HAL open science

DNA minicircles clarify the specific role of DNA structure on retroviral integration

Marc Pasi, Damien Mornico, Stevenn Volant, Anna Juchet, Julien Batisse, Christiane Bouchier, Vincent Parissi, Marc Ruff, Richard Lavery, Marc Lavigne

To cite this version:

Marc Pasi, Damien Mornico, Stevenn Volant, Anna Juchet, Julien Batisse, et al.. DNA minicircles clarify the specific role of DNA structure on retroviral integration. Nucleic Acids Research, 2016, 44 (16), pp.7830 - 7847. 10.1093/nar/gkw651 . pasteur-01416869

HAL Id: pasteur-01416869

https://pasteur.hal.science/pasteur-01416869v1

Submitted on 14 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

7830–7847 Nucleic Acids Research, 2016, Vol. 44, No. 16 Published online 20 July 2016
doi: 10.1093/nar/gkw651

DNA minicircles clarify the specific role of DNA structure on retroviral integration

Marco Pasi[1], Damien Mornico[2,†], Stevenn Volant[2,†], Anna Juchet[3], Julien Batisse[4], Christiane Bouchier[5], Vincent Parissi[6], Marc Ruff[4], Richard Lavery[1] and Marc Lavigne[3,*]

[1]MMSB UMR5086 University of Lyon I/CNRS, Institut de Biologie et Chimie des Protéines, 7 passage du Vercors, Lyon 69367, France, [2]Institut Pasteur—Bioinformatics and Biostatistics Hub—C3BI, USR 3756 IP-CNRS, Paris 75015, France, [3]Institut Pasteur, Unité de Virologie Moléculaire et Vaccinologie, UMR 3569 IP-CNRS, Paris 75015, France, [4]Institut de Génétique et de Biologie Moléculaire et Cellulaire, Dpt de Biologie Structurale Intégrative, UDS, U596 INSERM, UMR7104 CNRS, Illkirch 67400, France, [5]Institut Pasteur, PF1, Plate-forme Génomique—Pôle Biomics, Citech, Paris 75015, France and [6]Laboratoire de Microbiologie Fondamentale et Pathogénicité, UMR 5234 CNRS—Université de Bordeaux, Bordeaux 33000, France

Received May 25, 2016; Revised July 8, 2016; Accepted July 11, 2016

ABSTRACT

Chromatin regulates the selectivity of retroviral integration into the genome of infected cells. At the nucleosome level, both histones and DNA structure are involved in this regulation. We propose a strategy that allows to specifically study a single factor: the DNA distortion induced by the nucleosome. This strategy relies on mimicking this distortion using DNA minicircles (MCs) having a fixed rotational orientation of DNA curvature, coupled with atomic-resolution modeling. Contrasting MCs with linear DNA fragments having identical sequences enabled us to analyze the impact of DNA distortion on the efficiency and selectivity of integration. We observed a global enhancement of HIV-1 integration in MCs and an enrichment of integration sites in the outward-facing DNA major grooves. Both of these changes are favored by LEDGF/p75, revealing a new, histone-independent role of this integration cofactor. PFV integration is also enhanced in MCs, but is not associated with a periodic redistribution of integration sites, thus highlighting its distinct catalytic properties. MCs help to separate the roles of target DNA structure, histone modifications and integrase (IN) cofactors during retroviral integration and to reveal IN-specific regulation mechanisms.

INTRODUCTION

Integration of the DNA copy of the viral genome into the DNA genome of infected cells is an essential step of retroviral replication. This is performed by a viral-encoded enzyme, called integrase (IN), whose enzymatic and biochemical properties have been well characterized (reviewed in (1)). In the case of HIV-1, IN catalytic inhibitors are efficient anti-viral compounds included in highly active antiretroviral therapies (reviewed in (2)). A new generation of antiviral compounds, targeting the interaction between IN and cell host proteins, has emerged. In addition to their therapeutic properties, these molecules have revealed new roles of the targeted interactions during the viral replication cycle.

Retroviral integration is not random and the genomic distribution of integration sites differs between retroviruses. IN selectivity is regulated at different levels and this regulation requires specific virus-host protein–protein and protein–DNA interactions (reviewed in (3,4)). In the case of HIV-1, euchromatin domains located below the nuclear pores of the infected cells constitute a first level of IN selectivity, with a specific role of nuclear pore proteins (5,6). At a second level, HIV-1 IN targets the bodies of active and highly spliced genes present in gene-dense regions of chromosomes (7–10). Two cellular cofactors, LEDGF/p75 and CPSF6, interacting with HIV-1 IN and Capsid respectively, are involved in this selectivity (11–15). In the case of LEDGF/p75, its interaction with both HIV-1 IN and the H3K36me3 modified histone is responsible for the preferential integration in active genes bodies, enriched in this histone mark (16–18). *In vitro*, LEDGF/p75 specifically stimulates integra-

*To whom correspondence should be addressed. Tel: +33 1 40 51 65 77; Fax: +33 1 40 51 65 70, Email: marc.lavigne@pasteur.fr
†These authors contributed equally to the paper as second authors.
Present addresses:
Marco Pasi, School of Pharmacy and Centre for Biomolecular Sciences, University of Nottingham, Nottingham NG7 2RD, UK.
Marc Lavigne, Institut Cochin, Laboratoire ≪ Interaction Hôte-Virus≫, Inserm U1016-CNRS UMR8104-Université Paris Descartes, Paris 75014, France.

© The Author(s) 2016. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

tion in chromatin templates (19,20), and also stabilizes the tetrameric active form of HIV-1 IN (21).

The third and more local level of IN selectivity is the target DNA and its structure in nucleosomes. *In vitro*, HIV-1, MLV and PFV INs integrate preferentially into nucleosomes or in curved DNA sequences (22–27). Chromatin compaction, density and remodeling also affect the efficiency of integration in poly-nucleosome reconstituted templates (22,23,28). In infected cells, if a weak consensus nucleotide sequence is present around integration sites (9,29), the same sites show a stronger consensus for highly flexible DNA sequences (25,29–31). This flexibility (also termed bendability) allows the target DNA to be dramatically deformed by the enzyme, as observed in cryo-electron microscopy (cryo-EM) and crystal structures of the HIV-1, PFV and RSV intasomes (21,32,33). In infected cells, HIV-1 and MLV INs also target nucleosomes and more precisely their widened DNA major grooves (9,25,34). At this local level, both DNA and histones contribute to IN selectivity, but the distinct roles of each component are not clearly defined. For this purpose, we chose to construct DNA minicircles (MCs) that mimic nucleosome-induced DNA curvature in the absence of histones and to use them as integration substrates.

MCs (between 64 and 200 bp) have already been used to study DNA structural parameters and their influence on protein–DNA interactions (reviewed in (35)). Using the ligase-assisted MC accumulation (LAMA) protocol (36), they can be constructed with any nucleotide sequence and can include any protein-binding site or enzymatic target sequence (37). MCs have several advantages compared to small DNA oligonucleotides, large linear DNA fragments or supercoiled plasmids. Their structures can be studied by chemical or enzymatic digestion or by cryo-EM (36,38,39). Their small size is compatible with appropriate molecular modeling studies that can shed light on their structural properties as a function of their fixed sizes and different degrees of superhelicity (35,39–42). Indeed, MCs offer the possibility of controlling both curvature and torsion of the DNA helix, by simply varying the number of nucleotides in the circle. For example, a 75-bp MC should closely reproduce the DNA curvature present in a nucleosome. In addition, the rotational register of the MC (defined as the orientation of a given base pair with respect to the direction of curvature induced by circularization) (43) can be controlled by introducing phased poly-adenine tracts (A-tracts) into the MC sequence. This allows the orientation of DNA grooves to be defined with respect to the inside and outside surfaces of the MC, in a similar way to that imposed by protein–DNA interactions in nucleosomes, but in a histone-free manner.

MC structural properties and preferential retroviral integration into nucleosomes prompted us to use MCs to study the specific role of nucleosome-induced DNA curvature in this process. We observed a significant enhancement of HIV-1 integration in MCs with respect to linear DNA fragments (named Fts) of the same length and nucleotide sequence. This enhancement is slightly increased by LEDGF/p75, but is not changed by IN mutations that modify the recognition of target DNA flexibility. With or without LEDGF/p75, we observed that DNA circularization induces a major redistribution of HIV-1 integration sites toward the outer surface of MCs. This redistribution is favored by LEDGF/p75, even in the absence of histones. DNA circularization also enhances PFV integration, but this enhancement is not associated with a periodic redistribution of integration sites in MCs. Thus, structurally constrained and protein-free MCs provide novel data on the role of DNA structure in retroviral integration and reveal differential regulations of retroviral INs.

MATERIALS AND METHODS

Oligonucleotides

Sequence of the oligonucleotides used in to clone the 75, 75S, 86 and 86S fragments and to quantify and sequence integration products (IPs) are described in Supplementary Table S1.

Plasmids and protein purification

Cloning of the DNA fragments (Fts) for MC production: 75, 86, 75S and 86S sequences were cloned in two shifted Fts (NheI/XbaI or BamHI/BglII) necessary for the production of DNA MCs (Figure 1 and Supplementary Figure S1). These Fts were cloned as eight tandem copies in the pUC18 vector by the following protocol. Chronologically, four DNA hybrids were made with synthetic oligonucleotides H1(ML1F/ML1R), H2(ML2F/ML2R) H3(ML3F/ML3R) and H4(ML4F/ML4R) and digested with pairs of restriction enzymes (Acc65I/BglII or NheI/SalI for H1 and H2, Acc65I/XbaI or BamHI/SalI for H3 and H4). Digested hybrids were ligated by pairs (Acc65I/BglII H1 or H2 with BamHI/SalI H3 or H4 and Acc65I/XbaI H3 or H4 with NheI/SalI H1 or H2) and cloned between the Acc65I-SalI restriction sites of pUC18. This initial cloning step gave eight pUC18 derived plasmids containing one copy of NheI/XbaI Ft (H1 + H3, H2 + H3, H1 + H4, H2 + H4), or one copy of BamHI/BglII Ft (H3 + H1, H3 + H2, H4 + H1, H4 + H2) and corresponding to the 75, 86, 75S and 86S sequences respectively. The second step was the amplification of these Fts, from one to eight copies, using the Acc65I, XhoI, SalI restriction sites and the strategy described in (44). Briefly, at each amplification step, the Acc65I-SalI Ft of each plasmid was cloned between the Acc65I-XhoI sites of the same plasmid. This operation was repeated three times in order to obtain eight tandem-repeats. Plasmids obtained at intermediate and final steps of this cloning strategy were all sequenced.

The NheI/XbaI and BamHI/BglII Fts were obtained from the final plasmids by two steps. First, the Acc65I-SalI Ft of these plasmids (containing the eight copies) was purified by 1.5% agarose-TAE electrophoresis and gel extraction (nucleoSpin Gel and polymerase chain reaction (PCR) Clean-up Macherey Nagel kit). Second, the NheI/XbaI or BamHI/BglII Fts were obtained by extensive overnight digestion of the Acc65I-SalI Fts and purified by the same kit (with no electrophoresis step). The quality and yield of final DNA Fts were checked by 8% Polyacrylamide Gel Electrophoresis (Figure 1B) and nanodrop measurement.

To obtain standards of qPCR quantification of IP, the HIV-1 or PFV SupF-NdeI integration substrate (see In-

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

tegration assays) were filled with the Klenow Fragment and cloned in the filled NcoI or EcoRV restriction sites of plasmids containing one single copy of NheI/XbaI and BamHI/BglII Fts of 75 and 86 sequences (plasmids described above). Plasmids corresponding to the insertion of the SupF-NdeI integration substrate in the orientation that mimics a U5 integration were selected. The NheI/XbaI and BamHI/BglII Fts were obtained from the selected plasmids and purified by 1.5% agarose-TAE electrophoresis and gel extraction.

