa and ν=0.4 are typical values for the Young's modulus and Poisson's ratio for straight DNA, which fix the energy scale. Note that by definition this quantity is always positive, although the various stress components may have both negative and positive values.

At a more global scale, another important quantity that can be derived from the local stress tensor is the density of elastic energy at each point in space:

$$\varepsilon^{el}(\mathbf{r}) = \frac{1}{E}\sum_{i,j}\left[](\sigma_{ii})^2 - 2\nu(\sigma_{ii}\sigma_{jj}) + \frac{1+\nu}{2}(\sigma_{i\neq j}+\sigma_{j\neq i})^2\right] \tag{11}$$

By looking at the time evolution of this elastic energy at different positions in space, it can be observed how the elastic energy, progressively stored by the constant loading force, is redistributed.

	force (pN)	start filename	compress (ns) filename	relax (ns) filename
1	40	n183.pdb	12 183_40.pdb	
2	85		12 183_85.pdb	
3	125		100 183_125.pdb	150 183_125_rel.pdb
3’	125		15 183_125s.pdb	
4	150		20 183_200.pdb	
5	400		15 183_400.pdb	
1	40	n169.pdb	10 169_40.pdb	
1*	40	n169f.pdb	10 169_40f.pdb	
2	85		10 169_85.pdb	
2*	85		10 169_85f.pdb	
3	125		15 169_125.pdb	200 169_125_rel.pdb
3*	125		15 169_125f.pdb	200 169_125f_rel.pdb
4	150		15 169_200.pdb	
4*	150		25 169_200f.pdb	
5	400		15 169_400.pdb	

Table 1. Summary of the steered-MD simulations presented in this work. The upper and lower blocks correspond to the T183 and T169, respectively. For the latter, the ’*’ indicates configurations produced with alternate H3 tails initial configurations. The name of the pdb file containing the final configuration of the run, and available from the repository, is indicated whenever available. Note that water and ions are not included in the pdb.

T183

TACGTAATATTGGCCAGCTAGGA
TATCACAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAA
CGCACGTACGGA**A**TCCGTACGTGCGTTTAAGCGGTGCTAGAGCTGTCTACGACCAATTGA
GCGGCCTCGGCACCGGGATTGTGATA
TCCTAGCTGGCCAATATTACGTATGGCCAGCTAGGA
TATCACAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAA
CGCACGTACGGA**A**TCCGTACGTGCGTTTAAGCGGTGCTAGAGCTGTCTACGACCAATTGA
GCGGCCTCGGCACCGGGATTGTGATA
TCCTAGCTGGCCATACGTAATATTGGCCAGCTAGGA
TATCACAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAA
CGCACGTACGGA**A**TCCGTACGTGCGTTTAAGCGGTGCTAGAGCTGTCTACGACCAATTGA
GCGGCCTCGGCACCGGGATTGTGATA
TCCTAGCTGGCCAATATTACGTA

T169

ATCAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAA
CGCACGTACGCGC**T**GTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAG
GCACGTGTCAGATATATACATCGATTG
GATAGGCCCGGACGGCCTGGAT
ATCAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAA
CGCACGTACGCGC**T**GTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAG
GCACGTGTCAGATATATACATCGATTG
GATAGGCCCCAACGGCCTGGAT
ATCAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAA
CGCACGTACGCGC**T**GTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAG
GCACGTGTCAGATATATACATCGATTG

Widom-601

ATCGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAA
ACGCACGTACGCG**C**TGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCA
GGCACGTGTCAGATATATACATCCGAT

Table 2. DNA sequences of the two trinucleosome systems employed in this work; the original Widom-601 sequence is also reported for comparison. Red letters indicate the DNA linker regions; blue letters indicate the hanging DNA leads; boxed letters indicate the dyad.

