Supporting Information for:

Structure, mechanics, and binding mode heterogeneity of LEDGF/p75-DNA nucleoprotein complexes revealed by scanning force microscopy

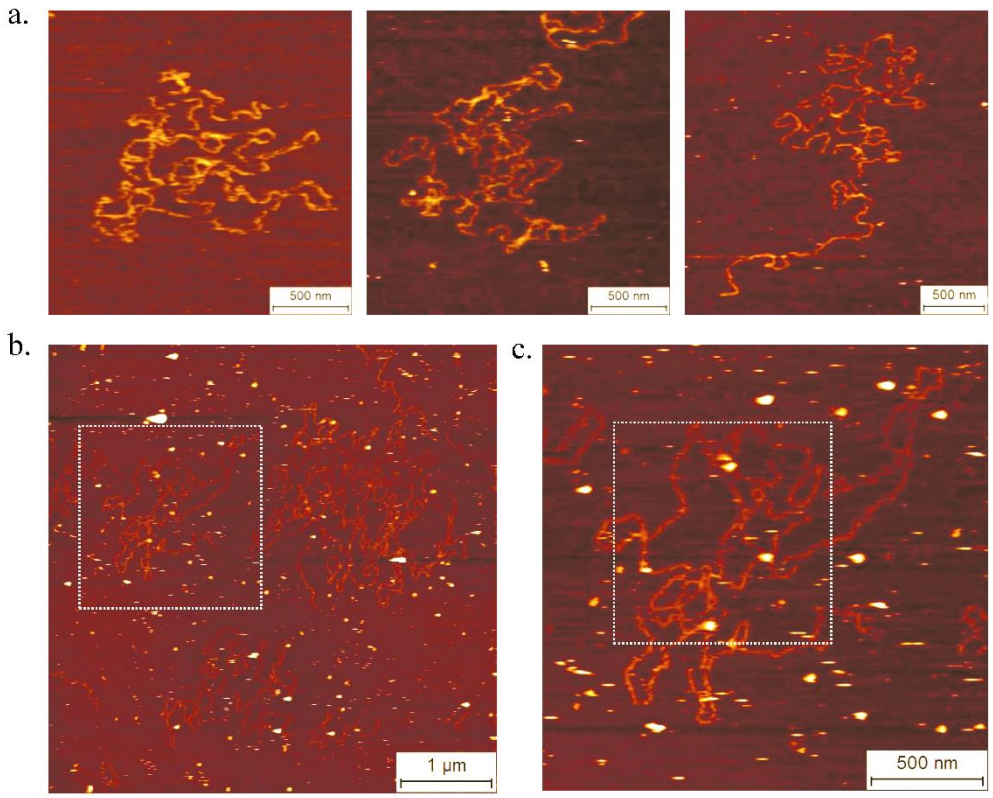
Willem Vanderlinden[a,*], Jan Lipfert[b], Jonas Demeulemeester[c], Zeger Debyser[c,*] and Steven De Feyter[a,*]

Table of Contents:

SFM topographs of phage lambda DNA in situ, in the absence and presence of LEDGF/p75

Supporting Figure S1a depicts three representative examples of phage lambda DNA molecules adsorbed onto freshly cleaved muscovite mica from a buffer containing 100 mM K-acetate, 50 mM Na-acetate, 10 mM, Mg-acetate and 10 mM Tris-HCl (pH = 8.0). Under these conditions, the chains are loosely bound to the surface, allowing in plane and out-of-plane dynamics (1). The adsorbed DNA chains appear slightly entangled and comprise regions which are transiently desorbed, both at the ends of the chain (so-called tails) as well as internally (so-called loops). Therefore, the DNA remains accessible to other molecules at the solid-liquid interface.

