

Characterization of the long-terminal repeat single-strand tail-binding site of Moloney-MuLV integrase by crosslinking

JORGE VERA¹, BEATRIZ VALENZUELA¹, MÓNICA J ROTH² and ÓSCAR LEÓN^{1*}

¹ Programa de Virología, Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile, Independencia 1027, Santiago

² Department of Biochemistry, University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, New Jersey 08854 USA

ABSTRACT

Processing of viral DNA by retroviral integrase leaves a dinucleotide single-strand overhang in the unprocessed strand. Previous studies have stressed the importance of the 5' single-stranded (ss) tail in the integration process. To characterize the ss-tail binding site on M-MuLV integrase, we carried out crosslinking studies utilizing a disintegration substrate that mimics the covalent intermediate formed during integration. This substrate carried reactive groups at the 5' ss tail. A bromoacetyl derivative with a side chain of 6 Å was crosslinked to the mutant IN 106-404, which lacks the N-terminal domain, yielding a crosslinked complex of 50 kDa. Treatment of IN 106-404 with N-ethylmaleimide (NEM) prevented crosslinking, suggesting that Cys209 was involved in the reaction. The reactivity of Cys209 was confirmed by crosslinking of a more specific derivative carrying maleimide groups that spans 8Å approximately. In contrast, WT IN was not reactive, suggesting that the N-terminal domain modifies the reactivity of the Cys209 or the positioning of the crosslinker side chain. A similar oligonucleotide-carrying iodouridine at the 5' ss tail reacted with both IN 106-404 and WT IN upon UV irradiation. This reaction was also prevented by NEM, suggesting that the ss-tail positions near a peptide region that includes Cys209.

Key terms: integrase, retrovirus, crosslinking.

INTRODUCTION

Integration of the reverse transcribed retroviral genome into the host chromosome is an essential step in the life cycle of retroviruses (Asante-Appiah and Skalka, 1999; Hindmarsh and Leis, 1999). This step is catalyzed by the viral-encoded integrase (IN) and requires the DNA sequences present in the long-terminal repeats (LTR) of the viral DNA (reviewed by Craigie, R. 2001). After reverse transcription of the viral RNA genome in the cytoplasm by reverse transcriptase, the linear DNA is cleaved at the 3' end of each strand, releasing two nucleotides and creating a two nucleotide 5' overhang. The 5' single-strand overhangs (5' ss-tail) on the LTRs have been implicated in

the stabilization of the IN-LTR complexes *in vitro* (Ellison and Brown, 1994; Vink et al., 1994). This end-processing exposes the conserved CA sequence at the 3' ends, found among all retroviruses and retrotransposons. The processed viral DNA migrates to the nucleus as a protein-DNA complex where it is integrated into the chromosome by an isoenergetic, staggered transesterification of the LTR termini into the target DNA (strand transfer) (Engelman et al., 1991). Strand transfer in Moloney murine leukemia virus (M-MuLV) IN joins each viral DNA end to sites on opposite strands that are separated by four bases, in a coordinated event, resulting in a gapped intermediate that generates short direct repeats flanking the provirus upon repair.

* Corresponding Author: Oscar Leon, Programa de Virología, ICBM, Facultad de Medicina, Universidad de Chile, Independencia 1027, Independencia, Santiago, Chile

Both end-processing and strand-transfer reactions have been reconstituted *in vitro* utilizing synthetic oligonucleotides containing the LTR termini (Craigie et al., 1990; Katzman et al., 1989; Sherman and Fyfe, 1990) and recombinant integrases. Assays for the concerted two-end integration have also been developed (Hindmarsh et al., 1999; Yang et al., 1999). The retroviral IN is also able to catalyze the disintegration of an intermediate containing LTR and target sequences (Chow et al., 1992). The integration and disintegration activity have been characterized for the M-MuLV IN proteins produced in bacteria as a GST fusion (Chow and Brown, 1994), renatured (Jonsson et al., 1993), and soluble forms (Villanueva et al., 2003).

Comparison of the amino acid sequences between different retroviral species show two regions of high similarity: a zinc-binding motif or HHCC region in the amino terminus and a central core region containing the DD(35)E motif. A third region located at the C-terminus is the least conserved among retroviruses. Mutational analysis of the human immunodeficiency virus (HIV-1) (reviewed by Chiu and Davies, 2004), avian retrovirus (Bushman and Wang, 1994; Katz et al., 1990; Khan et al., 1991; Kulkosky et al., 1992), and M-MuLV (Jonsson et al., 1996) indicate that all three regions are required for *in vitro* and *in vivo* activities. The N-terminal domain (HHCC) that coordinates a Zn²⁺ cation has been expressed as a separate domain. The functional role of this domain in M-MuLV IN has been delineated by chemical modification and complementation analysis (Jonsson et al., 1996). This domain can complement an N-terminal-deleted mutant (IN 106-404) in 3' processing, strand transfer and concerted two-end integration (Yang and Roth, 2001). The N-terminal domain was also required for the coordinated disintegration reactions catalyzed by the deletion mutant IN 106-404 on substrates lacking the 5' overhangs of the LTRs (ss-tail) (Donzella et al., 1996). Chemical modification of the HHCC domain by NEM impaired complementation of coordinated disintegration of the untailed substrates. Similarly, disintegration of untailed dumbbell substrate by the N-

terminal-deleted mutant also required the addition of the HHCC domain.

The central core of IN contains a triad of acidic amino acids (DD(35)E) that is required for activity both *in vitro* and *in vivo* (Drelich et al., 1992; Engelman and Craigie, 1992; Kulkosky et al., 1992; Leavitt, 1993; Vera et al., 2005). Crosslinking and mutagenesis studies indicate that nucleotides at the end of the LTR interact with amino acid residues close to the putative active site (Jenkins et al., 1997; Esposito and Craigie, 1998). Crosslinking studies of the C-terminal domain have identified peptides of this domain in close contact with substrates containing LTR sequences (Heuer and Brown 1997, 1998; Lutzke and Plasterk, 1998; Esposito and Craigie, 1998; Gao et al., 2001).

