Electronic Supplementary Material (ESI) for Nanoscale.
This journal is © The Royal Society of Chemistry 2016

Electronic Supporting Information

Force measurements reveal how small binders perturb the dissociation mechanisms of DNA duplex sequences

Anastasia Burmistrova, Barbara Fresch, Damien Sluysmans, Edwin De Pauw, Françoise Remacle and Anne-Sophie Duwez**

1. AFM force spectroscopy experiments

We have optimized the density of the molecules grafted on the surface and on the tip to have a very low probability of pulling more than one duplex at the same time. The lower is the concentration of molecules on the surface, the lower is the probability to probe more than one duplex, but also the higher is the probability to probe no molecule at all. It is a question of compromise between limiting the number of molecules interacting with the tip and keeping a reasonable probability of probing one molecule.

Most of the force curves (~95%) revealed flat profiles without peaks, indicating that no molecule had been probed. Complex profiles with multiple peaks were sometimes observed (~0.5% of cases), and were not considered. In ~3-5% of the experiments, a single peak was observed. Only such profiles were considered for further analysis. Histograms of the rupture length distributions are shown in Figure S1. The sharp position of the rupture events is evidenced and proves the specificity of these events.

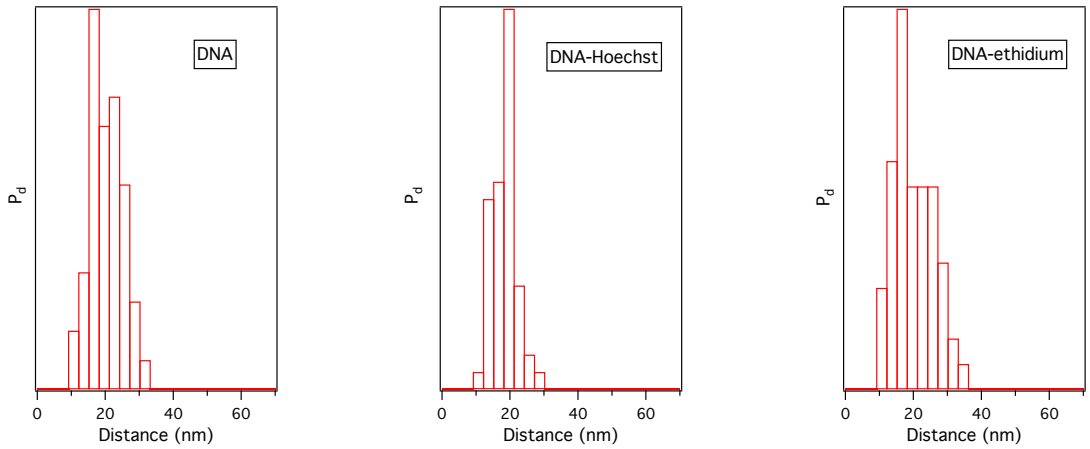


Figure S1: Histograms of the position of the rupture peak for the DNA duplex, DNA-Hoechst and DNA-ethidium complexes.

Each point of the most probable rupture force dependence on the loading rate has been obtained from at least 250 force curves (for low and high loading rates) up to 500-600 force curves (for loading rates around the critical loading rate threshold). 3 or 4 independent experiments were realized for each data point (different tip, different sample, different batch of functionalization, different batch of working medium). For the experiments involving the ligands, we always measured a reference point corresponding to DNA alone and checked the position of this point on the force vs. loading rate plot to detect any experimental inconsistency. Data from multiple experiments show the reproducibility of the results (Figures S2 and S3). We also realized a series of experiments on the three systems (DNA, DNA-Hoechst and DNA-ethidium) with the same cantilever to exclude any influence of spring constant calibration or any other instrumental or technical variations (data in grey in Figure S3).

The most probable rupture forces were obtained by fitting the distribution of raw data (fit of the Probability Density Function and use of a Gaussian mixture model), not by a curve fitting of the histograms whose result is dependent on the bin size used to build the histogram. The procedure is described in details in the next section.

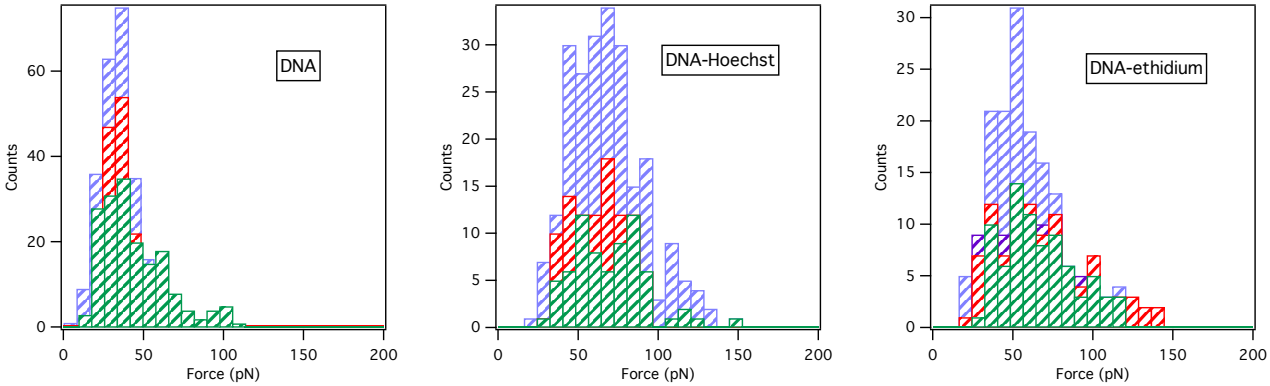


Figure S2: Histograms of the rupture forces from multiple experiments at a loading rate of 3.6 10^4 pN s^{-1} (near the critical threshold).