H3 (96,5% identity 98% positives)

Human	1	MARTKQTARKSTGGKAPRKQLATKAARKS **T** PST **C** GVK–PHRYRPGTVALREIRRYQKSTE	59
Xenop	1	MARTKQTARKSTGGKAPRKQLATKAARKS **A** PST **G** GVKKPHRYRPGTVALREIRRYQKSTE	60
Human	60	LLIRKLPFQRLVREIAQDF **N** TDLRFQSAA **V** GALQEASEAYLVGL **L** EDTNLCAIHAKRVTI	119
Xenop	61	LLIRKLPFQRLVREIAQDF **K** TDLRFQSAA **I** GALQEASEAYLVGL **F** EDTNLCAIHAKRVTI	120
Human	120	MPKDIQLARRIRGERA	135
Xenop	121	MPKDIQLARRIRGERA	136

H4 (100% identity)

Human	1	MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLK	60
Xenop	1	MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLK	60
Human	61	VFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG	103
Xenop	61	VFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG	103

H2A (93% identity 95% positives)

Human	1	MSGRGKQGGK **A** RAKAK **S** RSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVY **M** AAV	55
Xenop	1	MSGRGKQGGK **T** RAKAK **T** RSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVY **L** AAV	55
Human	56	LEYLTAEILELAGNAARDNKKTRIIPRHLQLA **I** RNDEELNKLLG **K** VTIAQGGVLPNIQ **A**	114
Xenop	56	LEYLTAEILELAGNAARDNKKTRIIPRHLQLA **V** RNDEELNKLL **G** GVTIAQGGVLPNIQ **S**	114
Human	115	VLLPKKTES **HHK** AK **G** K	130
Xenop	115	VLLPKKTES **AKS** AK **S** K	130

H2B (93% identity 98% positives)

Human	1	MPEP **S** KSAPAPKKGSKKA **I** TK **A** QKKDGKKR **KR** SRKESY **S** IYVYKVLKQVHPDTG	54
Xenop	1	MPEP **A** KSAPAPKKGSKKA **V** TK **T** QKKDGKKR **RK** SRKESY **A** IYVYKVLKQVHPDTG	54
Human	55	ISSKAM **G** IMNSFVND **I** FERIAGEASRLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSE	114
Xenop	55	ISSKAM **S** IMNSFVND **V** FERIAGEASRLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSE	114
Human	115	GTKAVTKYTS **S** K	126
Xenop	115	GTKAVTKYTS **A** K	126

Table 3. Histone sequence comparison between human (used in this work) and *Xenopus laevis* (from the 1KX5 experimental PDB dataset). Blue letters indicate conservative (positive) replacement, red letters non-conservative.

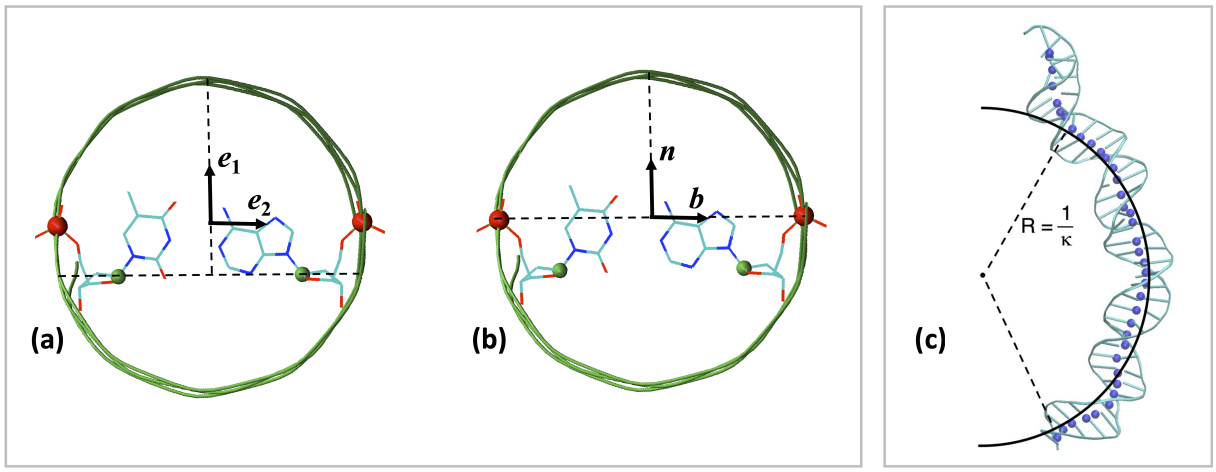

Figure S1. Conventional definition of (a) the basis-centered reference frame (Tsukuba set) used in the CURVES+ module, and (b) the second phosphate-centered reference frame also used in this work (NA-BAKE module). The two green spheres are the C1' carbon atoms of the base-pair, while the two red spheres indicate the P atoms. The P-centered frame is shifted by a half-rise (0.17 nm) along $\mathbf{e}_3$ with respect to the first one. (c) Example of calculation of the curvature from the procedure described in Suppl.Mat.Appendix I, as the inverse of the radius of the circle best-fitting the ensemble of midpoints (blue spheres) for a bent DNA linker.