We have employed this feature to study the dynamic interactions of LEDGF/p75 with DNA using time-resolved SFM imaging (Figure 2d-e). In such an experiment, first a buffered solution containing phage lambda DNA (0.5 ng/µL) was dropcasted onto a freshly cleaved muscovite mica surface and the DNA chains were left to adsorb and equilibrate for 15 minutes before a 10 µL solution containing 10 nM of His-LEDGF/p75 was added. Next, the sample was loaded in the SFM liquid cell and 250 µL of additional buffer solution. After engaging the SFM probe, a 5 µm x5 µm area was imaged in a time-resolved fashion by scanning the same sample area continuously from top to bottom at a rate of 1 frame/ 8 minutes and with a pixel size of ~ 5 nm x 5 nm. This approach allowed sampling the interactions of individual LEDGF/p75 molecules with several (typically 2-4) DNA chains simultaneously. Supporting Figure S1b shows a typical overview image, and the boxed region highlighting a single DNA molecule, is digitally zoomed and depicted in Supporting Figure S1c. The boxed region in Supporting Figure S1c corresponds to the region of 1 µm x 1 µm that is shown in a time-resolved fashion in Figure 2d.



Supporting Figure S1. SFM topographs of phage lambda DNA *in situ*, in the absence and presence of LEDGF/p75 (**a**.) Three representative examples of SFM topographs of phage lambda DNA molecules acquired under liquid, in the absence of LEDGF/p75. (**b**.) Typical 5 µm x 5 µm SFM topograph depicting four lambda DNA molecules in the presence of LEDGF/p75, which was added as a 10 µL solution (10 nM) before loading the sample in the SFM liquid cell. The boxed image is

digitally enlarged and presented in (c). (**c**.) Digital zoom of the area indicated in (b) depicting a single phage lambda DNA molecule in the presence of LEDGF/p75. The boxed area corresponds to the window which is shown in a time-resolved fashion in Figure 2d.

Model-independent tests for DNA chain equilibration in 2D

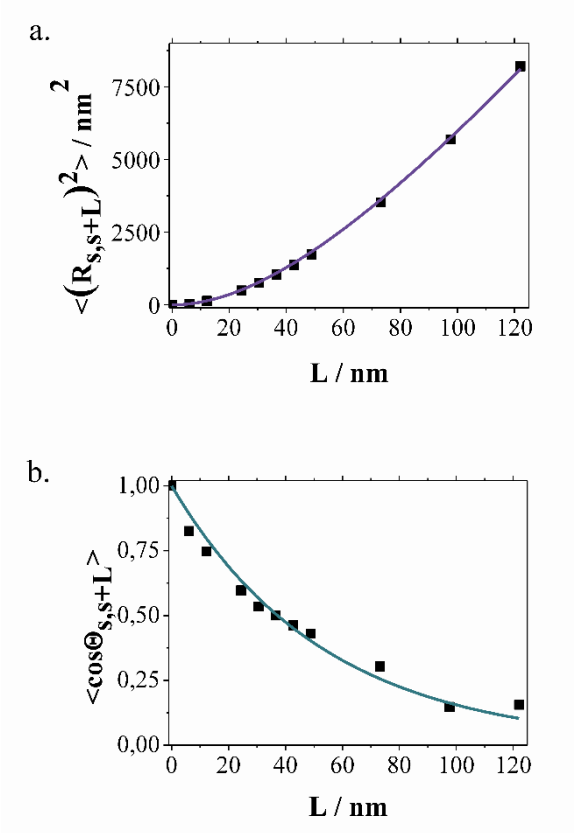
It is important to reflect on how the observation of surface-adsorbed structures compares to the fully dissolved solution structures. First, the adsorption of a 3D flexible object such as DNA onto a 2D surface greatly affects its structure, and this is particularly important in the case of (protein-induced) bend angles (2). The interaction strength between the DNA chain and the substrate plays a defining role in the correct interpretation of quantitative values obtained from SFM images. Only in case the bended DNA segment comprising the nucleoprotein complex is able to equilibrate in 2D at the length scale of the analysis, the mean value for the bend angle as well as its variance can reflect the situation in solution. As a first test, chain equilibration in 2D can be evaluated from model-independent polymer statistics by measuring the mean squared end-to-end distance $\langle\left(R_{s,s+L}\right)^2\rangle$ as a function of the separation L along the chain contour, parameterized by the coordinate s:

$$\langle\left(R_{s,s+L}\right)^2\rangle = 4P * \left(L + 2P\left(e^{-L/(2P)} - 1\right)\right) \qquad \textit{Equation S1}$$