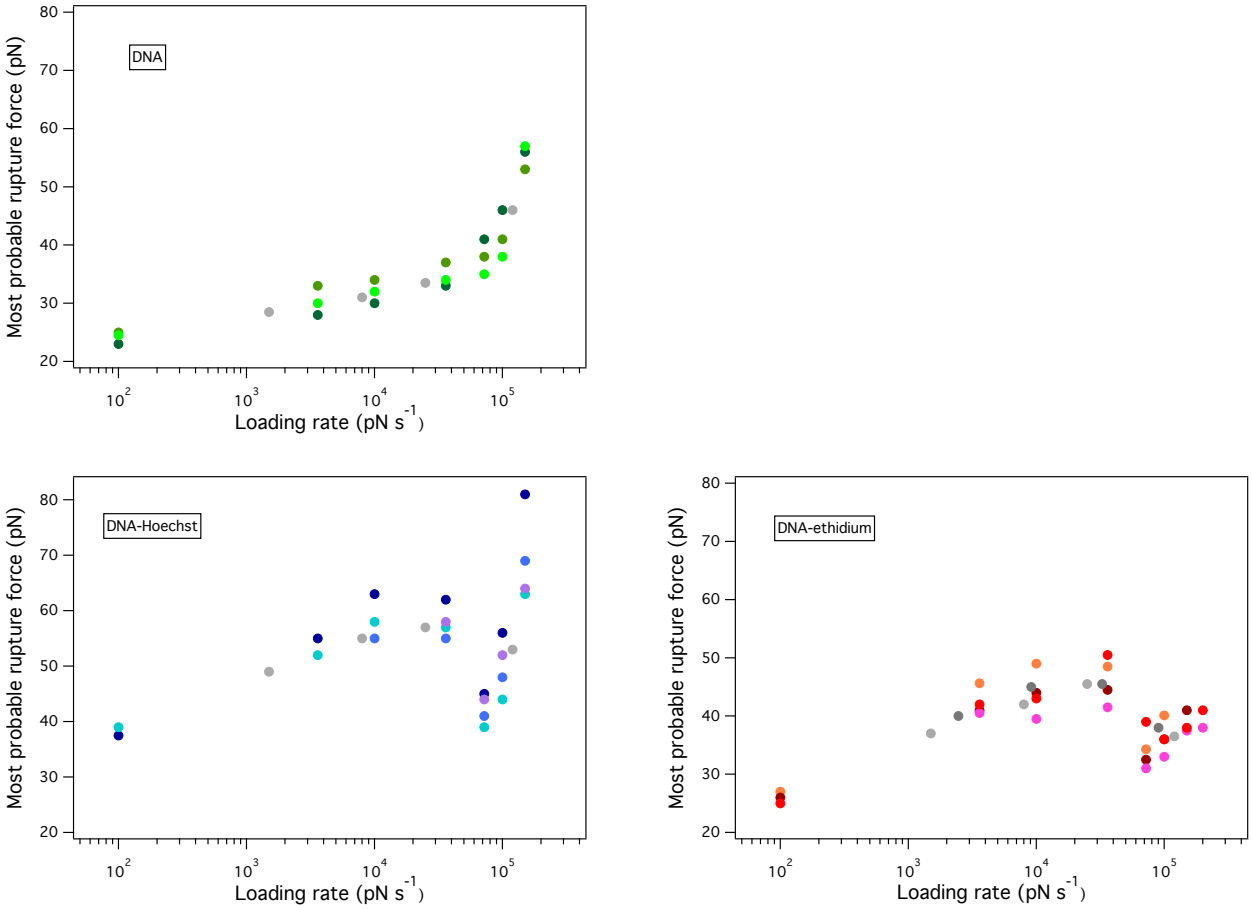


Figure S3: Dynamic force spectroscopy plots of the DNA duplex dissociation, over loading rates from 100 to 2 10^5 pN s^{-1}. Most probable rupture forces obtained from different sets of experiments (each color corresponds to a different independent series of experiment) for DNA, DNA-Hoechst, and DNA-ethidium. The data in grey have been obtained with the same cantilever for the three systems.

2. Statistical analysis of the rupture force data using a Gaussian mixture model

The rupture force data obtained for the DNA, DNA-Hoechst and DNA-ethidium complexes are shown as normalized histograms in Figure S4. The bin size h for the histograms was determined by the Freedman and Diaconis rule[1] based on the interquartile range (IQR=difference of the 75th and 25th percentiles) so that $h_{opt}^{FD} = 2\mathrm{IQR} \cdot \mathrm{N}^{-1/3}$, where N is the size of each data set and it is reported in Table S1. The data sets correspond to experiments realized with only one cantilever to avoid calibration variability issues.

The majority of the collected data sets show a distinct tail to the right, i.e. the force histograms are skewed towards high force. This is a common feature of rupture force measurements and it has been explained by invoking multiple attachment events[2] or heterogeneity in the chemical bonding and dynamical disorder[3,4]. Most probably, all these features contribute to the observed statistics of the rupture force. In order to account for multiple binding and other anomalies, we fit the recorded data with a Gaussian mixture model (GMM), that is, a weighted sum of M component Gaussian densities as given by the equation

$$P(\mathbf{x}|\lambda) = \sum_{i=1}^{M} p_i G(\mathbf{x}|\mu_i, \sigma_i) \qquad (1)$$

where **x** is the vector of the observed rupture forces at a given loading rate, $G(\mathbf{x}|\mu_i, \sigma_i)$ is the normalized Gaussian component with mean μ_i and variance σ_i and p_i is the weight of the i-th component. The weights satisfy the normalization condition $\sum_{i=1}^{M} p_i = 1$, and λ represents the set of all the parameters $\lambda = \{p_i, \mu_i, \sigma_i\}$ for i=1,…M that specify the model. The total probability density in eq.(1) allows the treatment of cases in which the observed data in **x** are sampled from M sub-populations. The mixture model is used to make statistical inferences about the properties of the sub-populations given only observations on the pooled population, without sub-population identity information. The set of parameters $\lambda = \{p_i, \mu_i, \sigma_i\}$ for i=1,…M are determined through the Expectation-Maximization algorithm[5] that maximize the likelihood of the GMM given the observed data through an iterative procedure. The GMM provides a smooth overall distribution fit of the observed data and its components can be used to detail the multi-modal nature of the density.

In practice, we first determined the optimal number of Gaussian components to fit the data sets by calculating the Bayesian (BIC) and the Akaike (AIC) Information Criterion[6] for models with different number of components M=1-4. A two component Gaussian model minimizes the information criteria for the DNA data set of all the loading rates, except for the highest one (150000 pN/s) for which a single Gaussian was sufficient. We thus use the Gaussian component with the higher weight to determine the most probable rupture force, while we identify the second component as describing the population of outliers composed of multiple bond rupture events or instrumental anomalies.

For the DNA-ligand complexes, we first fitted the data with a two-mode Gaussian mixture model for consistency of the analysis procedure used for the DNA data set. However, we noted that the minimization of BIC and AIK is achieved for M=3 for the intermediate loading rates.

Figure S4 shows the obtained two-mode Gaussian density superimposed to the histograms of all the rupture force data sets. In Table S1, we report the weight p_1 of the main component used to determine the most probable rupture force of each data set. In general the main component accounts for more than 60-70% of the overall observed population. For each estimated average rupture force, the ± 95% confidence interval was computed as $\pm 2\sqrt{\sigma^2/(p_1 N)}$ where σ^2 is the estimated variance of the main Gaussian component while $p_1 N$ represents the effective size of the population.