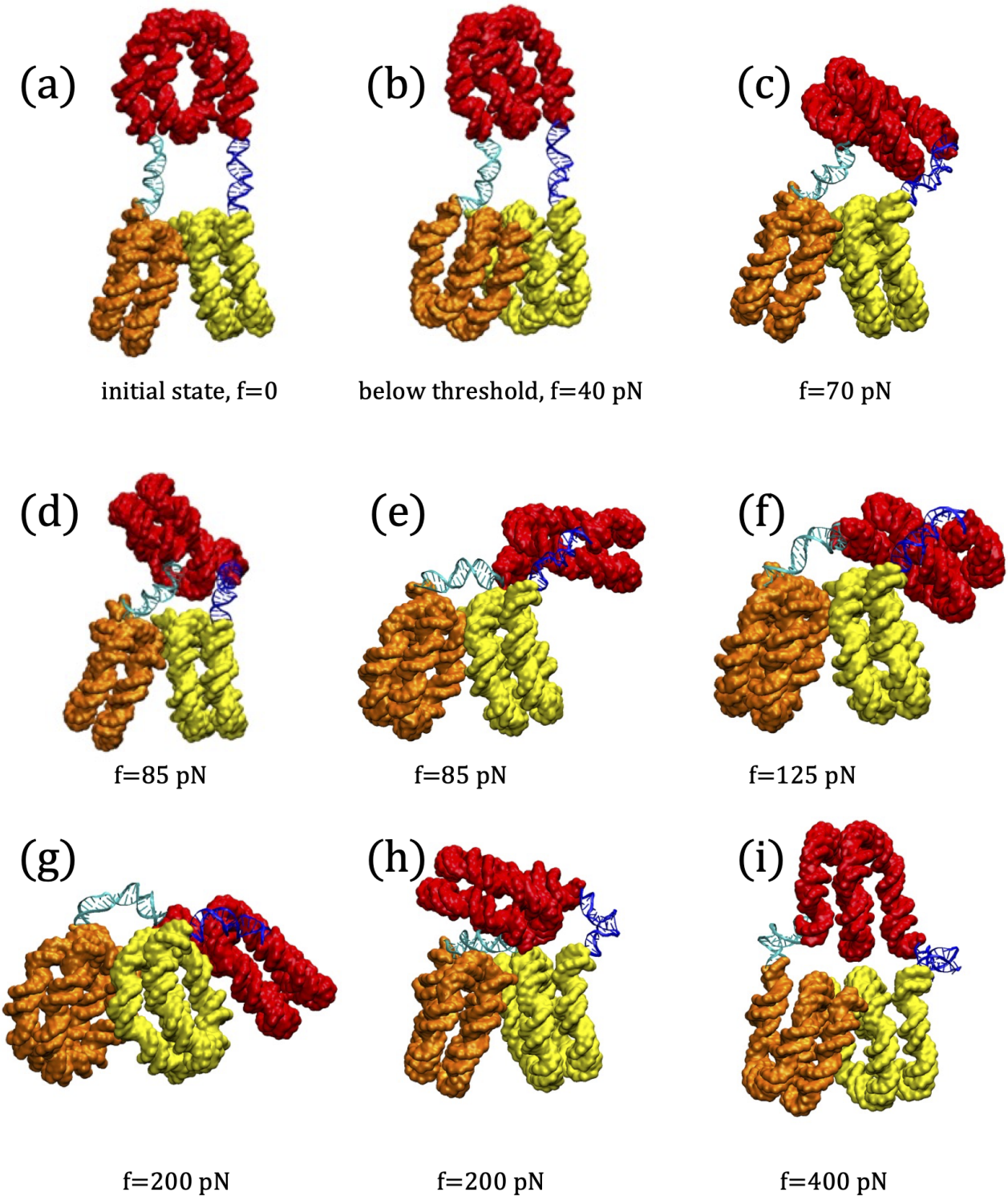

Figure S2. (a) Initial state for the T169 simulations. (b)-(i) Configuration at t=10 ns. Only DNA shown, same color codes as Fig.1-2 for nucleosomal and linker DNA.

