Title

- Manuscript title: HIV integrase compacts viral DNA into biphasic condensates.
- Short title: Viral genome compaction by HIV integrase

Authors

Pauline J. Kolbeck[1,2], Marjolein de Jager[2], Margherita Gallano[1], Tine Brouns[3,4], Ben Bekaert[3], Wout Frederickx[3], Sebastian F. Konrad[1], Siska Van Belle[4], Frauke Christ[4], Steven De Feyter[3], Zeger Debyser[4], Laura Filion[2], Jan Lipfert[1,2,*], Willem Vanderlinden[1,2,3,5,*]

Affiliations

[1]Department of Physics and Center for NanoScience, LMU Munich, Germany

[2]Soft Condensed Matter and Biophysics, Department of Physics and Debye Institute, Utrecht University, Utrecht, The Netherlands

[3]Division of Molecular Imaging and Photonics, KU Leuven, Belgium

[4]Department of Pharmaceutical and Pharmacological Sciences, KU Leuven, Belgium

[5]School of Physics and Astronomy, University of Edinburgh, Scotland UK

[*]Correspondence: willem.vanderlinden@ed.ac.uk or j.lipfert@uu.nl

Abstract

The human immunodeficiency virus (HIV) infects non-dividing cells and its genome must be compacted to enter the cell nucleus. Here, we show that the viral enzyme integrase (IN) compacts HIV DNA mimetics *in vitro*. Under physiological conditions, IN-compacted genomes are consistent in size with those found for pre-integration complexes in infected cells. Compaction occurs in two stages: first IN tetramers bridge DNA strands and assemble into "rosette" structures that consist of a nucleo-protein core and extruding bare DNA. In a second stage, the extruding DNA loops condense onto the rosette core to form a disordered and viscoelastic outer layer. Notably, the core complex is susceptible towards IN inhibitors, whereas the diffuse outer layer is not. Together, our data suggest that IN has a structural role in viral DNA compaction and raise the possibility to develop inhibitors that target IN-DNA interactions in disordered condensates.

Teaser

Single-molecule studies demonstrate the mechanism, dynamics, and drug-susceptibility of viral genome compaction by HIV integrase.

MAIN TEXT

Introduction

To establish infection, the human immunodeficiency virus (HIV) traverses multiple cellular micro-environments of human host cells, using only a limited arsenal of viral proteins. HIV infects non-dividing cells, and viral particles must enter the cell nucleus to perform integration, thereby irreversibly establishing a provirus in the human genome. While the molecular structures and mechanisms of reverse transcription (*1*,*2*) and integration (*3–7*) are reasonably well understood, the composition and structural assembly of the large nucleoprotein complexes wherein they occur, *i.e.* the reverse transcription complex and the pre-integration complex (PIC) respectively, remain obscure.

Recent work has dramatically impacted our view on the intracellular nature of the virus particle: in contrast to previous belief, HIV transport towards the nucleus occurs within the confines of a (largely) intact viral capsid (*8*, *9*). The new insights suggest that reverse transcription takes place in the confines of the capsid, which in turn requires constraining the increasing mechanical stresses due to the conversion of single-stranded HIV RNA into a much stiffer double-stranded copy-DNA (cDNA) of the HIV genome. Together, the observations suggest the need for a molecular mechanism to compact the DNA within the capsid.

In this context, it was found that the estimated copy-number of the viral protein integrase (IN) in PICs (*10*) is ~150-250, at least one order of magnitude larger than the number of IN (4-16 copies; (*4*, *11*)) required for intasome assembly and catalysis of integration. This raises the question of whether and how such a large excess of IN is beneficial during the replication cycle. While IN serves a functional role during the later stages of HIV replication, in particular during viral assembly and maturation through binding to viral RNA (*12*), IN mutants defective for non-specific DNA binding are impaired for nuclear import (*13*) and reverse transcription (*14*), suggesting a critical role of IN within the cytoplasmic viral capsid also during the early stages of infection.

Here, we provide evidence for an additional functional role of IN by showing that IN can effectively compact viral DNA mimetics under physiologically relevant conditions. Dimensions of the resulting condensates agree favorably with those of PICs purified from (*15*) or within the infected cell (*16*). Our data, obtained by single molecule force microscopy and spectroscopy techniques, and complemented with coarse-grained polymer simulations, suggest a bridging-induced attraction mechanism (*17–19*) that features an on-pathway “rosette” intermediate preceding full collapse. The fully collapsed condensates retain the mechanically robust rosette core and a diffuse viscoelastic outer layer, reminiscent of current models of eukaryotic genome folding (*20*) – albeit at a length scale orders of magnitude shorter. Importantly, we demonstrate a differential susceptibility towards allosteric IN inhibitors for the rosette folding intermediate *versus* the fully collapsed condensate.

Results

IN binds viral DNA ends with moderate selectivity

Binding of IN to viral DNA ends is a prerequisite for the catalysis of the integration reaction. Here, we use atomic force microscopy (AFM) imaging to quantify the selectivity of recombinant IN for the specific ends in short viral cDNA mimics (Fig. 1a), as a function of the level of 3' processing. From AFM images, we determine the position of the bound IN complexes with respect to the nearest DNA end in IN-DNA nucleoprotein complexes (Fig. 1 and Supplementary Fig. S1), which enables us to estimate the selectivity of end-binding (*21*) $S = \frac{P_{end}}{P_{background}}$. We first investigated binding of IN to a 491 bp cDNA mimic with blunt ends in buffer containing 1 mM Mg^{2+}. The distribution of binding events along the internal DNA is essentially homogenous (Supplementary Fig. S1), implying the absence of high affinity internal binding sites. Assuming (*22*) a footprint of 16 bp, there are 476 potential binding sites in the cDNA mimic, and we estimate a selectivity S = 64 ± 13 (from N=159 molecules; errors were estimated from counting statistics using standard error propagation; Fig. 1g). Next, we studied binding in buffer wherein Mg^{2+} is replaced by Ca^{2+} to impede catalysis of 3' processing (*23*) and we find a reduced selectivity S = 19 ± 7 (N = 122; Fig. 1c). Last, we quantify binding to HIV DNA mimetics with 3' pre-processed ends (see Methods), which results in an increased selectivity, S = 256 ± 15 (N = 129). Thus, the selectivity of binding to the specific blunt HIV ends increases by more than 10-fold by catalysis of 3' processing. Nevertheless, binding selectivity overall is only moderate; the probability of binding to a random position in genomic length viral DNA (~8-9 kbp) is still 1-2 orders of magnitude larger than binding to the specific ends.

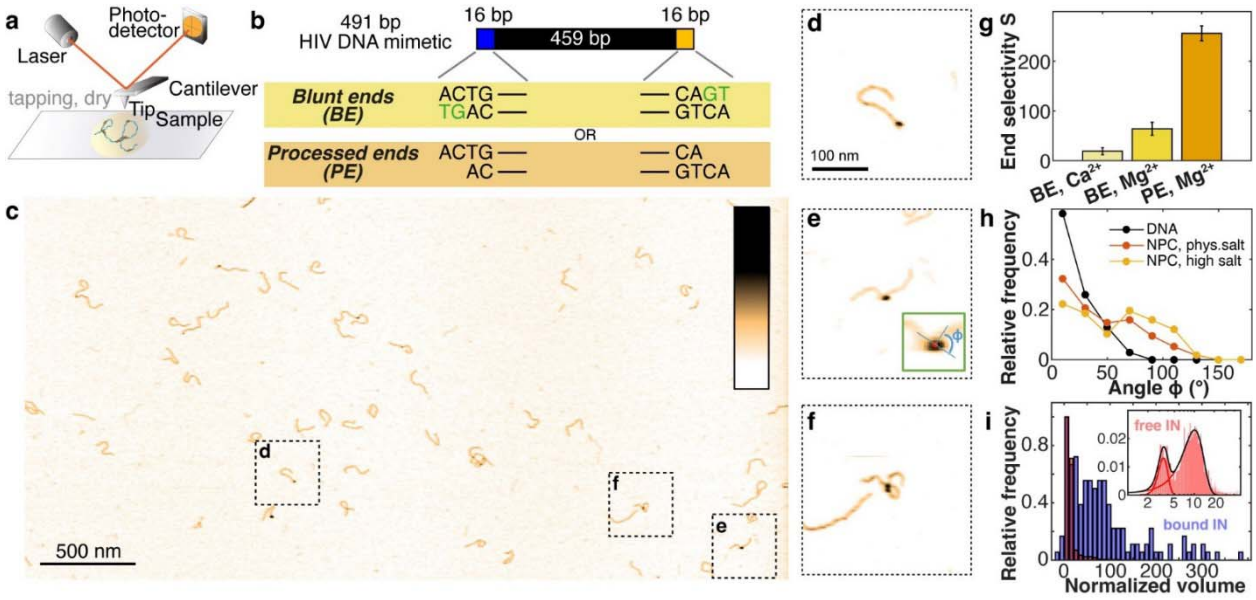

Figure 1. IN specific and aspecific binding to HIV DNA mimetics. a. Schematic representation of the experimental set up for AFM imaging experiments. **b.** Schematic of the 491 bp DNA used to mimic the HIV cDNA, both as a blunt ended (BE) or as a 3'-preprocessed substrate (PE). **c.** Overview AFM topograph of the 491 bp DNA in the presence of 100 nM IN. Color range is 1.5 nm. **d.** HIV DNA mimetic with end-bound IN. **e.** HIV DNA mimetic with integrase bound at internal sites and definition of bend angle (inset). **f.** HIV DNA mimetic with

integrase-mediated DNA looping. **g**. Selectivity for end-binding over binding to internal sites in the mimetic as a function of 3' processing and buffer conditions. From left to right: blunt DNA ends, in the presence of Ca^{2+}; blunt DNA ends, in the presence of Mg^{2+}; processed DNA ends, in the presence of Mg^{2+}. The IN binding site size is assumed to be 16 bp. Error bars are based on counting statistics. **h.** Bend angle distribution measured at 10 nm length scale for bare DNA and IN-DNA nucleoprotein complexes (NPCs) deposited from near-physiological salt buffer (50 mM NaCl, 1 mM $MgCl_2$, 10 mM Tris-HCl pH = 7.4; N = 232) or high salt buffer (250 mM NaCl, 5 mM $MgCl_2$, 10 mM Tris-HCl pH = 7.4; N = 109). **i.** Histogram of the normalized volume of IN free in solution (red) and IN bound to DNA (blue). In free solution, we find a small fraction of IN monomers and a dominant fraction of IN dimers (inset). For IN bound to DNA, we find the resulting protein volume distribution to be significantly broader and shifted to larger volumes as compared to the DNA-unbound IN population.

