The EMBO Journal Vol.16 No.24 pp.7511–7520, 1997

A large nucleoprotein assembly at the ends of the viral DNA mediates retroviral DNA integration

Shui-Qing Wei, Kiyoshi Mizuuchi and Robert Craigie[1]

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

[1]Corresponding author
e-mail: bobc@helix.nih.gov

We have probed the nucleoprotein organization of Moloney murine leukemia virus (MLV) pre-integration complexes using a novel footprinting technique that utilizes a simplified *in vitro* phage Mu transposition system. We find that several hundred base pairs at each end of the viral DNA are organized in a large nucleoprotein complex, which we call the intasome. This structure is not formed when pre-integration complexes are made by infecting cells with integrase-minus virus, demonstrating a requirement for integrase. In contrast, footprinting of internal regions of the viral DNA did not reveal significant differences between pre-integration complexes with and without integrase. Treatment with high salt disrupts the intasome in parallel with loss of intermolecular integration activity. We show that a cellular factor is required for reconstitution of the intasome. Finally, we demonstrate that DNA–protein interactions involving extensive regions at the ends of the viral DNA are functionally important for retroviral DNA integration activity. Current *in vitro* integration systems utilizing purified integrase lack the full fidelity of the *in vivo* reaction. Our results indicate that both host factors and long viral DNA substrates may be required to reconstitute an *in vitro* system with all the hallmarks of DNA integration *in vivo*.

Keywords: DNA transposition/footprinting/PCR/pre-integration complexes/retrovirus

Introduction

Retroviral virions are nucleoprotein assemblies that include many of the enzymes that are essential for subsequent steps in the viral replication cycle (Katz and Skalka, 1994). After entry into the host cell, reverse transcription of the viral RNA genome occurs within a higher order nucleoprotein complex derived from the core of the infecting virion. The resulting double-stranded viral DNA forms part of a large nucleoprotein structure, termed the pre-integration complex. This complex contains the information that is necessary for nuclear localization and the enzymatic machinery required to insert the viral DNA into the host genome. Although the protein composition and architecture of these complexes is not well understood, much progress has been made regarding the viral enzymes that catalyze the key steps in the viral replication cycle. First, protease cleaves the polyprotein precursor to generate the mature viral proteins. Reverse transcriptase then synthesizes a double-stranded DNA copy of the viral genome and integrase catalyzes the key DNA cutting and joining reactions that insert the viral DNA into the host genome.

Pre-integration complexes, isolated after infection of sensitive cells with either Moloney murine leukemia virus (MLV) (Brown *et al.*, 1987; Bowerman *et al.*, 1989) or human immunodeficiency virus type 1 (HIV-1) (Ellison *et al.*, 1990; Farnet and Haseltine, 1990), efficiently integrate their DNA into an exogenously added target DNA *in vitro*. Analysis of MLV DNA integration intermediates made *in vitro* has defined the DNA cutting and joining steps of retroviral DNA integration (Fujiwara and Mizuuchi, 1988; Brown *et al.*, 1989). First, two nucleotides are cut from each 3′ end of the initially blunt-ended viral DNA. The resulting recessed 3′ ends of the viral DNA are then joined to the 5′ ends of the target DNA at the site of integration, leaving the 5′ ends of the viral DNA and the 3′ ends of the target DNA unjoined. Cellular DNA repair enzymes are probably responsible for degrading the unpaired nucleotides at the 5′ ends of the viral DNA, filling in the single strand gaps and subsequent ligation to complete the integration process; it is also possible that ligation may be carried out by integrase itself (Roe *et al.*, 1997).

MLV and HIV pre-integration complexes preferentially integrate their DNA into an intermolecular target *in vitro*, whereas the corresponding avian retrovirus complexes mostly 'autointegrate' into their own DNA (Lee and Coffin, 1990). In the case of MLV complexes, high salt treatment removes a cellular factor, and the resulting stripped complexes then predominantly autointegrate (Lee and Craigie, 1994).

Purified integrase of several retroviruses, including MLV (Craigie *et al.*, 1990), HIV (Sherman and Fyfe, 1990; Engelman *et al.*, 1991) and Rous sarcoma virus (RSV) (Katzman *et al.*, 1989; Katz *et al.*, 1990), carries out both the 3′ processing and DNA strand transfer reactions with simple DNA substrates that mimic one of the ends of the viral DNA. These simplified reaction systems have proved invaluable for gaining detailed biochemical insights into the reaction mechanism. Both 3′ processing and DNA strand transfer occur by a one-step transesterification mechanism (Engelman *et al.*, 1991). These reactions are therefore DNA phosphoryltransfer reactions of the type used by a class of mobile genetic elements that includes phage Mu (Mizuuchi, 1992a,b, 1997). Other important insights are revealed by the functional and structural domain structure of retroviral integrases (reviewed in Andrake and Skalka, 1996; Rice *et al.*, 1996). A triad of acidic residues that are critical for catalysis lies within the central core of integrase, and the

crystal structure of this domain has been solved for both HIV-1 integrase (Dyda *et al.*, 1994) and RSV integrase (Bujacz *et al.*, 1995). Both the N- and C-terminal domains of HIV integrase are also required for 3′ processing and DNA strand transfer activity, but the roles of these domains are less well defined. The C-terminal domain binds DNA non-specifically (Vink *et al.*, 1993; Woerner and Marcus-Sekura, 1993; Engelman *et al.*, 1994), and the structure of this domain has been solved by NMR (Eijkelenboom *et al.*, 1995; Lodi *et al.*, 1995). The N-terminal domain is a zinc-binding domain (Burke *et al.*, 1992; Zheng *et al.*, 1996; Lee *et al.*, 1997) whose function is not yet understood. This domain has a very similar fold to the helix–turn–helix motif of a number of DNA-binding proteins (Cai *et al.*, 1997).

In contrast to the considerable progress made in understanding the structure and function of integrase protein, our knowledge of pre-integration complexes is still rudimentary; this is in part due to the limited quantity of material that can be readily obtained. Pre-integration complexes efficiently integrate their DNA *in vitro* with all the hallmarks of the authentic *in vivo* reaction, whereas current *in vitro* integration systems using purified integrase are deficient in some aspects. For example, the authentic reaction integrates a pair of viral DNA ends, one on each strand of the target DNA, with a spacing that is characteristic of the particular retrovirus. In contrast, the majority of products with purified integrase are insertions of single viral DNA ends into one strand of target DNA. Biochemical analysis of pre-integration complexes can be expected to reveal the components and assembly steps that are required for the fully authentic integration process.

We have probed the nucleoprotein organization of MLV pre-integration complexes using a novel footprinting technique that exploits a simplified *in vitro* phage Mu transposition system. We find that the DNA integration reaction is mediated by a large nucleoprotein structure that involves integrase, host factors and several hundred base pairs at each end of the viral DNA.

Results

Footprinting by Mu-mediated PCR

The limited quantity of pre-integration complexes that can be isolated from cells necessitated the development of a novel footprinting technique to study the interaction between DNA and proteins in the complexes. Established techniques, including ligation-mediated PCR, were either insufficiently sensitive or did not give a reproducible signal above background. We utilized the Mu transposase (Mu A protein) to insert oligonucleotides corresponding to the ends of Mu DNA into the DNA to be footprinted (Figure 1A). The Mu insertion reaction simultaneously cuts the target DNA and covalently attaches an oligonucleotide to the 5′ end at the site of insertion. In the case of ligation-mediated PCR (LM-PCR) (Mueller and Wold, 1989), the equivalent steps are carried out by a nuclease and a DNA ligase. With Mu-mediated PCR (MM-PCR) the cutting and ‘ligation’ steps are tightly coupled, and this probably accounts for the significantly higher signal to noise ratio of this technique in our hands. Subsequent to the Mu reaction, the distribution of insertions in the target DNA is analyzed in a similar manner to LM-PCR, using one primer specific to the region of interest and a second primer specific to the Mu end. The PCR products are resolved by gel electrophoresis and visualized by autoradiography. We first tested the technique by carrying out control experiments with Mu repressor and Mu operator DNA (data not shown). The observed footpint was similar to that previously seen with DNase I footprinting, indicating that the resolution of the two techniques is not significantly different.