To understand the architecture of the catalytic protein-DNA complex involved in concerted integration, several models for HIV-1 IN have been proposed (reviewed by Karki et al., 2004; Wielens et al., 2005; Chen et al., 2006), based on the tetramer representing the minimal oligomer to carry out the concerted two-end integration (Faure et al., 2005) and the distance between the active sites consistent with a five-base-pair separation of the cleavage sites in the target DNA. However, several differences are noticed in the organization of the N- and C-terminal domains. For example, in one of the models (Wielens et al., 2005), dimerization of either the N-terminal domains or C-terminal domain is not observed, whereas in another model (Podtelezhnikov et al., 2003), the N-terminal domain dimerizes in one pair of subunits. The latest model (Chen et al., 2006) shows a more symmetric organization. In this case, the N-terminal domains of all subunits dimerize. This model predicts the existence of two perpendicular grooves able to accommodate target and viral DNAs. It has been suggested that the flexible loop 140-150 stabilizes the integration complex and regulates the appropriate positioning of the 3'OH during strand transfer (Wielens et al., 2005). For HIV-1 IN, the mutation of Gly 140 to Ala decreases the disintegration activity with an increase in the rigidity of

the loop (Greenwald et al., 1999). Diketo acids such as 1-(5-chloroindol-3-yl)-3-(tetrazoyl)-1,3-propanedione-ene (5CITEP) that inhibit strand transfer bind near the flexible loop (Goldgur et al., 1999; Johnson et al., 2006). The HIV-1 IN mutant G140S was less sensitive to inhibition by diketo acids in agreement with that hypothesis (King et al., 2003). Alignment of the amino acid sequences of HIV-1, ASV, and M-MuLV INs showed high similarities in the core region that includes the flexible loop (Johnson et al., 1986).

In order to identify amino acid that are near the ss-tail of the M-MuLV LTRs, we carried out crosslinking studies, utilizing oligonucleotide derivatives carrying reactive groups at the 5' terminus. In the experiments described here, we used disintegration substrate (dumbbell) carrying reactive groups directed to cysteine. We found that in the presence of Mn²⁺, a mutant lacking the N-terminal domain produced a crosslinked complex of 50,000 MW consistent with the size of the protein (35 kDa) and the oligonucleotide (50 nucleotides long). In contrast, under similar conditions, WT IN was not reactive. Further characterization of the crosslinking reaction indicated that the core cysteine (Cys209) was the target of cysteine directed crosslinkers used in this work.

MATERIALS AND METHODS

Oligonucleotides, plasmids and bacterial strains

DNA oligonucleotides were prepared on an Applied Biosystems Model 380B DNA Synthesizer by the UMDNJ Biochemistry Department DNA Synthesis Facility. Oligonucleotides were purified by electrophoresis through 20% polyacrylamide denaturing gels, eluted from gel slices in 500 mM ammonium acetate, 10 mM magnesium acetate at 37°C overnight and ethanol precipitated. Oligonucleotide 6015, 5'-CATGAAAGCGTAAGCTTTCAACCTGCGTAAGCAGGTAGACCGTAAGGTCT was used for chemical crosslinking. Iodouracil (X) containing oligonucleotides

were used in photocrosslinking 8161, XXTGAAAGCGTAAGCTTTCAACCTGCGTAAGCAGGTAGACCGTAAGGTCT and 2899, XATGAAAGCGTAAGCTTTCAACCTGCGTAAGCAGGTAGACCGTAAGGTCT; both of these oligonucleotides lack one nucleotide at the 3' end to avoid disintegration. Other oligonucleotides used in this work were: 4166, 5-AATGAAAGTCTTTCACGCTGTCCTTGGAC; 4167, 59-AATGAAAGTCTTTTCAAGCGAGTCCTGGAC; 5467, 59-AATGAAAGTCTTTCACGCT; 5527, 5-TGAAAGTCTTTTCAAGCT; 4985, 59-CGCTTACCTGTTTACAGTA.

WT IN and IN 106-404 carrying a hexahistidine tag at the C-terminus were purified by expression of previously reported plasmids in *E. coli* BL21(DE3) cells (Jonsson et al., 1993).

Modification of the 5' terminal cytidine of dumbbell 6015

The attachment of an aminoethyl side chain to a cytidine residue in the 5' single-strand tail of oligo 6015 was performed as previously described (Schulman et al., 1981). Following gel purification and annealing (Donzella et al., 1998), 5 nmole of the oligonucleotide were dissolved in 50 µl sterile H₂O. Then, 200 µl of an aqueous solution containing 3 M NaHSO₃ and 1.5 M ethylenediamine pH 7.0 was added and incubated at 37°C for 70 h. At the end of this reaction, the oligonucleotides were separated from reactants by centrifugation through a Sephadex G-25 spin column. 1 M Tris pH 9.2 was then added to a 0.1 M final concentration and incubated at 37° for 8 h, followed by ethanol precipitation in 2 M ammonium acetate. All crosslinking substrates were 5'-end ³²P-labeled as previously described (Jonsson and Roth, 1993), with the exception that removal of the unincorporated [γ ³²P]-ATP was performed in a Sephadex G-25 spin column equilibrated in 0.2 M HEPES pH 7.8, to allow for efficient coupling of the crosslinking reagents. To determine the extent of the modification, 1 pmol of the ³²P-labeled aminoethylated dumbbell (AE6015) was heated at 95°C for 5 min,

cooled on ice, and subjected to hydrolysis with P1 nuclease (1 U), in 20 mM sodium acetate, pH 5.3 for 16 h at 37°C. A second aliquot of nuclease was added after heating the digest at 70°C to ensure complete hydrolysis. The resulting nucleotides monophosphates were analyzed by TLC on PEI cellulose plates, developed with 2 M LiCl and exposed to X-ray films.

Coupling of Bromoacetyl N-hydroxysuccinimide (BrAcNHS) or Maleimido Propionyl N-hydroxysuccinimide to the modified cytidine

20 pmol of AE 6015 oligonucleotide in 20 µl 0.2 M HEPES pH 7.8 were added to 20 µl of 6 mg/ml BrAcNHS dissolved in DMSO. The mixture was incubated at 25°C for 15 min. The BrAcNHS coupled oligonucleotide was ethanol precipitated, washed extensively with 80% ethanol and dried at room temperature. To couple maleimido acetyl N-hydroxysuccinimide to the modified cytidine, the procedure above was followed with the exception that the reaction was protected from light.

Disintegration Assays

All disintegration reactions were performed in 15 µl volume containing 20 mM piperazine-N,N'-bis(2-ethanesulfonic acid) (PIPES) pH 6.4, 10 mM DTT, 25 mM MnCl₂, 10 mM 3-[(3-cholamido-propyl)-dimethylammonio]-1-propane sulfonate (CHAPS), and 0.05% Nonidet P-40, at 30°C for 1 hr, typically with 0.5-1.0 pmol of dumbbell substrate and 5 to 20 pmol of M-MuLV IN.