Unspecific IN binding infers DNA bending and looping

As most IN-DNA complexes bind to internal sequences, *i.e.*, non-sequence specifically, we next investigate whether non-specific binding alters DNA structure. We quantify IN-induced DNA bending (Fig. 1e,h) and find that the bend angle distribution is much broader and shifted to larger bend angles at positions where IN is bound, as compared to the distribution of bare DNA bends. At higher ionic strength (250 mM NaCl, 5 mM $MgCl_2$), the fraction of complexes with large bend angles (> 50º bending; Fig. 1h) increases with respect to complexes with smaller bend angles, suggesting the presence of (at least) two binding modes: a first mode that is salt-sensitive and minimally invasive, and a second binding mode that is less sensitive to ionic strength and infers strong DNA bending.

Beyond bending, a small fraction (~5% of all internally bound complexes) forms synapses, i.e., intramolecular DNA loops mediated by IN (Fig. 1f). Using a longer viral DNA mimetic (1000 bp) and otherwise identical conditions (100 nM IN), the fraction of looped DNA increases to 30 ± 5% of all internally bound nucleoprotein complexes (N = 93). This increase is consistent with the expectation that looping is facilitated by the longer DNA length, since ~500 bp corresponds to only ~4 times the bending persistence length of DNA.

AFM volumetric analysis suggests that IN in the absence of DNA is monomeric and mostly dimeric, whereas in the DNA-bound state it takes on higher oligomers, in particular tetramers (Fig. 1i and Supplementary Fig. S2). Together, our results demonstrate that IN binding to random positions along the DNA is prevalent, induces higher order IN oligomerization, and can alter DNA conformation by bending and looping.

Two-stage compaction of genomic length viral DNA

To investigate how IN binding affects DNA of lengths similar to the HIV genome, we titrated linearized mini-HIV DNA(*7*) of different lengths (3.4 kbp, 4.8 kbp, 9.1 kbp) with IN and quantify the resulting complexes using AFM (Fig. 2a and Supplementary Fig. S3 to S5). To directly relate

the conformations of adsorbed complexes to their solution state, we performed AFM imaging under kinetic trapping conditions, where the resulting images are close to a 3D to 2D projection (*24–26*).

We first used a 4.8 kbp mini-HIV DNA(*7*) for titration with IN in the concentration range [IN] = 0-2 μM. At IN concentrations of < 500 nM, we observe localized IN binding and IN-induced looping. Increasing the concentration to [IN] = 500 - 1200 nM results in the formation of large nucleoprotein complexes featuring a central nucleoprotein core surrounded by bare DNA loops. Because of their appearance, we refer to these complexes as "rosettes". At even higher IN concentrations (> 1200 nM), the DNA is further compacted, and the nucleoprotein complexes are predominantly in globular conformations wherein individual DNA strands are no longer visible. On occasion, these intramolecular complexes self-associate to form large conglomerates comprising multiple DNA copies (Supplementary Fig. S4f, S5d,e,f).

We quantified DNA compaction upon titration of mini-HIV DNA with IN by automated analysis of the height-weighted 2D radius of gyration $R_g = \sqrt{\frac{\sum_{i=1}^{n} h_i \cdot s_i^2}{\sum_{i=1}^{n} h_i}}$ where s_i is the distance of pixel i with respect to the molecular center of mass, and h_i is the height value in pixel i. Consistent with the qualitative observations of DNA-IN complex geometries, the R_g data reveal that IN significantly compacts DNA in two successive, highly cooperative transitions: a first transition occurs at [IN] ~ 400 nM and reduces R_g by approximately 40% (from 75 nm at [IN] ≤ 250 nM, to 45 nm for [IN] = 500-1200 nM). A second transition is observed at [IN] ~ 1400 nM, and further reduces R_g ~ two-fold (Fig. 2 and Supplementary Fig. S6).

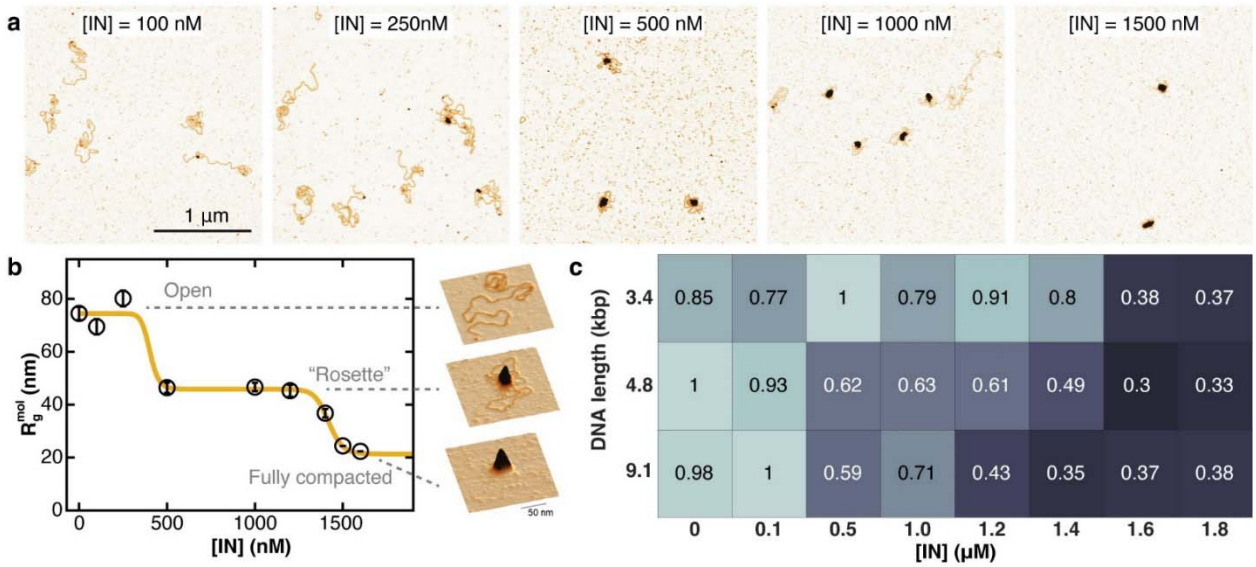

Figure 2. IN compacts long viral DNA mimetics into biphasic condensates. a. AFM topographs of 4.8 kbp DNA in the presence of (from left to right) increasing IN concentrations. **b.** Radius of gyration of the nucleoprotein complex for 4.8 kbp DNA at 60 mM ionic strength as a

function of [IN]. The data reveal two distinct compaction transition and are fitted by a double-Hill-fit: the first transition from open state to rosette state occurs at ~ 500 nM, the second transition from rosette state to fully compacted state at ~ 1100 nM. Representative AFM images are shown as insets on the right. Symbols and error bars are the mean and s.e.m. from typically ≥ 100 molecules per condition. **c.** Phase diagram depicting the normalized radius of gyration of the nucleoprotein complexes formed as a function of [IN] and DNA length, normalized to the largest R_g observed for each DNA length. With increasing DNA length, the compaction transitions shift to lower IN concentration. Intriguingly, at DNA length > 3.4 kbp a biphasic compaction transition is observed, whereas at 3.4 kbp the transition occurs in one step.

Condensation of nucleic acids with proteins and the formation of nucleoprotein complexes are in general highly dependent on environmental parameters. We mapped out the dependence of IN-induced DNA condensation on environmental conditions and observed that compaction is independent of incubation temperature (we tested 37 °C and room temperature, i.e. 22.5 °C; Supplementary Fig. S7) and incubation time (Supplementary Fig. S8), but strongly dependent on ionic strength(*27*) (Supplementary Fig. S9 and S10). We constructed a phase diagram of condensate formation as a function of integrase and salt concentration (Fig. 2c and Supplementary Fig. S11). The data show that both rosette formation and full collapse are disfavored by high salt concentration but occur under physiologically relevant conditions. As expected from the modest end-binding selectivity (Fig. 1g), we find no effect of the specific miniHIV terminal sequences on compaction efficiency (Supplementary Fig. S12).

Having characterized the compaction behavior of a 4.8 kbp HIV DNA mimic, we next explore how DNA length alters IN-induced condensation. We find that the degree of compaction is independent of the DNA contour length and reduces R_g approximately 3-fold at full compaction compared to bare DNA without IN present (Supplementary Fig. S13). Nevertheless, the critical concentrations for the compaction transitions increases for shorter DNA lengths (Fig. 2c and Supplementary Fig. S6). Most notably, full compaction of the 3.4 kbp HIV mimetic occurs in a single transition (i.e., the rosette state is no longer observed), in contrast to the biphasic behavior observed for 4.8 kbp and 9.1 kbp DNA. The absence of a rosette state with DNA loops for the 3.4 kbp construct is in line with a previously identified DNA length threshold for bridging induced looping (*19*). Last, the radius of gyration of the fully compacted state for 9.1 kbp DNA (approximately the length of the viral genome) is ~ 32 nm, close to the Stokes' radius found for purified pre-integration complexes (*15*) (28 nm) and in agreement with the molecular dimensions of pre-integration complexes in infected cells (*16*). Taken together, our quantitative AFM analysis indicates that genomic length viral DNA is compacted via two distinct transitions into condensates with dimensions that are in quantitative agreement with those found in infected cells.

Simulations suggest bridging-mediated attraction mechanism with a critical role for IN-IN interactions

To obtain a mechanistic understanding of how local interactions lead to DNA compaction, we turn to Monte Carlo (MC) simulations (Fig. 3; Supplementary Methods). We model DNA as a chain of beads, each representing 12 bp (equal to the approximate binding site size of IN) joined by springs. A bending potential between subsequent beads is added to reproduce the experimentally determined bending stiffness of DNA. IN tetramers are similarly modeled as beads of the same size as the DNA beads, with an isotropic binding interaction with the DNA beads. We capture the tetrameric binding nature of the active IN complex by introducing a maximum valence of 4: each IN bead can bind to at most 4 DNA beads (Fig. 3a) (see Supplementary Information). We determine the interaction strength of IN binding to a single DNA segment to be $\varepsilon = 5\ k_B T$ (where k_B is Boltzmann's constant and T the absolute temperature) by direct quantitative comparison (*21*) to experimental data (Fig. 3h). For the IN-IN interaction we consider two cases: (i) no IN-IN interaction (beyond excluded volume interactions) and (ii) IN-IN interactions (i.e. IN-IN attractions in addition to the excluded volume interactions). We perform extensive simulations at different IN concentration for both cases and analyze the DNA-IN conformations as a function of time. During these simulations, the IN concentration is controlled via a grand canonical ensemble (see Supplementary Information) to ensure that the concentration of free IN remains constant. To identify different states of compaction and the pathways between them, we develop a classification pipeline using principal component analysis of 15 different structural parameters and a Gaussian mixture model (Fig. 3b-g and Supplementary Information).

