Experimental details

Unless otherwise stated, chemicals were purchased from Sigma-Aldrich (St. Louis, MO). DNA oligonucleotides were purchased from Syntezza (Jerusalem, Israel).

Expression and purification of HIV-1 IN and LEDGF/p75

The histidine-tagged IN and LEDGF/p75 expression vectors were a generous gift from Prof. Alan Engelman (Harvard Medical School, Boston, MA). The expression and purification of LEDGF/p75 and IN 1-288 containing the mutations F185K and C280S were performed as described previously[1, 2].

DNA molecules

dsDNA 753 bp molecules were generated by Polymerase Chain Reaction (PCR). The PCR product was dsDNA containing two viral LTR sequences at their termini and a random sequence in between. Primers were designed to pick up the YFP gene from a pEYFP-c1 plasmid that was used as a template. 21 bp of each U3 and U5 viral LTRs were added by the primers. The LTR sequences were:

U3: 5'-ACTGGAAGGGCTAATTCACTC-3'.

U5: 5'-GTGTGGAAAATCTCTAGCAGT-3'.

Phusion High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA) was used. DNA molecules containing U3 and U5 sequences, U3 only, U5 only and no LTRs, were generated in 30 cycles at the following conditions: 98 °C, 20 sec; 62 °C, 20 sec; 72 °C, 30 sec. For DNA containing the U5 sequence at both ends, the cycling conditions were changed to: 98 °C, 1 min and 72 °C, 50 sec (30 cycles). The primers sequences for all the above DNA molecules are:

Desired LTR in the DNA substrate	Primers sequence
U3+U5	Forward: 5'-ACTGGAAGGGCTAATTCACTCATGGTGAGCAAGGGCGAGGAG-3' Reverse: 5'-ACTGCTAGAGATTTTCCACACCAGCTCGTCCATGCCGAGAG-3'
U5	Forward: 5'-ATGGTGAGCAAGGGCGAGGAG-3' Reverse: 5'-ACTGCTAGAGATTTTCCACACCAGCTCGTCCATGCCGAGAG-3'
U3	Forward: 5'-ACTGGAAGGGCTAATTCACTCATGGTGAGCAAGGGCGAGGAG-3' Reverse: 5'-CAGCTCGTCCATGCCGAGAG-3'
U5+U5	Forward: 5'-ACTGCTAGAGATTTTCCACACATGGTGAGCAAGGGCGAGGAGCTGTTCAC-3' Reverse: 5'-ACTGCTAGAGATTTTCCACACCAGCTCGTCCATGCCGAGAGTGATCCC-3'
No LTRs	Forward: 5'-ATGGTGAGCAAGGGCGAGGAG-3' Reverse: 5'-CAGCTCGTCCATGCCGAGAG-3'

The underlined bases show the LTR sequences and the other part of the primer is the complementary sequence to the start and the end of the YFP gene.

PCR products were confirmed using 1% Agarose Gel Electrophoresis and DNA sequencing. The DNA concentration was determined using NanoDrop spectrophotometer.

Synthesis of the IN inhibitory peptide LEDGF/p75 361-370

The peptide LEDGF/p75 361-370 was synthesized as described previously [3]. The peptide sequence is: WNSLKIDNLDV. The N terminal tryptophan is not part of the original sequence and was added to enable measurement of the peptide concentration by UV spectroscopy.

In vitro integration reaction

The in vitro integration reaction mixture contained 40 µl of 20 mM HEPES pH=7.5, 10 mM MgCl2, 175 mM NaCl, 10 mM DTT, 5% PEG-8000, 10% DMSO, 20 nM DNA and 2.3 µM IN. Each time, fresh reaction ingredients were used. The reaction mixture was incubated at 37 °C for 10, 30, 60, 90, 120 or 180 minutes and terminated by adding sodium dodecyl sulfate (SDS) to 0.5% and heating at 72 °C for 10 minutes. When the LEDGF/p75 protein was used, it was added to the reaction mixture from the beginning at 1:1 (IN:LEDGF/p75) ratio, reaching a final concentration of 2.3 µM. When IN inhibitor was used, 200 µM of the peptide LEDGF/p75 361-370 were incubated for 15 minutes at

37 °C with all the reaction ingredients, except the DNA, and then transferred into ice for 2 minutes. To start the reaction, DNA was added and the mixture was transferred back to 37 °C for 2 hours.