Chemical crosslinking

Crosslinking reactions were performed under the same conditions for disintegration, with the exception that DTT was omitted from the reaction buffer to lower its concentration. Reactions were terminated by the addition of 10 µl stop buffer (95% formamide, 1 mM EDTA), heated to 95°C for 3 to 5 min, and then run on 20% sequencing gels. Dried gels were exposed to Kodak X-OMAT X-ray film.

RESULTS

The productive interaction of M-MuLV IN with a unimolecular dumbbell disintegration substrate was investigated through chemical crosslinking studies. A dumbbell substrate containing the sequence 5' CA at the LTR ss-tail instead of normal 5' AA was synthesized to attach a crosslinker (6015, Fig. 1A). M-MuLV IN maintained high levels of disintegration on substrates bearing this mutation, efficiently releasing the 15-mer LTR product (Fig. 1B, lane 2). Ethylenediamine was coupled to the terminal cytidine of the LTR 5' ss-tail to provide reactive amino groups for the introduction of bifunctional crosslinking reagents, directed toward different amino acid side chains. Thus, the relative proximity of the LTR 5' ss-tail to specific residues or regions of M-MuLV IN could be delineated.

Synthesis of the crosslinking derivatives

The synthesis of the probes for crosslinking was carried out in two steps. In the first step, the amino group of the single-stranded cytidine, at the 5' end of the annealed dumbbell substrate (6015), was replaced by ethylenediamine by a transamination reaction catalyzed by sodium bisulfite (Schulman et al., 1981). TLC analysis of the P1 nuclease hydrolysis products showed that under the conditions of the modification, more than 80% of the label migrated similarly to modified [³²P]-dCMP (not shown). As expected, a minor spot migrates in the position of [³²P]-UMP, a side product of the modification with bisulfite, but no radioactivity was detected at the position of dCMP. Modification of the terminal cytidine did not affect disintegration, as shown in Figure 1B (lane 4). The change in mobility of the substrate, readily visible in the released AE 15-mer LTR product (Fig. 1B, lane 4), can be attributed to the introduction of a positive charge by ethylenediamine. In a second step, bifunctional crosslinkers containing N-hydroxysuccinimide ester groups were coupled to the reactive amino group of the modified cytosine (Fig. 1C).

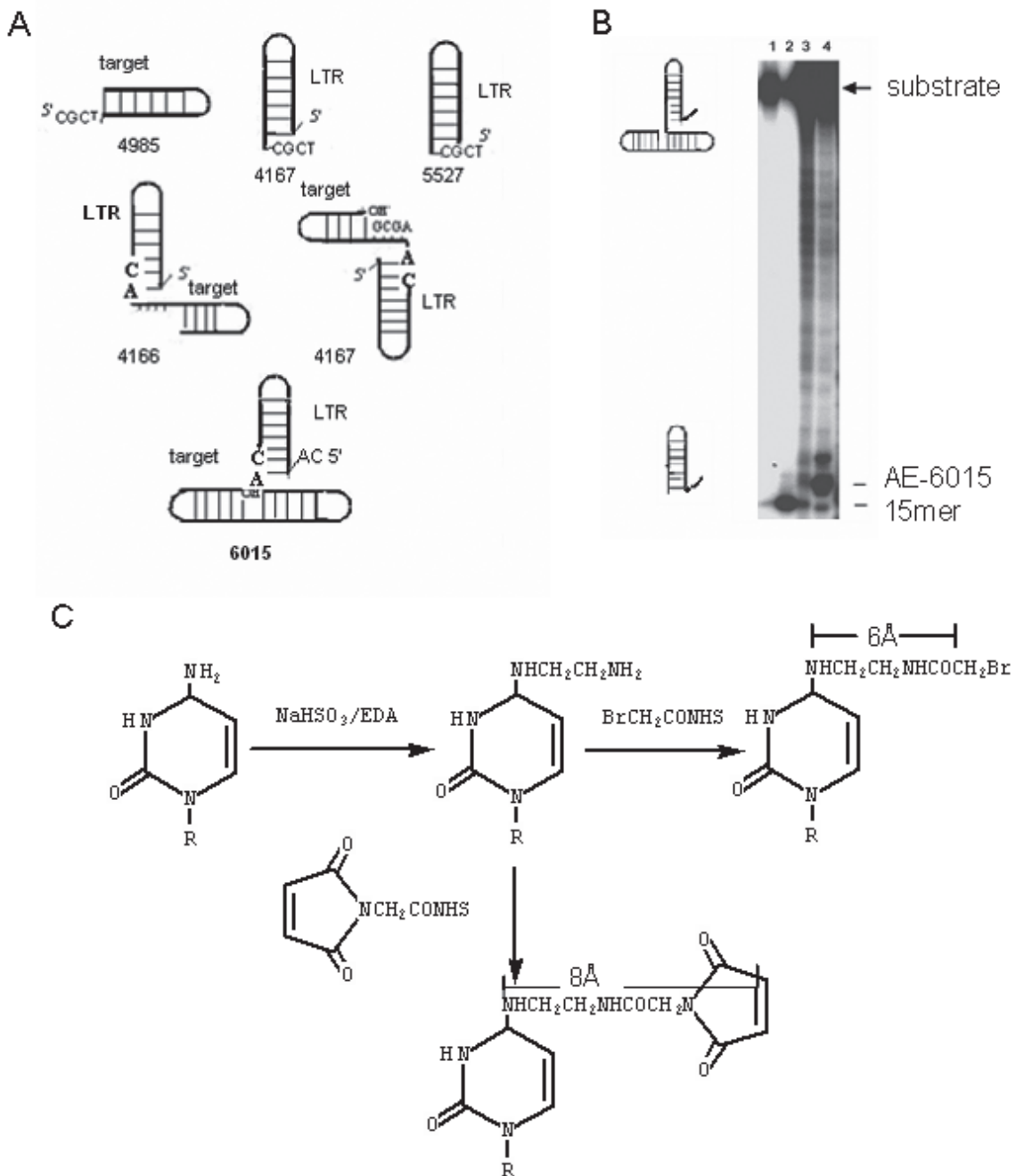


Figure 1: Panel A. Oligonucleotides used in this work.

Panel B. Disintegration activity of WT IN on the dumbbell substrate (6015) modified with ethyldiamine/bisulfite. The assay was done as described in “Materials and Methods” using 1 pmole of the [³²P]-labeled oligonucleotide and 15 pmoles of WT IN. Lanes 1 and 2: unmodified substrate. Lanes 3 and 4: modified substrate. Lanes 1 and 3: no enzyme. The position of the substrates and products of the reaction are indicated.

