Published online 15 February 2019 *Nucleic Acids Research, 2019, Vol. 47, No. 7* ***3607–3618***
doi: 10.1093/nar/gkz091

Human H4 tail stimulates HIV-1 integration through binding to the carboxy-terminal domain of integrase

Eric Mauro[1,2,3], Paul Lesbats[1,2,3], Delphine Lapaillerie[1,2,3], Stephane Chaignepain[3,4], Benoit Maillot[5], Oyindamola Oladosu[5], Xavier Robert[3,6], Francesca Fiorini[3,6], Bruno Kieffer[5], Serge Bouaziz[3,7], Patrice Gouet[3,6], Marc Ruff[3,5] and Vincent Parissi[1,2,3,*]

[1]Fundamental Microbiology and Pathogenicity Laboratory, UMR 5234 CNRS-University of Bordeaux, SFR TransBioMed. Bordeaux, France, [2]International Associated Laboratory (LIA) of Microbiology and Immunology, CNRS/University de Bordeaux/Heinrich Pette Institute-Leibniz Institute for Experimental Virology, France, [3]Viral DNA Integration and Chromatin Dynamics Network (DyNAVir), France, [4]Université de Bordeaux, UMR CNRS 5248 CBMN (Chimie Biologie des Membranes et Nanoobjets), 33076 Bordeaux, France, [5]IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire), Département de Biologie Structurale Intégrative, UDS, U596 INSERM, UMR7104 CNRS, Strasbourg, France, [6]MMSB-Institute of the Biology and Chemistry of Proteins, UMR 5086 CNRS-Lyon 1 University, Lyon, France and [7]CiTCoM, CNRS, UMR 8038, Université Paris Descartes, Sorbonne Paris Cité, Paris, France

Received September 28, 2018; Revised January 28, 2019; Editorial Decision January 30, 2019; Accepted February 04, 2019

ABSTRACT

The integration of the retroviral genome into the chromatin of the infected cell is catalysed by the integrase (IN)•viral DNA complex (intasome). This process requires functional association between the integration complex and the nucleosomes. Direct intasome/histone contacts have been reported to modulate the interaction between the integration complex and the target DNA (tDNA). Both prototype foamy virus (PFV) and HIV-1 integrases can directly bind histone amino-terminal tails. We have further investigated this final association by studying the effect of isolated histone tails on HIV-1 integration. We show here that the binding of HIV-1 IN to a peptide derived from the H4 tail strongly stimulates integration catalysis *in vitro*. This stimulation was not observed with peptide tails from other variants or with alpha-retroviral (RAV) and spuma-retroviral PFV integrases. Biochemical analyses show that the peptide tail induces both an increase in the IN oligomerization state and affinity for the target DNA, which are associated with substantial structural rearrangements in the IN carboxy-terminal domain (CTD) observed by NMR. Our data indicate that the H4 peptide tail promotes the formation of active strand transfer complexes (STCs) and support an activation step of the incoming intasome at the contact of the histone tail.

INTRODUCTION

Retroviruses are single stranded RNA plus-enveloped virus characterized by a reverse transcription step of the viral RNA genome into DNA and its stable insertion into the host chromosomes. After entry of the viruses and reverse transcription the generated viral DNA (vDNA) assembles with the viral integrase (IN), diverse additional viral and cellular host factors to form the pre-integration complex (PIC). After entering the nucleus, the synaptic complex formed between IN and vDNA ends, called the intasome, will bind the host target DNA (tDNA) into the host chromatin and catalyse integration (1). Integration site selectivity in the host genome is a complex process depending on the retrovirus and multiple parameters such as the nuclear entry pathway, the nuclear architecture, targeting cellular cofactors such as CPSF6 and LEDGF/P75, and the chromatin structure surrounding the insertion site (2). Indeed, HIV-1 integration occurs in the outer shell of the nucleus in close correspondence with the nuclear pore and preferentially into transcriptionally active regions of the chromatin (3,4). HIV-1 integration distribution correlates positively with histone modifications that are enriched in actively transcribed genes and negatively with repressive epigenetic marks. While IN is sufficient to promote integration *in vitro* cellular cofactors direct integration in preferential regions of the host genome in infected cells. LEDGF/p75, a chromatin-tethering factor, binds directly IN and targets HIV-1 integration into highly spliced and intron rich genes by specifically recognizing the H3K36me3 mark enriched in these regions (5–7). It has also been shown that CPSF6 is determinant for HIV-1 integration site selection by bind-

*To whom correspondence should be addressed. Tel: +33557571740; Fax: +3357571766; Email: vincent.parissi@u-bordeaux.fr

ing to the viral capsid protein (CA) and directing the PIC toward transcriptionally active and open chromatin region of the euchromatin (8).

After reaching the suitable region of the chromatin, incoming viral intasomes bind host nucleosomes. This final association between the intasome and the nucleosome is also a critical step in the tDNA capture and is governed by the need for the integration complex to bend the tDNA (9–12). The functional intasome/nucleosome association was found to be modulated by the structure of both the incoming intasome and the chromatin surrounding the targeted nucleosome (13). Furthermore, cellular chromatin remodelling activity at the integration site was shown to regulate access to the targeted nucleosome (14). Recently, the FACT histone chaperone complex has been identified as a modulator of both ALV and HIV-1 integration (15,16). Its nucleosome dissociation activity was shown to promote HIV-1 integration in initially refractory compact chromatin (16). Taken together, these data suggest that specific intasome/nucleosome contacts are required for optimal integration and may be regulated by chromatin compaction and remodelling. This hypothesis is further supported by the demonstration of direct IN/histone interactions in the solved cryoEM structure of the PFV intasome/nucleosome (9) and the recent finding of the direct binding of HIV-1 IN to histone tails, especially histone 4 (H4), promoting nucleosomal integration (17). In both cases the interactions between INs and histones occur *via* the carboxy-terminal domains (CTD) of the retroviral enzymes. Moreover, mutations in the CTD of the INs that impair their binding to histones also impair their functional association with the nucleosomes as well as cellular integration efficiency and selectivity providing biological evidence for the relevance of the IN/histone interaction (9,17).

We have, thus, investigated here the molecular mechanism involved in the regulation of HIV-1 integration by histone tails. Our work shows that HIV-1 integration is strongly stimulated by peptides derived from the H4 tail *in vitro*. This stimulation was correlated with the stimulation of both IN oligomerization and its binding to tDNA. This stimulation was found to be highly specific for the H4 tail and HIV-1 since the other histone tails had no effect on integration and the other retroviral INs were not found to be stimulated, which was consistent with the recently reported IN/H4 interaction. Based on these data, we propose that an intasome stimulation step occurs at the contact of the histone tails with the nucleosomal substrate.

MATERIALS AND METHODS

Proteins, peptides and antibodies

Wild-type (WT) and mutated full-length HIV-1 INs were purified as previously reported (17). PFV and maedi-visna virus (MVV) INs were purified as previously described (respectively (18,19). The Rous-associated virus 1(RAV-1) IN expression vector provides a cleavable N-terminal hexahistidine tag (His tag). The protein was expressed in *Escherichia coli* BL21 (DE3) Rosetta cells (Novagen) grown in LB medium supplemented with 1% of glucose and induced 3h at 37°C with 1mM IPTG. The cells were lysed in buffer A (50 mM Tris–HCl pH 7.5, 1 M NaCl, 20 mM Imidazole, 2 mM 2-mercaptoethanol and 10% (w/v) glycerol) supplemented with 100 μg/ml of egg white lysozyme (Sigma-Aldrich) and with 1× protease inhibitor cocktail EDTA-Free (Sigma-Aldrich). Soluble lysate was applied to a prepacked nickel column (HisTrap HP, GE Healthcare) and fractioned on an Äkta Purifier (GE Healthcare) using a linear gradient from buffer A to B (buffer A added of 0.5 M imidazole and containing only 0.5 M NaCl) over 20 column volumes. His tag was cleaved by adding TEV protease at a 1:50 molar ratio and incubating the mixture for 3 h at 34°C. The digestion mixture was loaded onto nickel column to eliminate histidine-tagged TEV protease and undigested IN. IN protein was further purified on a heparin column by using buffer composed of 50 mM Tris–HCl pH 7.5, 0.5 M NaCl, 2 mM 2-mercaptoethanol and 10%(w/v) glycerol and eluted with the same buffer containing 2 M NaCl. Rav-1 IN was subsequently dialyzed against the storing buffer (50 mM HEPES pH 7.5, 1 M NaCl, 5 mM 2-mercaptoethanol and 5% (w/v) glycerol) and aliquots were snap freezed by using liquid nitrogen. LEDGF/p75 was purified following the previously reported protocol (17). Polyclonal anti-HIV-1 IN antibody was purchased from Bioproducts MD (Middletown, MD, USA). Monoclonal anti-HIS was purchased from Abcam. HRP-conjugated secondary anti-rabbit antibody was purchased from Sigma and anti-mouse from Jackson ImmunoResearch. The 5′-biotinylated 601 sequence was purchased from TEBU-Bio. Biotinylated histone tail peptides were purchased from Eurogentech (Angers, France). The sequences of the peptides used in this work are reported in Supplementary Figure SI1.