As explained in the main text, the use of the three-component GMM allowed us to gain insights into the sub-populations following the two different dissociation pathways at loading rates bracketing the observed discontinuity in the rupture force-loading rate dependence. In figure S4 we report the total density resulting from the fitting of the three-component GMM together with the two-component GMM for the DNA-Hoechst and DNA-ethidium complexes at loading rates of 36000 and 72000pN/s. These fits are plotted in red. The three

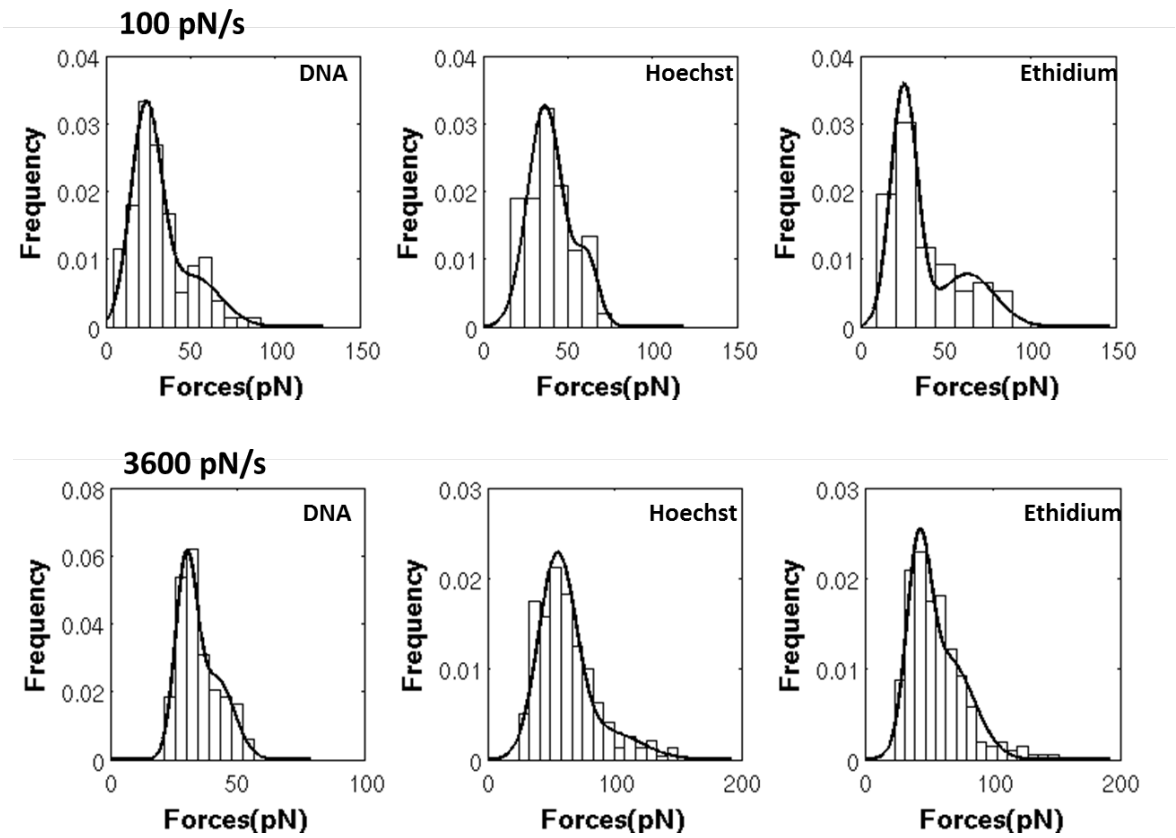
components model suggests the existence of three sub-populations with considerable weight, supporting the co-existence of the two dissociation pathways in the set of measurements at this loading rate.

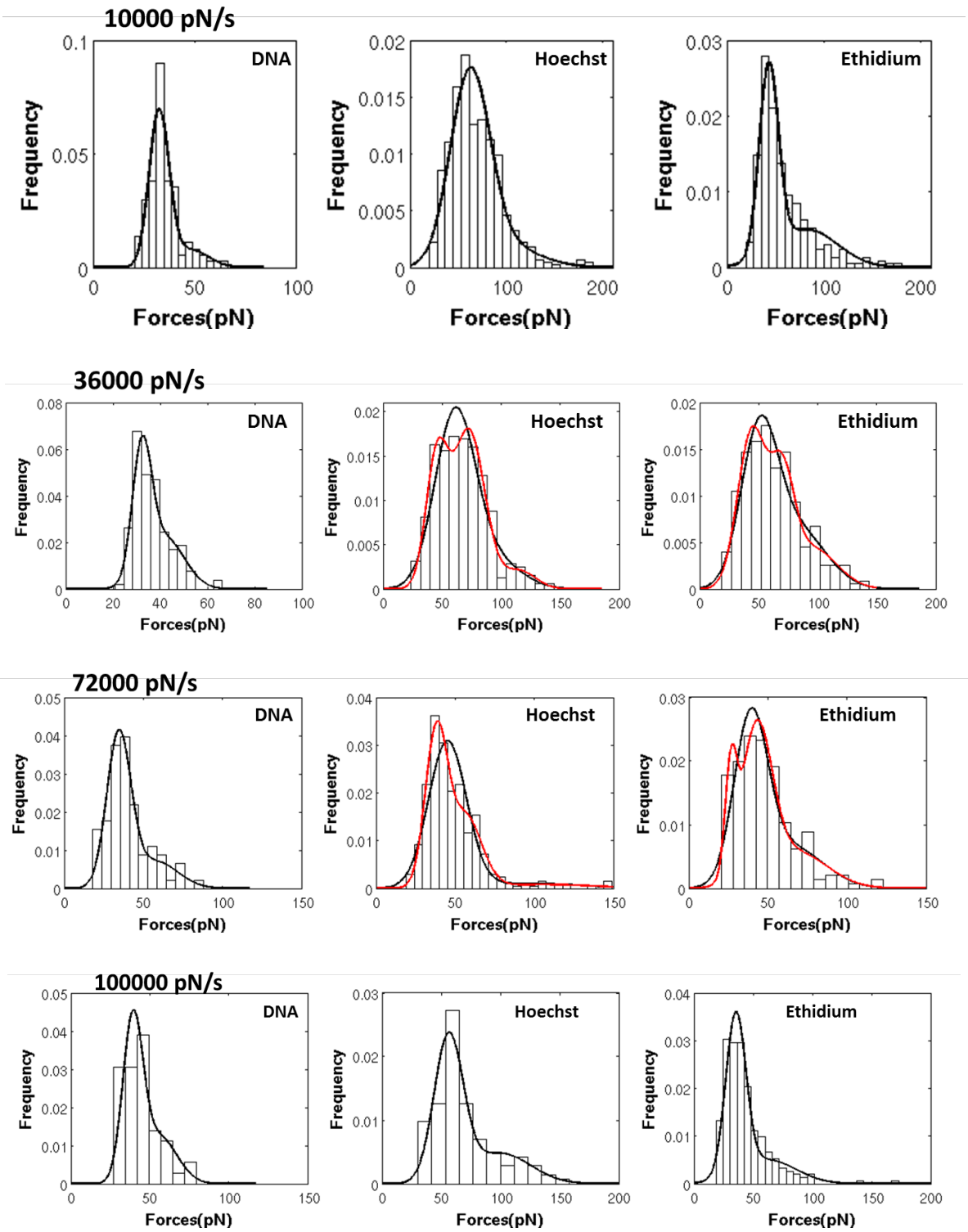
To evidence the two pathways also for the other loading rates and for DNA alone, we finally used the three-component GMM for all the data sets. This three-component GMM nicely describes the data in agreement with the Friddle model (see next section, Figure S5).

Table S1: Weight p_1 of the main component used to determine the most probable rupture force of each data set. The data sets correspond to experiments realized with only one cantilever to avoid calibration variability issues.