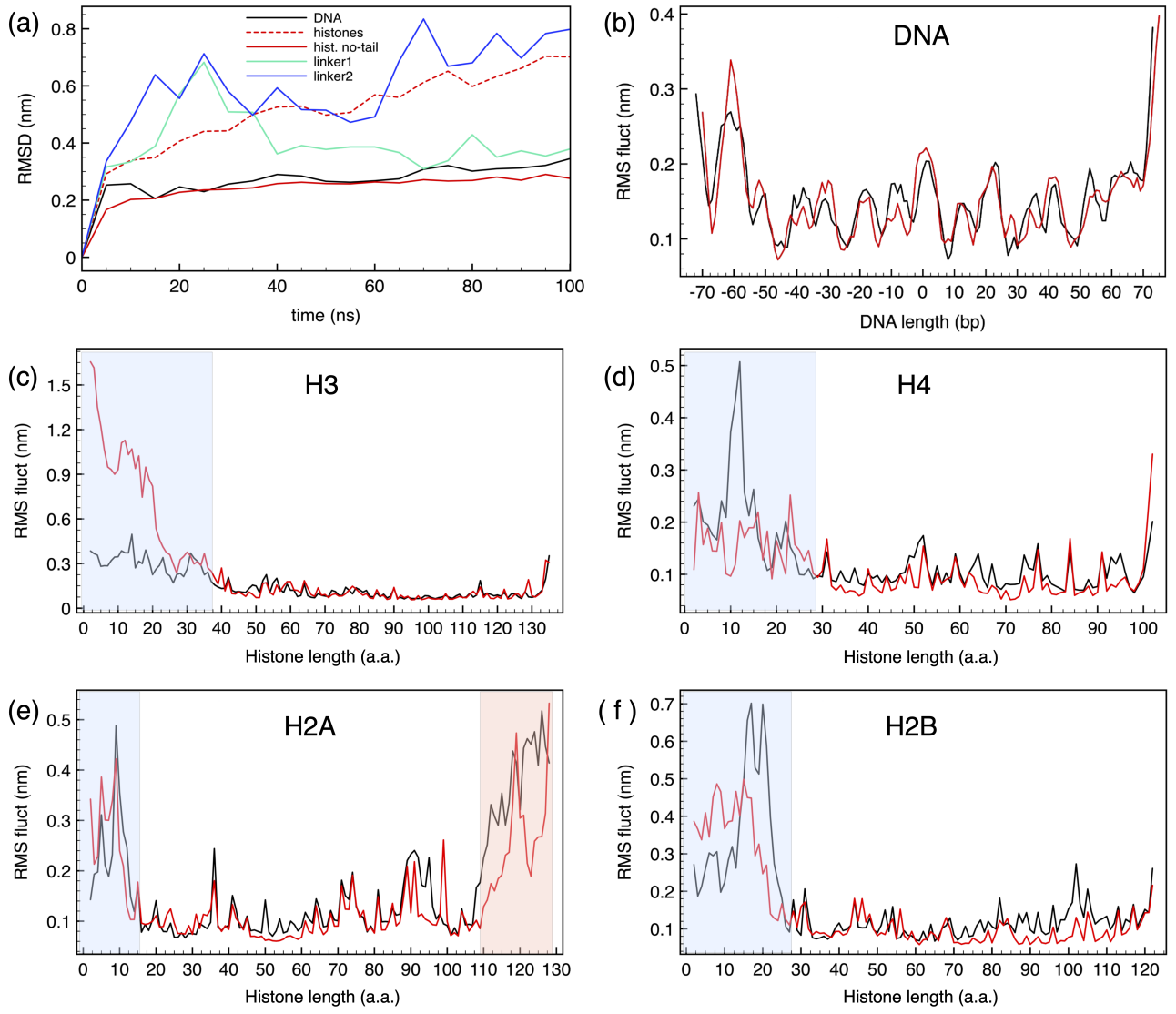

Figure S3. MD compression simulation of the T183 system at F=125 pN. (a) Root-mean-squared deviation (RMSD) from the initial configuration as a function of time (compression, 0 to 100 ns), for the different components of the central C nucleosome to which the force is applied. (b) Root-mean-squared fluctuation (RMSF) per residue, averaged over the compression time (0-100 ns) for the nucleosomic DNA (in red/black the two strands, 147 bp ranging from -73 to +73, 0 being the dyad); (c-f) same as (b), for the H3, H4, H2A, H2B pairs (red/black) of histones; N-terminal tail regions indicated by light-blue shading, C-terminal tail regions by light-red shading.

