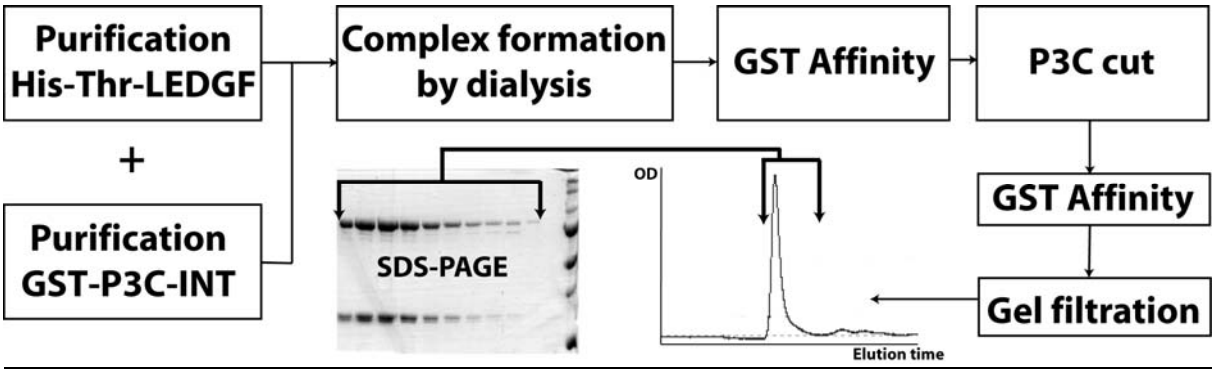
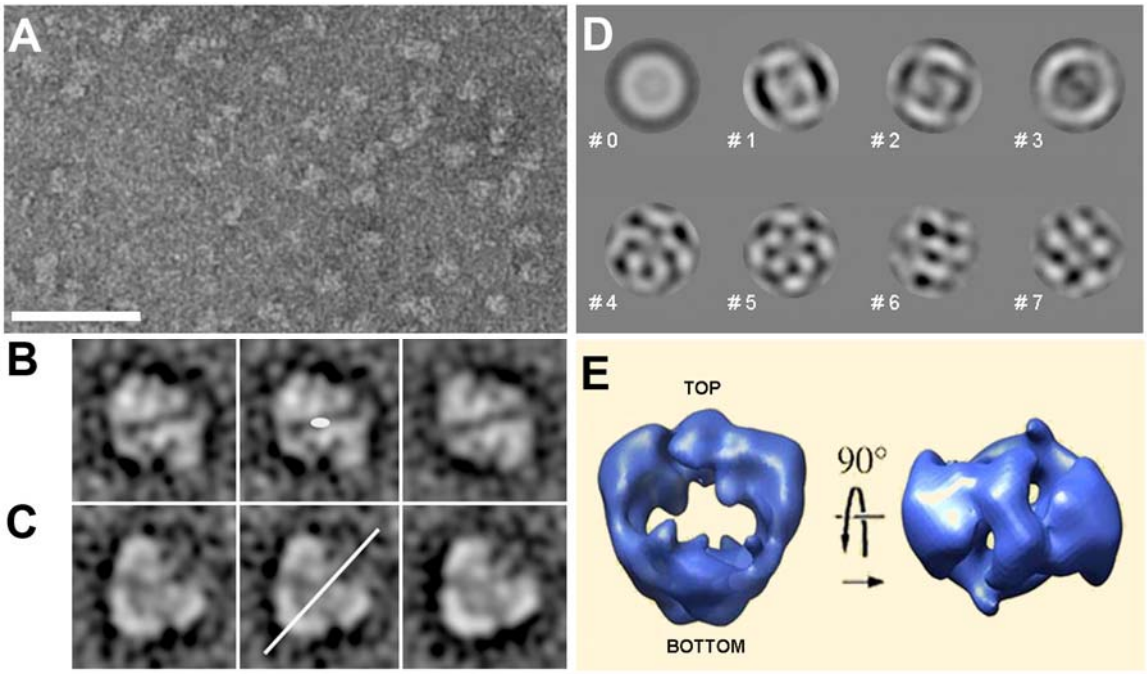
Supplementary Figures 1-10