Since the long terminal repeat (LTR) sequence forms a direct repeat at each end of the viral DNA, PCR using primers within the LTR region would result in the superimposition of two footprints, e.g. the footprint at the U3 sequence at the left end of the viral DNA would be superimposed on that of the internal U3 sequence at the right end (Figure 1B). To overcome this problem, we carried out two rounds of PCR. For the first round of PCR, a primer outside of the LTR region was used together with the Mu end primer, SW0; primers SW1 and SW0 were used to amplify the left end LTR DNA and primers SW4 and SW0 were used to amplify the right end. The product of the first PCR, which is specific for Mu insertions into either the left or right end LTR DNA, was then used as the template for the second round of PCR with primers specific to the U3 or U5 region.

Integrase is required to form a higher order nucleoprotein structure at the ends of the viral DNA in the pre-integration complex

Purified integrase protein binds DNA non-specifically and exhibits little or no preference for viral DNA end sequence. Since between 50 and 100 copies of integrase are estimated to be associated with each virion, integrase could potentially play a structural role in the pre-integration complex in addition to catalyzing the integration reaction. To address this question, we made pre-integration complexes by infecting cells with either wild type MLV or an integrase-minus mutant of MLV. The mutant virions were produced from a cell line carrying an MLV provirus with a deletion in the integrase gene (dl5401); the truncated integrase that would be expressed from the mutant gene is not detected in virions (Tanese *et al.*, 1986), presumably because the protein is unstable. The mutant virions are normal for cell entry and reverse transcription, but the resulting viral DNA does not integrate (Schwartzberg *et al.*, 1984). The footprint patterns on both viral end DNA sequences and on internal regions of the viral DNA were then compared.

Figure 2A and B show the footprints on the U3 and U5 ends of the viral DNA respectively. Both protection and enhancements are seen at each end of the viral DNA in the complexes made by infection with integrase plus virus (compare lanes 1 and 2 of each panel). The protection extends ~500 bp from the ends of the viral DNA, although the border of the protected region is not sharply defined. A particularly strong enhancement is located ~20 nucleotides from each end (Figure 2A and B, lane 2, panel c); in the case of the U3 end, an additional strong enhancement is seen at ~100 nucleotides from this end of the viral DNA. This pattern of protection and enhancement was not observed when the complexes were made by infection with integrase-minus virus, although some differences were apparent compared with the naked DNA control

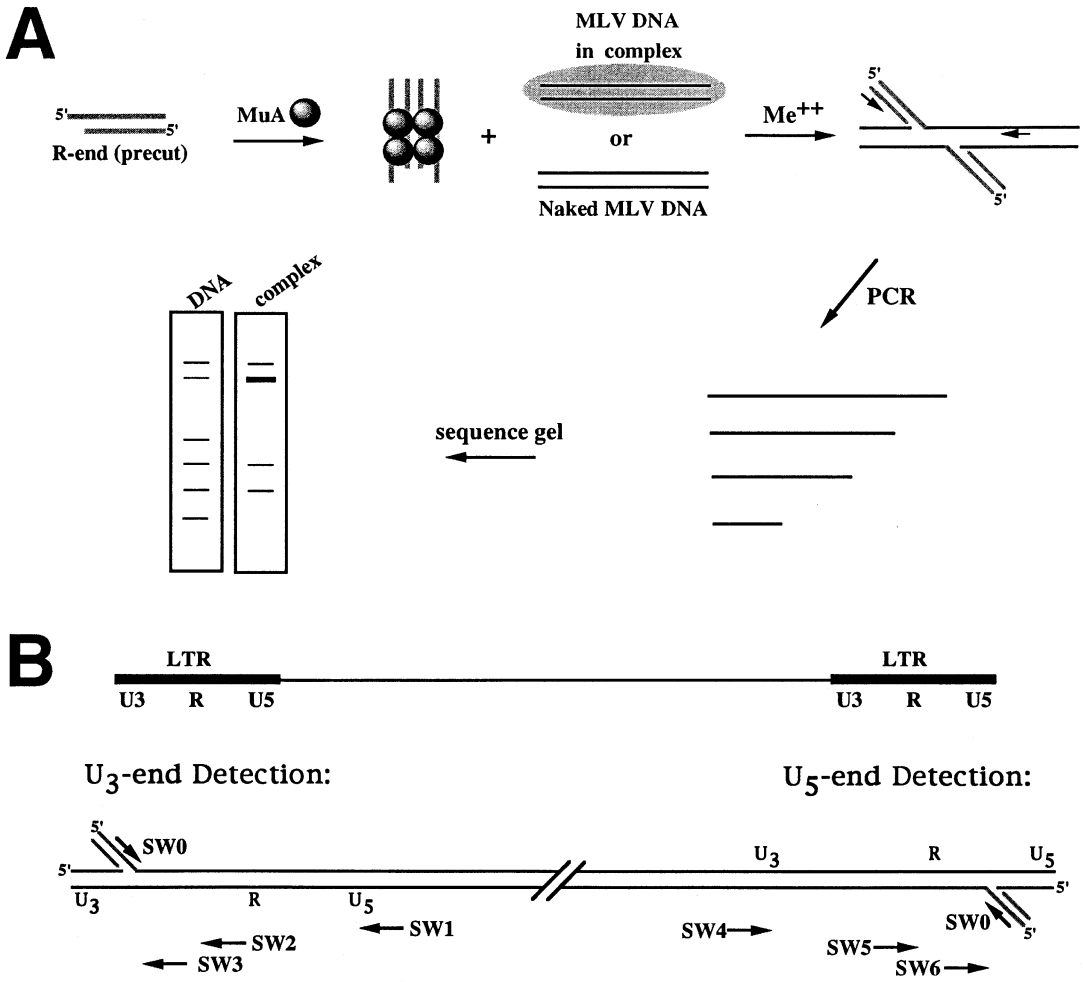

Fig. 1. (**A**) Footprinting of MLV pre-integration complexes by Mu-mediated PCR. Mu A protein is mixed with pre-cut Mu R-end DNA in the absence of divalent metal ions. Upon incubation with MLV pre-integration complexes and a divalent metal ion, the Mu ends are inserted into the viral DNA by Mu A protein. The distribution of insertion sites is determined by PCR using one primer specific for the target DNA region of interest and a second primer that is specific for the Mu ends. The products are analyzed by gel electrophoresis. The distribution of insertions is compared with a parallel reaction using naked MLV DNA as a control. (**B**) PCR primers for probing the U3 or U5 ends of the viral DNA. The U3, R and U5 sequences that comprise the LTR are present at both ends of the viral DNA. To amplify the U3 end specifically, the PCR reaction included the SW1 primer and the Mu end primer, SW0. A second round of PCR was then carried out using a primer within the LTR sequence (e.g. SW2 or SW3) and the Mu end primer. The U5 end was probed in the same way except that the first round of PCR included the SW4 primer, and the SW5 or SW6 primers were used in the second PCR.

(compare lanes 1 and 3 of each panel). We conclude that integrase is required for formation of a higher order nucleoprotein structure that involves several hundred base pairs at each end of the viral DNA. In contrast, no major differences were observed between integrase-plus and integrase-minus complexes in the distribution of Mu insertions in internal regions of the viral DNA, although some local differences were noticed. No regions with protection or enhancements comparable with those seen at the ends of the viral DNA were observed in the *gag* gene (Figure 2C, lanes 2 and 3), or in several other internal regions probed (data not shown). However, differences in the footprints of both complexes compared with the naked viral DNA (Figure 2C, lane 1) suggest that other proteins that interact with the internal regions of the viral DNA may interfere with Mu insertion.