Reaction products were separated by 1% Agarose gel electrophoresis at 100 V for 2 hours in Tris Acetate EDTA (TAE) buffer, and the gels were stained with ethidium bromide. Contrast adjusting and analyzing the gels images, e.g., bands intensity, were performed using ImageJ software (http://imagej.nih.gov/). OriginPro 8 software (Microcal Software, Northhampton, MA) was used for statistical analysis.

Analysis of integration products by AFM imaging

DNA was extracted from the gels using Zymoclean Gel DNA Recovery Kit (Zymo Research Corporation, Orange, CA, USA). A 10 μl drop, containing the extracted DNA and 10 mM MgCl2, was placed on a freshly cleaved mica substrate. After 2 minutes of incubation, the mica was washed with triple distilled water (TDW) and dried with a flow of N2 gas.

Commercial AFM (Nanotec Electronica, Madrid, Spain) was used for imaging. All the samples were morphologically characterized by dynamic mode, using silicon nitride tips, OMCL-RC800PSA (Olympus Optical Co., Tokyo, Japan), with resonant frequency of 71 kHz, spring constant of 0.76 N/m, and a tip diameter of 15 nm, at ambient conditions. Images were analyzed by the WSxM software [4] . OriginPro 8 software was used for statistical analysis.

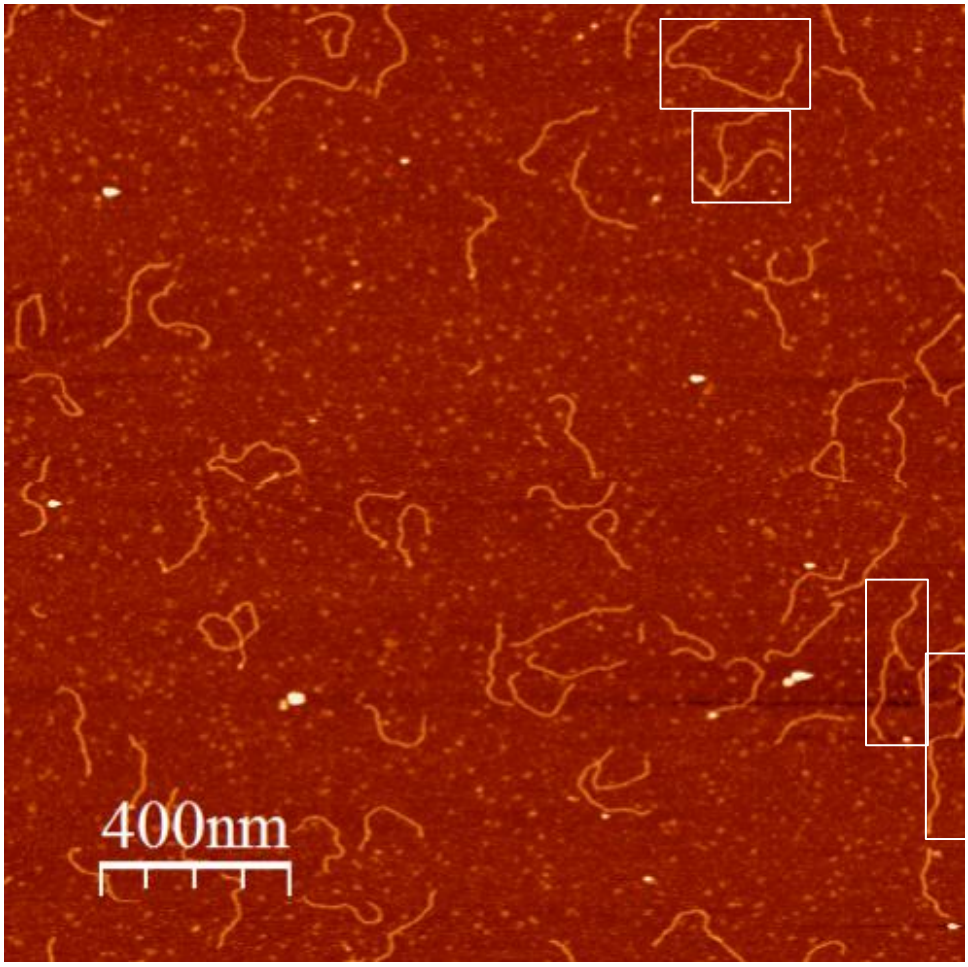

Figure S1. AFM imaging of a reaction mixture. 5 µl were taken out of the 40 µl of the reaction mixture after 2 hours of incubation at 37 °C and directly diluted 10 fold with dilution buffer containing 10 mM $MgCl_2$ and 20 mM HEPES pH=7.5. A 10 µl drop of the diluted reaction mixture was immediately adsorbed on a freshly cleaved mica substrate. The frames indicate molecules corresponding to linear and branched integration products.