HIV-1 IN, the HIV-1 IN–LEDGF/p75 complex and PFV IN were purified as described in (21,45,46). S119G, S119T, R231G and S119T/R231G mutations were introduced in the pET15b-His-HIV-1 IN expression vector (45), using the QuikChange II Site-Directed Mutagenesis Kit (Agilent) and specific mutated oligonucleotides. Mutated HIV-1 IN were expressed and purified similarly as WT HIV-1 IN (45).

MCs synthesis and quality control

MCs of 75, 86, 75S and 86S sequences were produced following the LAMA protocol described in (36). Briefly, this protocol relies on repeated denaturation, annealing and ligation of two DNA Fts that have complementary sequences shifted by half of their size (Figure 1A and Supplementary Figure S1). In our case, pairs of Fts corresponding to each MC, are bordered by cohesive restriction ends (NheI/XbaI or BamHI/BglII). Eight repeats of these fragments were cloned in a pUC plasmid (strategy adapted from the one used to amplify DNA segments for MN assembly Dyer *et al*., (44), 'Materials and Methods' section) and obtained by restriction digestion and gel purification. No PCR amplification was used, in order to avoid mutations or deletion in the A-tracts. After the LAMA protocol, DNA MC were digested by both Exonucleases I and III to remove non circularized DNA Fts and purified with the nucleoSpin Gel and PCR Clean-up Macherey Nagel kit.

MCs were digested by Bal31 and S1 nucleases from NEB. A total of 50 ng of MC were digested in NEB reaction buffers during 1 h at 25°C by 0.015 U/μl or 0.005 U/μl of Bal31 and 0.6 U/μl or 0.2 U/μl of S1 nuclease and separated by 8% polyacrylamide gel electrophoresis (PAGE) in Tris Borate EDTA buffer (Supplementary Figure S2).

Structural modeling and molecular dynamics simulation studies of MCs

The three-dimensional (3D) structures of the four MC constructs (MC75, MC86, MC75S and MC86S) were predicted using the internal coordinate program JUMNA (47). In particular, a closed, planar circular DNA molecule was created using the standard value of rise (3.38 Å), and twist corresponding to linking number 7 expected for 75 bp with a helical repeat of 10.5 bp (i.e. 33.6°). A similar procedure was used to construct an 86 bp-long MC with linking number 8, using a value of twist of 33.5°. The presence of phased A-tracts in the sequence of the four constructs is expected to limit variations in their rotational register. In order to identify the most stable register, the sequence of each of the four constructs was threaded through the MC, by replacing each base pair with the preceding and then subjecting the resulting structures to conjugate gradient energy minimization using the AMBER parm99 force field (48) with the parmBSC0 modifications (49). This is equivalent to progressively varying the register at each step of the procedure by the value of the imposed twist. The resulting final energies as a function of sequence step show, as expected, a periodic behavior, with sharp minima separated by about 10 bp. The minimum energy structures across the whole register range were selected (Figure 2) and used throughout this work.

This structure prediction procedure was validated through comparison with explicit-solvent molecular dynamics (MDs) simulation for the MC75 construct. The predicted structure was centered in a truncated octahedral box filled with SPC/E water (50) with the same sodium chloride (51) concentration used in the integration experiments. Simulations were run with GROMACS 5 (52–54) using periodic boundary conditions, under controlled temperature and pressure (298K and 1 atm), using the Bussi thermostat (55) and the Berendsen barostat (56). Long-range electrostatics were treated using the particle mesh Ewald method (57) and bond lengths involving hydrogen atoms were restrained using P-LINCS (58,59) to allow for a 2-fs time step. The system was carefully minimized and thermalized following a standard protocol (60) before collecting 100 ns production simulation. Only the final 50 ns of the simulation were used for all further analyses. Molecular graphics in Figures 2 and 7 were produced using Chimera (61).

Analysis of minicircle curvature

In order to define whether the major groove a given location along the MC is facing toward the inside or the outside of the MC, we studied the local curvature of the predicted MC structures, using a procedure that uses the helical axis calculated by the Curves+ program (62,63). First, a locally optimal circle is fit through three equally spaced points along the helical axis, corresponding to base pair levels. The spacing of these points determines the sensitivity of the results to local deformations of the helical axis: best results for the present application were obtained using a spacing of 5 bp. Next, the center of the fitted circle is located, and its curvilinear helicoidal coordinates are measured using a procedure similar to that described in (64). In particular, φ is defined as the angle between the major groove dyad and the vector connecting the helical axis to the center of the locally optimal circle, measured in the plane of the central base pair (note that $\varphi = A\text{-}\pi/2$, where A is the curvilinear helicoidal angular coordinate of the center of the fitted circle). φ can cover the range 0–360°: values around 0° indicate that the major groove is facing toward the inside of the MC at that base-pair location, while values around 180° imply that the minor groove is facing inward and the major groove is facing outside. As stated previously, several phased A-tracts were inserted in the sequence of the four constructs considered in this work in order to bias the rotational register of the MCs. This will limit the range of accessible values of φ at each base-pair level, most strongly in the immediate vicinity of the A-tracts. In order to quantify the residual variability in φ, we calculated the circular standard deviation of

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

φ, $S(\varphi)$, for each base pair, during the second half of the 100 ns MD trajectory of the MC75 construct described in the previous section. The rose plots representations shown in Figures 5C, 5D and 6D were produced using Matplotlib v1.5.0 (65).

Low-resolution 3D modeling

The structure of HIV-1 IN–LEDGF/p75 in complex with MC75 and MC86 was modeled (Figure 7), starting with the 14 Å-resolution cryo-EM density map described in (21), on which we fitted the modeled structures of the MC75 and MC86 constructs. Almost four turns of atomic-resolution target DNA were fitted into the cryo-EM density as described (21). To obtain the best relative position of the MCs with respect to the enzymatic complex, two turns of DNA from the MC were rigid-body least-square fitted to the central two turns of the atomic-resolution DNA (using only phosphate atoms), before measuring the heavy-atom RMSD (root-mean-square deviation). This procedure was repeated for each possible position and orientation of the MC, and then the position with lowest RMSD was selected. The RMSD values calculated during the procedure show a clearly periodic behavior, with troughs roughly every ∼10 bp (data not shown). The procedure for MC86 was recorded in a movie using Chimera (61), and is available as supporting information (Supplementary Movie SM1).

IN integration assays

Integration assays were performed using a protocol adapted from (19,25). HIV-SupF-NdeI or PFV-SupF-NdeI donor substrates were mixed with HIV-1 IN, HIV-1 IN–LEDGF/p75 or PFV IN (1 μm, 1 μm and 100 nM of monomeric concentration respectively) in the integration buffer (20 mM Hepes pH 7.4, 50 mM NaCl, 5 mM MgCl2, 10 uM ZnCl2, 1 mM 1,4-Dithiothreitol and 100 μg/ml bovine serum albumin). After 10 mn at 4°C and 10 mn at 37°C, acceptor linear fragment or MC (50 nM) was added to the reaction mixtures and incubated during 1 h at 37°C. Reaction was stopped by addition of 0.1% sodium dodecyl sulphate, 10 mM ethylenediaminetetraacetic acid, 1 μg/ul PNK enzyme and one hour incubation at 37°C. IPs were purified with the nucleoSpin PCR Clean-up kit (Macherey Nagel) and digested by EcoRV (for IPs obtained in MCs and NheI-XbaI Fts) or NcoI (for IPs obtained in MCs and BamHI/BglII Fts). This digestion step allows the comparison of integration efficiencies between linear fragments and MCs.

IPs were quantified by real-time PCR using a set of two primers, one complementary to the donor substrate (P1HIV, P3HIV of P3HIV) and one complementary to the acceptor template and bordering the target integration segment (MCF, MCR3 or MCR4) (Supplementary Figure S1). Artificial IPs, containing one integrated copy of viral donor substrate cloned in the EcoRV or NcoI restriction site of the 75 and 86 bp constructs, were used for the qPCR standard curves. Each integration sample was analyzed in triplicate by real-time PCR and integration assays were repeated between two and six times. A *t*-test was used to compare the weighted means of ratios of integration efficiency in MC compared to Ft (named Ratio (MC/Ft)), for each condition, with a level of significance of 5%. The Benjamini–Hochberg method was used to correct for multiple testing.

IPs were amplified by regular PCR using primers pairs identical as for real-time PCR, with Pfusion DNA polymerase and a small number of amplification cycles. Amplification products were quantified and mixed to obtain the resulting sequencing libraries.

Sequencing librairies were performed using the NEXTflex PCR-Free DNA-Seq kit for Illumina (Bio Scientific) and sequenced on MiSeq machine (Illumina) in single reads of 250 bases. Sequence files were generated using Illumina Analysis Pipeline version 1.8 (CASAVA).

Mapping of integration sites and correlative studies

In order to identify the absolute position of the integration sites for each sequence of the 24 libraries, a custom python script was implemented to read these sequences and count, for each position within the region of interest, the number of integrations achieved. These positions are calculated from the distances between the acceptor and the donor primers. The output is a list of counts per position, for each library.

In this study, 24 libraries of FastQ files have been analyzed, containing from 15 000 to 74 125 sequences. A first step of filtering excluded sequences without the two good primers (acceptor and donor) expected. A second filtering allowed the exclusion of sequences having the integration site outside the region of interest. Around 52% of the initial sets of sequences were finally conserved for this analysis.

As a pre-processing step, the data were log-transformed and the median normalization scheme was used for its robustness to outliers. A Spearman correlation coefficient was calculated between the integration sites of each pair of libraries. This leads to a clustering of all the libraries (Supplementary Figure S5). To quantify the effect of DNA circularization on the distribution of integration sites, we calculated the ratio of normalized integration frequencies in the MC versus the normalized integration frequencies in the linear Ft, at each position along the integration target sequence (ITS). The $\log_{10}$ value of this ratio (called CAI for circular activated integration) was plotted along the studied sequences (Figures 4B, 6A and B). To evaluate the periodicity of the CAI values, we calculated the cross-correlation between CAI measured on same strands (US/US, LS/LS) or complementary strands (LS/US), for each pair of target sequence and IN (Supplementary Figure S6).

RESULTS

DNA minicircles mimic nucleosome-induced DNA curvature

MCs were constructed to mimic the major structural deformations of DNA within a nucleosome, without requiring histones. Four sequences were selected, two of 75 bp (75 and 75S) and two of 86 bp (86 and 86S). In fact, 75 bp MCs correspond to seven helical turns of DNA and mimic one of the two DNA gyres in a nucleosome (66), while 86 bp MCs correspond to a slightly larger circular DNA (eight helical turns) and are characterized by a smaller average curvature angle (45° per helical turn for 86 bp MCs compared to 51.4° per helical turn for 75 bp MCs). The four MCs should not