Supplementary Figure 1: IN/LEDGF complex preparation



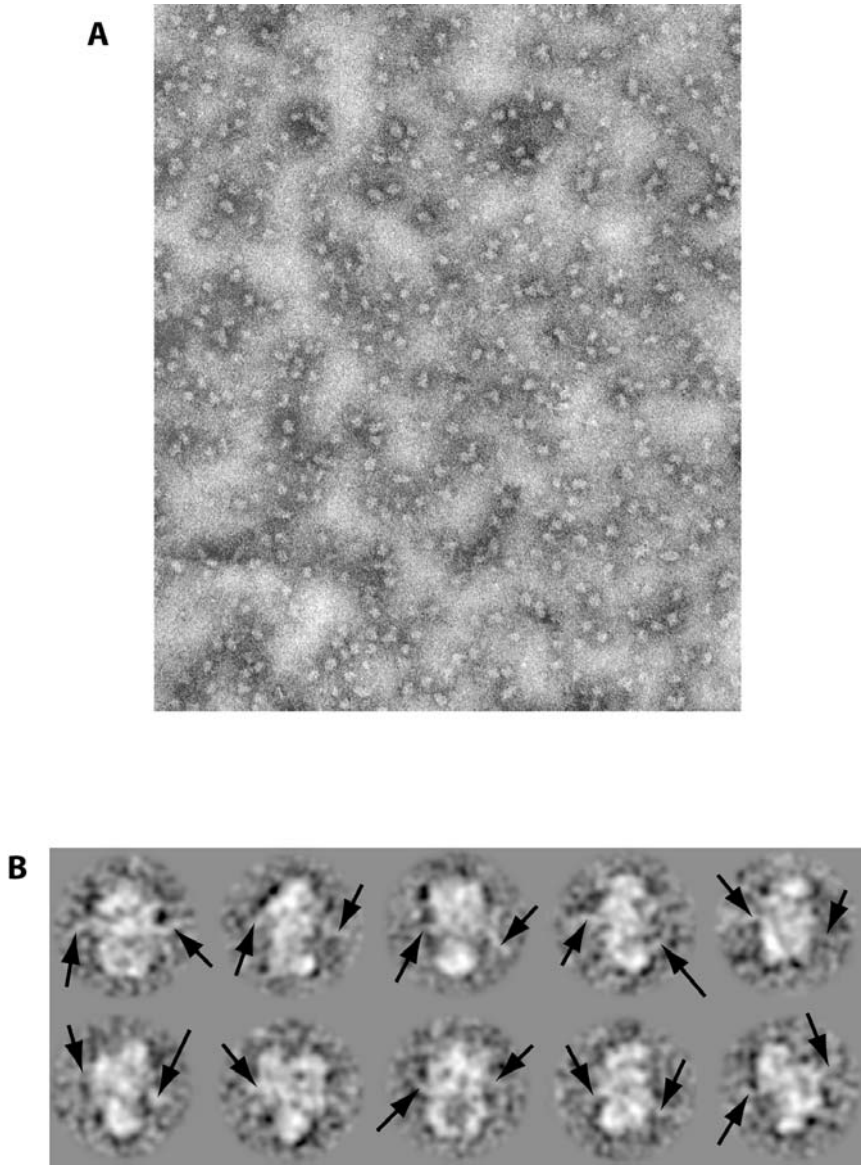
The integrase and LEDGF are purified separately and mixed to assemble the complex. After mixing the complex is formed by removing the solubilization agents by dialysis. The complex is then purified on a GST affinity column; the GST tag is cut by the P3C protease and removed on a GST affinity column. Finally the complex is purified on a gel filtration column. The gel filtration profile and coomassie blue stained SDS-PAGE demonstrate the homogeneity of the IN/LEDGF complex.

<u>Supplementary Figure 2:</u> Starting model obtained from negatively stained IN/LEDGF complexes.