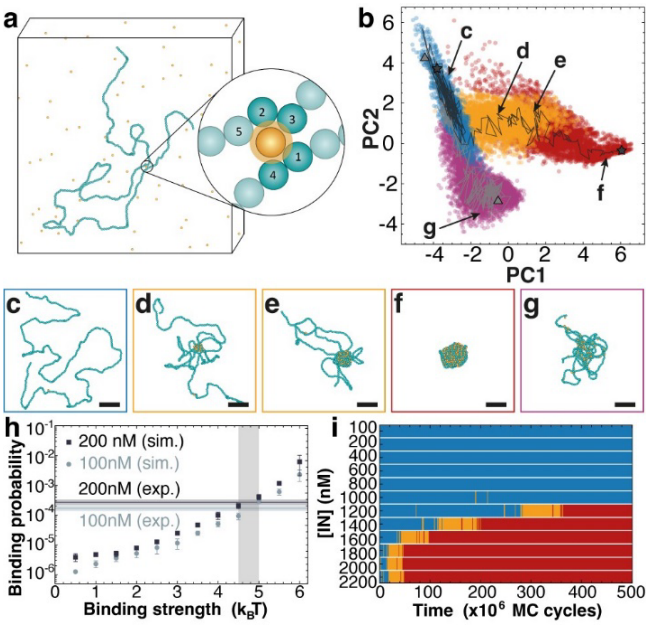

Figure 3. Monte Carlo (MC) simulations show that IN protein-protein interactions are required for biphasic compaction. a. Schematic representation of the setup for MC simulations. DNA beads (4 nm) are colored in light blue and IN beads (same size) are colored in yellow. The box size is not to scale. **b.** Principal component analysis separates the large dataset of DNA-IN complexes into several categories of DNA-IN complexes (depicted as different colors, which

were assigned using a Gaussian mixture model). The grey trajectories indicate two typical simulation paths of 4.8 kbp DNA (408 beads). Without IN-IN attraction, the left path is taken (triangles at start and end point), here no rosette intermediates are formed. Only if IN-IN attractions are added (right path; stars at start and end point), rosette formation (yellow) as well as full compaction are possible (red). **c-g.** Some typical conformations along the pathways in panel b. The scale bar indicates 40 nm. Free IN are not depicted. **h.** IN-DNA binding probability (mean ± SD from 5 independent simulations) as a function of binding strength (in k_BT) from MC simulations. The horizontal lines indicate the experimentally determined binding probability for 100 nM and 200 nM IN from AFM images. The crossing area of experiment and simulations allows to estimate the IN-DNA binding energy to be 4.5-5 k_BT. **i.** Time evolution of the state of compaction (same color as in panel b) of a 4.8 kbp DNA during a MC simulation with IN-IN attractions for varying IN concentrations, well in line with the experimental data (Fig. 2). Each simulation was selected as the most typical one from 12 simulations per IN concentration.

A key result is that we observe distinct pathways in the presence and absence of IN-IN attractions: MC simulations that only include IN-DNA binding (*17*), but no IN-IN attraction, exhibit DNA compaction, but fail to reproduce the two-phase compaction and the experimentally observed rosette conformations. In contrast, if IN-IN attraction is included (*18*) (Fig. 3i), we observe transitions and structures that are in good agreement with the experimental observations (Fig. 2). In summary, our simulations demonstrate that a simple model that includes both IN-DNA and attractive IN tetramer-tetramer interactions can semi-quantitatively reproduce the observed DNA compaction.

IN-DNA condensates feature a rigid core surrounded by a soft coat

To further characterize the IN-DNA condensates and to probe their mechanical properties that likely play a role in the interior of the capsid and/or during nuclear entry, we use AFM force-volume mapping (*28*, *29*). In this modality, a full force-distance curve (i.e., an approach and retract cycle) is recorded in every pixel of the image (Fig. 4a). Localized interaction forces on the tip are recorded as it moves towards and away from the sample surface (Fig. 4b) and can be directly correlated to the sample topography. We analyze the approach force curves to extract the indentation stiffness k, the retract curves to quantify elastic deformation energy $E_{Elast.}$, and the hysteresis between both curves to obtain a measure of plastic deformation $E_{Plast.}$ (Fig. 4a, inset and Supplementary Fig. S14). From the elastic and plastic contributions to the deformation energy, we can quantify the dimensionless viscoelasticity index(*30*) $\eta = 1 - \frac{E_{Elast.}}{E_{Elast.}+E_{Plast.}}$, where $\eta = 0$ for a perfectly elastic deformation and $\eta = 1$ for a completely viscous deformation, i.e. where the bonds broken during indentation are not reformed during the stress release.

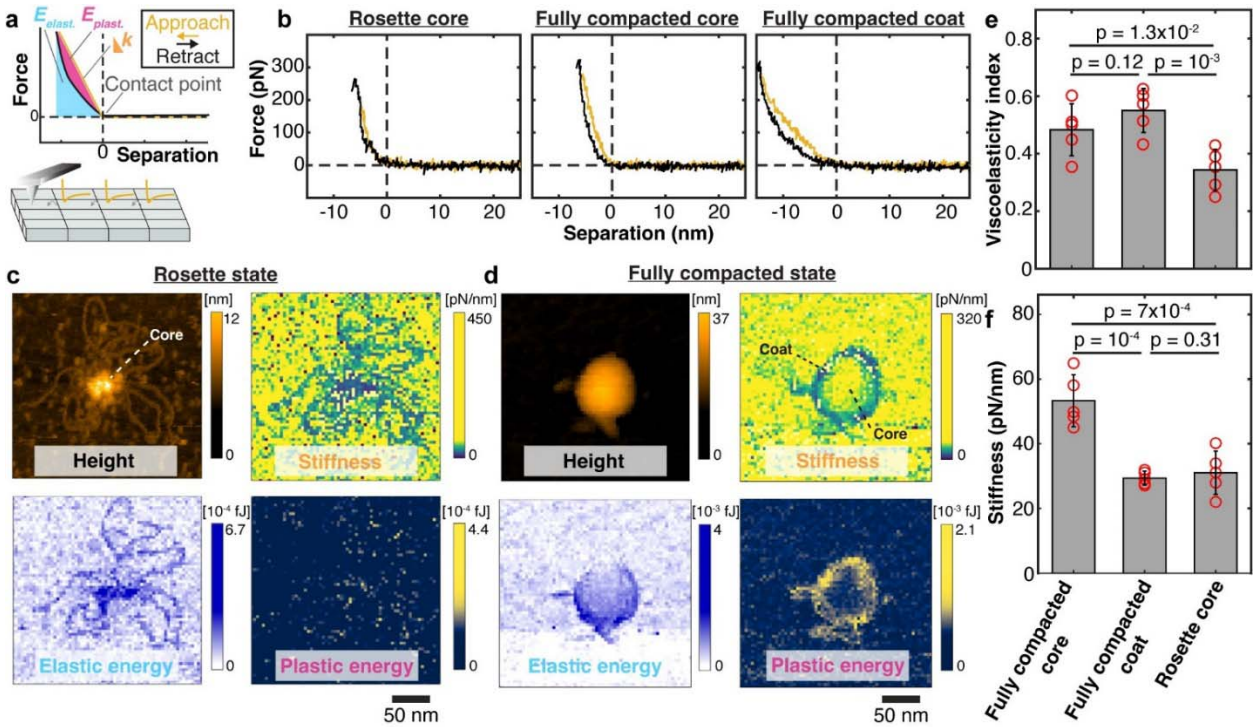

Figure 4. AFM force-volume based multiparametric imaging of IN-DNA condensates. a. Schematic depiction of AFM-based force-volume mapping, and derived parameters k (indentation stiffness), $E_{\text{plast.}}$ (plastic deformation energy), and $E_{\text{elast.}}$ (elastic deformation energy). **b.** Example force distance curves (yellow: approach; black: retract). **c.** Multiparametric imaging of a condensate in the rosette state: zero-force height (top left), stiffness (top right), elastic energy (bottom left), plastic energy (bottom right). **d.** Multiparametric imaging of a DNA-IN condensate in the fully compacted state, indicating a soft "coat" surrounding a rigid "core". **e.** Viscoelasticity index (mean ± SD) of the core and coat of fully compacted state, and of core of rosette state. *p*-values are based on 2-sample Kolmogorov-Smirnov test. **f.** Indentation stiffness (mean ± SD) of the core and coat of fully compacted state, and of core of rosette state. *p*-values are based on 2-sample Kolmogorov-Smirnov test.

We first perform force-volume mapping on complexes in the rosette state (Fig. 4c). The approach and retract curves are uniform across the nucleoprotein core and exhibit little hysteresis. The fitted values of k, $E_{Elast.}$, and $E_{Plast.}$ from at least 100 approach and retract curves are pooled and averaged per complex. Averaging over different rosette complexes, we find a mean stiffness $\langle k \rangle = 31 \pm 7$ pN/nm (mean ± SD from $N = 5$ complexes) and a mean viscoelasticity index $\langle \eta \rangle = 0.34 \pm 0.07$ (Fig. 4e,f).

In a second set of experiments, we use force-volume mapping on fully collapsed condensates (Fig. 4d; $N = 5$). In contrast to force curves recorded on rosettes, we consistently observe substantial hysteresis between the approach and retract curves. Furthermore, we find the overall slopes of the force curves to vary depending on their radial position: at constant peak forces, larger indentations are observed near the edges of the condensates. Indentation curves do not

show breakthrough events, and because one would expect decreased indentations due to finite AFM tip size effects near steep topographic features we interpret the spatially resolved features in the condensates to result from complex material properties.(*31*, *32*) The spatially resolved stiffness map is used to identify a "core" and outer layer or "coat". We pool force curves in the core and find $\langle k \rangle$ = (53 ± 8) pN/nm (error is SD) and a mean viscoelasticity index $\langle \eta \rangle$ = 0.48 ± 0.09 over the different condensates (Fig. 4e,f). An analogous treatment for force curves recorded on the condensate coats gives $\langle k \rangle$ = (29 ± 2) pN/nm (error is SD), significantly less than the stiffness of the rosette, and a mean viscoelasticity index $\langle \eta \rangle$ = 0.55 ± 0.08.

We conclude that the fully collapsed condensates feature a relatively rigid, stiff core and more viscous, softer outer layer. The observed stiffness of the fully compacted condensates are on the same order of magnitude, but still significantly lower (by at least ~5-fold) than the reported stiffnesses of full virus particles (*33*), consistent with the possibility that they could remain in the capsid during transport to and into the nucleus. The pronounced viscoelasticity and hysteresis between approach and retract curves suggests that intrinsic dynamics and re-arrangement of bonds in the complex are much slower than the time scale of the force application and release by AFM (~5 ms).