LR(pN s^{-1})	DNA		DNA-Hoechst		DNA-ethidium	
	N	p_1	N	p_1	N	p_1
100	108	0.71	62	0.84	67	0.68
3600	110	0.60	287	0.80	252	0.50
10000	104	0.82	533	0.85	455	0.61
36000	140	0.55	397	0.82	424	0.62
72000	78	0.75	398	0.91	200	0.66
100000	48	1	123	0.69	256	0.70
150000	45	1	169	0.94	282	0.25*
200000					186	0.78

* An anomalous small weight of the first Gaussian component (p_1=0.25) resulted for the data set relative to the DNA-ethidium complex at a loading rate of 150000pN/s. The high weight of the high-force tail of this distribution is interpreted as an anomaly resulting from the particular experiment, since the distribution recorded for the further point, at 200000pN/s, displays again a majority population that confirms the trend of the average rupture forces as a function of the loading rate.




10000 pN/s
DNA
Hoechst
Ethidium
36000 pN/s
DNA
Hoechst
Ethidium
72000 pN/s
DNA
Hoechst
Ethidium
100000 pN/s
DNA
Hoechst
Ethidium
Frequency
Forces(pN)

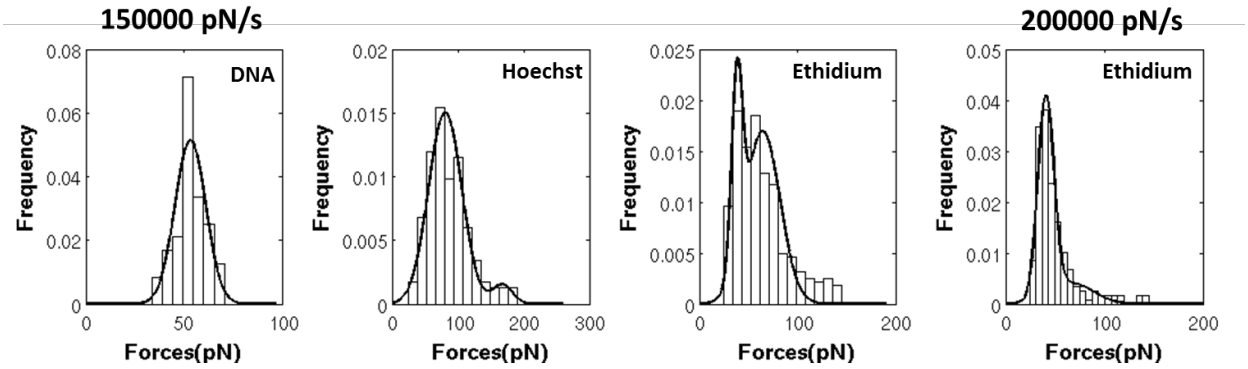


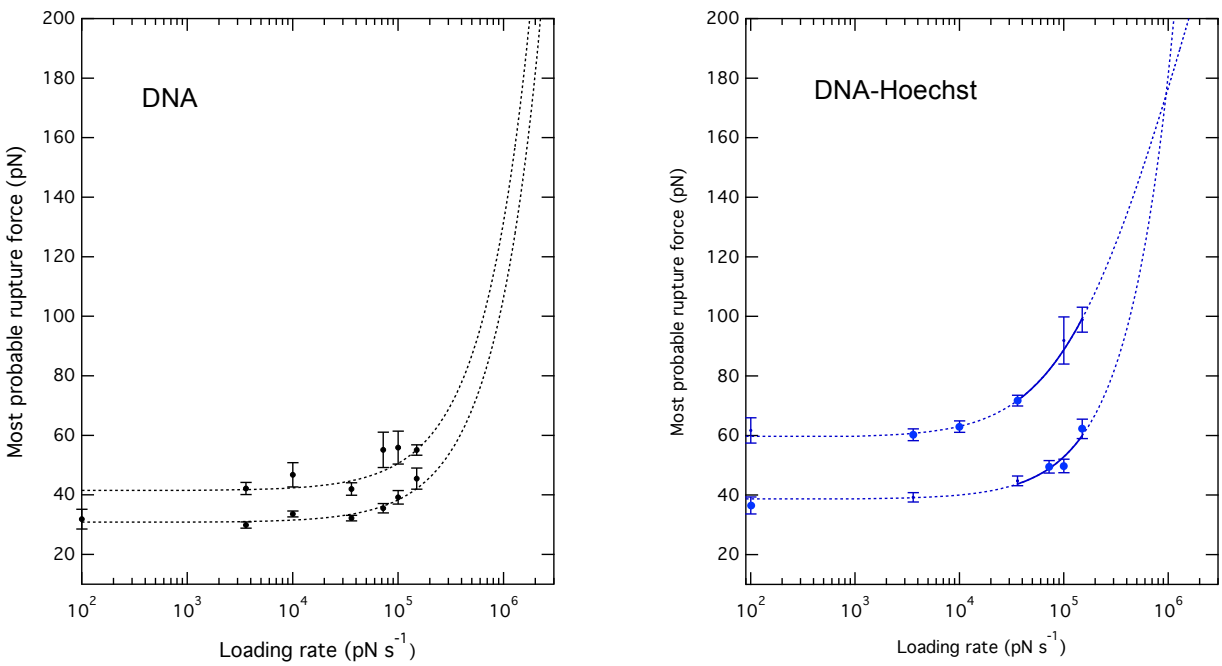
Figure S4: Histograms of the rupture forces and corresponding probability distribution.

3. Analysis of the two dissociation pathways with the single barrier model of Friddle *et al.*

By observing separately the two sub-populations evidenced above by the three-component GMM analysis at each loading rate, we evidenced a non-linear evolution with the logarithm of the loading rate, consistent with the two-regimes trend proposed by Friddle and coworkers.[7] The equation they propose to describe the evolution of the most probable rupture force <f> with the loading rate r is

$$\langle f \rangle = f_{eq} + f_{\beta}^{app} \cdot \ln[1+e^{-\gamma}(r/(f_{\beta}^{app}.k_u.f_{eq}))] \quad (2)$$

where f_{eq} is the equilibrium rupture force, $f_{\beta}^{app} = k_B T/x_t$ (x_t=distance to the transition state), k_u is the kinetic unbinding rate and γ is Euler's constant. This relation describes well the behavior of both sub-populations and supports the presence of two distinct pathways (Figure S5). For DNA, the values are very close for both mechanisms, so that it is hazardous to further discuss their occurrence on this basis. For the DNA-ligand complexes, the split between the two sub-populations is much more significant and evidences that around a critical threshold there is a switch in dominant dissociation pathway.