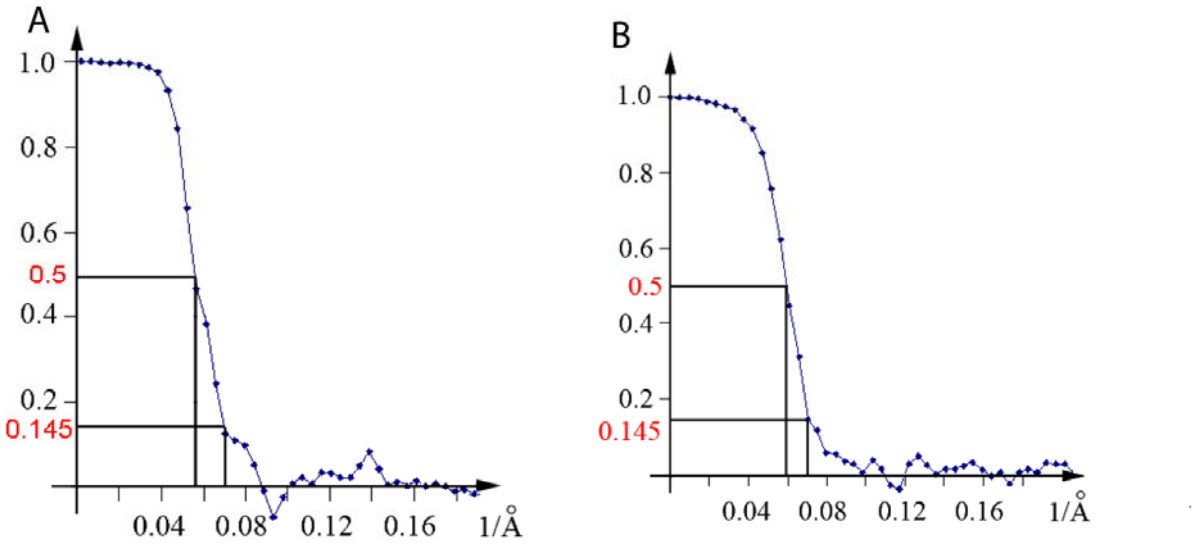
(**A**) Electron micrograph of negatively stained IN-LEDGF complexes. (**B, C**) Views revealing a two fold-symmetry axis within the complex. In (B) the symmetry axis is perpendicular to the plane whereas in (C) the axis is in plane. The left panel represents the original view, the middle panel shows the position of the symmetry axis (highlighted in white) and the view is symmetrized in the right panel. The bar represents 50 nm in (**A**) and 11 nm in (**B, C**). (**D**) Eigenvectors obtained upon multi statistical analysis of the starting, unaligned, image dataset. The two first eigenvectors are similar, show an internal two-fold symmetry and are rotated one with respect to the other by 45°. This indicates that a two-fold symmetry operator is present in the dataset. (**E**) Surface representation of the 3-D envelope of the HIV integrase-LEDGF complex obtained by negative staining

<u>Supplementary Figure 3:</u> Electron microscopy of cryo negatively stained complexes