Force spectroscopy of IN-DNA reveals heterogeneous dynamics of compacted loops

To investigate the dynamics and stability of IN-DNA condensates, we performed magnetic tweezers (MT) force spectroscopy experiments. In our assay, we tethered long linear DNA molecules (21 kbp) between the flow cell surface and superparamagnetic beads (Fig. 5a). Using external magnets, we can apply precisely calibrated stretching forces to the tethered molecules (*34*). To investigate the forces involved in IN-DNA condensation, we first introduced IN (2 μM) in the flow cell to interact with tethered DNA molecules under very low stretching forces (F < 0.01 pN) for 10 min, to allow for condensate formation, and then increase the applied forces to mechanically probe the stability and structure of the condensate (Fig 5b). In a first set of experiments, we gradually increase the force from 0.1 to 5 pN with plateaus of constant force of 60 min (0.1 to 0.9 pN) or 30 min (1 to 5 pN), respectively (Fig 5b and Supplementary Fig. S15), while tracking the extension of the DNA tether. In the presence of IN, the DNA is shortened, compared to control experiments in the absence of IN and to the prediction of the WLC model (red line in Fig 5b), in particular in the force range 0.5 – 1 pN (Fig. 5b, inset). At very low forces (< 0.5 pN), we observe slow increases and decreases in DNA extension at constant force (Fig. 5c and Supplementary Fig. S15c-f), suggesting that the breakage and re-formation of interactions is in equilibrium under these conditions, pointing towards a dynamic and reversible nature of the interactions involved in loop compaction. Notably, the largest deviations from bare DNA behavior are observed in the same force range in force-extension experiments (Fig 5b, inset). The observed reversibility of compaction in the MT is consistent with the observation that rosette formation and compaction are reversible in AFM imaging experiments upon dilution.

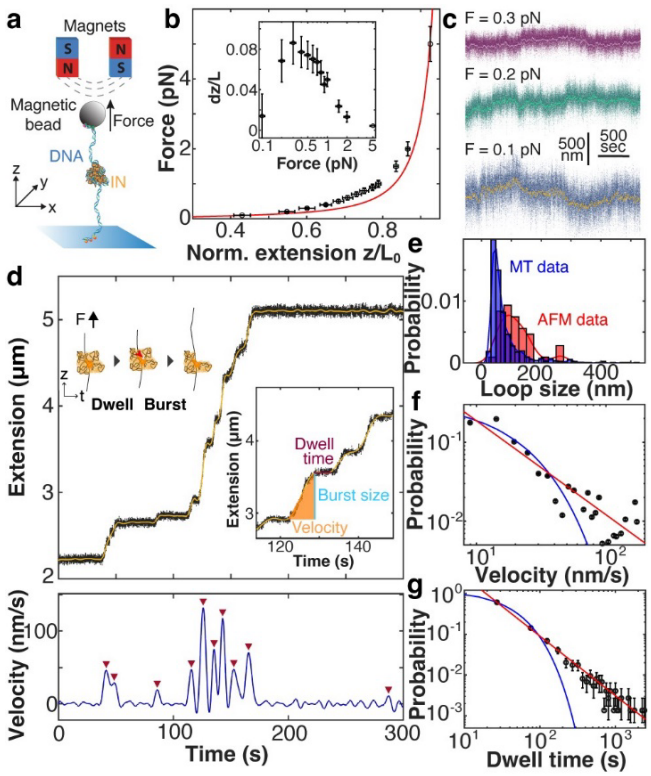

Figure 5. Single-molecule force spectroscopy quantifies dynamics condensate unfolding. a. Schematic depiction of the MT force spectroscopy measurements. 21 kbp DNA is tethered between a flow cell surface and magnetic beads and (partially) compacted by IN. External magnets enable the application of precisely calibrated stretching forces. **b.** Force-extension relation at forces ≤5 pN. Extension z is normalized to L_0, the extension of bare DNA at F = 5 pN. Red line is a worm-like chain (WLC) fit of bare DNA. The inset shows the deviation of the experimental data (mean ± SD) in the presence of IN from bare DNA WLC behavior; the maximum deviation occurs at ~ 0.3 pN. **c.** Extension-time traces of IN-DNA condensates at constant forces ≤0.3 pN exhibiting dynamic compaction and extension. **d.** Top: Extension-time trace following a force jump from 0.01 to 1 pN. Raw data (68 Hz) are depicted in black, yellow line is smoothed data (Butterworth filter, see Methods and SI for filter parameters). Insets depict a schematic of the events governing dwells (corresponding to extension plateaus), and extension bursts following dissociation of IN-DNA bridge in the core of the complex (red arrow). Bottom: Velocity-time trace, as obtained by differentiation of the filtered extension trace. Red marks indicate extension bursts detected by thresholding. **e.** Burst size distribution (blue bars; N= 777 events from 41 independent DNA tethers). For comparison, the loop size distribution obtained by AFM imaging are shown (red bars). A two sample t-test suggest no significant difference between both distributions (2-sample t-test; p = 0.5) **f.** Weighted bursts velocity distribution fitted with exponential decay (blue line; fitted decay constant 12 ± 3 nm/s) and power law (red line; scaling exponent –1.9 ± 0.3) for 1 nm/s < v < 100 nm/s. **g.** Distribution of dwell times between extension bursts and fits using an exponential decay (blue line; decay time of 24 ± 3 s) and power

law (red line; scaling exponent –1.0 ± 0.1). Errors reflect 95% confidence intervals on the fit parameter.

In a second set of experiments, we again incubated tethered DNA molecules at a very low force (F < 0.01 pN) in presence of 2 μM IN for 5 min, to allow for the formation of IN-DNA condensates. Then, we apply a force-jump to a constant stretching force of 1 pN. We observe that the extension only increases very gradually – in contrast to control experiments with bare DNA, where the DNA extension increases essentially instantaneously (≤1 s) to the full extension of DNA at the given force (Supplementary Fig. S16). During force-induced extension in the presence of 2 μM IN, we observe extension increments or "bursts", interspersed by long (>min) pauses or "dwells" where extension remains constant (Fig 5d, top panel). We quantify the extension time traces at constant force by determining i) the dwell times between extension events and for each extension event, ii) the total increase in tether length for the event, and iii) the extension velocity (Fig 5d, bottom panel and Supplementary Fig. S17 to S19).

The increase in length per extension event is in reasonable agreement with the loop sizes found in AFM measurements with short DNA (Fig. 1 and Fig. 5e), suggesting that individual extension events correspond to the release of compacted loops. The extension velocities are much smaller (by at least two orders of magnitude) than what would be expected for the release of free DNA loops (*35*, *36*) and show a broad velocity distribution (Fig. 5f), suggesting that IN-compacted DNA loops are released under forces of ~1 pN, but that the release is slowed down by protein interactions that cause an effective friction and exhibit heterogeneous dynamics, consistent with the viscous nature of the soft coat region observed by AFM probing. The distribution of dwell times between extension events has a mean lifetime of (24 ± 3) s and does not follow a single exponential, but rather a power law with scaling exponent −1.0 ± 0.1 (Fig. 5g), suggesting heterogeneous interactions and dynamic disorder (*37*).

The finding that extension plateaus (characterized by long dwell times ~1 min) separate bursts of extension increments (which occur on the ~s timescale), suggests that force is first exerted on mechanically stable and long-lived DNA-protein bonds that shield the hidden length in compacted loops from opening up. Once these long-lived bonds break, compacted loops are released and open under the applied force, but still experience friction originating from protein-protein and protein-DNA interactions. Together, our MT data further confirm the biphasic nature of fully compacted condensates, and indicate heterogeneous interactions with high levels of dynamic disorder.

Allosteric IN inhibitors interact differentially with rosettes versus collapsed condensates

Our experimental data suggest that interactions in the core are mediated by IN tetramers and higher order assemblies thereof (Fig. 1i). As IN tetramers are the preferred target for allosteric integrase inhibitors (*38*), we tested the susceptibility of condensates in the rosette and fully compacted states towards the small molecule, quinoline-derivative inhibitor CX014442, an integrase inhibitor originally developed to inhibit association with the coactivator LEDGF/p75

(LEDGIN; this class of inhibitors has also been referred to as ALLINI for *allosteric integrase inhibitor*) (*39*). In the following, we will refer to CX014442 from now on as the "inhibitor" for simplicity.

In a first set of experiments, we introduced the inhibitor at 1 and 2 μM, at fixed [IN] = 1 μM, i.e. under condition where the rosette is formed, and studied the resulting complexes via AFM imaging (Supplementary Fig. S20). At equimolar [inhibitor]:[IN] ratio, the typical single nucleoprotein clusters of the rosette conformations appear partly disintegrated (Fig. 6a). We observe an increase in the radius of gyration of the entire molecule, albeit statistically not significant (p = 0.35; Fig. 6b). In contrast, the radius of gyration of the core decreases significantly (p = 0.024; Fig. 6c and Supplementary Fig. S21). At 2:1 [inhibitor]:[IN] ratio, the rosette core is disintegrated completely: individual IN complexes are still bound along the DNA contour but do not exhibit bridging interactions. Consistently, we find significant changes in cluster size and integrity (via $R_{g,core}$; $p = 10^{-25}$), as well as significant changes in compaction of the condensate (via $R_{g,mol}$; $p = 2{\cdot}10^{-14}$).

In a second set of experiments, we introduced the inhibitor at 1.5 and 3 μM, at a fixed [IN] = 1.5 μM, i.e. to fully compacted condensates (Fig. 6d). Strikingly, independent of the inhibitor concentration, the nucleoprotein assemblies retain their fully compacted conformations akin to those found in the absence of inhibitor and we observe no significant changes in cluster size (Fig. 6e,f). We note that the influence of the inhibitor on rosette or fully compacted conformations does not depend on order of addition, indicating that steric effects on inhibitor binding cannot explain our observations (Supplementary Fig. S21).

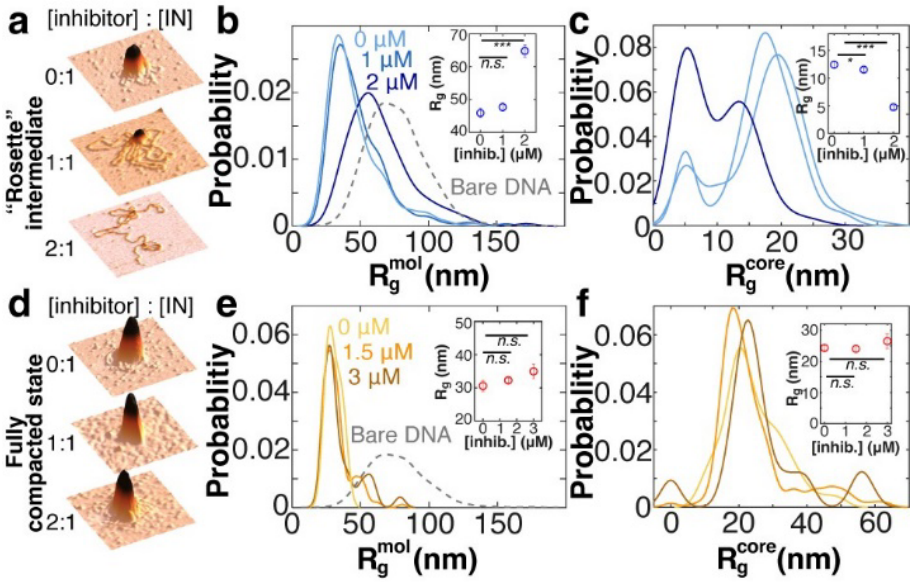

Figure 6. Allosteric inhibitors affect compaction behavior of long viral DNA mimetics. a. AFM topographs of the 4.8 kbp DNA in the presence of 1000 nM IN (rosette state) and no, 1000 nM, 2000 nM inhibitor. The inhibitor was added after complex formation. **b.** R_g-distributions for the entire molecule at 1000 nM IN and no, 1000 nM, 2000 nM inhibitors. As a reference, the distribution for bare DNA is shown as well. With increasing inhibitor concentration, less

compaction is observed. **c.** Same conditions and color coding as in panel b only the R_g-distributions for the core are shown. **d.** AFM topographs of the 4.8 kbp DNA in the presence of 1500 nM IN (fully compacted state) and no, 1500 nM, 3000 nM inhibitor. The inhibitor was added after complex formation. **e.** R_g-distributions at 1500 nM IN and no, 1500 nM, 3000 nM inhibitors. As a reference, the distribution for bare DNA is shown as well. **f.** Same conditions and color coding as in panel e only the R_g-distributions for the core are shown. Even at high inhibitor concentrations, the fully compacted DNA-IN complexes do not disassemble under influence of the inhibitor. The insets in panels b, c, e, and f shows the means ± SEM of the respective distributions. Significance values are based on 2-sample Kolmogorov-Smirnov test: * indicates $p < 0.05$; *** indicate $p < 0.001$, *n.s.* indicates not significant.