While the strong enhancements and overall footprint at the ends of the viral DNA are highly reproducible between experiments, we note that some variability is observed in the precise pattern of Mu insertion at the nucleotide level. This may in part result from variability in sample preparations that is amplified by the PCR methodology. Therefore, we focus on the overall pattern of protection and enhancements, and interpret minor differences in the pattern of Mu insertion with caution.

To investigate further the effect of integrase on the overall structure of the pre-integration complexes, we compared the sedimentation profiles in sucrose gradients. The profiles of the pre-integration complexes made by infection with integrase-plus or integrase-minus MLV do not differ significantly (Figure 3), demonstrating that the viral DNA within the integrase-minus complex is still associated with protein factors. However, since velocity sedimentation can only probe the global structure of the complex, it is possible that there are significant differences in the organization of complexes with and without integrase. The footprinting and sedimentation data together suggest that integrase is associated principally with the viral DNA ends in the pre-integration complex and is not required for the complex to behave as a large nucleoprotein structure.

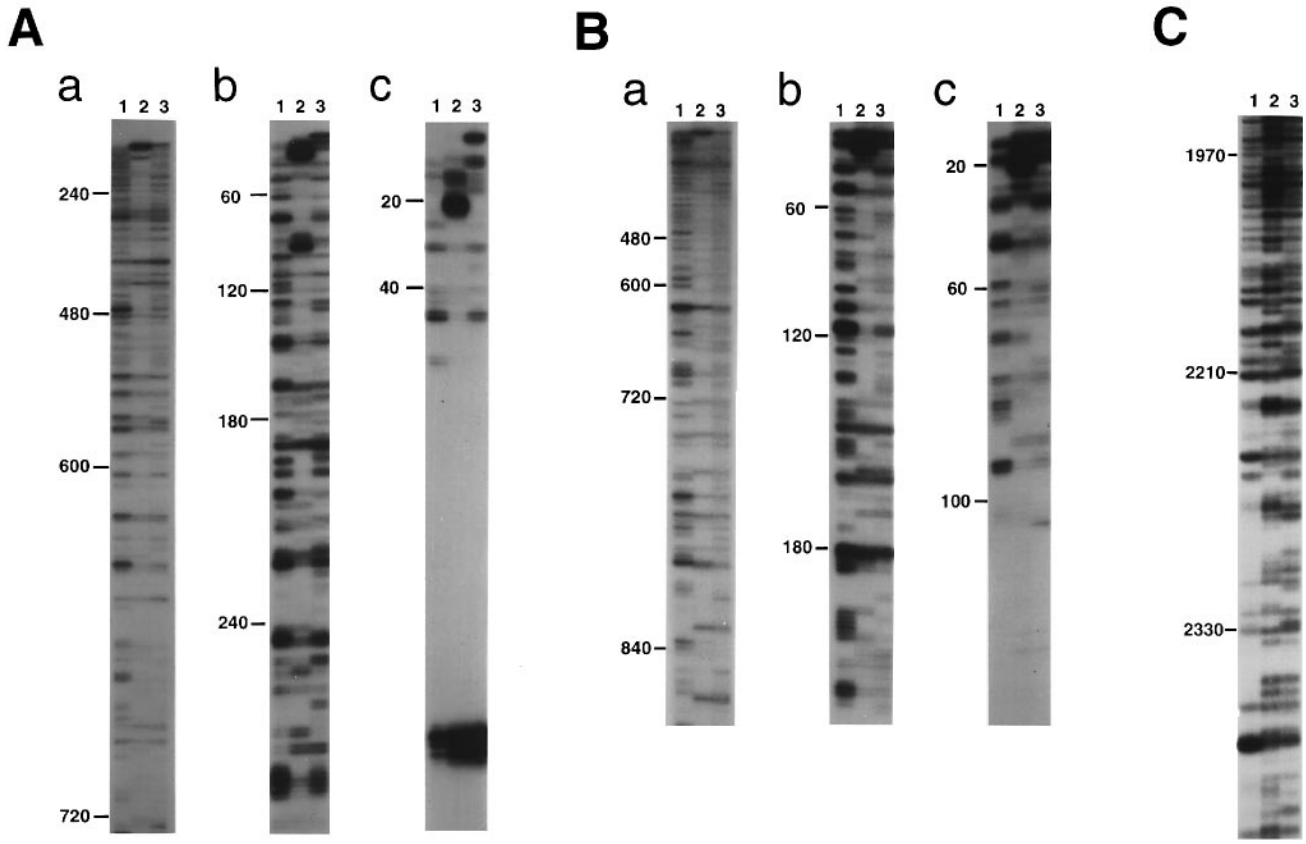

Fig. 2. Footprints on the U3 and U5 ends of the viral DNA. (**A**) Footprint on the U3 end. Panels a, b and c are PCR reactions carried out with end-labeled primers SW1, SW2 and SW3 respectively. (**B**) Footprint on the U5 end. Panels a, b and c are PCR reactions carried out with end-labeled primers SW4, SW5 and SW6 respectively. (**C**) Footprinting on an internal region of the viral DNA. Numbers on the left of each panel indicate the distance from the very end of the viral DNA. Samples are: lane 1, naked viral DNA; lane 2, MLV pre-integration complexes; lane 3, integrase-minus MLV pre-integration complexes. Prominent enhancements are located at ~20 and 100 bp from the U3 end (A) and at ~20 bp from the U5 end (B).

The strong enhancements and regions of protection at each end of the viral DNA indicate a specific association of protein factors with the viral DNA termini. We shall refer to this complex as the intasome to distinguish it from the entire pre-integration complex. Although the protein composition of the intasome other than integrase remains to be determined, the enhancements appear to be a hallmark of the functional complex (see below).

Kinetics of intasome assembly

In order to analyze the kinetics of intasome formation, we isolated MLV pre-integration complexes at various times after semi-synchronous infection of NIH-3T3 cells. Footprinting of complexes isolated 4.5 h after infection did not reveal significant protection at the ends of the viral DNA, and the intensity of the enhancement was considerably less than at the later time (18–19 h) after infection; the distribution of Mu insertions resembled that observed with a naked DNA target (Figure 4, compare lane 3 with lanes 1 and 2). By 7 h post-infection, the characteristic enhancements were observed at both the U3 (Figure 4, lane 4) and U5 (data not shown) ends of the viral DNA. We conclude that assembly of the intasome, following synthesis of the viral DNA by reverse transcription, is a relatively slow step.

High salt treatment disrupts the intasome

MLV pre-integration complexes efficiently integrate their DNA into an intermolecular target DNA *in vitro*. After

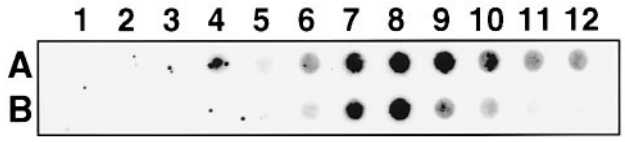

Fig. 3. Velocity sedimentation profile of wild-type (**A**) and integrase-minus (**B**) pre-integration complexes in a 15–30% sucrose gradient. Aliquots of gradient fractions numbered 1 (top) to 12 (bottom) were transferred to a membrane and hybridized with a probe specific for the viral DNA.

treatment with high salt, these complexes lose their ability to integrate their DNA efficiently into an intermolecular target and instead integrate into their own DNA (Lee and Craigie, 1994), a reaction termed autointegration. Salt-stripped complexes can restore their protection against autointegration, and preference for intermolecular integration, upon incubation with an extract of NIH-3T3 cells (Lee and Craigie, 1994). High salt treatment therefore strips a cellular factor from the complex that prevents self-destructive autointegration. One way this host factor might work is to physically block access to the viral DNA. We thus examined the accessibility of the viral DNA to Mu insertion before and after high salt treatment. Although some differences were observed in the footprint on internal DNA regions between the untreated complexes and those incubated with up to 750 mM KCl (Figure 5A, compare lanes 2 and 5), they were not comparable with the drastic