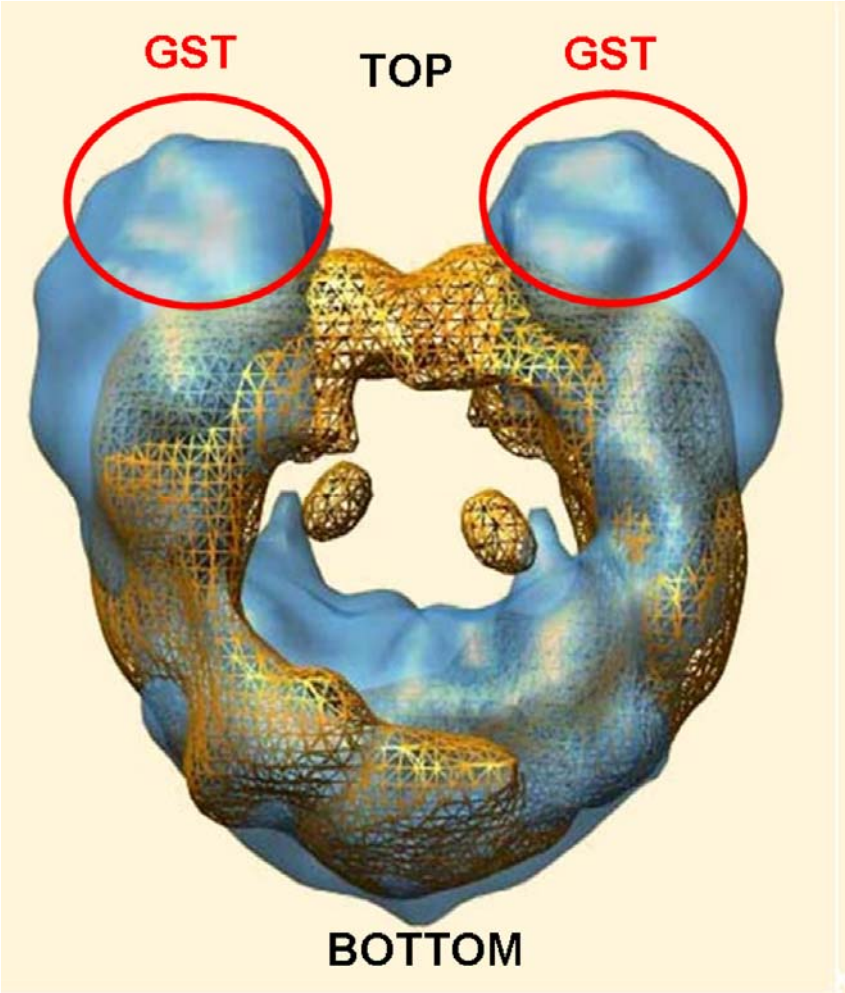
(**A**) Field of cryo negatively stained IN/LEDGF complexes. (**B**) Average images of cryo negatively stained IN/LEDGF complexes incubated with 21bp long oligonucleotides. DNA protruding filaments are marked by arrowheads.

Supplementary Figure 4: Resolution tests



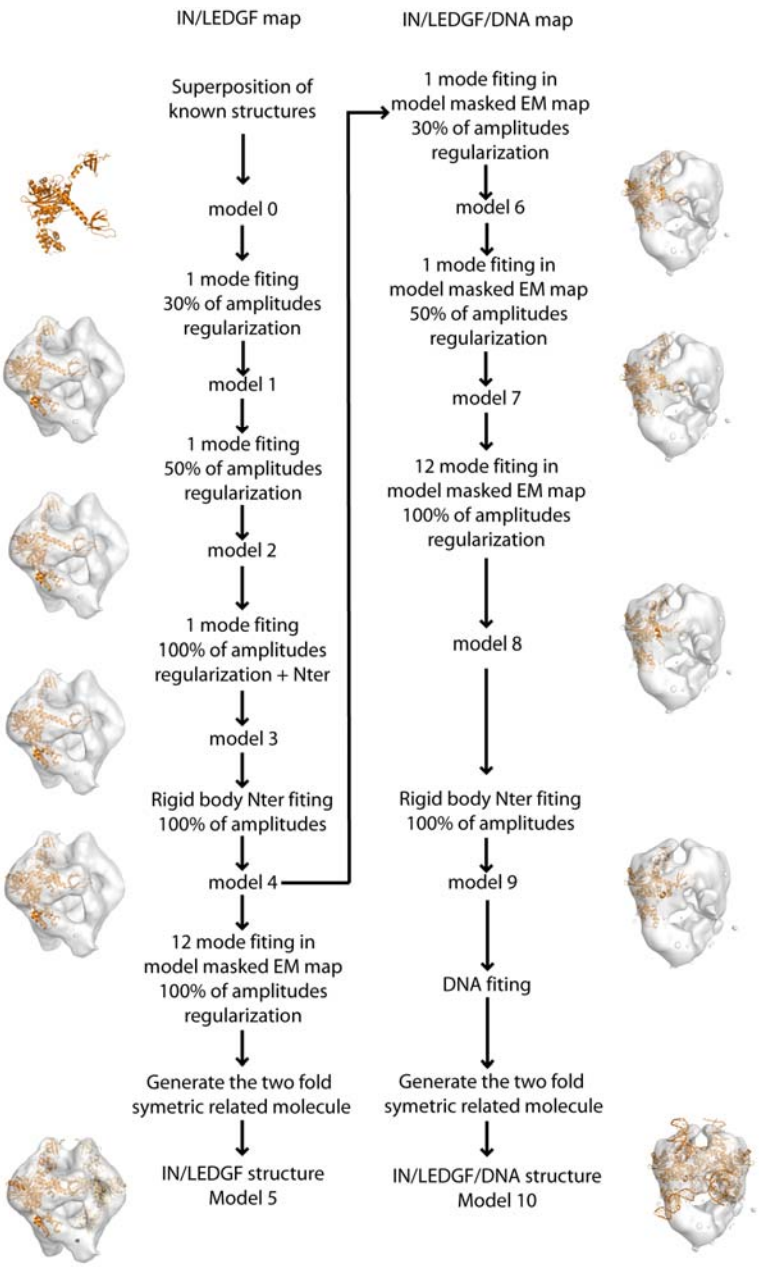
(**A**) Fourier Shell Correlation Function obtained by comparing two distinct IN/LEDGF reconstructions obtained by splitting the data set in two. The 0.5 FSC criterion gives a resolution of 17.5 Å whereas the Rosenthal and Henderson criterion (0.145 FSC) gives a resolution of 14.2 Å and the 3 σ criterion gives a value of 13.8 Å (not shown). (**B**) Fourier Shell Correlation Function obtained for the IN/LEDGF/DNA complex. The 0.5 FSC criterion gives a resolution of 17.2 Å whereas the Rosenthal and Henderson criterion (0.145 FSC) gives a resolution of 14.1 Å