In vitro integration assay with recombinant integrases

Recombinant purified IN (200 nM final concentration) was incubated with radiolabelled vDNA containing U5 viral ends (10 nM) and 50 ng of pBSK-derived p481 acceptor plasmid DNA. The reaction was initiated by the addition of a reaction buffer (final concentration: 17% DMSO; 6 mM DTT; 11 mM $MgCl_2$; 22 μM $ZnCl_2$; 8% PEG; 100 mM NaCl) for 2 hours at 37°C. After the reaction, the resultant integration products were deproteinized by proteinase K treatment (Promega) and phenol/chloroform/isoamyl alcohol (25/24/1 v/v/v) extraction before loading onto a 1% agarose gel. Nucleic acids were fixed with 5% TCA solution, and the gel was then dried and imaged by autoradiography with a PharosFX. The bands corresponding to free substrate (S), donor/donor, linear two-end full site integration products (FSI) and circular one-end half site integration products (HSI) + FSI products were quantified using Quantity One software. The circular FSI products were specifically quantified by cloning them into bacteria and determining the numbers of ampicillin-, kanamycin- and tetracycline-resistant clones as percentages of the integration reaction control, which was performed using the WT enzyme.

In vitro integration assay with intasomes

For MVV assays, 10 nM of the purified intasome was incubated with 200 ng of naked p481 plasmid tDNA and 7.5 mM $MgSO_4$ in a final volume of 40 μl. Similar amounts of

intasome and tDNA were used in PFV assays but with 125 mM NaCl, 5 mM $MgCl_2$, 10 mM DTT, 4 μM $ZnCl_2$ and 25 mM BTP pH 7.5 (optimal conditions set up in (19). Reactions were run for 5 min at 37°C and stopped with 5.5 μl of a mix containing 5% SDS, 0.25 mM EDTA pH 8.8 and 2 μl of proteinase K (Promega) for 1 h at 37°C. Nucleic acids were then precipitated with 150 μl of ethanol overnight at −20°C. Prior to electrophoresis in a 0.8% agarose gel, samples were spun at top speed for 45 min at 4°C, and the pellets were dried and then resuspended with DNA loading buffer.

3′ Processing assay

Processing assays were performed using the same conditions used for the integration assay, except that the concentration of IN was 400 nM and a short 21 bp DNA was used as the substrate (20 nM). The reaction was stopped by the addition of 90 μl of a buffer containing 750 mM urea, 4.5 mM EDTA, 0.75× TE, 300 mM AcNa and 0.45 mg/l glycogen. The products were then extracted by phenol/chloroform and precipitated with ethanol overnight before loading on a 12% polyacrylamide denaturing gel. An image of the gel was obtained by autoradiography with a PharosFX, and the bands corresponding to the processed 19 bp DNA were quantified using Quantity One software.

Intasome assembly

MVV and PFV intasomes were assembled and purified as previously described (18,19).

Solubility assay

The assay was performed as previously described (20). Recombinant IN (3 μM) was incubated with either LEDGF/P75 or the H4 histone tail for 30 min at room temperature in a buffer containing 20 mM HEPES pH 7.5, 10% glycerol, 5 mM DTT, 1 mM EDTA and 100 mM NaCl. After centrifugation at 15 000 g for 15 min, the soluble protein in the supernatant and the insoluble protein in the pellet were separated and analysed by SDS-PAGE. The amount of soluble IN was then quantified using ImageJ software.

DNA binding assay

The effect of H4 on the binding of IN with either 601 (tDNA) or HU5 (vDNA) biotinylated DNA was evaluated using the pull down technique. Briefly, 10 pmol of IN was incubated with the H4 histone tail for 20 min on ice. The DNA was then added (1 pmol), and interaction was allowed for 20 min on ice in a final volume of 10 μl before the addition of a 12.5 μl aliquot of DynabeadsMyOne Streptavidin T1 (Invitrogen, ref. 65601) in a total volume of 300 μl of interaction buffer (50 mM HEPES pH 7.5, 1 μg/ml BSA, 1 mM DTT, 0.1% Tween 20, 10% glycerol and 200 mM NaCl) and incubated at 4°C for 1 h under rotation. The beads were washed three times with 300 μl of interaction buffer, and the products were resuspended in Laemmli buffer, then separated by SDS-PAGE. Interacting proteins were detected by western blot analysis using anti-HIV-1 IN antibody and quantified using ImageJ.

BS3 crosslinking assay

IN was first diluted to 5 μM in a buffer containing 1 M NaCl, 20 mM HEPES pH 7, 400 μM $ZnCl_2$ and 10 mM DTT to render the IN soluble prior to crosslinking. Then, 1 μl of 5 μM IN was mixed with the peptides and incubated for 30 min on ice. The samples were then crosslinked with BS3 (final concentration 20 mg/l) in a final volume of 10 μl for 30 min at 25°C and analysed on 6% SDS-PAGE in order to detect high-molecular-weight complexes by western blotting with an antibody directed against the appropriate IN (anti-IN for HIV-1 or anti-HIS for PFV). The oligomerization was then quantified by measuring the amount of complexes larger than monomers using imageJ.

NMR analysis

A sample containing the uniformly $^{15}N/^{13}C$-labelled carboxy-terminal domain (CTD) of the integrase in 25 mM HEPES buffer, 150 mM NaCl, 5 mM DTT and 10% D2O was introduced into a 3 mm NMR tube at the concentration of 280 μM in a final volume of 200 μl. In order to achieve sequential assignment the following experiments were performed: $^{1}H–^{15}N$ HSQC (21), $^{1}H–^{13}C$ HSQC (21), HNCA (22,23), HNCO (22,23), HNCACB (22,23) and CBCACONH (24,25). All NMR experiments were carried out at 303 K on an Avance III Bruker spectrometer operating at 600.13 MHz in proton mode, equipped with a cryogenic probe. The acquisition and processing of the experiments were performed using the Bruker software Topspin (version 3.5 Bruker, Karlsruhe, Germany). The spectral analysis and assignment were performed with CCPNMR 2.4.2 (26). For the interaction experiments, 10 μl of a 48 mM H4K20me1 peptide solution was added to 200 μl of solution containing the CTD at the concentration of 280 μM resulting in a final peptide concentration of 2.28 mM and a peptide/protein ratio of 7.85. The same experiments were conducted using 1 μl of the same peptide solution added to 200 μl of solution containing the CTD at the concentration of 150 μM, resulting in a final peptide concentration of 238 μM and a peptide/protein ratio of 1.58. 2D ^{15}N SOFAST-HMQC (27) spectra were acquired to ensure that the protein was well folded and uniformly ^{15}N labeled. On the later sample, 2D ^{15}N SOFAST-HMQC experiments were performed at different time points after adding the peptide (1, 2, 3, 6 and 8 days) to monitor the interaction of the CTD with the peptide over time.

RESULTS

In vitro HIV-1 integration is specifically stimulated by a peptide mimicking the histone tail H4

A previous study showed that the presence of histone tails, especially histone H4 tail, strongly improves HIV-1 nucleosomal integration efficiency and IN/nucleosome association (17). To obtain additional clues on the molecular mechanism of histone tail involvement in the catalysis of integration, we investigated the effect of histone tail-derived peptides in *in vitro* integration reaction. A peptide corresponding to the 1–23 region of the histone amino-terminal tail H4 mono-methylated on lysine 20 (H4K20me1), previously