Panel C. Scheme of the reactions for the coupling of N-hydroxysuccinimide esters to the 5' terminal cytidine of the oligonucleotide 6015.

Chemical crosslinking

Bromoacetyl groups in the modified oligonucleotide are capable of crosslinking to proteins mainly with appropriately oriented cysteine SH groups and histidine, although reactions with other amino acids have also been reported (Hartman and Brown, 1976). This crosslinker is expected to react at a distance of 6Å from the base when it is extended. Incubation of the [³²P]-labeled-bromoacetyl dumbbell with a mutant lacking the HHCC domain, IN 106-404 in the presence of Mn²⁺, resulted in a rapid formation of a covalent complex of protein and nucleic acid, as determined by SDS-PAGE (Fig. 2). The reaction was over after 5 min (lane 3). The size of a predominant band (50 kDa) matches the expected molecular mass of a 1: 1 protein-nucleic acid complex. Under these conditions, WT IN did not react (see results below).

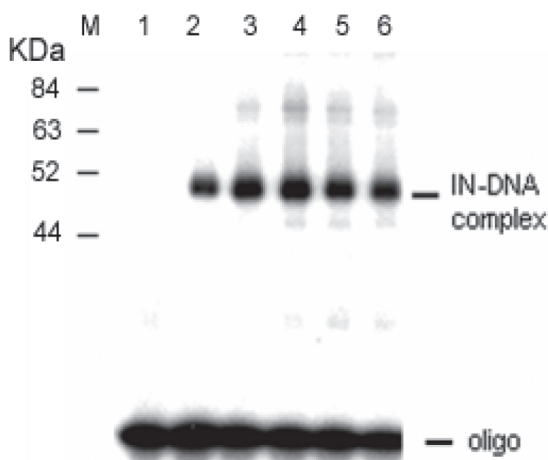


Figure 2: Time course of the crosslinking reaction between BrAc6015 and IN 106-404. 5'-[³²P]-BrAc-6015 (0.1 mM) and IN 106-404 (1.0 mM) in 30 ml final volume were incubated at 25°C under the conditions described in Materials and Methods. Five microliters were removed at the desired times and analyzed by electrophoresis on 12% polyacrylamide-SDS gels: Lane M: Prestained molecular mass markers; 84, 63, 52, and 44 kDa. Lane 1: no enzyme, lane 2: 1 min, lane 3: 5 min, lane 4: 10 min, lane 5: 20 min, lane 6: 30 min.

The specificity of the crosslinking reaction was examined by determining the effect of several unmodified oligonucleotides that are shown in Figure 1A. As it is shown in Figure 3, the crosslinking reaction was competed by unmodified dumbbell and oligonucleotides containing both a viral LTR and a target sequence (4166 and 4167). Oligonucleotides containing the tailed (5467) or untailed (5527) LTR or the target sequence (4985) had no significant effect on the crosslinking reaction. These results indicate that simultaneous binding of the LTR and target sequences are required to displace the dumbbell substrate.

In order to eliminate the possibility of non-specific reaction of the bromoacetyl groups, crosslinking was performed in the presence of lysine-quenched bromoacetyl NHS. The addition of lysine-quenched bromoacetyl NHS at concentrations up to 50 times higher than the crosslinking probe was unable to prevent crosslinking, indicating that the reaction is dependent on DNA binding (not shown).

The rapid kinetics of the crosslinking reaction suggested that the SH group of the core Cys209 in IN 106-404 could be the target. To address this possibility, IN 106-404 was chemically modified with N-ethylmaleimide (NEM) prior to crosslinking. Preincubation of IN 106-404 with 10 mM NEM resulted in a dramatic loss of the crosslinking (Fig. 4A), confirming our prediction. The low level of reaction of NEM-treated IN 106-404 is most likely resultant of crosslinking to unreacted cysteines, although we cannot rule out the participation of other amino acid side chains since bromoacetyl groups are not specific for cysteine. Maleimide is known to react specifically with SH groups, therefore a derivative containing maleimide groups at the 5' terminal end of the dumbbell oligonucleotide was then used in crosslinking. Figure 4B shows that this derivative reacts with IN 106-404, yielding a covalent adduct of the expected size for a 1: 1 enzyme: DNA complex (lane 3). Since Cys209 is the only cysteine present in IN 106-404, we conclude that this residue is the target in the reaction with the NEM moiety.

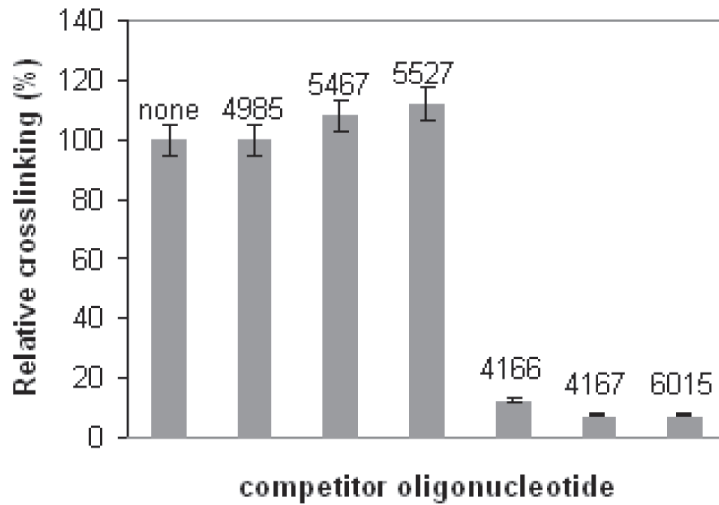


Figure 3: Competition of crosslinking between BrAc-6015 and IN 106-404 by DNA oligonucleotides containing target and LTR sequences. These experiments were carried out as described in Fig. 2, except that unlabeled oligonucleotides at 1 mM were included in the reaction. The oligonucleotides used are described in Fig. 1A and are indicated by number. After 30 min, the products of the reaction were separated by electrophoresis on 12% acrylamide gels with SDS. The relative amounts of the protein-DNA complex was determined in a phosphorimager.