<u>Supplementary Figure 5:</u> Mapping of the N-terminal GST-tagged integrase molecules within the complex.



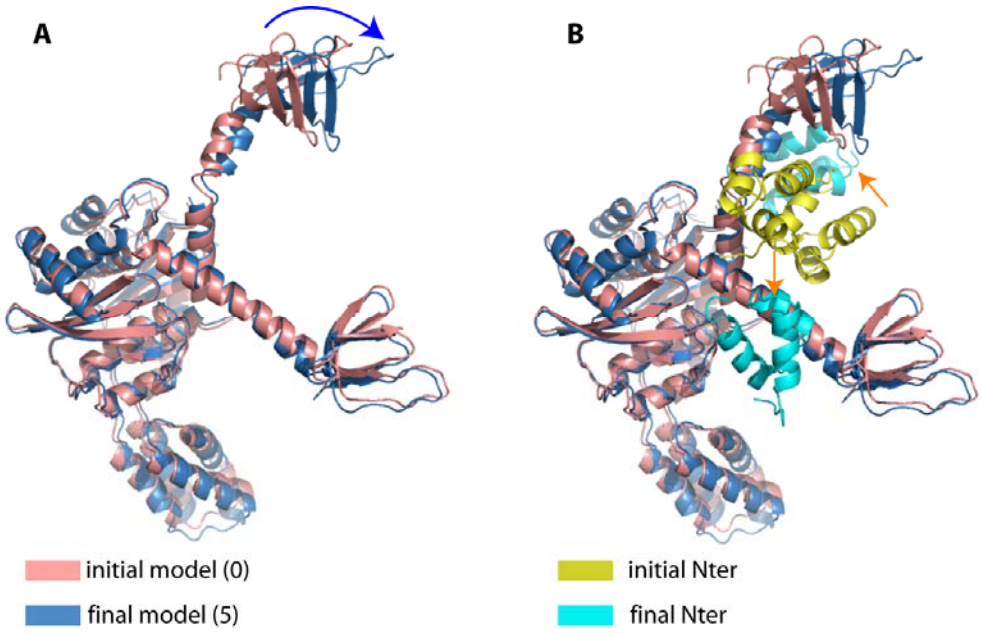
Three dimensional models of two IN/LEDGF complexes calculated from negatively stained molecules. The untagged complex (orange) is superimposed to a complex in which the Integrase is GST-tagged at its N-ternimus (blue). The additional density circled in red corresponds to the GST moieties, revealing the position of the integrase N-termini within the complex.

<u>Supplementary Figure 6:</u> Procedure used for the molecular dynamic flexible fitting.