Where P is a model-dependent parameter. In the worm-like chain model, which describes a polymer as an isotropic elastic rod wherein bending deformations are characterized by an energy penalty according to Hookes law, P corresponds to the persistence length. The persistence length is defined by the correlation of the angles θ between tangent vectors along the chain contour:

$$\langle cos\ \theta_{s,s+L}\rangle = e^{-L/(2P)} \qquad \textit{Equation S2}$$

The exponential decay of $\langle cos\ \theta_{s,s+L}\rangle$ as a function of L represents a second means to evaluate 2D chain equilibration. The experimental SFM data, obtained by depositing 500 bp linear DNA fragments (Eurogentec; 0.25 ng/µL) onto poly-L-lysine coated mica from a buffer containing 200 mM Na-acetate, 10 mM Tris-HCl (pH = 8.0) for 30 seconds, were read out employing a home-written routine in Matlab (Mathworks) based on the algorithms described in reference (3) From Supporting Figure S2 it is clear that the condition of 2D equilibration is certified up to contour lengths of 120 nm, which is far beyond the length scale required for bend angle analysis (15 nm). The fits to the experimental data yield P = 28.00 ± 0.37 nm (error is SEM) and P = 26.87 ± 2.70 nm (error is SEM) for equation S1 and equation S2 respectively. In the framework of the WLC model, P corresponds to the persistence length which is typically measured to be ~ 50 nm (4-6). The significant difference with the value we measure here relates to the electrostatic contribution to the persistence length: on charged surfaces, such as the poly-L-lysine coated mica surface we have employed in this study, the charges on the DNA backbone are substantially neutralized, resulting in a smaller value for P (7). This increased DNA bending flexibility helps negatively supercoiled plasmids to maintain their plectonemic structure on 3D to 2D transition, without affecting chain equilibration (7) on short length scales which is required for proper bend angle analysis.



Supporting Figure S2. Model-independent tests for DNA chain equilibration in 2D**.** (**a**.) mean-square separation of pairs of points located at contour length s and s + L from the end of the molecule as a function of the contour length separation L between these points. The solid line represents the best fit (R^2 = 0.999) to the data according to equation S1. (**b**.) the tangent–tangent correlation $\langle \cos \theta_{s,s+L} \rangle$ as a function of L along the chain contour, where $\theta_{s,s+L}$ is the angle between tangent vectors at a pair of points separated by L. The solid line represents the best fit (R^2 = 0.978) to the data according to equation S2.

SFM topographs of plasmid DNA substrates I-IV in the absence of LEDGF/p75

In order to evaluate the conformations and bend angle distributions of the plasmid DNA samples (plasmids I-IV), SFM imaging was employed. Moreover, because plasmids II, III and IV were generated by enzymatic treatments (see materials and methods), it was necessary to make sure that all samples were properly purified from contaminating DNA-bound protein particles which might otherwise interfere with the analysis of LEDGF/p75 nucleoprotein complexes. Depending on the sample, one to five rounds of purification and subsequent SFM imaging were performed to achieve this criterium. Supporting Figure S3 depicts representative images of plasmids I-IV obtained in the absence of LEDGF/p75.

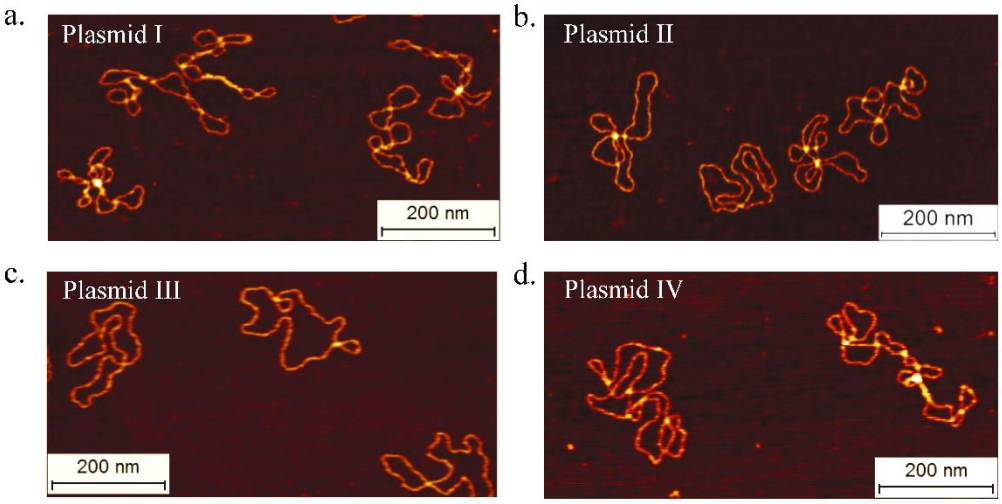
Plasmid I, i.e. native negatively supercoiled pUC19, was obtained commercially and was used after one round of purification using spin columns (Supporting Figure S3a). The adsorbed molecules exhibit many different crossovers, resulting from the chiral coiling of the DNA helical axis in 3D-space. This coiling reduces partly the untwisting (as compared to B-DNA form) of the double helix.