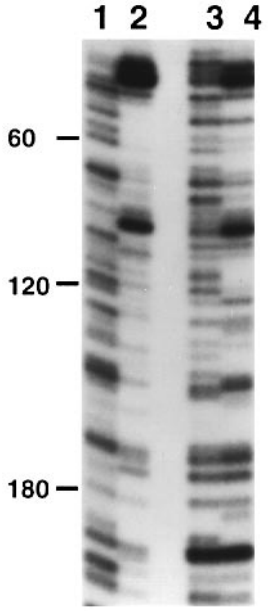

Fig. 4. Kinetics of the appearance of enhancements and footprints at the U3 end of the viral DNA. NIH-3T3 cells were semi-synchronously infected with MLV by incubation with medium from a culture of MLV-producing clone 4 cells. After replacing the clone 4 supernatant with fresh virus-free medium, cytoplasmic extracts were made at various times post-infection. The pre-integration complexes within each extract were partially purified by centrifugation in a Nycodenz density gradient and then subjected to Mu-mediated PCR footprinting. Samples are: lane 1, naked viral DNA; lane 2, MLV pre-integration complexes; lanes 3 and 4, MLV pre-integration complexes isolated 4.5 and 7 h post-infection.

changes in the footprint pattern at both the U3 and U5 ends of the viral DNA (Figure 5B, compare lanes 2 and 5). Both the protection and enhancements diminished with increasing concentrations of salt treatment, and with 750 mM KCl the pattern of Mu insertion was very similar to that observed with naked DNA. Furthermore, the disappearance of the protection and enhancements correlated with the conversion from intermolecular integration to autointegration in the activity assay (Figure 5C). Although the footprints disappear after high salt treatment, some integrase must remain associated with the pre-integration complex because the complex is competent for integrase-dependent autointegration; moreover, reconstitution of a functional intasome, after separating the pre-integration complex from free proteins, does not require addition of integrase (Lee and Craigie, 1994).

Reconstitution of the functional intasome

Disappearance of the enhancements and protection upon titrating KCl paralleled a decrease in intermolecular integration activity and an increase in autointegration. We therefore speculated that loss of the host factor previously shown to prevent autointegration might be involved in elimination of the enhancements and protection. Since the bias for intermolecular integration can be reconstituted by addition of host cell extract to the salt-stripped complex, we asked whether the characteristic footprinting pattern of the functional intasome is also restored in the reconstitution reaction. Pre-integration complexes were treated with high salt and then separated from free proteins. Reconstitution reactions were then performed with either an extract of MLV particles or a cytoplasmic extract of NIH-3T3 cells. The enhancement was restored by incubation with an extract of NIH-3T3 cells, but not with an extract of disrupted MLV particles (Figure 6, lanes 4 and 5). Thus, a host factor is required for both reconstitution of the enhancements and intermolecular integration activity (Lee and Craigie, 1994). We note that although the enhancement is restored to near its original level, the protection was only partially restored. This may be due to less than complete reconstitution among the population of pre-integration complexes. Addition of host extract to salt-stripped integrase minus complexes did not affect the pattern of Mu insertion (data not shown), indicating that the footprint requires the presence of both the host factor and integrase.

Functional analysis of the extent of terminal viral DNA sequence required for intermolecular integration

Although footprints extend several hundred base pairs from each end of the viral DNA in the pre-integration complex, the experiments described above do not establish unambiguously the functional relevance of the entire protected region to the integration reaction. For example, the functional unit for integration could consist of integrase associated with the very ends of the viral DNA, and the more extensive footprint could reflect protein–DNA interactions that are not essential for integration; in addition, both active and inactive complexes may contribute to the observed footprint. In an attempt to address these issues, we carried out the retroviral integration reaction after the Mu reaction was first used to insert Mu DNA sequences in the viral DNA within pre-integration complexes. The pre-integration complexes containing Mu insertions were then used as the viral DNA substrate for a retroviral integration reaction with a biotinylated target DNA. The biotin tag on the target DNA was used to separate reacted (integration-competent) and unreacted complexes. The distribution of Mu insertions in the reacted complexes was then compared with their distribution in the unreacted complexes (Figure 7A). Complexes that integrated into the target DNA showed a striking absence of Mu insertions within ~300 bp from the ends of the viral DNA compared with the complexes that remained unintegrated (Figure 7B, lanes 3 and 4 of each panel). Weaker interference of Mu insertion extended for an additional 200 bp from the ends of the viral DNA. These results indicate that Mu insertion into the viral DNA sequence within several hundred base pairs of the ends of the viral DNA is either blocked in active complexes, or that Mu insertion into these regions interferes with the viral DNA integration reaction. The DNA–protein interactions involving these extensive regions at the ends of the viral DNA thus appear to be functionally important for intermolecular DNA integration.

Discussion

The ends of the viral DNA are organized in a large nucleoprotein assembly

We have shown that retroviral DNA integration is mediated by a large nucleoprotein complex that involves several