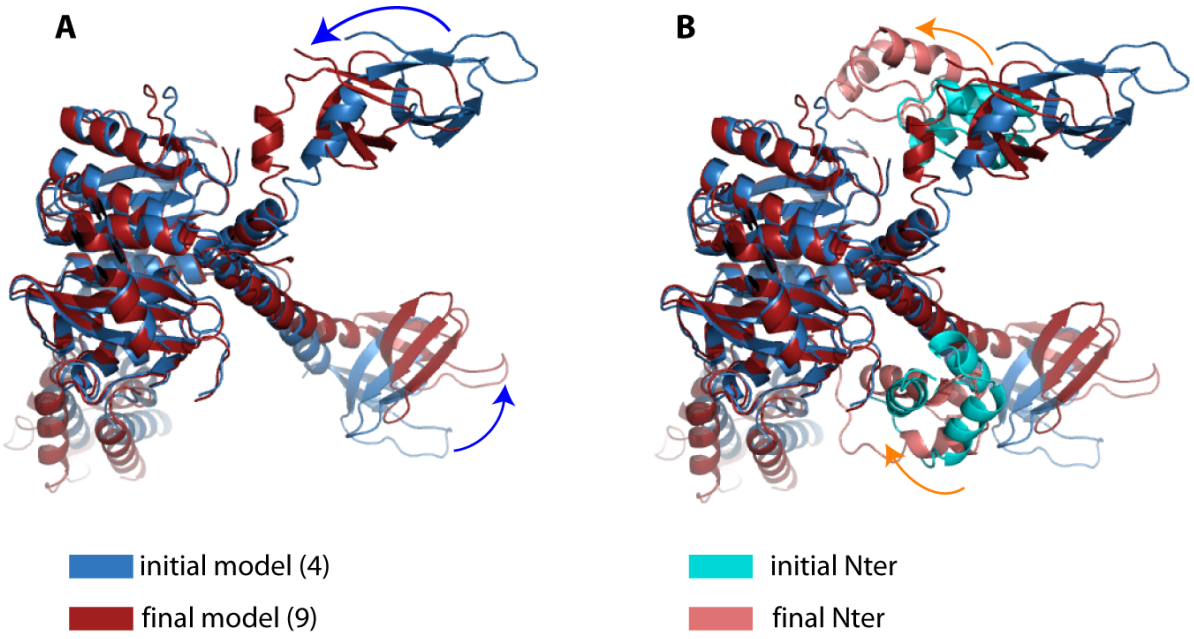
Summary of the fitting procedure. The 3D models are represented at each step of the fitting procedure. When 1 mode fitting is indicated, the fitting has been done using the lowest frequency mode with NORMA (mode 7). In the case of 12 mode fitting, the fitting was done using 12 normal modes with NORMA (mode 7 to 18).

<u>Supplementary Figure 7:</u> Superposition of the initial model (model 0, in pink) and the final model (model 5, in blue) after normal modes flexible fitting into the cryoEM map without DNA.