In general, LEDGINs have multiple mechanisms of action during HIV replication. By interference with the binding of LEDGF/p75 they redirect integration into less optimal sites. LEDGINs also allosterically interfere with the enzymatic activity of integrase and induce aberrant multimerization (*40*). Aberrant multimerization of IN during virus assembly results in defective virus particles. LEDGINs do affect IN-IN interactions and, in particular induce the loss of IN into inactive aggregates at high IN concentrations (*12*, *41*). Consistent with this view, we observe protein aggregates by AFM imaging (Supplementary Fig. S20). However, the fact that at 1.5 μM IN no visible changes on fully compacted nucleoprotein complexes are observed in the presence of inhibitor strongly suggests that the major effect on IN-induced condensation is not due to the loss of IN in inactive aggregates. In addition, the differential effects of the inhibitor on rosette formation and the fully collapsed state indicate that the protein-protein and/or protein-DNA interactions involved are different for the different compaction stages.

Discussion

Here we use single molecule experiments in combination with Monte Carlo simulations to analyze nucleoprotein condensates formed between viral cDNA mimics and recombinantly expressed HIV IN. IN induces (flexible) DNA bends and mediates the formation of DNA loops. Furthermore, DNA-bound IN complexes were found to be larger than unbound complexes, suggesting that DNA binding induces conformational changes leading to higher order oligomerization (Fig. 7). At concentrations 500-1000 nM, IN induces formation of large DNA complexes with multiple DNA loops extending from the center that we refer to as rosettes. At higher IN concentrations 1000-1500 nM, a second compaction transition occurs in which the extruding loops of the rosette structure collapse onto the core of the condensate and full compaction is achieved. The resulting fully collapsed condensates drastically (~30-fold) reduce the volume occupied by the viral genome. In addition, our data indicate that in the fully collapsed condensate, a rigid central core is surrounded by a diffuse outer layer with relaxation times on the ~min timescale. It is interesting to note that this architecture resembles a polymer brush, the model that similarly describes the folding of eukaryotic chromosomes (*42*) that are orders of magnitude larger and also plays key roles in the organization of prokaryotic genomes (*43*). Importantly, we experimentally demonstrate that the center of the rosette and the outer layer of the fully compacted IN-DNA complexes interact differently with the inhibitor CX014442: while

the center of the rosette dissolves in an inhibitor concentration-dependent manner, the outer layer of the fully compacted state is not susceptible to the inhibitor in the concentration range probed, irrespective of the order of addition of the inhibitor (Supplementary Fig. S21). Intriguingly, this drug-susceptibility (Fig. 6) is in stark contrast to the mechanical (Fig. 4 and 5) and thermodynamic stability (Fig. 2).

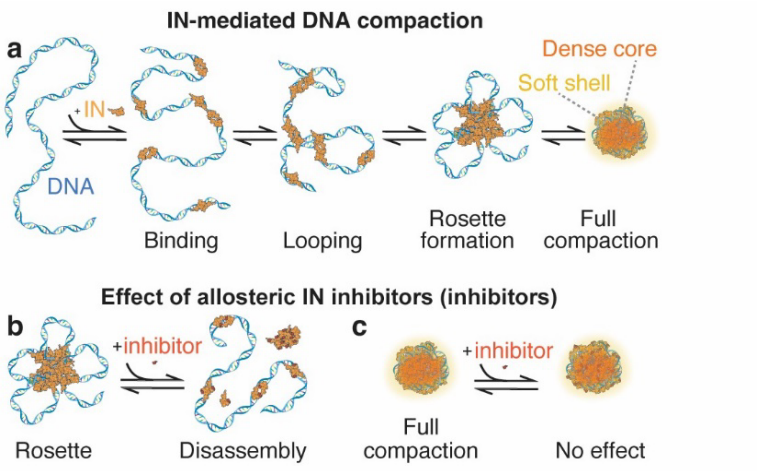

Figure 7. Model for IN-mediated DNA compaction. a. Schematic model of DNA compaction by IN. At increasing IN concentrations, the protein binds to the DNA and can form DNA loops. At intermediate concentration, loop-extruding complexes, so-called rosettes, form. At high IN concentrations, full compaction occurs. **b.** The allosteric IN inhibitor CX014442 can de-compact rosettes. **c.** The allosteric IN inhibitor CX014442 cannot de-compact fully compacted DNA-IN condensates. We observe that all compaction and de-compaction processes are reversible.

Provided that our experiments give a very reductionist view of HIV DNA genome compaction, it is important to evaluate their potential relevance with regards to viral condensates in the infected cell. In this regard, the size of IN-DNA condensates in vitro compare very favorably to sizes found by super-resolution microscopy in vivo and to those of PICs purified from infected cells (*16*). In addition, IN plays a crucial role during capsid uncoating in vivo (*44*) and IN mutants defective for non-specific DNA binding are impaired for nuclear import (*13*) and reverse transcription (*14*). Thus, it appears that IN plays a critical role within the cytoplasmic viral capsid during the early stages of infection. Maybe, these effects involve an architectural role for IN as a DNA compaction agent – similar to the established role of IN during folding of the RNA genome in nascent virus particles (*12*). Nevertheless, we stress that our data do not rule out the involvement of other DNA interacting viral proteins that have been shown to induce DNA compaction in vitro, such as the nucleocapsid (*45*, *46*) and VpR proteins (*47*).

Finally, we are convinced that the highly quantitative approaches presented here will be instrumental in future efforts to unravel the molecular mechanisms governing retroviral

nucleoprotein condensation, e.g. to further establish the role of IN and other DNA-binding viral proteins in compaction of viral DNA. We especially anticipate breakthrough insights into PIC architecture by combining bottom-up (i.e. reconstitution) and top-down (i.e. purification) approaches. Similarly, it would be intriguing to use similar methods to probe the effects of IN on HIV RNA compaction, which has an established biological role during virus maturation (*12*). Last, we expect our experimental platform to be valuable for the in-depth mechanistic investigation of strand transfer and allosteric inhibitors of IN on DNA compaction, and potentially uncover the mode of action of future inhibitors targeting disordered phases of IN and its condensates.

Materials and Methods

HIV-1 integrase expression and purification

For all experiments, we used recombinantly expressed C-terminal or N-terminal His-tagged HIV-1 integrase as described previously.(*48*) The small-molecule allosteric IN inhibitors CX014442 were obtained from KU Leuven (Cistim-CD3).

AFM imaging with short DNA

We used two different short DNA constructs: a blunt-end construct, and a 3' pre-processed construct. For the blunt-end 491 bp viral cDNA mimic, we used a custom-synthesized double-strand DNA (gblock, IDT; see below) without further purification. For the 3' pre-processed HIV DNA mimetic, a gblock of 750 bp (gblock, IDT; see below) was first PCR amplified with forward primer 5'- ACT GGA AGG GCT AAT TCA CTC -3' and reverse primer 5'- ACT GCT AGA GAT TTT CCA CAC -3'. 2 µg of this DNA was cut with the restriction enzymes EcoRI and XhoI. This fragment is used in a ligation reaction with an U3 and U5 pre-processed fragment. The U3 pre-processed fragment was generated by first hybridizing the oligo's: 5'- /5Phos/TC GAG TTT TAG TCA GAG TGA ATT AGC CCT TCC A -3' and 5'- ACT GGA AGG GCT AAT TCA CTC TGA CTA AAA C -3' and subsequently cutting it with the restriction enzyme XhoI and the U5 pre-processed fragment was generated by hybridizing the oligo's: 5'- /5Phos/AA TTC TTT TAG TCA GTG TGG AAA ATC TCT AGC A -3' and 5'- ACT GCT AGA GAT TTT CCA CAC TGA CTA AAA G -3' and subsequently cutting it with the restriction enzyme EcoRI. The ligated fragment was run on a gel and the correct band was cut out and eluted from the gel.

The AFM samples are prepared by incubating 2.5 ng/ µL viral cDNA mimics (491 bp with 16 bp terminal sequences corresponding to viral U3 and U5 ends; see sequence below) and recombinant (C-terminal His-tag) HIV integrase (IN) at different concentrations, in high sodium buffer (10 mM Tris-HCl pH = 7.6; 250 mM NaCl; 5 mM $MgCl_2$) and at 37° C for 30 min, and subsequent 5-fold dilution in low sodium buffer (10 mM Tris-HCl pH= 7.6) just prior to drop-casting onto poly-L-lysine-modified mica(*49*, *50*). Additional measurements were performed in the absence of Mg^{2+}, using a high salt buffer comprising $CaCl_2$ (10 mM Tris-HCl pH= 7.6; 250 mM NaCl; 5 mM $CaCl_2$). After allowing adsorption for 30 s, mica substrates were rinsed with milliQ water and dried using a gentle flow of filtered N_2 or Argon, respectively. AFM imaging was performed in amplitude-modulation ("tapping") mode in air using a Nanoscope IV Multimode AFM.