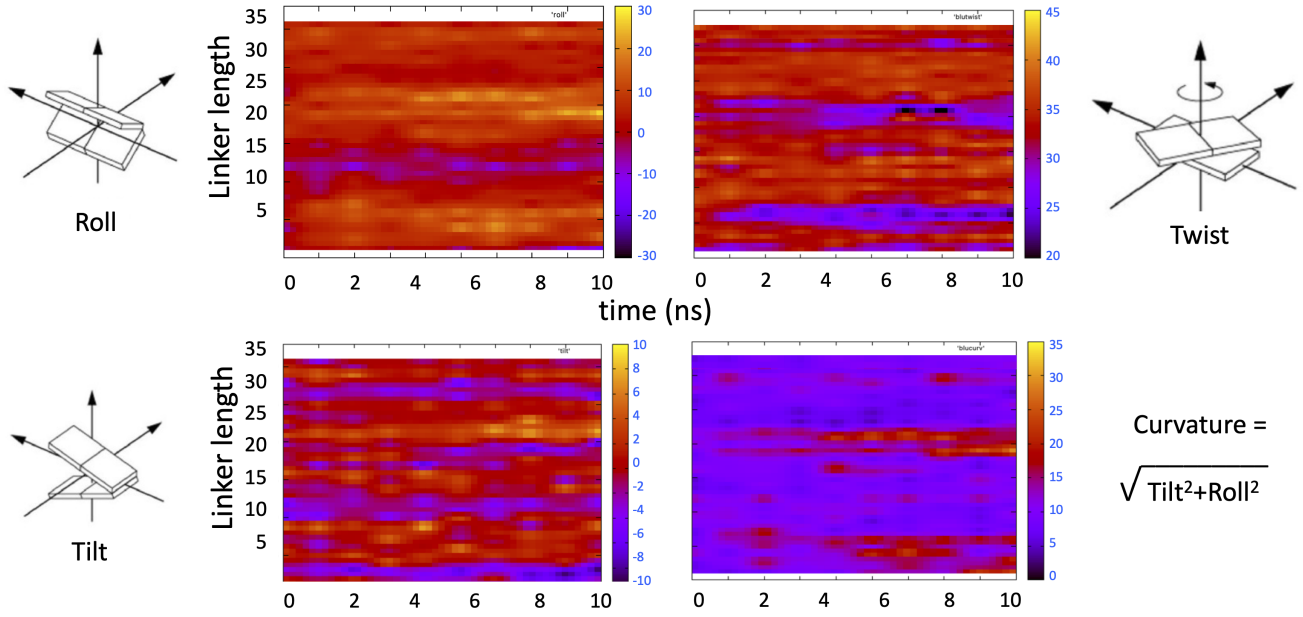

Figure S4. Time-plots for the roll (top left), twist (top right), tilt (bottom left) intra-bp helical parameters, and the resulting curvature (bottom right), for the R (blue) DNA linker of Fig.2b. See Fig3 of the main text for nomenclature.