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

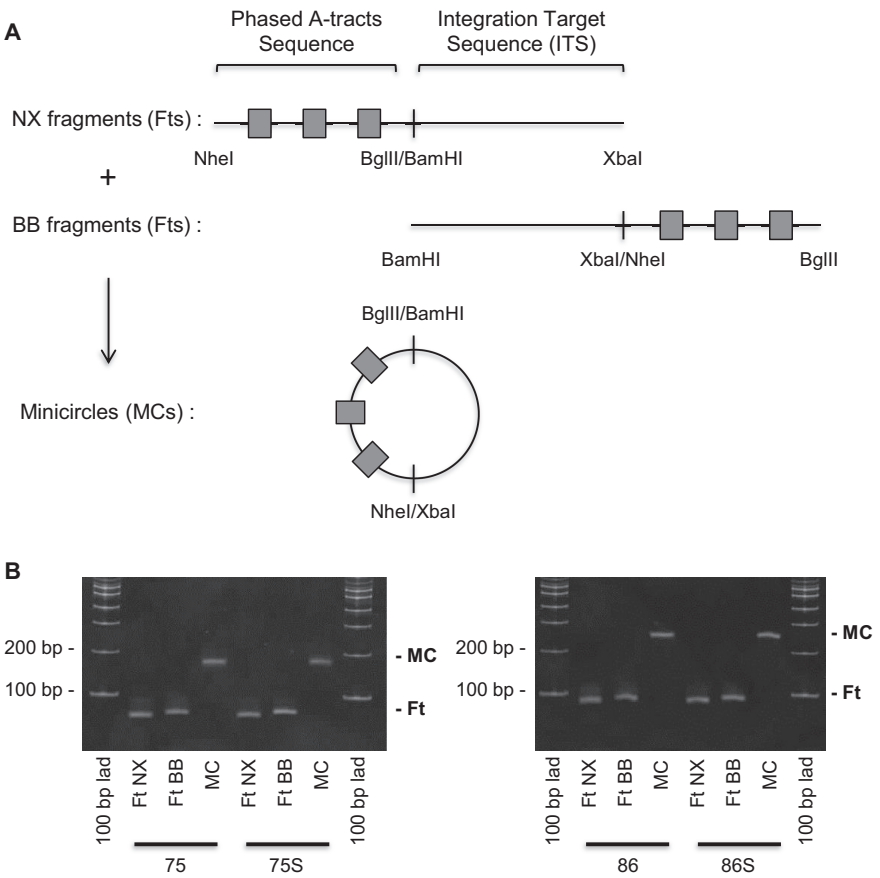

Figure 1. Linear DNA fragments (Fts) and Minicircles (MCs) used for the study. (**A**) Selected sequences and circularization strategy. The four selected sequences, called 75, 86, 75S and 86S, were obtained as two overlapping Fts, NheI/XbaI (NX) and BamHI/BglII (BB), that were used to synthesize the corresponding MCs. (**B**) Control of the circularization. Ethidium-bromide staining of linear Fts and MCs separated by 8% PA-TBE electrophoretic migration.

exhibit significant superhelicity since their lengths were chosen to achieve an integer number of helical turns. These MC structures can be predicted at atomic resolution (41) and the modeled structures can be used to study the specific role of DNA structure in the selectivity of nucleosome-binding proteins (or enzymes), in the absence of histones.

The 75, 75S, 86 and 86S sequences are composed of two domains (Figure 1 and Supplementary Figure S1). One contains three or four phased A-tracts that bias the rotational register of the MCs (43). The second contains a 40-bp sequence that will be studied for its integration properties (the Integration Target Sequence, or ITS). These ITSs contain an identical 23 bp segment, derived from the TPT construct (67) and inserted at two different distances from the phased A-tracts. These distances are shifted by 5 bp between the 75/86 and 75S/86S constructs. This 5 bp shift changes the rotational register of this segment by roughly 180°, providing information on the specific effects of sequence and structure on integration.

MCs were produced by the LAMA protocol (36) with pairs of NheI/XbaI and BamHI/BglII Fts (Figure 1A, Supplementary Figure S1 and 'Materials and Methods' section). MCs were digested by Exonucleases I and III to remove non-circularized DNA and their quality was checked by PAGE in native conditions (Figure 1B). MCs electrophoretic migration was retarded with respect to Fts of the same sequence. MC75 and MC86 were resistant to Bal31 and S1 nucleases digestion (at dilutions used in (36)) suggesting the absence of unpaired bases, DNA kinks, backbone lesions or B-Z junctions in these constructs (Supplementary Figure S2). Therefore, the DNA helix in the 75 and 86 bp MCs is expected to adopt a canonical B conformation, at least within the ITS. Before using these MCs as integration substrates, we studied their structure using molecular modeling and MDs simulation.

Structural properties of the MCs

In order to rationalize differences in integration efficiency between linear and circular constructs, we predicted a representative 3D structure of the four MCs at atomic resolution. Phased A-tracts inserted in the sequence of the four constructs should restraint the rotational freedom (register) of DNA within the MCs. To evaluate this effect, we calculated the deformation energy for each construct as a function of register variation. The results are presented in Figure 2A (see Methods for details of calculations). Deformation energies show a roughly periodic behavior and sharp minima separated by roughly 360°, confirming that the phased

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

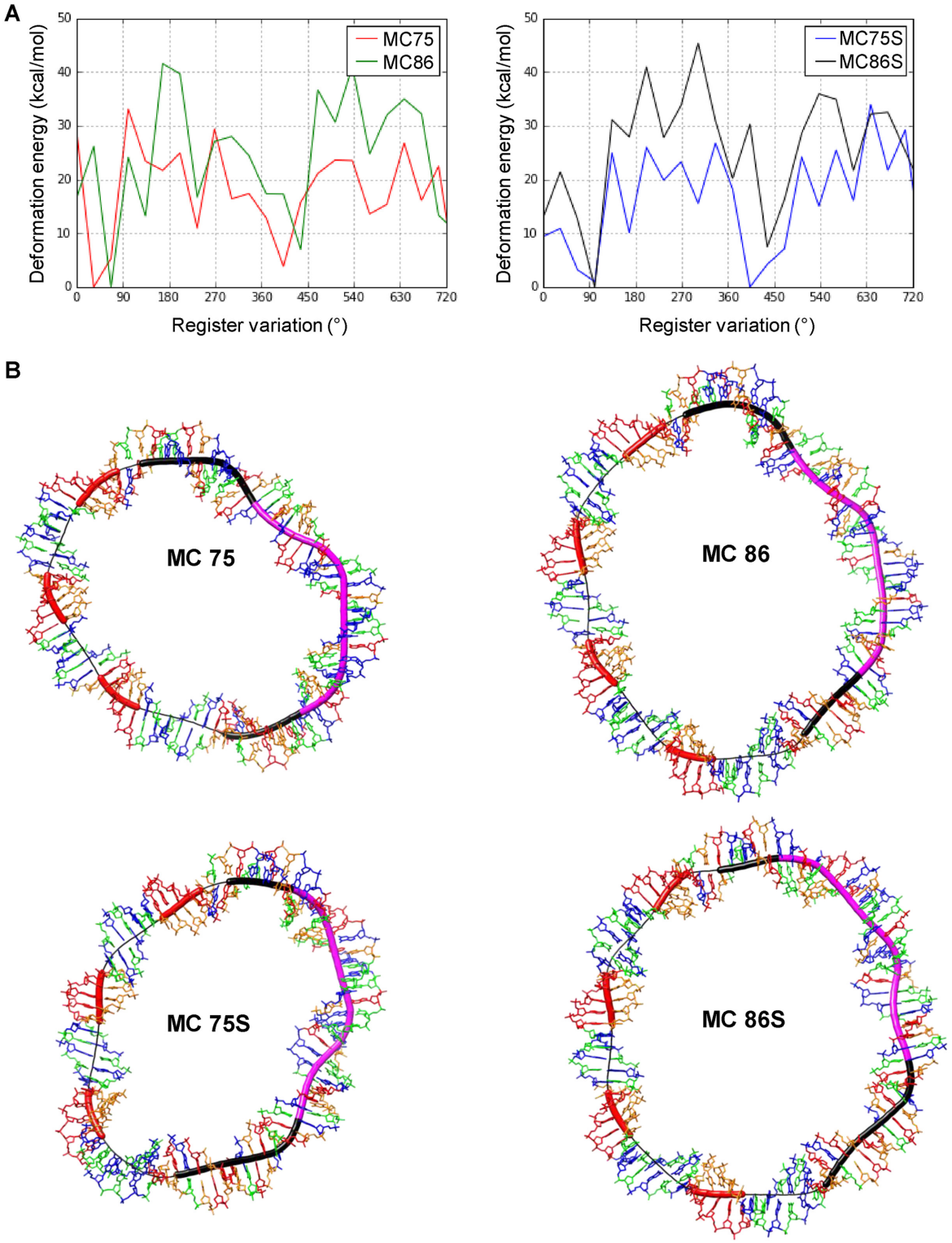

Figure 2. Structural modeling of DNA MCs. (**A**) Deformation energy as a function of register variation for the MC75 and MC86 constructs and the MC75S and MC86S constructs, showing sharp troughs every ~10 bp. (**B**) Representative 3D structures of the MC75, MC75S, MC86 and MC86S constructs shown as sticks, colored according to the base type (A red, C green, G blue and T orange); the Curves+ DNA helical axis is shown as a thin black line, with thicker regions to highlight specific portions of the MCs: black and purple for the ITS (purple is specific of the 23-bp segment common to the four constructs) and red for the phased A-tracts. Molecular graphics were produced using Chimera.

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

A-tracts indeed induce a strong preference for a restricted range of register values.

The structures with the smallest deformation energy were chosen as representative for each MC (Figure 2B). Overall, the predicted structures are roughly planar, as expected for superhelically relaxed MCs. Curvature is rather uniformly distributed throughout each MC, apart from small local heterogeneities due the different propensity of DNA to bend toward the major or the minor groove. These small variations in curvature give rise to slightly irregular MC contours, in particular for the 75 bp constructs. We note that the extent of these variations is well within the range of curvature values resulting from thermal fluctuations (see below). In addition, we observe no major departures from canonical B-DNA, such as sharp kinks (involving base unstacking), or broken base pairs.

The predictions presented above will serve as a basis for the structural interpretation of differences in integration efficiency between MCs and Fts of the same sequence. In this context, these predictions suffer from two main limitations. First, MCs in solution at room temperature will undergo significant structural fluctuations; using a single structure gives no information regarding how stable the predicted structural features are to these fluctuations. Second, necessary approximations (notably in the treatment of the effects of solvent and counterions) made in the prediction procedure may hinder the accuracy of the results. We therefore performed a 100-ns explicit-solvent MD simulation of MC75 to gain insight into the impact of these limitations. The duration of the simulation, although much shorter than the time scale of the experiments, nevertheless would allow us to observe significant conformational rearrangements of the MC, including register and curvature variations, as well as kinking (41,68). The simulation mimics the salt concentration and temperature conditions of the experiments, and provides us with an accurate and reasonably complete picture of the conformations accessible to MC75 during the integration assays.

Analysis of the MD simulation of MC75 confirms the accuracy of the static structural predictions, especially in terms of the predicted register along the MC. Furthermore; we were able to quantify the effectiveness of phased A-tracts in MC75, in terms of the residual variability in rotational register. With this aim, we followed the time evolution of the φ angle, which measures the direction of curvature at each position along the MC (see 'Materials and Methods' section). Results confirm the register selectivity observed previously (Figure 2A), and in particular that within the ITS, the average extent of register fluctuations (in terms of circular standard deviation) is limited to $\pm 18^\circ$, and φ values lie within 20° of the predicted values during 70% of the simulation time.

Target DNA circularization enhances HIV-1 integration

Our goal is to evaluate how DNA curvature induced by its circularization, and similar to that present in the nucleosome, affects HIV-1 integration. We started this study with the HIV-1 IN–LEDGF/p75 complex (21) and the 250-bp HIV-SupF-NdeI viral donor substrate bordered by the pre-processed U3 and U5 ends (69). IPs were quantified by qPCR using primers that border the ITS, and primers complementary to the U5 or U3 ends (Supplementary Figure S1). This strategy allows us to avoid considering IPs within A-tracts that were already shown to affect integration (70,71). It also distinguishes IPs obtained in the two strands of the target substrate and from U3 and U5 viral ends. Finally, IPs within MCs can be opened using a specific restriction enzyme after integration and before the qPCR quantification (see 'Material and Methods' section), allowing a quantitative comparison of integration efficiencies in linear Fts and MCs.

We first observed an enhanced integration, from the U5 end, on both strands of MC75 and MC86, compared to Ft75 and Ft76 respectively (Supplementary Figure S3). The enhancement ratios (MC/Ft) decrease with increasing concentrations of acceptor substrates and a plateau of 10- to 20-fold enhancement is reached at a five to ten times excess of acceptor versus donor substrate. In the following studies, we chose to work at 50 nM of acceptor and 10 nM of donor substrates, which corresponds to these plateau conditions.