Plasmid II (partially relaxed negatively supercoiled pUC19; Supporting Figure S3b) was generated by treatment of plasmid I with wheat germ topoisomerase Ib in the presence of 1 μM of chloroquine phosphate, a well-known compound which unwinds the double helix locally by intercalation in between the basepairs of DNA. To ensure complete reaction equilibrium, a high enzyme concentration (10 U/ 100 μL) was used, and the reaction was executed for 5 hours at 37 degrees C.

The reaction products were purified in several steps: first, the sample was passed through a spin column. Next, the sample (100 μL) was dialyzed extensively against Tris-HCl buffer (10 mM, pH = 8.0; volume: 5 L) in a micro dialysis step to remove the chloroquine intercalator. The dialysate was then subjected to two more rounds of purification using spin columns before use. The SFM topographs indicate that the reaction products are still to some extend supercoiled, even though the number of crossovers per molecule is substantially reduced and the regular plectonemic structure disappeared.

Plasmid III (torsionally relaxed pUC19; Supporting Figure S3c) was generated by treatment of plasmid I with wheat germ topoisomerase Ib at room temperature. To ensure complete reaction equilibrium, a high enzyme concentration (10 U/ 100 μL) was used, and the reaction was executed for 12 hours at 25 degrees C. Three rounds of spin column -based purification yielded samples free of DNA-bound protein, as judged from SFM topography data. The SFM topographs depict open circular conformations with few local loops. These loops originate from bending and torsional fluctuations of the double helix during the relaxation reaction as well as during the transfer from 3D to 2D.

Plasmid IV (positively supercoiled pBR322; Supporting Figure S3d) was obtained commercially. This plasmid was prepared from torsionally relaxed pBR322 using reverse gyrase. It was found that this sample contained large amounts of protein contamination. Although most of these contaminating proteins were not bound to DNA (and probably originated from the large concentration of BSA used during the reaction), we have used five rounds of spin column-based purification in order to obtain satisfactory clean samples. It should be noted that, whereas some spherical features are apparent in the background of the SFM topographs, no DNA-bound protein particles were apparent after purification. The DNA structure is characterized by fairly regular plectonemes, as for native negatively supercoiled plasmids. However, in the present case the 3D coiling of the helical axis serves to relieve torsional strain resulting from overwinding of the helix (as compared to B-DNA form). Only in rare cases it is possible to unambiguously distinguish the chirality of the crossovers from SFM data.