<u>Sequence of the 491 bp viral cDNA mimic:</u>

ACTGG AAGGG CTAAT TCACT CGGGC GAATT CGAGC TCGGT ACCCG GGGAT CCTCT AGAGT CCATG CAAGC TTGAG TATTC TATAG TGTCA CCTAA ATAGC TTGGC GTAAT CATGG TCATA GCTGT TTCCT GTGTG AAATT GTTAT CCGCT CACAA TTCCA CACAA CATAC GAGCC GGAAG CATAA AGTGT AAAGC CTGGG GTGCC TAATG AGTGA GCTAA CTCAC ATTAA TTGCG TTGCG CTCAC TGCCC GCTTT CCAGT CGGGA AACCT GTCGT GCCAG CTGCA TTAAT GGAGG CGGTT TGCGT ATTGG GCGCT CTTCC GCTTC CTCGC TCACT GACTC GCTGC GCTCG GTCGT TCGGC TGCGG CGAGC GGTAT CAGCT CACTC AAAGG CGGTA ATACG GTTAT CCACA GAATC AGGGG ATAAC GCAGG AAAGA ACATG TGAGC AAAAG GCCAG CAAAA GGCCA GGAAC GTGTG GAAAA TCTCT AGCAG T

<u>Sequence of the 750 bp gblock:</u>

ACTGGAAGGGCTAATTCACTCTGACTAAAAGGATCCCAGTATTTGGTATCTGCGCTC
TGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAA
CCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAA

AAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACG
AAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCTCGAGGA
TACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTAAAGTATATATGAGTAAACTT
GGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATT
TCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGG
CTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCC
AGATTTATCAGCAAAGCTTCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTG
CAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAT
CTAGACAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTTCGCTCA
CGCTCGTCGTTTGGTATGGCTTCATTCAGAATTCTTTTAGTCAGTGTGGAAAATCTCT
AGCAGT

AFM volume analysis of proteins

To estimate the molecular weight of protein particles, we constructed a calibration curve of protein volumes as a function of known protein molecular weights. To correct for the finite size of the AFM tip, we normalize the measured volumes by the volume per nanometer length of co-adsorbed DNA. The normalized volumes of protein particles measured by AFM are well-described by a power law fit: $V_{NORM} = 0.0198^{0.0291}_{0.0105} \times Mw^{1.52^{1.61}_{1.43}}$ (for $Mw \leq 632$ kDa; $\chi_{\mathrm{Red}}^2 = 1.14$; the sub- and superscript numbers indicate the 95% confidence intervals from the fit). To quantify to the volume of DNA-bound proteins, we quantified the volume of the entire protein-DNA complex and subtracted the average volume of surrounding free DNA molecules. Volume normalization using free DNA as a fiducial marker is done as descried above.

AFM imaging with long DNA

As viral cDNA mimic, we used the miniHIV plasmid, which has been described previously.(*7*) Using the 4751 miniHIV plasmid(*7*) as a parental plasmid, we generated two other viral cDNA mimics (3437 bp and 9112 bp) using blunt end cloning (BEC) and Gibson assembly, respectively. The resulting sequences can be found as pU3U5_miniHIV_4p8kbp_seq.txt, pU3U5_restrictedto3p4kbp_seq.txt, and pU3U5_enlargedto9p1kbp_seq.txt as part of the zip archive Plasmid_data_seq.zip in the Supporting Information. All three plasmids exhibit two 180 bp terminal sequences corresponding to viral U3 and U5 ends. The plasmids are linearized using the cutting enzyme ScaI. AFM samples are prepared by incubating 5 ng/µl DNA and IN at different concentrations in high sodium buffer (HSB; 250 mM NaOAc, 25 mM Tris, 5 mM $Mg(OAc)_2$; pH=7.6) at room temperature, and subsequent 5-fold dilution with 25 mM Tris. The sample was deposited on freshly cleaved mica modified with aminopropylsilatrane (APS)(*26*). The sample was incubated for 30 s before being washed with 20 ml of MilliQ water and dried with a gentle stream of filtered argon gas.

After sample preparation, AFM images were acquired in tapping mode at room temperature using a Nanowizard Ultraspeed 2 (JPK, Berlin, Germany) AFM with silicon tips (FASTSCAN-A, drive frequency 1400 kHz, tip radius 5 nm, Bruker, Billerica, Massachusetts, USA). Images were scanned over (5 × 5) μm^2 or smaller areas at a line scan speed of 5 Hz. The free amplitude was adjusted to the complex heights, usually 20 nm. The amplitude set point was set to ~80% of the free amplitude and adjusted to maintain good image resolution.

Force-volume experiments were performed in aqueous buffer (50 mM NaOAc, 25 mM Tris, 1 mM $Mg(OAc)_2$; pH= 7.6) using the Nanowizard Ultraspeed 2 (JPK, Berlin, Germany) this time with silicon tips (BL-AC40TS, drive frequency 25 kHz in water, tip radius 10 nm, Olympus,

Tokyo, Japan). Cantilever spring constants and optical lever sensitivities were determined by fitting the frequency spectrum of thermal fluctuations, using a Lorentzian function. Imaging uses a peak force of 300 pN, loading rate of 2 μm/s, and a sampling rate of 20 kHz. Images were scanned over different fields of view and with pixel sizes between 4 and 10 nm (indicated for each image).

AFM data analysis

The post-processing of the AFM data was carried out using SPIP software (v.6.4, Image Metrology, Hørsholm, Denmark) to smooth and line-correct the images. Next, the images were analyzed using custom-written Python code (available from the authors upon request) to determine the radii of gyration for the core of the IN-DNA clusters and the clusters themselves.

Analysis of AFM based force-volume data

Processing of force volume maps was performed using MountainsSPIP v9 (Digital Surf) software and includes baseline z-correction using a first order polynomial, and x-alignment of the curves using the approach segment of the force curves. The contact point for x-alignment is taken the first crossing of the deflection (y-) axis. The measured piezo height is converted to lateral tip position by correction for the bending of the cantilever. To this end the cantilever deflection (measured in units of length) is subtracted from the piezo height.

Individual force curves were analyzed in terms of stiffness, elastic deformation energy, and plastic deformation energy. Stiffness k was determined by taking the slope of a linear fit to the approach segment within the force regime of 0-200 pN. Elastic deformation energy $E_{Elast.}$, (defined as the area enclosed by the retract segment above the x-axis, i.e. at zero force), and the plastic deformation energy $E_{Plast.}$ (defined as the area enclosed by the approach segment above the y-axis), were used to define a viscoelasticity index

$$\eta = 1 - \frac{\mathrm{E}_{Elast.}}{\mathrm{E}_{Plast.}+\mathrm{E}_{Elast.}}.$$

Maps of the spatially resolved stiffness and deformation energies were generated and used to identify and extract areas corresponding to nucleoprotein complexes. The pixel values were pooled per particle and the midpoint of the Abott-Firestone curve (i.e. at 50% material ratio) was used to determine means and standard deviations. Significance tests used two-sample t-test.

Magnetic tweezers setup

Force spectroscopy experiments were performed on a custom-built MT setup described previously(*51*). Two magnets (5 × 5 × 5 mm^3; W-05-N50-G, Supermagnets, Switzerland) are mounted vertically(*34*, *52*) on a motorized arm with a translational motor (M-126.PD2 motor with C-863.11-Mercury controller, PI, Germany) and a rotational motor (C-150.PD motor with C-863.11-Mercury controller, PI, Germany) to control the rotation and z-position of the magnets. The outlet of the flow cell is connected via an adapter to a peristaltic pump (ISM832C, Ismatec) for fluid flushing. The setup is controlled by Lab-VIEW software (National Instruments) described previously (*53*). The flow cells were constructed from two glass coverslips (24 × 60 mm, Carl Roth, Germany). The lower coverslip was pre-functionalized with (3-glycidoxypropyl)trimethoxysilane (abcr GmbH, Germany) as described(*54*). The silanized slides were then covered with 75 μL of a 5000-fold diluted stock solution of polystyrene beads (Polysciences, USA) in ethanol (Carl Roth, Germany) and slowly air dried to later serve as reference beads for drift correction. The upper coverslip was equipped with two holes, ~1 mm in diameter, drilled with a laser cutter to allow for fluid exchange in the flow cell. The lower and

upper coverslips were assembled with a single layer of parafilm (Carl Roth, Germany), pre-cut to form a ∼50 µL channel connecting the inlet and outlet ports of the flow cell, on a heating plate at 80 °C for 30s. After flow cell preparation, 100 µl of 200 µg/mL anti-digoxigenin (Abcam, Germany) in 1× phosphate-buffered saline (PBS) was introduced into the flow cell and incubated overnight. The flow cell was then rinsed with 800 µL of 1× PBS and passivated with BSA (25 mg/mL; Carl Roth, Germany) for 1 hour to minimize non-specific interactions. The flow cell was then rinsed with 800 µL of 1× PBS to remove any residual material.

DNA constructs and magnetic beads for magnetic tweezers

For the MT experiments, we used a 21-kb DNA construct, prepared as described previously (*55*). The two ends of the DNA segments were provided with handles (∼600 bp) containing several biotins and several digoxigenins, respectively, to bind magnetic beads and the bottom surface of the flow cell. The magnetic beads used were streptavidin-coated MyOne beads of 1 µm diameter (Thermo Fisher Scientific, Germany). The DNA construct was attached to the beads by incubating 0.5 µL picomolar DNA stock solution and 2 µL of MyOne beads in 200 µL PBS for 30 s. The DNA bead solution was then introduced into the flow cell to allow the multiple digoxigenin-anti digoxigenin bonds to form on the surface. The flow cell was then rinsed with 1 mL PBS to remove any unbound beads.

Magnetic tweezers measurements protocol

Before introducing IN to the flow cell, DNA-bound beads were examined for the presence of multiple DNA tethers by measuring their response to force and torque. To find out whether a magnetic bead is bound to the surface by more than one DNA tether, we introduce negative rotations under high tension (F = 5 pN). In the case of a single double-stranded DNA, this level of force prevents the formation of plectonemes upon underwinding due to DNA melting, and consequently no height change is observed. In contrast, if a bead is bound to two or more double-stranded DNA molecules, braiding occurs when the bead is rotated, resulting in a decrease in z-extension. Beads bound by multiple tethers are excluded from further analysis.

The length of the DNA tethers was checked by varying the force from 5 pN to <0.1pN. At 5 pN the DNA constructs are stretched to >95% of the contour length, whereas at <0.1 pN they drop to the surface and the extension is close to zero. Only the tethers of a length close to the expected contour length (~7 µm) were considered for the experiment. Once the DNA tethers for the experiment were selected, the flow cell was flushed with 800 µl of low sodium buffer (LSB, 25 mM Tris-HCl pH= 7.6; 5 mM $MgCl_2$; 50 mM NaCl; pH= 7.6), followed by 50 µl of 2 mM IN in LSB flushed at 1.7 µl/min, while holding the magnet to exert a force of 5 pN onto the beads. Subsequently, the force was lowered to have a control value of the tether length at 1pN for the analysis. The magnet was next moved to exert a very low force (>0.1 pN) in order to allow for integrase-DNA interaction. Force-jump measurements then involved magnet movement after 5 min of incubation, to cause a sudden force jump to 1 pN. The z-extension of the DNA tethers was monitored over 3-4 h or overnight. For force-extension measurements, the magnet is moved stepwise to induce force-increments of 0.1 pN (for forces smaller than 1 pN) and 0.5 pN (for forces larger than 1 pN). Extension-time traces are recorded at each force plateau for 60 min (for forces smaller than 1 pN) and 30 min (for forces larger than 1 pN).