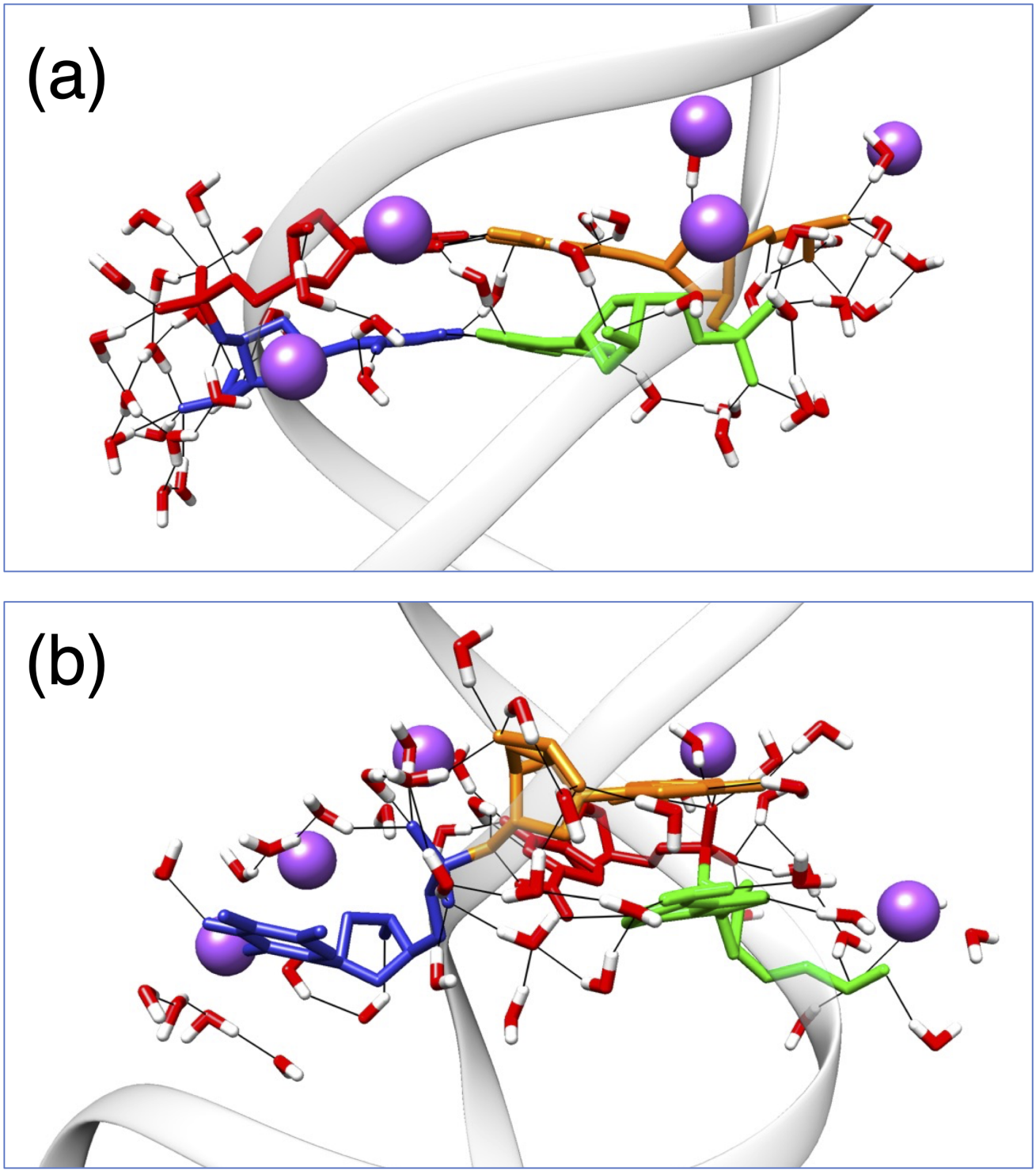

Figure S5. Schematic of the hydrogen-bonded water network surrounding an adjacent C-G/A-T dinucleotide sequence in the T183 linker. (a) Configuration of a normally-stacked pair. (b) Configuration of the flipped-out T95 base (see also Fig.5 in the main text). Color code: T blue, A green, G orange, C red licorice-sticks; Na ions, purple spheres; water, red-white sticks. The grey ribbons in background depict the local arrangement of the DNA backbone.