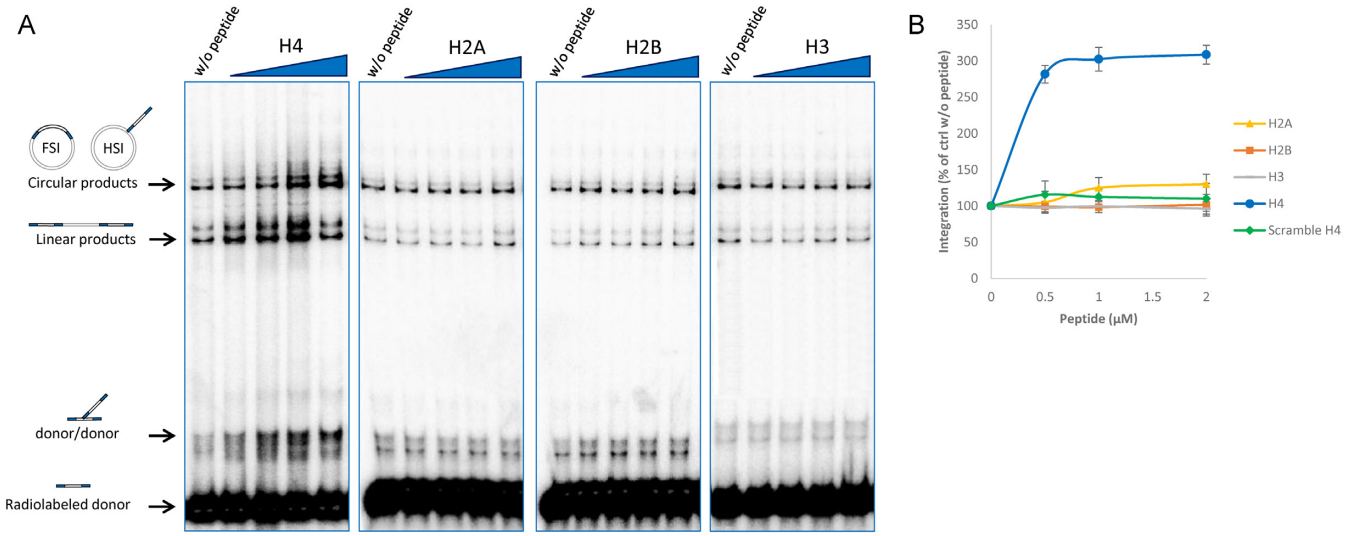

Figure 1. (**A**) Concerted integration assays performed on naked plasmid acceptor DNA, P^{32} radiolabeled 247 bp donor DNA, HIV-1 IN and increasing concentrations of peptides derived from histone tails (H4, H2A, H2B and H3). Integration products are represented on the left (FSI: full site integration; HSI: half site integration). (**B**) Quantification of heterointegration products (circular + linear products) using Quantity One software. All values are shown as the mean ± standard deviation (error bars) of at least three independent sets of experiments.

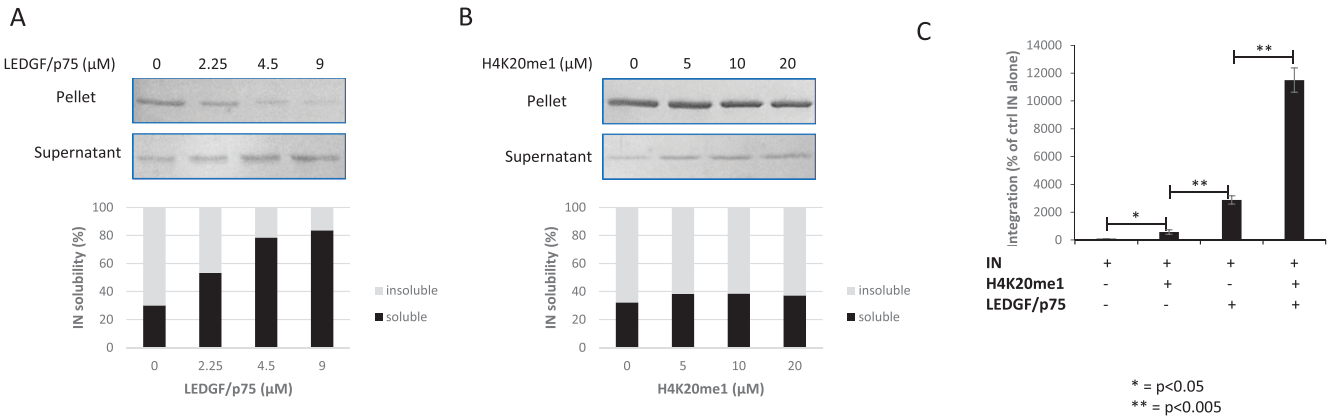

Figure 2. (**A**) Effects of LEDGF/P75 and H4K20me1. (**B**) on the solubility of HIV-1 IN. Briefly, after incubating HIV-1 IN with either LEDGF/P75 or H4 in 100 mM NaCl, proteins were spun, and then the pellets (insoluble) and the supernatant (soluble) were separated and loaded onto SDS-PAGE. Quantification of the soluble (dark) and insoluble (grey) fractions is shown in the lower part. Representative experiments are shown in (A) and (B). (**C**) Quantification of heterointegration from the concerted integration assay performed with HIV-1 IN without PEG and DMSO. The results from three independent experiments are shown, and the values are reported as the mean ± standard deviation (error bars).

identified as the best HIV-1 IN binder among various histone modifications (17), was first added in a typical *in vitro* concerted integration assay performed on naked DNA. As shown in Figure 1, integration was stimulated, ∼3-fold, in the presence of the H4 peptide. Peptides derived from the other H2A, H2B and H3 tails, previously reported not to bind HIV-1 IN (17), did not affect the catalysis, indicating that the integration stimulation was specific to the H4 tail and might be dependent on a direct binding of the peptide to the IN protein. Additionally, a scramble peptide, which had the same size and amino acid composition as the H4 tail-derived peptide but randomly distributed, showed no effect on integration, indicating that the stimulation was highly specific to the native H4 tail sequence (Figure 1B). 'Time of addition' experiments reported in Supplementary Figure SI2 indicated that pre-incubation of the peptide with the DNA before adding IN completely prevented its stimulatory effect leading to a slight inhibition of the integration reaction. This suggests that the peptide can also bind DNA and compete with IN for its binding to its substrate when added in saturating concentrations.

Previous HIV-1 IN stimulation by LEDGF/p75, the main co-factor in HIV integration, has been shown to be, at least partially, due to an increase in the IN solubility (20). Consequently, we next checked whether the H4-induced IN stimulation might be due to a similar effect. As reported in Figure 2A, while LEDGF/p75 clearly increased the IN solubility at a low NaCl concentration (100 mM), the H4 tail did not impact the solubility of the enzyme (Figure 2B). This result suggests that the stimulation of integration induced by the H4 tail *in vitro* is not due to better solubility of the IN and supports a different stimulatory mecha-