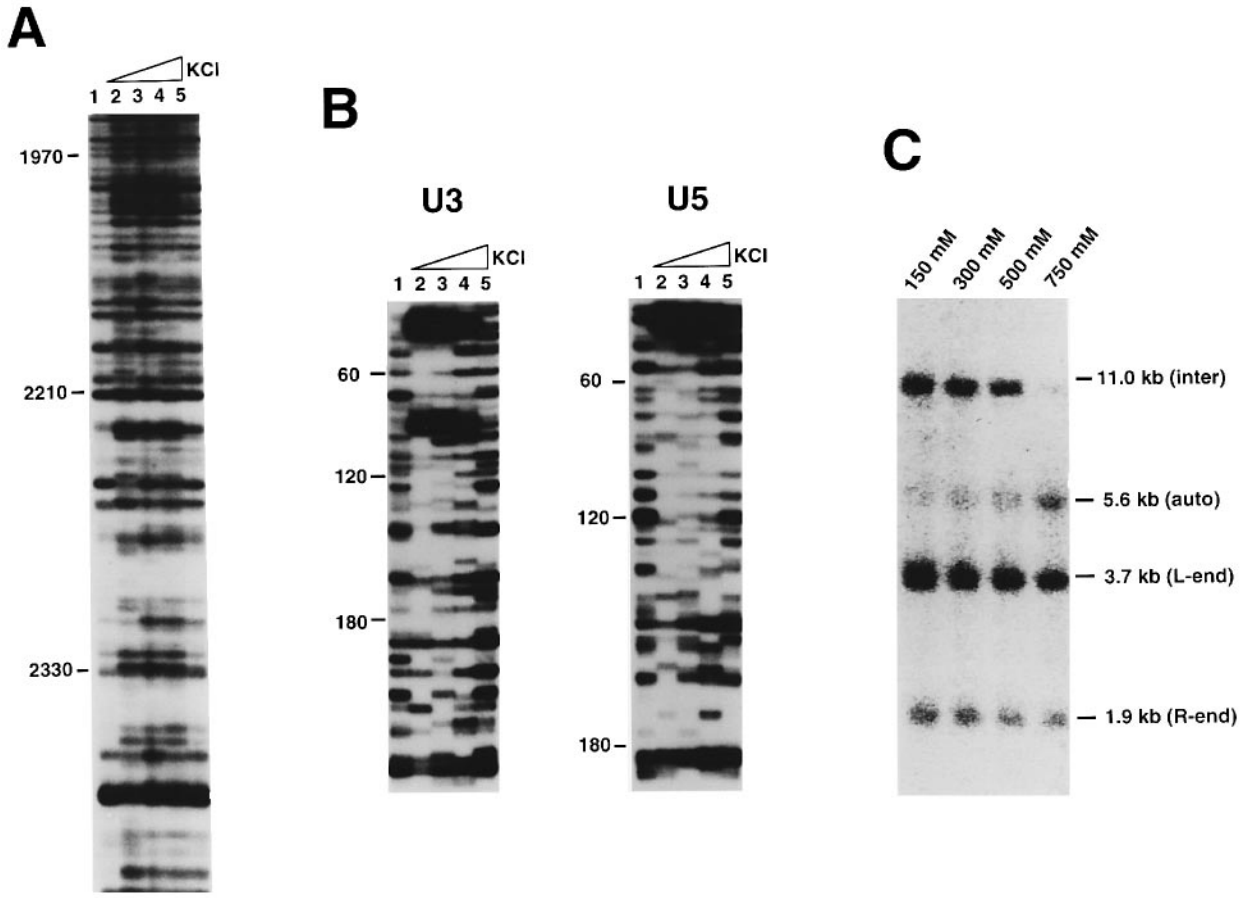

Fig. 5. (**A**) Effect of high salt treatment on the footprint on an internal region of the viral DNA. Cytoplasmic extract containing MLV pre-integration complexes was incubated in the presence of 150 (lane 2), 300 (lane 3), 500 (lane 4) or 750 mM (lane 5) KCl for 30 min on ice. The pre-integration complexes were then partially purified in a Nycodenz density gradient and an internal region of the viral DNA was footprinted by MM-PCR. Lane 1 shows a control with deproteinized MLV DNA. (**B**) The lanes are the same as in (A), except that primers were selected to footprint the U3 or U5 ends of the viral DNA. (**C**) *In vitro* integration activities of the KCl-treated pre-integration complexes after partial purification in a Nycodenz density gradient. Products of reactions with ϕX174 RF DNA as the target DNA were digested with *Bam*HI and electrophoresed in an agarose gel. The DNA was transferred to a membrane and probed for viral DNA. The 11 kb band results from intermolecular integration of the viral DNA into ϕX174 RF DNA. The 5.6 kb band and the smear below it result from autointegration of the viral DNA into itself. The 3.7 and 1.9 kb bands are derived from unreacted viral DNA.

hundred base pairs at each end of the viral DNA, integrase and other protein factors. The major differences between the footprints of the complexes with or without integrase were located at the ends of the viral DNA; the internal regions exhibited more minor differences in the pattern of Mu insertion. We infer that integrase is associated specifically with the ends of the viral DNA in the pre-integration complex. We observe strong hotspots for Mu insertion ~20 nucleotides away from each end of the viral DNA; presumably the viral DNA within the complex is distorted at this position in such a way as to make it a better target for Mu insertion. Significant protection also extends several hundred nucleotides from each end of the viral DNA. While the enhancements at the ends of the viral DNA were consistently strong, the footprint exhibited some variation between different preparations of pre-integration complexes; this may indicate that a subpopulation of the complexes is inactive. We do not observe any apparent periodicity in the pattern of Mu insertion as has been reported for DNase I protection experiments with purified avian integrase and LTR sequence-containing restriction fragments (Misra *et al.*, 1982).

Insertions into the ends of the viral DNA were strikingly absent in the intermolecular integration products when the complexes were first subjected to Mu insertion. The functional footprint should be interpreted somewhat qualitatively since the distribution of Mu insertions before the integration reaction was not random. Two distinct mechanisms may contribute to the footprint in this experiment. First, bound proteins within the pre-integration complex will diminish, or enhance, the insertion of Mu as with standard footprinting. Secondly, the insertion of Mu into some positions within the viral DNA may 'interfere' with the subsequent retroviral integration reaction. Therefore, both protection and interference contribute to the observed protection. However, the absence of the strong enhancements near the ends of the viral DNA must be due to Mu interference. The near absence of insertions in other regions near the end of the viral DNA (Figure. 7B), compared with the much lesser degree of protection observed with standard footprinting (Figure 2A), indicate that (i) Mu interference is a major component of the footprint shown in Figure 7B, and/or (ii) the more robust footprint may be due to exclusion of a subpopulation of inactive complexes. In either case, this functional analysis demonstrates the participation of several hundred base

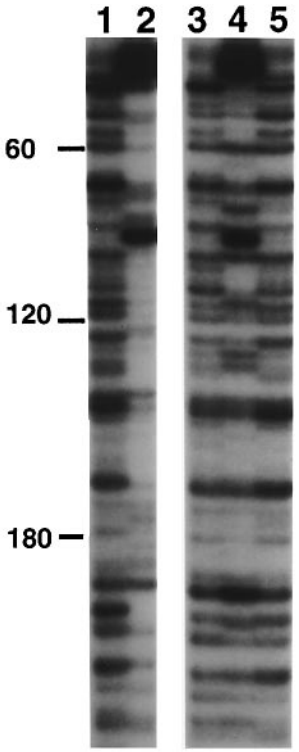

Fig. 6. Reconstitution of the footprint on the U3 end of the viral DNA. Cytoplasmic extract containing MLV pre-integration complexes was incubated with 750 mM KCl for 30 min on ice and the pre-integration complexes were then partially purified by centrifugation in a Nycodenz density gradient. Salt-stripped complexes were incubated on ice with various factors for 60 min, followed by MM-PCR. Lanes 1 and 2 are controls showing the patterns of Mu insertion into naked viral DNA and MLV DNA within pre-integration complexes. Reconstitution reactions included: lane 3, buffer only control; lane 4, cytoplasmic extract of NIH-3T3 cells; lane 5, extract of disrupted MLV virus particles.

pairs of terminal viral DNA sequence in the normal intermolecular integration reaction.

Further evidence for the relative inaccessibility of the viral DNA ends, compared with internal DNA regions, comes from experiments with restriction endonucleases (data not shown). Internal DNA sites are cleaved quite readily, whereas sites near the ends of the viral DNA are relatively refractory to cleavage.

Both integrase and host factors are required for intasome assembly

Both integrase and host factors may contribute to the footprint at the ends of the viral DNA. Although the characteristic protection and enhancements are dependent on integrase, we do not think the observed footprint is due simply to binding of integrase. In fact, the disappearance of the footprint after high salt treatment, and its restoration by a host factor, suggest that it does not only reflect the binding of integrase. At least some integrase must remain associated with the complexes after high salt treatment, since they are competent for autointegration; in addition, the functional complex for the intermolecular reaction can be reconstituted, after separation from free proteins, without addition of integrase. If the footprints are entirely due to integrase, high salt treatment must modify its interaction with the viral DNA so as to make it invisible to footprinting by MM-PCR.

What might be the identity of the required factor? Several host proteins have been suggested to be involved in retroviral DNA integration. HMG I(Y) recently has been implicated as being involved in HIV-1 pre-integration complexes (Farnet and Bushman, 1997), and some host proteins, such as HMG 1 (Aiyar *et al.*, 1996), have been reported to increase the efficiency of *in vitro* reactions by a factor of several-fold. In view of these observations, we tested both HMG I(Y) and HMG 1 for reconstitution of the footprint at the ends of the viral DNA in salt-stripped complexes. Neither of these factors restored the footprint (data not shown). A clue to the possible identity of the host factor comes from the abolition of the footprint in parallel with a conversion from complexes that carry out predominantly intermolecular integration to those that mainly autointegrate. This observation suggests that the factor required for reconstitution of the footprint may be identical to the previously described factor that protects retroviral DNA against autointegration (Lee and Craigie, 1994).

Implications for reconstituting a high-fidelity integration reaction from purified components

Although purified integrase proteins from several retroviruses have been shown to carry out 3′ end processing and DNA strand transfer reactions, reconstitution of an efficient integration system with all the hallmarks of the authentic reaction has proved elusive. In particular, many reaction products result from insertion of a single viral DNA end into one strand of target DNA, although relatively efficient two-ended integration has been observed with the avian myeloblastosis virus integrase (Vora *et al.*, 1994; Vora and Grandgenett, 1995). At the other extreme, the efficiency of two-ended integration events with HIV integrase is so low that they have only been detected using a highly sensitive genetic assay (Bushman *et al.*, 1990). In addition, protection of viral DNA substrate from autointegration (Lee and Craigie, 1994) has not been observed in reactions with purified integrase protein. Clearly, these simplified reaction systems are in some way deficient. Study of pre-integration complexes may provide valuable insights into the nature of these differences.