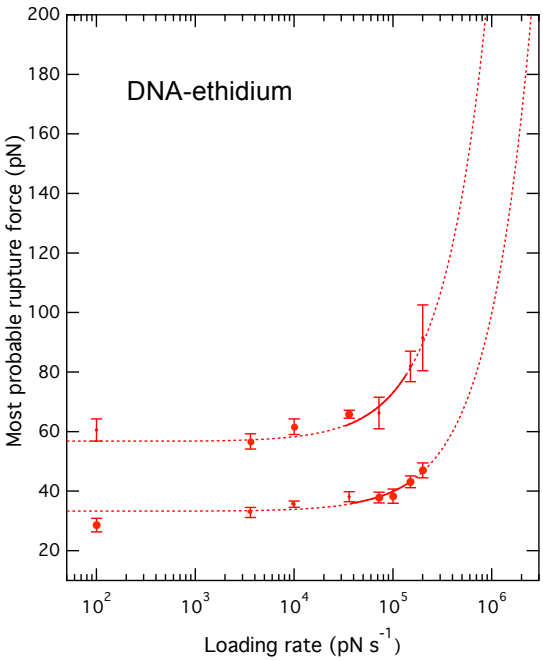


Figure S5: Most probable rupture force of the two main Gaussian components evidenced by the three-component GMM analysis. Each component follows the single barrier model of Friddle *et al.*[7]. The continuous line is a fit of the data at intermediate loading rates (for which the minimization of BIC and AIK is achieved for M=3) to equation 2. The dotted line is the extrapolation of the fit at low and higher loading rates. The size of the dots reflects the weight of the component.

4. Molecular Dynamics simulations

4.1 Complexes between DNA and ligands (Hoechst and Ethidium cation)

It is well established that the favored binding site of Hoechst in the minor grove of DNA corresponds to AT rich portions[8]. For a dodecamer duplex similar to the one used here, only one unique site of fixation for Hoechst 33258 has been found and only one Hoechst molecule binds to the duplex (1:1 complex).[8-10] We performed preliminary unconstrained equilibrium simulations on two different bonding configurations (see Figure S6A) in order to identify a good bonding site in the AT rich central region of the oligonucleotides. MM-GBSA analysis (see next section) has been used to evaluate the relative binding free energy of the two binding sites. The binding results in two complexes, with Complex 2 being about 4 Kcal mol^{-1} more stable than Complex 1. Thus, Complex 2 has been used in all the subsequent simulations. An H-bond analysis of the free complex reveals that the interaction occurs mainly through the formation of hydrogen bonding between two of the N-donors in the HOE (N1,N3) and the free oxygen of the thymine bases T19 -20-7, a weaker hydrogen bond is also present with the nitrogen N3 of the adenine base A6. In particular during the trajectory N3 –H of the Hoechst interacts with T19 and T7 while N1-H establishes a hydrogen bridge with T20 and the nitrogen of A6. The same hydrogen bonds are persistent in the trajectory corresponding to the pinned simulations at all the elongations between 0 and 25 Å. As already reported in other studies both experimental[1] and theoretical[2], the binding with Hoechst does not lead to significant distortion of the B-Helix structure of the free oligonucleotide. In particular, the average end-to-end distance (calculated as the distance between the –OH termination of the guanine and cytosine residues at the 5' termini) results of 41.9Å and 43.6Å in the free oligonucleotide and the complex with Hoechst, respectively.

Intercalators, like ethidium, follow the nearest neighbor exclusion principle, which means that at saturation, every second base pair stack is occupied by a drug molecule.[10,11] The stoechiometry of the complex is

dependent on the concentration of the ligand. At the concentration we used, the 1:3 complex is by far the most abundant one.[10,11] The DNA-ethidium complex 1:3 (Figure S6B) has been obtained by performing preliminary 5ns simulations on the 1:1 and 1:2 complexes in order to allow the helix structure to adapt to each intercalating unit. It is known that intercalations strongly perturb the double helix structure, entailing stretching and partial unwinding.[10,11] From the equilibrium unconstrained simulation of the 1:3 DNA-ethidium complex the average end to end distance results in 54.9 Å, which corresponds to an elongation of about 13 Å with respect to the free oligonucleotide.

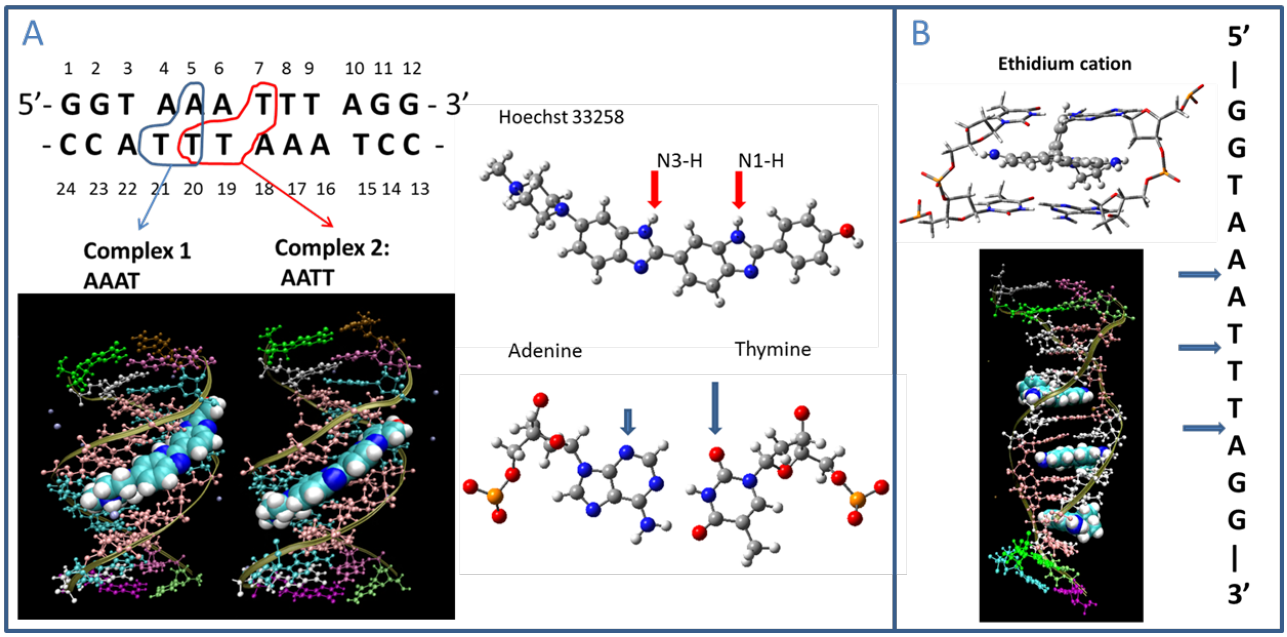


Figure S6: Complexes between the DNA dodecamer and the binders. Panel A: On the left, the oligomer sequence is shown together with two possible interaction sites for Hoechst. The corresponding structures shown in the bottom panel are snapshots from the un-constrained equilibrium MD. On the right we report the molecular structure of Hoechst 33258 and of the A-T bases with arrows indicating the atoms involved in the H-bond interactions between the drug and the DNA sequence. Panel B: Molecular structure of the ethidium cation intercalated between an AA base pair step (top). The complex used for the simulations is shown on the bottom; the arrows indicate the three intercalation sites in the sequence.