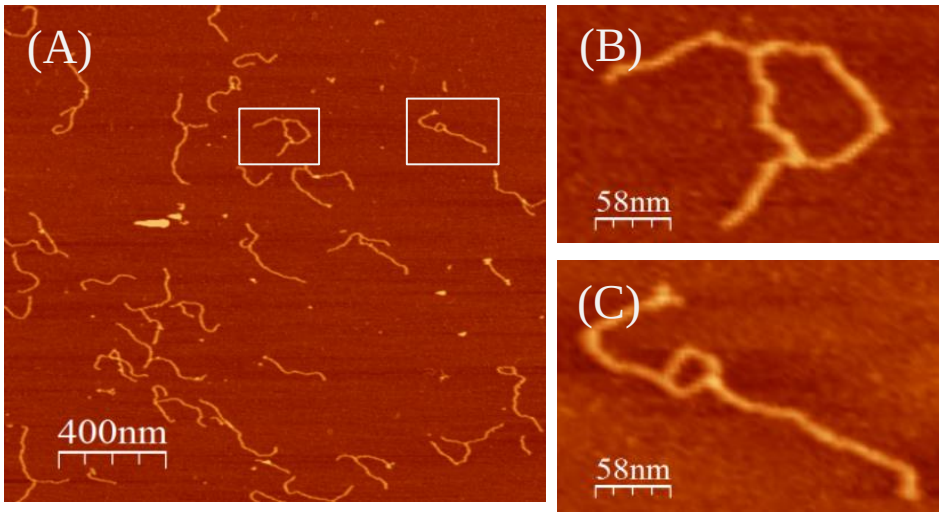

Fig. S2. AFM imaging of rare integration products. (A) *In vitro* integration reaction was performed with DNA molecules containing U3 and U5 viral LTR sequences at their termini. The reaction products were separated on 1% agarose gel. The images are of integrated DNA extracted from the 1500 bp band. One product corresponds to strand transfer of one molecule into another followed by self-integration (B). The other is a product of intramolecular integration followed by strand transfer of another molecule (C).

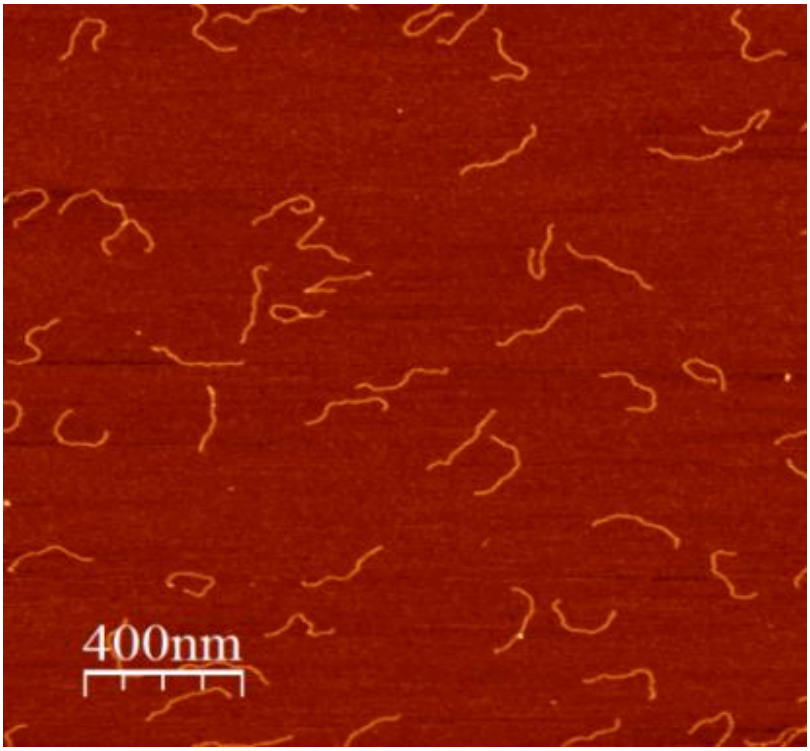

Fig. S3. AFM imaging of unreacted DNA, extracted from the 750 bp band. Each AFM image represents at least 15 different scanned areas, 2-4 μm^2 each.