Integrase protein binds to DNA non-specifically, but it appears to be associated specifically with the viral DNA ends in the pre-integration complex. Furthermore, the association of the integrase molecules that carry out the chemistry of the integration reaction is highly stable, unlike simple binding of integrase to viral end DNA sequence, which can be disrupted by relatively low concentrations of KCl. Studies of closely related prokaryotic transposition systems (Mizuuchi, 1992b), such as phage Mu, suggest paradigms for the assembly of the retroviral intasome. Although Mu has specific binding sites for transposase at the ends of its DNA, simple binding of transposase to these sites is quite weak and easily reversible. Formation of a highly stable complex involving the Mu DNA ends and transposase normally requires additional factors: the host-encoded protein HU, negatively supercoiled DNA and the Mu enhancer sequence (IAS). These factors appear to act as ‘catalysts’ that promote stable complex formation by lowering an energy barrier on the reaction pathway. Our results indicate that host factors are involved in assembling a functional retroviral

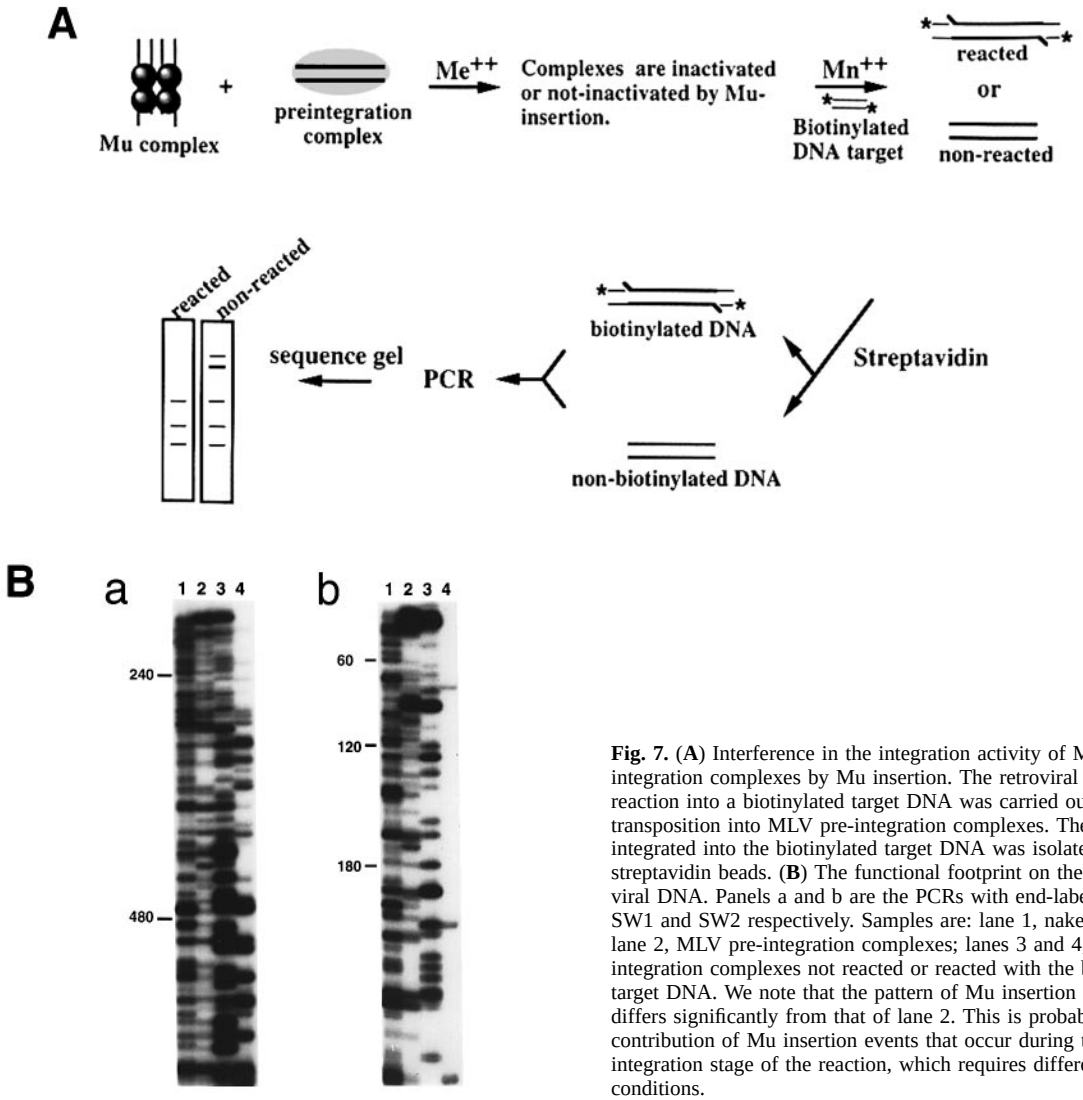

Fig. 7. (**A**) Interference in the integration activity of MLV pre-integration complexes by Mu insertion. The retroviral integration reaction into a biotinylated target DNA was carried out after Mu transposition into MLV pre-integration complexes. The viral DNA integrated into the biotinylated target DNA was isolated with streptavidin beads. (**B**) The functional footprint on the U3 end of the viral DNA. Panels a and b are the PCRs with end-labeled primers SW1 and SW2 respectively. Samples are: lane 1, naked viral DNA; lane 2, MLV pre-integration complexes; lanes 3 and 4, MLV pre-integration complexes not reacted or reacted with the biotinylated target DNA. We note that the pattern of Mu insertion in lanes 3 and 4 differs significantly from that of lane 2. This is probably due to a contribution of Mu insertion events that occur during the retroviral integration stage of the reaction, which requires different buffer conditions.

intasome, and it is tempting to speculate that similar mechanisms are involved. Missing host factors may account for the deficiencies of current *in vitro* retroviral integration systems. However, unlike the case of Mu, retroviral intasomes do not self-assemble from soluble components *in vivo*. Rather, the key components comprise part of a nucleoprotein structure derived from the core of the infecting virion. If structural transitions within pre-integration complexes play a key role in intasome assembly, the proper complex may not assemble readily by simply mixing the mature components.

MM-PCR as a tool to probe nucleoprotein structures

Mu has been used previously as a tool for *in vivo* footprinting of *Escherichia coli* chromosomal DNA (Wang and Higgins, 1994) and retroviral pre-integration complexes have been used as a probe for chromatin structure (Pryciak *et al.*, 1992a,b; Pruss *et al.*, 1994). Although not conceptually unique, the MM-PCR methodology we have developed for footprinting retroviral pre-integration complexes may be useful for probing other nucleoprotein complexes. The *in vitro* reaction system is highly efficient, and the substrates, Mu transposase and Mu end DNA, are easy to prepare. The high efficiency of insertion, together with the tight coupling of target DNA cleavage with Mu end insertion, result in a signal to noise ratio that is considerably better than LM-PCR in our hands; this may be especially true when only small quantities of material are available for footprinting.