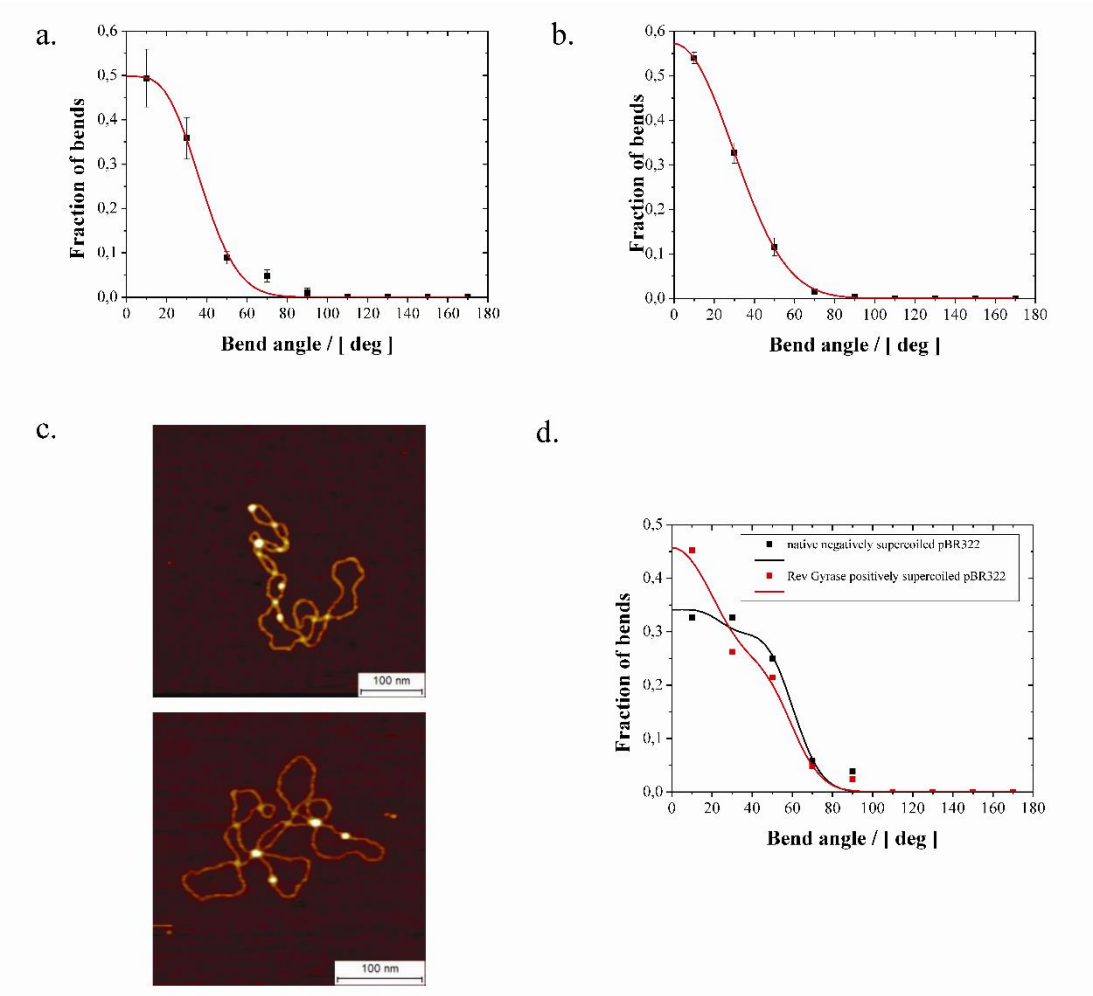


Supporting Figure S3. SFM topographs of covalently closed circular plasmid DNA substrates in the absence of LEDGF/p75. Plasmids are adsorbed onto poly-L-lysine coated mica from a buffer solution containing 200 mM Na-acetate and 10 mM Tris-HCl (pH = 8.0). (**a**.) Native negatively supercoiled pUC19 (plasmid I); (**b**.) Partially relaxed negatively supercoiled pUC19 (plasmid II); (**c**.) Torsionally relaxed pUC19 (plasmid III) and (**d**.) Positively supercoiled pBR322 (plasmid IV).

Proof-of-principle of bend angle determination on supercoiled plasmid DNA: EcoRV binding to cognate and non-cognate sites in pBR322 plasmid DNA

Bend angle distributions of nucleoprotein complexes have been analyzed previously via SFM imaging, and the mean bend angles obtained through data fitting compared in many cases favorably to