Under these conditions, we studied the integration of HIV-SupF-NdeI donor substrate into the four MC constructs, and with linear Fts of the same sequence. Integration of the U5 end in both strands of the acceptor substrate revealed a similar enhancement of integration in the lower strand (orange bars, Figure 3A). On the upper strand, integration is also enhanced on the four constructs (green bars, Figure 3A), but this enhancement is larger on the 75 and 86 than on the 75S and 86S constructs. This difference is due to a lower integration efficiency in the 75S and 86S MC and not to an increased integration in linear Fts.

We then performed integration in the absence of LEDGF/p75. HIV-1 IN alone also integrates more efficiently in the 75 and 86 MCs than in Fts (Figure 3B). However, the MC/Ft integration enhancement ratios measured with IN alone are significantly lower than the ones measured with the IN–LEDGF/p75 complex (5- to 10-fold instead of 15-fold, P-values < 0.001). Therefore, LEDGF/p75 activates integration into 'nucleosome-like' curved DNA, in the absence of histones. This activation is smaller than the one previously observed between naked and chromatinized DNA (19,25) but the use of different integration assays does not allow us to directly compare LEDGF/p75-dependent activations in MCs and nucleosomes.

Serine 119 (S119) and Arginine 231 (R231) of HIV-1 IN are involved in the selectivity of integration (30,72). INs mutated at these residues (S119G, S119T, R231G and S119T/R231G) target nucleotide sequences with different propensities of the DNA major or minor grooves to be compressed. We tested these mutants for the property of enhanced integration in MCs. Integration was performed with the WT and mutated INs, in Fts and MCs of 75 and 86 constructs and the IPs were quantified by qPCR (Figure 3C). We did not observe any significant variation of MC/Ft integration enhancement ratios between the WT and mutated INs. Therefore, the selected mutations do not affect the IN recognition of MCs and target DNA curvature is probably not the only parameter responsible for the changes of IN selectivity observed in infected cells (72).

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

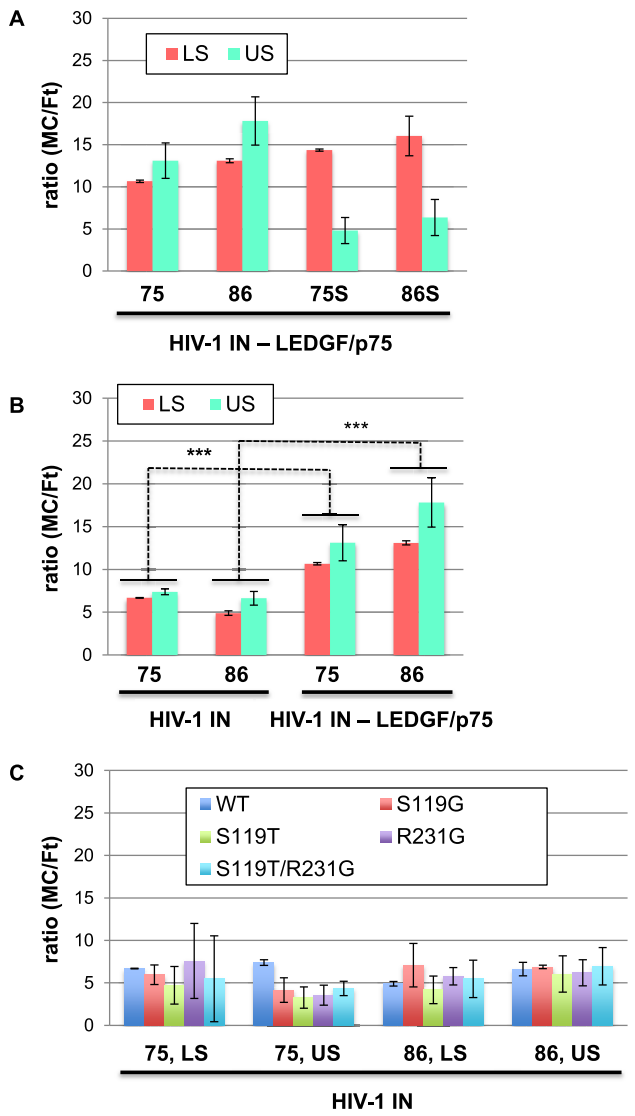

Figure 3. Effect of DNA circularization on the efficiency of HIV-1 integration. (**A**) MC/Ft ratio of integration efficiency by the HIV-1 IN–LEDGF/p75 complex, in the lower strand (LS) and upper strand (US) of the 75, 86, 75S and 86S constructs. (**B**) MC/Ft ratio of integration efficiency by HIV-1 IN and HIV-1 IN–LEDGF/p75 complex, in the LS and US of the 75 and 86 constructs (*** corresponds to P-values < 0.001). (**C**) MC/Ft ratio of integration efficiency by the WT, S119G, S119T, R231G and S119T/R231G HIV-1 INs, in the LS and US of the 75 and 86 constructs. Error bars correspond to 95% confidence intervals of the calculated ratios (A–C).

In summary, MCs are better HIV-1 IN substrates than Fts, reproducing the favored integration into nucleosome-induced curved DNA. MCs also highlight the role of other parameters in the LEDGF/p75-dependent activation of integration into chromatin templates.

MCs reveal a structure-dependent selectivity of HIV-1 integration

We next investigated how DNA deformations induced by its circularization could affect the distribution of HIV-1 integration sites. We started this study with integration by the HIV-1 IN–LEDGF/p75 complex, in the four selected sequences, either linear or circularized (Libraries 1–8, Table 1). IPs were amplified by regular PCR performed independently with four pairs of primers (U3/MC-For, U3/MC-Rev, U5/MC-For and U5/MC-Rev). These PCR conditions specifically amplify IPs from one viral end and in one strand of the acceptor substrate (Supplementary Figure S1). The PCR products obtained for each integration condition were mixed and sequenced on a MiSeq sequencing system. The obtained sequences were analyzed to identify junctions with viral ends (U3 or U5) in order to determine the position of integration sites on each strand of the acceptor substrate (see 'Materials and Methods' section). Integration events were quantified at each positions along the 40 bp ITS and normalized by the total number of events along the whole sequence in order to compare the distribution of integration sites between the different libraries.

We first observed that integration frequencies as a function of sequence clearly correlate between U3 and U5 ends, in all four selected libraries, in both Fts and MCs (correlation coefficients of 0.989 and 0.992 respectively) (Supplementary Figure S4). These correlations show that the selectivity of integration does not depend on the viral end, as already observed on naked and chromatinized DNA templates (25). This justifies summing integration events from both viral ends mapped at identical sites and the use of these combined values in our studies of IN selectivity (Figure 4A).

On both Fts and MCs, we observed a clear superposition of integration profiles between the 75 and 86 constructs (75/86 pair), or the 75S and 86S constructs (75S/86S pair), confirmed by the correlation values measured between the corresponding libraries of integration sites (Supplementary Figure S5). These superpositions are expected for the Fts, since the 40-bp target sequence is identical within each pair of constructs (75/86 or 75S/86S). They show that the PCR and sequencing steps do not modify the positions or frequencies of integration along the selected sequences and validate our protocol for mapping integration sites. On the other hand, in MCs, the superpositions of integration profiles within the 75/86 or 75S/86S pairs could reflect both sequence- and structure-dependent integration selectivity. Our modeling results show that the register along the target sequence is similar within both the 75/86 and 75S/86S pairs of constructs (since it is largely determined by the distance between the target sequence and the phased A-tracts), while the average angle of curvature is different (respectively 4.8 and 4.2° per bp in the 75 bp and 86 bp constructs). These results suggest that the rotational orientation of DNA curvature, rather than the magnitude of the curvature itself (in the range tested here), is a key parameter of IN selectivity.

For each tested construct, integration profiles were clearly different between linear DNA and MCs (Figure 4A and Supplementary Figure S5). These differences correspond to integration sites that were either favored or disfavored by circularization. To distinguish between the sequence- and structure-dependent effects of circularization, we took advantage of the 23-bp segment present in the ITS of the four constructs, which is shifted by 5 bp between the 75/86 and 75S/86S constructs (represented in purple in Figure 4A). Within this 23-bp segment, sequence-specific sites should be shifted, while structure-specific sites should

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

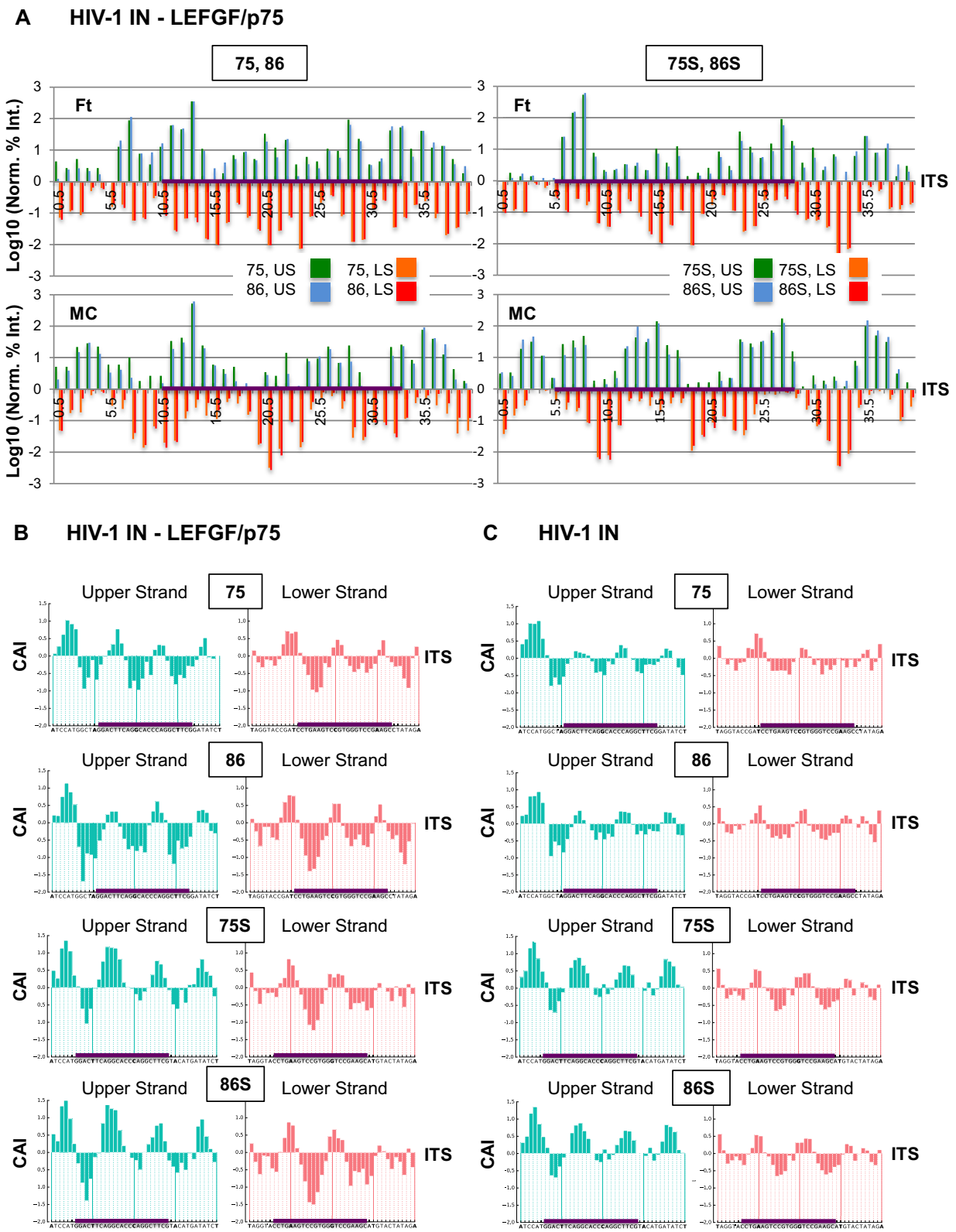

Figure 4. Effect of DNA circularization on HIV-1 IN selectivity of integration. (**A**) Distribution of HIV-1 IN–LEDGF/p75 integration sites along Fts and MCs. Log_{10} value of normalized efficiencies of integration were plotted along the ITSs of 75, 86, 75S, 86S constructs, in Ft and MC conformations. Integration in the US and LS are represented as positive and negative values respectively. The purple line along the abscissa axis represents the 23-bp segment common to the four constructs. (**B** and **C**) Effect of DNA circularization on the HIV-1 integration sites. The CAI values were calculated from integration frequencies measured with HIV-1 IN–LEDGF/p75 (B) or HIV-1 IN (C) and were plotted along the ITSs of the 75, 86, 75S and 86S constructs. Integration in the US and LS are represented on the left and right panels, respectively. The purple line along the abscissa axis represents the 23-bp segment common to the four constructs. The ITSs are represented by their position with respect to the phased attracts domain (A) or their nucleotide sequence (B and C).