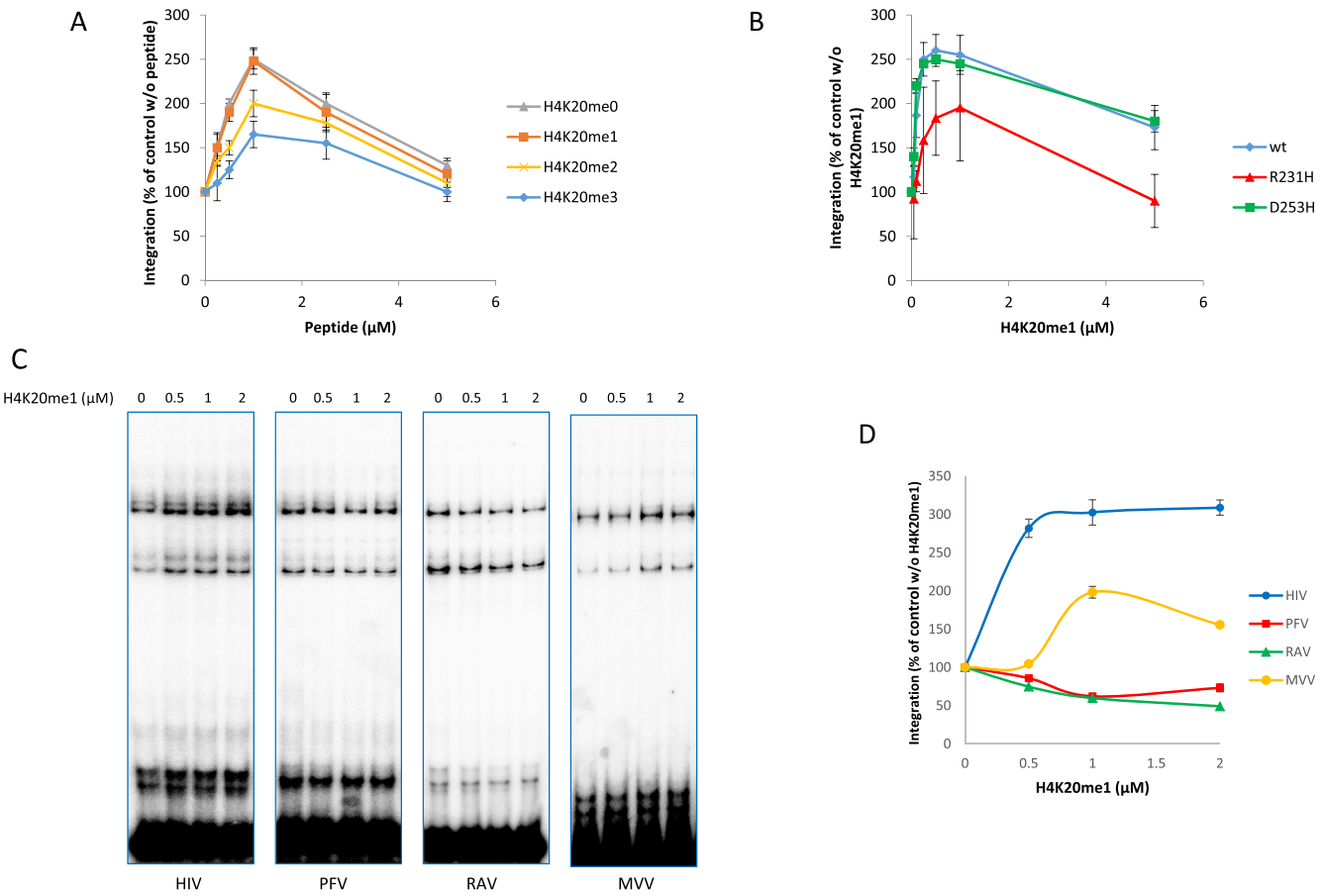

Figure 3. (**A**) Quantification of heterointegration from the concerted integration assay performed with HIV-1 IN in the presence of peptides derived from the H4K20 tail with different methylation states on K20 (non-, mono-, di- or trimethylated). (**B**) Quantification of heterointegration from the concerted integration assay performed with mutants of HIV-1 IN (R231H and D253H, which respectively bind with the H4K20me1 tail to a lesser or a greater extent than the WT IN) in the presence of the H4K20me1 peptide. (**C**) Concerted integration assays performed with either HIV-1, PFV, RAV or MVV IN in the presence of increasing concentrations of the H4K20me1 peptide and their quantification (**D**). All values are shown as the mean ± standard deviation (error bars) of at least three independent sets of experiments.

nism for LEDGF/p75 and H4. This question was further addressed by testing whether the integration stimulation by the H4 peptide could be modulated by the addition of LEDGF/p75. For this purpose, we performed *in vitro* concerted integration using non-optimized conditions allowing the detection of LEDGF/p75-mediated stimulation (without PEG and DMSO). As seen in Figure 2C and Supplementary Figure SI3, when LEDGF/p75 and H4 were added at the same time, the stimulation of the integration was higher than when either was added alone. This further supports the idea that LEDGF/p75 and H4 integration stimulation occurs by different and not mutually exclusive mechanisms.

In vitro HIV-1 integration stimulation induced by the H4 tail requires IN/H4 tail interaction

The interaction between HIV-1 IN and the H4 tail was shown to be modulated by the methylation state of lysine K20 from H4 (17). Thus, we next tested whether K20 modification could also affect the integration stimulation of the peptide. As shown in Figure 3A, we found that K20 methylation also induced changes in the peptide effect on integration, as dimethylation and trimethylation of H4K20 decreased IN stimulation, which parallels the previously reported IN/peptide interaction. HIV-1 IN R231H, previously shown to be deficient for optimal binding with H4, but not for DNA binding nor catalytic activity, (17), was found to be less sensitive to stimulation (Figure 3B). In contrast, the D253H mutant, which was found to bind the H4 tail better than the WT without being affected for the DNA binding nor catalytic activity, showed elevated insertion efficiency similar to that of the WT IN (Figure 3B). We next assessed the specificity of the stimulatory effect on different retroviral INs. Alpha-retroviral RAV and spuma-retroviral PFV INs were thus tested. As shown in Figure 3C and D, neither RAV nor PFV integration was stimulated by the peptide. Moreover, MVV lentiviral IN was found also stimulated by the H4 peptide. However, comparison of the effect of H4 on HIV-1 and MVV INs showed that both the optimal peptide concentration and the range of stimulation observed for MVV are lower than what observed for HIV-1 suggesting a different affinity of MVV IN for H4 tail. To confirm this effect we tested the other histones and H4

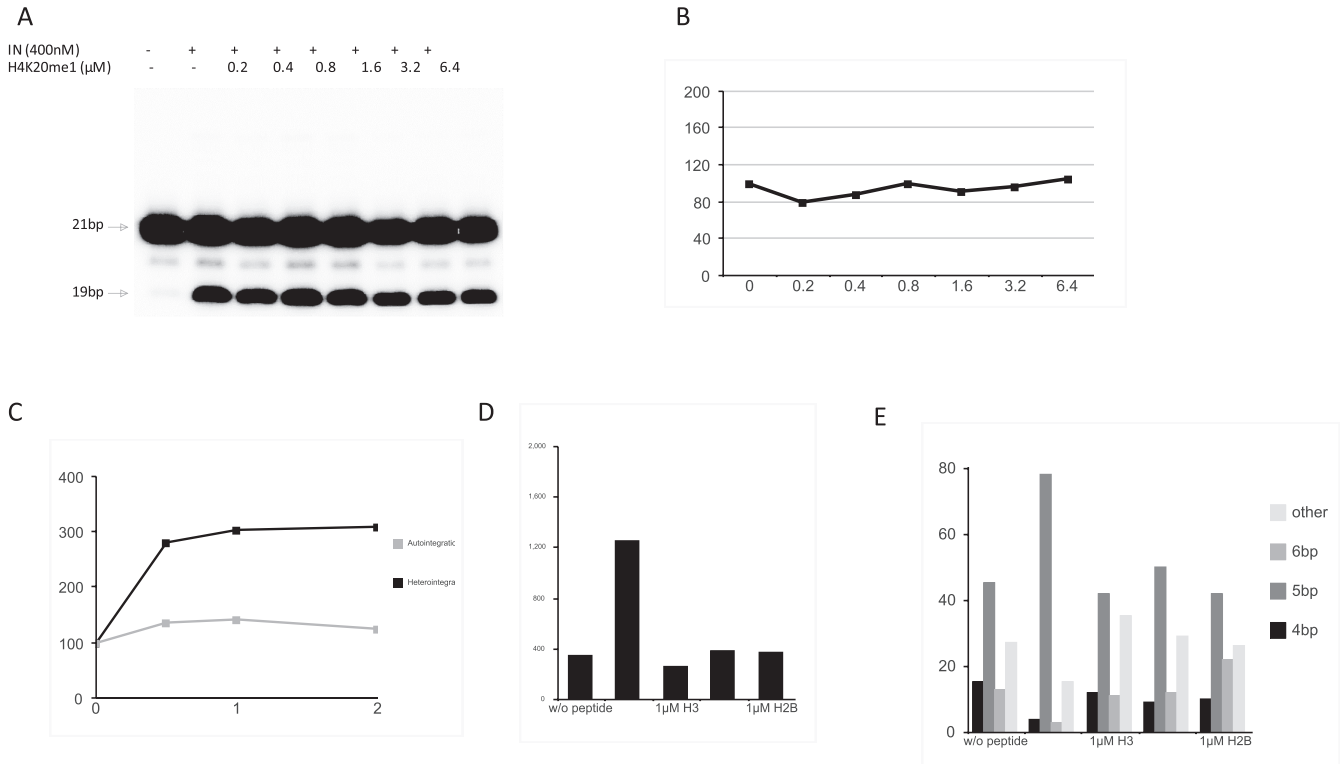

Figure 4. (**A**) 3′ processing assay using 21 bp radiolabelled DNA fragment, HIV-1 IN and increasing concentrations of the H4K20me1 peptide. The 19 bp processed product is quantified in (**B**). (**C**) Quantification of different integration products (heterointegration and autointegration) from a concerted integration assay performed with HIV-1 IN and increasing concentrations of the H4K20me1 peptide. (**D**) Selection of replicative integration FSI forms in bacteria obtained in concerted integration assays performed in the presence or absence of histone peptides. (**E**) Structure of base pair duplication in the sequenced integrated products. All values are shown as the mean ± standard deviation (error bars) of at least three independent sets of experiments.

peptides derivatives on MVV integration. Data reported in Supplementary Figure SI4 confirmed the reproducible stimulation of MVV by H4 tail as well as the preference of this enzyme for H4K20me1 and H4K20me0 over H4K20me2 and H4K20me3. These results indicate that MVV is also stimulated by the H4 tail and that it shares similar preference profile with HIV-1 IN but to a lesser extent probably due to differential affinity.

The H4 tail increases the amount of strand transfer complexes (STCs)

To better characterize the H4 peptide-induced stimulation of HIV-1 integration, we studied its effect on each step of the reaction, i.e. 3′ processing and strand transfer. For this purpose, we first added H4 peptide in a typical 3′ processing assay, as reported in Figure 4. Under standard conditions, no effect of the peptide on this step was detected (Figure 4A and quantification in B). We next further analysed the strand transfer products obtained in concerted integration reactions. Specific quantification of the self- and heterointegration products obtained in our concerted integration assays performed in the presence of the H4 peptide showed that heterointegration was more sensitive to H4 stimulation than donor/donor self-integration (Figure 4C). To specifically analyze the effect of the H4 tail on the FSI events, circular products were cloned and sequenced. The results showed an increased proportion of FSI integration products containing the correct 5 bp duplications (Figure 4D and E).