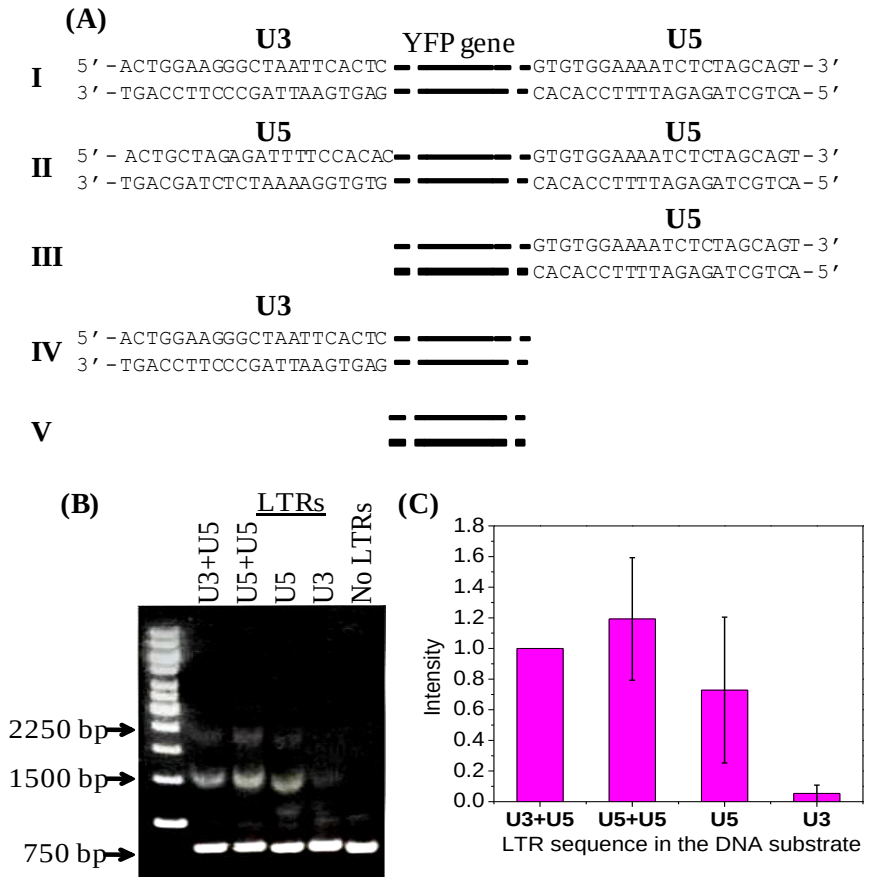

Fig. S4. The integration reaction is dependent on the presence of LTR sequences in the DNA molecule. **(A)** A schematic description of the 750 bp dsDNA PCR products containing different LTR sequences. **I**. U3 and U5 LTRs; **II**. U5 LTR at both termini; **III**. U5 LTR only at one terminus; **IV**. U3 LTR only at one terminus; **V**. No LTRs. The PCR of a DNA molecule with U3 LTR sequence at both termini was unsuccessful. **(B)** Agarose gel electrophoresis separation of the *in vitro* integration reaction products. **(C)** Quantification of the 1500 bp bands intensity. Values were normalized with respect to the 1500 bp band intensity of the reaction with DNA that contained U3 and U5 sequences. The average was calculated based on three independent repeats. The error bars are the standard deviation.