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

Table 1. Libraries of integration sites

				reads in the 40 bp "integration target" sequence						
Library	Integrase	Acceptor DNA		total	US	LS	US, U3	US, U5	LS, U3	LS, U5
1	HIV-1 IN - LEDGF/p75	75	Linear	15056	1186	10080	839	347	7887	2193
2			MC	17458	1212	10682	906	306	8419	2263
3		86	Linear	65658	4287	49482	3025	1262	35841	13641
4			MC	64548	3882	38533	2766	1116	26890	11643
5		75S	Linear	27786	2563	20349	1814	749	15641	4708
6			MC	57902	4719	33362	3647	1072	23022	10340
7		86S	Linear	52166	4359	39796	2660	1699	26870	12926
8			MC	47997	2485	31672	1560	925	22747	8925
9	HIV-1 IN	75	Linear	41497	6503	8515				
10			MC	20571	3727	3661				
11		86	Linear	39582	10127	12193				
12			MC	46343	7507	8561				
13		75S	Linear	25651	6712	5825				
14			MC	19209	2782	3069				
15		86S	Linear	74125	21301	26600				
16			MC	59922	11134	11907				
17	PFV IN	75	Linear	23719	3029	4923				
18			MC	16992	3204	3283				
19		86	Linear	49825	4524	7814				
20			MC	43956	6530	10500				
21		75S	Linear	22995	5144	6043				
22			MC	16680	3626	4044				
23		86S	Linear	51200	2321	7557				
24			MC	58502	9776	14444				

maintain the same position with respect to the phased A-tracts and register of rotational orientation in MCs. As expected, integration sites present in this common segment, superimpose well in all four linear Ft constructs, where the effect of shifting the phased A-tracts is negligible. However, this superposition is lost in the MCs. As an example, consider the efficiency of integration at positions 14.5 and 15.5 in lower strand (LS) of the 75/86 constructs, which is decreased in the MCs with respect to linear Fts (Figure 4A). The corresponding sites at positions 9.5 and 10.5 in LS of the 75S/86S constructs, which have the same sequence but a different value of register, are instead enhanced in the MCs. Conversely, sites at positions 14.5 and 15.5 in LS of the 75S/86S constructs show decreased efficiency upon circularization. A similar effect is observed on the upper strand (US), for example at position 20.5 of 75/86, which corresponds to the same nucleotide sequence as position 15.5 of 75S/86S. These observations imply that IN selectivity into the ITS of MCs depends more on its relative position (and thus rotational orientation) with respect to the phased A-tracts than it does on its nucleotide sequence.

To quantify the effect of DNA circularization on IN selectivity, we define R(MC/Ft) as the ratio of the normalized integration frequencies for the MC versus the linear Ft, at each position along the ITS. The $\log_{10}$ values of these R(MC/Ft), named CAI, were plotted along the 40-bp ITS, for both the US and LS of the four constructs (Figure 4B). The resulting plots show a clearly sinusoidal profile, with integration sites periodically favored or disfavored in the MCs (positive and negative peaks respectively). Auto-correlation analysis of the CAI profiles on each strand (US or LS), show a periodic behavior with a first peak around 11 bp (Supplementary Figure S6). This suggests that circularization favors integration on the same side of the DNA helix. Correlation analysis between the CAI profiles on complementary strands (US/LS), also reveals a periodic signal with a first peak around 6 bp (Supplementary Figure S5), a value consistent with integration in the DNA major groove. Finally, the maxima and minima of CAI can nearly be superimposed between the four constructs (Figure 4B) and have the same phasing with respect to the A-tracts.

We also mapped integration frequencies from the U5 end obtained with HIV-1 IN alone, in Fts and MCs of the four constructs (Librairies 9–16, Table 1). CAI was calculated from these frequencies and plotted along the 40-bp ITS (Figure 4C). On both strands of the four constructs, CAI shows sinusoidal profiles similar to the ones observed previously with the HIV-1 IN–LEDGF/p75 complex (Figure 4B). Correlation analysis of CAI profiles also confirm the initial peaks at 11 bp and 6–7 bp for sites identified on the same strand and complementary strands, respectively (Supplementary Figure S5, middle panels). Furthermore, the maxima and minima of CAI are located at the same distance with respect to the phased A-tracts for the four constructs.

Taken together, these results show that, in the absence and presence of LEDGF/p75, there is a redistribution of HIV-1 integration sites in MCs with respect to linear Fts and that integration selectivity depends to a greater extent on DNA structural parameters than on the nucleotide sequence.

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

HIV-1 integration sites are enriched outside of the minicircles and the LEDGF/p75 protein enhances this redistribution

We used the modeled structure of the four MCs (Figure 2) to investigate the structural determinants of IN selectivity. The CAI values, obtained for the HIV-1 IN–LEDGF/p75 complex and for HIV-1 IN alone, were plotted on the structure of the ITS, using a color gradient to represent integration sites favored or disfavored in the MC (yellow to purple, respectively, see Figure 5A and B). For the four constructs and both enzymatic complexes, these representations clearly show that favored integration sites are enriched on the outer surface of the MCs, whereas the inner surface disfavors integration. This in/out distribution is not affected by the 5 bp shift of the target sequence between the 75/86 and 75S/86S constructs, which confirms that the rotational orientation of DNA deformation is the major factor of IN selectivity.

A more quantitative picture of the impact of deformation on IN selectivity was obtained by calculating the rotational orientation of each base pair along the MCs with respect to their curvature. We measured the angle φ ($0^\circ < \varphi < 360^\circ$) between the major groove dyad and the vector connecting the helical axis to the center of the MC, in the base-pair plane, at each position along the sequence (see 'Materials and Methods' section). Values of φ around 0° correspond to the major groove facing toward the inside of the MC (in), while values around 180° correspond to the major groove facing outside (out). We plotted the CAI values at each integration site as a function of the φ angle measured at the center of the corresponding IN binding site, which is located 2.5 bp away from the integration site in the 3′ direction of the same strand. These 'rose plot' representations are shown in Figure 5C and D for integration sites on both strands along the 'target integration' segment of the four constructs. For both HIV-1 IN–LEDGF/p75 and HIV-1 IN, they confirm the non-uniform distribution of the MC-favored sites along the DNA helix axis. Vectors with a positive CAI value (going out of the black circle, Figure 5C and D) are indeed significantly enriched in the top half of the dial, which corresponds to the outer surface of the circles. On average, integration by the HIV-1 IN–LEDGF/p75 complex on the outside surface of MCs is increased 4- to 7-fold, with respect to the inside (Figure 5C). This preference is less pronounced in the absence of LEDGF/p75, with an average increase of 2- to 4-fold (Figure 5D).

Therefore, combined molecular modeling of MCs and quantitative analysis of MC enhanced integration sites show that DNA circularization clearly reorients HIV-1 integration sites toward the outer surface of MCs. This result supports the role of DNA curvature as a major parameter of IN selectivity in nucleosomes and excludes simple steric effects due to the presence of histones. Results obtained with IN alone or in complex with LEDGF/p75 also show that this IN cofactor does not change the distribution of integration sites enhanced in MCs, but rather increases the amplitude of this redistribution. Therefore, in addition to its histone-dependent role, LEDGF/p75 can modulate HIV-1 IN selectivity in curved DNA, in the absence of histones, implying an additional histone-independent role for this IN cofactor.

MCs reveal differences between HIV-1 and PFV integrases

Similarly to HIV-1 IN, PFV IN preferentially integrates into flexible and highly curved DNA as well as in nucleosomes (24,31,32). However, both the distance between concerted integration sites (31) and the curvature angle of target DNA in the intasome differ between these two enzymes (21,32). We wondered if DNA MCs would reveal new differences between the catalytic properties of these two enzymes. We performed integration assays in 75 and 86 linear Fts and MCs constructs, with PFV IN. The results show that this IN also integrates more efficiently into the MCs (Figure 6A), with a MC/Ft enhancement, which is larger than that measured using HIV-1 IN, either alone or in complex with LEDGF/p75 (Figure 3). Furthermore, PFV IN is more reactive on the 75 bp MC (compared to the 86-bp MC), which corresponds more closely to the structure of DNA in the nucleosome.

We also compared PFV IN U5 integration sites between linear Fts and MCs of the four constructs (Figure 6B). Surprisingly, CAI values do not show the sinusoidal profile observed for HIV-1 IN or the HIV-1 IN–LEDGF/p75 complex, as confirmed by auto-correlation analysis (Supplementary Figure S5, lower panels). Finally, as shown in (Figure 6C and D), PFV IN shows only a very slight enhancement of integration efficiency on the outer surface (the increase is on average between 25 and 100% with respect to the inner surface). We remark that for PFV IN, φ angles are measured differently with respect to HIV IN, taking the distance between the integration site and the center of the IN binding site to be 2 bp. Altogether, these analyses of PFV integration sites show that the large enhancement of PFV integration in DNA MCs is a global effect of DNA circularization and is not associated with a preference of the enzyme for one surface of the curved DNA helix. We will return to an analysis of this result in the following 'Discussion' section.

DISCUSSION

Retroviral integration targets nucleosomes (9,22,25,34), but the respective roles of DNA and histones in this selectivity are still a matter of debate. In infected cells, HIV-1 integrates preferentially into flexible DNA sequences (30,31,72) that can adapt their conformation to the DNA trajectory in the intasome (25). *In vitro*, specific target DNA sequences can reproduce the effects of DNA curvature, major groove opening and flexibility on IN selectivity (26,27). However, these studies could not distinguish between the roles of nucleotide sequence and structure induced by this sequence. On the 'protein side', histones can play a passive or an active role in selectivity of integration. Passively, they can induce a DNA deformation recognized by IN and this effect can be reproduced with other DNA binding proteins (70). Actively, histones can be directly bound by INs as shown by H2B–PFV IN interaction (24) or by a IN cofactor as shown by the H3K36me3–LEDGF/p75 interaction (16–18).