These data suggest that the H4 peptide exerts its stimulatory effect preferentially on the tDNA capture or strand transfer steps. To confirm this finding, we next tested the peptide on fully assembled retroviral intasomes. Lentiviral MVV intasomes were tested in concerted integration assay since their assembly was previously fully described (18,19). As shown in Figure 5, integration catalysed by the purified MVV intasome was also stimulated by the H4 peptide. To confirm the specificity of this stimulation, PFV intasomes were also tested, and no effect of the H4 tail was detected.

Taken together, these results suggest an effect of the peptide on the structured intasome complexes and a capability of the histone H4 tail to channel the integration towards concerted two-end integration. This increase in the catalytic activity can be attributed to either stabilization or promotion of the formation of active STCs via H4 tail binding to IN. Accordingly, the effect of the H4 tail on the IN-tDNA association was further investigated.

H4 tail increases the formation of IN-DNA complexes and IN oligomers

We first tested the effect of the H4 peptide on IN/DNA binding by pull down experiments performed with HIV-1 recombinant IN and a biotinylated 601 DNA fragment, mimicking the tDNA. As shown in Figure 6A and B, the

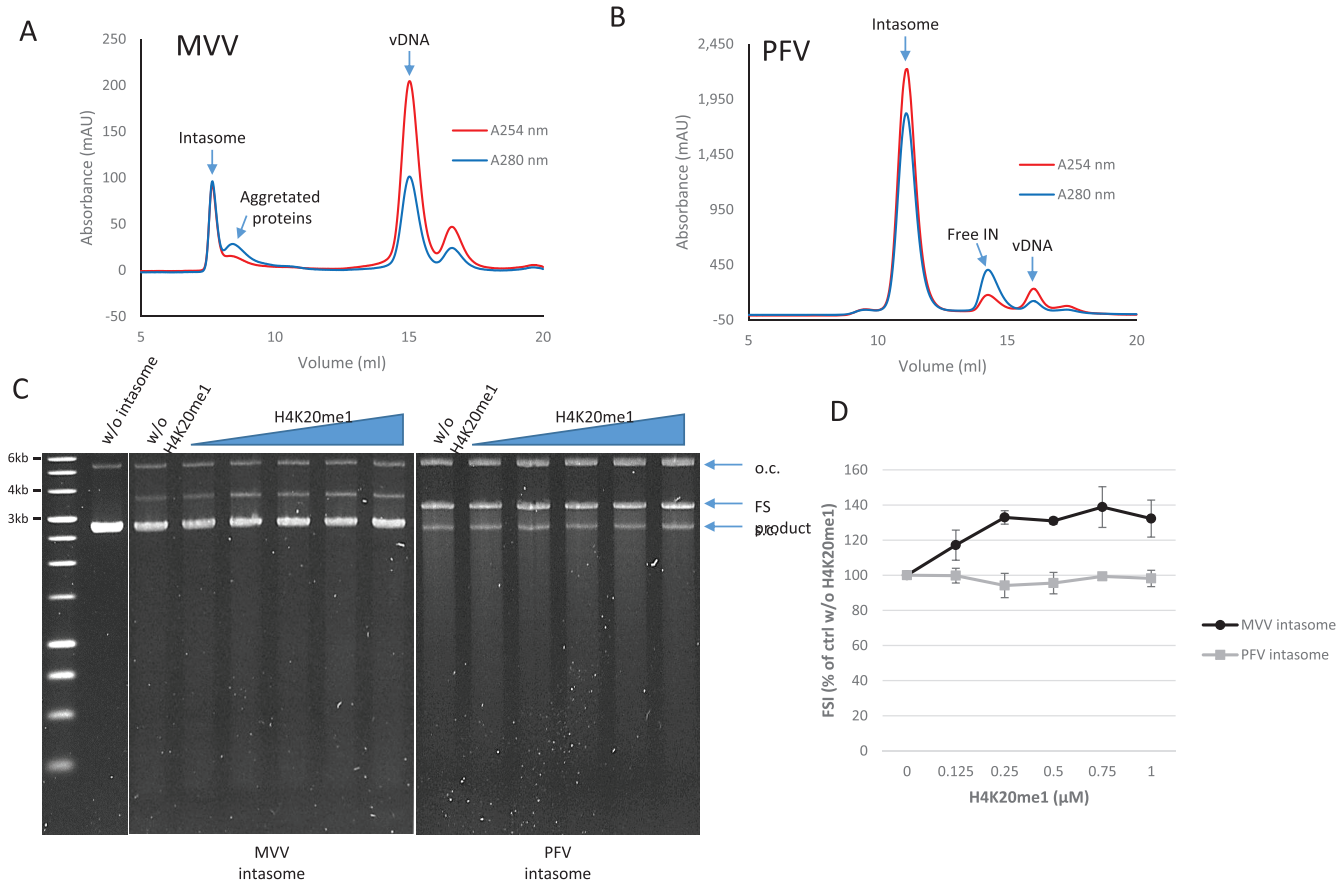

Figure 5. (**A**) Integration assays performed on naked plasmid acceptor DNA using purified intasomes (MVV and PFV) in the presence of increasing concentrations of H4K20me1 peptide. The legends of the different bands are shown on the right (o.c.: open circular; s.c.: supercoiled; FS product: full site product). (**B**) Quantification of the FSI (full site integration) band using Quantity One software. The results from three independent experiments are shown, and the values are reported as the mean ± standard deviation (error bars).

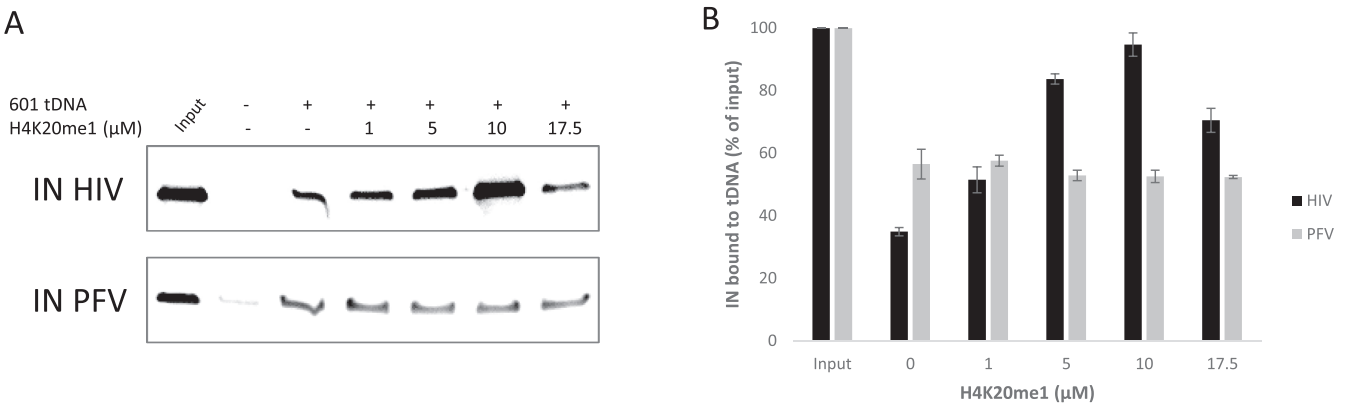

Figure 6. (**A**) Pull down experiment performed with 147 bp biotinylated 601 DNA fragment and HIV-1 or PFV IN in the presence of increasing concentrations of the H4K20me1 peptide at 200 Mm NaCl. The fractions interacting with the DNA were loaded on SDS-PAGE and analysed by western blotting. (**B**) Quantification of the amount of IN interacting with the DNA using Quantity One software. The results from two representative sets of independent experiments are shown, and the values are reported as the mean ± standard deviation (error bars).

addition of the peptide induced an increase in the quantity of IN retained on the DNA without changing the non-specific IN/beads binding (data not shown). An optimum effect was detected at 1 μM of peptide which parallels well with the optimum observed in the integration stimulation reported in Figure 3A. This was assumed to be due to the DNA binding property of the peptide as suggested by the 'time of addition' experiments performed in Supplementary Figure SI2. The increase in IN/DNA binding was not observed with the control PFV IN, suggesting that the stimulation of DNA binding was dependent on the interaction between IN and the peptide.

Since the binding to DNA is dependent on IN oligomerization, we assessed the effect of H4 on the formation of IN multimeric forms. For this purpose, we performed HIV-1 IN BS3 crosslinking experiments in the presence of the H4 peptide. As reported in Figure 7A, the H4 tail induced an increase in the amount of dimers, tetramers and higher-order oligomeric forms present in HIV-1 IN solution in contrast to PFV IN (Figure 7B and comparative quantifications in 7C). The other histone tails showed no effect under these conditions (Figure 7D). This result confirmed the specific effect of H4 on HIV-1 IN and the correlation between oligomerization and the binding of the histone tail to the viral enzyme.