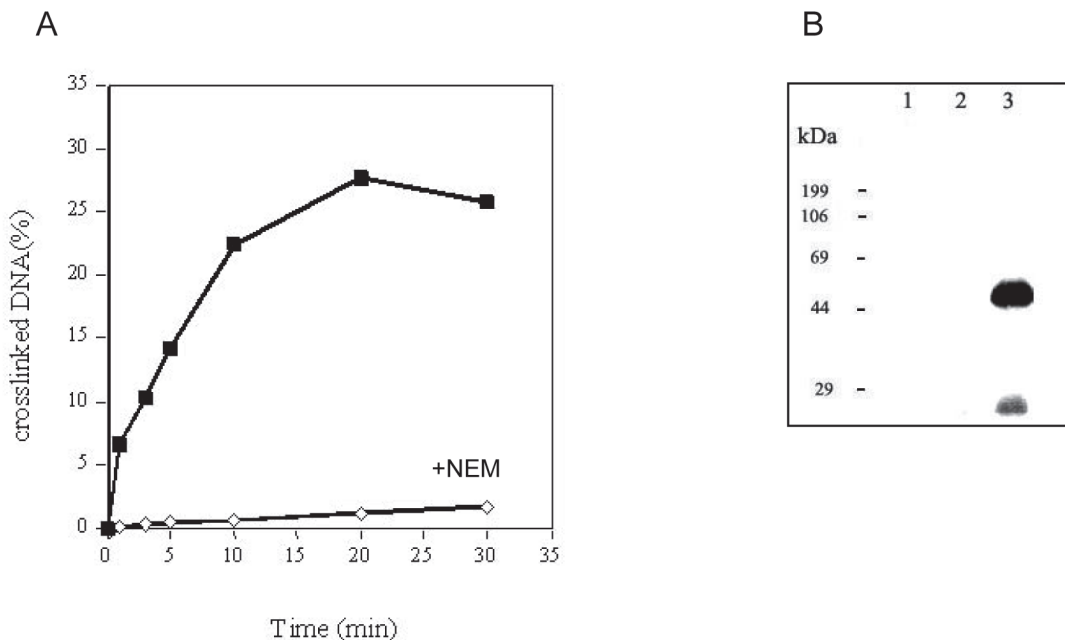


Figure 4: Panel A. Effect of NEM on the crosslinking of BrAc6015 to IN 106-404. Crosslinking was carried out as described in Fig. 2. For NEM treatment, IN 106-404 was preincubated with 20 mM NEM for 30 min on ice in the disintegration reaction buffer before adding the reactive oligonucleotide. The products of the reaction were separated by electrophoresis on 12% acrylamide gels with SDS and quantified in a phosphoimager.

Panel B. Crosslinking of NEM-Ac6015 to WT IN and IN 106-404. The proteins and NEM-Ac6015 were incubated in the disintegration conditions. Lanes 1: no enzyme. Lane 2: WT IN. Lane 3: IN 106-404. Positions of the molecular mass standards are marked on the left.

In contrast, no crosslinked product was observed in the presence of WT IN (lane 2), indicating that this cysteine is not available for reaction in the full-length enzyme. One explanation for this result is that the crosslinker side chain is sterically blocked and cannot assume the appropriate orientation to target the cysteine.

Photocrosslinking

To address the question of steric interference of the maleimide-modified 5' ss-tail substrate with the WT IN, photocrosslinking using dumbbell oligonucleotides containing iodo U (IdU) at the 5' terminus was utilized. IdU reacts upon UV light irradiation with several amino acids at a short distance (Meisenheimer and Koch et al., 1997). The results of this experiment are shown in Figure 5A. A covalent complex IN-DNA that migrates with a molecular mass near 62 kDa was observed (lane 5). This size corresponds to a 1: 1 IN: DNA complex. Treatment of the reaction products with nuclease P1 that cleaves DNA to 5'-nucleosides monophosphates decreases the size of the complex to approximately 45 kDa (lane 6). A crosslinking reaction with a dumbbell oligonucleotide without IdU was also run as a control. A low amount of crosslinking was observed (lane 1); however, after treatment with nuclease, the radioactivity associated to the complex disappeared (lane 2). These results indicate that crosslinking is specifically due to the presence of the iodouridine at the 5' ss tail. Similar results were observed when we used IN 106-404. A covalent adduct of an approximate size of 50 kDa (lane 7) is observed that decreased to near 35 kDa upon treatment with nuclease P1 (lane 8). Again, the crosslinking of the unmodified dumbbell did not result in a covalent adduct upon treatment with nuclease P1 (lane 4), indicating that the crosslinking observed (lane 3) does not occur at the 5' ss tail of the dumbbell. These results show that the 5' ss tail binds to both WT IN and IN 106-404.

Since NEM blocked the crosslinking reaction of IN 106-404 with the cysteine

directed derivatives, we examined the effect of NEM on photocrosslinking. In this experiment, WT IN was preincubated with NEM as described in materials and methods before photocrosslinking. Then the reaction mixture was treated with nuclease P1, and the products separated by gel electrophoresis with SDS. In Figure 5B, we can observe that when IN was treated with 20 mM NEM, the specific crosslinking to the 5' ss tail was prevented.

Cys209 is located in a region that shows high similarity with a flexible loop, seen in HIV-1 and ASV IN. Modification of Cys 209 of IN 106-404 by NEM has been shown to impair disintegration. From our crosslinking results, we can conclude that the 5' ss tail of the integration intermediate binds near the flexible loop of M-MuLV IN and that modification of Cys209 by NEM alters the positioning of the 5'-terminal nucleotide of the disintegration substrate, explaining the lack of activity by the modified enzyme.

DISCUSSION

In vitro studies using short oligonucleotides has allowed the characterization of IN activities, including 3' processing, strand transfer, and disintegration. The disintegration represents the reversal of integration and can be determined by utilizing a substrate (Y or dumbbell) that contains both the target and LTR sequences. This is a useful model to determine the domains of the protein involved in binding/recognition of elements, including the 3' CA and the 5' ss tail of the LTR and the target DNA. Binding of the CA to the catalytic core is supported by biochemical and crosslinking data (Gerton and Brown, 1997; Heuer and Brown 1997; 1998, Jenkins, et al., 1997; Kulkosky et al., 1995).