A substrate mimicking nucleosome-induced DNA curvature, but lacking the histones, is therefore required to distinguish between the roles of DNA and histones in retroviral integration. We chose 75 and 86 bp MCs as these substrates

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

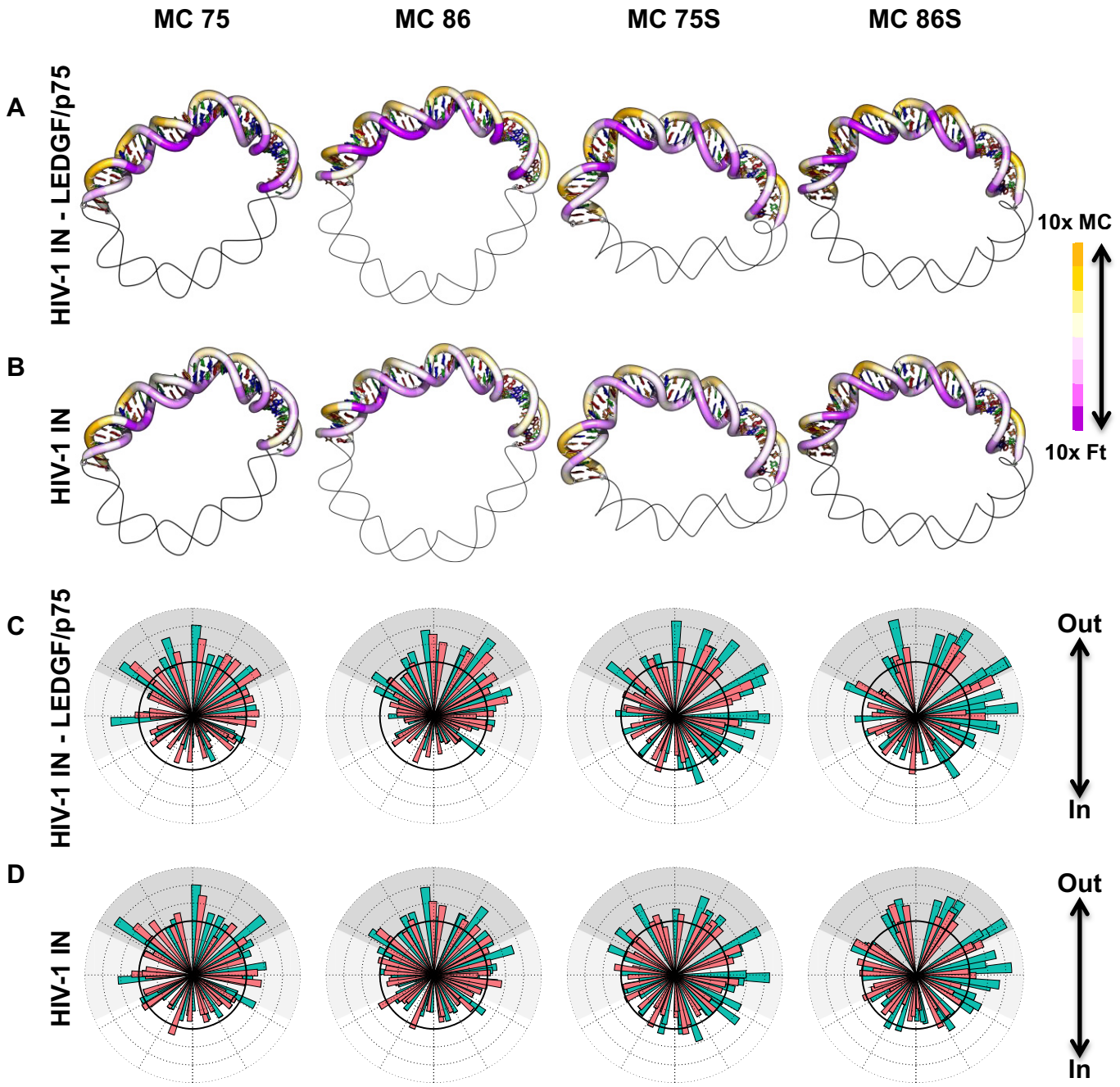

Figure 5. Distribution of MC-regulated HIV-1 integration sites along the MC structures. (**A** and **B**) CAI values obtained with the HIV-1 IN–LEDGF/p75 complex (A) or HIV-1 IN (B) are plotted on the 3D structures of the MC75, MC75S, MC86 and MC86S using a color gradient code where yellow/purple reflects enhanced integration in the MCs/Fts. Only values pertaining to the ITSs are shown: bases are shown as sticks, colored according to the base type (see caption to Figure 2). (**C** and **D**) Rose-plot representation of MC enhanced integration efficiencies on the upper strand (green bars) and the lower strand (red bars). CAI values obtained with the HIV-1 IN–LEDGF/p75 complex (C) or HIV-1 IN (D), for each base-pair step along the ITSs are plotted radially as a function of the φ angle of the center of the corresponding IN binding site. Values above the thick dark circle indicate MC-enhanced integration. The top half of the dial (shaded dark gray) corresponds to the major groove facing toward the outside of the MC (Out), while the bottom half of the dial corresponds to the major groove facing inside (In). Lightly shaded areas at the boundaries of the In and Out regions represent the uncertainty in the in/out definition, measured as the average fluctuation of the φ angle measured during a MD simulation of MC75 (see main text and 'Materials and Methods' section for details). Molecular graphics and rose plots were produced using Chimera and Matplotlib v1.5.0., respectively.

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

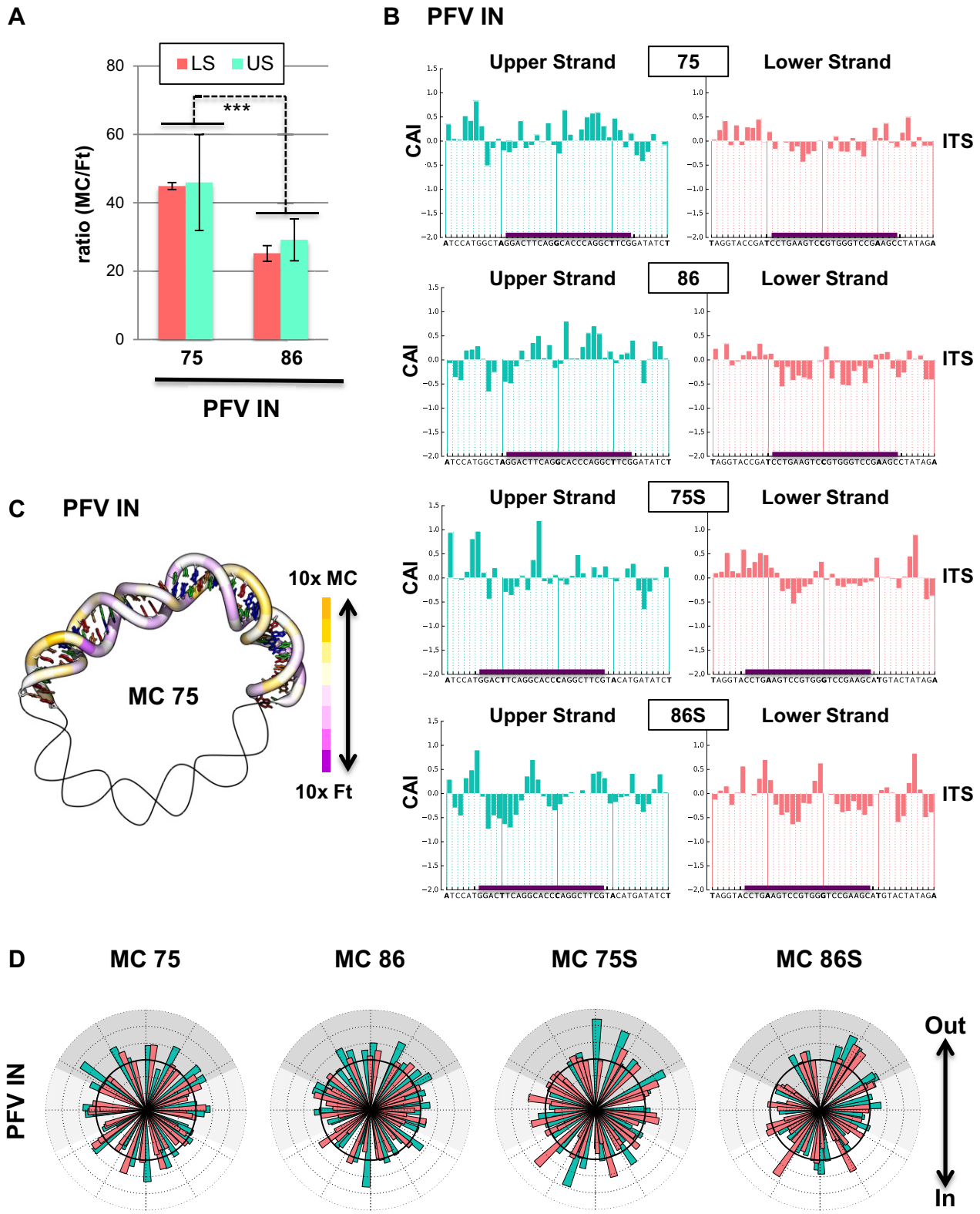

Figure 6. Effect of DNA circularization on the efficiency and selectivity of PFV integration. (**A**) MC/Ft ratio of integration efficiency by PFV IN, in the LS and US of the 75 and 86 constructs (error bars correspond to 95% confidence intervals of the calculated ratios and *** corresponds to *P*-values < 0.001). (**B** and **C**) CAI values obtained with PFV IN were plotted along the ITSs of 75, 86, 75S and 86S constructs (B) (similarly as in Figure 4B and C) or on the 3D-modeled structures of MC75 (C) (using a similar color gradient code as in Figure 5A). (**D**) Rose-plot representation of MC enhanced PFV integration efficiencies (CAI values) on the upper strand (green bars) and the lower strand (red bars) of the four selected constructs. This representation is similar as in Figure 5B.

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

and developed a procedure to use them in integration studies. We introduced phased A-tracts in one part of the MCs in order to freeze the rotational register of DNA curvature and then performed PCR amplification to quantify and map the IPs in the remaining part of the MC, termed the ITS (Integration Target Sequence, see also Supplementary Figure S1). By varying the length of the sequence separating the phased A-tracts and this ITS, we were able to specifically study the effect of two different DNA orientations with respect to the direction of curvature, while maintaining the same base sequence. Finally, MC structures were modeled at atomic resolution, allowing us to compare the impact of DNA structure on integration efficiency. Our PCR strategy does not allow us to distinguish between half- and full-site IPs, but the chosen experimental conditions (25) and the PAGE analysis of IPs (data not shown) indicate that most of our conclusions correspond to the half-site integration process. Our results show that MCs offer a unique system to study the structural determinants of retroviral integration selectivity.

MCs effectively mimic nucleosomal DNA

Experimental and molecular modeling results agree in demonstrating that our sequence design was successful in producing MCs that effectively mimic nucleosomal DNA. The 75 and 86 bp MCs, which were chosen to reproduce DNA curvatures, respectively, similar and slightly larger than those present in nucleosomes, exhibit roughly planar structures, with curvature distributed rather uniformly along their length. The absence of sharp kinks (i.e. involving base unstacking) and base-pair opening is confirmed by the MCs' resistance to *in vitro* Bal31 and S1 nuclease digestion (Supplementary Figure S2), performed as previously described (36).

Our modeling results also confirm that the phased A-tracts are sufficient to constrain the rotational register of the MCs to a narrow range of values, which also contributes to restraining fluctuations in curvature. These results, taken together, indicate that our MC constructs are homogeneous and stable in solution under the chosen experimental conditions, allowing us to establish detailed relations between their structure and the observed integration efficiencies. Using phased A-tracts in MC construction has also the double advantage of allowing total freedom in the design of the ITS as well as providing accurate comparisons with linear Fts of the same sequence, where the effect of A-tracts on the ITS is expected to be very small. Finally, it is worth noting that our design strategy, as well as being used to study the effects of the orientation of the ITS with respect to the direction of curvature (by varying its position relative to the phased A-tracts), could also be adapted to study the effects of DNA supercoiling, by producing MCs with non-zero superhelical density.

HIV-1 integration is favored in MCs

In the present study, we observed an enhanced HIV-1 integration into MCs with respect to Fts. This enhancement is more pronounced with the HIV-1 IN–LEDGF/p75 complex than with IN alone. Similarly, *in vitro* studies have shown an enhanced HIV-1 integration in mono- and polynucleosomes templates and this enhancement was increased in the presence of LEDGF/p75 (19,25,26). These results suggest that curved DNA is indeed a better substrate for HIV-1 IN even in the absence of histones. The LEDGF/p75 effect is more surprising since this IN cofactor is known to interact with histones and this interaction is involved in LEDGF-dependent activation of integration into chromatin (19,20). A possible explanation is that the DNA binding domains of LEDGF/p75, such as its A/T hook, could directly interact with curved DNA and stabilize the IN–MC complex. Alternatively, LEDGF/p75 stabilizes the HIV-1 IN tetramer, which could become more sensitive to curved DNA. Indeed, the target DNA trajectory in the HIV-1 IN–LEDGF/p75 intasome structure (21), fits with the DNA trajectory in a nucleosome, or in our MC constructs (see Figure 7). Further investigations using LEDGF/p75 mutants are required to clarify the histone-independent impact of LEDGF/p75.