H4 tail induces strong structural changes in the CTD of HIV-1 IN

Previous work using HIV-1 IN truncations unambiguously mapped the interaction within the CTD domain of the protein (17). Furthermore, the modulation of the H4 peptide mediated stimulation by the R231H and D253H CTD IN/H4 mutants suggests that the association of the H4 tail with IN may induce structural rearrangements *via* its binding to the CTD. To characterize the interaction between these two partners and determine the effect of the peptide on the structure of IN, we undertook an NMR structural study of the interaction between the CTD of integrase, IN(I220-D270), and the shorter version of the monomethylated peptide RHRK(me1)VLR which is also capable of stimulating *in vitro* integration (see Supplementary Figure SI5).

Analyses were first performed with a long version of the CTD domain of integrase IN (I220-D288) in 25 mM HEPES buffer, 1 M or 150 mM NaCl, and 5 mM DTT at pH 7 or pH 8. The data indicated that this domain is unstable because of the presence of the completely disordered end Y271-D288. Thus, a shorter version of the CTD domain, IN (I220-D270), was used. This domain lacking the Y271-D288 end appeared to be more stable under NMR conditions, and various NMR experiments could be carried out on this free domain alone or in the presence of the H4K20me1 peptide. The SOFAST-HMQC (27) spectra show that the protein is fully structured with an extensive chemical shift dispersion of ^{1}H and ^{15}N resonance peaks between 6.7 and 10.0 ppm and between 104 and 130 ppm, respectively (Figure 8A), validating our approach.

The assignment of the CTD was performed and allowed to identify 38 amino acids among the 51 of the CTD (Figure 8A). Prolines P233, P238 and P261 were not visible on the ^{15}N HSQC spectra because they lack the amide proton. The remaining missing residues Q221, N222, S230, R231, D232, L242, E246, G247, N254 and S255 could not be assigned unambiguously (Figure 8A). The mapping of these missing residues on the 3D structure of the CTD published previously by other groups (28,29) shows that they are located in very flexible loops with the exception of the residue L242 located in the β2 strand. These amino acids are exposed to the solvent in the three dimensional structure, they are in fast exchange with the solvent and do not appear on the spectrum. Other residues, whose corresponding peaks show an intense signal on the spectrum could not be assigned unambiguously (Figure 8A). The signals observed on the SOFAST-HMQC experiment have a high intensity and we do not observe a broadening of the resonance lines except for residues I220, D229, A248 and R263, whose wide peaks show a fast exchange with the solvent due to a greater exposure. It is interesting to note that these residues are not only located in loops as D229 or R263 but also in β3 strand (A248) (see Supplementary Figure SI6). In the presence of the peptide H4K20me, several effects are observed on the NMR spectra of the CTD (I220-D270). We note a large decrease in the intensity of the peaks over time affecting almost all residues (Figure 8B; Supplementary Figure SI7), a variation of chemical shifts of certain resonances, the disappearance or appearance of peaks and peaks with a large linewidth (Figure 8B). These later peaks could move on the spectrum either because of their interaction with the peptide or because of a change in the structure of the CTD provoked by the peptide, in particular the oligomerization of the CTD. The disappearance of peaks is due to the increase of the size of the oligomeric form of the CTD and to an intermediate exchange between different oligomerization states of the CTD (Figure 8B).

These data confirm that the peptide H4K20me1 interacts with the CTD domain IN (I220-D270). A possible explanation for the decrease in signal intensity would be the formation of oligomers. Indeed, it can be expected that these higher molecular weight complexes will cause a decrease in signal intensity due to faster relaxation rates. The peptide H4K20me1 facilitates the oligomerization or the aggregation of the CTD as shown by the appearance of broad peaks at the center of the spectrum (between 8 and 8.5 ppm in Figure 8C) and the decrease of the peak intensity over time (Supplementary Figure SI7). This hypothetical self-association parallels the oligomerization effect of the peptide observed on the full-length enzyme.

DISCUSSION

In recent years, the literature has reported direct interactions between incoming retroviral intasomes and histone components of the nucleosome that participate in integration (9,16,17). In particular, histone tails appeared to be candidates for interactions with the integration complexes and were found to modulate viral DNA insertion. By investigating the special case of HIV-1 IN and the H4 amino-terminal tail, we show here that the binding of the H4 tail to IN promotes the integration process. In particular, H4-derived peptides were shown to specifically stimulate the FSI of both long terminal repeats (LTRs) into the tDNA

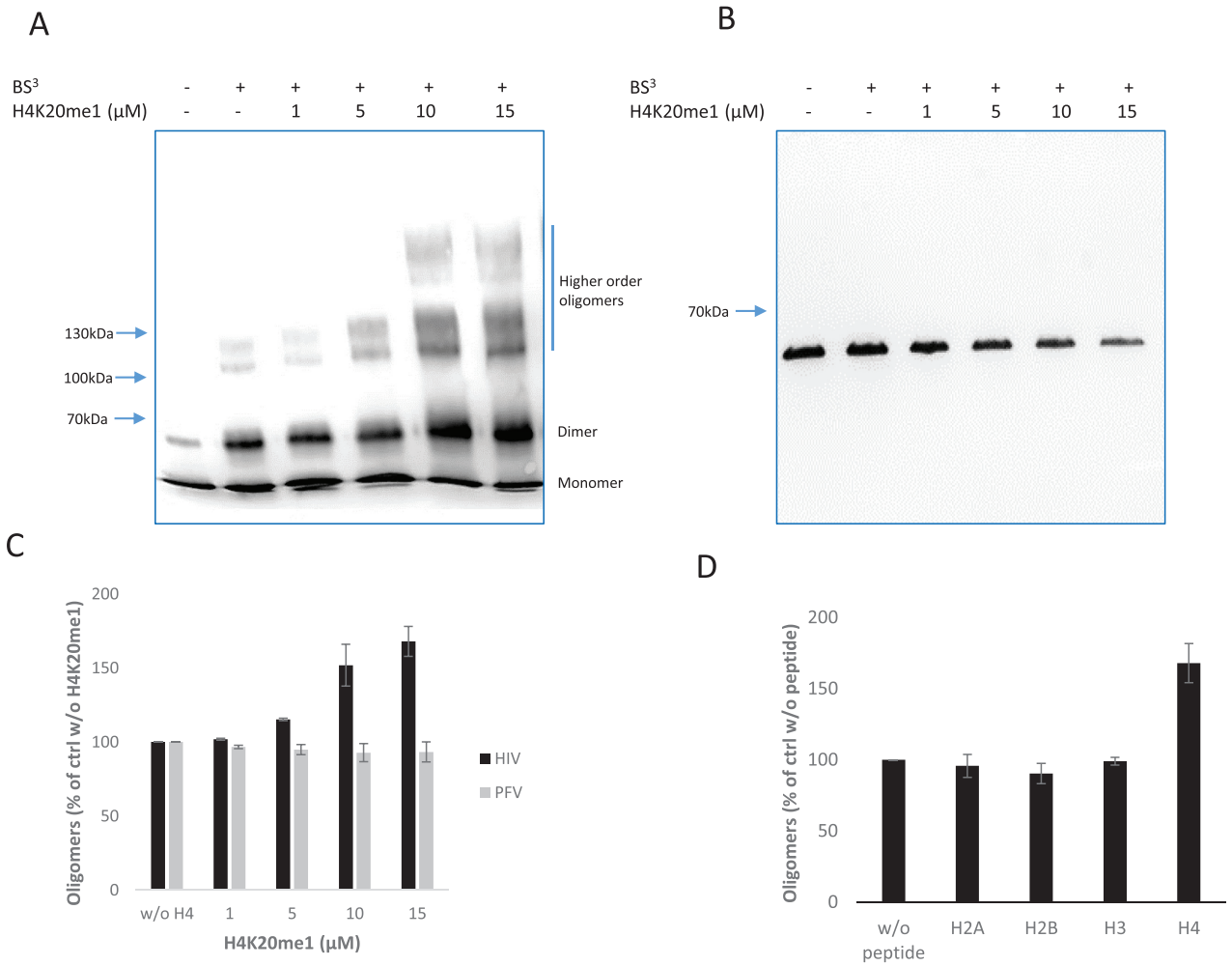

Figure 7. (**A**) Chemical crosslinking of HIV-1 IN or (**B**) PFV IN in the presence of increasing concentrations of H4K20me1 peptide. Samples were loaded on SDS-PAGE and analysed by western blotting. The different species are shown on the right side of the HIV-1 gel. (**C**) Quantification of the amounts of oligomers (dimers and higher-order species) using Quantity One software. (**D**) Quantification of HIV-1 IN oligomers in the presence of different peptides derived from histone tails. The results from three independent experiments are shown, and the values are reported as the mean ± standard deviation (error bars).

with only a slight effect on the processing activity. Importantly, this stimulation was also shown to increase the integration fidelity, leading to a strong stimulation of the proportion of correct 5 bp duplication insertion events. This effect was not due to an unspecific increase in the enzyme solubility and closely paralleled with the stimulation of IN binding to the tDNA by the H4 peptide, while the association with the viral DNA substrate was not changed. This stimulation was found to be dependent on the direct peptide binding to IN since IN/H4 interaction mutants or other retroviral INs unable to bind H4 are not sensitive. A biochemical analysis of the effect of the H4 peptide on IN showed that it also increased the oligomerization of the enzyme both in the presence and the absence of DNA. Taken together, our data support a specific effect of the peptide on the association between the intasome and the tDNA by improving the formation of the active STC, probably by increasing the stability of the active IN oligomers on the target substrate or by allowing further oligomerization stages.