Analysis of magnetic tweezers data

Magnetic tweezers data were analyzed using custom written code in Matlab (R2022a, Mathworks). Force-extension measurements were analyzed by correcting the time traces for drift by subtracting the (x,y,z) time traces of surface attached reference beads. For each force plateau, the mean extension was computed and plotted against previously calibrated forces. To enable averaging over multiple beads and account for different offsets, we normalize the extension with respect to the end-end distance for each bead at F = 5 pN. In order to identify deviations from the worm-like chain induced by IN, we compare the data in presence of IN to the fitted (*56*) force extension behavior in the absence of IN.

For force-jump measurements, the individual tethers time traces were drift corrected using the x,y,z-coordinates of reference beads, de-spiked by linear interpolation using the *filloutliers* function in Matlab, and smoothed using a Butterworth filter. The resulting extension-time traces are numerically differentiated to yield velocity-time traces, and subsequently analyzed using the peak finder function. Peaks in the velocity data correspond to "extension bursts" where the effective DNA length increases, due to de-compaction of the DNA from DNA-IN clusters. Dwell times between bursts are calculated from successive peak locations, burst velocities are evaluated at the peak maximum, and burst sizes are calculated by peak integration.

Monte Carlo simulations of DNA-IN interactions and conformations

We simulated the DNA-IN systems using coarse-grained Monte Carlo (MC) simulations. We refer to reference *(57)* for further details. In the simulations, we modelled the DNA with a bead size of 12 bp of DNA. The DNA beads interact via an excluded-volume interaction to prevent overlapping and subsequent beads are connected by springs. To reproduce the experimentally determined bending stiffness of DNA, we introduce a bending potential, which penalizes sharp bends between subsequent springs. The IN proteins are modelled as beads of the same size as the DNA beads. The IN beads correspond to tetramers and can bind to DNA beads via an isotropic, short-ranged binding potential, where the tetrameric nature of IN is captured by imposing a maximum valence of 4, i.e. each IN bead can bind to at most four DNA beads. To determine the binding strength between IN and DNA, we systematically simulated the binding probability of IN to DNA as a function of the binding strength. These data are then compared to experimental data, where we have directly determined the binding probability from AFM images (Supporting Figure Fig. S22). Briefly, we quantified the fraction of IN-DNA complexes with respect to the total number of binding sites (assuming a site size of 16 bp, the linear DNA has 1000 – 16 + 1 = 985 sites per molecule). We quantify the fraction of bound sites in the presence of 100 nM and 200 nM IN. The best fitting binding energy of IN to DNA was determined to be 5 $k_B T$ (Fig. 3 in the main text).

In order to investigate the role of the IN-IN interaction in the IN-induced compaction of DNA, we considered two cases for IN-IN interaction: (i) a purely excluded-volume interaction (same as for the DNA beads), and (ii) an excluded-volume interaction with an added short-ranged attraction. For both cases, we performed multiple simulations for a range of IN concentrations and DNA lengths. In all simulations, the DNA was simulated in the canonical ensemble, while we simulated the IN in the grand-canonical ensemble to ensure that the free IN is kept constant at the value assigned at the start of the simulation. From the simulations, we obtained time series showing different compaction pathways. To classify these pathways and the individual states along these pathways, we developed an unsupervised machine learning pipeline. For this we composed a training dataset of roughly 15,000 snapshots from the simulation time series. Then, for each snapshot in this dataset, we computed 15 different parameters which captured the

structural configuration of the DNA-IN complex, including the radius of gyration of the DNA chain, the DNA bend angles, the number of proteins bound to the DNA, and the number and size of protein clusters. Next, we used principal component analysis (PCA) to reduce the dimensionality of the 15-dimensional input. We found that 3 dimensions were sufficient to retain all important features of the input data. Furthermore, we observed distinct clusters in this 3-dimensional PC distribution. Therefore, we used a clustering algorithm –specifically a Gaussian mixture model (GMM)– to provide an unsupervised and objective clustering of the 3-dimensional space into seven categories for (compacted) DNA-IN complexes. Using the trained unsupervised machine learning pipeline, we then classified every configuration of each time series.

Data analysis and statistics

AFM force-volume data we statistically assessed using a 2-sample *t*-test. Statistical significance of the effect of inhibitor CX014442 was tested using 2-sample Kolmogorov-Smirnov test. All statistical tests were performed in Matlab (Mathworks). *P* values of <0.05 were considered significant.

References

1. S. G. Sarafianos, B. Marchand, K. Das, D. M. Himmel, M. A. Parniak, S. H. Hughes, E. Arnold, Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J. Mol. Biol.* **385**, 693–713 (2009).

2. E. A. Abbondanzieri, G. Bokinsky, J. W. Rausch, J. X. Zhang, S. F. J. Le Grice, X. Zhuang, Dynamic binding orientations direct activity of HIV reverse transcriptase. *Nature*. **453**, 184–189 (2008).

3. R. Craigie, "Nucleoprotein Intermediates in HIV-1 DNA Integration: Structure and Function of HIV-1 Intasomes" in *Virus Protein and Nucleoprotein Complexes* (Springer, Singapore, 2018), pp. 189–210.

4. D. O. Passos, M. Li, R. Yang, S. V Rebensburg, R. Ghirlando, Y. Jeon, N. Shkriabai, M. Kvaratskhelia, R. Craigie, D. Lyumkis, Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science*. **355**(6320), 89-92.

5. W. Vanderlinden, T. Brouns, P. U. Walker, P. J. Kolbeck, L. F. Milles, W. Ott, P. C. Nickels, Z. Debyser, J. Lipfert, The free energy landscape of retroviral integration. *Nat. Commun*. **10**, 4738 (2019).

6. B. Van Maele, Z. Debyser, HIV-1 integration: an interplay between HIV-1 integrase, cellular and viral proteins. *AIDS Rev*. **7**, 26–43 (2005).

7. P. Cherepanov, D. Surratt, J. Toelen, W. Pluymers, J. Griffith, E. De Clercq, Z. Debyser, Activity of recombinant HIV-1 integrase on mini-HIV DNA. *Nucleic Acids Res*. **27**, 2202–2210 (1999).

8. R. C. Burdick, C. Li, M. Munshi, J. M. O. Rawson, K. Nagashima, W.-S. Hu, V. K. Pathak, HIV-1 uncoats in the nucleus near sites of integration. *Proc. Natl. Acad. Sci.* **117**, 5486–5493 (2020).

9. V. Zila, E. Margiotta, B. Turoňová, T. G. Müller, C. E. Zimmerli, S. Mattei, M. Allegretti, K. Börner, J. Rada, B. Müller, M. Lusic, H.-G. Kräusslich, M. Beck, Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell*. **184**, 1032-1046.e18 (2021).

10. J. A. G. Briggs, M. N. Simon, I. Gross, H.-G. Kräusslich, S. D. Fuller, V. M. Vogt, M. C. Johnson, The stoichiometry of Gag protein in HIV-1. *Nat. Struct. Mol. Biol*. **11**, 672–675 (2004).

11. A. Ballandras-Colas, D. P. Maskell, E. Serrao, J. Locke, P. Swuec, S. R. Jónsson, A. Kotecha, N. J. Cook, V. E. Pye, I. A. Taylor, V. Andrésdóttir, A. N. Engelman, A. Costa, P. Cherepanov, A supramolecular assembly mediates lentiviral DNA integration. *Science* **355**, 93-95 (2017)

12. J. J. Kessl, S. B. Kutluay, D. Townsend, S. Rebensburg, A. Slaughter, R. C. Larue, N. Shkriabai, N. Bakouche, J. R. Fuchs, P. D. Bieniasz, M. Kvaratskhelia, HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis. *Cell*. **166**, 1257-1268.e12 (2016).

13. S. De Houwer, J. Demeulemeester, W. Thys, S. Rocha, L. Dirix, R. Gijsbers, F. Christ, Z. Debyser, The HIV-1 Integrase Mutant R263A/K264A Is 2-fold Defective for TRN-SR2 Binding and Viral Nuclear Import *. *J. Biol. Chem.* **289**, 25351–25361 (2014).

14. X. Li, Y. Koh, A. Engelman, Correlation of Recombinant Integrase Activity and Functional Preintegration Complex Formation during Acute Infection by Replication-Defective Integrase Mutant Human Immunodeficiency Virus. *J. Virol*. **86**, 3861-3879 (2012).

15. M. D. Miller, C. M. Farnet, F. D. Bushman. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J. Virol.* **71** (1997).

16. M. Lelek, F. Di Nunzio, R. Henriques, P. Charneau, N. Arhel, C. Zimmer, Superresolution imaging of HIV in infected cells with FlAsH-PALM. *Proc. Natl. Acad. Sci.* **109**, 8564-8569 (2012).

17. C. A. Brackley, S. Taylor, A. Papantonis, P. R. Cook, D. Marenduzzo, Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization. *Proc. Natl. Acad. Sci*. **110**, E3605–E3611 (2013).

18. M. Ancona, C. A. Brackley, Simulating the chromatin-mediated phase separation of model proteins with multiple domains. *Biophys. J.* **121**, 2600–2612 (2022).

19. JK. Ryu, C. Bouchout, H.W. Liu, E. Kim, M. Minamino, R. de Groot. Ralph, A. J. Katan, A. Bonato, D. Marenduzzo, D. Michieletto, F. Uhlman, C. Dekker, Bridging-induced phase separation induced by cohesin SMC protein complexes. *Sci. Adv.* **7**, eabe5905 (2022).

20. J. H. Gibcus, K. Samejima, A. Goloborodko, I. Samejima, N. Naumova, J. Nuebler, M. T. Kanemaki, L. Xie, J. R. Paulson, W. C. Earnshaw, L. A. Mirny, J. Dekker, A pathway for mitotic chromosome formation. *Science*. **359**, eaao6135 (2018).

21. Y. Yang, L. E. Sass, C. Du, P. Hsieh, D. A. Erie, Determination of protein–DNA binding constants and specificities from statistical analyses of single molecules: MutS–DNA interactions. *Nucleic Acids Res.* **33**, 4322–4334 (2005).

22. J. J. Kessl, J. C. McKee, O. J. Eidahl, N. Shkriabai, A. Katz, M. Kvaratskhelia, HIV-1 Integrase-DNA Recognition Mechanisms. *Viruses* . **1** (2009).

23. D. J. Hazuda, P. J. Felock, J. C. Hastings, B. Pramanik, A. L. Wolfe, Discovery and Analysis of Inhibitors of the Human Immunodeficiency Integrase. *Drug Des Discov.* **15**(1),17-24 (1997).