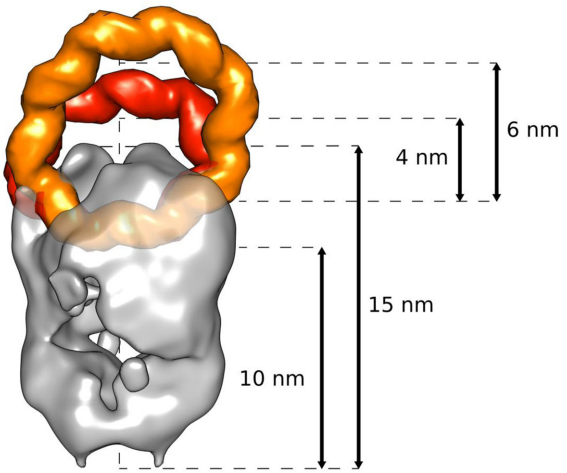

Figure 7. Low-resolution 3D model of the HIV-1 IN–LEFGF/p75 complex bound to the 75 and 86 bp MCs. The 14 Å-resolution cryo-EM density map of the HIV-1 IN–LEFGF/p75 complex (21) is shown as a semi-transparent gray isodensity surface; densities at the same resolution obtained from the modeled structures of MC75 and MC86 are shown as red and orange isodensity surfaces, respectively. Measurements on the right of the picture are approximate, and provided to guide the comparison of relative sizes. The image was produced using Chimera (61).

Enhanced integration in MCs is robust with respect to mutations that affect HIV-1 IN selectivity toward sequences with different curvature propensities and groove orientations (72). *In vitro,* in our experimental conditions, these mutants (S119G, S119T, R231G and R231G/S119T) integrate more efficiently into MCs, in a similar way to the WT enzyme (Figure 3C). This suggests that these residues are involved in recognition processes that are not relevant for integration into MCs. On one hand, in the case of the R231G mutant, the changed selectivity observed *in vivo* probably results from the loss of a histone–IN direct interaction (V. Parissi, personal communication). On the other hand, we can speculate that the DNA trajectory in MCs, although similar to the one in the nucleosome, is not optimal for the insertion of the IN S119 side chains into the minor grooves surrounding the central duplicated sequence, as proposed

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

in structural models of HIV-1 intasome. Alternatively, these mutations could affect the distribution of integration sites in the MCs, without changing the global enhancement of integration.

MCs modify the distribution of HIV-1 integration sites and reveal the structural parameters of IN selectivity

We observe a major redistribution of integration sites upon circularization of the target DNA substrate for HIV-1 IN alone or in complex with LEDGF/p75 (Figure 4A). This redistribution is confirmed by correlations measured between libraries of integration sites and correlative clusters formed between them (Supplementary Figure S5). Three major conclusions can be made from these correlative studies. First, the high correlation values measured between libraries of integration sites identified in identical target DNA sequence (75/86 or 75S/86S) validate our protocol for mapping integration sites. Second, clusters formed between libraries of sites identified in similar conformations of target DNA (linear or circular) highlight the important role of curvature in integration selectivity. Finally, the high correlations measured between HIV-1 IN and HIV-1 IN–LEDGF/p75 integration site libraries, contrast with the low correlations measured between HIV-1 and PFV integration site libraries. This result confirms the different selectivities of HIV-1 and PFV IN at the local DNA level (4).

To quantify the role of the target DNA conformation on IN selectivity, we calculated CAI values, which measure the variation of integration frequencies induced by circularization, on a logarithmic scale, at each position along the target sequences. Profiles of CAI, along the ITSs, on both DNA strands (Figure 4B and C) show a periodicity of roughly 11 bp (Supplementary Figure S6, upper and middle panels), for both HIV-1 IN alone and in complex with LEDGF/p75, although the magnitude of the variation is less pronounced in the absence of this cofactor. It is interesting to note that, on both the upper and lower strands, the periodicity of the CAI profiles is not affected by the sequence (compare 75 and 86 with their shifted versions, 75S and 86S). This suggests that the effects of circularization depend mainly on the phasing with the A-tracts, and therefore on the structure of the MCs, rather than the sequence itself.

Our modeling studies have confirmed that the phased A-tracts are effective at restraining the rotational register of the MCs. We can therefore relate integration efficiency results to the structural features of the MCs. Our results show that the major parameter for the redistribution of HIV-1 integration sites from linear DNA to MCs, is the orientation of the integration site with respect to the inner or outer surfaces of the MCs (Figure 5). In particular, integration sites enhanced in MCs are significantly more likely to occur on the outer surface of these MCs (Figure 5C and D), especially in the presence of LEDGF/p75. These values confirm that although LEDGF/p75 does not change the global distribution of integration sites along the circular DNA, it does increase the amplitude between favored and disfavored integration sites, even in the absence of histones.

It is tempting to interpret these results with simple steric considerations. The HIV-1 IN may be too large to correctly bind to major grooves facing inside the MCs, and this effect could be enhanced by the presence of LEDGF/p75 (our modeling suggests this may indeed be the case, see Figure 7). However, this steric interpretation does not explain why the enzymatic complex preferentially integrates at the major grooves facing outside the MCs rather than on the sides of the MCs, which are still accessible. The deformation of the major groove at the center of the IN binding site is another parameter involved in its selectivity. Major grooves facing outward, i.e. forming the outside surface of the MC, are deformed by curvature in a way that likely fits better the IN binding site. Interestingly, autocorrelations measured between MC-enhanced integration sites identified on complementary strands (Supplementary Figure S6A and B) show a broad peak centered around 6 or even 7 bp, which is in good agreement with the typical width of a DNA major groove. We note that this distance is unrelated to the 5 bp length characteristic of the central sequence duplicated by HIV-1 IN in full-site integration reactions, since our observations consist chiefly of half-site reactions.

Finally, LEDGF/p75 may have a steric or allosteric role on the redistribution of integration sites. On one hand, LEDGF/p75 may increase the volume of the integration complex and reduce its accessibility to the inner surface of the MCs. On the other hand, we speculate that LEDGF/p75 may stabilize a specific IN conformation, linked with its oligomerization state, that is more sensitive to DNA curvature or major groove width. Structural studies of an HIV-1 IN–LEDGF/p75–MC complex are required to test these proposed interpretations.

MCs reveal differences between PFV and HIV integration in curved DNA templates

Circularization of target DNA affects differently PFV and HIV-1 INs, at least at two levels. First, in terms of efficiency, PFV IN is more reactive to DNA circularization than HIV-1 IN (Figure 6A), and shows a much more pronounced sensitivity to the size of the MCs, with a preference for higher curvature (smaller radius). Second, in terms of selectivity, CAI profiles for PFV IN show very little periodicity (Figure 6B and Supplementary Figure S6B) and in fact the direction of curvature seems to have very little effect on its activity, in stark contrast with the results on HIV-1 IN (Figure 6B–D).

The target DNA is highly curved in both the intasomes formed by HIV-1 IN–LEDGF/p75 and PFV IN (21,32), but this curvature is higher with PFV IN. This DNA deformation is consistent with a shorter central duplicated sequence (4 bp for PFV instead of 5 bp for HIV), as well as with higher enrichment of flexible dinucleotides within this sequence (31). Recently, the cryo-EM structure of a PFV intasome–nucleosome complex has revealed that the target DNA is slightly extruded from the usual nucleosomal trajectory (24). This result suggests that PFV IN is able to adapt the target DNA structure to a conformation that is optimal for integration, even when DNA is confined by interactions with histones. We speculate that the observed lack of effect of curvature on PFV IN selectivity may be due to a similar ability to locally adapt the target DNA curvature within an MC, even in the presence of phased A-tracts. In contrast, the energetic cost to modify the curvature of DNA in our MCs is apparently too high for HIV-1

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

IN, which thus preferentially integrates on the outside surface of MCs.

DNA curvature induced by circularization can affect differently the efficiency and selectivity of retroviral INs, as observed with HIV-1 and PFV INs. Curved DNA can globally stimulate the strand-transfer activity and it can target this activity to regions optimal for integration. These two effects are not exclusive and could be explained by a scan and lock mechanism. In this mechanism, an initial IN interaction with target DNA is followed by a scan along this target until an optimal sequence is found for integration catalysis. In the case of PFV IN, the curved DNA conformation in the MC would favor the initial interaction or the scan of the target DNA, but it would not affect the choice of integration sites. In the case of HIV-1 IN, a curved target DNA would have a lower effect on the initial steps (interaction and/or scan), but it would largely favor the interaction or the catalytic activity at specific sites along the curved DNA molecule.

SUMMARY AND PERSPECTIVES

DNA circularization increases the efficiency of retroviral integration and, in the case of HIV-1, it modifies its selectivity toward the outer surface of a curved DNA. These effects are similar to those observed with nucleosomal DNA, although here they are obtained using MCs and in absence of histones. Therefore, MCs allow the role of target DNA structure on retroviral integration to be specifically studied. MCs also reveal a histone-independent role for LEDGF/p75 and unexpected differences in the integration mechanisms of two retroviral INs. In the future, MCs can be used to test various retroviral INs, characterized by different oligomeric states, lengths of the central duplicated sequence or propensities to interact with nucleosomes. Their use can also be extended to other DNA binding proteins or enzymes characterized by an indirect (structure- or mechanically-sensitive) readout of the DNA helix.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank M. Naughtin, F. Di Nunzio, P. Charneau and Céline Ralec for scientific discussions and L. Ma and C. Calmels for technical help. High-throughput sequencing has been performed on the Genomics Platform, member of 'France Génomique' consortium. We acknowledge GENCI for computer time on the CINES supercomputer OCCIGEN.

FUNDING

ANRS [AO2014-02 to V.P., M.L.] and [AO2012-02 to M.R. and J.B.]; SIDACTION [AI25-1-02332 to M.P., A.J., M.R., R.L., M.L.] and to M.R. and J.B.; ANR, Genomics Platform, member of 'France Génomique' consortium [ANR10-INBS-09-08 to C.B.]; French Infrastructure for Integrated Structural Biology [FRISBI, ANR-10-INSB-05-01 to M.R., J.B.]; INSTRUCT, part of the European Strategy Forum on Research Infrastructure (ESFRI) supported by national members subscription (to M.R., J.B.). Funding for open access charge: ANRS [A02014-02].
Conflict of interest statement. None declared.