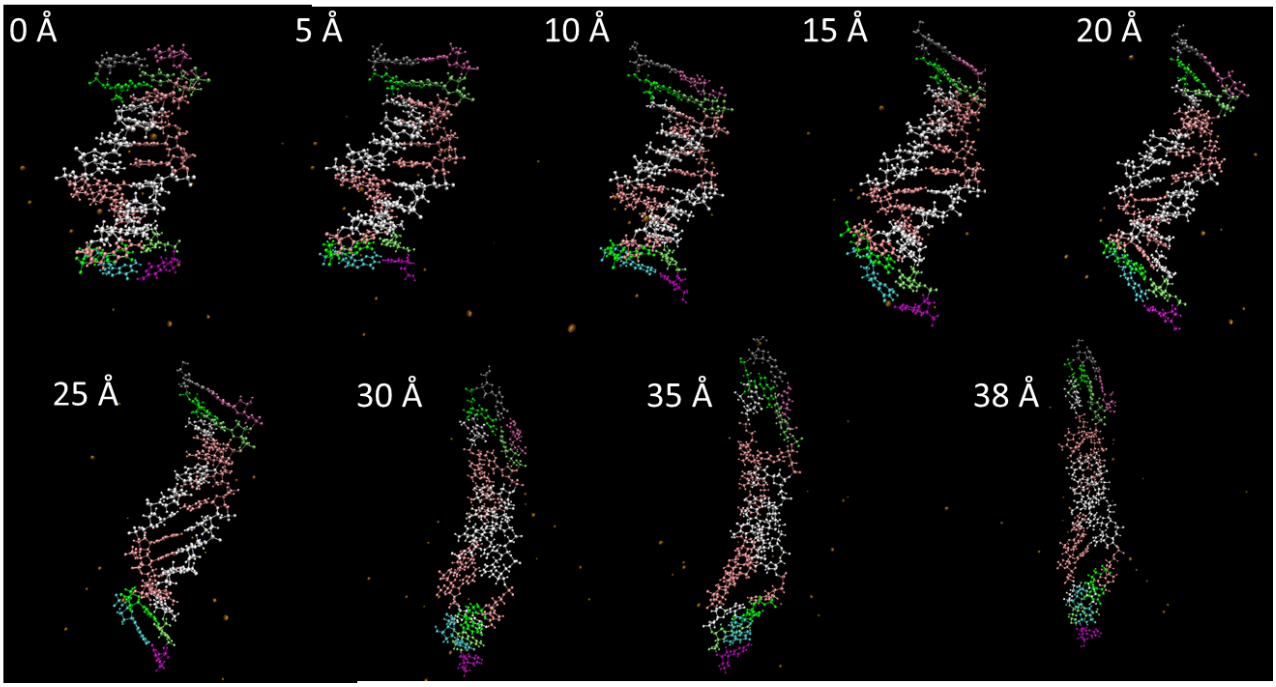


Figure S7: Representative structures at different elongation during a SMD simulation of the free DNA oligonucleotide.

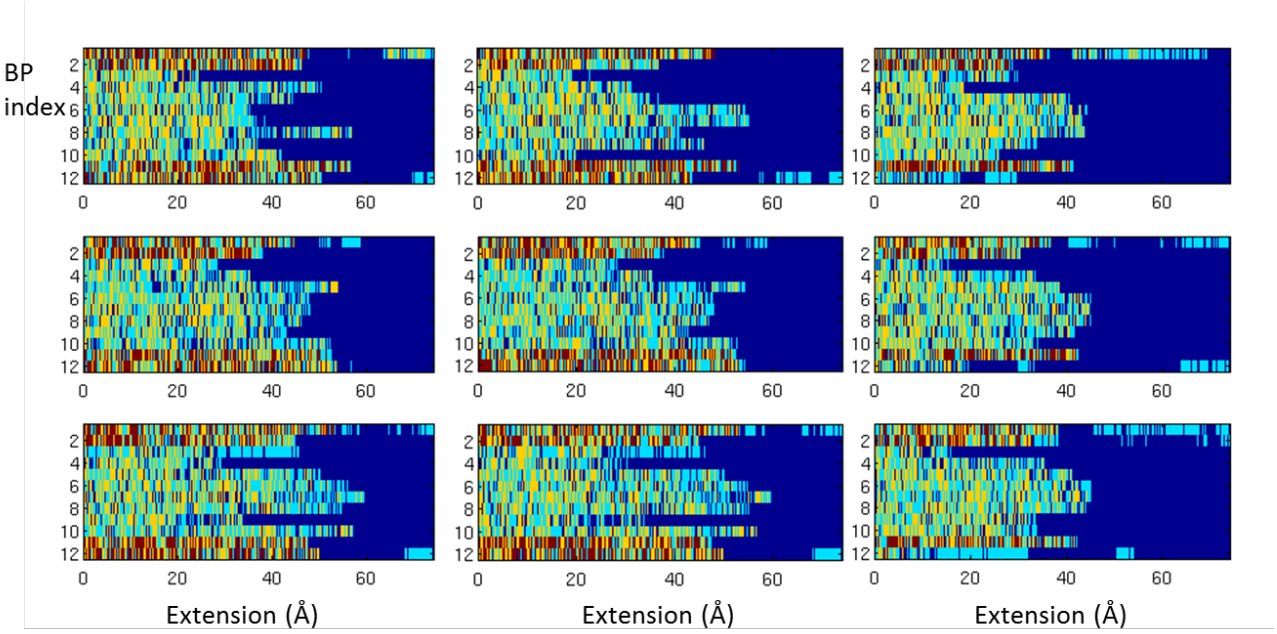


Figure S8: base pairing in term of H-bonds as a function of the elongation in three SMD simulations of the DNA oligonucleotides (A), complex with Hoechst 33258 (B) and complex with ethidium cations (C). (red: 3 H-bonds, yellow: 2 H-bonds, light blue: 1 H-bond, blue: no H-bond)

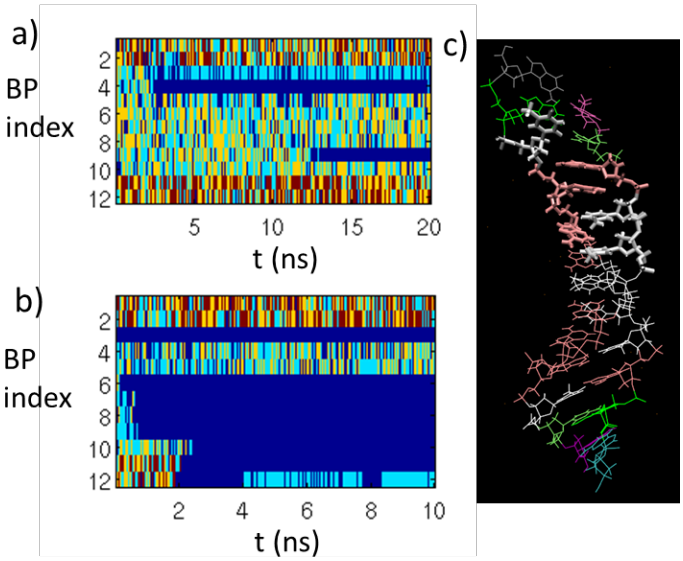


Figure S9: a) H-bond pattern as a function of the simulation time for the MD of the oligonucleotide constrained at a relative elongation of 20Å. b) H-bond pattern as a function of the simulation time for the MD of the oligonucleotide constrained at a relative elongation of 37Å, the BPs still intact in the structure obtained from the SMD quickly disintegrate in few ns of relaxation. c) snapshot from 20Å constrained simulation showing the melting process to start from the rupture of internal BPs TA-3 and AT-4.