bend angles as found in atomic resolution maps of protein-DNA cocrystal structures determined by X-ray diffraction, gel bandshift assays, ligase-catalyzed cyclization or electric birefrigence (2). However, in all these reports, SFM data were obtained on proteins complexed onto linear DNA substrates. In this case, the bend angle distribution along the 2D equilibrated DNA chain is analytically predicted to be Gaussian. In contrast, no analytical expressions for bend angle distributions of topologically restricted DNA circles are available, and the Gaussian fits to the bend angle distributions of the adsorbed bare DNA substrates is a good approximation at best. Since bend angle determination of nucleoprotein complexes on circular DNA substrates has not been described before, we have validated our methodology using EcoRV, a restriction enzyme which forms well-characterized nucleoprotein complexes with both cognate as well as non-cognate DNA sequences. In the presence of certain divalent cations, the EcoRV scans in 3D and in 1D along the DNA double helix searching for its sequence-specific restriction site. When bound to DNA non-sequence specifically, it rotates along the helix during a 1D sliding search mechanism. In this case, the protein does not induce DNA structural deformations. However, on encountering its restriction site, the protein forms a specific complex with DNA, bending the DNA to allow efficient DNA restriction. Interestingly, in the presence of Ca^{2+} as a divalent cation instead of Mg^{2+}, DNA binding can occur, but catalysis of DNA restriction is impeded. We incubated EcoRV under such conditions with both positively as well as negatively supercoiled pBR322 plasmid DNA, which contains a single EcoRV restriction site. These samples were further prepared for SFM imaging in an identical manner as employed for the LEDGF/p75–DNA binding. For many adsorbed plasmid molecules, several nucleoprotein complexes were observed (Supporting Figure S4c). As the pBR322 plasmid only contains a single restriction site, this observation implies that both specific as well as non-specific nucleoprotein complexes are observed. In a first step, the bend angle distributions for naked DNA substrates were analyzed (Supporting Figure S4a,b). For the specific EcoRV-DNA complex we find, after global fitting (Supporting Figure S4d), an optimized mean and standard deviation of 49 ± 1 degrees and 13 ± 1 degrees, respectively. The value for the mean is in very good accordance with reported values for the specific nucleoprotein cocrystal structures (8). Therefore, our methodology to determine bend angles of nucleoprotein complexes on circular DNA substrates appears robust and accurate.



Supporting Figure S4. Bend angle determination of cognate EcoRV-DNA complexes on supercoiled plasmids. (**a**.) Bend angle distribution of native negatively supercoiled pBR322 plasmid adsorbed for 30 seconds onto poly-L-lysine coated mica from a buffer containing 200 mM Na-acetate, 1 mM $CaCl_2$ and 10 mM Tris-HCl (pH = 8.0), averaged over six molecules. (**b**.) Bend angle distribution of positively supercoiled pBR322 plasmid adsorbed for 30 seconds onto poly-L-lysine coated mica from a buffer containing 200 mM Na-acetate, 1 mM $CaCl_2$ and 10 mM Tris-HCl (pH = 8.0), averaged over six molecules. (**c**.) Representative SFM topograph of EcoRV complexes on native negatively supercoiled pBR322 (top) and positively supercoiled pBR322 (bottom). (**d**.) Bend angle distributions of DNA in EcoRV nucleoprotein complexes on negatively supercoiled pBR322 (black symbols) and positively supercoiled pBR322 (red symbols). The solid lines represent the fits to a double folded Gaussian, wherein the mean and standard deviation are fixed at the values corresponding to the respective naked DNA substrates, and the mean and standard deviation of the second Gaussian are optimized using global fitting over both datasets.

Supporting References

1. W. Vanderlinden and S. De Feyter, *Nanoscale,* 2013, **5**, 2264.
2. C. Bustamante and C. Rivetti, *Annu Rev Biophys Biomol Struct*, 1996, **25**, 395.
3. P.A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker and P.C. Nelson, *Nat Nanotechnol*, 2006, **1**, 137.
4. C. Bustamante,C., J.F. Marko, E.D. Siggia, and S. Smith, *Science*, 1994**, 265,** 1599.
5. P.J. Hagerman, *Annu Rev Biophys Biophys Chem*, 1988, **17**, 265.
6. C. Rivetti, M. Guthold and C. Bustamante, *J Mol Biol*, 1996, **264,** 919.

7. A. Podestà,A., M. Indrieri, D. Brogioli, G.S. Manning, P. Milani, R. Guerra, L. Finzi and D. Dunlap, *Biophys J*, 2005, **89,** 2558.
8. .K. Winkler, D.W. Banner, C. Oefner, D. Tsernoglou, R.S. Brown, S.P. Heathman, R.K. Bryan, P.D. Martin, K. Petratos and K.S. Wilson, *EMBO J*, 1993, **12**, 1781.