Materials and methods

DNA

All oligonucleotides were gel purified prior to use. Mu R-end DNA was generated by annealing oligos MM429 (5′-GTT TTC GCA TTT ATC GTG AAA CGC TTT CGC GTT TTT CGT GCG CCG CTT CA) and MM568 (5′-GAT CTG AAG CGG CGC ACG AAA AAC GCG AAA GCG TTT CAC GAT AAA TGC GAA AAC). SW0 (5′-GCA TTT ATC GTG AAA CGC TTT CGC G) is Mu R-end specific primer. For U_3 end detection, SW1 (5′-TCC ACC ACG GGT CCG CCA GAT ACA G), SW2 (5′-ACT GCT GAG GGC TGG ACC GCA TCT GGG GA) and SW3 (5′-TCT CAG TTA TGT ATT TTT CCA TGC CTT GC) were used. For U_5 end detection, SW4 (5′-GCG GAC CAC ACA GGA CTA GTG AGA G), SW5 (5′-ATT TGA ACT AAC CAA TCA GTT CGC TTC TC) and SW6 (5′-TGA GTC GCC CGG GTA CCC GTG TAT CCA AT) were used. For the internal region (*gag* gene) detection, SW7 (5′-cgc ggg atc ctt aCA ATA GCT TGC TCA TCT CTC TAT G) was used. 5′ End-labeling was carried out with T4 polynucleotide kinase and [γ-^{32}P]ATP (NEN).

Reagents

Most enzymes used in this study were from New England BioLabs. MuA protein was purified as described (Baker *et al.*, 1993). Cells were grown in Dulbecco's modified Eagle's medium (Gibco) supplemented with 10% (v/v) calf serum (ICN). Nycodenz was purchased from Accurate Chemical & Scientific Corp.

MuA end complexes

Assembly of the complex between the Mu DNA ends and MuA protein was carried out essentially as described (Savilahti *et al.*, 1995) with 40 nM Mu R-end DNA fragment and 260 nM MuA protein in 25 µl of buffer containing 25 mM Tris–HCl pH 8.0, 15% (w/v) glycerol, 15% (v/v) dimethylsulfoxide (DMSO), 0.1% (w/v) Triton, 100 mM KCl and 100 µg/ml bovine serum albumin (BSA). The mixture was pre-incubated on ice for 10 min, followed by dilution (1:3) in the same buffer. Although pre-incubation on ice or at 30°C gave identical results in our experiments, we now pre-incubate at 30°C because the Mu complex assembly is more efficient at this temperature.

MoMLV pre-integration complexes

Standard preparation of MLV pre-integration complexes, after 18 h of co-culturing, was as previously described (Lee and Craigie, 1994) except for the time course shown in Figure 4 (see below). The complexes were purified further on a 10–50% (w/v) Nycodenz gradient made in the same buffer. After centrifugation at 50 000 r.p.m. in a Rotor TLS55 (Beckman) for 16 h, fractions (200 µl each) were collected starting from the top of the gradient. Fractions containing pre-integration complexes were identified by the presence of full-length viral DNA and *in vitro* integration activity.

Pre-integration complexes derived from mutant MLV were obtained by co-culture of NIH-3T3 cells and corresponding mutant MoMLV-producing cell lines. The producer cell line for MLV with a deletion in the integrase gene (dl5401) (Schwartzberg *et al.*, 1984) was kindly provided by Dr Alice Telesnitsky.

Mu transposition

Equal volumes (25 µl each) of diluted MuA complexes and MLV pre-integration complexes were mixed on ice, and $CaCl_2$ was added to final concentration of 10 mM. The reaction was carried out at 30°C for 30 min and stopped by addition of 3 vol of 0.05% SDS. DNA was precipitated after phenol and chloroform extraction and resuspended in 20 µl of water. For comparison, equivalent amounts of deproteinized viral DNA were substituted for pre-integration complexes in the MuA transposition reaction. Naked viral DNA was isolated from pre-integration complexes after proteinase K digestion, phenol extraction and ethanol precipitation.

Detection of sites of Mu insertion

Round 1 PCR (50 µl total volume) was carried out with 3 µl of resuspended DNA template from the transposition reaction in buffer containing 10 mM KCl, 20 mM Tris–HCl pH 8.8, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100, 100 µg/ml BSA, 200 µM dNTP, 0.4 µM of each primer and 2 U of Vent polymerase. The reaction mixture was subjected to 25 cycles on a thermal cycler (Perkin Elmer Cetus) pre-heated to 96°C.

Round 2 PCR (22 µl total volume) was carried out with 2 µl of reaction mixture from Round 1 PCR in the same reaction buffer except 0.25 µM of each primer, 0.15 µM of 5′ end-labeled primer (viral DNA specific) and 2 U of Vent (exo^-) polymerase. After 20 cycles of PCR, the reactions were stopped by adding an equal volume of sequencing gel loading buffer. Samples were electrophoresed in a 4% urea–polyacrylamide gel.

Velocity sedimentation with sucrose gradient

Sucrose gradients (15–30%) were made by layering 2.8 ml each of 30, 25, 20 and 15% of sucrose and equilibrated at 4°C overnight. Then 0.8 ml of samples were loaded onto the top of the gradient followed by centrifugation for 2 h at 35 000 r.p.m. in a SW40Ti rotor (Beckman). Fractions (1 ml each) were collected from the top of the gradient and 0.5 ml of each fraction was denatured and transferred to the Genescreen Plus membrane (Du Pont) with a Bio-Dot apparatus (Bio-Rad). The position of the pre-integration complexes in the gradient was monitored by the presence of the viral DNA using MLV DNA-specific probes.

High salt treatment and reconstitution of the intasome

Pre-integration complexes were mixed with KCl to a final concentration of 300, 500 or 750 mM. The mixtures were incubated on ice for 30 min and then applied to a 10–50% Nycodenz gradient as described above. Preparation of cytoplasmic extracts of NIH-3T3 cells and reconstitution reactions were carried out as previously described (Lee and Craigie, 1994) except that the KCl concentration was 240 mM. Approximately 10 µg of protein was added to each 50 µl reconstitution reaction.

Time course of appearance of the footprint at the ends of the viral DNA

Supernatant from ~10^7 clone 4 cells (MLV-producing cells) was filtered and layered onto a NIH-3T3 monolayer (~2×10^7 cells) in the presence of polybrene (8 µg/ml). After incubation at 37°C for 2 h, the supernatant was removed. The monolayer was washed once and then incubated with fresh medium. Cytoplasmic extracts containing pre-integration complexes were prepared at various time points and purified on the Nycodenz gradient prior to footprinting.

Functional analysis of the extent of terminal viral DNA sequence required for intermolecular integration

After a Mu transposition reaction with 75 µl of MLV pre-integration complexes, the retroviral integration reaction was carried out as previously described (Lee and Craigie, 1994) except that 1.5 µg of 5′-biotinylated DNA fragment (205 bp) was used as the target. Reactions were stopped by addition of 0.05% SDS. DNA was recovered after phenol and $CHCl_3$ extraction and ethanol precipitation. To separate biotinylated and non-biotinylated DNA, samples were resuspended in 0.3 ml of 0.5× SSC and incubated with Streptavidin MagneSphere Paramagnetic particles (Promega) at room temperature for 30 min. Particles were washed twice with buffer containing 0.5 M NaCl, 20 mM Tris–HCl pH 8.0, 1 mM EDTA and 0.05% Triton X-100, and twice with 0.1× SSC. Non-biotinylated DNA was recovered from the supernatant by ethanol precipitation. Biotinylated DNA was eluted from the particles with concentrated NH_4OH after heating to 60°C for 30 min. Biotinylated and non-biotinylated DNA were then used as templates for MM-PCR.

Acknowledgements

We thank Myung Soo Lee for helpful discussion during this study. We also thank Martin Gellert, Regis Krah, Kevin Hiom and Myung Soo Lee for comments on the manuscript. This work was supported in part by the National Institutes of Health Intramural AIDS Targeted Antiviral Program.

References

Aiyar,A., Hindmarsh,P., Skalka,A.M. and Leis,J. (1996) Concerted integration of linear retroviral DNA by the avian sarcoma virus integrase *in vitro*: dependence on both long terminal repeat termini. *J. Virol.*, **70**, 3571–3580.

Andrake,M.D. and Skalka,A.M. (1996) Retroviral integrase, putting the pieces together. *J. Biol. Chem.*, **271**, 19633–19636.