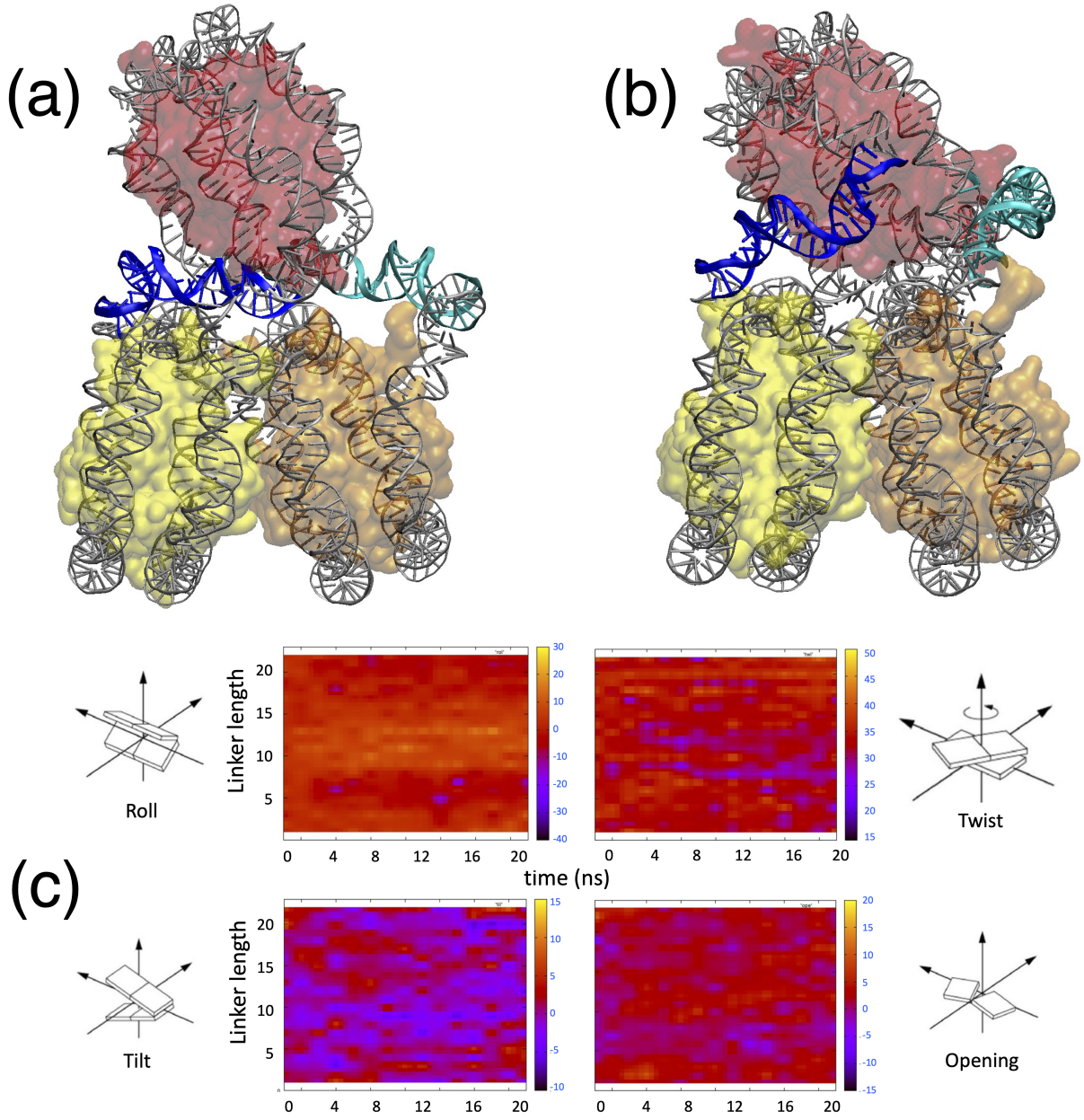

Figure S6. Results of MD simulations of the T169 without histone tails. (a) Final configuration after 20 ns compression at 50 pN. (b) Final configuration after 20 ns compression at 100 pN. (c) Main helical parameters along the 22 base-pairs of the cyan DNA linker for the 50 pN simulation.

Supplementary Movie 1 - Slow compression (100 ns) of the T183 trinucleosome, followed by free relaxation (150 ns), at a force of 125 pN. The two linkers are represented in cyan and blue, as in the main text. The red nucleotide on the cyan linker is the thymine-95, which flips out in extrahelical position some time after the beginning of compression, and induces the kinking (Brazier-like instability). Water molecules and ions not shown.

Supplementary Movie 2 - Comparison of the rapid compression (10 ns) of the T169 trinucleosome, followed by free relaxation (200 ns), at a force of 125 pN. The left and right panels correspond to two identical simulations, starting with different initial configurations of the H3 histone tails. Only the H3 histone tails implicated in the contacts are shown as colored surfaces (red, orange and yellow for the C,A1,A2 nucleosome) plus the H4 tail (grey) from C nucleosome, while the rest of the proteins is shadowed; water molecules and ions are not shown.