Further NMR analysis of the interaction between the H4-derived peptide and the CTD of IN confirmed that the peptide H4K20me1 derived from H4 and the CTD domain IN (I220-D270) interact with each other. Furthermore, our result revealed that the peptide H4K20me1 could induce oligomerization of the CTD as it has been demonstrated with the full length enzyme. The formation of these oligomers in an intermediate exchange is supported by the broadening of the resonance lines observed between 8.0 and 8.5 ppm as well as by the overall decrease in the intensity of the signals observed in the presence of the peptide H4K20me1 over time due to faster relaxation rates (Figure 8B, C and Supplementary Figure SI7). Further investigations should be undertaken to fully understand the role of this phenomenon in stimulating integration. However, since these results are perfectly consistent with the biochemical data, we propose that the oligomerization induced by H4 tail binding to the IN CTD is the main mechanism involved in the integration stimulation observed *in vitro*. Indeed, the

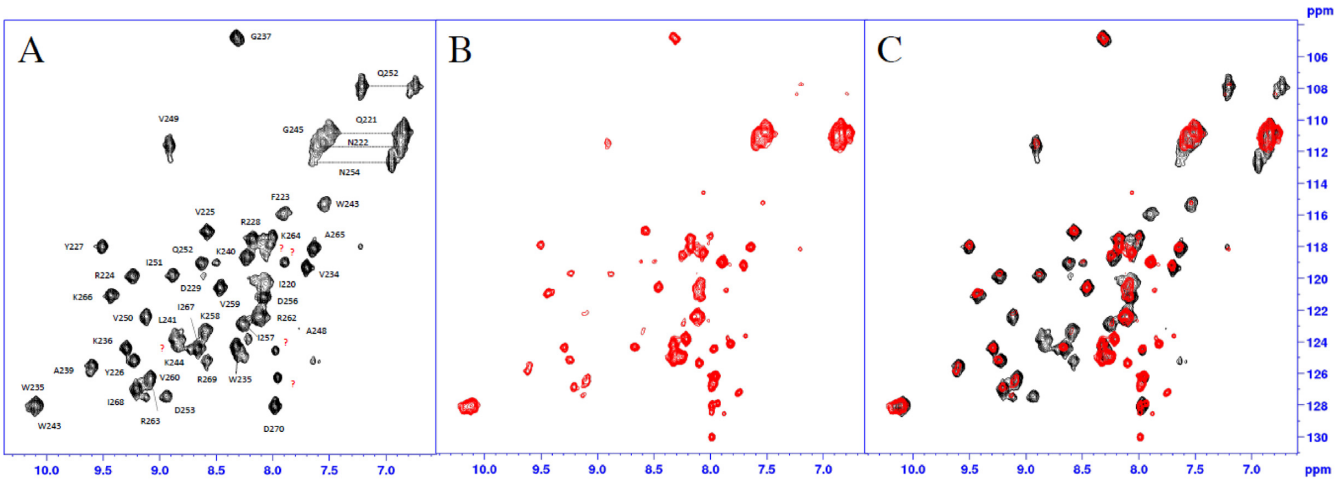

Figure 8. 600 MHz ^{1}H–^{15}N SOFAST-HMQC (16) spectra of the uniformly $^{15}N/^{13}C$-labelled carboxy-terminal domain (CTD) of integrase IN (I220-D270), at the concentration of 150 μM, in 200 μl of a 25 mM HEPES buffer, containing 150 mM NaCl, 5 mM DTT and 10% D2O, at pH 8.0 and 303 K. (**A**) SOFAST-HMQC (16) spectra of the free CTD show that the protein is well structured with a broad spreading of ^{1}H and ^{15}N resonances between 6.7 ppm and 10.0 ppm in ^{1}H NMR and between 104 ppm and 130 ppm in ^{15}N. The signals are all of comparable intensity except the I220, D229, A248 and R263 residues. (**B**) SOFAST-HMQC (16) spectra of the CTD (150 μM) in the presence of the peptide RHRK(me1)VLR (2.38 μM) after 8 days. There is a large decrease in the intensity of the peaks up to the disappearance of some of them. There is a variation of the chemical shifts of other peaks, the disappearance or appearance of peaks and broaden peaks appear between 8 and 8.5 ppm. (**C**) Superimposition of the SOFAST-HMQC experiments of the free IN CTD (black) and in the presence of the peptide H4K20me1 (red) eight days after mixing the two partners, showing the decrease of the peak intensity in the presence of the peptide, the disappearance and the appearance of new peaks and a global widening of resonance peaks.

intermediate exchanges detected by NMR in the presence of the H4-derived peptide highlight the highly dynamic structure of the HIV-1 IN CTD. We can speculate and suggest that this domain may adopt various structures depending on the binding partner. The reporting H4 peptide may also serve as a useful tool to further characterize these structural intermediates.

Taken together, our results suggest that an activation step of the incoming intasome occurs after its docking to the targeted nucleosomal substrate. As summarized in Figure 9, we propose that the final functional association between the intasome and the nucleosome, leading to optimal integration, occurs in steps. The first step consists of anchoring of the intasome to the nucleosome by interaction with histone tails and nucleosomal DNA, as previously proposed (11). The second step corresponds to structural changes in the IN CTD in the intasome, promoting optimal tDNA binding and leading to an activated state of the integration complex that can channel the reaction to FSI. Remodelling activities, previously shown to activate HIV-1 integration into polynucleosomal dense substrates (14,16), would increase the availability of the histone tails for both the docking and activation steps. Since the incoming intasome is expected to be fully assembled, the simulation of IN oligomerization observed *in vitro* may account for an effect of the peptide in the stabilization of the active IN oligomers that could also promote tDNA binding. Another possibility could be that low-order IN oligomers are induced to undergo additional oligomerization at the contact of the nucleosome and the tails, leading to high-order, fully active oligomers. This mechanism would be in agreement with the different IN oligomeric forms recently isolated (30). Interestingly, recent works reported quaternary structures changes of MLV IN at the contact of the chromatin in infected cells (31) suggesting that such structural modifications in the incoming intasomes during their association with chromatin may be a conserved process among retroviruses.

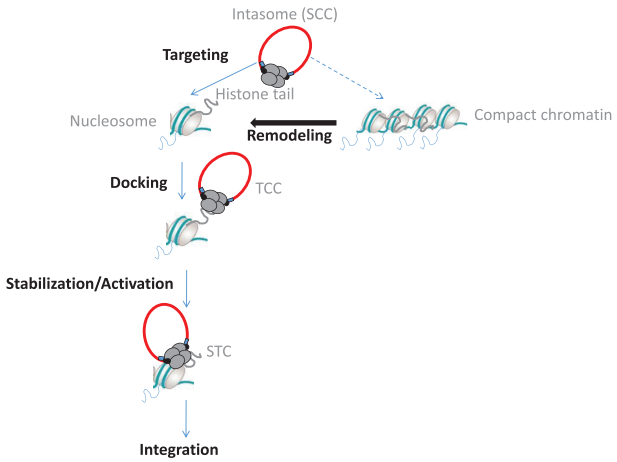

Figure 9. Proposed model of the final association between the incoming intasome and the nucleosome. The intasome is targeted towards uncondensed active chromatin due to the cumulative action of targeting cellular cofactors, nuclear import pathways and nuclear spatial organization. Upon reaching the insertion region, the intasome docks to the nucleosome via LEDGF/p75•H3K36me3. Our data suggest that additional protein/protein contacts between the target capture complex (TCC) and the nucleosome may occur upon IN•H4 tail interaction. This interaction may trigger intasome structural changes that lead to a fully active strand transfer complex (STC). Compact chromatin known to be refractory for HIV-1 integration may obstruct the availability of the histone tails for intasome association, and chromatin remodelling may modulate this accessibility.