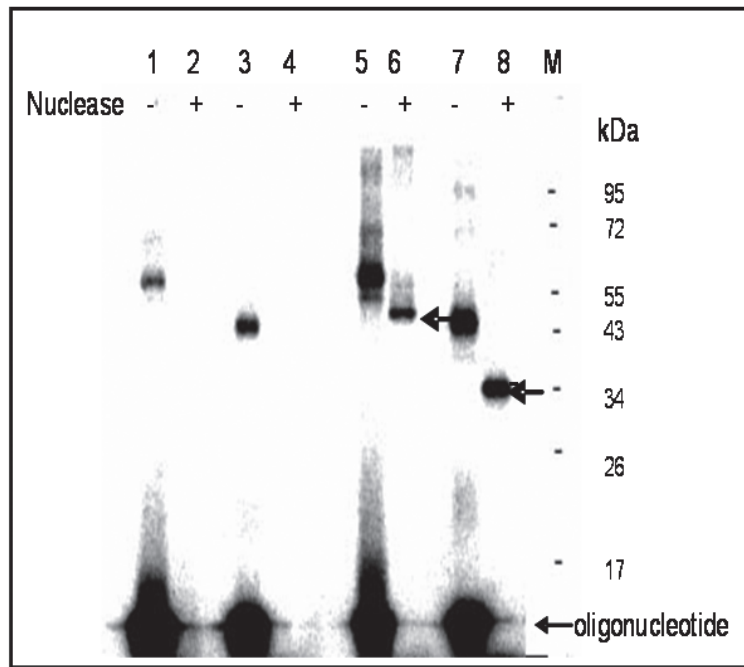
The importance of the ss tail in double- and single-coordinated disintegration has been established in IN mutants lacking the HHCC domain (Donzella et al., 1996, 1998). In the absence of the ss tail, IN 106-404 was unable to catalyze these reactions. However, double and single disintegration

of untailed crossbones can be restored by adding the HHCC domain (IN 1-105), suggesting a functional relationship between the ss-tail and the HHCC domain (Donzella et al., 1993). Disintegration of the untailed dumbbell requires both the core and the HHCC domain. Modification of IN 106-404 by NEM blocks complementation by the unmodified HHCC domain (Donzella et al. 1998; Yang et al., 1999).

In this work, photo and chemical crosslinking studies were carried out to

determine the peptide regions near the ss DNA tail-binding site. In this approach, we synthesized an oligonucleotide carrying a non-specific crosslinker capable of reacting with several functional groups. This derivative reacted with IN 106-404 in a rapid fashion, and the crosslinking reaction was blocked by NEM, suggesting Cys209 as the main target. This conclusion was confirmed by using a specific crosslinker on the substrate directed towards cysteine, such as NEM.

A



B

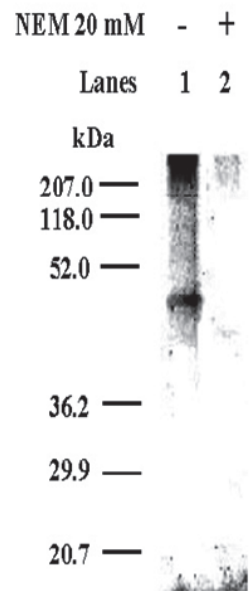


Figure 5: Panel A. Photocrosslinking of dumbbell oligonucleotides to WT IN and IN 106-404. 5'-[³²P] 8161 (0.1 mM) and IN (1.0 mM) in 30 ml of disintegration buffer were irradiated at 300 nm for 30 min. 15 ml samples were run on a 12% acrylamide gel with 0.1% SDS (lanes 1, 3, 5, and 7), 15 ml of the samples was treated with 2 units of nuclease P1 in buffer 20 mM sodium acetate pH 6.0, 1 mM zinc chloride for 16 h at 37°C (lanes 2, 4, 6, and 8). Lanes 1 to 4 show the crosslinking with the unmodified dumbbell substrate 6015 to wt IN (lanes 1 and 2) and to IN 106-404 (lanes 3 and 4). Lanes 5 to 8 show the crosslinking of the IdU modified oligonucleotide to WT IN (lanes 5 to 6) and IN 106-404 (lanes 7 and 8). Positions of the bands after nuclease digestion are indicated by arrows.

Panel B. Effect of NEM on crosslinking of WT IN with the IdU containing oligonucleotide. Photocrosslinking was carried out as described in panel A. To determine the effect of NEM, the enzyme was preincubated with 20 mM NEM on ice for 30 min and quenched with DTT, as described in Methods. Samples were digested with 2 units of nuclease P1 for 16 hrs and subjected to electrophoresis in a 12% polyacrylamide-SDS gel. Positions of the molecular mass standards are marked on the left.

A model for the structure of the region comprising residues P178 to S238 of M-MuLV IN (residues Val110 to Glu170 in HIV-1 IN, PDB, 1BL3) was generated using the program Swiss Model Prot (Fig. 6). Cys 209 is located in a region called the flexible loop (His 208-Ser 216). The role of this region in the catalytic activity of IN is not clear, however, it has been suggested that it is involved in the regulation of the strand transfer step during concerted integration (Wielens et al., 2005; Chen et al., 2006). Photocrosslinking studies between HIV-1 and an LTR substrate locate the 5' end of the non-processed strand close to the residues Tyr143 and Gln148 (Esposito and Craigie, 1998). Johnson et al. (2006) proposed that in HIV-1, a specific interaction takes place between the second C located at the 5' end of the unprocessed strand of the LTR and Gln 148. Chen et al., (2006) have proposed that the flexible loop could be involved in regulating binding of the target DNA. The results presented in this work indicate that in M-MuLV IN, the terminal base of the 5' ss tail localizes

within 6 Å of Cys 209, suggesting an interaction between the flexible loop and the 5' single-strand region. The lack of reactivity of Cys209 in WT IN to the cysteine directed crosslinkers could be explained by a modification of the environment of this residue in the presence of the HHCC domain.

Alternatively, the orientation of the crosslinker reactive groups in the assembled DNA-enzyme complex could be altered in the full-length IN. However, photocrosslinking occurs since DNA binding is not affected. Photocrosslinking studies reported in this work indicate that the 5' ss tail of the dumbbell oligonucleotide has access to the core and that modification of Cys 209 by NEM blocked photocrosslinking at the 5' ss tail. Whether this effect is due to a direct steric interference of the 5' ss-tail binding or to a conformational change of the flexible loop remains to be defined.