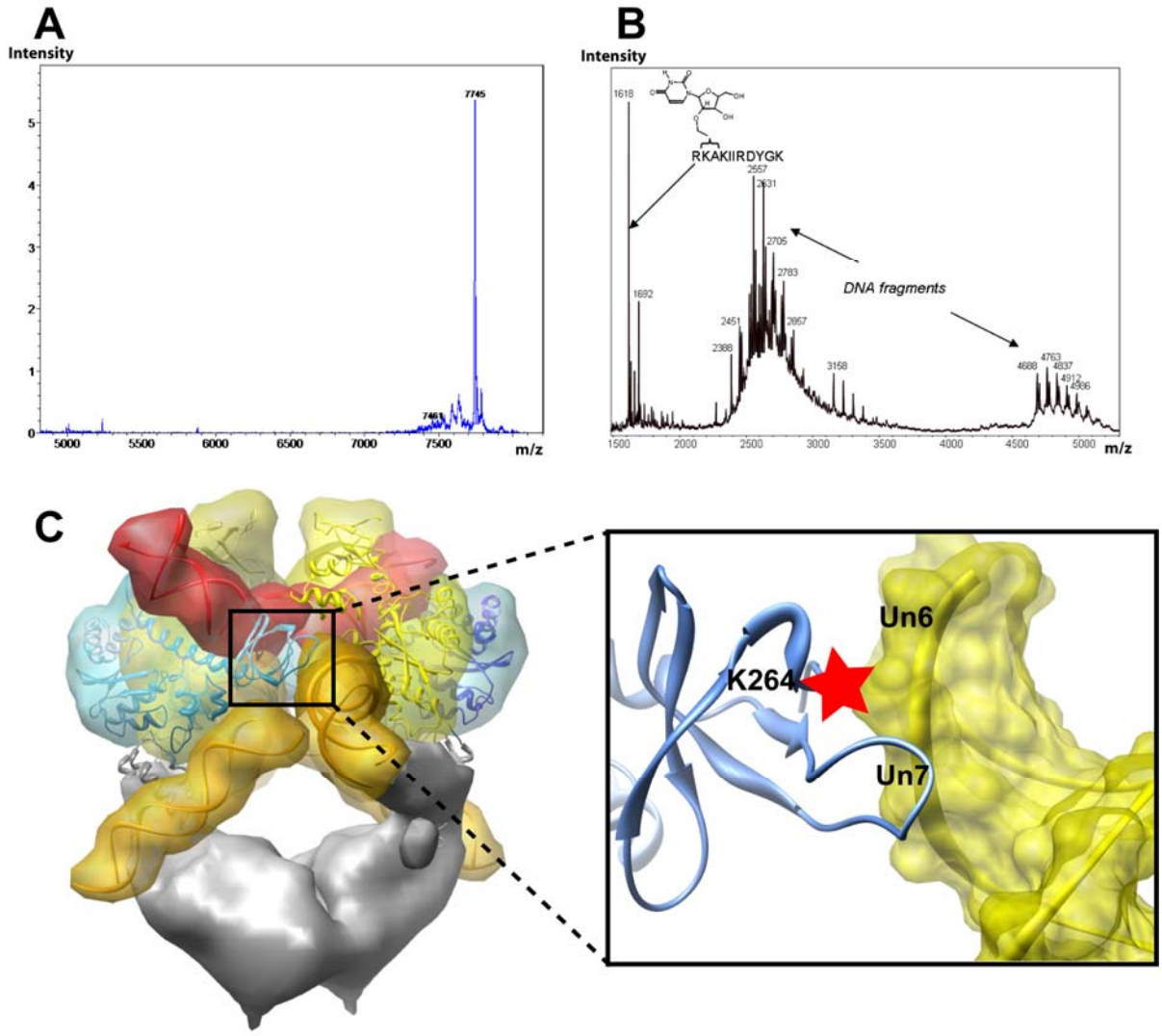
(**A**) Model without the IN-NTD. The blue arrow indicates the displacement of the IN-CTD. (**B**) Model with the IN-NTD. The two IN-NTD are shown in yellow in the initial position and in light blue in the final positions after rigid body fitting. The orange arrows indicate the displacement of the two IN-NTD.

<u>Supplementary Figure 8:</u> Superposition of the initial model (model 4, in blue) and the final model (model 9, in red) after normal modes flexible fitting into the cryoEM map with DNA.