REFERENCES

1. Engelman,A. and Cherepanov,P. (2014) Retroviral integrase structure and DNA recombination mechanism. *Microbiol. Spectr.*, **2**, 1–22.
2. Di Santo,R. (2014) Inhibiting the HIV integration process: past, present, and the future. *J. Med. Chem.*, **57**, 539–566.
3. Demeulemeester,J., De Rijck,J., Gijsbers,R. and Debyser,Z. (2015) Retroviral integration: site matters: mechanisms and consequences of retroviral integration site selection. *Bioessays*, **37**, 1202–1214.
4. Serrao,E. and Engelman,A.N. (2015) Sites of retroviral DNA integration: from basic research to clinical applications. *Crit. Rev. Biochem. Mol. Biol.*, **51**, 26–42.
5. Lelek,M., Casartelli,N., Pellin,D., Rizzi,E., Souque,P., Severgnini,M., Di Serio,C., Fricke,T., Diaz-Griffero,F., Zimmer,C. *et al.* (2015) Chromatin organization at the nuclear pore favours HIV replication. *Nat. Commun.*, **6**, 6483.
6. 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.
7. Cattoglio,C., Pellin,D., Rizzi,E., Maruggi,G., Corti,G., Miselli,F., Sartori,D., Guffanti,A., Di Serio,C., Ambrosi,A. *et al.* (2010) High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors. *Blood*, **116**, 5507–5517.
8. 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.
9. Wang,G.P., Ciuffi,A., Leipzig,J., Berry,C.C. and Bushman,F.D. (2007) HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.*, **17**, 1186–1194.
10. Singh,P.K., Plumb,M.R., Ferris,A.L., Iben,J.R., Wu,X., Fadel,H.J., Luke,B.T., Esnault,C., Poeschla,E.M., Hughes,S.H. *et al.* (2015) LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. *Genes Dev.*, **29**, 2287–2297.
11. 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.
12. Marshall,H.M., Ronen,K., Berry,C., Llano,M., Sutherland,H., Saenz,D., Bickmore,W., Poeschla,E. and Bushman,F.D. (2007) Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *PLoS One*, **2**, e1340.
13. Shun,M.-C., Ragahvendra,N.K., Vandergraaf,N., Daigle,J.E., Hughes,S., Kellam,P., Cherepanov,P. and Engelman,A. (2007) LEDGF/p75 functions downstream of preintegration complex formation to effect gene-specifc HIV-1 integration. *Genes Dev.*, **21**, 1767–1778.
14. Rasheedi,S., Shun,M.C., Serrao,E., Sowd,G.A., Qian,J., Hao,C., Dasgupta,T., Engelman,A.N. and Skowronski,J. (2016) The Cleavage and polyadenylation specific factor 6 (CPSF6) subunit of the capsid-recruited pre-messenger RNA Cleavage factor I (CFIm) complex mediates HIV-1 integration into genes. *J. Biol. Chem.*, **291**, 11809–11819.
15. 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.
16. Daugaard,M., Baude,A., Fugger,K., Povlsen,L.K., Beck,H., Sorensen,C.S., Petersen,N.H., Sorensen,P.H., Lukas,C., Bartek,J. *et al.* (2012) LEDGF (p75) promotes DNA-end resection and homologous recombination. *Nat. Struct. Mol. Biol.*, **19**, 803–810.
17. 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.

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

18. 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.
19. Botbol,Y., Raghavendra,N.K., Rahman,S., Engelman,A. and Lavigne,M. (2008) Chromatinized templates reveal the requirement for the LEDGF/p75 PWWP domain during HIV-1 integration in vitro. *Nucleic Acids Res.*, **36**, 1237–1246.
20. Shun,M.C., Botbol,Y., Li,X., Di Nunzio,F., Daigle,J.E., Yan,N., Lieberman,J., Lavigne,M. and Engelman,A. (2008) Identification and characterization of PWWP domain residues critical for LEDGF/p75 chromatin-binding and human immunodeficiency virus type 1 infectivity. *J. Virol.*, **82**, 11555–11567.
21. Michel,F., Crucifix,C., Granger,F., Eiler,S., Mouscadet,J.F., Korolev,S., Agapkina,J., Ziganshin,R., Gottikh,M., Nazabal,A. *et al.* (2009) Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor. *EMBO J.*, **28**, 980–991.
22. 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.
23. 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.
24. 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.
25. 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.
26. Pruss,D., Bushman,F.D. and Wolffe,A.P. (1994) Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 5913–5917.
27. Pryciak,P.M. and Varmus,H.E. (1992) Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell*, **69**, 769–780.
28. Taganov,K.D., Cuesta,I., Daniel,R., Cirillo,L.A., Katz,R.A., Zaret,K.S. and Skalka,A.M. (2004) Integrase-specific enhancement and suppression of retroviral DNA integration by compacted chromatin structure in vitro. *J. Virol.*, **78**, 5848–5855.
29. Wu,X., Li,Y., Crise,B., Burgess,S.M. and Munroe,D.J. (2005) Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J. Virol.*, **79**, 5211–5214.
30. 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.
31. Serrao,E., Ballandras-Colas,A., Cherepanov,P., Maertens,G.N. and Engelman,A.N. (2015) Key determinants of target DNA recognition by retroviral intasomes. *Retrovirology*, **12**, 39.
32. Maertens,G.N., Hare,S. and Cherepanov,P. (2010) The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature*, **468**, 326–329.
33. Yin,Z., Shi,K., Banerjee,S., Pandey,K.K., Bera,S., Grandgenett,D.P. and Aihara,H. (2016) Crystal structure of the Rous sarcoma virus intasome. *Nature*, **530**, 362–366.
34. Roth,S.L., Malani,N. and Bushman,F.D. (2011) Gammaretroviral integration into nucleosomal target DNA in vivo. *J. Virol.*, **85**, 7393–7401.
35. Bates,A.D., Noy,A., Piperakis,M.M., Harris,S.A. and Maxwell,A. (2013) Small DNA circles as probes of DNA topology. *Biochem. Soc. Trans.*, **41**, 565–570.
36. Du,Q., Kotlyar,A. and Vologodskii,A. (2008) Kinking the double helix by bending deformation. *Nucleic Acids Res.*, **36**, 1120–1128.
37. Lionberger,T.A. and Meyhofer,E. (2010) Bending the rules of transcriptional repression: tightly looped DNA directly represses T7 RNA polymerase. *Biophys. J.*, **99**, 1139–1148.
38. Demurtas,D., Amzallag,A., Rawdon,E.J., Maddocks,J.H., Dubochet,J. and Stasiak,A. (2009) Bending modes of DNA directly addressed by cryo-electron microscopy of DNA minicircles. *Nucleic Acids Res.*, **37**, 2882–2893.
39. Lionberger,T.A., Demurtas,D., Witz,G., Dorier,J., Lillian,T., Meyhofer,E. and Stasiak,A. (2011) Cooperative kinking at distant sites in mechanically stressed DNA. *Nucleic Acids Res.*, **39**, 9820–9832.
40. Harris,S.A., Laughton,C.A. and Liverpool,T.B. (2008) Mapping the phase diagram of the writhe of DNA nanocircles using atomistic molecular dynamics simulations. *Nucleic Acids Res.*, **36**, 21–29.
41. Lankas,F., Lavery,R. and Maddocks,J.H. (2006) Kinking occurs during molecular dynamics simulations of small DNA minicircles. *Structure*, **14**, 1527–1534.
42. Mitchell,J.S., Laughton,C.A. and Harris,S.A. (2011) Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled DNA. *Nucleic Acids Res.*, **39**, 3928–3938.
43. Sanghani,S.R., Zakrzewska,K., Harvey,S.C. and Lavery,R. (1996) Molecular modelling of (A4T4NN)n and (T4A4NN)n: sequence elements responsible for curvature. *Nucleic Acids Res.*, **24**, 1632–1637.
44. Dyer,P.N., Edayathumangalam,R.S., White,C.L., Bao,Y., Chakravarthy,S., Muthurajan,U.M. and Luger,K. (2004) Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.*, **375**, 23–44.
45. Leh,H., Brodin,P., Bischerour,J., Deprez,E., Tauc,P., Brochon,J.C., LeCam,E., Coulaud,D., Auclair,C. and Mouscadet,J.F. (2000) Determinants of Mg2+-dependent activities of recombinant human immunodeficiency virus type 1 integrase. *Biochemistry*, **39**, 9285–9294.
46. Valkov,E., Gupta,S.S., Hare,S., Helander,A., Roversi,P., McClure,M. and Cherepanov,P. (2009) Functional and structural characterization of the integrase from the prototype foamy virus. *Nucleic Acids Res.*, **37**, 243–255.
47. Lavery,R., Zakrzewska,K. and Sklenar,H. (1995) JUMNA (Junction Minimization of Nucleic Acids). *Comput. Phys. Commun.*, **91**, 135–158.
48. Cheatham,T.E. 3rd, Cieplak,P. and Kollman,P.A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.*, **16**, 845–862.
49. Pérez,A., Marchand,I., Svozil,D., Sponer,J., Cheatham,T.E. 3rd, Laughton,C. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.*, **92**, 3817–3829.
50. Berendsen,H.J.C., Grigera,J.R. and Stratsma,T.P. (1987) The missing term in effective pair potentials. *J. Phys. Chem.*, **91**, 6269–6271.
51. Dang,L.X. (1995) Mechanism and thermodynamics of ion selectivity in aqueous-solutions of 18-crown-6 ether—a molecular dynamics study. *J. Am. Chem. Soc.*, **117**, 6954–6960.
52. Berendsen,H.J.C., Van der Spoel,D. and Van Druner,R. (1995) GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.*, **91**, 43–56.
53. Lindahl,E., Hess,B. and Van der Spoel,D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.*, **7**, 306–317.
54. Van der Spoel,D., Lindahl,E., Hess,B., Groenhof,G., Mark,A.E. and Berendsen,H.J.C. (2005) GROMACS: fast, flexible, and free. *J. Comput. Chem.*, **26**, 1701–1718.
55. Bussi,G., Donaldio,D. and Parinello,M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.*, **126**, 014101.
56. Berendsen,H.J.C., Postma,J.P.M., Van Gunsteren,W.F., DINola,A. and Haak,J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
57. Essmann,U., Perera,L., Berkowitz,M.L., Darden,T., Lee,H. and Pedersen,L.G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.*, **103**, 8577–8593.
58. Hess,B., Bekker,H., Berendsen,H.J.C. and Fraaije,J.G. (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.*, **18**, 1463–1472.
59. Hess,B., Kutzner,C., Van der Spoel,D. and Lindahl,E. (2008) Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, **4**, 435–447.
60. Pasi,M., Maddocks,J.H., Beveridge,D., Bishop,T.C., Case,D.A., Cheatham,T. III, Dans,P.D., Jayaram,B., Lankas,F., Laughton,C. *et al.* (2014) muABC: a systematic microsecond molecular dynamics

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016

study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.*, **42**, 12272–12283.
61. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
62. Lavery,R., Moakher,M., Maddocks,J.H., Petkeviciute,D. and Zakrzewska,K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.*, **37**, 5917–5929.
63. Blanchet,C., Pasi,M., Zakrzewska,K. and Lavery,R. (2011) CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res.*, **39**, W68–W73.
64. Lavery,R., Maddocks,J.H., Pasi,M. and Zakrzewska,K. (2014) Analyzing ion distributions around DNA. *Nucleic Acids Res.*, **42**, 8138–8149.
65. Hunter,J.D. (2007) Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.*, **9**, 90–95.
66. Luger,K., Mader,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution [see comments]. *Nature*, **389**, 251–260.
67. Schnitzler,G., Sif,S. and Kingston,R.E. (1998) Human SWI/SNF interconverts a nucleosome between its base state and a stable remodeled state. *Cell*, **94**, 17–27.
68. Curuksu,J., Zacharias,M., Lavery,R. and Zakrzewska,K. (2009) Local and global effects of strong DNA bending induced during molecular dynamics simulations. *Nucleic Acids Res.*, **37**, 3766–3773.
69. Faure,A., Calmels,C., Desjobert,C., Castroviejo,M., Caumont-Sarcos,A., Tarrago-Litvak,L., Litvak,S. and Parissi,V. (2005) HIV-1 integrase crosslinked oligomers are active in vitro. *Nucleic Acids Res.*, **33**, 977–986.
70. Muller,H.P. and Varmus,H.E. (1994) DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *EMBO J.*, **13**, 4704–4714.
71. Pruss,D., Reeves,R., Bushman,F.D., Wolffe,A.P., Carteau,S., Hoffmann,C. and Bushman,F. (1994) The influence of DNA and nucleosome structure on integration events directed by HIV integrase. *J. Biol. Chem.*, **269**, 25031–25041.
72. Demeulemeester,J., Vets,S., Schrijvers,R., Madlala,P., De Maeyer,M., De Rijck,J., Ndung'u,T., Debyser,Z. and Gijsbers,R. (2014) HIV-1 integrase variants retarget viral integration and are associated with disease progression in a chronic infection cohort. *Cell Host Microbe*, **16**, 651–662.

Downloaded from http://nar.oxfordjournals.org/ at Institut Pasteur MediathÃ¨que Scientifique on December 14, 2016