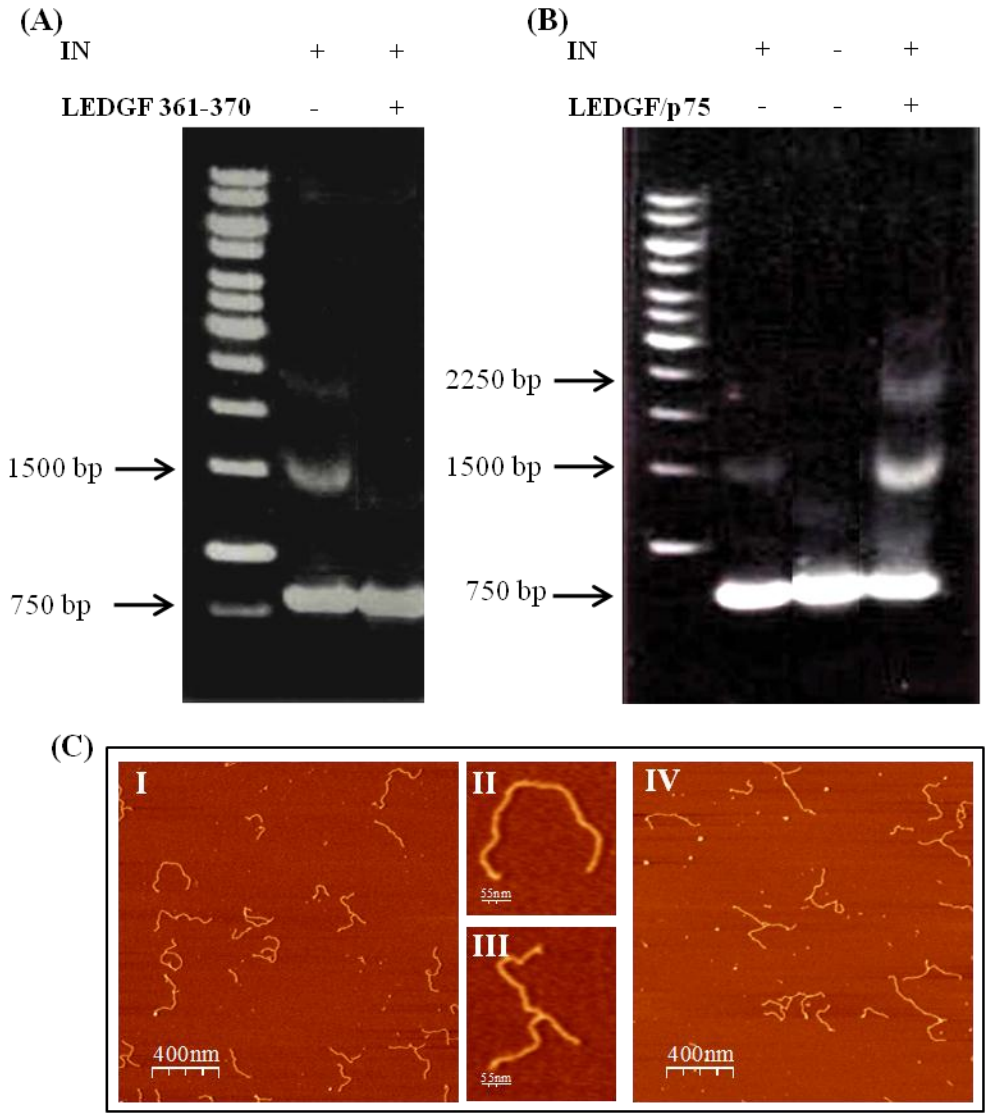

Fig. S5. **(A)** Inhibition of the *in vitro* integration reaction by the peptide LEDGF/p75 361-370. **(B)** Activation of the *in vitro* integration reaction by the cellular protein LEDGF/p75. **(C)** AFM imaging of the extracted DNA from the gel in a reaction with LEDGF/p75. **I**. 1500 bp integrated DNA. **(II, III)**. Close-up views of the 1500 bp integrated DNA are indicated by the white squares. **I**V. 2250 bp integrated DNA. Each AFM image represents at least 15 different scanned areas, 2-4 μm^2 each. The incubation time was 2 hours at 37 °C and DNA with viral U3 and U5 LTRs was used in all cases.

References

1. T. M. Jenkins, A. Engelman, R. Ghirlando and R. Craigie, *Journal of Biological Chemistry*, 1996, **271**, 7712-7718.

2. F. Turlure, G. Maertens, S. Rahman, P. Cherepanov and A. Engelman, *Nucleic Acids Research*, 2006, **34**, 1653.

3. Z. Hayouka, J. Rosenbluh, A. Levin, S. Loya, M. Lebendiker, D. Veprintsev, M. Kotler, A. Hizi, A. Loyter and A. Friedler, *Proceedings of the National Academy of Sciences*, 2007, **104**, 8316-8321.

4. I. Horcas, R. Fernandez, J. Gomez-Rodriguez, J. Colchero, J. G mez-Herrero and A. Baro, *Rev. Sci. Instrum.*, 2007, **78**, 013705.