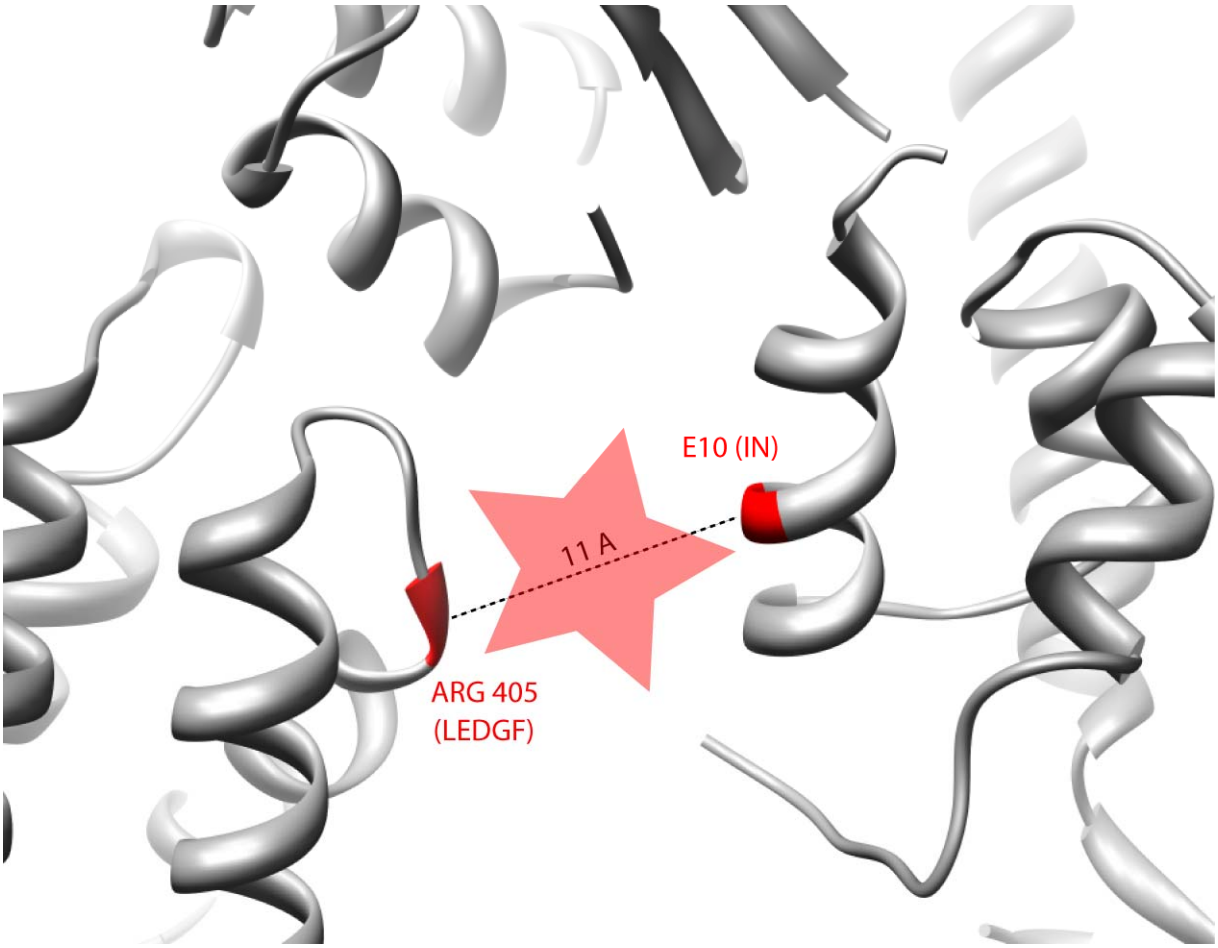
(A) Model without the IN-NTD. The blue arrow indicates the displacement of the IN-CTD. (**B**) Model with the IN-NTD. The two IN-NTD are shown in light blue in the initial positions and in pink in the final positions after rigid body fitting. The orange arrows indicate the displacement of the two IN-NTD.

<u>Supplementary Figure 9:</u> Model validation by DNA-protein cross – linking experiments:



Identification of the peptide crosslinked to DNA by mass spectrometry. (**A**) Mass spectrum of the trypsin digested fragment reveals one major peak. (**B**) Mass spectrum after degradation of the bound oligonucleotide identifying unambiguously the sequence of the peptide (263-273). (**C**) View of the IN/LEDGF/DNA complex. The magnified area shows that in the fitted model the position of K264 is at a distance compatible with cross-linking to the phosphate between nucleotide 6 and 7 of the viral-processed DNA strand.

Supplementary Figure 10:



Close view of the IN(4)LEDGF(2) complex showing the interaction between the CTD and the IBD of LEDGF as published recently by (Hare et al., 2009). The C alpha atoms of the LEDGF R405 and IN E10 are at a distance of 11 Å, compatible with a direct interaction of the side chains as described in (Hare et al., 2009).