4.2 Estimation of the free energy profiles along the end-to-end distance coordinate

An estimation of the free energies difference between different elongation states of the DNA was calculated on the basis of the long-time constrained simulation by using the MM-GBSA method[12], in connection with an evaluation of the entropic contribution by means of quasi-harmonic analysis[13]. The MM-GBSA is a post-processing method of evaluating binding free energies of molecules in solution by the analysis of a set of structures collected by the equilibrated portion of a MD trajectory. Solute configurations are sampled as "snapshots" from a molecular dynamics simulation using explicit solvent. For each solute configuration, the "gas-phase" energy is estimated using the same molecular mechanics potential that was used to perform the simulation, but all solvent molecules are ignored, and no cutoffs are used in evaluating the nonbonded interactions. Free energies of solvation are then re-introduced by using a numerical Generalized-Boltzmann (GB) calculation for the electrostatic portion and a surface-area-dependent term (SA) for non-electrostatic contributions to solvation. This method is commonly used to estimate relative binding free energy of different ligands. Here we use it to calculate relative free energy differences amongst different elongation state of the DNA dodecamer as first suggested in ref. 14. In practice we sample 90 configurations from an equilibrated portion of 9ns of pinned simulations of the DNA at a given elongation. For each ensemble of configurations the estimated free energy can be expressed as

$$G = \langle E_{MM} \rangle + \langle G_{sol} \rangle - T \langle S \rangle \qquad (1)$$

Where the first two terms on the rhs are the molecular mechanics energy without solvent and the solvation free energy as provided by the MM-GBSA as explained above and the brackets denotes the average over the sampled configurations. The evaluation of the entropic contribution is a difficult problem. We applied the method based on quasi-harmonic analysis[13-15], which requires the calculation of the eigenvalue ω of the mass-weighted coordinate covariance matrix from the simulation. By defining the corresponding scaled energy as $\alpha_i = \hbar\omega_i / (kT)$ the entropy is calculated by applying the formula for a quantum mechanical oscillator, as suggested by Andricioaei and Karplus[16]

$$S = k\sum_{i=1}^{3N-6}\left[\frac{\alpha_i}{e^{\alpha_i}-1} - \ln\left(1-e^{-\alpha_i}\right)\right] \quad (2)$$

For a system containing N atoms it is formally necessary to include 3N-6 configurations in the quasi-harmonic analysis to ensure completeness of the data set. We thus evaluate the first value of the entropy by analyzing the first 5ns of the equilibrated trajectory. The configurational entropy of the system relates to the volume of conformational space accessible to a molecule at a given temperature. Since in a finite-time MD simulation the system might not explore all the accessible region of such a space, the entropy calculated according to eq.(2) depends on the length of the sampling window used to calculate the covariance matrix. In the case in which the system explore a smooth energy landscape during all the simulation time the entropy can be obtained by fitting its time dependence by the empirical function

$$S = S_\infty - At^{-2/3} \quad (3)$$

where S_∞ is the asymptotic value of the entropy formally corresponding to a simulation of infinite length. As pointed out in ref 17 to get an idea of the error intrinsic to this methodology the value of S_∞ obtained from different simulations should be compared. In figure S10 we show the entropy calculated from the first portion (1-10ns) and the last portion (11-20ns) of the simulation pinned at the equilibrium end-to-end distance and at 5 Å of relative elongation (upper and lower panel, respectively). In the first case quite similar value results from the two portions for each window length, nonetheless the different growing trend as a function of time leads to asymptotic values, which correspond to a difference in the TS term of 8.9 kcal mol^{-1}. This corresponds to the intrinsic uncertainty of this method and is compatible with the error derived in ref 17. The entropy values calculated for the two portions of the 5 Å elongation simulation show a similar time dependence but the curve corresponding to the second part of the simulations is characterized by higher value of the entropy for all the considered times. This is a consequence of a reversible opening of the terminal base-pair observed at 14ns of the simulation. The corresponding difference of the two asymptotic values corresponds to a difference in the evaluated TS of about 15.1 kcal mol^{-1}.

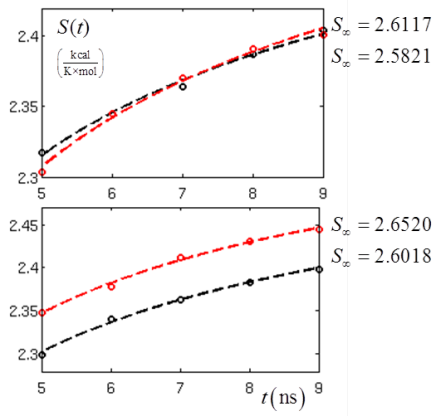


Figure S10: Entropy as a function of the sampling window length and corresponding fitting with equation (3). The black trace refers to the first portion of the simulation (1-10ns) while the red trace is calculated on the second portion (11-20ns). Upper panel: data from the simulation constrained at the equilibrium end-to-end distance (0Å of relative elongation). Lower panel: data from the simulation constrained at a relative elongation of 5Å.

If the portion of energy landscape explored during the MD timescale is rough and presents multiple minima the evaluation of the entropic contribution is less reliable: firstly the assumption that every degree of freedom can be treated as a quantum harmonic oscillator, intrinsic in the definition of the entropy eq. (2), breaks down adding additional uncertainty to the obtained value[18]. Moreover the time dependence of the calculated entropy cannot be fitted by a simple function like eq. (3), as one can see from the evolution of the entropy as a function of time during the 15Å simulation showed in figure 5 of the main text. Despite the fact that one cannot obtain a

quantitative estimation of the entropic contribution to the free energy change, the analysis of the time dependence of the entropy during the simulation time gives information about the relation between structural changes and the gain/loss of motional freedom of the duplex. In Table S2 we report the value of $T\Delta S_\infty$ obtained for different elongations as calculated from the first and the second portion of the 20ns long trajectories. We also report the value of TΔS calculated by using the value of the entropy at the last point (i.e. for a window length of 9ns). The relative stability of the DNA duplex at different elongation results from the sum of two large contributions: the internal plus solvation energy given by the MM-GBSA calculation and the entropic contribution. While the error in the former is of few kcal mol^{-1} the error on the evaluation of the entropic contribution is significant (not less than 10kcal mol^{-1} in the absence of relevant conformational change like base-pair breaking or un-staking) and prevents any quantitative comparison. Nonetheless the values reported in Table S1 should be considered as significant trend supporting the analysis presented in the main text and giving a qualitative picture of the thermodynamical driving force of the stretching and the fraying/unpeeling processes. As a general trend we remark that the elongation of the free oligonucleotide until 10 Å is not expensive from the point of view of the GBSA free energy (around 7-9 kcal mol^{-1}) and the entropic change (TΔS) tends to be slightly negative or positive but always small in absolute value. The further elongation to 15Å requires comparatively more energy (around 24 kcal mol^{-1} if no base pairs are disrupted) and the structure starts to be dynamically unstable.