24. C. Rivetti, M. Guthold, C. Bustamante, Scanning Force Microscopy of DNA Deposited onto Mica: Equilibration *versus* Kinetic Trapping Studied by Statistical Polymer Chain Analysis. *J. Mol. Biol.* **264**, 919–932 (1996).

25. T. Brouns, H. De Keersmaecker, S. F. Konrad, N. Kodera, T. Ando, J. Lipfert, S. De Feyter, W. Vanderlinden, Free Energy Landscape and Dynamics of Supercoiled DNA by High-Speed Atomic Force Microscopy. *ACS Nano* **12**(12),11907-11916 (2018).

26. L. S. Shlyakhtenko, A. A. Gall, A. Filonov, Z. Cerovac, A. Lushnikov, Y. L. Lyubchenko, Silatrane-based surface chemistry for immobilization of DNA, protein-DNA complexes and other biological materials. *Ultramicroscopy* **97** (1-4),279-287. (2003).

27. J. Lipfert, S. Doniach, R. Das, D. Herschlag, Understanding Nucleic Acid–Ion Interactions. *Annu. Rev. Biochem.* **83**, 813–841 (2014).

28. D. Alsteens, D. J. Müller, Y. F. Dufrêne, Multiparametric Atomic Force Microscopy Imaging of Biomolecular and Cellular Systems. *Acc. Chem. Res.* **50**, 924–931 (2017).

29. R. Garcia, Nanomechanical mapping of soft materials with the atomic force microscope: methods, theory and applications. *Chem. Soc. Rev.* **49**, 5850–5884 (2020).

30. H.-J. Butt, B. Cappella, M. Kappl, Force measurements with the atomic force microscope: Technique, interpretation and applications. *Surf. Sci. Rep.* **59**, 1–152 (2005).

31. S. Wegmann, I. D. Medalsy, E. Mandelkow, D. J. Müller, The fuzzy coat of pathological human Tau fibrils is a two-layered polyelectrolyte brush. *Proc. Natl. Acad. Sci.* **110**, E313–E321 (2013).

32. S. Boeynaems, A. S. Holehouse, V. Weinhardt, D. Kovacs, J. Van Lindt, C. Larabell, L. Van Den Bosch, R. Das, P. S. Tompa, R. V Pappu, A. D. Gitler, Spontaneous driving forces give rise to protein−RNA condensates with coexisting phases and complex material properties. *Proc. Natl. Acad. Sci.* **116**, 7889–7898 (2019).

33. N. Kol, Y. Shi, M. Tsvitov, D. Barlam, R. Z. Shneck, M. S. Kay, I. Rousso, A Stiffness Switch in Human Immunodeficiency Virus. *Biophys. J.* **92**, 1777–1783 (2007).

34. J. Lipfert, X. Hao, N. H. Dekker, Quantitative Modeling and Optimization of Magnetic Tweezers. *Biophys. J.* **96**, 5040–5049 (2009).

35. D. A. Koster, V. Croquette, C. Dekker, S. Shuman, N. H. Dekker, Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB. *Nature*. **434**, 671–674 (2005).

36. H. Bai, J. E. Kath, F. M. Zörgiebel, M. Sun, P. Ghosh, G. F. Hatfull, N. D. F. Grindley, J. F. Marko, Remote control of DNA-acting enzymes by varying the Brownian dynamics of a distant DNA end. *Proc. Natl. Acad. Sci.* **109**, 16546–16551 (2012).

37. H. P. Lu, L. Xun, X. S. Xie, Single-Molecule Enzymatic Dynamics. *Science*. *282*, 1877–1882 (1998).

38. P. C. Koneru, A. C. Francis, N. Deng, S. V Rebensburg, A. C. Hoyte, J. Lindenberger, D. Adu-Ampratwum, R. C. Larue, M. F. Wempe, A. N. Engelman, D. Lyumkis, J. R. Fuchs, R. M. Levy, G. B. Melikyan, M. Kvaratskhelia, HIV-1 integrase tetramers are the antiviral target of pyridine-based allosteric integrase inhibitors. *Elife*. **8**, e46344 (2019).

39. F. Christ, S. Shaw, J. Demeulemeester, B.A. Desimmie, A. Marchand, S. Butler, W. Smets, P. Chaltin, M. Westby, Z. Debyser, C. Pickford, Small-Molecule Inhibitors of the LEDGF/p75 Binding Site of Integrase Block HIV Replication and Modulate Integrase Multimerization. *Antimicrob. Agents Chemother*. **56**, 4365–4374 (2012).

40. L. S. Vranckx, J. Demeulemeester, S. Saleh, A. Boll, G. Vansant, R. Schrijvers, C. Weydert, E. Battivelli, E. Verdin, A. Cereseto, F. Christ, R. Gijsbers, Z. Debyser, LEDGIN-mediated Inhibition of Integrase-LEDGF/p75 Interaction Reduces Reactivation of Residual Latent HIV. *eBioMedicine*. **8**, 248–264 (2016).

41. K. Gupta, A. Allen, C. Giraldo, G. Eilers, R. Sharp, Y. Hwang, H. Murali, K. Cruz, P. Janmey, F. Bushman, G. D. Van Duyne, Allosteric HIV Integrase Inhibitors Promote Formation of Inactive Branched Polymers via Homomeric Carboxy-Terminal Domain Interactions. *Structure*. **29**, 213-225.e5 (2021).

42. S. Brahmachari, J. F. Marko, Chromosome disentanglement driven via optimal compaction of loop-extruded brush structures. *Proc. Natl. Acad. Sci. U. S. A.* **116**(50), 24956-24965 (2019).

43. J. Harju, M. C. F. van Teeseling, C. P. Broedersz, *bioRxiv* (2023) doi:10.1101/2023.06.30.547230.

44. B. M. S., D. C. W., C. S. A., Role of Human Immunodeficiency Virus Type 1 Integrase in Uncoating of the Viral Core. *J. Virol.* **84**, 5181–5190 (2010).

45. H. Gien, M. Morse, M. J. McCauley, J. P. Kitzrow, K. Musier-Forsyth, R. J. Gorelick, I. Rouzina, M. C. Williams, HIV-1 Nucleocapsid Protein Binds Double-Stranded DNA in Multiple Modes to Regulate Compaction and Capsid Uncoating. *Viruses*. **14** (2022).

46. K. Jiang, N. Humbert, S. K.K., I. Rouzina, Y. Mely, F. Westerlund, The HIV-1 nucleocapsid chaperone protein forms locally compacted globules on long double-stranded DNA. *Nucleic Acids Res.* **49**, 4550–4563 (2021).

47. S. Lyonnais, R. J. Gorelick, F. Heniche-Boukhalfa, S. Bouaziz, V. Parissi, J.-F. Mouscadet, T. Restle, J. M. Gatell, E. Le Cam, G. Mirambeau, A protein ballet around the viral genome orchestrated by HIV-1 reverse transcriptase leads to an architectural switch: From nucleocapsid-condensed RNA to Vpr-bridged DNA. *Virus Res.* **171**, 287–303 (2013).

48. J. Demeulemeester, C. Tintori, M. Botta, Z. Debyser, F. Christ, Development of an AlphaScreen-Based HIV-1 Integrase Dimerization Assay for Discovery of Novel Allosteric Inhibitors. *SLAS Discov.* **17**, 618–628 (2012).

49. M. Bussiek, N. Mücke, J. Langowski, Polylysine□coated mica can be used to observe systematic changes in the supercoiled DNA conformation by scanning force microscopy in solution. *Nucleic Acids Res.* **31**, e137–e137 (2003).

50. W. Vanderlinden, J. Lipfert, J. Demeulemeester, Z. Debyser, S. De Feyter, Structure, mechanics, and binding mode heterogeneity of LEDGF/p75–DNA nucleoprotein complexes revealed by scanning force microscopy. *Nanoscale*. **6**, 4611–4619 (2014).

51. F. Kriegel, N. Ermann, R. Forbes, D. Dulin, N. H. Dekker, J. Lipfert, Probing the salt dependence of the torsional stiffness of DNA by multiplexed magnetic torque tweezers. *Nucleic Acids Res.* **45**, 5920–5929 (2017).

52. F. Kriegel, N. Ermann, J. Lipfert, Probing the mechanical properties, conformational changes, and interactions of nucleic acids with magnetic tweezers. *J. Struct. Biol.* **197**, 26–36 (2017).

53. J. P. Cnossen, D. Dulin, N. H. Dekker, An optimized software framework for real-time, high-throughput tracking of spherical beads. *Rev. Sci. Instrum.* **85**, 103712 (2014).

54. F. Kriegel, W. Vanderlinden, T. Nicolaus, A. Kardinal, J. Lipfert, "Measuring Single-Molecule Twist and Torque in Multiplexed Magnetic Tweezers" in *Nanoscale Imaging: Methods and Protocols* (Springer, New York, 20178), pp. 75–98.

55. J. Lipfert, D. A. Koster, I. D. Vilfan, S. Hage, N. H. Dekker, "Single-Molecule Magnetic Tweezers Studies of Type IB Topoisomerases" in *DNA Topoisomerases: Methods and Protocols* (Humana Press, Totowa, 2009), pp. 71–89.

56. C. Bouchiat, M. D. Wang, J.-F. Allemand, T. Strick, S. M. Block, V. Croquette, Estimating the Persistence Length of a Worm-Like Chain Molecule from Force-Extension Measurements. *Biophys. J.* **76**, 409–413 (1999).

57. M. de Jager, *et al.*, https://doi.org/10.1101/2024.03.28.587201 (2024)

Acknowledgments
We thank Thomas Nicolaus for laboratory assistance, and Enrico Skoruppa, Enrico Carlon, Michael Hafner, Frank Smallenburg, Chris Brackley, Rinske Alkemade, Gerhard Blab, Wolfgang Ott, Paulo Onuchic, and Arthur Ermatov for helpful discussions.

Funding:
Funds for Scientific Research Flanders (FWO) post-doctoral fellowship (WV)
Funds for Scientific Research Flanders (FWO) doctoral fellowship (WF)
Funds for Scientific Research Flanders (FWO) grant G0A5316N
Funds for Scientific Research Flanders (FWO) grant SBO-Saphir
Deutsche Forschungsgemeinschaft (German Research Foundation) grant SFB 863, Project 111166240 A11 (JL)
KU Leuven grant IDO/12/08 (SD, ZD)
KU Leuven grant C14/17/095-3M170311 (ZD)

Author contributions:

Conceptualization	WV
Formal Analysis	PJK, WF, BB, WV
Funding acquisition	ZD, SDF, WV, WF, LF, JL
Investigation	PJK, MdJ, MG, TB, BB,
Methodology	PJK, MdJ, TB, LF, JL, WV
Resources	SVB, FC, SDF, ZD, LF, JL
Software	SFK, WV
Supervision	SDF, LF, JL, WV
Writing – original draft	PJK, MdJ, JL, WV

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data are available in the main text or the supplementary materials.

Supplementary Materials

Supporting Figures S1-S22.
Supporting Data S1