Collectively, the results presented hereby indicate that the flexible loop of the core domain of M-MuLV IN would be involved in positioning of the ss-DNA tail. These

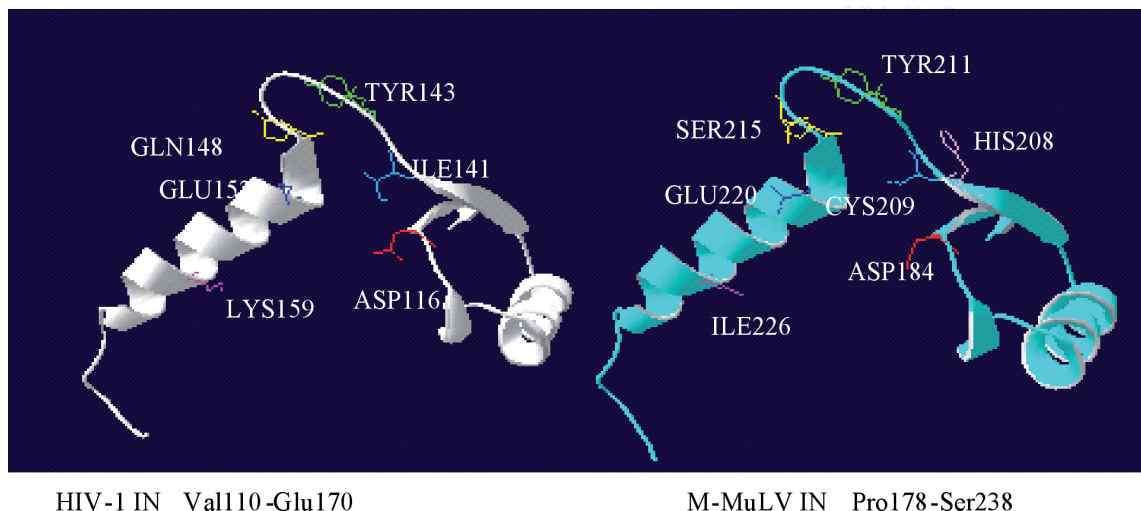


Figure 6: Ribbon structures of a predicted model of catalytic domain of M-MuLV IN (right), comprising residues P178 to S238, in comparison with the structure of the catalytic domain of HIV-1 IN (left), comprising residues V110 to E170, (PDB format: 1BL3), generated by Swiss Model Prot Program. Energy Minimization was done by the program MODELLER. The residues of the putative active site (D116 and E152 in HIV-1 IN and D184 and E220 in M-MuLV IN) and residues of the loop (I141, Y143, N148 in HIV-IN and H208, C209, Y211 and S215 in M-MuLV IN) are labeled to show the proximity and spatial relationship to each other and some more distant residues (K159 in HIV-1 IN and I226 in M-MuLV IN).

observations agree with the suggestion that this loop may have an important role in the separation of both LTRs through binding of the ss tail during concerted integration and stabilization of the preintegration complexes. Mutation of His 208 to Ala in M-MuLV IN yields a phenotype that is similar to the drug inhibition pattern seen in HIV-1 IN (King et al., 2003), since strand transfer and disintegration were completely abolished, although processing was reduced by 90% (O. Leon, unpublished results). This results suggest that 5'ss tail could bind to IN in different contexts throughout the catalysis involving conformational changes of the flexible loop. A more systematic study of the flexible loop in M-MuLV is currently underway in our laboratory.

ACKNOWLEDGEMENTS

This work was funded by grants Fondecyt 1040409 (O.L.) and National Institutes of Health RO1 GM070837 (M.J.R.)

REFERENCES

- ASANTE-APPIAH E, SKALKA AM (1999) HIV-1 integrase: Structural organization conformational changes and catalysis. *Adv Virus Res* 52: 351-69
- BUSHMAN FD, WANG B (1994) Rous sarcoma virus integrase protein: Mapping functions for catalysis and substrate binding. *J Virol* 68: 2215-23
- CHEN A, WEBER IT, HARRISON RW, LEIS J (2006) Identification of amino acids in HIV-1 and avian sarcoma virus integrase subsites required for specific recognition of the long terminal repeat ends. *J Biol Chem* 281: 4173-82
- CHIU TK, DAVIES DR (2004) Structure and function of HIV-1 integrase. *Curr Top Med Chem* 4: 965-77
- CHOW SA, BROWN PO (1994) Juxtaposition of two viral DNA ends in a bimolecular disintegration reaction mediated by multimers of human immunodeficiency virus type 1 or murine leukemia virus integrase. *J Virol* 68: 7869-78
- CHOW SA, VINCENT KA, ELLISON V, BROWN PO (1992) Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science* 255: 723-6
- CRAIGIE R (2001) HIV integrase, A brief overview from chemistry to therapeutics. *J Biol Chem* 276: 23213-6
- CRAIGIE R, FUJIWARA T, BUSHMAN F (1990) The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*. *Cell* 62: 829-37
- DONZELLA GA, JONSSON CB, ROTH MJ (1996) Coordinated disintegration reactions mediated by Moloney murine leukemia virus integrase. *J Virol* 70: 3909-21
- DONZELLA GA, JONSSON CB, ROTH MJ (1993) Influence of substrate structure on disintegration activity of Moloney murine leukemia virus integrase. *J Virol* 67: 7077-87
- DONZELLA GA, LEÓN O, ROTH MJ (1998) Implication of a central cysteine residue and the HHCC domain of Moloney murine leukemia virus integrase protein in functional multimerization. *J Virol* 72: 1691-8
- DRELICH M, WILHELM R, MOUS J (1992) Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein *in vitro*. *Virology* 188: 459-68
- DYDA F, HICKMAN AB, JENKINS TM, ENGELMAN A, CRAIGIE R, DAVIES DR (1994) Crystal structure of the catalytic domain of HIV-1 integrase: Similarity to other polynucleotidyl transferases. *Science* 266: 1981-6
- ELLISON V, BROWN PO (1994) A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration *in vitro*. *Proc Natl Acad Sci U S A* 91: 7316-20
- ENGELMAN A, CRAIGIE R (1992) Identification of conserved amino acid residues critical for human immunodeficiency virus type-1 integrase function *in vitro*. *J Virol* 66: 6361-9
- ENGELMAN A, HICKMAN AB, 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: 5911-7
- ENGELMAN A, MIZUUCHI K, CRAIGIE R (1991) HIV-1 DNA integration: Mechanism of viral DNA cleavage and DNA strand transfer. *Cell* 67: 1211-21
- ESPOSITO D, CRAIGIE R (1998) Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J* 17: 5832-43
- FAURE A, CALMELS C, DESJOBERT C, CASTROVIEJO M, CAUMONT-SARCOS A, TARRAGO-LITVAK L, LITVAK S, PARISSI V (2005) HIV-1 integrase crosslinked oligomers are active *in vitro*. *Nucleic Acids Res* 33: 977-86
- GAO K, BUTLER S L, AND BUSHMAN F (2001) Human immunodeficiency virus type-1 integrase: Arrangement of protein domains in active cDNA complexes. *EMBO J* 20: 3565-76
- GERTON JL, BROWN PO (1997) The core domain of HIV-1 integrase recognizes key features of its DNA substrates. *J Biol Chem* 272: 25809-15
- GOLDGUR Y, CRAIGIE R, COHEN GH, FUJIWARA T, YOSHINAGA T, FUJISHITA T, SUGIMOTO H, ENDO T, MURAI H, DAVIES DR (1999) Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: A platform for antiviral drug design. *Proc Natl Acad Sci USA* 96: 13040-3
- GREENWALD J, BUTLER SL, BUSHMAN FD, CHOE S (1999) The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity. *Biochemistry* 38: 8892-8
- HARTMAN FC, BROWN J P (1976) Affinity labeling of a previously undetected essential lysyl residue in class-I fructose bisphosphate aldolase. *J Biol Chem* 251: 3057-62
- HEUER TS, BROWN PO (1997) Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photocrosslinking. *Biochemistry* 36: 10655-65
- HEUER TS, BROWN PO (1998) Photocrosslinking studies suggest a model for the architecture of an active human immunodeficiency virus type-1 integrase DNA complex. *Biochemistry* 37: 6667-78