Table S2

Elongation (Å)	ΔE_GBSA (1-10ns)	$T\Delta S_\infty$	TΔS(9ns)	ΔE_GBSA (11-20ns)	$T\Delta S_\infty$	TΔS(9ns)
2	1.65	-10	-3	-0.57	-4	1
5	4.92	6	-2	8.83	12	12
10	7.00	5	1	9.27	0	-1
15	23.54	17	0	51.91*	62*	65*
20	33.59	35	34	36.33	48	40
25	43.79	95	61	41.77	44	30

*values calculated after the fraying of the terminal base-pairs

4.3 Computational methodology

MD simulations of the DNA, DNA-Hoechst 33258 complex and DNA-ethidium cation (1:3) were performed using AMBER 10.0 with the AMBER ffbsc0 force field and the TIP3P water model. The force fields for the ligand have been prepared with ANTECHAMBER and corrected according to the parameterization developed in ref 19 for Hoechst and in ref 20 for the ethidium cation. The initial structures were solvated in a cubic periodic box, with a minimum buffer of 10Å between any DNA or solute atom and the closest box edge. Sodium counterions were added to establish charge neutrality. The resulting system was energy minimized using steepest decent and conjugate gradient methods to relieve any residual unfavorable steric interactions introduced during the solvation procedure. First, the systems (DNA and DNA complexes) were restrained (500 kcal/mol-$Å^2$) while the water and counterions were subjected to 10000 cycles of minimization. Then, the full system was allowed to relax during an additional 10000 cycles of unrestrained minimization. The DNA/Hoechst 33258/Ethidium cations systems were restrained (25 kcal/mol$Å^2$) during a 20 ps, constant volume MD simulation (NVT), during which water and the Na+ atoms were allowed to move freely and the temperature was raised from 0 to 300K using a Langevin temperature control. Next, the system was subjected to 150 ps of constant pressure (NPT) MD to achieve proper density and 1ns of relaxation before production runs. During MD the long-range electrostatic interactions were treated with the particle-mesh Ewald method[21] using a real-space cutoff distance of $r_{cuttoff}$ = 9 Å. The SHAKE algorithm[22] was used to constrain bond vibrations involving hydrogen atoms, which allowed a time step of 2 fs. Frames are recorded every 1000 steps

(2ps). In the H-bond analysis the cutoff distance is 3A and the cutoff angle between the acceptor, hydrogen and donor atoms has been set to 135°.

References

1 Freedman, D.; Diaconis, P. On the histogram as a density estimator:L 2 theory. *Z. Wahrscheinlichkeitstheorie verw Gebiete* **1981**, *57,* 453-476.

2 Guo, S.; Ray, C.; Kirkpatrick, A.; Lad, N.; Akhremitchev, B.B. Effects of Multiple-Bond Ruptures on Kinetic Parameters Extracted from Force Spectroscopy Measurements: Revisiting Biotin-Streptavidin Interactions. *Biophysical Journal* **2008**, *95,* 3964-3976.

3 Raible, M. *et al.* Theoretical Analysis of Single-Molecule Force Spectroscopy Experiments: Heterogeneity of Chemical Bonds. *Biophysical Journal* **2006**, *90*, 3851-3864.

4 Hyeon, C., Hinczewski, M.; Thirumalai, D. Evidence of Disorder in Biological Molecules from Single Molecule Pulling Experiments. *Physical Review Letters* **2014**, *112,* 138101.

5 Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* **1977**, *39,* 1-38.

6 Burnham, K.P.; Anderson, D.R. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research* **2004**, *33,* 261-304.

7 Friddle, R. W.; Aleksandr Noy, A.; De Yoreo, J. Interpreting the widespread nonlinear force spectra of intermolecular bonds. *Proc. Natl. Acad. Sci.* **2012**, *109*, 13573–13578.

8 Haq, I.; Ladbury, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chaires, J. B. Specific binding of hoechst 33258 to the d(CGCAAATTTGCG)2 duplex: calorimetric and spectroscopic studies. *Journal of Molecular Biology* **1997**, *271*, 244-257.

9 Bostock-Smith, C. E.; Harris, S. A.; Laughton, C. A.; Searle, M. S. Induced fit DNA recognition by a minor groove binding analogue of Hoechst 33258: fluctuations in DNA A tract structure investigated by NMR and molecular dynamics simulations. *Nucleic Acids Research* **2001**, *29,* 693-702.

10 Haq, I. Thermodynamics of drug–DNA interactions. *Archives of Biochemistry and Biophysics* **2002**, *403,* 1-15.

11 Strekowski, L.; Wilson, B. Noncovalent interactions with DNA: An overview. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **2007**, *623,* 3-13.

12 Tsui, V.; Case, D. A. Molecular Dynamics Simulations of Nucleic Acids with a Generalized Born Solvation Model. *Journal of the American Chemical Society* **2000**, *122,* 2489-2498.

13 Harris, S. A.; Laughton, C. A. A simple physical description of DNA dynamics: quasi-harmonic analysis as a route to the configurational entropy. *Journal of Physics: Condensed Matter* **2007**, *19,* 076103.

14 Harris, S. A.; Sands, Z. A.; Laughton, C. A. Molecular Dynamics Simulations of Duplex Stretching Reveal the Importance of Entropy in Determining the Biomechanical Properties of DNA. *Biophysical Journal* **2005**, *88,* 1684-1691.

15 Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. Cooperativity in Drug−DNA Recognition: A Molecular Dynamics Study. *Journal of the American Chemical Society* **2001**, *123,* 12658-12663.

16 Andricioaei, I.; Karplus, M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *The Journal of Chemical Physics* **2001**, *115,* 6289-6292.

17 Piana, S. Structure and energy of a DNA dodecamer under tensile load. *Nucleic Acids Research* **2005**, *33,* 7029-7038.

18 Schafer, H.; Mark, A. E.; van Gunsteren, W. F. Absolute entropies from molecular dynamics simulation trajectories. *The Journal of Chemical Physics* **2000**, *113,* 7809-7817.

19 Furse, K. E.; Lindquist, B. A.; Corcelli, S. A. Solvation Dynamics of Hoechst 33258 in Water: An Equilibrium and Nonequilibrium Molecular Dynamics Study. *The Journal of Physical Chemistry B* **2008**, *112,* 3231-3239.

20 Kubař, T.; Hanus, M.; Ryjáček, F.; Hobza, P. Binding of Cationic and Neutral Phenanthridine Intercalators to a DNA Oligomer Is Controlled by Dispersion Energy: Quantum Chemical Calculations and Molecular Mechanics Simulations. *Chemistry – A European Journal* **2006**, *12,* 280-290.
21 Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98,* 10089-10092.
22 Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **1992**, *13,* 952-962.