The importance of the activation step in the integration process occurring in infected cells remains to be fully inves-

tigated. However, the previously reported impairment of the H4-mediated stimulation of viral replication and integration stage by mutations (17) fully supports a role of this activation process under physiological conditions. The importance of this process in viral replication may indicate that it could be of interest for future antiviral strategies aiming to decrease integration efficiency. Additionally, the strong effect of the H4 peptide in promoting the formation of highly active IN•DNA complexes makes it a good tool for further structure/function studies aimed at elucidating the mechanism leading to the formation of these complexes as well as good candidates for the resolution of their structures. Indeed, the intermediate exchanges detected by NMR in the presence of the H4-derived peptide highlight the highly dynamic structure of the HIV-1 IN CTD, suggesting that this domain may adopt various structures depending on the binding partner. The reporting H4 peptide may also serve as a useful tool to further characterize structural intermediates that may constitute new and interesting therapeutic targets.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The manuscript was edited by NPG Language Gold Editing

FUNDING

French National Research Agency against AIDS (ANRS) [AO 2016-2, ECTZ18624]; SIDACTION [AO-27-1 10465, 16-1-AEQ-10465]; French Infrastructure for Integrated Structural Biology (FRISBI) [ANR-10-INSB-05-01]; Instruct, a part of the European Strategy Forum on Research Infrastructures (ESFRI); Centre National de la Recherche Scientifique (CNRS); the University Victor Segalen Bordeaux 2; ECOS-CONICYT C12B03 program. Funding for open access charge: ANRS [AO 2016-2].
Conflict of interest statement. None declared.

REFERENCES

1. Lesbats,P., Engelman,A.N. and Cherepanov,P. (2016) Retroviral DNA integration. *Chem. Rev.*, **116**, 12730–12757.
2. Kvaratskhelia,M., Sharma,A., Larue,R.C., Serrao,E. and Engelman,A. (2014) Molecular mechanisms of retroviral integration site selection. *Nucleic Acids Res.*, **42**, 10209–10225.
3. Marini,B., Kertesz-Farkas,A., Ali,H., Lucic,B., Lisek,K., Manganaro,L., Pongor,S., Luzzati,R., Recchia,A., Mavilio,F. *et al.* (2015) Nuclear architecture dictates HIV-1 integration site selection. *Nature*, **521**, 227–231.
4. Schroder,A.R., Shinn,P., Chen,H., Berry,C., Ecker,J.R. and Bushman,F. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*, **110**, 521–529.
5. Eidahl,J.O., Crowe,B.L., North,J.A., McKee,C.J., Shkriabai,N., Feng,L., Plumb,M., Graham,R.L., Gorelick,R.J., Hess,S. *et al.* (2013) Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. *Nucleic Acids Res.*, **41**, 3924–3936.
6. Pradeepa,M.M., Sutherland,H.G., Ule,J., Grimes,G.R. and Bickmore,W.A. (2012) Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genet.*, **8**, e1002717.
7. Ciuffi,A., Llano,M., Poeschla,E., Hoffmann,C., Leipzig,J., Shinn,P., Ecker,J.R. and Bushman,F. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.*, **11**, 1287–1289.
8. Sowd,G.A., Serrao,E., Wang,H., Wang,W., Fadel,H.J., Poeschla,E.M. and Engelman,A.N. (2016) A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E1054–E1063.
9. Maskell,D.P., Renault,L., Serrao,E., Lesbats,P., Matadeen,R., Hare,S., Lindemann,D., Engelman,A.N., Costa,A. and Cherepanov,P. (2015) Structural basis for retroviral integration into nucleosomes. *Nature*, **523**, 366–369.
10. Serrao,E., Krishnan,L., Shun,M.C., Li,X., Cherepanov,P., Engelman,A. and Maertens,G.N. (2014) Integrase residues that determine nucleotide preferences at sites of HIV-1 integration: implications for the mechanism of target DNA binding. *Nucleic Acids Res.*, **42**, 5164–5176.
11. Naughtin,M., Haftek-Terreau,Z., Xavier,J., Meyer,S., Silvain,M., Jaszczyszyn,Y., Levy,N., Miele,V., Benleulmi,M.S., Ruff,M. *et al.* (2015) DNA physical properties and nucleosome positions are major determinants of HIV-1 integrase selectivity. *PLoS ONE*, **10**, e0129427.
12. Pasi,M., Mornico,D., Volant,S., Juchet,A., Batisse,J., Bouchier,C., Parissi,V., Ruff,M., Lavery,R. and Lavigne,M. (2016) DNA minicircles clarify the specific role of DNA structure on retroviral integration. *Nucleic Acids Res.*, **44**, 7830.
13. Benleulmi,M.S., Matysiak,J., Henriquez,D.R., Vaillant,C., Lesbats,P., Calmels,C., Naughtin,M., Leon,O., Skalka,A.M., Ruff,M. *et al.* (2015) Intasome architecture and chromatin density modulate retroviral integration into nucleosome. *Retrovirology*, **12**, 13.
14. Lesbats,P., Botbol,Y., Chevereau,G., Vaillant,C., Calmels,C., Arneodo,A., Andreola,M.L., Lavigne,M. and Parissi,V. (2011) Functional coupling between HIV-1 integrase and the SWI/SNF chromatin remodeling complex for efficient in vitro integration into stable nucleosomes. *PLoS Pathog*, **7**, e1001280.
15. Winans,S., Larue,R.C., Abraham,C.M., Shkriabai,N., Skopp,A., Winkler,D., Kvaratskhelia,M. and Beemon,K.L. (2017) The FACT complex promotes avian leukosis virus DNA integration. *J. Virol.*, **91**, e00082-17.
16. Matysiak,J., Lesbats,P., Mauro,E., Lapaillerie,D., Dupuy,J.-W., Lopez,A.P., Benleulmi,M.S., Calmels,C., Andreola,M.-L., Ruff,M. *et al.* (2017) Modulation of chromatin structure by the FACT histone chaperone complex regulates HIV-1 integration. *Retrovirology*, **14**, 39.
17. Benleulmi,M.S., Matysiak,J., Robert,X., Miskey,C., Mauro,E., Lapaillerie,D., Lesbats,P., Chaignepain,S., Henriquez,D.R., Calmels,C. *et al.* (2017) Modulation of the functional association between the HIV-1 intasome and the nucleosome by histone amino-terminal tails. *Retrovirology*, **14**, 54.
18. Hare,S., Gupta,S.S., Valkov,E., Engelman,A. and Cherepanov,P. (2010) Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature*, **464**, 232–236.
19. Ballandras-Colas,A., Maskell,D.P., Serrao,E., Locke,J., Swuec,P., Jónsson,S.R., Kotecha,A., Cook,N.J., Pye,V.E., Taylor,I.A. *et al.* (2017) A supramolecular assembly mediates lentiviral DNA integration. *Science*, **355**, 93–95.
20. Busschots,K., Vercammen,J., Emiliani,S., Benarous,R., Engelborghs,Y., Christ,F. and Debyser,Z. (2005) The interaction of LEDGF/p75 with integrase is Lentivirus-specific and promotes DNA binding. *J. Biol. Chem.*, **280**, 17841–17847.
21. Schanda,P. and Brutscher,B. (2005) Very fast Two-Dimensional NMR spectroscopy for Real-Time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.*, **127**, 8014–8015.
22. Favier,A. and Brutscher,B. (2011) Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J. Biomol. NMR*, **49**, 9–15.
23. Solyom,Z., Schwarten,M., Geist,L., Konrat,R., Willbold,D. and Brutscher,B. (2013) BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J. Biomol. NMR*, **55**, 311–321.
24. Grzesiek,S. and Bax,A. (1992) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson. 1969*, **99**, 201–207.
25. Grzesiek,S. and Bax,A. (1993) Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J. Biomol. NMR*, **3**, 185–204.

26. Schanda,P., Kupče,Ē. and Brutscher,B. (2005) SOFAST-HMQC experiments for recording Two-Dimensional deteronuclear correlation spectra of proteins within a few seconds. *J. Biomol. NMR*, **33**, 199–211.
27. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,M., Ulrich,E.L., Markley,J.L., Ionides,J. and Laue,E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–96.
28. Eijkelenboom,A.P., Lutzke,R.A., Boelens,R., Plasterk,R.H., Kaptein,R. and Hard,K. (1995) The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nat. Struct. Biol.*, **2**, 807–10.
29. Lodi,P.J., Ernst,J.A., Kuszewski,J., Hickman,A.B., Engelman,A., Craigie,R., Clore,G.M. and Gronenborn,A.M. (1995) Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry (Mosc.)*, **34**, 9826–9833.
30. Passos,D.O., Li,M., Yang,R., Rebensburg,S.V., Ghirlando,R., Jeon,Y., Shkriabai,N., Kvaratskhelia,M., Craigie,R. and Lyumkis,D. (2017) Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science*, **355**, 89–92.
31. Borrenberghs,D., Zurnic,I., De Wit,F., Acke,A., Dirix,L., Cereseto,A., Debyser,Z. and Hendrix,J. (2018) Post-mitotic BET-induced reshaping of integrase quaternary structure supports wild-type MLV integration. *Nucleic Acids Res.*, doi:10.1093/nar/gky1157.