- HINDMARSH P, LEIS J (1999) Retroviral DNA integration. *Microbiol Mol Biol Rev* 63: 836-43
- HINDMARSH P, RIDKY T, REEVES R, ANDRAKE M, SKALKA AM, LEIS J (1999) HMG protein family members stimulate human immunodeficiency virus type-1 and avian sarcoma virus concerted DNA integration *in vitro*. *J Virol* 73: 2994-3003
- JENKINS TM, ESPOSITO D, ENGELMAN A, CRAIGIE R (1997) Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photocrosslinking. *EMBO J* 16: 6849-59
- JOHNSON AA, SANTOS W, PAIS GC, MARCHAND C, AMIN R, BURKE TR, JR, VERDINE G, POMMIER Y (2006) Integration requires a specific interaction of the donor DNA terminal 5'-cytosine with glutamine 148 of the HIV-1 integrase flexible loop. *J Biol Chem* 281: 461-7
- JOHNSON MS, MCCLURE MA, FENG DF, GRAY J, DOOLITTLE RF (1986) Computer analysis of retroviral pol genes: Assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. *Proc Natl Acad Sci USA* 83: 7648-52
- JONSSON C B, DONZELLA G A, GAUCAN E, SMITH CM, ROTH MJ (1996) Functional domains of Moloney murine leukemia virus integrase defined by mutation and complementation analysis. *J Virol* 70: 4585-97
- JONSSON CB, DONZELLA GA, ROTH MJ (1993) Characterization of the forward and reverse integration reactions of the Moloney murine leukemia virus integrase protein purified from *Escherichia coli*. *J Biol Chem* 268: 1462-9
- JONSSON CB, ROTH MJ (1993) Role of the His-Cys finger of Moloney murine leukemia virus integrase protein in integration and disintegration. *J Virol* 67: 5562-71
- KARKI R, TANG GY, BURKE TR, JR, NICKLAUS MC (2004) Model of full-length HIV-1 integrase complexed with viral DNA as template for anti-HIV drug design. *J Comput Aided Mol Des* 18: 739-60
- KATZ RA, MERKEL G, KULKOSKY J, LEIS J, SKALKA AM (1990) The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*. *Cell* 63: 87-95
- KATZMAN M, KATZ RA, SKALKA AM, 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: 5319-27
- KHAN E, MACK JP, KATZ RA, KULKOSKY J, SKALKA AM (1991) Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res* 19: 851-60
- KING PJ, LEE DJ, REINKE RA, VICTORIA JG, BEALE K, ROBINSON WE, JR, (2003) Human immunodeficiency virus type-1 integrase containing a glycine to serine mutation at position 140 is attenuated for catalysis and resistant to integrase inhibitors. *Virology* 306: 147-61
- KULKOSKY J, JONES KS, KATZ RA, MACK JP, SKALKA AM (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol Cell Biol* 12: 2331-8
- KULKOSKY J, KATZ RA, MERKEL G, SKALKA AM (1995) Activities and substrate specificity of the evolutionarily conserved central domain of retroviral integrase. *Virology* 206: 448-56
- LEAVITT A, SHIUE DL, VARMUS HE (1993) Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions *in vitro*. *J Biol Chem* 268: 2113-9
- LUTZKE RA, PLASTERK R H (1998) Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type-1 integrase: Critical residues for protein oligomerization and DNA binding. *J Virol* 72: 4841-8
- MEISENHEIMER KM, KOCH TH (1997) Photocrosslinking of nucleic acids to associated proteins. *Crit Rev Biochem Mol Biol* 32: 101-40
- PODTELEZHNIKOV AA, GAO K, BUSHMAN FD, MCCAMMON JA (2003) Modeling HIV-1 integrase complexes based on their hydrodynamic properties. *Biopolymers* 68: 110-20
- SCHULMAN LH, VALENZUELA D, PELKA H (1981) Reversible inactivation of *Escherichia coli* methionyl-tRNA synthetase by covalent attachment of formylmethionine tRNA to the tRNA binding site with a cleavable crosslinker. *Biochemistry* 20: 6018-23
- SHERMAN PA, FYFE, JA (1990) Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. *Proc Natl Acad Sci USA* 87: 5119-23
- VERA J, PARISSI V, GARCÍA A, ZÚÑIGA R, ANDREOLA ML, CAUMONT-SARCOS A, TARRAGO-LITVAK L, LEÓN O (2005) Yeast system as a model to study Moloney murine leukemia virus integrase: Expression mutagenesis and search for eukaryotic partners. *J Gen Virol* 86: 2481-8
- VILLANUEVA RA, JONSSON CB, JONES J, GEORGIADIS MM, ROTH MJ (2003) Differential multimerization of Moloney murine leukemia virus integrase purified under non-denaturing conditions. *Virology* 316: 146-160
- VINK C, LUTZKE RA, PLASTERK RH (1994) Formation of a stable complex between the human immunodeficiency virus integrase protein and viral DNA. *Nucleic Acids Res* 22: 4103-10
- WIELENS J, CROSBY IT, CHALMERS DK (2005) A three-dimensional model of the human immunodeficiency virus type 1 integration complex. *J Comput Aided Mol Des* 19: 301-17
- YANG F, LEÓN O, GREENFIELD NJ, ROTH MJ (1999) Functional interactions of the HHCC domain of moloney murine leukemia virus integrase revealed by nonoverlapping complementation and zinc-dependent dimerization. *J Virol* 73: 1809-17
- YANG F, ROTH MJ (2001) Assembly and catalysis of concerted two-end integration events by Moloney murine leukemia virus integrase. *J Virol* 75: 9561-70