Baker,T.A., Mizuuchi,M., Savilahti,H. and Mizuuchi,K. (1993) Division of labor among monomers within the Mu transposase tetramer. *Cell*, 74, 723–733.

Bowerman,B., Brown,P.O., Bishop,J.M. and Varmus,H.E. (1989) A nucleoprotein complex mediates the integration of retroviral DNA. *Genes. Dev.*, **3**, 469–478.

Brown,P.O., Bowerman,B., Varmus,H.E. and Bishop,J.M. (1987) Correct integration of retroviral DNA *in vitro*. *Cell*, **49**, 347–356.

Brown,P.O., Bowerman,B., Varmus,H.E. and Bishop,J.M. (1989) Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc. Natl Acad. Sci. USA*, **86**, 2525–2529.

Bujacz,G., Jaskolski,M., Alexandratos,J., Wlodawer,A., Merkel,G., Katz,R.A. and Skalka,A.M. (1995) High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J. Mol. Biol.*, **253**, 333–346.

Burke,C.J., Sanyal,G., Bruner,M.W., Ryan,J.A., LaFemina,R.L., Robbins,H.L., Zeft,A.S., Middaugh,C.R. and Cordingley,M.G. (1992) Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase. *J. Biol. Chem.*, **267**, 9639–9644.

Bushman,F.D., Fujiwara,T. and Craigie,R. (1990) Retroviral DNA integration directed by HIV integration protein *in vitro*. *Science*, **249**, 1555–1558.

Cai,M., Zheng,R., Caffrey,M., Craigie,R., Clore,G.M. and Gronenborn, A.M. (1997) Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nature Struct. Biol.*, **4**, 567–577.

Craigie,R., Fujiwara,T. and Bushman,F. (1990) The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*. *Cell*, **62**, 829–837.

Dyda,F., Hickman,A.B., Jenkins,T.M., Engelman,A., Craigie,R. and Davies,D.R. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, **266**, 1981–1986.

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. *Nature Struct. Biol.*, **2**, 807–810.

Ellison,V., Abrams,H., Roe,T., Lifson,J. and Brown,P. (1990) Human immunodeficiency virus integration in a cell-free system. *J. Virol.*, **64**, 2711–2715.

Engelman,A., Mizuuchi,K. and Craigie,R. (1991) HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell*, 67, 1211–1221.

Engelman,A., Hickman,A.B. and Craigie,R. (1994) The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *J. Virol.*, **68**, 5911–5917.

Farnet,C.M. and Bushman,F.D. (1997) HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of pre-integration complexes *in vitro*. *Cell*, **88**, 483–492.

Farnet,C.M. and Haseltine,W.A. (1990) Integration of human immunodeficiency virus type 1 DNA *in vitro*. *Proc. Natl Acad. Sci. USA*, **87**, 4164–4168.

Fujiwara,T. and Mizuuchi,K. (1988) Retroviral DNA integration: structure of an integration intermediate. *Cell*, **54**, 497–504.

Katz,R.A. and Skalka,A.M. (1994) The retroviral enzymes. *Annu. Rev. Biochem.*, **63**, 133–173.

Katz,R.A., Merkel,G., Kulkosky,J., Leis,J. and Skalka,A.M. (1990) The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*. *Cell*, **63**, 87–95.

Katzman,M., Katz,R.A., Skalka,A.M. and Leis,J. (1989) The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the *in vivo* sites of integration. *J. Virol.*, **63**, 5319–5327.

Lee,M.S. and Craigie,R. (1994) Protection of retroviral DNA from autointegration: involvement of a cellular factor. *Proc. Natl Acad. Sci. USA*, **91**, 9823–9827.

Lee,S.P., Xiao,J.M., Knutson,J.R., Lewis,M.S. and Han,M.K. (1997) Zn^{2+} promotes the self-association of human immunodeficiency virus type-1 integrase *in vitro*. *Biochemistry*, **36**, 173–180.

Lee,Y.M. and Coffin,J.M. (1990) Efficient autointegration of avian retrovirus DNA *in vitro*. *J. Virol.*, **64**, 5958–5965.

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*, **34**, 9826–9833.

Misra,T.K., Grandgenett,D.P. and Parsons,J.T. (1982) Avian retrovirus pp32 DNA-binding protein. I. Recognition of specific sequences on retrovirus DNA terminal repeats. *J. Virol.*, **44**, 330–343.

Mizuuchi,K. (1992a) Polynucleotidyl transfer reactions in transpositional DNA recombination. *J. Biol. Chem.*, **267**, 21273–21276.

Mizuuchi,K. (1992b) Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu. Rev. Biochem.*, **61**, 1011–1051.

Mizuuchi,K. (1997) Polynucleotidyl transfer reactions in site-specific DNA recombination. *Genes to Cells*, **2**, 1–12.

Mueller,P.R. and Wold,B. (1989) *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science*, **246**, 780–786.

Pruss,D., Reeves,R., Bushman,F.D. and Wolffe,A.P. (1994) The influence of DNA and nucleosome structure on integration events directed by HIV integrase. *J. Biol. Chem.*, **269**, 25031–25041.

Pryciak,P.M., Muller,H.P. and Varmus,H.E. (1992a) Simian virus 40 minichromosomes as targets for retroviral integration *in vivo*. *Proc. Natl Acad. Sci. USA*, **89**, 9237–9241.

Pryciak,P.M., Sil,A. and Varmus,H.E. (1992b) Retroviral integration into minichromosomes *in vitro*. *EMBO J.*, **11**, 291–303.

Rice,P., Craigie,R. and Davies,D.R. (1996) Retroviral integrases and their cousins. *Curr. Opin. Struct. Biol.*, **6**, 76–83.

Roe,T., Chow,S.A. and Brown,P.O. (1997) 3′-End processing and kinetics of 5′-end joining during retroviral integration *in vivo*. *J. Virol.*, **71**, 1334–1340.

Savilahti,H., Rice,P.A. and Mizuuchi,K. (1995) The phage Mu transpososome core: DNA requirements for assembly and function. *EMBO J.*, **14**, 4893–4903.

Schwartzberg,P., Colicelli,J. and Goff,S.P. (1984) Construction and analysis of deletion mutations in the *pol* gene of Moloney murine leukemia virus: a new viral function required for productive infection. *Cell*, **37**, 1043–1052.

Sherman,P.A. and Fyfe,J.A. (1990) Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. *Proc. Natl Acad. Sci. USA*, **87**, 5119–5123.

Tanese,N., Roth,M.J. and Goff,S.P. (1986) Analysis of retroviral *pol* gene products with antisera raised against fusion proteins produced in *Escherichia coli*. *J. Virol.*, **59**, 328–340.

Vink,C., Oude Groeneger,A.M. and Plasterk,R.H. (1993) Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. *Nucleic Acids Res.*, **21**, 1419–1425.

Vora,A.C. and Grandgenett,D.P. (1995) Assembly and catalytic properties of retrovirus integrase–DNA complexes capable of efficiently performing concerted integration. *J. Virol.*, **69**, 7483–7488.

Vora,A.C., McCord,M., Fitzgerald,M.L., Inman,R.B. and Grandgenett, D.P. (1994) Efficient concerted integration of retrovirus-like DNA *in vitro* by avian myeloblastosis virus integrase. *Nucleic Acids Res.*, **22**, 4454–4461.

Wang,X. and Higgins,N.P. (1994) 'Muprints' of the lac operon demonstrate physiological control over the randomness of *in vivo* transposition. *Mol. Microbiol.*, **12**, 665–677.

Woerner,A.M. and Marcus-Sekura,C.J. (1993) Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis. *Nucleic Acids Res.*, **21**, 3507–3511.

Zheng,R., Jenkins,T.M. and Craigie,R. (1996) Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc. Natl Acad. Sci. USA*, **93**, 13659–13664.

Received on July 25, 1997; revised on September 9, 1997